Mechanisms of drug resistance in *T. brucei*; Beyond the P2 Transporter…

Daniel Bridges

2007

This thesis is submitted for the degree of Doctor of Philosophy

Division of Infection and Immunity

Faculty of Biomedical and Life Sciences

University of Glasgow

April 2007

Abstract

The principal aim of this project was to investigate mechanisms of drug resistance in *Trypanosoma brucei*, the causative agent of disease in humans (sleeping sickness) and livestock (nagana), which affects large areas of sub-Saharan Africa. While recent figures suggest that the number of infections in humans may have declined in the last 10 years, the emergence of resistance to some of the first line drugs threatens to undermine this progress, as no acceptable alternative treatments exist. By understanding the mechanisms of resistance, the useful life of current therapies (of which there are only a few) may be extended, diagnostics to identify resistant parasites could be developed and the design of novel therapies aided. We therefore developed parasites with high levels of resistance to the clinically important drug pentamidine, which is the first-line treatment for early stage West African sleeping sickness and is closely related to the main veterinary treatment diminazene aceturate (Berenil). The characterisation of this strain revealed that the resistance phenotype was at least in part due to the loss of the previously characterised high affinity pentamidine transporter (HAPT). To investigate the protein(s) responsible for HAPT activity, and to identify any other proteins contributing to the resistance phenotype, we employed a proteomic approach. To achieve this, a series of plasma membrane separation and protein digestion techniques as well as various membrane protein enrichment techniques were established. To aid us in estimating success of each technique we performed whole genome bioinformatic analyses. By combining the data from all these approaches and removing proteins localised to other compartments associated with, but unlikely to be part of, the plasma membrane (e.g. cytoskeleton and flagellum), the plasma membrane sub-proteome (TbPM) of long slender bloodstream form trypanosomes was defined. A number of interesting observations were made from TbPM, and it will no doubt be of benefit to the greater scientific community. One example is the positive identification of many proteins hitherto designated as putative. A quantitative approach was then employed to analyse the resistant parasites using isotope-coded affinity tagging (I-CAT) and difference gel electrophoresis (DiGE), including a novel combination of DiGE and 16-BAC protein separation technologies. Both the plasma membrane subproteome and the soluble proteome were investigated, and a number of regulated proteins identified. No confirmed transporters were differentially expressed in the resistant strain, and therefore no HAPT candidate identified. However, when the differentially regulated proteins were classified by function, relatively few functional classes were represented. The role of some proteins, with potentially relevant functions, such as a kinase, adenylate cyclase and a protein involved in kinetoplast stability, should be further investigated as they might have a

modulatory effect on the intracellular target of pentamidine or on HAPT1 activity. While time constraints prevented the further elucidation of the role of these proteins in the resistance phenotype, the work described in this thesis provides us with important new tools and information to achieve that goal.

Table of Contents

List of Tables

List of Figures

Acknowledgements

There are a huge number of people to thank, but first and foremost I would like to acknowledge my two supervisors - Harry De Koning and Richard Burchmore for their support, guidance, time and encouragement over the years. You have both helped to develop me into the young scientist that I am today and for that I am very grateful. I would also like to thank Mike Barrett - my adopted supervisor. Thank you for your time, ideas and open door.

I would also like to thank Andy Pitt for help and advice on all things mass spectrometry, Karl Burgess and Jennifer Kee for invaluable help (and a lot of patience) in helping me write Perl scripts. A huge thank you to Matt for giving so freely of your time to do the *in vivo* work. Finally to all the members of the various labs and offices who have always been there to lend a hand – Morag, Jo, Pius, Gordon, Ru, Christine, Mohammed and Mhairi. To anyone else who I have missed out, my deepest apologies.

Niall, you have helped to guide me in so many ways. You have been a fantastic sounding board for all my crazy ideas and aspirations. I thank you for your boldness, wisdom and prayers.

To my whole family, your unwavering support, encouragement and love over the years through the good and the bad has meant the world to me. I hope I make you as proud of me as I am of you. I would particularly like to thank my parents for their complete devotion to all of us children, and for being role models that we can only aspire to be.

And finally my attention turns to the most important person in my life – my beautiful wife Katherine. Knowing you has made me a better person. Your witness profoundly challenges me. You are more than I ever imagined in every way and I thank you for your love. We may not know what the future holds, but I know I can step boldly into it with you by my side.

Author's Declaration

I declare that the results presented in this thesis are my own work, except when stated otherwise.

Daniel James Bridges

March 2007

Miscellaneous

Some of the work presented within this thesis has been previously presented at the following international meetings, as well as various other more local meetings:

2006 – ICOPA, Glasgow, Scotland

2004 – BSP Meeting, Ceske Budjovice, Czech Republic

Published Papers

De Koning, H. P., D. J. Bridges, and R. J. Burchmore. 2005. Purine and pyrimidine transport in pathogenic protozoa: from biology to therapy. FEMS Microbiol.Rev. **29**:987-1020.

Bridges, D.J., Gould, M.K., Nerima, B., Mäser, P., Burchmore, R.J., De Koning, H.P. 2007. Loss of the High Affinity Pentamidine Transporter is responsible for high levels of cross-resistance between arsenical and diamidine drugs in African trypanosomes. Mol Pharmacol.

Abbreviations

16-BAC	Benzyl hexadecyl ammonium chloride
2DE	Two dimensional gel electrophoresis
ACN	Acetonitrile
AmBic	Ammonium bicarbonate
APS	Ammonium persulphate
BLAST	Basic Local Alignment Search Tool
bp	base pair
BSA	Bovine serum albumin
Bsf	Blood stream form
CAI	Codon adaptation index
CHAPS	3-[3-(cholamidopropyl)-dimethyl-1-ammonio]-1-propane sulphate
CHCA	α cyano-4-hydroxycinamic acid
Ci	Curie
CNS	Central nervous system
CSF	Cerebrospinal Fluid
DAPI	4', 6-Diamidino-2-phenylindole
DB75	2.5 bis (4-amidinophenyl) furan
DEAE	Diethylaminoethyl cellulose
DFMO	Difluoromethyl ornithine
dH_20	double distilled de-ionised water
DiGE	Differential Gel Electrophoresis
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylene diamine tetra-acetic acid
ESI	Electrospray Ionisation
GRAVY	Grand average of hydrophobicity
HAPT	High affinity pentamidine transporter
HAT	Human African trypanosomiasis
HPLC	High pressure liquid chromatography
IC_{50}	50% Inhibitory Concentration
IEF	Iso-electric focusing
IMP	Integral membrane protein
I.P.	Intra-peritoneal
IPG	Immobilised pH gradient
kb	Kilobase
LAPT	Low affinity pentamidine transporter
LB	Luria Bertani medium
MALDI	Matrix Assisted Laser Desorption Ionisation
Mb	Megabase
MelB	Melarsoprol
MelCy	Cymelarsen
MelOx	Melarsen oxide
mM	milli-Molar
MOWSE	Molecular Weight Search
mRNA	Messenger RNA
MS	Mass Spectrometry
MSR	Membrane spanning region
mtDNA	Mitochondrial DNA
MuDPIT	Multi-dimensional protein identification technology
NCBI	National Centre for Biotechnology Information
n-OG	n-octyl glucoside

OD	Optical Density
ODC	Ornithine decarboxylase
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate Buffered Solution
Pcf	Procyclic form
PCR	Polymerase Chain Reaction
PDB	Protein data bank
pI	Iso-electric point
PMF	Peptide mass fingerprint
PMSF	Phenylmethylsulfonyl fluoride
PMT	Photomultiplier tube
PTRE	Post treatment reactive encephalopathy
RNA	Ribonucleic acid
RT-PCR	Reverse transcription PCR
SDS	Sodium dodecyl sulphate
SEM	Standard error of mean
SILAC	Stable isotope labelling by amino acids in cell culture
SIT	Sterile insect technique
Taq	*Thermus Aquaticus*
TbAT1	*T. brucei* adenosine transporter 1
TBS	Tris buffered saline
TCEP	Tris carboxylethylphosphine
TEAB	Tri-ethyl ammonium bicarbonate
TEMED	N,N,N',N'-tetramethyl ethylenediaminide
TES	N-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid
TFA	Trifluoroacetic acid
TMD	Transmembrane domain
ToF	Time of flight
TPCK	N-Tosyl-L-phenylalanine chloromethyl ketone
TSB	Trypanosome separation buffer
UTR	Untranslated region
V	Volt
WHO	World Health Organisation
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside

Chapter 1

General Introduction

Trypanosomes are the causative agent of Human African Trypanosomiasis (HAT) a disease that if left untreated is invariably fatal. A number of control measures aimed at alleviating the disease burden of this parasite through reducing vector numbers etc have been very successful at various times. However, chemotherapies remain the core method of combating this disease. Unfortunately, drug resistance to current therapies is seriously undermining this stronghold. New drugs are being developed. However, even these are susceptible to cross-resistance with the existing agents. It is therefore of interest that unlike all other drugs, pentamidine resistance in the field is yet to be reported, although it has been generated *in vitro*. By understanding the mechanism(s) of resistance to this drug, we are likely to gain insights into rationally designed therapeutic strategies to evade current resistance mechanisms, target systems whereby resistance acquisition is reliant on a series of adaptations rather than a single mutation, or where it is associated with a fitness cost.

1.1 The Kinetoplastida

Trypanosomes belong to the order – kinetoplastida, which possess a unique organelle consisting of an intercatenated DNA network that resides within the mitochondrial matrix called the kinetoplast. Interestingly the kinetoplast can be absent (dyskinetoplastidy) and parasites remain viable (Shapiro and Englund, 1995). One of the sub-orders within the kinetoplastida is the trypanosomatid family (Maslov *et al.*, 2001). The trypanosomatids are a group of extracellular and intracellular protozoan parasites with a single flagellum and a small kinetoplast. Their hosts may be plants (*Phytomonas*), invertebrates (*Crithidia* and *Leptomonas*) or vertebrates (*Trypanosoma* and *Leishmania*). The latter two genera cause important human disease, are spread by insect vectors, and have plagued man over the millennia.

The genus *Trypanosoma* is capable of infecting every vertebrate class (Maslov *et al.*, 2001), however only two species cause human disease. *Trypanosoma cruzi* is exclusively located in South and Central America where it causes Chagas' disease. On the other hand *Trypanosoma brucei* is isolated to the African continent where it causes sleeping sickness throughout sub-Saharan Africa (Barrett *et al.*, 2003).

1.2 Trypanosoma brucei species

There are three morphologically indistinguishable subspecies of *Trypanosoma brucei* (*T. b. brucei*, *T. b. rhodesiense* and *T. b. gambiense*) that differ in their host specificity and range, clinical onset of disease and geographical localisation (Barrett *et al.*, 2003). Interestingly, only two groups of tsetse fly can transmit this parasite, with the *G. palpalis* and the *G. morsitans* group responsible for the majority of the propagation of *T. b. gambiense* and *T. b. rhodesiense* respectively.

1.2.1 Lifecycle

Trypanosomes have a di-genetic lifecycle split between their mammalian host and their insect vector - the tsetse fly (from the genus Glossina). A number of different lifecycle stages can be found each with a specialised biochemistry and morphology (Figure 1.1). The cycle begins with an infected tsetse fly taking a blood meal. From this meal, metacyclic trypomastigotes are introduced into the mammalian host (A). These transform into the replicative human infective long slender (LS) bloodstream form trypanosomes (B), which then disseminate throughout the host vasculature. A small proportion of LS trypanosomes continue to the non-replicative insect-infective short stumpy (SS) stage (C). Unlike the LS trypanosomes that cause the disease symptoms, it is thought that the transition from LS to SS developed in part to prolong an infection by reducing the parasite burden on the host, and therefore increasing the chances of transmission (Seed and Wenck, 2003). To continue the lifecycle, a tsetse fly ingests some of the SS form trypanosomes while taking a blood meal (D). Inside the fly midgut, the short stumpy forms transform into the proliferative metacyclic insect infective form (E). Over time, these cells exit the midgut, to transform into epimastigotes (F). These epimastigotes migrate to the salivary gland of the fly where they replicate to once again generate the human infective metacyclic trypomastigote form (G) ready for transmission to the next mammalian host. The entire lifecycle is extracellular and therefore trypanosomes are constantly exposed to the immune system of both hosts.

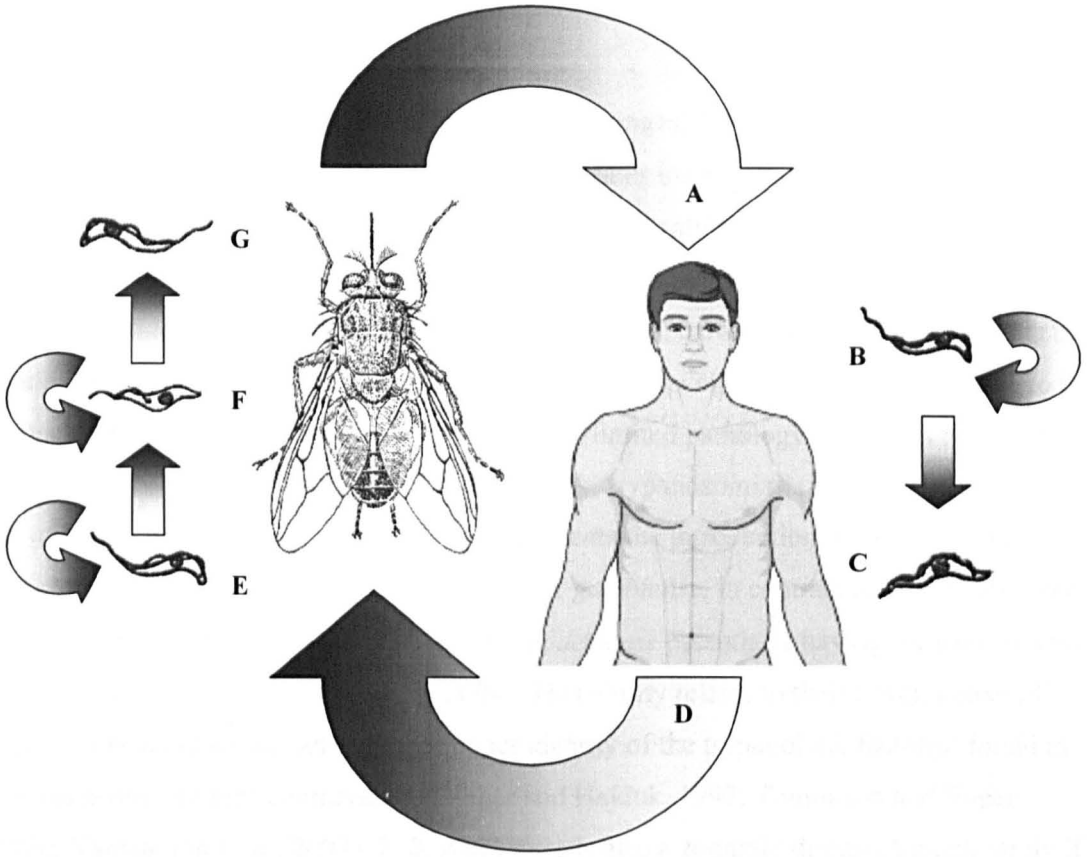


Figure 1.1 – The digenetic lifecycle of *T. brucei* in the mammalian host and tsetse fly insect vector. A number of lifecycle stages have been identified some of which are replicative (B), (E) & (F), while others are non-replicative (C) & (G). See text for details.

1.2.2 Host Range

The three *T. brucei* subspecies have different host ranges. *T. b. brucei* is unable to infect humans, but is able to infect a wide range of ungulates including many domestic and game animals, where along with *T. b. rhodesiense*, and other salivarian trypanosomes such as *T. congolense* and *T. vivax* it causes trypanosomiasis or nagana (the Zulu name). Nagana is a chronic illness that prevents or reduces milk production, calving, and very often results in death. Over time, local cattle breeds have adapted and developed an innate resistance to the parasite, where infection occurs but leads to very limited pathology as long as the animal is otherwise healthy (Murray *et al.*, 1990). However, trypanosomiasis still claims a high toll on milk and meat production in Africa and prevents the introduction of high yield breeds to endemic areas. Both *T. b. rhodesiense* and *T. b. gambiense*, in contrast to *T. b. brucei*, can infect both humans and animals, with *T. b. rhodesiense* essentially having the same animal reservoir as *T. b. brucei* (Njiru *et al.*, 2004). This ability relates to their development of human serum resistance, although the exact identity of the trypanolytic factor(s) found in human serum remains controversial (Hager and Hajduk, 1997; Tomlinson and Raper, 1996; Vanhamme *et al.*, 2003). *T. b. gambiense* is also a zoonotic disease, though rarely if ever found in cattle. The main animal reservoir appears to be various wild fauna (Njiokou *et al.*, 2006) and domestic pigs (Mehlitz *et al.*, 1982). Interestingly, the first case of human trypanosomiasis caused by *Trypanosoma evansi* was reported last year (Joshi *et al.*, 2005).

1.2.3 Clinical presentation

Following the bite of an infected tsetse fly, a localised chancre or trypanome may develop, but this occurs predominantly in *T. b. rhodesiense* infections (Barrett *et al.*, 2003). From the bite, parasites rapidly disseminate throughout the body causing a systemic illness that progresses through two defined clinical stages. This early-stage infection presents as a general malaise with a non-specific fever that not only prevents straightforward diagnosis, but also enables carriers to remain mobile, thereby increasing the chances of disease propagation. Historically, slave traders identified sleeping sickness sufferers by their enlarged posterior cervical lymph nodes; Winterbottom's sign - a phenomenon that is generally only observed in *T. b. gambiense* infections (Barrett *et al.*, 2003). Over time the parasite crosses the blood brain barrier (BBB) to invade the central nervous system, causing late stage or second stage sleeping sickness. At this point due to irreversible demyelination and the compromised integrity of the BBB, a number of various neurological dysfunctions become manifest (Enanga *et al.*, 2002). These include profound

disturbances in circadian sleep / wake patterns (hence the derivation 'sleeping sickness'), reduced higher cognitive functions and an increasingly pronounced lethargy. If left untreated, victims inexorably slide into a terminal somnolent state that culminates in coma and death (Stich *et al.*, 2002), although possible self-cure may occur in *T. b. gambiense* infections (anecdotal evidence). Unfortunately, acquiring a definitive diagnosis of early or late-stage HAT in the field, particularly with *T. b. gambiense*, is often difficult due to the low levels of parasitaemia in the CNS (for review see (Chappuis *et al.*, 2005)).

Both human infective trypanosomes generate an identical clinical disease, however they vary markedly in their rate of onset. *T. b. rhodesiense* tends to present as an acute disease with a rapid onset of only a few weeks to late stage disease. However there is considerable geographical variation, which could relate to regional differences in parasite virulence or reflect differences in human genetics of the different populations. In contrast *T. b. gambiense* can take months or years to manifest late stage trypanosomiasis, during which time carriers remain asymptomatic. Considering that the host reservoir associated with this species is small, low transmission rates would be expected, and therefore protracted infections would maximise propagation to the insect vector and subsequent host.

1.2.4 Geographical distribution

HAT is found in 36 sub-Saharan countries in around 250-300 distinct foci (Figure 1.2B). The two human-infective sub-species of *T. brucei* have a distinct geographical distribution. *T. b. gambiense* is restricted to central and western sub-Saharan Africa, whereas *T. b. rhodesiense* is found in eastern and southern Africa (Figure 1.2A). However the two sub-species are both found in close proximity to one another in Uganda and any small expansion could lead to an overlap, creating complications in diagnosis and treatment (Picozzi *et al.*, 2005).

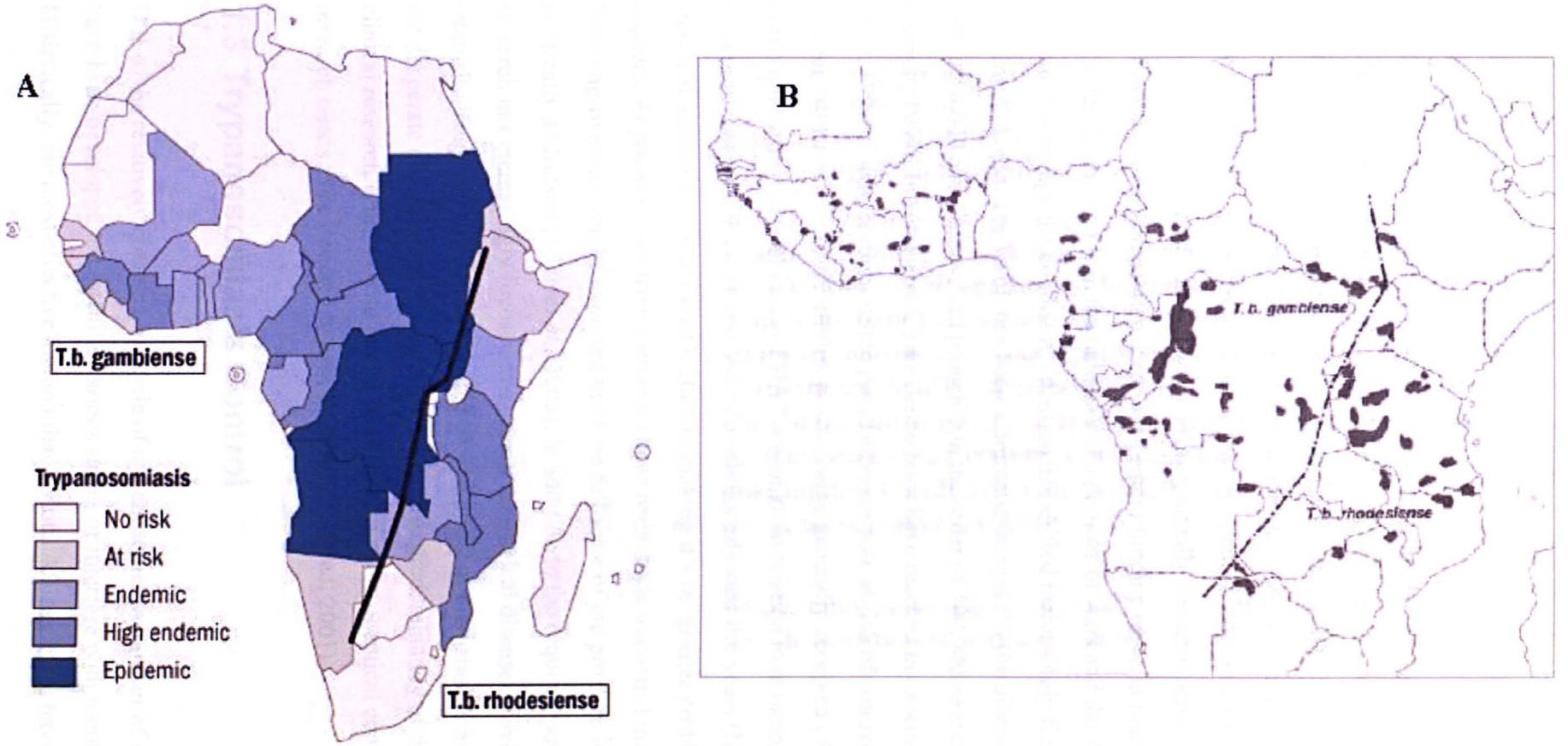


Figure 1.2 – Geographical distribution of *T. b. gambiense* and *T. b. rhodesiense* in Africa, indicating relative risk (A) and major endemic foci of sleeping sickness in 1995 (B). Reproduced from (WHO, 2000a) and (WHO, 1998), respectively.

1.2.5 History

Sleeping sickness is a controllable disease when a combined approach employing vector population control, wild animal reservoir culls, peri-domestic vegetation clearance, human screening, use of insecticides and effective treatment of infected individuals is instigated (Seed, 2000). Indeed such was the success of these implementations in the early 20th century that HAT was practically wiped out, with less than 100 cases per 100,000 in all endemic countries (TDR Strategic Direction for Research, 2002). This success story was achieved through well-funded colonial-power established programmes that centred on active surveillance and subsequent treatment of infected individuals (Cattand *et al.*, 2001). From the 1950s onwards, independence was gradually declared across Africa. Unfortunately with independence came a general funding crisis in many of these fledgling economies. Combined with the outbreak of a number of wars and the AIDS epidemic that continues to ravage the continent, attention has shifted increasingly further away from sleeping sickness. The result has been a dramatic increase in the disease, with an estimated 300 - 450,000 cases today, and probably in the order of 100,000 new cases per annum (Barrett, 1999). In some areas, prevalence rates have reached in excess of 70% (Ekwanzala *et al.*, 1996). Large sections of the population are now at risk of contracting HAT, with current estimates of ~55-60 million people being generally accepted (WHO, 1998). Although the numbers of cases of HAT are small in relation to malaria and some other infectious diseases, the cost in terms of disability adjusted life years (DALY's) lost to HAT stands at around 2 million (WHO, 2000b), making it a far greater problem than it at first appears. At present, only trivial attempts have been made to control this epidemic, and little improvement can be expected until the full scale of the problem is accepted and politically addressed (Stich *et al.*, 2003). In addition to the required political willpower, research has increasingly focused away from the clinical disease, providing us with exquisite insights into the lifecycle and biochemistry of the parasite, but failing to address the desperate clinical need that exists. As a disease predominantly of the rural poor, such clinical research is unlikely to attract interest from pharmaceutical companies and so research needs to be directed to maximise results (Seed, 2001).

1.3 Trypanosomiasis Control

Due to the relatively complex lifecycle of trypanosomes, a number of control measures have been developed to target the various stages or interfere with transmission. Historically, the most effective method that various African tribes have employed,

particularly those who rear cattle, is the avoidance of tsetse-infected areas. This is dramatically shown in Figure 1.3, and hints at the economic impact that nagana has on many sub-Saharan countries, restricting not only cattle farming, but also the use of draught animals such as horses, oxen and donkeys which are vital to proper agricultural development. A less radical approach than avoidance is to reduce numbers of tsetse flies. Historically, the use of bush clearance and game culling provided an effective albeit environmentally unsound solution. These methods have now been superseded by the development of traps that specifically attract tsetse flies (Torr *et al.*, 2005), and which are effective on a local scale as long as properly maintained and regularly impregnated with insecticide. More controversially, Sterile Insect Technique (SIT), which involves breeding large numbers of single sex flies for release into the wild, where they out compete their wild counterparts thus reducing the tsetse population (Vreysen, 2001), has been proposed as a means of eradicating tsetse flies from Africa (Schofield and Maudlin, 2001). Aside from its use in the removal of tsetse from the island of Zanzibar (Reichard, 2002), the Southern Cone Initiative (Schofield and Dias, 1999) rolled out in South America to control Triatominae is used as an example of the power of SIT. However, opinion on the potential effectiveness of such an approach even when combined with other techniques remains divided (Molyneux, 2001; Rogers and Randolph, 2002; Vinhaes and Schofield, 2003). The reason for this division is that treatment is very costly, it needs to be applied simultaneously across the continent to prevent re-invasion, it also needs to be individually applied to each tsetse species. Unfortunately, the feasibility of such a massive global approach considering the infrastructure and funding in many African nations remains poor.

In terms of agriculture, attempts have been made (through in-breeding and cross-breeding) to produce trypanotolerant dairy cattle (D'Ieteren *et al.*, 1998), although this solution completely fails to address the human disease and has had very limited success to date. All of these approaches help to combat this disease and are underpinned by chemotherapeutic options that form the mainstay of trypanosomiasis control.

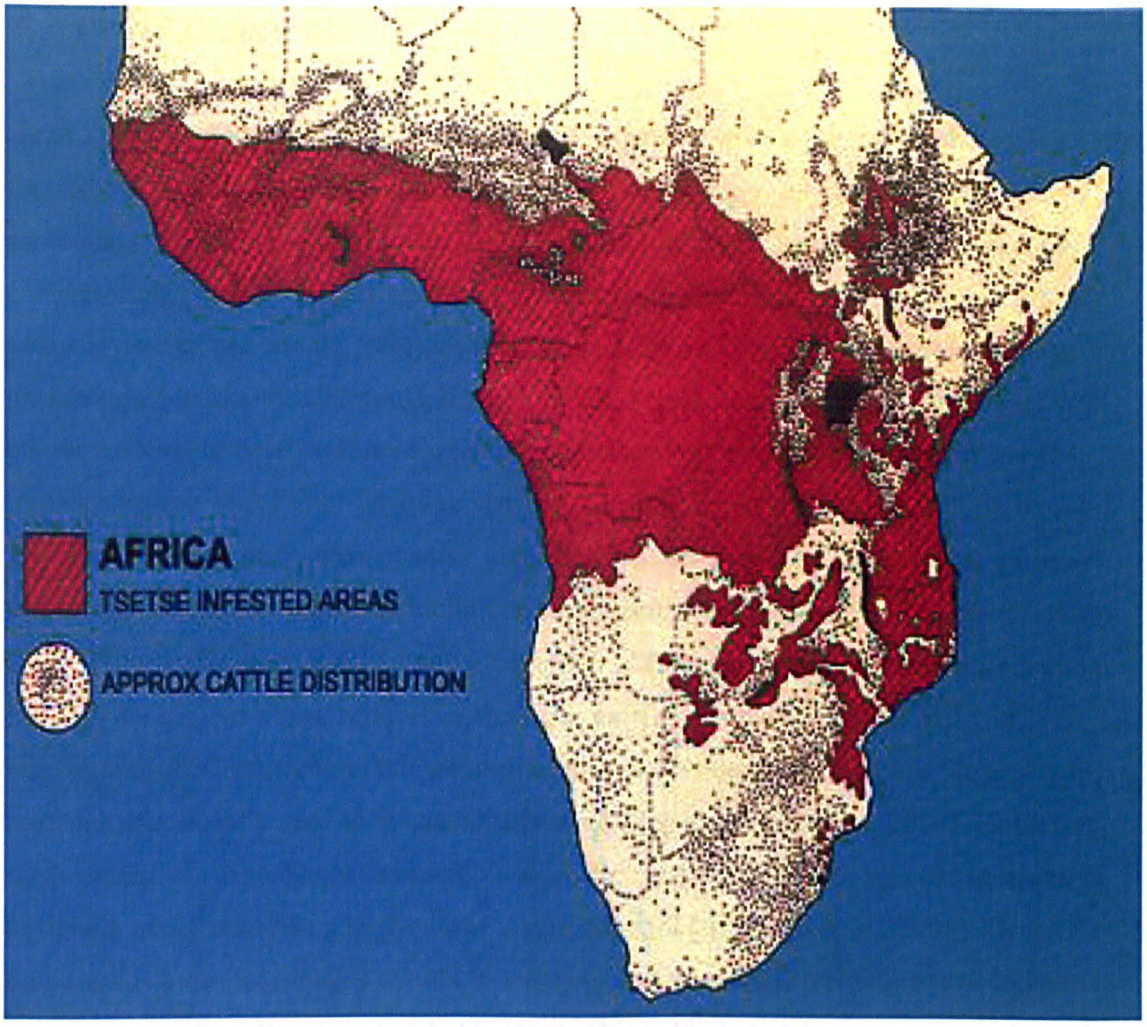


Figure 1.3 - Map of Africa illustrating the effect of the presence of tsetse flies on sustainability of rearing cattle. Taken from http://www.iaea.org/About/Policy/GC/GC45/SciProg/sftsetse.html

1.3.1 HAT Diagnosis

Since identifying trypanosomes as the causative agent of sleeping sickness, microscopy has formed the backbone of diagnostic techniques. However analysis of blood films, lymph aspirates, and CSF often fail to identify infections with low parasitaemias. Advances in concentrating parasites e.g. using mini-anion exchange columns, quantitative buffy coat analysis (Bailey and Smith, 1994), or microhaematocrit centrifugation techniques have improved detection rates, although many infections (principally *T. b. gambiense*) still fall below the level of detection. Serological diagnosis using the card agglutination test for trypanosomiasis (CATT) or the more sensitive LATEX (Penchenier *et al.*, 2003), remains problematic where only weak positive identifications are made. PCR represents a highly sensitive and specific diagnostic technique for detecting the presence of *Trypanosoma spp.* DNA (Penchenier *et al.*, 2000), but confirmation of living parasites is still required, and it can only be applied where suitable laboratories exist. Considering the difficulty in implementing PCR in Africa, it is perhaps at first surprising that proteomic fingerprinting has been successfully used to diagnose infections (Papadopoulos *et al.*, 2004). However, this technique is more likely to identify diagnostic markers for which specific tests can be designed, rather than being applied as a diagnostic tool (Agranoff *et al.*, 2005). Despite obvious advances, challenges remain in ensuring that diagnostic techniques are highly sensitive, specific, affordable and field robust. This is likely to become crucial in Uganda where a potential overlap of the two human infective sub-species, with their different treatment regimes is likely to cause major problems (Picozzi *et al.*, 2005).

1.3.2 HAT treatment

There are four drugs that are currently used to treat HAT and have been the focus of a number of in-depth articles (see references (Docampo and Moreno, 2003; Pepin and Milord, 1994; Wang, 1995)for reviews). All of these compounds have one or more serious drawbacks in terms of their activity / safety profile, cost, availability or drug resistance, and clearly there is a desperate need for new affordable, safe and effective drugs.

It is interesting to note that all of these therapeutic agents (Figure 1.4) have been shown to be active against neoplastically transformed cells, posing an interesting association between cancer and HAT (Barrett and Barrett, 2000).

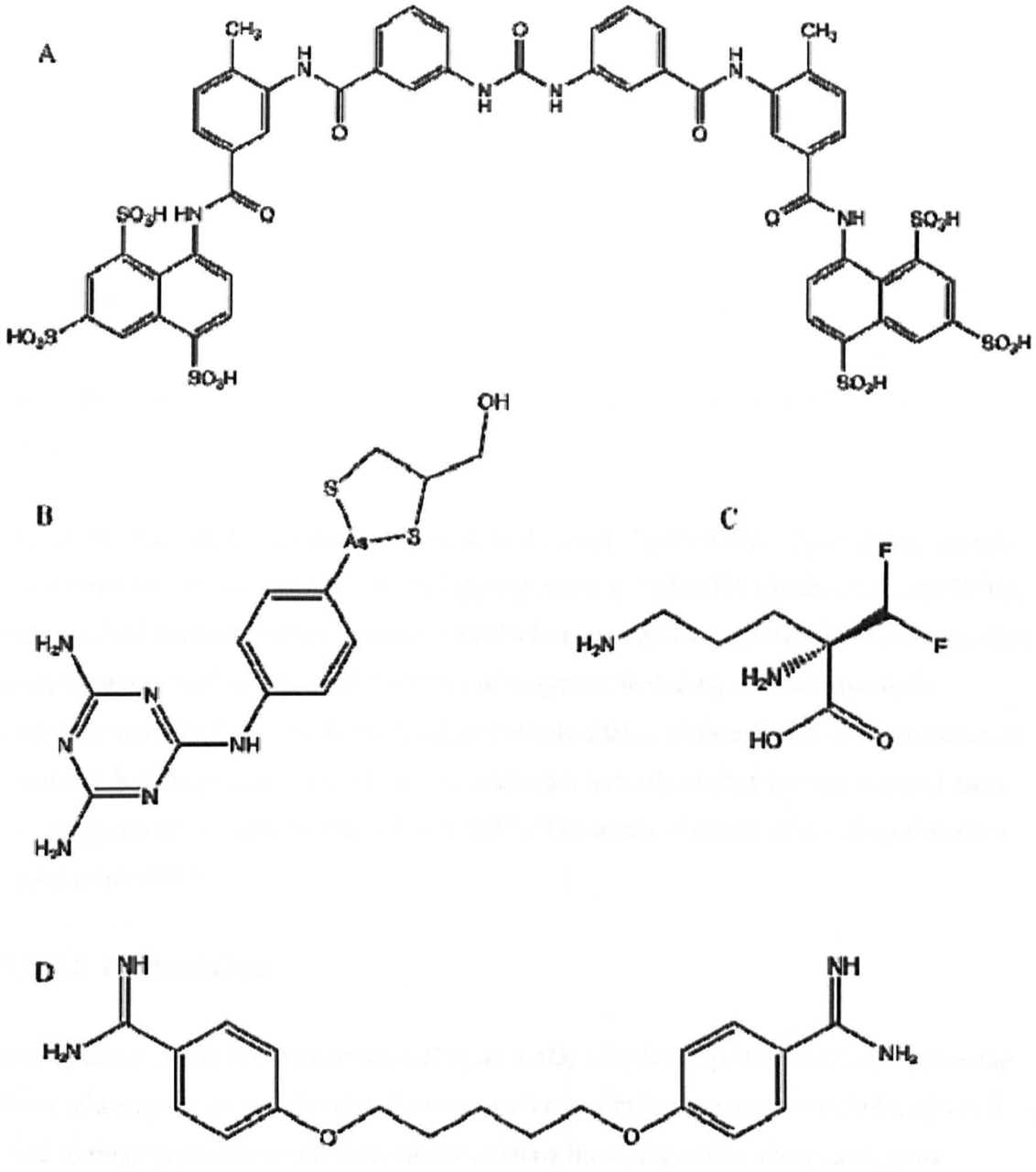


Figure 1.4 - Chemical structures of Suramin (A), Melarsoprol (B), Eflornithine / DFMO (C) and Pentamidine (D), the four drugs used to treat HAT.

1.3.2.1 Suramin

Historically, the first trypanocide used was Suramin - a symmetrical polyanionic sulfonated naphthylamine (Figure 1.4A) discovered in 1916, by chemically modifying the dye trypan red (Bouteille *et al.*, 2003). Suramin is very effective against *T. b. rhodesiense* in early stage disease, where it rapidly clears parasites from the blood. It also has a long half-life and trypanocidal drug levels remain at prophylactic levels for up to 3 months from a single dose (Apted, 1970). However, suramin can cause immediate life-threatening collapses involving renal damage, haemolytic anaemia and severe diarrhoea (Fairlamb, 2003).

Drug resistance has been well documented in the field (Matovu *et al.*, 2001c), but, despite a number of resistance mechanisms being proposed e.g. TbMRPA (Shahi *et al.*, 2002), the basis of field resistance remains undefined (Delespaux and De Koning, 2006). Suramin has been shown to bind to and inhibit a series of enzymes, including all nine glycolytic enzymes found in *T. brucei*. Glycolysis, as the sole energy source in BSF trypanosomes, is essential, but drug access to these enzymes appears unlikely as they are sequestered away in the glycosome (Opperdoes and Borst, 1977). The mode of action of this drug therefore remains unknown.

1.3.2.2 Pentamidine

The aromatic diamidine pentamidine (Figure 1.4D) was developed in 1937, and forms the basis of a number of closely related compounds e.g. diminazene aceturate (DA), which is used to treat trypanosomiasis in domestic animals including cattle, sheep, and goats. However during prolonged drug shortages, DA has been used for *T. b. gambiense* and *T. b. rhodesiense* infections in humans (Van Nieuwenhove, 1999). Pentamidine is used to treat early stage *T. b. gambiense* infections, and like suramin, is likely to exert its effect via multiple cellular targets as it reaches millimolar concentrations within the cell (Damper and Patton, 1976a; Damper and Patton, 1976b). Whilst the mechanism of action in *T. brucei* remains unclear, studies in *Saccharomyces cerevisiae* suggest that the target is of mitochondrial origin (Ludewig *et al.*, 1994).

Pentamidine, along with a number of other nitro-heterocyclic compounds such as ethidium bromide, strongly binds to the minor groove of DNA, where in combination with type II topo-isomerase's, mini-circle protein cleavable complexes may form. Interestingly, these complexes only form with kinetoplast DNA (kDNA) and not genomic DNA despite drugs

reaching similar levels (as measured by fluorescence of intercalated DNA) in both compartments (Shapiro and Englund, 1990). Indeed, the fluorescent properties of some of the diamidine drugs when intercalated with DNA now forms the basis of a diagnostic test for identifying drug resistant parasites, who acquire fluorescence at a far slower rate (Stewart *et al.*, 2005). Overall then this all suggests that the mechanism of action of pentamidine centres on its interaction with kDNA, possibly via interrupting opening / closing of the circular kDNA (Wilson *et al.*, 2005), although the exact mechanism remains unknown.

1.3.2.3 Melarsoprol

A huge breakthrough was made in the 1940's, with the discovery of the melaminophenyl arsenical compounds, of which melarsoprol (Figure 1.4B) was the prototypic member. For the first time, a trypanolytic agent that was able to cross the blood-brain barrier could be administered, and therefore effect a cure in late-stage HAT. Unfortunately this class of drugs contains arsenic, a highly poisonous heavy metal. As a result melarsoprol's safety profile is very poor, as it causes Post Treatment Reactive Encephalopathy's (PTRE) in around 5-10% of patients, with a 50% mortality rate (Docampo and Moreno, 2003). In addition, melarsoprol is poorly soluble in aqueous solutions and is therefore administered in propylene glycol, which destroys tissue and veins at the injection site. Anti-inflammatories such as prednisolone are therefore often co-administered to reduce drug toxicity (Legros *et al.*, 2002). Data on the pharmacokinetics of the trypanolytic drugs remains scarce. As a result most treatment regimes are based on empirical data and remain relatively unchanged since their introduction. Modern studies are likely to reveal ways to minimise side-effects, reduce hospitalisation times and identify more effective formulations. For example a new melarsoprol treatment schedule for late-stage *T. b. gambiense* sleeping sickness that reduced hospitalisation time (down to 10 days) and drug quantity (by 30%), was trialled in Angola and found to be as effective as the standard schedules (Burri *et al.*, 2000). Unfortunately resistance has occurred for some time in the field (Bales, Jr. *et al.*, 1989; Ogada, 1974), and appears to be on the increase (Brun *et al.*, 2001), threatening treatment in many areas.

1.3.2.4 DFMO

Discovered in 1977, DFMO (Figure 1.4C) was nicknamed the 'miracle drug' or 'resurrection drug' for its ability to cure patients with severe late stage *T. b. gambiense* HAT unresponsive to melarsoprol. To date, DFMO is the only trypanocide whose

mechanism of action has been elucidated: it is an irreversible inhibitor of ornithine decarboxylase (ODC), most probably via covalent binding of the drug to a residue within the active site of the enzyme (Metcalf *et al.*, 1978). ODC is an essential enzyme that converts ornithine to putrescine as the first committed step in polyamine biosynthesis. Inhibition of this pathway *in vivo*, reduces cellular putrescine levels to undetectable levels after 12hours (Fairlamb *et al.*, 1987). The result is the complete abrogation of cell proliferation and perhaps more importantly a huge reduction in VSG synthesis (Bitonti *et al.*, 1988; Li *et al.*, 1998). Thus although the parasites remain viable in the presence of this trypanostatic drug, their ability to replicate and evade the immune system is removed, enabling the host to rapidly overcome the infection. DFMO also blocks mammalian ODC, however this is turned over far more rapidly than the *T. brucei* ODC (approximately every 20 min). This combined with the short drug half-life ensures that the parasite is selectively targeted. An additional factor may be that mammalian cells (unlike trypanosomes) are able to effect high-affinity polyamine uptake when facing polyamine starvation, thus bypassing the effect of eflornithine (Fairlamb, 2003). The inherent resistance observed in *T. b. rhodesiense* to eflornithine (Iten *et al.*, 1995) is likely to be due to increased ODC turnover in this subspecies, relative to *T. b. gambiense*.

Unfortunately DFMO's short half-life requires correspondingly high dosages i.e. 400 mg/kg^{-1} per day, the administration of which (four 2 hour intravenous infusions per day for 7 or 14 days) requires extensive hospitalisation (Fairlamb, 2003). DFMO is also very expensive - with costs in excess of $300-500 per treatment (TDR Strategic Direction for Research, 2002), and its supply has not been secured in the long term (Wickware, 2002). Although clinical resistance to this drug has not yet been documented, given that increased ODC turnover would be expected to confer resistance, it is unlikely that this will continue indefinitely.

1.3.2.5 Future Drugs and Prognosis

As outlined above, all of the drugs currently registered for treatment of HAT have associated problems, stressing the desperate need for new drugs. However, the development of drugs is long, expensive (estimated at $800 million per drug), and increasingly difficult (Croft *et al.*, 2005). Considering that HAT is a disease exclusively associated with the developing world, with a small commercial market worth <£1 million p.a. in 1983 (Gutteridge, 1985), the incentive for the large pharmaceutical companies to tackle this disease remains low. Indeed, despite very substantive advances in trypanosome research, most notable the publication of the *T. brucei* genome (Berriman *et al.*, 2005),

knowledge has not been translated into new therapies and progress on the whole has been disappointing (Naula and Burchmore, 2003; Wirth, 2001). One exception to this trend is the discovery of a new orally administered pro-drug; DB-289 (Yeates, 2003), which is now undergoing stage III human trials in Angola (Mathis *et al.*, 2006). If clinically proven in the field, the ease of oral administration would go some way to aiding treatment. Combination therapies also need to be further investigated (Legros *et al.*, 2002), for example DFMO has been found to act synergistically with 9-deazainosine or suramin to treat late-stage *T. b. rhodesiense* (Matovu *et al.*, 2001c). Clinical trials are also under way using the longstanding Chagas disease treatment nifurtimox (Lampit®) in combination with melarsoprol or DFMO (Delespaux and De Koning, 2006; Van Nieuwenhove, 1992).

1.3.2.6 Vaccine

As extracellular and therefore immune-accessible parasites, it was hoped that a unique *T. brucei* protein marker could be found from which a vaccine could be developed. However, the parasite has the ability to alter its antigenic signature by switching expression from one variable surface glycoprotein (VSG) to another (see (Donelson, 2003) for review) - a process called antigenic variation. The VSG coat forms a dense surface array that sterically hinders the approach of antibodies, thus protecting against complement attack and exposure of highly conserved membrane proteins. There are around 1000 VSG genes, although many of these are pseudo- or truncated genes. Indeed the VSG complement is constantly changing through mosaic recombination, creating an almost limitless stock of antigens. VSG transcription is very tightly regulated, possibly in an extra-nucleolar body, so that only one VSG is transcribed at any one point. This vast immunogenic repertoire enables the parasite to evade the immune response, and makes development of a vaccine highly unlikely. However, VSGs are anchored to the membrane via a unique glycophosphatidyl (GPI) anchor, and differences between mammalian and parasitic GPI biosynthetic pathways have revealed a number of potential (Ferguson *et al.*, 1999) and validated (Smith *et al.*, 2004) drug targets, prompting new hope of a vaccine, although current efforts are aimed towards chemotherapies.

1.4 Drug resistance

There are essentially two mechanisms by which an organism can become resistant to a drug. The first involves a change in the drug target whereby the drug interactions are abolished / altered so that the drug no longer has the desired effect, or the target is over-expressed to compensate for the drugs effect. The second involves a reduction in the level

of active drug in the target compartment. This may be effected by reduced uptake, drug inactivation / activation failure, sequestration, or increased drug efflux. Excluding drug inactivation, all of these mechanisms involve transporters. Transporters can potentially mediating drug resistance through a number of mechanisms. Such mechanisms are most likely to be observed where a drug elicits its effect(s) via multiple intracellular targets or with a invariable target e.g. DNA, that cannot alter its structure to affect drug interactions. In this case, it could be envisaged that the most effective acquisition of resistance would be through altering the action of a single transporter to prevent drugs accessing, or remaining in, the compartment where they elicit their effect.

1.4.1 Drug resistance in T. brucei

Before looking at drug resistance in *T. brucei*, it is important to bear in mind that a refractory / relapse case does not definitively indicate the emergence of drug resistance (Brun *et al.*, 2001). More subtle parasitic mechanisms such as sequestration in the CNS or other tissue niches that are relatively drug-inaccessible may be involved. The other obvious aspect is the occurrence of differences between hosts: e.g. CNS drug levels may vary from patient to patient thus promoting or preventing a relapse.

Pentamidine, suramin, and melarsoprol have all been in widespread clinical use for more than 50 years. Considering such protracted and extensive applications using treatment schedules that have not been optimised to reduce selection for resistance, it is no surprise that drug resistance has emerged (Bacchi, 1993). For example, melarsoprol relapse rates across Africa have historically been in the region of 5-8%. Regrettably, far higher relapse rates are now being reported e.g. 30% in Arua, Uganda (Legros *et al.*, 1999), although fortunately melarsoprol resistance has only been observed in *T. b. gambiense*, where eflornithine remains as a backup drug. The chilling thought is that if resistance also develops in *T. b. rhodesiense* there will be no alternative treatment (Kaminsky and Mäser, 2000). Indeed, there are anecdotal reports of this already being the case, but as a result of the massive underreporting of East African HAT, reliable data are not available. Perhaps more surprising is the case of pentamidine, which – apart from its curative use – has been used prophylactically to treat almost entire populations for a number of decades and yet resistance has not been observed in the field (Bray *et al.*, 2003), although along with Melarsoprol (Carter and Fairlamb, 1993), Cymelarsen (Scott *et al.*, 1996) and DFMO (Phillips and Wang, 1987), pentamidine resistance has been successfully induced *in vitro* (Berger *et al.*, 1993). While it is difficult to quantify the selection pressure that

pentamidine exerts on trypanosomes in the field this observation suggests that multiple mutations are required to confer resistance.

Interestingly, there are examples of cross-resistance between diamidines and melarsoprol (Frommel and Balber, 1987), suggesting a common mode of action (Barrett and Fairlamb, 1999). In this particular paper, resistance appeared to be achieved via reduced cellular accumulation of the drug. Another strain of melarsen-resistant *T. brucei* displayed increasing diamidine cross-resistance as interatomic distance between amidine carbons decreased (Fairlamb *et al.*, 1992).

1.4.2 Drug Efflux

Drug efflux as a mechanism of resistance has been intensely studied in human cancers. Many of these ABC (ATP binding cassette) transporters (Klein *et al.*, 1999), which include P-glycoprotein proteins (pgp) from the MRP family (Borst *et al.*, 1999; Borst *et al.*, 2000), have been shown to affect human tumour chemosensitivity and are capable of extruding a diverse range of compounds from tumour cells to confer multi-drug resistance. A number of inhibitors and functional knock-down strategies have been developed to 're-sensitize' these cells although no effective method has so far been found to do so therapeutically (Fojo and Bates, 2003).

T. brucei contains at least three ABC transporters, which were found by data mining its genome (Mäser and Kaminsky, 1998). 100-fold over-expression (not confirmed by Western blot) of one of these transporters (TbMRPA) *in vitro* was shown to mediate a 10-fold increase in melarsoprol resistance (Lüscher *et al.*, 2006; Shahi *et al.*, 2002), although so far no drug resistant lab- or field-derived strains have been shown to up-regulate TbMRPA expression. Nevertheless, this work demonstrates the potential role that efflux transporters may play in resistance.

There are, however, many well-documented examples of drug resistance being mediated by efflux pumps in other parasitic protozoa (see (Klokouzas *et al.*, 2003; Ullman, 1995) for overview). Some of the best examples comes from *Leishmania*, where a pgp antibody was used to detect a putative pgp protein in *L. donovani*, that conferred arsenite-resistance when over-expressed (Kaur and Dey, 2000). Similarly a genetic screen identified an ABC transporter that conferred pentamidine resistance in *L. major* (Coelho *et al.*, 2003), and while not confirmed, evidence does suggest that efflux pumps are also important in the field (Singh, 2006). Likewise, in *P. falciparum*, an efflux pump has been shown to mediate

resistance via the verapamil sensitive extrusion of chloroquine (Sanchez *et al.*, 2003; Sanchez *et al.*, 2005), although this the gene responsible for this component of transport does not belong to the ABC super-family (Howard *et al.*, 2002).

1.4.3 Nucleoside Transporters

Trypanosomes are incapable of *de novo* purine synthesis, and along with all parasitic protozoa scavenge these compounds from their host (see (De Koning *et al.*, 2005) for review). In *T. b. brucei* bloodstream forms, purine nucleoside uptake consists of two components (P1 and P2), which can be differentiated based on their kinetics and substrate affinities: whereas P1 is sensitive to competitive inhibition by inosine, P2 is selectively inhibited by adenine. A loss of P2 transport was observed in an arsenical-resistant line (Carter and Fairlamb, 1993), suggesting its involvement in drug uptake. This is confirmed by the ability of the P2 transporter to recognise a range of diverse unrelated structures e.g. melarsoprol, adenosine, and pentamidine (De Koning and Jarvis, 1999). Further work using a range of structural analogues to determine each compound's binding energy for P1 or P2 allowed substrate recognition profiles for both transporters to be formulated. The P2 recognition criteria were satisfied by both pentamidine and the melaminophenyl arsenicals (Barrett and Fairlamb, 1999; De Koning and Jarvis, 1999).

The gene responsible for the P2 component of purine transport was identified and named as TbAT1 (*T. b. brucei* adenosine transporter 1). This was achieved by transforming yeast defective in purine biogenesis with a *T. b. brucei* cDNA library and then selecting for clones that grew in media containing adenosine as the sole purine source (Mäser *et al.*, 1999). This study also sequenced the *TbAT1* genes from a *T. b. brucei* isogenic pair. Two alleles were defined which could be differentiated by restriction fragment length polymorphism (RFLP) analysis using the restriction enzyme SfaNI. This RFLP analysis was then used on some culture-adapted field isolates from Uganda, where a significant association was found between RFLP status and relapse, although the correlation was not perfect (Matovu *et al.*, 2001a). Further work using a combination of RFLP analysis, SSCP (single strand conformation polymorphism) and direct sequencing has revealed a remarkably small number of TbAT1 variants. In fact, a common set of nine mutations were found in different *T. brucei* subspecies and from different geographical locations (Matovu *et al.*, 2001b). However a number of relapse patients retained the wild-type TbAT1 gene suggesting that additional factors were involved. Nonetheless all of this was strong evidence of TbAT1's involvement in drug resistance. Definitive verification came with the generation of a TbAT1 knock out (KO) line, which showed a significant reduction in

sensitivity to various arsenicals and diamidines (Matovu *et al.*, 2003), confirming a mechanism of cross-resistance between these two classes of structurally distinct drugs. Perhaps what was most surprising however was that the resistance phenotype varied considerably from drug to drug. For example while TbAT1 KO cells showed almost a 20-fold increase in resistance to diminazene, pentamidine resistance had only increased by 2-3 fold. While the functional loss of TbAT1 may be clinically relevant, this result clearly demonstrates that TbAT1 is not the only gene capable of modulating pentamidine resistance. Diminazene on the other hand appears to be almost exclusively transported via the P2 transporter and hence the acquisition of resistance is far simpler (De Koning *et al.*, 2004).

1.4.4 Additional Pentamidine Transporters

For a while, it has been known that pentamidine can enter trypanosomes via transporters other than TbAT1 (De Koning and Jarvis, 2001). This transport is adenosine-insensitive and has two components of transport termed the high and low affinity pentamidine transporters (HAPT1 and LAPT1 respectively). Although the genes encoding these putative transporters are yet to be identified, a biochemical characterisation has enabled some crucial kinetic details to be elucidated (Table 1.1), from which an overall picture of pentamidine transport in *T. brucei* has now been formulated (Figure 1.5). These kinetic details allow insights and predictions to be made as to the relative importance of the different components in terms of their impact on drug transport (Table 1.1). For example, the P2 component of transport due to its low K_m is the primary route of entry at low pentamidine concentrations (10nM). HAPT1 is also important at low pentamidine concentrations, however its low V_{max} means that it is soon saturated. LAPT1, on the other hand with its high V_{max} and low K_m becomes very important at higher drug concentrations. The mere presence of these three pentamidine transport components helps to explain why resistance to pentamidine has so far not been observed in the field. This becomes even more understandable when it can be seen how rapidly their contribution to pentamidine transport changes with drug concentration - a situation that would very likely occur spatially and temporally in a patient undergoing treatment. Similarly it is easy to see why cross-resistance to pentamidine in arsenical- and diminazene-refractory strains is also rare (De Koning, 2001), although as stated previously, some cross-resistance via loss of a functional TbAT1 is possible.

Frustratingly, the physiological substrates of HAPT1 and LAPT1 remain unknown despite testing a wide range of purines, pyrimidines (De Koning, 2001; De Koning and Jarvis,

2001) and other unrelated molecules including choline (De Koning *et al.*, 2005). This is in contrast to *P. falciparum*, where the choline transporter in combination with the New Permeation Pathway (NPP) has been shown to be the major route of entry into the parasite (Biagini *et al.*, 2004). Intriguingly, the first arginine transporter characterised from *L. donovani* can be inhibited by pentamidine, although the drug itself doesn't seem to be transported (Shaked-Mishan *et al.*, 2006). Considering that the *T. brucei* genome so far contains 43 ORFs annotated as amino acid transporters, it is possible that one or more of these could encode HAPT1 and or LAPT1.

It is intriguing to note that a gene (PNT1) that confers resistance to pentamidine in *S. cerevisiae*, is not a transporter (Ludewig and Staben, 1994). PNT1 may however be involved in membrane protein expression.

	HAPT1	LAPT1	P2
Kinetic Parameters			
K_m (μM)	0.036 ± 0.06	56.2 ± 8.3	0.26 ± 0.03
V_{max} [a]	0.0044 ± 0.0004	0.85 ± 0.15	0.068 ± 0.007
K_i values for inhibitors			
Adenosine	N.E.	N.E.	0.8 ± 0.12
Propamidine	4.6 ± 0.7	>250	1.9 ± 0.8
Component of uptake (%)			
at 10nM	26	4	69
at 1μM	6	20	74
at 10μM	2	65	33

[a] in pmol.(10^7cells)$^{-1}$.s^{-1}

Table 1.1 - Summary of kinetic parameters, inhibition constants (μM) and component of substrate uptake (%) of the three pentamidine transporters in *T. brucei*. HAPT1: high-affinity pentamidine transporter 1; LAPT1: low-affinity pentamidine transporter 1; P2: purine transporter 2 encoded by TbAT1 (GeneDB name AF152369). N.E. - No Effect. Values are taken from (Bray *et al.*, 2003; De Koning, 2001).

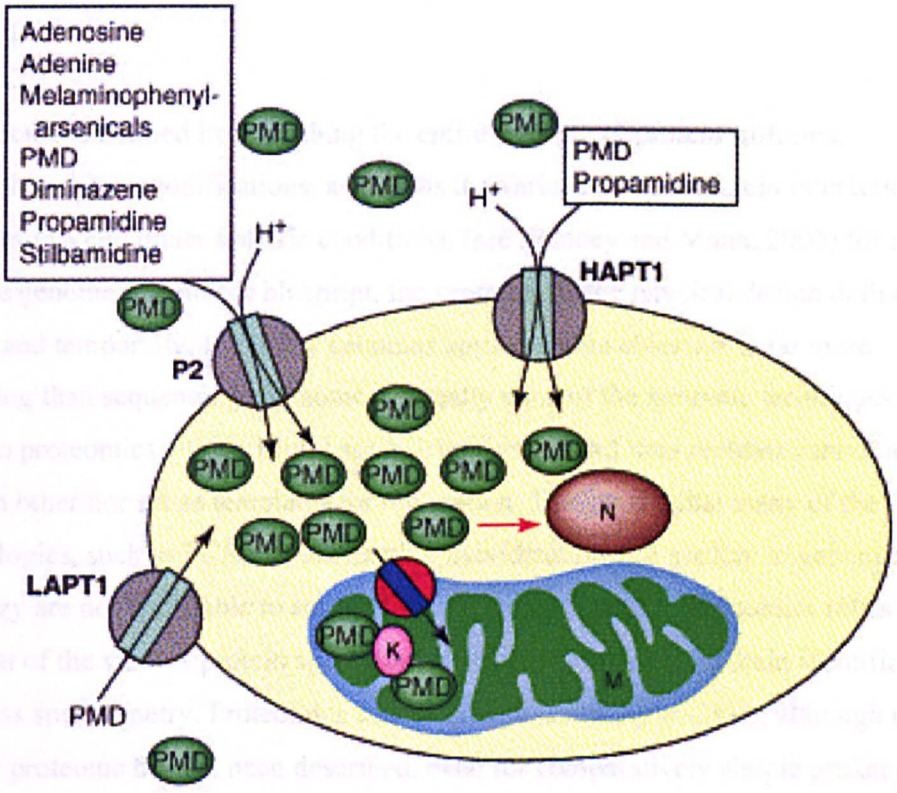


Figure 1.5 – Pentamidine uptake routes in *T. brucei*. Adapted from (Bray *et al.*, 2003). LAPT1, low affinity pentamidine transporter 1; HAPT1, high affinity pentamidine transporter 1; PMD, Pentamidine; N, Nucleus; M, Mitochondrion; K, Kinetoplast.

1.5 Proteomics

The genomic era has revolutionised biosciences, with molecular biology forming the cornerstone of most laboratory techniques, indeed the large and ever increasing number of genomes that have been sequenced is testament to the importance of such information. However, though genomes give an array of vital information, it is becoming increasingly clear that little functional data is evident from the genome sequence alone. Micro-array analysis, while useful, is unable to give definitive data on the expression of a protein due to the poor correlation between mRNA and protein levels. In the kinetoplastids such as *T. brucei*, this correlation is likely to be even weaker than in other eukaryotes due to their polycistronic gene expression, where gene regulation takes place almost entirely at the post-transcriptional level (Clayton, 2002). To ask meaningful biological questions in these organisms, in the context of analysing changes in protein expression, a *de novo* protein analysis must be performed. The quest to achieve this has given rise to the advent of the next 'omics - Proteomics (see (Aebersold and Mann, 2003; Mann *et al.*, 2001) for overview).

The proteome is defined by describing the entire panoply of protein isoforms, polymorphisms and modifications, as well as the various protein-protein interactions that are present in a cell under specific conditions, (see (Pandey and Mann, 2000) for review). While the genome is a simple blueprint, the proteome is the physical design defined both spatially and temporally. It quickly becomes apparent, this objective is far more challenging than sequencing a genome. Virtually none of the genomic techniques can be applied to proteomics – there is no base-pairing system and thus proteins cannot hybridize with each other nor act as templates for replication. This means that many of the methodologies, such as PCR and subtractive hybridisation that are key to genomic technology are not applicable to study of the proteome. Instead, proteomics relies on the separation of the various protein species in a sample, followed by protein identification using mass spectrometry. Proteomics as a discipline is rapidly evolving although no complete proteome has yet been described, even for comparatively simple prokaryotes. This is testament to the challenge posed by such an approach and a number of obstacles to harnessing the power of proteomics remain (Tyers and Mann, 2003).

1.5.1 Instability

Unlike static genomes, proteomes are highly dynamic and can alter dramatically in response to stimuli or simply as a function of time. If answering a biological question, it is therefore important to analyse the proteome under biologically relevant conditions and within a biologically relevant timeframe. In the context of trypanosomes, the relevant lifecycle stage needs to be defined as well as the environment that the parasite is maintained in.

1.5.2 Complexity

It is worth noting that an organism's proteome is far more complex than its corresponding genome due to protein isoforms, alternative splicing of introns, alternative recombination of subunits, glycosylation, phosphorylation, proteolytic cleavage and other post-translational modifications. For example, the human genome contains around 30,000 genes, yet the number of protein species thought to be present in human serum alone is likely to far exceed this number, despite only a relatively small subset of the total human genome being expressed in this cell type (Aebersold, 2003). In the same way, many *T. brucei* proteins will only be expressed in a specific lifecycle stage or under certain environmental cues, but the total number of protein species expressed is likely to be far greater than the number of ORF's in the genome. Virtually none of this complexity can be accessed purely through knowledge of the genome sequence. This tremendous knowledge gap drives the need to study cells at the protein level.

1.5.3 Dynamic Range

Protein expression levels (both in terms of absolute, active, and local levels) can vary tremendously. For example 50% of the total protein content in yeast comes from just 100 genes, while low abundance proteins make up 80% of the predicted yeast proteome (Pedersen *et al.*, 2003). Different proteins have a huge dynamic range, with some present at less than 50 copies per cell through to highly abundant proteins present at $>10^6$ copies per cell.

1.5.4 Pre-fractionation

To reduce the complexity of a sample and to focus in on certain protein classes, sample pre-fractionation is highly desirable. It is particularly advisable where the target class of

proteins are poorly expressed. In this case pre-fractionation will serve to enrich the sample for those target species, by removing many of the others.

There are a large number of ways to pre-fractionate a sample (see (Righetti *et al.*, 2005) for review). Many proteomic investigations focus on a particular organelle or cellular structure. In this case a physical fractionation approach is favoured e.g. cell lysis followed by differential centrifugation (Schirmer *et al.*, 2003). Where a specific class of proteins are being investigated, fractionation based on the physicochemical properties of the proteins can be performed. For example hydrophobic / hydrophilic proteins can be separated due to their differential solubilisation in detergent e.g. Triton X-114 (Brusca and Radolf, 1994). Chromatographic separations are able to utilise a wide range of solid-phase materials to enable separation based on size-exclusion, anion / cation exchange, reverse-phase etc. Indeed, stacked sorbents combine a number of these different chemistries in series to sort the entire proteome into discrete sub-fractions, each with a different proteome bias (Righetti *et al.*, 2005). In some cases, where a handful of proteins dominate the proteome, rather than trying to enrich for the underlying target proteins, it may be more effective to try and deplete the highly expressed proteins. For example in *T. brucei*, there are ~10^7 copies of a single VSG gene per cell membrane (Ferguson *et al.*, 1999).

1.5.5 Protein Separation

Even with pre-fractionation, any sample is likely to contain a great many protein species. To be able to identify each of these, it is necessary to separate them. Fortunately, proteins are highly heterogeneous, for instance they all have different hydropathies, isoelectric points (pI), molecular weights and post-translational modifications. All of these differences can be utilised to resolve a proteome. Historically, proteomics has been synonymous with the use of two dimensional polyacrylamide sodium dodecyl sulphate gel electrophoresis (2D-GE), first described in 1975 (O'Farrell, 1975). This technique separates proteins based on their pI (first dimension) and molecular weight (second dimension). Proteins in the gel can then be visualised by staining e.g. with Coomassie blue, before picking spots for protein identification. On a typical 2D-GE gel ~1000 spots, each containing a different protein species, may be seen depending on the stain used. By running samples on a series of narrow pH gradient (zoom) gels, proteome coverage can be maximised. These advances have made 2D-GE an attractive technique when interrogating the soluble proteome. Unfortunately hydrophobic (i.e. membrane bound / associated) proteins are generally poorly represented on 2D-GE due to their tendency to aggregate during the electrofocusing step. The addition of a co-solvent appears to improve the situation (Deshusses *et al.*, 2003),

however questions over the validity of a 2D-GE approach for membrane proteins remains, although many still believe it to be the best technique available (Fey and Larsen, 2001). The shortfalls of 2D-GE has forced some groups to return to using 1D-GE, despite its reduced resolution, because of its enhanced compatibility with hydrophobic polytopic membrane proteins (Galeva and Altermann, 2002).

Liquid chromatography (LC) represents the other main method for separating proteins or peptides. A sample can be separated in this way based on hydrophobicity, charge, size, etc. The real advantage of using an LC approach is that any number of separation procedures can be performed in sequence- the so-called multi-dimensional protein identification technology (MuDPIT) (Liu *et al.*, 2002; Washburn *et al.*, 2001). Conventionally MuDPIT is performed using a two-dimensional separation, although a three-dimensional approach has also been used (Wei *et al.*, 2005). LC systems are also now highly automated, enabling accurate and reproducible analysis. A MudPIT based approach may even have greater resolution that 2D-GE. In addition it is able to identify proteins across a dynamic range of 10,000 to 1 (Wolters *et al.*, 2001).

1.5.6 Membrane Proteins

This class of proteins performs a large number of functions including cell-signalling, substrate transport, maintaining ion gradients, cellular interactions, cell compartmentalization, and energy generation. Such is the importance of this class of proteins that despite representing at most 30% of a given proteome, they constitute 70% of all known pharmaceutical drug targets (Wu and Yates, III, 2003). With such interest in these proteins, their proteomic analysis has been a high priority, but technically demanding for a number of reasons. Firstly, membrane proteins are typically of very low-abundance. Fortunately, fractionation of cells, to isolate the organelle or structure of interest (Taylor *et al.*, 2003), combined with selective enrichment for hydrophobic proteins (Bordier, 1981; Jones, 1999; Jones *et al.*, 1990) can often solve this problem. A far greater challenge lies in the manipulation of the proteins post-isolation. This is due to the fact that all membrane proteins by definition interact with the lipid bilayer, and to do so have regions that are highly hydrophobic. Most proteomic techniques were developed for hydrophilic proteins and employ aqueous environments where hydrophobic regions (unless stabilised) will tend to aggregate together causing the protein to precipitate. As a result a number of solutions, have been developed to enable proteomic projects directed at membrane proteins to be successfully pursued (Blonder *et al.*, 2004; Molloy *et al.*, 2000).

1.6 Mass spectrometry (MS)

Mass spectrometry lies at the heart of proteomics and while there are a huge array of different instruments available all with their own inherent advantages and disadvantages (see (Baldwin, 2005) for review), all MS machines are based on the same fundamental design. While the MS analysis of intact proteins (the so-called top-down approach) looks extremely attractive it is difficult to introduce intact proteins to the MS machine. In addition, MS of intact proteins requires very high mass resolution and, at present, this can only be achieved by the use of very expensive Fourier transform ion cyclotron resonance (FT-ICR) machines (Guerrera and Kleiner, 2005; Loo *et al.*, 1990). Currently developed high throughput techniques amenable for global proteome analysis therefore focus almost exclusively on peptides.

MS is built around four steps. Firstly the peptide sample is introduced into the machine and then it is ionised. Ions of a particular mass and charge ratio pass through a mass analyser before the ion detector measures their mass charge ratio (*m/z*). It is important to note that *m/z* is measured rather than mass.

1.6.1 Ionisation

In terms of ionisation, high-energy ionisation was historically used in MS. Unfortunately this sort of ionisation caused labile molecules such as proteins to be degraded. It was therefore only after the introduction of 'soft' ionisation techniques that MS could be applied to protein samples. Soft ionisation transfers protein sample from the solid / liquid phase into the gas phase while imparting a charge to the molecule, without fragmenting it. By applying a potential difference across the generated ion field, the ions will 'fly' in a direction relative to the polarity of the voltage applied. MS can therefore be performed in positive or negative mode. Most proteins are analysed under positive ionisation conditions. There are essentially two ionisation techniques that achieve this in two very different ways, although they are complementary, and the advantages and disadvantages of each should be examined in relation to the experimental aims.

1.6.1.1 Electrospray Ionisation (ESI)

ESI (Fenn *et al.*, 1989) uses sample in the liquid phase. It is an extremely efficient technique at ionising sample and, for proteomic studies, is typically combined with a High Pressure Liquid Chromatography (HPLC) system which makes its application very

powerful (Romijn *et al.*, 2003). To generate ionic species, the analyte is dissolved in a volatile solvent and sprayed out of a fine needle at a high electrical potential. This produces a stream of fine droplets that are aimed towards the counter electrode at the opening of the MS. As the solvent evaporates, the ensuing droplets shrink, and electrical charge increases until electrostatic repulsion exceeds the surface tension (Rayleigh limit) and the droplet blows apart. This either creates smaller droplets (for which the whole process can be repeated) or discrete solvated ions (Lane, 2005).

1.6.1.2 Matrix Assisted Laser Desorption / Ionisation (MALDI)

MALDI (Hillenkamp and Karas, 1990) involves co-crystallising peptide sample with an excess of organic matrix onto a target plate. MALDI is therefore essentially an off-line technique that allows a user to repeatedly analyse the same sample and therefore focus on a particular spectral region or ion species (Baldwin, 2005). The matrix contains a chromophore that can be excited with a laser beam. Upon laser excitation, a portion of the matrix along with the co-crystallised sample is vaporised and ionised, although the exact mechanism remains unclear (Karas and Kruger, 2003). Pulses of laser are used to generate discontinuous ion packets. MALDI is therefore often coupled to an MS instrument that can trap all ions or measure a complete mass spectrum (see section 1.6.3). The matrix absorbs almost all of the laser energy (being in vast excess) thus preventing sample degradation. Different matrices favour the formation of larger or smaller ions (Wysocki *et al.*, 2005), due to differences in how much energy is imparted to the analyte. In addition, this form of ionisation invariably produces only singly charged ions (Karas *et al.*, 2000).

1.6.1.3 Ambient MS

Both MALDI and ESI require sample preparation. To study living systems *in situ* a new class of techniques e.g. desorption electrospray ionisation (DESI) has emerged that can perform MS analysis on native samples (Cooks *et al.*, 2006). It is easy to envisage how this technology can be applied to imaging mass spectroscopy (Guerrera and Kleiner, 2005), where different proteomes are defined across a section of tissue.

1.6.2 Fragmentation

Both soft ionisation techniques ionise and vaporise sample, but due to the low internal energy transfer, do not fragment the sample. Without fragmentation only intact peptide masses can be measured. When working with simple peptide mixtures, a protein identity

may be assigned from this information alone, however as sample complexity and therefore spectral complexity increases (even with high mass accuracy) protein assignments will be harder to make (see section 1.6.5). Many machines therefore now fragment peptides to yield primary structure amino acid sequence information (Figure 1.6D). This process of scanning the mass range, selecting a specific ion, fragmenting it and then measuring the resultant fragments is called tandem mass spectrometry (MS/MS).

When proteins are fragmented a series of different daughter ions are created from the parent, although *b* and *y* ions (Figure 1.6E) generated by breakage of the amide bond are generally favoured. MS/MS data from ESI is often superior to MALDI MS/MS as ESI generates multiply charged peptides, the fragmentation of which in turn generates more ionic species than from singly charged peptides.

Fragmentation massively increases spectral complexity and therefore needs to be exclusively applied to the selected species, otherwise the spectra generated will contain so much data that an ID can not be assigned. Fragmentation itself is normally achieved via collision induced dissociation (CID). This involves colliding the analyte with another molecule (usually an inert gas e.g. N_2 or argon). To do this the ion must be isolated, fragmented and the resultant ions detected. MS/MS can be achieved in a number of ways inherent to the design of each MS machine.

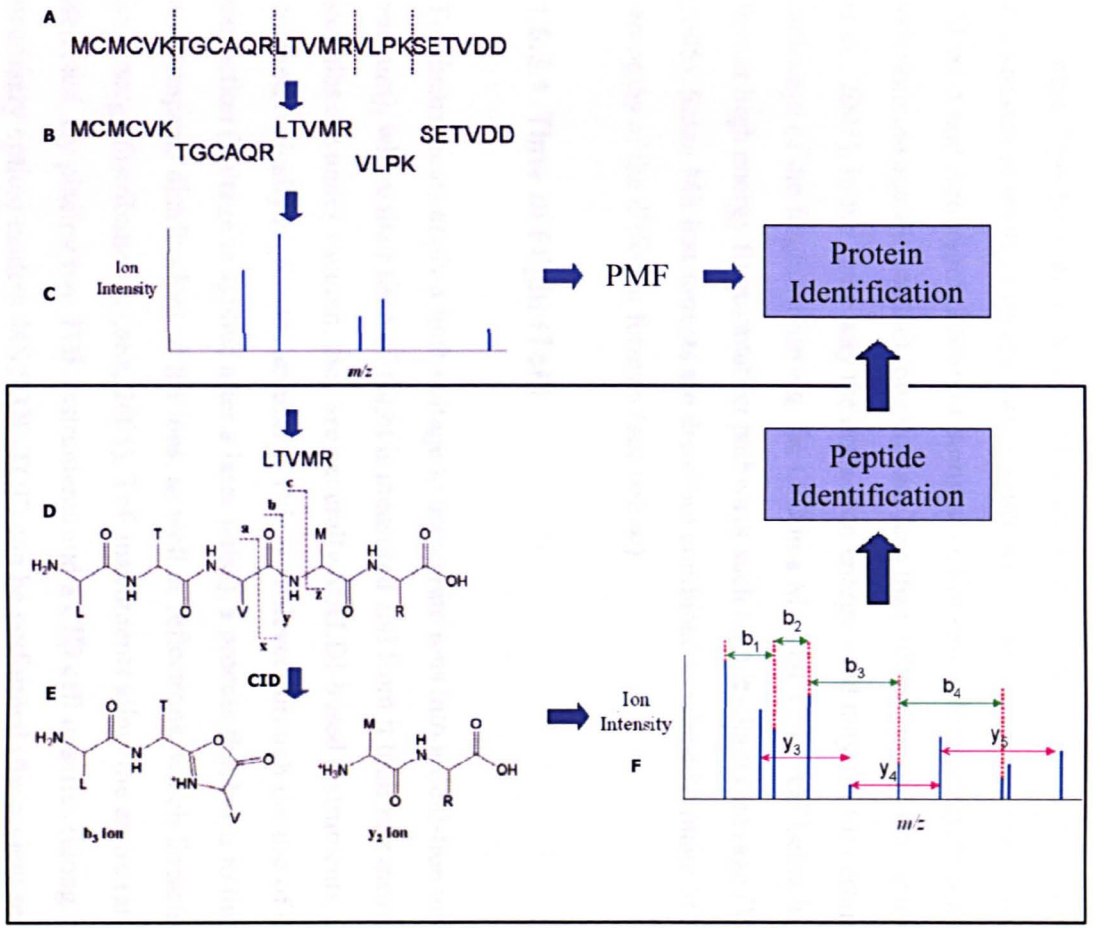


Figure 1.6 – Standard MS and tandem MS (boxed) proteomic analysis workflows. Firstly sample protein is cleaved (A) e.g. with trypsin, to produce peptides (B). These are analysed by MS (C) to generate a peptide mass fingerprint (PMF) for protein ID assignment. Alternatively in tandem MS, individual peptides (D) are selected for fragmentation by collision induced dissociation (CID). Fragmentation can occur anywhere along the peptide backbone (---), with the resultant ions named accordingly. For simplicity a single *b* and *y* ion pair is shown (E). Daughter ions are detected (F) to enable peptide and subsequently protein identification.

1.6.3 MS Instruments

The key component in any MS machine (apart from the ionisation source – see section 1.6.1) is the mass analyser, which enables the selection and detection of target ion species. There are many types of MS instruments, although these can be summarised into four types - In the interests of brevity, Fourier transform ion cyclotron resonance (FT-ICR) MS machines will not be discussed here. No single MS machine can perform every type of analysis and the use of more then one instrument is likely to provide greater sample coverage. This is because the type of instrument used will have both strengths and weaknesses in terms of its resolution, sensitivity, mass accuracy, and dynamic range (see (Domon and Aebersold, 2006) for summary). For example quadrupole ion traps discriminate against product ions that are less than 30% of the precursor ion *m/z* (Wysocki *et al.*, 2005). In the same way the activation energy used may favour certain cleavage pathways of the fragment ion e.g. the CID in a MALDI TOF-TOF being high energy may favour high energy fragmentation pathways such as side-chain cleavage (Wysocki *et al.*, 2005). Some MS instruments are therefore combined to take advantage of the relative strengths of the different formats (see below).

1.6.3.1 Time of Flight (ToF)

ToF instruments apply a high voltage to accelerate ions into a field-free tube (i.e. under vacuum), where their time-of-flight is measured and from it their *m/z* measured. Because samples are under vacuum, they are generally MALDI-based instruments, and have a characteristically high mass accuracy. This is achieved through the use of delayed extraction (voltage is applied after a laser pulse), a process that helps to linearise the spatial and temporal distributions of the ions, as well as reflectrons, which linearise the various ion energy distributions (Lane, 2005). ToF instruments allow the entire range of ions to be detected. By placing two TOF instruments and a CID cell in series (along with the necessary optics) tandem MS (TOF-TOF) can be performed (for review see (Vestal and Campbell, 2005)).

1.6.3.2 Quadrupoles

A quadrupole instrument consists of four exactly parallel rods arranged around a central axis through which the ion stream passes. There are two positive rods (facing each other) and two negative rods. Both pairs have a direct current (DC) as well as an alternating current or radio frequency (RF) voltage component (Lane, 2005). A high (positive rods)

and low (negative rods) *m/z* filter can be applied to the ion beam, by manipulating the voltage (higher voltage increases the *m/z* value) and DC / RF (resolution) components. Similarly to TOF instruments, multiple quadrupoles can be put in series to perform tandem MS. In this case, three units are required (triple-quadrupole), where the first and third scan the ion beam, while the second performs CID. Quadrupoles can even be combined with TOF instruments to generate a hybrid Q-ToF. Such a machine is extremely versatile as it is suitable for ESI or MALDI generated ion sources (for review see (Ens and Standing, 2005)).

1.6.3.3 Ion Traps

Ion traps, as the name suggests, trap ions in a defined space. These machines have been referred to as 'quadrupole ion traps' as they can be thought of as a reworked quadrupole (Lane, 2005). Like standard quadrupoles, ions are manipulated by altering the RF frequency applied. However *m/z* ratios are measured through the selective ejection of ions into the detector. Therefore if a particular *m/z* ratio is left unobserved, over time ions within that *m/z* range will accumulate. This retention of ions enables analytes present at low concentration to be pooled. Unlike other MS machines that require multiple units to perform MS analysis, by manipulating the radio frequency and voltage across the trap, all ions bar the target species can be selectively discarded. The target can then be fragmented and all daughter ions measured by the detector. This allows ion traps to perform multiple rounds of fragmentation (MSn).

Ion traps do suffer from low mass accuracy, and although linear ion traps (larger volume) have improved this (Lane, 2005), again a hybrid system with quadrupoles offers many more benefits (Hopfgartner *et al.*, 2004).

1.6.4 Proteolysis

Before MS analysis, proteins are cleaved into a series of peptides. Proteolysis can either be specific i.e. cleavage occurs after a particular residue or sequence, or non-specific. The two will produce two different peptide populations, with the latter generating a far more complex population of overlapping peptides. This can be an advantage as it provides increased sequence coverage, however increased computing power is needed to cope with the sample complexity (Wu and Yates, III, 2003). Cleavage is therefore conventionally performed using the enzyme trypsin due to its high cleavage specificity. This high specificity enables the compilation of *in silico* generated peptide spectra for comparison

and therefore protein identification. Trypsin cleaves at arginine and lysine residues, except if followed by proline. This means that the amino terminal residue of all tryptic peptides is basic, and as a result these often carry a double or triple charge, which is good for tandem MS. Other proteolytic agents include enzymes e.g. proteinase K and chemicals e.g. cyanogen bromide. Each cleaves at different amino acid residues, with the relative abundance of these residues in any organism therefore having a bearing on the peptide sizes generated by the different proteolytic agents.

When considering the proteolysis of native membrane proteins it is possible that enzymes may have difficulty accessing all cleavage sites, particularly if they are located in a transmembrane α-helix. In this case, steric hindrance may prevent complete proteolysis and therefore a reduction in protein coverage. Small molecules e.g. CNBr, or alternative proteolysis conditions e.g. trypsin digestion in 60% MeOH may be pursued to counteract this effect.

1.6.5 Data Analysis

There are a number of different programmes for automatically assigning protein identifications from MS or MS/MS spectra. Two of the most widely used are SEQUEST (http://www.thermo.com) and MASCOT (http://www.matrixscience.com/). Although the details vary from program to program, they all compare the experimentally derived mass spectrum against a theoretical spectrum. When using MS data this consists of identifying all peptide masses in a sample and then determining all proteins that could have generated that specific series of peptide masses. This is known as peptide mass fingerprinting (PMF) and when analysing simple samples a diagnostically unique series of peptide masses and therefore the protein identity can usually be identified. However as sample complexity increases, PMF searching becomes unreliable, if at all possible, and has therefore been largely superseded by tandem MS analysis. When searching tandem MS data, a number of different models (classified as descriptive, interpretative, stochastic or probability-based matching) can be used (see (Sadygov *et al.*, 2004) for review). For example, the fragmentation ion series of all peptides with a similar mass to the parent ion can be compared. Alternatively the ion intensity of the theoretical spectrum can be calculated based on peptide fragmentation theory or empirically derived data. Whichever method is used, deriving a protein identity from high quality tandem MS data is relatively straightforward. However dealing with lower quality data is far more problematic and must take into account the limitations of the MS instrument used. Care must also be taken to

apply the correct search criteria to reduce false positives and negatives (Keller *et al.*, 2002).

1.6.6 Post-translational modifications

The positive detection of a protein within a simple sample is fairly trivial based on identifying a few peptides unique to that protein. However, many proteins undergo post-translational modifications (PTM) e.g. phosphorylation or glycosylation. These may be very important in regulating protein function. Therefore while total protein content may remain level, the functional species may vary tremendously. Discriminating between the two (or more species) by MS is not possible unless the peptides incorporating the PTM are detected by the MS (Guerrera and Kleiner, 2005). Even if a PTM peptide is detected it may be very difficult to identify that a PTM is present, particularly if the stoichiometry is low or if it is a complex sample. This is because many search programs will be unable to process the additional potential spectral complexity of PTM's. By including a range of PTM's in the searches (and there are a huge number of them) particularly in combination with poor data, protein ID assignment is likely to be reduced (Mann and Jensen, 2003). One therefore has to generally look for a PTM to see it. However some PTM's e.g. phosphorylation, appear to be 'more important' than others, although this may simply reflect current understanding. Many researchers have therefore focused on isolating these 'more interesting' species (Mann and Jensen, 2003). This can be achieved through the chemical derivitisation of the modifying group to allow subsequent affinity purification (Guerrera and Kleiner, 2005). An alternative to enrichment is to perform two rounds of MS. Initially unmodified peptides are searched to generate a list of protein identities in the first round. This list is then used in the second round to search for all possible modifications of the proteins on the list (Lane, 2005). Care must also be taken to ensure that the mass differences observed are not artefacts that are introduced during the analysis e.g. free cysteine residues can react with a variety of compounds to effect a mass change (Wysocki *et al.*, 2005).

1.6.7 Quantitative analysis

All of the above techniques have focused on the qualitative identification of proteins. However to investigate the cellular response to a defined variable e.g. the introduction of a drug, a quantitative analysis must be performed, whereby two cell populations can be quantitatively compared. Historically this was achieved by running each sample on a separate 2D gel. After staining, the spot patterns could be compared and quantitative

differences estimated. Unfortunately gel-to-gel reproducibility often leaves much to be desired, making it difficult to reconcile multiple gel spot patterns and therefore preventing any firm conclusions from being drawn. As a result, an increasing number of techniques are becoming available that allow the differential labelling of two samples. While labelling can be performed using a variety of tags and targets, the key point is that they allow the two samples to be mixed and analysed in combination thus enabling direct comparisons between them to be made. Labelling can take place at any stage during sample preparation e.g. in culture, post sub-cellular fractionation etc. The earlier that labelling is performed the less likely that artefacts will be introduced, as samples can be combined post-labelling, and all subsequent steps performed on the mixture.

One technique, called stable isotope labelling with amino acids in culture (SILAC) labels cell populations while in culture (for review see (Julka and Regnier, 2005)). It does this by growing the two cell populations in defined media, one of which exclusively contains an isotopically 'heavy' essential amino acid, while the other contains the corresponding isotopically 'light' essential amino acid (Ong *et al.*, 2002). By using an essential amino acid and growing cells long enough in the defined media, all proteins (and therefore all peptides) from each population that contain one or more residues of the essential amino acid will be tagged as 'light' or 'heavy'. This complete labelling of sample allows smaller samples to be used than a non-saturation labelling technique. However there are a few drawbacks to this technique. Firstly, the spectra produced from the samples will be more complex due to the additional isotopic peaks, often making it necessary to perform additional fractionation to enable spectral deconvolution. In addition if labelling is not 100% in the 'heavy' sample, then any unlabelled 'heavy' sample will add to the 'light' (Ong *et al.*, 2003a). However the biggest drawback to this technique is its expense, and as a result has in the main only really been successfully applied to mammalian cell lines (Ong *et al.*, 2003b), although recent papers show it can be applied to *A. thaliana* (Gruhler *et al.*, 2005), *C. elegans* and *D. melanogaster* (Krijgsveld *et al.*, 2003).

There are now a large number of ways to chemically modify a protein sample. For example, isotope-coded affinity tagging (I-CAT) reagents react with free cysteine residues creating 'light' and 'heavy' mass tagged samples (Gygi *et al.*, 1999a), with a solid-phase labelling format also having been developed (Zhou *et al.*, 2002). One advantage of the I-CAT technology is that the label contains a biotin moiety allowing the selective purification of labelled peptides, thus reducing sample complexity. This is likely to improve proteome coverage, although proteins that do not contain a cysteine obviously cannot be detected. Fortunately this is unlikely to be a big problem e.g. 96.1% of all genes

in the human proteome contain a cysteine residue (Yan and Chen, 2005), therefore to increase proteome coverage other techniques targeting more abundant residues may need to be used. For example I-TRAQ labels lysine residues and terminal amine groups (Zieske, 2006), however targeting more abundant residues can lead to over-complex spectra, requiring a balance to be struck between complexity and information yielded. To address this issue an I-TRAQ like system that only targets the terminal amines has also been devised (Kuhn *et al.*, 2005). There are now four different I-TRAQ mass tags allowing multiple sample analysis (Yan and Chen, 2005). Another interesting system involves trypsin digesting one sample in ^{18}O enriched water. During peptide bond hydrolysis, the ^{18}O is transferred to the peptide thus labelling it (Wang *et al.*, 2001). An alternative approach to mass tagging is difference gel electrophoresis (DiGE), that involves tagging protein samples with different charge and mass matched fluorophores (for overview see (Wu, 2006)). Multiple samples can then be run and fluorescence intensity compared from a single gel (Ünlü *et al.*, 1997). Being a gel-based separation technique, there is the added benefit that different species of the same protein arising from post-translational modifications are likely to be resolved on the gel due to a change in mass / pI.

All of the above techniques are excellent in defining the relative change in protein expression, but they are unable to give precise information regarding the absolute level of a protein. Absolute quantitation (AQUA) was therefore developed and involves the addition of a defined amount of synthetic labelled peptide that is diagnostic for a specific protein. The synthetic peptide acts as an internal calibrant to define the amount of endogenous protein in the original sample (Gerber *et al.*, 2003). These peptides are however expensive and therefore can only be used in a focused rather than proteome wide analysis.

It is worth bearing in mind that with any quantitative approach where the comparative samples undergo manipulation in isolation, it is possible to introduce variability, although the high reproducibility of most of these techniques does not preclude such an approach.

1.6.8 Target Validation

In summary it can be seen that proteomics represents a suite of techniques to interrogate organisms at the molecular level. It is a particular powerful technique in terms of its ability to define quantitative differences across the entire proteome of two cell populations. However as with any technique it is not infallible and does require confirmation (e.g. by western blot) and phenotype validation (e.g. by RNAi knock-down) post analysis.

1.7 Parasite proteomics

Finally it is worth mentioning the power of proteomics with regard to developing therapies to parasitic diseases. Most parasitic diseases (and HAT is a classic example) affect the poorest people of the developing world. Their inherent lack of money means that diseases like HAT will never attract large scale investment from the pharmaceutical industry. Any money put into studying these diseases must therefore be fully utilised (Cowman and Crabb, 2003). Value for money and clinically relevant results both in terms of prolonging the life of a therapy and developing novel therapies are therefore of key concern (Klokouzas *et al.*, 2003). Analysing drug resistance mechanisms is likely to fulfil both criteria. Firstly by understanding a resistance mechanism, diagnostics tests can be developed to enable appropriate treatment regimes to be implemented on a patient-to-patient basis. Indeed, the identification of TbAT1 as a factor in arsenical resistance has led to the development of a diagnostic test capable of detecting the presence or absence of a functional TbAT1 gene (Stewart *et al.*, 2005). If possible, potential modulators of resistance may also be gained once the underlying molecular mechanism has been elucidated (Pradines *et al.*, 2005). In cancer therapy, this approach has undergone various trials and while the clinical results have so far been disappointing (Fojo and Bates, 2003), it should be easier to specifically target parasitic cells rather than cancer cells (Pradines *et al.*, 2005). Finally, analysing drug resistance mechanisms will inevitably lead to the identification of new drug targets. As a by-product of pursuing a proteomic approach, a large amount of information regarding the protein composition of the *T. brucei* plasma membrane will be generated. This will range from simple confirmation of the expression of ORF's identified in the genome as putative ORF's through to more relevant identification of which proteins are expressed in a particular lifecycle stage. Defining the plasma membrane proteome in this way is analogous to defining the *T. brucei* genome sequence in terms of the information it is likely to yield.

1.7.1 Investigating resistance mechanisms

There are a huge number of ways in which drug resistant mechanisms can be investigated, particularly when working with a system where resistance has been induced under controlled conditions *in vitro*. Strains developed *in vitro* from a clonal parent, should be truly isogenic i.e. identical to the parent line except in relation to changes relating to its ability to survive in the presence of drug. Free from the difficulties of working with field-derived strains, it is possible to compare isogenic lines. A number of options were

available to us to investigate the acquisition of resistance, from classical genetics whereby genes are positionally cloned, through to proteomic approaches. Each approach offers advantages and disadvantages, and some have already been successfully applied to *T. brucei* (Mäser *et al.*, 1999), however a number of mechanisms remain unknown and are unlikely to be uncovered by repeating previous work. To that end a novel approach utilising proteomics was investigated. The aim of this approach was to develop isogenic pentamidine-sensitive and resistant lines from both wild-type s427 strain and the TbAT1 KO strain generated from s427. Pentamidine resistance is thought to be multifactorial. So far though only TbAT1 has been demonstrated to be involved, although the deletion of this gene only resulted in a 2 to 3-fold loss of sensitivity. However by using the TbAT1-KO strain as the parent, it is anticipated that higher resistance levels may be induced more easily than in a wild-type parent. Defining the mechanism(s) of resistance may also shed light on the antiparasitic action of pentamidine as well as possible reasons for high-level pentamidine resistance not being observed in the field.

When looking at mechanisms of resistance, it is important to use isogenic lines. These lines consist of a drug resistant line, generated by the application of a selective pressure (i.e. sub-curative drug doses), and its parental line. In theory this isogenic pair will only differ in one respect – their sensitivity to a chosen drug, thereby allowing easier identification of a drug resistance mechanism.

1.8 Aims

There is a desperate requirement for more effective HAT therapies in the field due to increasing drug resistance. There are very few drugs being developed and novel approaches are therefore required to tackle HAT. The aim of this project is therefore to use proteomics, particularly aimed at the plasma membrane to try and define the mechanism(s) of pentamidine resistance in *T. brucei* parasites. Our aims can be summarised as follows:.

- Develop / employ proteomic techniques to specifically analyse membrane proteins
- Define a *T. brucei* plasma membrane proteome
- Generate isogenic pentamidine resistant and sensitive lines
- Perform a quantitative proteomic analysis on the isogenic lines

- Confirm any target proteins identified from the isogenic lines by conventional technologies e.g. RNAi

Chapter 2

Analysing the *T. brucei* genome for rare polytopic membrane proteins

2.1 Introduction

T. brucei has been the focus of prolonged intense biochemical and genetic study, culminating in the publication of the *T. brucei* genome in 2005 (Berriman *et al.*, 2005). It was therefore a surprise that of the 9-10,000 open reading frames (dependant on genome release version), approximately 50% had no known function. This suggested that there are as yet a huge number of pathways and functions that remain unknown. It was therefore highly likely that many of these proteins would be encountered in an analysis of the trypanosome plasma membrane sub-proteome. Characterising membrane proteins is highly challenging, and we therefore needed a way to assess our success in identifying this class of proteins. In a similar manner when investigating the pentamidine resistant strains, we wanted to exclude proteins not associated with the phenotype. To do this we therefore needed to try and classify all proteins, even if they were annotated as having an unknown function. Fortunately a huge number of bioinformatic approaches exist that harness advanced computing power, apply it to large datasets e.g. an entire genome, and from primary DNA sequence alone make various predictions. In terms of defining the trypanosome membrane proteome, we needed to be able to predict all integral membrane proteins, together with their relative expression levels. This information would then allow us to measure the efficiency with which our proteomic analysis had covered the potential membrane protein repertoire. In the same way, when asking biological questions in a quantitative analysis (e.g. of drug resistant / sensitive lines), this would allow us to reduce the number of targets based on their properties. For example where a transporter was thought to be involved in the drug resistance phenotype, all non-polytopic proteins could be excluded.

2.1.1 Transmembrane Prediction

Perhaps the most useful bioinformatic tool that has been developed for membrane protein characterisation is transmembrane prediction software. A number of these programmes have been developed, all with the aim of successfully predicting the presence of a

transmembrane domain. They do this by looking for two things. The first attempts to define membrane spanning regions i.e. hydrophobic regions that would interact favourably with the lipid environment. This can be achieved in a number of ways, but has classically focused on using a very simple sliding window approach, where residues are assigned a hydrophobic value from a weighting matrix (Edelman, 1993) or neural network (Casadio *et al.*, 1996). Areas that satisfy a hydrophobic criteria e.g. +1.6 hydrophobicity, for a particular length e.g. 15-25 residues (although many fall outside this accepted length (Chen and Rost, 2002)), then qualify as a transmembrane domain (TMD). Another 'clue' to identifying TMD's includes the association of positive charge with the cytoplasmic side of the protein (von Heijne, 1994). Many analyses are now employing even more sophisticated knowledge of TMD signatures e.g. preference for amphipathic aromatic amino acids at the membrane solute interface (Ulmschneider and Sansom, 2001).

Knowledge of these TMD signatures has given rise to a plethora of prediction methods, many of which have been evaluated against each other (Cuthbertson *et al.*, 2005; Möller *et al.*, 2001). Unfortunately, no polytopic *T. brucei* membrane protein structure has been solved yet, although the structure for the membrane-anchored VSG has been (Blum *et al.*, 1993; Chattopadhyay *et al.*, 2005). There is therefore no polytopic protein dataset of known TMD's against which the various prediction software programmes can be evaluated. In the same way the lipid composition of the *T. brucei* membrane has not been extensively studied. However there is no reason to suppose that TMD predictions cannot be made in *T. brucei* using these methods. However no single prediction method is infallible (Möller *et al.*, 2001), and for maximum accuracy it is therefore advisable to use more than one programme to aid in assigning TMD's. To that end, we decided to test two Hidden Markov Models (Ikeda *et al.*, 2002; Tusnády and Simon, 1998), that consistently appear to be the best prediction programmes (Cuthbertson *et al.*, 2005). It is worth noting that none of the above algorithms are able to predict the presence of β-barrel TMD's (although prediction software does exist (Bigelow *et al.*, 2004)) and they are therefore exclusively concerned with α-helix TMD predictions. However β-barrel TMD's have only been observed in bacterial outer membrane proteins (Schulz, 2002), and none have so far been found in *T. brucei*.

2.1.2 Codon Adaptation Index

DNA's limited four base repertoire requires a DNA triplet to encode each of the 20 amino acids. As a result, each amino acid may be defined by one or more codons e.g. methionine

is encoded for exclusively by one codon (ATG), whereas there are six different codons that encode the amino acid leucine. These synonymous codons are completely interchangeable and doing so does not affect protein sequence. However, certain synonymous codons have been observed to be preferentially used in high abundance proteins. The association between the synonymous codon usage in a protein and its expression level is known as a translational bias. If present, the translational bias can be used to predict any proteins expression level. To make these predictions quantitatively, the codon adaptation index (CAI) was developed (Sharp and Li, 1987). This index uses a reference set of highly expressed proteins (the original set in yeast comprised of 16 ribosomal proteins, 7 enzymes and one elongation factor (Sharp *et al.*, 1986)). Each codon in this set is assigned a 'weight' (Equation 2.1) based on the observed frequency against the expected frequency (assuming equal usage of all synonymous codons).

$$RSCU_{ij} = \frac{x_{ij}}{\frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}}$$

Equation 2.1 - Calculation of relative synonymous codon usage (RSCU). Where x_{ij} is the occurrence of jth codon for ith amino acid. From (Sharp and Li, 1987).

This reference table can then be used to describe the 'relative adaptiveness' of any codon. The 'relative adaptiveness' of any ORF is determined by calculating the mean RSCU score for each codon in that ORF, and dividing by the maximum possible CAI score i.e. using $RSCU_{imax}$ for each codon.

$$w_{ij} = \frac{RSCU_{ij}}{RSCU_{imax}} = \frac{x_{ij}}{x_{imax}}$$

Equation 2.2 - The relative adaptiveness of a codon (w_{ij}) is defined by its relationship to the most frequently used synonymous codon (imax), from (Sharp and Li, 1987).

In *S. cerevisiae*, where most protein expression levels have been experimentally determined (Ghaemmaghami *et al.*, 2003), CAI analysis was shown to reliably predict protein expression levels. CAI analysis thus represents a powerful tool, but it does make two major assumptions. The first is the assumption that there is a translational codon bias in the genome under analysis. Calculating CAI alone gives no indication as to whether this is a valid assumption. Indeed, considering the number of genomes that have been / are being sequenced, there may be little corroborating biological evidence to test the confidence of the results. The second assumption is that the chosen reference set of highly expressed proteins is representative of the codon bias i.e. that their codon usage is typical

of highly expressed proteins - again this may be hard to validate. To address these concerns, a relatively new approach using the CAI algorithm (albeit a revised version that accounts for the increased gene length in eukaryotes) was employed that aims to detect the existence of a dominating codon bias, by iteratively finding a set of genes (S) that have the highest CAI values (Carbone *et al.*, 2003). By consistently detecting the same S, the presence of a codon bias and the identity of a representative set of genes expressing that codon bias can be confidently assumed. Once S has been identified, biological data is needed to determine whether the bias is translational or not. Therefore if S is comprised of genes known to be highly expressed the bias can be assumed to be translational.

Interestingly the molecular basis of a translational CAI bias remains enigmatic (Sharp *et al.*, 1993). It has been postulated that the different codons have higher or lower efficiencies in binding and effecting protein translation i.e. high efficiency codons are reserved for high copy number genes to enable sufficient expression. Alternatively, it is possible that the bias merely reflects the relative levels of the different tRNA species e.g. rarely used codons have correspondingly low tRNA levels. What is clear however is that whatever the mechanism a large number of genomes display a codon bias. Whether this codon bias always gives rise to a translation bias seems to vary in different organisms (Gygi *et al.*, 1999b; Stenoien, 2005).

2.1.3 Other Indices

A number of other indices have been proposed such as the Grand Average score of Hydrophobicity: GRAVY (Kyte and Doolittle, 1982). This defines the hydrophobicity of a protein based on a composite score of the hydrophobicity of each amino acid, and has previously been used in proteomic investigations (Marmagne *et al.*, 2004). However GRAVY scores can only indicate the overall nature of the protein and not its membrane association. Indeed many of the most hydrophobic proteins contain none or very few TMD's (Blonder *et al.*, 2002). It is therefore potentially useful in evaluating the compatibility of a technique with hydrophobic proteins, but in the context of this analysis is unlikely to be very informative. Therefore in the interests of being able to simplify and summarise data, we decided to limit the number of parameters associated with each protein, and concentrated solely on TMD and CAI predictions.

2.2 Materials and Methods

Although a number of different *T. brucei* genome releases were used during these investigations, all gene ID's used in this thesis are derived from the latest *T. brucei* genome release (version 4 downloaded from ftp://ftp.sanger.ac.uk/pub/databases/T.brucei_sequences/ T.brucei_genome_v4/), to prevent confusion. However, it is worth noting that two files were used from the release. The first (TbProt) contains the amino acid sequence of every ORF predicted to be expressed by the Sanger Centre (9210 genes). The second (TbDNA) contains the DNA sequence of all predicted ORFs (11,049 genes). The gene number discrepancy is due to the inclusion of all pseudogenes and genes located on unordered contigs. This set is therefore likely to contain all potential genes, with some redundancy.

Various Perl scripts were written to concatenate, re-format, and extract proteins of interest during the various analyses.

2.2.1 Transmembrane prediction

The *T. brucei* genome was analysed using two TMD prediction programmes - HMMTOP (Tusnády and Simon, 2001) (run on a local server - http://fun-gen.gla.ac.uk) and TMHMM v 2.0 (Krogh *et al.*, 2001) (available from http://www.cbs.dtu.dk/services/TMHMM/).

2.2.2 CAI

gCAI values were calculated from TbDNA using a Java script (Carbone *et al.*, 2003) (http://www.ihes.fr/~materials/cai_dominant.html), while standard CAI values were generated from a codon usage table (produced using Emboss). Both programmes were implemented by Janssen Genomics (Glasgow, UK).

2.2.3 tRNA analysis

All genes annotated as a tRNA were identified from the *T. brucei* genome by searching for the word 'tRNA' using an in-house perl script.

2.3 Results

2.3.1 Transmembrane predictions

Analysis of all ORFs predicted to be expressed from the *T. brucei* genome using the two algorithms revealed few meaningful differences as shown in Figure 2.1. It can be seen that total numbers of proteins predicted to have up to 12 TMD's show a strong correlation between the two methods, with TMHMM consistently being slightly more conservative in its assignments. Above 12 TMD's, the two predictions vary more markedly.

To test whether individual proteins were assigned similar values from both algorithms and whether the predictions made sense biologically, all 43 predicted amino acid transporters (which are assumed to contain between 10 and 12 TMD's) were compared using the two methodologies. 32 of these transporters were assigned an identical number of TMD's, while of the remainder; only one differed by more than one TMD between the two values. As stated above, TMHMM was slightly more conservative with an average of 10.7 TMD's, compared to 11.1 from HMMTOP.

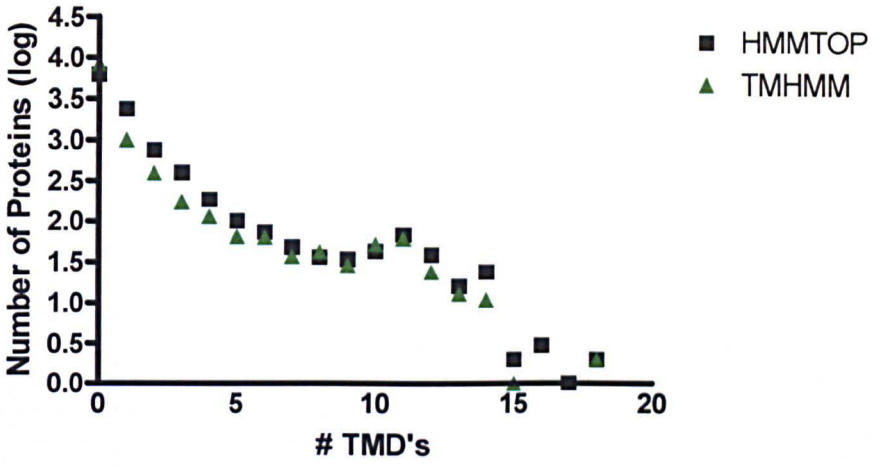


Figure 2.1 - Number of TMD's per protein as predicted by HMMTOP or TMHMM. Proteins predicted to contain 20 or more TMD's were excluded form the graph for clarity.

# TMD's	Number of proteins		Proteins with one or more predicted TMD's (%)	
	HMMTOP	TMHMM	HMMTOP	TMHMM
0	6194	8311	40.5	20.1
1	2380	1007	17.6	10.5
2	751	394	10.4	6.7
3	398	174	6.6	5.0
4	187	116	4.8	3.9
5	102	65	3.8	3.3
6	74	64	3.1	2.7
7	48	37	2.6	2.3
8	36	42	2.3	1.9
9	34	29	2.0	1.6
10	43	52	1.5	1.1
11	69	62	0.9	0.5
12	39	24	0.5	0.3
13	16	13	0.4	0.2
14	24	11	0.1	0.1
15	2	1	0.1	0.1
16	3	-	0.1	0.1
17	1	-	0.1	0.1
18	2	2	0.0	0.0
20	-	1	0.0	0.0
22	-	2	0.0	0.0
24	1	-	0.0	0.0
25	-	1	0.0	-
27	1	-	0.0	-
32	1	-	0.0	-
43	1	-	0.0	-
47	1	-	-	-

Table 2.1 - Number of TMD's predicted per protein for the entire *T. b. brucei* genome, using the HMMTOP and TMHMM algorithms.

2.3.2 CAI Values

A codon usage table was derived from Equation 2.1. However, instead of using a reference set (as in (Sharp and Li, 1987)), the entire putative proteome was analysed to compile a 'pseudo' codon usage table. From this table, a 'pseudo CAI' value was then generated for each ORF. These could then be ranked to enable easy identification (Figure 2.2A) of all ORFs with a synonymous codon usage that deviated from the normal codon usage observed across the entire genome. For example, in the *T. brucei* genome almost all genes had a ORF CAI value of 0.82 – 0.86, while some had values as low as 0.72 and as high as 0.9.

As outlined previously (Carbone *et al.*, 2003), an iterative approach was also implemented, with a total of 15 iterations undertaken. A stable set (S) was identified after 9 iterations and is summarised in Table 2.2 (see Appendix I for gene identities). S was then used to generate an RSCU table again allowing all ORFs to be ranked in terms of their codon usage bias (Figure 2.2B).

From these two analyses, it can be seen that there is a very clear codon bias present in the *T. brucei* genome. By plotting the percent usage of the various synonymous codons for a particular amino acid, information on the range i.e. the presence of a bias for that amino acid can be gained. For example the six codons that encode leucine show very strong preferential usage, whereas there doesn't appear to be a bias in the overall usage of codons encoding serine (Figure 2.3A). This doesn't appear to be related to the overall usage of an amino acid as all three of these amino acids are highly expressed throughout the genome (Figure 2.3B).

We also compared the number of tRNA's for each amino acid against the genome codon usage table, and noted that there was a correlation (Figure 2.4).

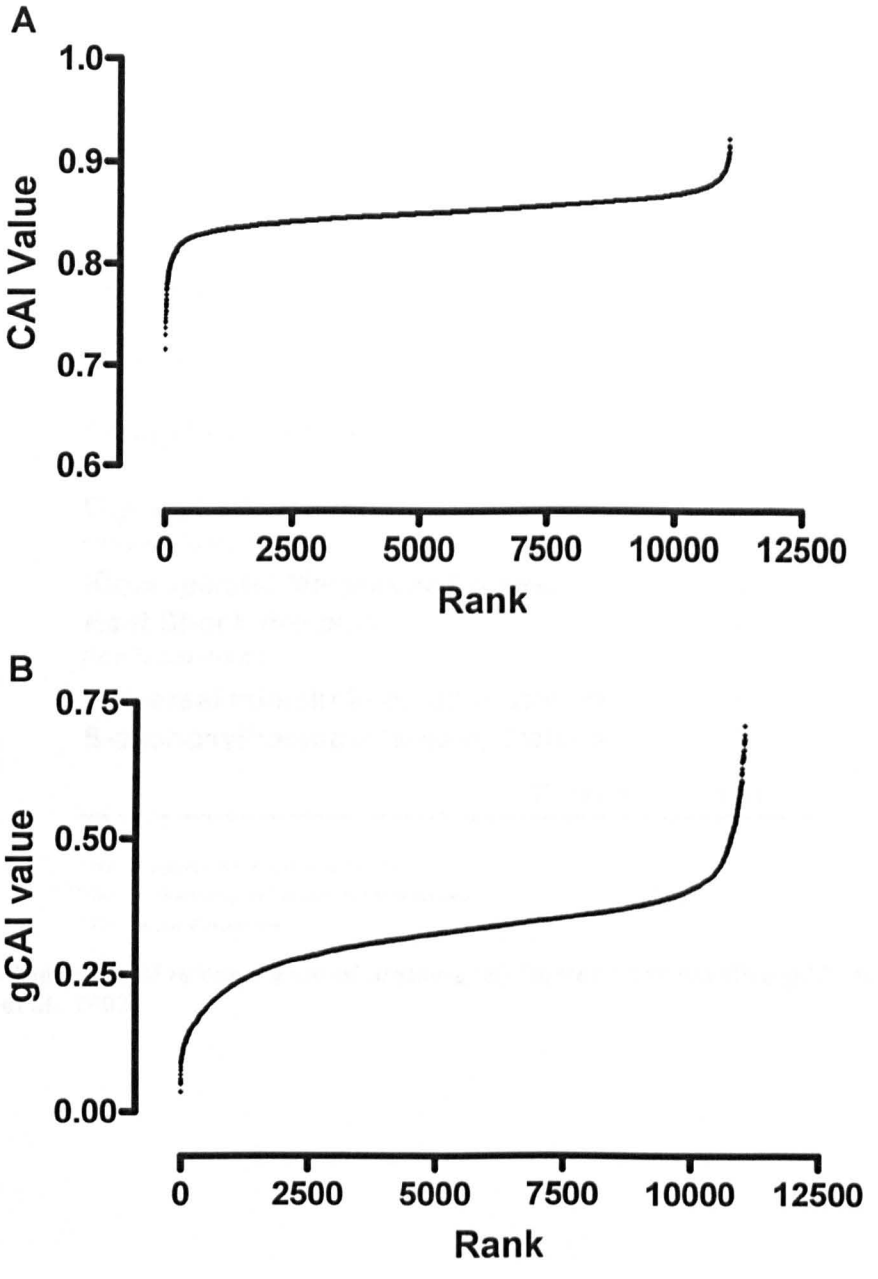


Figure 2.2 - CAI values for entire *T. brucei* genome calculated from the codon usage table (A) and from the iterative gCAI approach (B).

Type of Protein	Number
Ribosomal Proteins (40s, 60s, and S27a)	37
Histones (2A, 2B and 4)*	37
Paraflagellar Rod Proteins (69 and 73 kDa)	10
Tubulins (α and β)	9
Elongation Factors (1a and 2)	5
Glycolytic enzymes (Aldolase, GAPDH#, PFK+)	5
Kinetoplastid Membrane Protein	3
Heat Shock proteins (Hsp 70 and Hsp 83)	2
Universal minicircle binding protein	1
S-adenosylhomocysteine hydrolase	1
Total =	**110**

* Histone 3 genes are found in the top 3% gCAI

GAPDH - Glyceraldehyde 3 -phosphate dehydrogenase

+ PFK - Phosphofrructokinase

Table 2.2 - Summary of reference set of proteins (S) derived from iterative gCAI approach (Carbone *et al.*, 2003).

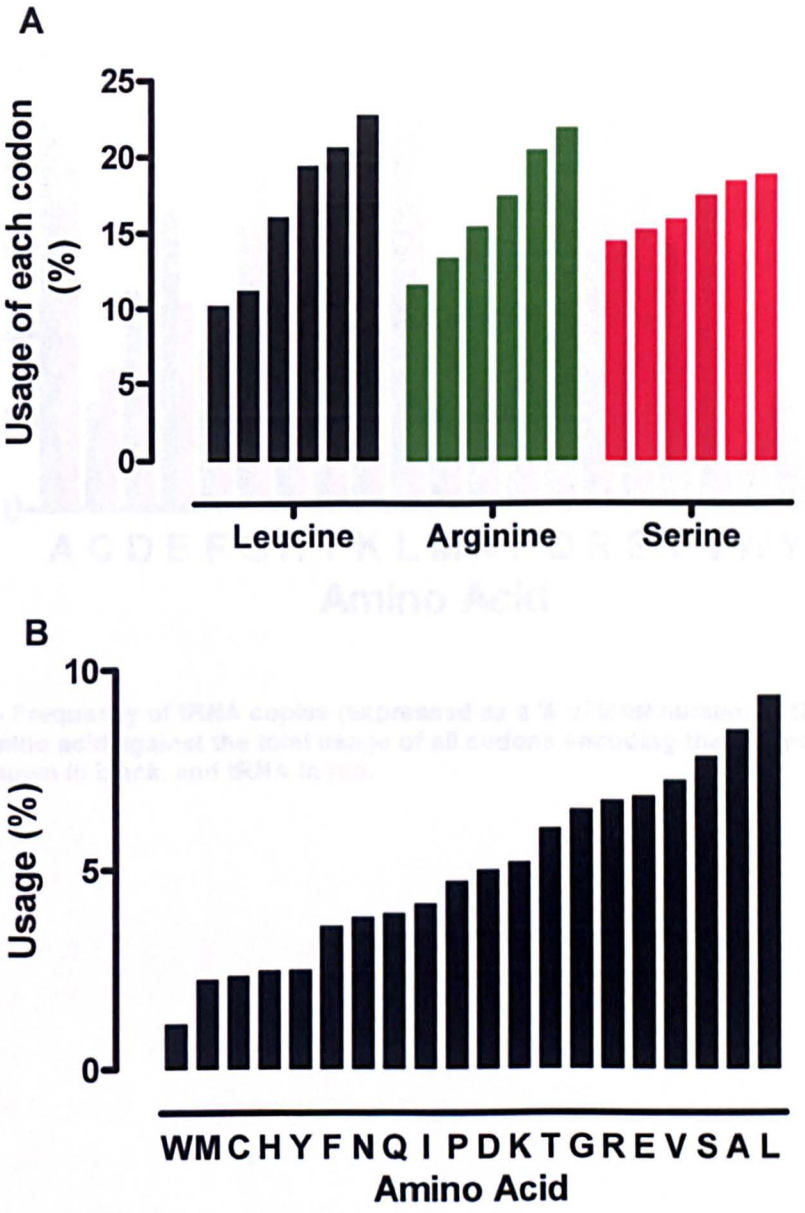


Figure 2.3 – Preferential synonymous codon usage for three amino acids (A) and overall amino acid usage (B). Values were generated using the gCAI approach that analyses the entire genome.

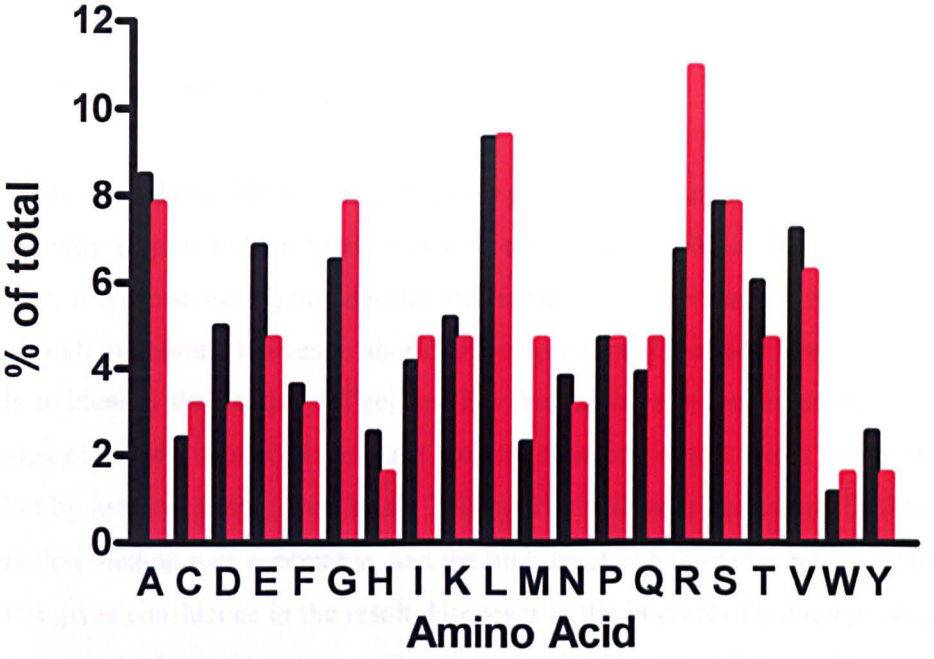


Figure 2.4 - Frequency of tRNA copies (expressed as a % of total number of tRNA copies) for each amino acid against the total usage of all codons encoding that amino acid. Codon usage is shown in black, and tRNA in red.

2.4 Discussion

2.4.1 Transmembrane Prediction

There are a number of different methods for predicting the existence of TMD's. The best consistently employ hidden Markov models and we assessed two of these for the *T. brucei* genome. It is worth noting that despite the fact that even the best prediction methods only identify all membrane helices in about 70% of proteins (Chen and Rost, 2002), they are likely to identify the majority of helices. For the purposes of this investigation the exact location of a TMD or indeed the exact number of helices (most errors under- or over predict by just one TMD (Chen *et al.*, 2002)) is by and large irrelevant. To that end, either prediction method was acceptable, and the high level of agreement between the two models gives confidence in the result. However in the interest of reducing list complexity, only one method could be chosen. To reduce the likelihood of false positive assignments, the more conservative TMHMM v2.0 algorithm was chosen as the prediction method of choice.

2.4.2 CAI Calculations

Like a number of prokaryotes, but unlike other eukaryotes, trypanosomes employ polycistronic transcription and regulate expression primarily at the post-transcriptional rather than translational level (for review see (Clayton, 2002)). Regulation appears to be conferred in the main by a 3' untranslated region (UTR) in the transcript. While this does not negate the potential for predictive information to be gained from CAI, the unusual nature of trypanosomes gene regulation (in relation to other eukaryotes) suggests that caution should be exercised in applying dogma derived from eukaryotic systems. Therefore before employing the predictive ability of CAI it was crucial to determine the presence or absence of a bias in the usage of synonymous codons in *T. brucei*. Two different methodologies ('pseudo-CAI' and iterative gCAI) were therefore employed. The 'pseudo-CAI' gives information on the frequency that each of the synonymous codons are used in the proteome. Due to the simple theory behind this analysis, only genes that have a gross non-normal synonymous codon usage are identified i.e. the range of values is compressed (see Figure 2.2A). However despite the naivety of this approach, it does strongly indicate the presence of a codon usage bias. The gCAI approach represents a far more refined iterative system that also identified a strong codon bias (see Figure 2.2B). The presence of a bias in codon usage could therefore be confirmed with confidence. We then also looked

at whether amino acid usage related to the number of tRNA's present in the genome and indeed, in general the more an amino acid is seen in the genome the more tRNA copies for that amino acid there are (Figure 2.4).

To identify whether the codon bias conferred a translational bias the stable reference set (S - summarised in Table 2.2), identified by the gCAI approach was investigated. Approximately equal numbers of H2A, H2B and H4 histones are present in S, although none have all members of each family present. These proteins along with a fourth histone - H3, form octamers which associate with DNA in complexes called nucleosomes (Felsenfeld, 1978). H3 was not found in the reference set S, however all members were found very close to S (within the 300 most biased genes). Similarly we looked at the tubulins that were identified, and indeed only the highly expressed α and β tubulin were identified in S, rather than other tubulins e.g. γ tubulin, which are known not to be highly expressed (Gallo and Precigout, 1988). The identification of kinetoplastid membrane protein in S was at first unexpected, but closer examination revealed that this group of proteins are also known to be highly expressed (Stebeck *et al.*, 1995). In summary we were confident that S consisted of a group of highly expressed genes. None were hypothetical, all are annotated, and all have been extensively studied. Certain multiple arrays of very similar / identical ORFs are over-represented in S, but this is also likely to reflect their translational bias. From this and the 'pseudo CAI' parameter (neither of which require nor use any biological information on gene expression), we suggest that CAI relates directly to a translational bias, and that the gCAI parameter can therefore be used to predict the expression level of any protein from the *T. brucei* genome.

It is worth noting that CAI values reflect a translational bias, but cannot predict when a gene may be expressed. For example, gene x and y have identical CAI values, however gene x is highly expressed during cellular stress, whereas gene y is only (highly) expressed in the absence of gene x. The application of this observation is important when considering that *T. brucei* has a digenetic lifecycle. Any protein identification technique is therefore highly unlikely to identify all proteins with a high CAI. However the predictive CAI value appears to be a strong indicator of a protein's expression level in *T. brucei.*

Chapter 3

Inducing pentamidine-resistance

3.1 Introduction

The aromatic diamidine pentamidine has been one of the mainstays of HAT treatment for the last five decades or so (Pepin and Milord, 1994). Our aim in this project was to identify proteins involved in pentamidine resistance. In terms of defining a clinically relevant resistance mechanism, it would be advantageous to use field isolates. Unlike many of the other trypanocides e.g. melarsoprol (Brun *et al.*, 2001; Ogada, 1974), or the closely related veterinary diamidine diminazene (Geerts *et al.*, 2001), these isolates are not available as resistance has not been observed in the field. Even if such resistant lines were identified, they must be adapted to axenic culture and are often difficult to maintain *in vitro*. More importantly however, when performing a comparative analysis isogenic lines are required, but are unlikely to be found in the genetically heterogeneous *T. brucei spp.* field population.

Even though field resistance remains conspicuous by its absence, there can be little doubt that resistance will ultimately arise with the continued use of pentamidine. Confirmation of this comes from the relative ease with which resistance has been induced by applying a drug selection pressure in the laboratory to generate isogenic lines (Berger *et al.*, 1995; Damper and Patton, 1976a). It would be naïve to assume that resistance can only occur via a single mechanism, and indeed a number of targets have been implicated in pentamidine resistance (Berger *et al.*, 1993). Of these targets, only TbAT1 has been definitively shown to be involved in resistance, although TbAT1 alone does not confer high-level resistance (Matovu *et al.*, 2003). This suggests that several genes are involved in resistance. In the same way we expect that lines undergoing multiple rounds of exposure to ever-increasing concentrations of pentamidine will acquire multiple adaptations of varying importance. To ensure that a further investigation of pentamidine resistance mechanisms would not just identify the link with TbAT1 again, we considered it prudent to develop isogenic lines from both the TbAT1 KO line and the wild-type s427 line from which it was derived. By using both lines we hoped to increase the likelihood of success in generating drug resistant strains and gain new insights in potential mechanisms of diamidine resistance.

Once generated, it was important to try and define the basis of resistance e.g. via a change in drug access to the cell, through altered access of the drug to its target or through changes to the intracellular target. To that end we needed to fully characterise the drug resistance phenotype. Considering the structural similarity of the diamidines we expected to observe a degree of cross-resistance. In contrast, we did not necessarily expect to see cross-resistance to the melaminophenylarsenicals such as melarsen oxide. These compounds are structurally distinct from the diamidines and while both are recognised by the P2 / TbAT1 transporter (De Koning and Jarvis, 1999), there was no evidence to suggest that they shared a second route of entry or drug target.

We were particularly interested in the involvement of membrane proteins i.e. transporters or efflux pumps, in the drug resistance phenotype. A wide range of diamidines, including the widely used DNA stain DAPI (4',6-diamidino-2-phenyl-indole, Figure 3.1A) and the early diamidine trypanocide stilbamidine (Fig. 1.1B), are transported across the plasma membrane into cells, where they bind to nuclear and kinetoplast DNA (Kapuscinski, 1995). Both of these compounds are strongly fluorescent and act as DNA probes, allowing convenient visualisation of both the nucleus and kinetoplast in trypanosomes. In terms of a functional analysis, monitoring the time taken for fluorescence acquisition can identify variations in membrane permeability to these compounds between different cell lines. Indeed, this forms the basis of a new diagnostic drug resistance test, assessing the presence or absence of P2 transport activity (Stewart *et al.*, 2005). Unfortunately pentamidine is one of the few trypanocidal diamidines that does not fluoresce. However, two other closely related diamidines do fluoresce; stilbamidine and DB75, whose structures are shown in Figure 3.1A. Previous work identified a reduced accumulation of a number of diamidines (including another DNA stain called Hoechst 33342; Figure 3.1B) in melarsoprol resistant *T. b. brucei* and *T. b. rhodesiense* (Frommel and Balber, 1987). Analysis of the accumulation of these compounds was postulated to identify differences in membrane permeability between the isogenic lines.

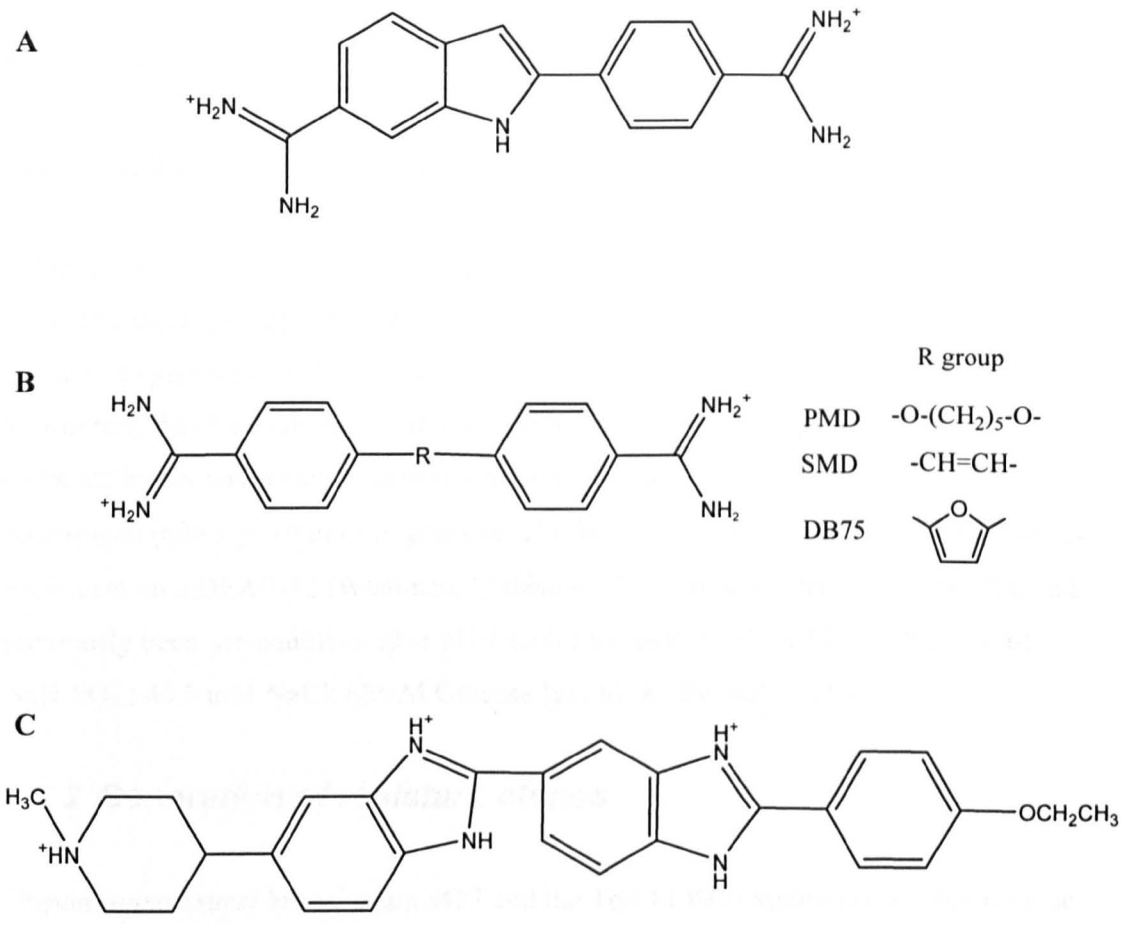


Figure 3.1 - Diamidine structures of selected trypanocides. (A) DAPI; (B) pentamidine (PMD), stilbamidine (SMD) and DB75; (C) Hoechst stain 33342. With the exception of pentamidine, all of these compounds are fluorescent.

3.2 Materials and Methods

3.2.1 Culturing trypanosomes

All blood stream form trypanosomes were cultured *in vitro* in HMI-9 media supplemented with 10% foetal bovine serum as previously described (Hirumi and Hirumi, 1989). For transport experiments, cells were either cultured on a large scale *in vitro*, or from infected Wistar rats. Adult female rats were infected by intraperitoneal injection, and blood collected by cardiac puncture at peak parasitaemia under terminal anaesthesia. Blood was centrifuged (600 × *g*, 10 min) to generate a buffy coat. Parasites were separated from the buffy coat on a DEAE-52 (Whatman, Maidstone, UK) anion-exchange column that had previously been pre-equilibrated to pH 8 with PSG buffer (45 mM Na_2HPO_4; 4 mM NaH_2PO_4 ; 43.5 mM NaCl; 60mM Glucose [pH 8]) as the mobile phase.

3.2.2 Generation of resistant clones

Trypanosoma brucei brucei strain s427 and the TbAT1 KO (Matovu *et al.*, 2003) clone derived from this strain, were selected for resistance to the diamidine drug pentamidine by stepwise selection *in vitro*. Initially, the parental strains were exposed to a range of pentamidine concentrations to identify the maximum tolerated drug concentration. During selection, cells were maintained in the maximal tolerated pentamidine concentration and in parallel exposed to double the drug concentration. As soon as a strain became viable in the higher concentration of drug, the strain was then maintained in that concentration and in parallel exposed to double the drug concentration. Strains were grown in 24 well plates in a volume of 1.5 ml. Cultures were diluted down to ~2 x 10^5 cells /ml in a fresh well whenever they reached 2-3 x 10^6 cells / ml. Cells were stabilated in liquid nitrogen at each stage. Additionally, strains were cloned when their resistance phenotype had increased 25 fold. Strains were cloned by limiting doubling dilution. This consists of taking a dilute culture (1 x 10^5 cells / ml) and serially diluting them two-fold across a 24-well plate. Isolated wells (in regards to the dilution series) were considered to contain a clonal population. From this work, four lines (labelled A to D) were initially established for drug selection for each of the two strains, thus aiming to provide a number of independently derived lines for analysis. Only a sub-set of the 8 lines initially established acquired a stable drug resistance phenotype (Figure 3.2).

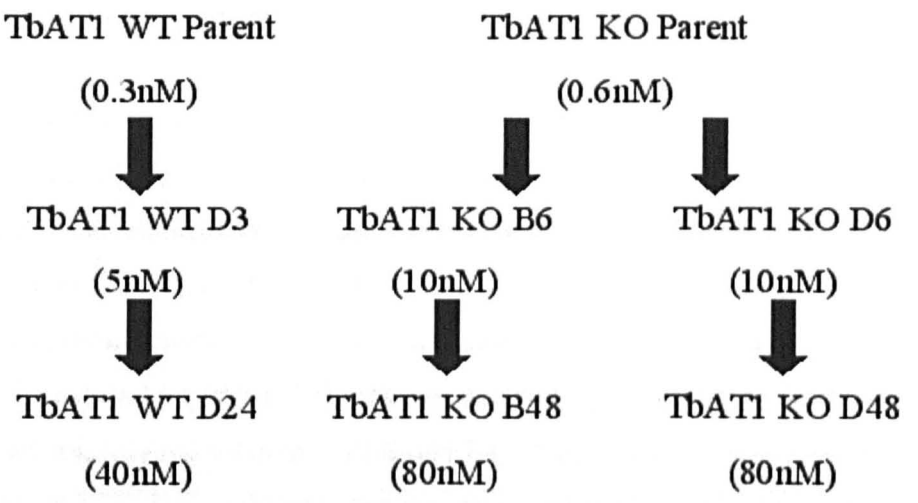


Figure 3.2 - Lineage of cell lines selected for drug resistance. The s427 line is labelled as TbAT1 WT in this figure. The drug concentration that each cell line was able to survive in is shown in parentheses and after each lettered clone (in ng/ml). Each arrow represents a number of rounds of constant drug selection followed by limiting dilution to clonality. Further selection pressure was then only applied to the subsequently derived clone.

3.2.3 Characterisation of lines

3.2.3.1 Alamar Blue assays

Alamar blue assays were performed as previously described (Räz *et al.*, 1997). Briefly, drugs were serially diluted in 100 µl of complete HMI-9 media across a 96-well micro-tire plate. Unless limited by solubility the top drug concentration used was 1 mM. BSF cultures grown to a maximum density of 2×10^6 cells/ml were washed twice in complete HMI-9 media, re-suspended at 2×10^5 cells/ml, and 100 µl added across the plate. Plates were then incubated for 48 h at 37°C, before the addition of 20 µl 10% Alamar Blue® or 5 mM Resazurin (Sigma) solution in PBS (pH 7.4). Plates were incubated for an additional 24 hours at 37°C, before fluorescence was measured using a LS 55 Luminescence Spectrometer (Perkin Elmer Instruments at 530 nm excitation and 590 nm emission wavelengths).

3.2.3.2 Growth Curves

Growth curves were performed *in vitro* using an initial concentration of 1×10^5 cells/ml. Parasite concentrations were determined using a haemocytometer.

3.2.3.3 Transport Assays

Transport assays were performed using a rapid oil stop protocol. Trypanosomes were harvested, washed twice in assay buffer (14 mM Glucose, 33.5 mM HEPES, 24 mM MOPS, 24 mM $NaHCO_3$, 4.6 mM KCl, 300 µM MgCl, 97.5 mM NaCl, 5.8 mM NaH_2PO4, 275 µM $CaCl_2$, 80 µM $MgSO_4$, pH 7.3), and resuspended at 1×10^8 cells/ml. Cells were then incubated with [^{3}H]-Pentamidine (3.4 TBq/mmol; Amersham, UK) in the presence or absence of an unlabelled inhibitor, or for varying lengths of time on top of a bed of oil. To arrest uptake, 1 ml of ice-cold 1 mM pentamidine in AB was added, followed by centrifugation (13,000g, 1 min) to pellet the cells below the oil layer. The cell pellet was collected by flash freezing the tube with liquid nitrogen and cutting off the base, including the cell pellet. Radioactivity measurements were made after solubilisation in 2% SDS, using liquid scintillation counting.

3.2.3.4 Uptake of fluorescent diamidines

T. brucei cells were pelleted (600 × *g*, 10 min at room temperature) before being re-suspended in fresh HMI-9 media containing DAPI, DB75 or Hoechst 33342. Cells were incubated at room temperature and assessed by fluorescence microscopy. Glass slides of the culture were prepared and visualised using differential interference contrast (DIC) and fluorescence microscopy ($\lambda_{ex=365nm}$, $\lambda_{em=445nm}$) on an Axioplan 2 imaging microscope (Carl Zeiss, Germany) using Volocity v 3.7 software (Improvision, Coventry). Where necessary, parasites were fixed by incubating in PBS containing 2.5% glutaraldehyde, for 20 min. Cells were then washed in 0.05 M glycine in PBS, before being mounted on slides.

3.2.3.5 *In vivo* characterisation

Two *in vivo* characterisation procedures were performed as follows, with assistance from M. Gould.

The first procedure consisted of infecting groups of 5 female Wistar rats (weighing ~200 g each) with 1×10^4 cells via i.p. injection. A total of 6 groups were infected (B6, Group A; B48, Group B; D6- Group C; D48- Group D; TbAT1 KO, Group E; TbAT1 KO - Group F). TbAT1 KO cells derived from continuous culture during drug resistance acquisition (group E) and cells grown up fresh from stabilates (group F) where tested.

The second procedure used female ICR mice (weighing ~30 g) that had been pre-treated with 250 mg/kg cyclophosphamide 24 hours prior to infection. This time, 3 groups (TbAT1 KO, Group A; B48, Group B; D48, Group C) of three mice were inoculated with 1×10^4 trypanosomes by i.p. injection.

Parasitaemia in both experiments was estimated daily by the rapid "matching" method (Herbert and Lumsden, 1976) from examinations of wet blood films on microscope slides, collected from tail pricks, using a phase-contrast microscope.

3.2.4 Data Manipulation

All graphical manipulations were performed using Prism version 4.00 software (GraphPad).

3.3 Results

3.3.1 Acquisition of resistance

A number of drug-adapted *Trypanosoma brucei* lines have been previously developed, e.g. (Carter and Fairlamb, 1993; Phillips and Wang, 1987; Scott *et al.*, 1996), including one developed for pentamidine resistance (Berger *et al.*, 1993). In this paper, a change in the intracellular drug target was postulated to be the basis of resistance. However this relied on indirect evidence as pentamidine's cellular target(s) has not been established (other than its ability to bind to DNA). We therefore decided to generate new isogenic pentamidine resistant cells, by pursuing a similar approach, but using two different parental strains (wild-type s427 and a genetically altered s427 clone where the TbAT1 gene had been knocked out (Matovu *et al.*, 2003)) on which to apply the selection pressure. Both strains were exposed to increasing concentrations of pentamidine during continuous *in vitro* culture. To maximise the potential for the successful generation of drug resistant lines, we applied our selection pressure to four independent cultures (labelled A to D) for each of the two parental strains. From these 8 cultures that were initially established, only 3 lines (TbAT1 WT D, TbAT1 KO B and TbAT1 KO D) went on to develop a stable high-level drug resistance phenotype. During selection an strain exhibiting an intermediate pentamidine resistance phenotype was also isolated (see Figure 3.2). The strains displaying the highest levels of resistance, remarkably, were able to survive in pentamidine concentrations eighty-fold higher than those tolerated by their parental strain. During each round of selection, it was observed that the initial acquisition of resistance in the presence of drug was associated with a reduced growth phenotype (Figure 3.3A), both in terms of doubling time and maximum cell density. This was reversed to a parental growth phenotype by maintaining the same selection pressure for an additional period of time (Figure 3.3B). Once the resistance phenotype was well established, removal of pentamidine from the media had no effect on growth. It was also observed that resistance acquisition was non-linear (Figure 3.4A). For example the B clone apparently increased from 20-fold to 89-fold more resistant than the parent within 48 hours. This suggests that a finite number of discrete adaptations were responsible for the total increase in resistance, with individual adaptations mediating each sudden increase in resistance. In a comparable study that generated diminazene resistance in *T. evansi* (Osman *et al.*, 1992), a similar trend of exponential increases in resistance were observed (Figure 3.4B).

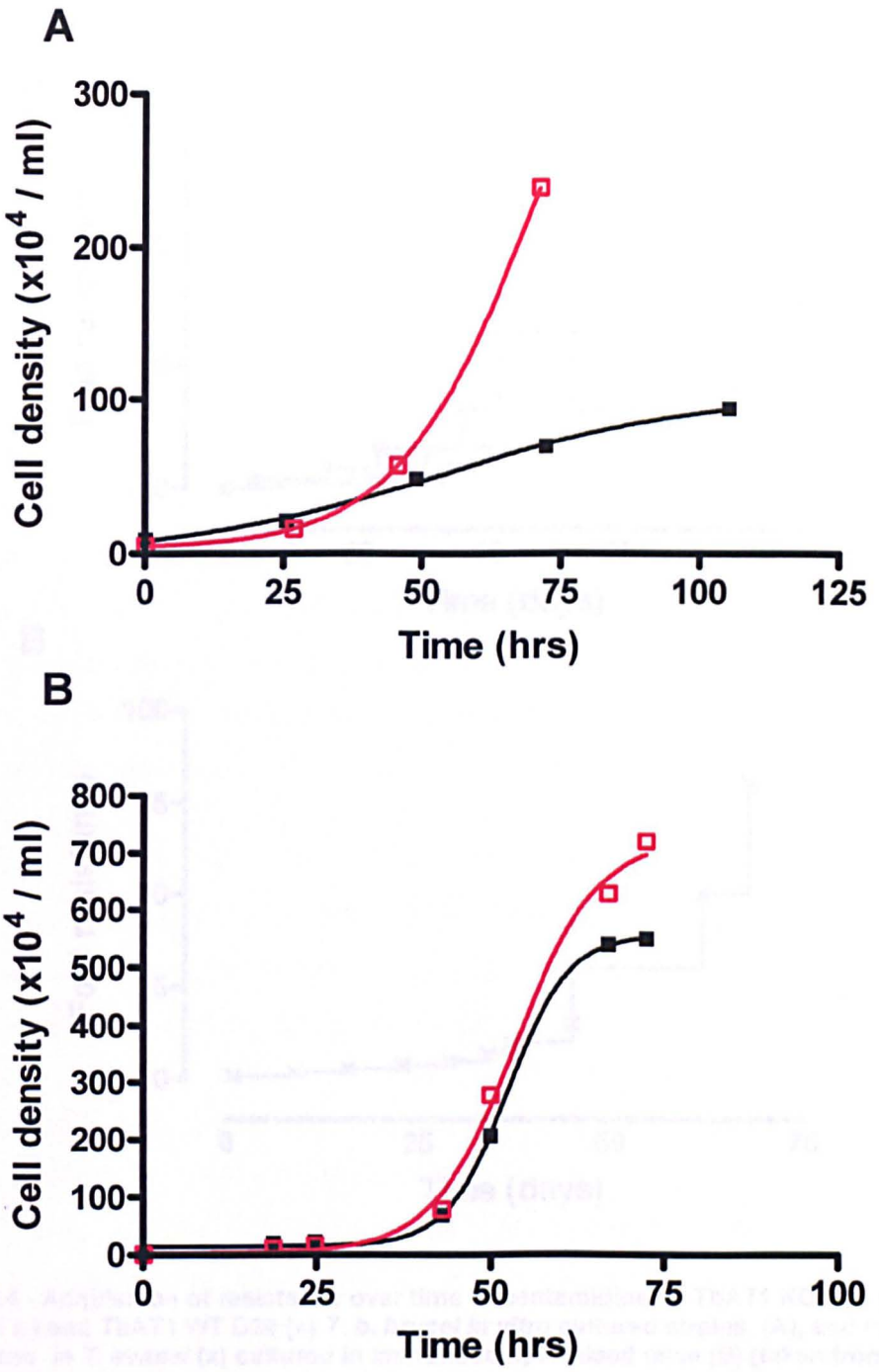


Figure 3.3 - Growth curves of B48 in the presence (■) and absence (□) of pentamidine when first clonally isolated (A) and 2 months later after being maintained in 80nM pentamidine (B).

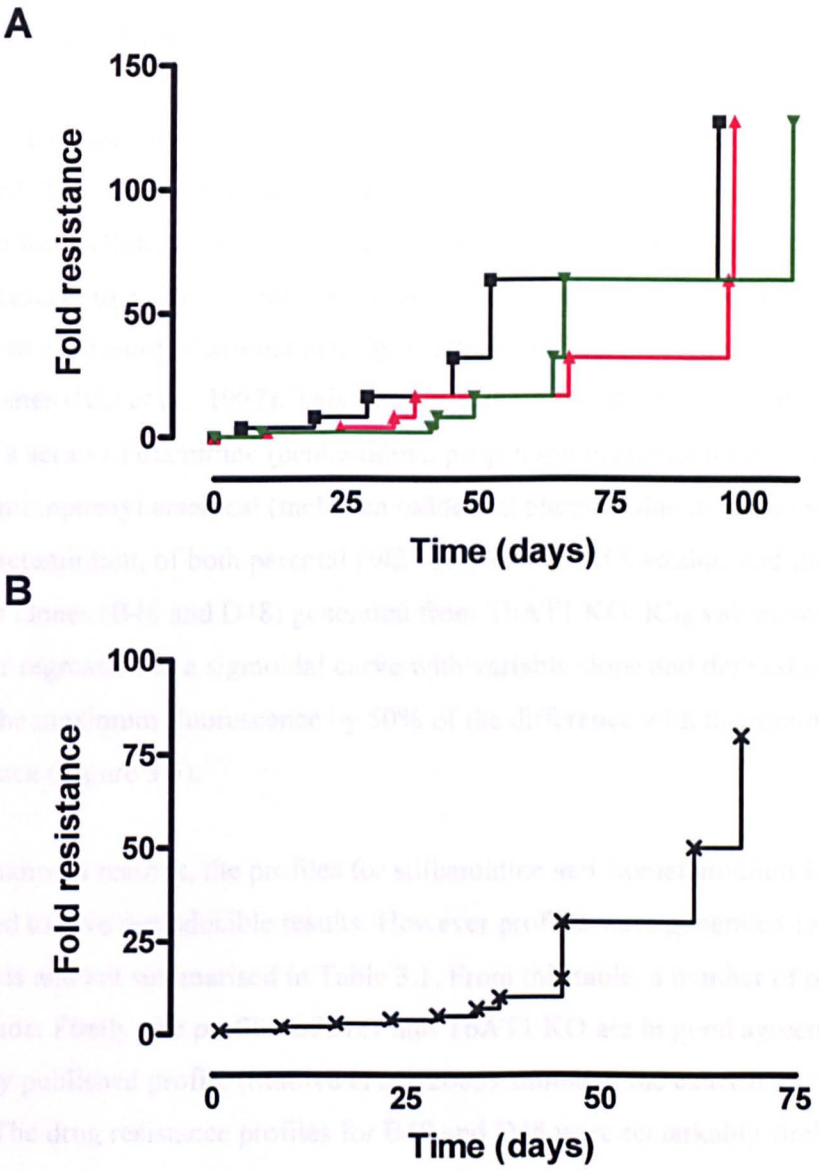


Figure 3.4 - Acquisition of resistance over time to pentamidine, in TbAT1 KO B48 (■),TbAT1 KO D48 (▲) and TbAT1 WT D24 (●) *T. b. brucei in vitro* cultured strains, (A), and to diminazene in *T. evansi* (x) cultured in immunocompromised mice (B) (taken from (Osman *et al.*, 1992)).

3.3.2 Drug resistance phenotype

3.3.2.1 Alamar Blue assays

Following drug adaptation, the drug resistance profile for each of the strains needed to be determined. The Alamar blue assay is a convenient test that measures the metabolic activity of intracellular enzymes via their conversion of a non-fluorescent blue dye to a pink fluorescent dye. This simple, fast, reproducible, and non-invasive technique has been validated as a measure of cellular activity in a host of different organisms, including BSF trypanosomes (Räz *et al.*, 1997). This assay was therefore employed to determine the IC_{50} values of a series of diamidine (pentamidine, propamidine, stilbamidine and diminazene) and melaminophenyl arsenical (melarsen oxide and phenylarsine oxide) compounds, along with isometamidium, of both parental (s427 and TbAT1 KO) strains, and the two high resistance clones (B48 and D48) generated from TbAT1 KO. IC_{50} values were obtained by non-linear regression to a sigmoidal curve with variable slope and defined as the value that reduced the maximum fluorescence by 50% of the difference with the minimum fluorescence (Figure 3.5).

Due to unknown reasons, the profiles for stilbamidine and isometamidium in these cell lines failed to give reproducible results. However profiles were generated for all other compounds and are summarised in Table 3.1. From this table, a number of observations can be made. Firstly, the profiles of s427 and TbAT1 KO are in good agreement with their previously published profile (Matovu *et al.*, 2003), although the exact IC_{50} values do vary slightly. The drug resistance profiles for B48 and D48 were remarkably similar, with resistance to pentamidine having dramatically increased for both, in this case 35- and 22-fold respectively. A far more interesting result from the Alamar blue analysis was the modest increase in resistance to melarsen oxide. It is highly likely that this effect was related to changes in melarsen oxide transport, as sensitivity to phenylarsine oxide, which is membrane permeable, was unchanged. This result does not distinguish between a change in drug export or import. However, melarsen oxide and pentamidine have previously been shown to share at least one route of entry (namely via the P2 transporter), despite their disparate chemical structure (Matovu *et al.*, 2003). In addition, pentamidine protects against melarsen-oxide induced lysis of TbAT1 KO bloodstream forms, with a potency that suggests involvement of HAPT1 (Matovu *et al.*, 2003). However, melaminophenyl arsenicals do not display high affinity for either the HAPT1 or LAPT1 component of transport (De Koning, 2001). The only way to reconcile both sets of data is to postulate that these cells take up melaminophenyl arsenicals very slowly through HAPT1, consistent

with the low affinity, but when exposed for a sufficiently long time, as in an Alamar Blue assay, still succumb to the drug. Consistent with this model, we also observed an increased resistance to propamidine in B48 and D48. This hypothesis is further supported by the propamidine-insensitive accumulation of [^{3}H]-pentamidine in B48 (Figure 3.7B). In contrast, there was no significant difference in diminazene sensitivity in B48 or D48. This is in keeping with previously reports, in which diminazene was shown to enter the trypanosome almost exclusively via the P2 transporter (De Koning *et al.*, 2004) and was unable to affect HAPT1 or LAPT1 transport (De Koning, 2001).

3.3.2.2 Phenotype Stability

To assess the relative stability of the resistance phenotype, B48 and D48 were cultured *in vitro* in the absence of pentamidine. After 3 months without any exposure to pentamidine, the drug resistant phenotype to diminazene, propamidine, pentamidine and melarsen oxide, was re-tested and found to be unchanged (data not shown). This suggests that the adaptation(s) to pentamidine are most likely to be defined at the genetic rather than protein level. In terms of performing a proteomic analysis (see later chapters) this stability is very important.

It was necessary at this point to select one clonal line for further biochemical and proteomic analysis. The high-resistance phenotype (B48 and D48) were preferred over the intermediate resistance phenotype (B6 and D6), as they were likely to contain more drug adaptations. In the same way, B48 displays a higher resistance phenotype than D48 and was therefore selected for all further analyses.

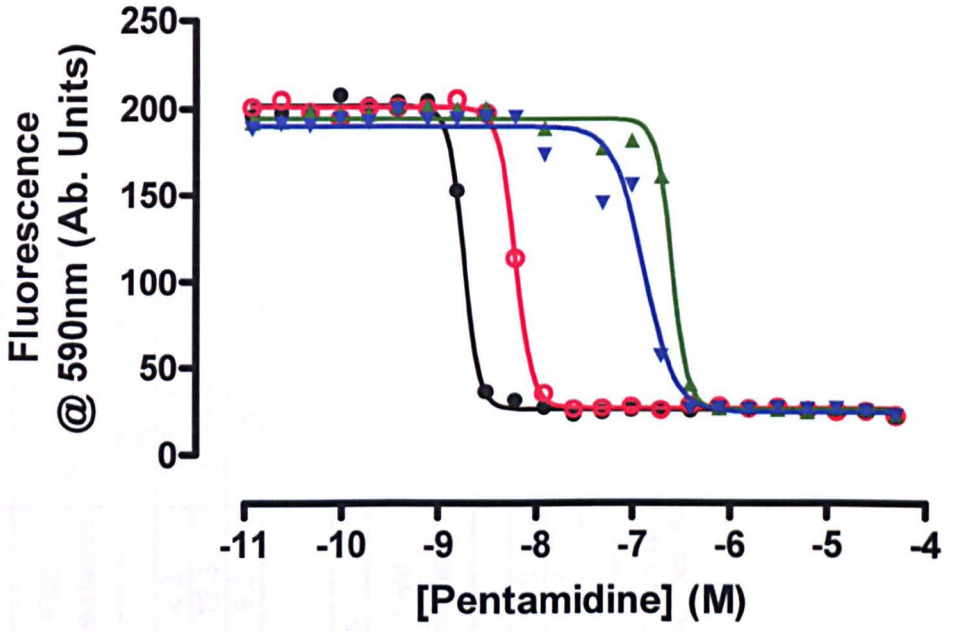


Figure 3.5 - Representative Alamar blue assay result demonstrating the differences in drug sensitivities between the four strains, s427 (●); TbAT1 KO (○); D48 (▼); and B48 (▲).

A

Cell line	Pentamidine			Propamidine			Diminazene		
	IC_{50} (nM)	SEM	Fold Resistance	IC_{50} (nM)	SEM	Fold Resistance	IC_{50} (nM)	SEM	Fold Resistance
s427	2.1	0.2	-	66.4	6.0	-	64.9	14.0	-
TbAT1 KO	7.9	1.1	3.7	317.8	26.6	4.8	1155.2	238.0	17.8
B48	274.8	17.3	128.4	813.4	33.9	12.3	1229.8	301.2	19.0
D48	176.1	13.4	82.3	627.5	60.1	9.5	1029.3	51.7	15.9

B

Cell line	Melarsen Oxide			Phenylarsine Oxide		
	IC_{50} (nM)	SEM	Fold Resistance	IC_{50} (nM)	SEM	Fold Resistance
s427	7.3	0.7	-	1.0	0.1	-
TbAT1 KO	18.3	1.3	2.5	0.8	0.1	0.8
B48	112.7	31.1	15.4	0.8	0.1	0.8
D48	78.2	15.7	10.7	0.9	0.1	1.0

Table 3.1 - Summary of drug sensitivity for wild-type (s427), parental (TbAT1 KO) and drug-resistant strains (B48 and D48) to various diamidine (A) and arsenical (B) drugs determined using the Alamar blue assay. Fold resistance for the two drug resistant lines (B48 and D48) and the TbAT1 KO line are calculated relative to the s427 cell line. All assays were performed in triplicate (minimum).

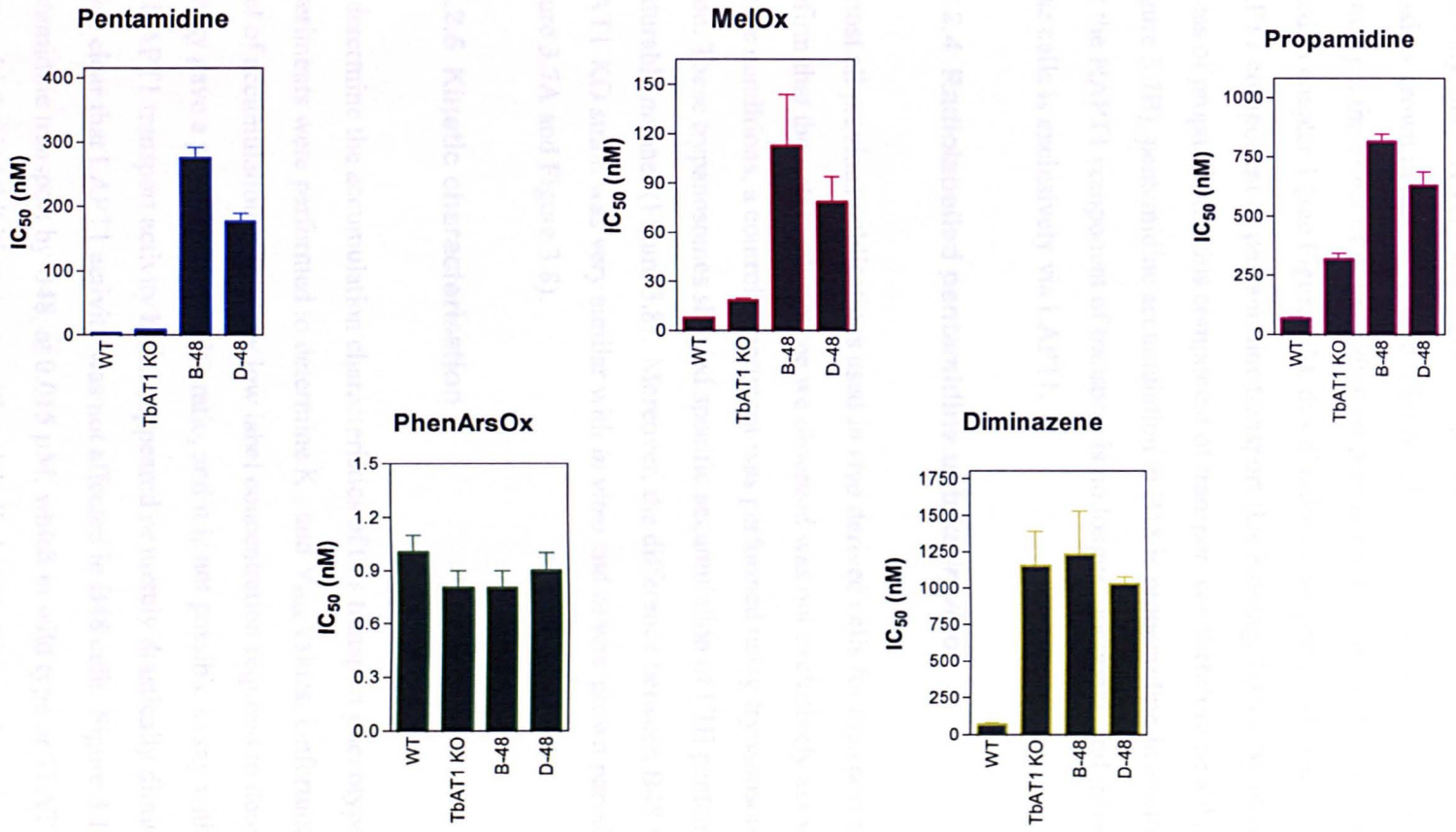


Figure 3.6 - Graphical summary of drug sensitivity (using IC_{50} values ± SEM) for wild-type (s427), parental (TbAT1 KO) and drug-resistant strains (B48 and D48) to various diamidine and arsenical drugs determined using the Alamar blue assay. MelOx, melarsen oxide; PhenArsOx, phenylarsine oxide.

3.3.2.3 Radiolabelled pentamidine uptake *in vitro*

To look for any changes in the relative rates of pentamidine uptake between the parental (TbAT1 KO) and drug adapted (B48) line, radiolabelled [^{3}H]-pentamidine uptake was determined over 5 minutes (Figure 3.7A). This figure clearly shows massively reduced pentamidine accumulation in the pentamidine adapted B48 line. This suggests that either uptake has become compromised, or that pentamidine is being removed from the cell following entry. A very similar difference was observed between the two lines when parasites grown in rats were used (Figure 3.8). To further examine this transport phenotype, the effect of propamidine on pentamidine accumulation was investigated. As stated in chapter 1 (see Figure 1.5), this diamidine has previously been shown to block the HAPT1 component of pentamidine transport (De Koning, 2001). By incubating with an excess of propamidine, this component of transport can therefore be ablated. As shown in (Figure 3.7B), pentamidine accumulation in B48 is propamidine-insensitive. This suggests that the HAPT1 component of transport is no longer functional, and therefore entry into these cells is exclusively via LAPT1.

3.3.2.4 Radiolabelled pentamidine uptake *in vivo*

Almost all previous publications used *in vivo* derived cells for transport studies. To confirm that the uptake phenotype we observed was not exclusively associated with axenic culture conditions, a control experiment was performed using trypanosomes isolated from blood. These trypanosomes showed specific accumulation of [^{3}H] pentamidine over time in a saturable manner (Figure 3.8). Moreover, the difference between B48 and the parental TbAT1-KO strain was very similar with *in vitro* and *in vivo* grown parasites (compare Figure 3.7A and Figure 3.8).

3.3.2.5 Kinetic characterisation

To determine the accumulation characteristics of this transport phenotype, a series of experiments were performed to determine K_m and V_{max} values. Unfortunately, the low level of accumulation at the very low label concentration required to detect any HAPT1 activity gave a poor signal to noise ratio, and it is not possible to say with certainty whether the HAPT1 transport activity had disappeared or merely drastically diminished. However, it was clear that LAPT1 activity was not affected in B48 cells. Figure 3.10 shows [^{3}H]-pentamidine transport by B48, at 0.015 μM, which in wild type or TbAT1 KO cells would show, a biphasic inhibition curve with unlabelled pentamidine, depicting >50% HAPT1

activity (De Koning, 2001; Matovu *et al.*, 2003). In B48, however, little or no HAPT1 activity was observed in three different experiments, but a K_m value for LAPT1 activity could be determined as 55.6 ± 6.6 μM with a V_{max} of 0.82 ± 0.20 pmol(10^7 cells^{-1})s^{-1} (n=3). These values are virtually identical to those previously reported for LAPT1 in WT bloodstream forms (De Koning, 2001).

3.3.2.6 Radiolabelled 2-deoxy-glucose (2DG) uptake *in vitro*

To test whether the transport phenotype observed regarding pentamidine accumulation was indicative of a general down-regulation in transport, it was important to look at transport in an unrelated system. BSF trypanosomes rely exclusively on glucose as their energy source (Opperdoes and Borst, 1977). Glucose transport in this lifecycle stage is correspondingly high and has been well characterised (Albert *et al.*, 2005). To confirm that transport activity had not as a whole been down regulated, we looked at the uptake of the glucose analogue [^{3}H]-2DG in B48 and the parental TbAT1 KO line (Figure 3.9). The rates of 2DG uptake in the two strains were not significantly different, suggesting that the reduction in pentamidine accumulation is not due to a general down regulation in transport activity. This is not surprising considering that both strains replicate at equivalent rates and therefore their glucose requirements would be expected to be the same.

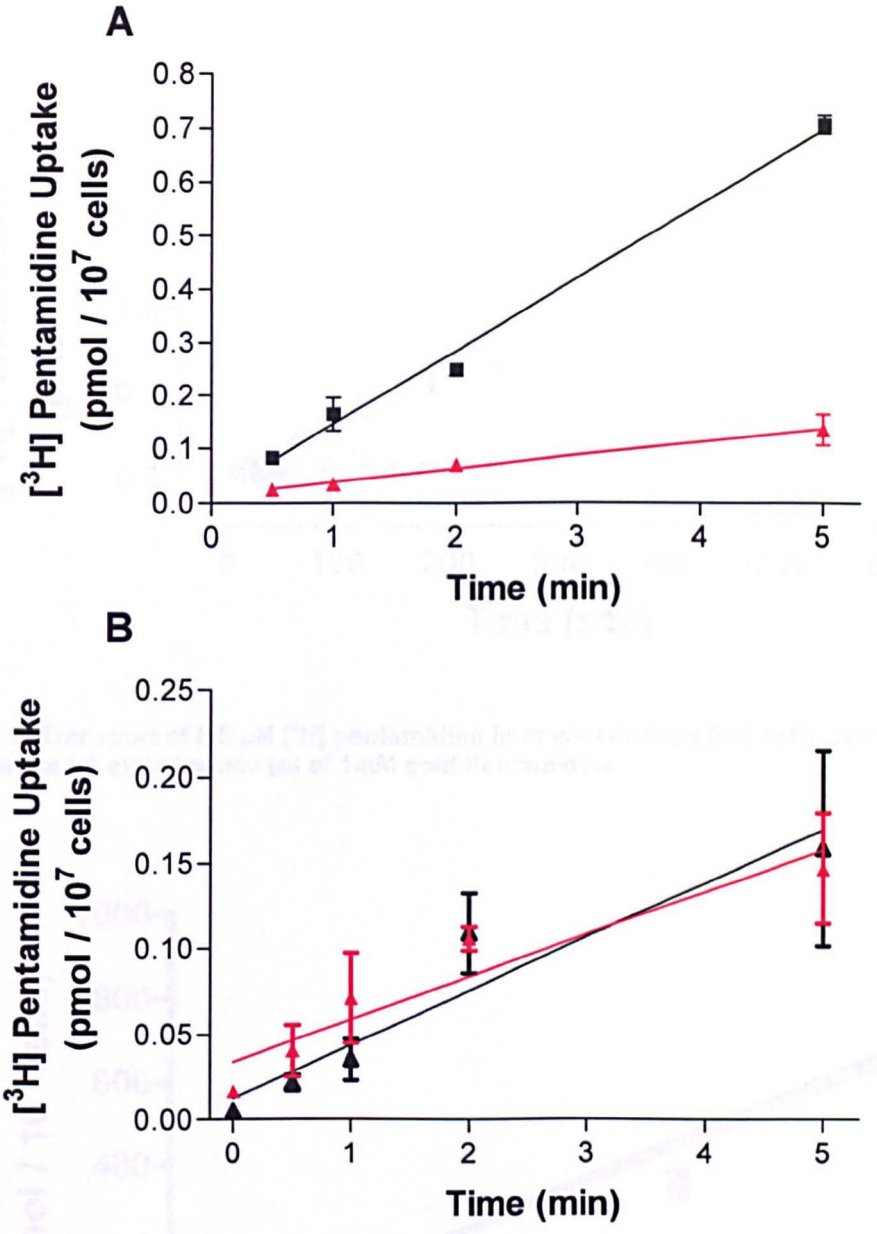


Figure 3.7 –(A) Transport of 0.5 μM [^{3}H] pentamidine over 5 min in TbAT1 KO (■) and B48 (▲). (B) Uptake of 0.5 μM [^{3}H] pentamidine in B48, in the presence (▲) and absence (▲) of 1 mM propamidine. Representative experiments are shown (n=2).

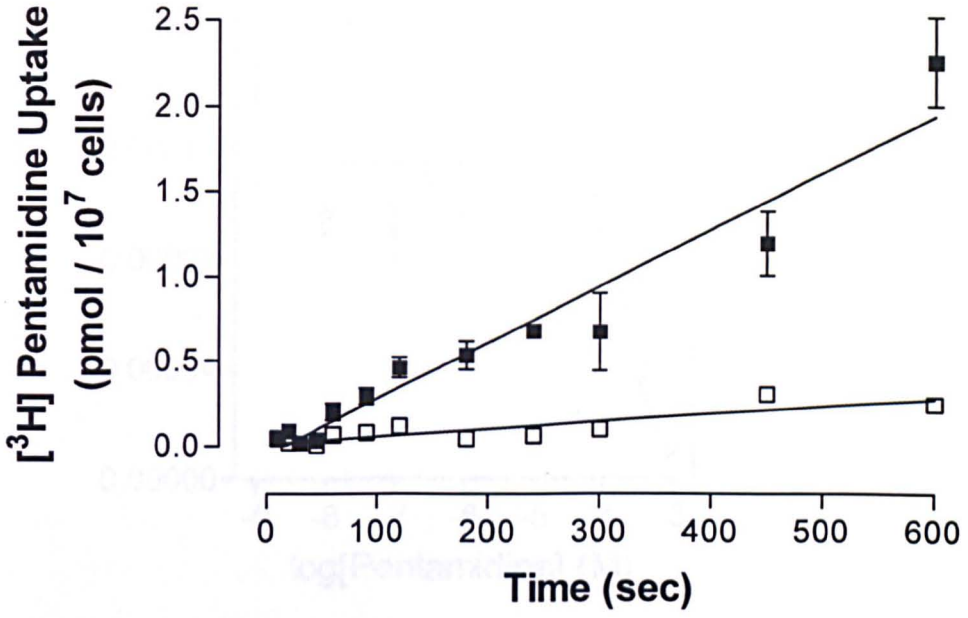


Figure 3.8- Transport of 0.5 μM [^{3}H] pentamidine in *in vivo* derived B48 cells over 10min in the presence (□) and absence (■) of 1mM cold Pentamidine.

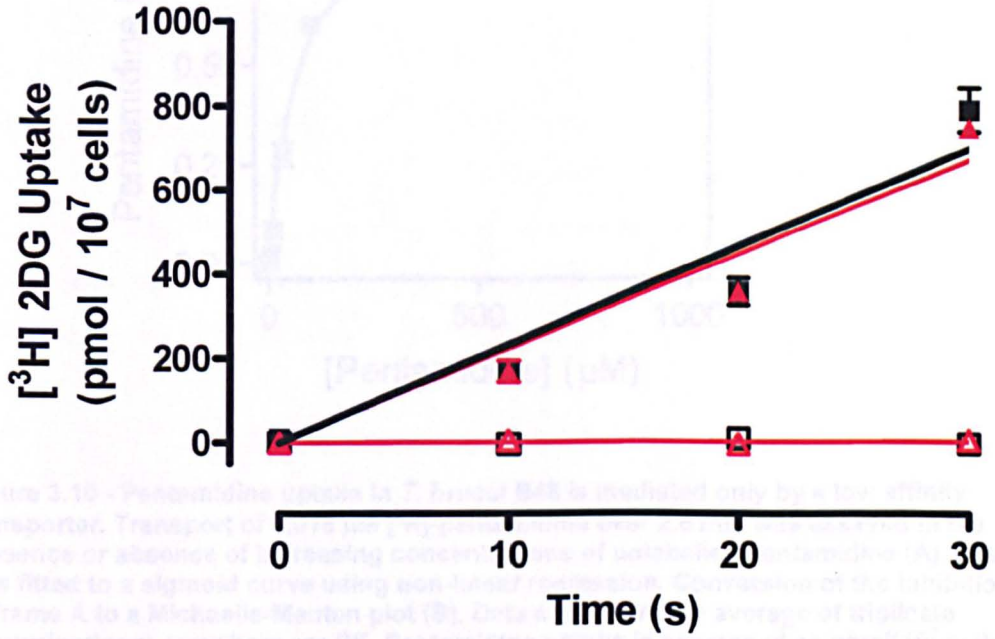


Figure 3.9- Transport of 100μM [^{3}H] 2-deoxy-glucose (2DG) in *in vitro* derived BSF trypanosomes over 30 seconds. Parental TbAT1 KO (■) and B48 (▲) cells, in the absence of inhibitor, and in the presence of 1 mM cold 2DG (□ and △ respectively).

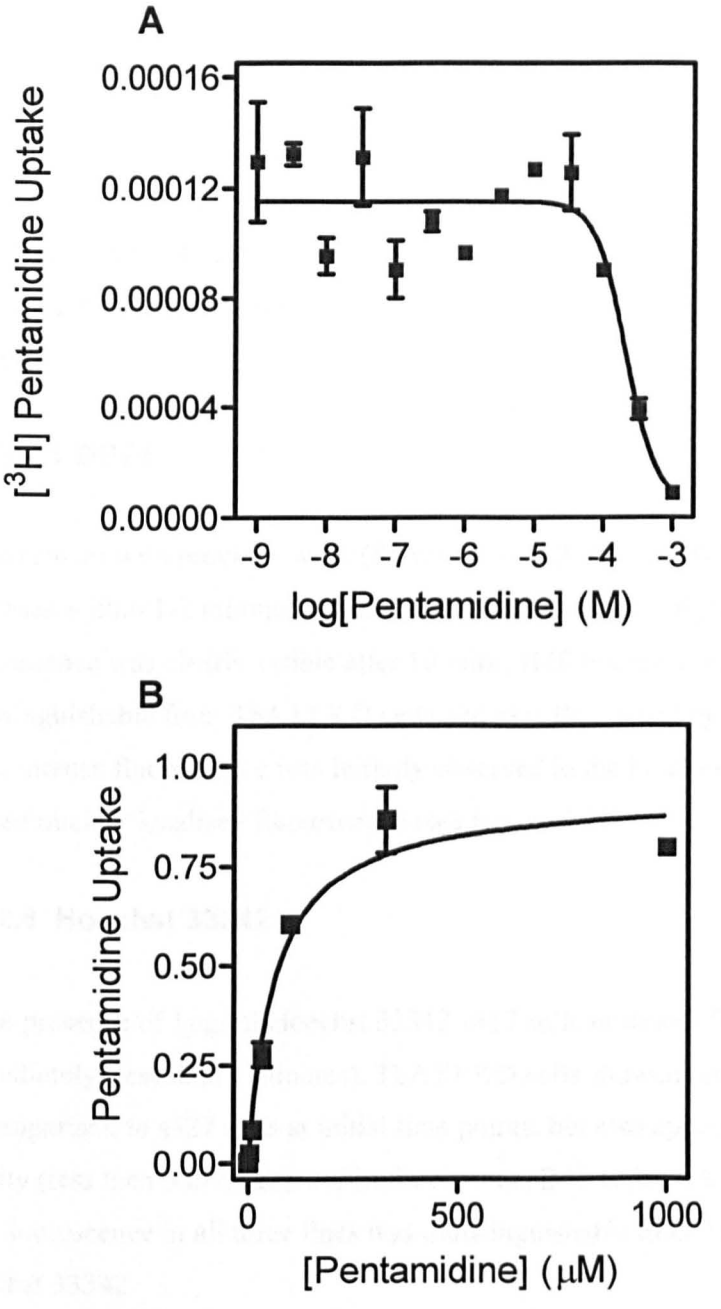


Figure 3.10 - Pentamidine uptake in *T. brucei* B48 is mediated only by a low affinity transporter. Transport of 0.015 μM [^{3}H]-pentamidine over 2.5 min was assayed in the presence or absence of increasing concentrations of unlabelled pentamidine (A). The data was fitted to a sigmoid curve using non-linear regression. Conversion of the inhibition data in frame A to a Michaelis-Menten plot (B). Data shown are the average of triplicate determinations; error bars are SE. Pentamidine uptake is expressed as pmol(10^7 cells)$^{-1}$s^{-1}.

3.3.2.7 Fluorescence microscopy

The route(s) of entry for the many fluorescent diamidines, with the exception of TbAT1 for DB75 (Stewart *et al.*, 2005), remain largely unknown. A previous publication reported that melarsoprol resistant trypanosomes displayed reduced diamidine accumulation. To ascertain whether B48 cells displayed a change in membrane permeability from their parent lines, accumulation of the various compounds was measured at a series of time points.

3.3.2.7.1 DB75

In agreement with previous work (Stewart *et al.*, 2005), all s427 cells were seen to fluoresce within 1-2 minutes. While TbAT1 KO cells took significantly longer, strong fluorescence was clearly visible after 10 mins. B48 fluorescence acquisition was indistinguishable from TbAT1 KO cells and also fluoresced by 10 minutes. For all cell lines, intense fluorescence was initially observed in the kinetoplast, followed by more diffuse nuclear localised fluorescence (see Figure 3.11).

3.3.2.8 Hoechst 33342

In the presence of 1μg/ml Hoechst 33342 s427 cells intensely fluoresced almost immediately (less than 3 minutes). TbAT1 KO cells showed reduced fluorescence intensity in comparison to s427 cells at initial time points, but also appeared to fluoresce very rapidly (less than 3 minutes) within. In contrast, B48 cells took around 10min to fluoresce. The fluorescence in all three lines was indistinguishable after 15min incubation with Hoechst 33342.

3.3.2.8.1 DAPI

As with DB75, s427 and TbAT1 KO cells displayed virtually identical fluorescence pictures over time (Figure 3.12) in the presence of 2.8μM DAPI. By 5 minutes, fluorescence could clearly be seen, and by 10 minutes, staining was very intense in both nucleus and kinetoplast. In contrast, B48 cells showed only very weak fluorescence in both the kinetoplast and nucleus after 20 min staining. Even after 60 min incubation, fluorescence intensity was still far below that observed in the s427 and TbAT1 KO lines at 10 minutes. This reduced accumulation of DAPI in the cells was not observed in dead B48 cells (Figure 3.13).

In an attempt to ascertain whether DAPI fluorescence could be selectively blocked by pentamidine, cells were incubated with 2.8μM DAPI in the absence or presence of a low (10μM) or high (1mM) concentration of pentamidine, and fluorescence acquisition in living cells was observed over time (Figure 3.14A). Unfortunately due to high motility it was very difficult to capture well-focused fluorescent DAPI images, hence images from live cells are slightly blurred. To show true fluorescence levels, gluteraldehyde-fixed specimens following 60 min incubation are also shown (Figure 3.14B). This figure clearly shows that the presence of a low concentration of pentamidine has little or no effect on DAPI fluorescence. In contrast incubating with a large excess of pentamidine practically abolishes the acquisition of fluorescence over the time analysed.

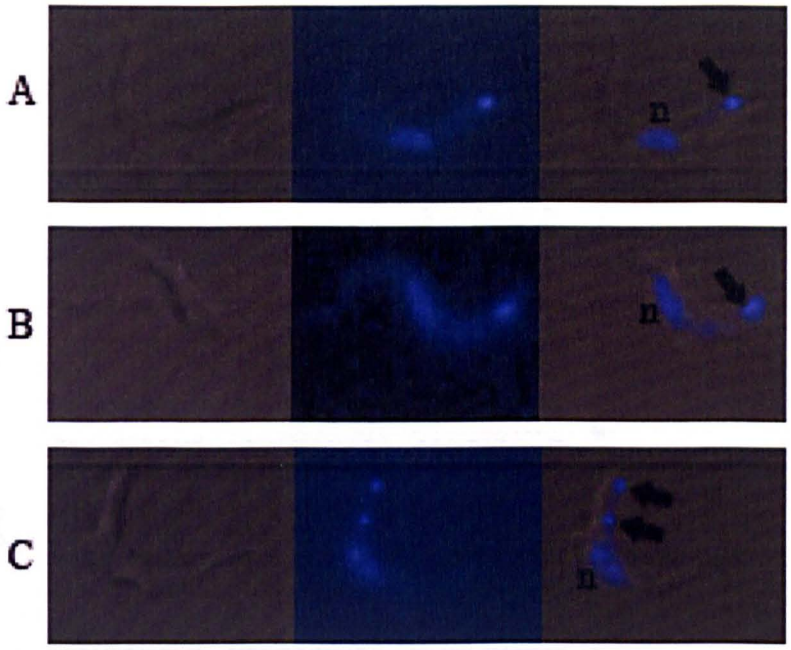


Figure 3.11 – DIC, fluorescence ($\lambda_{ex=365nm}$, $\lambda_{em=445nm}$) and overlay microscopy pictures after 5 min incubation with 10μM DB75 in (A) s427, (B) TbAT1 KO, and (C) B48. To enable image overlay, trypanosomes had to be non-motile and samples were therefore fixed in glutaraldehyde for 20 min. Fluorescence is clearly visible in the nucleus (n) and kinetoplasts (arrow), of which there are two in B48, as the cell is in the early stages of cell division.

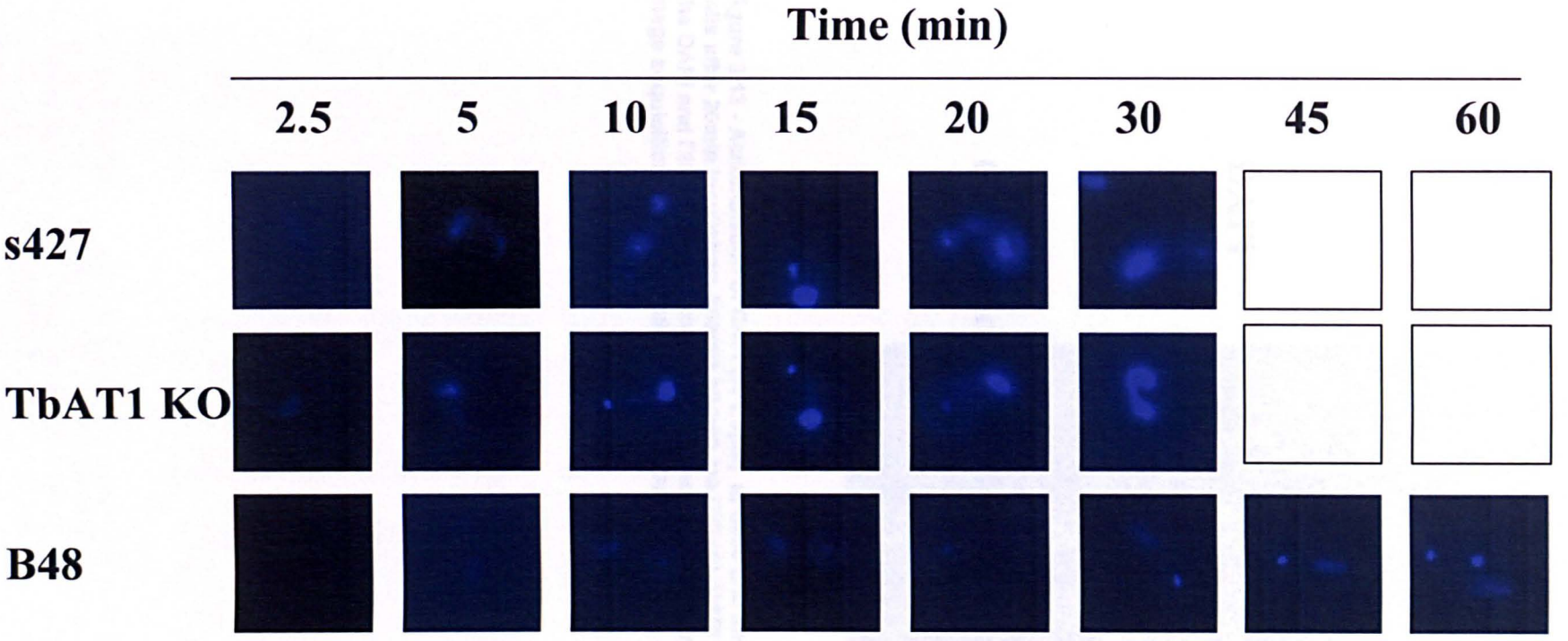


Figure 3.12 - Acquisition of DAPI (at 2.8μM) fluorescence over time in live cell lines; s427 (A), TbAT1 KO (B) and B48 (C). Cells were highly motile and only DAPI fluorescence was measured ($\lambda_{ex=365nm}$, $\lambda_{em=445nm}$).

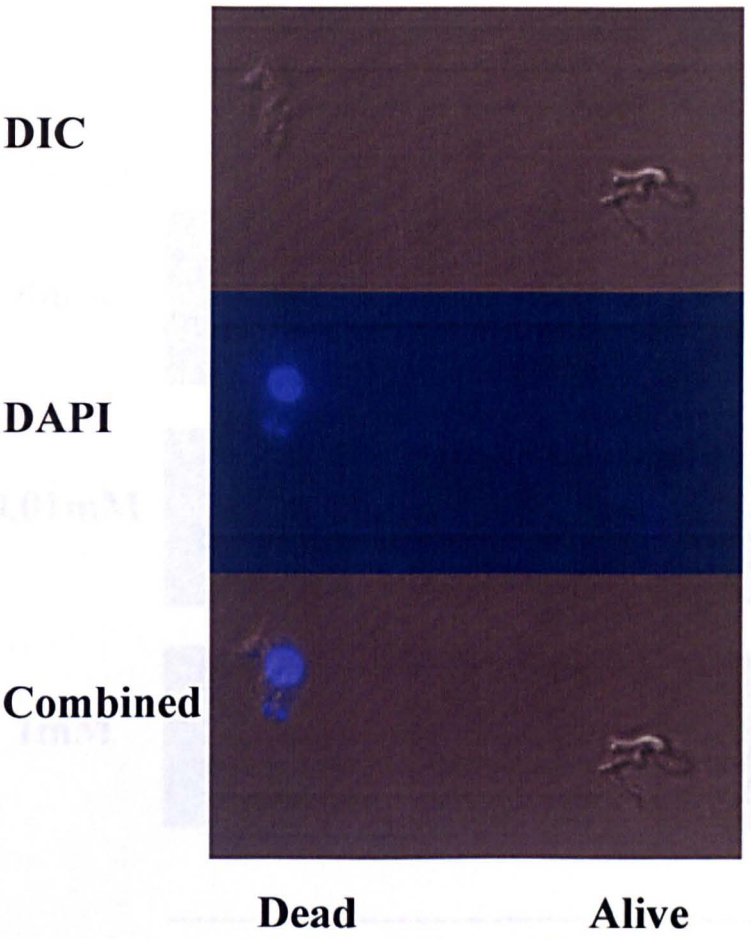


Figure 3.13 - Accumulation of DAPI (at 2.8μM) in dead and alive B48 pentamidine resistant cells after 20min incubation. Images shown are DIC (A), DAPI (B) and the overlay image (C). The DAPI and DIC images do not overlay perfectly for the living cell, as it was motile and image acquisition was consecutive not simultaneous.

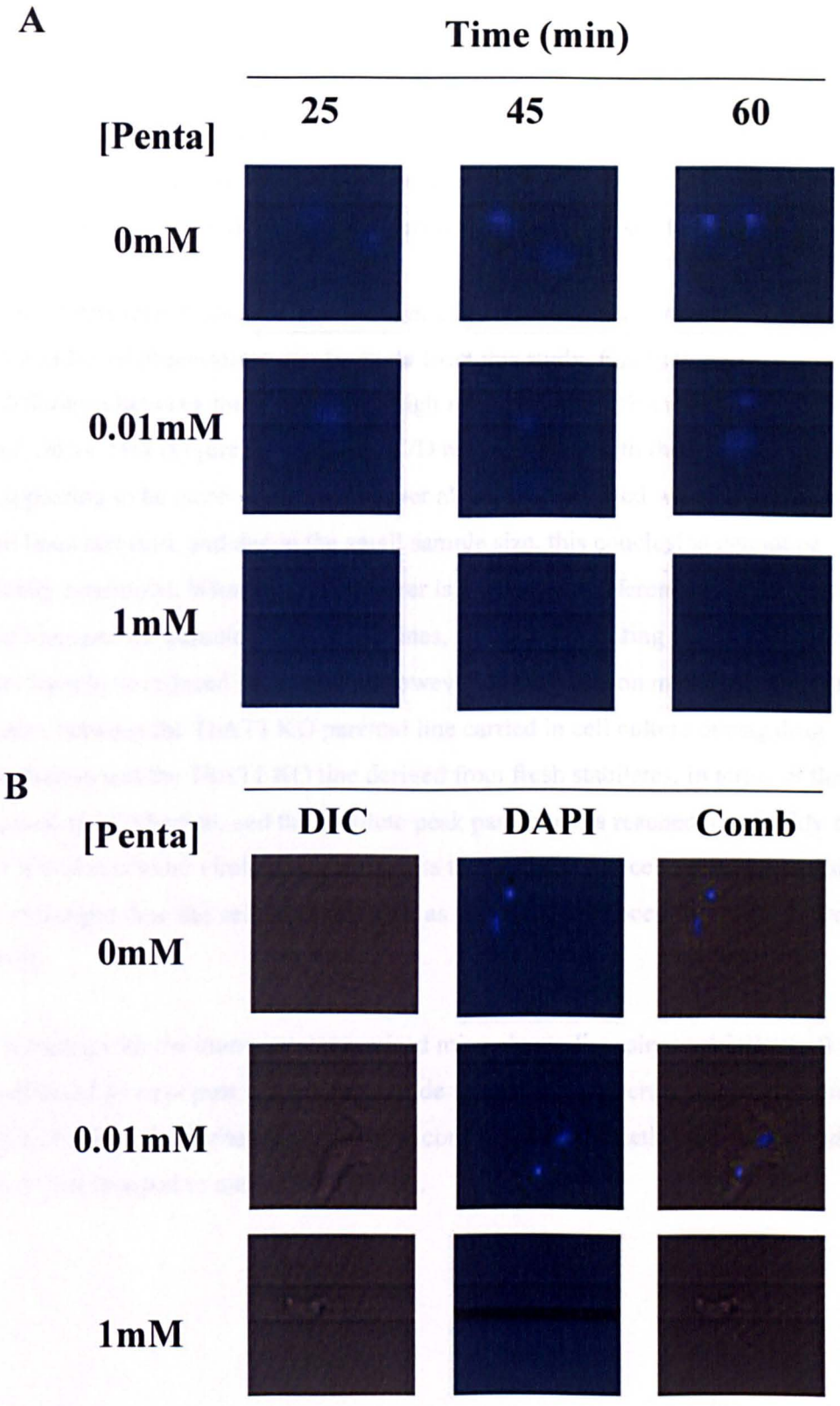


Figure 3.14 - DAPI fluorescence acquisition in live B48 cells incubated in the absence or presence of a low (0.01mM) or high (1mM) concentration of pentamidine over time (A). Separate images (DIC, DAPI and combined overlay) from gluteraldehyde fixed parasites after 60min incubation with DAPI ± pentamidine (B). Penta, pentamidine; Comb, combined overlay.

3.3.2.9 *In vivo* characterisation

During *in vivo* culturing, it was observed that the drug resistant strains appeared to exhibit reduced virulence in comparison to the parental wild type. To assess this observation further, two larger scale studies were performed as described in section 3.2.3.5.

Infections in rats were tracked for a minimum of 20 days and are summarised in Figure 3.15. A number of observations can be made from this study. Firstly there appears to be a slight difference between the medium and high resistant strains from each line, i.e. B6 vs. B48 and D6 vs. D48 (Figure 3.15A/B and C/D respectively), with the higher resistant strain appearing to be more virulent. However all animals infected with the pentamidine adapted lines survived, and due to the small sample size, this conclusion cannot be confidently confirmed. What is clear however is the marked difference between any of the adapted lines and the parental TbAT1 KO lines, strongly suggesting that the pentamidine resistant lines have reduced in virulence. However this conclusion must be tempered by the difference, between the TbAT1 KO parental line carried in cell culture during drug acclimatisation and the TbAT1 KO line derived from fresh stabilates. In terms of the speed of progression of infection, and the absolute peak parasitaemia reached, the freshly derived TbAT1 KO line is more virulent. This suggests that maintaining cells in axenic culture for such a prolonged time has reduced virulence, as has previously been reported (Berger *et al.*, 1995).

When working with the immunocompromised mice, due to licensing restrictions all mice were sacrificed 14 days post cyclophosphamide treatment, but were tracked up to this point (Figure 3.16). This time constraint prevented confirmation of whether infections had truly cleared or just dropped to undetectable levels.

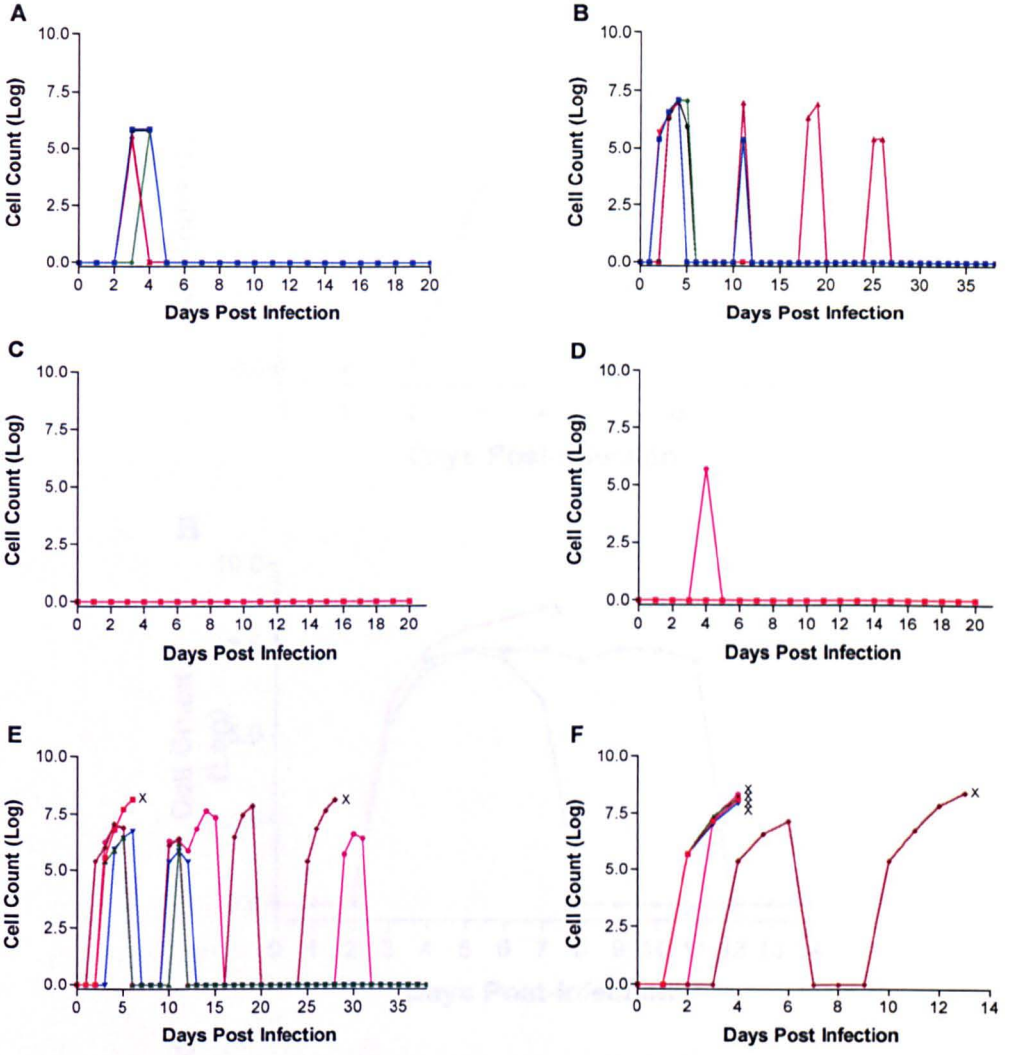


Figure 3.15 - Progression of infections in groups of 5 rats following i.p. inoculation with B6 (A), B48 (B), D6 (C), D48 (D), TbAT1 KO (maintained in culture) (E), TbAT1 KO (freshly resurrected from stabilates) (F) strains. Animals reaching a parasitaemia (cell count) of 8 (log) or more were euthanased (X). Each colour represents an individual rat within each group.

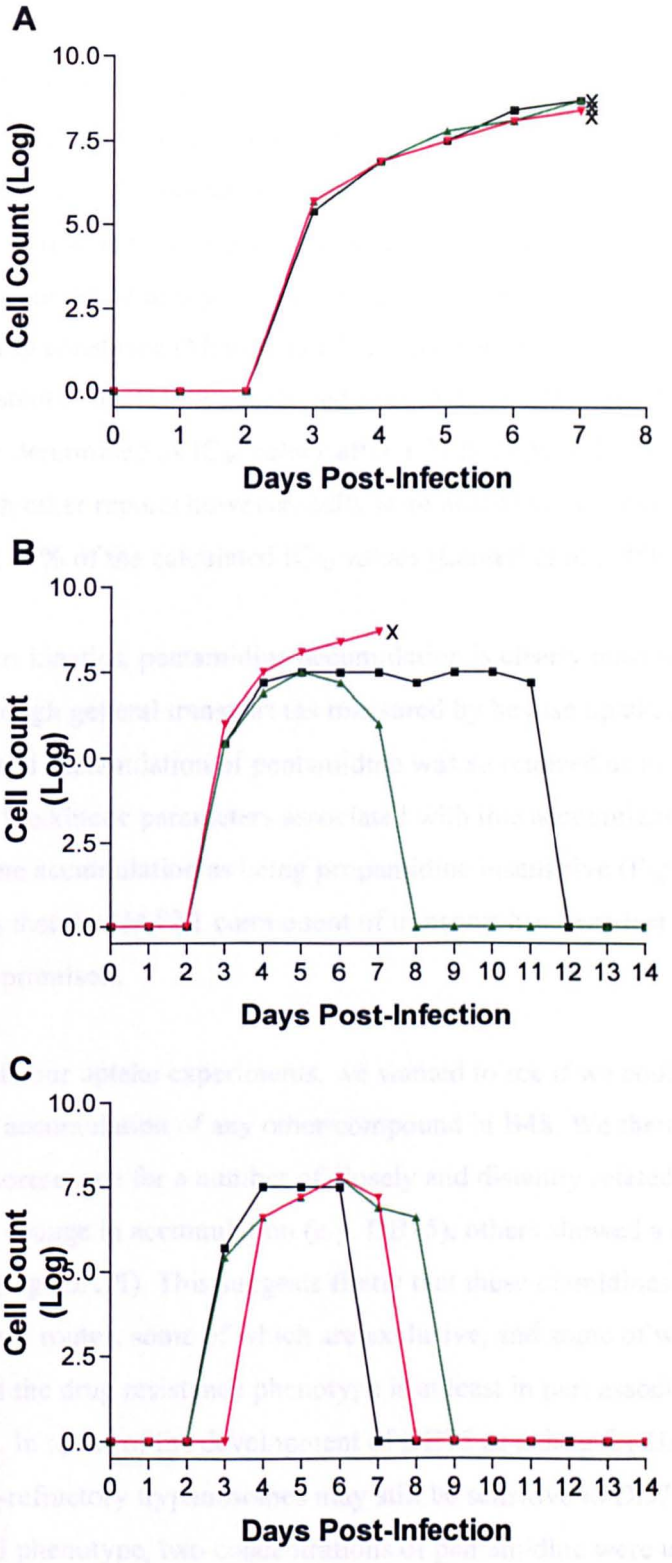


Figure 3.16 - Progression of *T. brucei* infections in groups of 3 immunocompromised mice following i.p. inoculation with TbAT1 KO (A), B48 (B), or D48 (C). Animals reaching a parasitaemia load of log 8 or more were euthanased (X). Each colour represents an individual rat within each group.

3.4 Discussion

A number of strains were generated from various parental lines and of varying pentamidine resistance (Figure 3.2). Lines displaying maximal drug resistance were most likely to contain the largest number of pentamidine resistance adaptations. Attention therefore focused on these strains, and in particular those derived from the TbAT1 KO parent as they lacked the P2 component of transport, the impact of which has already been well documented in drug resistance (Matovu *et al.*, 2003). For the same reasons, the higher pentamidine resistant clone B48 was selected over D48 for complete characterisation. Drug sensitivities were determined as IC_{50} values after a 72 hour period of incubation with drug. In agreement with other reports however, cells were unable to be maintained in drug concentrations at 10% of the calculated IC_{50} values (Lanteri *et al.*, 2006).

From the transport kinetics, pentamidine accumulation is clearly massively reduced in B48 (Figure 3.7), although general transport (as measured by hexose uptake) was unaffected (Figure 3.9). Indeed accumulation of pentamidine was so reduced as to prevent determination of the kinetic parameters associated with this accumulation. However we were able to define accumulation as being propamidine insensitive (Figure 3.8). This strongly suggests that the HAPT1 component of transport has been lost or in some way functionally compromised.

To further validate our uptake experiments, we wanted to see if we could observe alterations in the accumulation of any other compound in B48. We therefore analysed the acquisition of fluorescence for a number of closely and distantly related diamidines. While some showed no change in accumulation (e.g. DB75), others showed a marked reduction in accumulation (e.g. DAPI). This suggests firstly that these diamidines enter the cell via a number of different routes, some of which are exclusive, and some of which are shared. It also suggests that the drug resistance phenotype is at least in part associated with a change in the membrane. In terms of the development of DB75 as a drug for HAT, it does suggest that pentamidine-refractory trypanosomes may still be sensitive to DB75. To further analyse the DAPI phenotype, two concentrations of pentamidine were used to test whether fluorescence could be selectively blocked. At the low (10μM) concentration of pentamidine only HAPT1 (if present) would be saturated, while at the higher concentration (1mM) LAPT1 would also be saturated (De Koning, 2001). The fluorescence strongly suggests that a low affinity transporter is responsible for uptake of DAPI into the cell.

The observation that B48 and D48 displayed reduced virulence *in vivo* was not entirely unexpected, as it had previously been observed in diminazene resistant *T. brucei* (Egbe-Nwiyi *et al.*, 2005). The question was whether this reduction in virulence was associated with the resistant phenotype or with the strains having been maintained in axenic culture for more than a year. It appears from the control study using TbAT1 KO cells maintained in culture and fresh cells resurrected from stabilates stored in liquid N_2 that both have a role to play. Our interpretation of this data is that the pentamidine adaptations are more important in determining the avirulent phenotype than being carried in culture, although the latter does appear to reduce virulence (Figure 3.15). This is further supported by the observation that B48 is more virulent in immunocompromised mice than immunocompetent rats. Only in immunocompromised animals did the parasitaemia reach a terminal level, and high parasitaemia levels were maintained for longer. This strongly suggests that the avirulent phenotype is immune-mediated.

A previous publication has shown that developing resistance to a trypanocide is greatly aided *in vivo* by compromising the immune system (Osman *et al.*, 1992). Considering that these trypanosomes have essentially not been exposed to a functional immune system, it may be that the adaptations acquired *in vitro*, are essential for *in vivo* virulence.

From all of the transport, Alamar blue data and fluorescence microscopy work, a summary diagram of relevant transporters in *T. brucei* (Figure 3.17) has been constructed. In the drug-adapted B48 line, the transport components that are proposed to remain are shown in Figure 3.17C. We therefore propose that the B48 line has lost the high affinity pentamidine transporter (HAPT1).

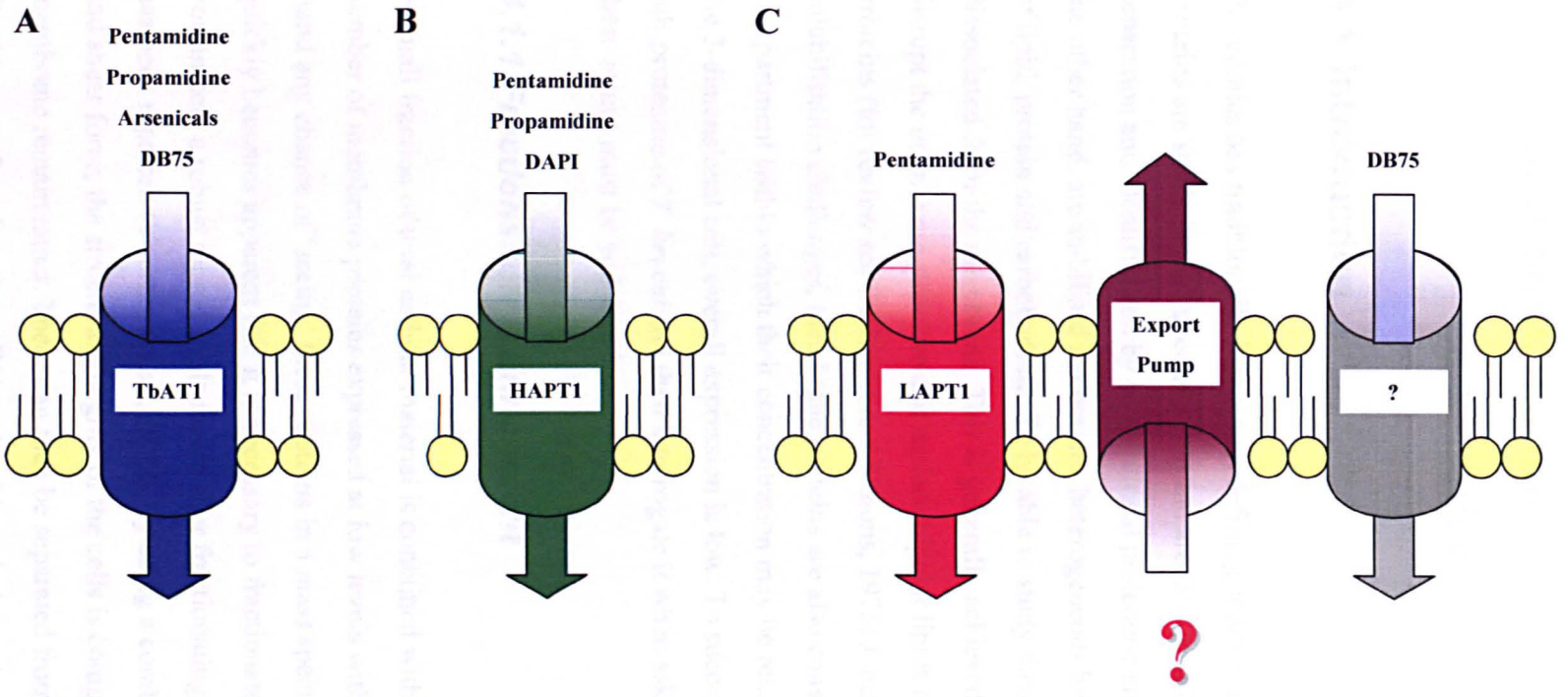


Figure 3.17 – Summary diagram of transporters involved or potentially involved in the pentamidine resistance phenotype. TbAT1; *T. brucei* Adenosine Transporter 1 (A) is known to transport a wide range of trypanocides and purine sources. HAPT1; High affinity pentamidine transporter 1 (B) is able to transport pentamidine, propamidine and we propose DAPI. LAPT1, other uptake mechanisms and any possible drug extrusion pumps (C) remain completely unknown.

Chapter 4

The plasma membrane sub-proteome: a technical challenge

4.1 Introduction

Proteomics has traditionally focused on defining soluble proteomes. By definition these proteins are stable in the aqueous phase and are thus readily extracted and amenable to separation and identification by conventional proteomic analysis. Membrane proteins on the other hand, are stabilised in complex heterogeneous biological membranes consisting of lipid, protein and carbohydrate. To be able to study these proteins, they need to be dissociated from the membrane. This is generally achieved with detergents, which need to disrupt the membrane structure and replace bipolar lipids to solubilise the membrane proteins (for review see (Helenius and Simons, 1975; Jones, 1999)). In addition to the solubilisation challenges, membrane proteins are also constrained in a 2-dimensional compartment within which their concentration may be relatively high, but in the context of the 3-dimensional cell, overall expression is low. To successfully define the membrane sub-proteome of *T. brucei* and then interrogate it when asking biological questions, both of these issues must be addressed.

4.1.1 Fractionation & Enrichment

A small fraction of total cellular material is contained within the plasma membrane, with a number of membrane proteins expressed at low levels within this structure. To realistically stand any chance of 'seeing' these proteins in a mass spectrometry based approach it quickly becomes apparent that it is necessary to fractionate / enrich for these components. Fortunately a robust reproducible method for fractionating *T. brucei* plasma membranes has been reported (Voorheis *et al.*, 1979). By using a combination of hypotonic swelling and shear force, the structural integrity of the cells is compromised, but sheets of membrane remain intact. These can then be separated from the cellular milieu by sucrose density centrifugation. Serendipitously, this technique also removes almost all variable surface glycoprotein, of which there are an estimated 10^7 copies per cell (Ferguson *et al.*, 1999), from the plasma membrane preparation presumably via the action of phospholipase

C (Carrington *et al.*, 1998). It is therefore likely that most other GPI-anchored proteins will also be removed from the plasma preparation.

Trypanosomes, being motile flagellated organisms, have a highly structured cytoskeleton that is intimately associated with the plasma membrane, and is co-purified with the plasma membranes. The majority of the cytoskeletal components are structural proteins, although a number of other species are present (Gull, 1999). To maximise our potential to identify rare polytopic membrane proteins, we hoped to test a series of techniques to selectively enrich for plasma membrane proteins over cytoskeletal proteins. Such separation techniques generally isolate membrane proteins based on hydrophobicity. Many of these utilise detergents to differentially solubilise membrane proteins. For example the non-ionic detergent Triton X-114 can be used to enrich for membrane proteins by promoting phase partitioning between aqueous and detergent phases (Bordier, 1981; Brusca and Radolf, 1994), and this technique has already been used to investigate GPI anchored proteins in Chagas disease (Añez-Rojas *et al.*, 2005).

Another approach effectively combines the fractionation and enrichment steps by exposing live cells to a reactive compound that is membrane-impermeable and therefore only tags proteins exposed to the extracellular fluid. Surface exposed proteins can then be isolated by affinity purification (Zhang *et al.*, 2003). However this approach would not be able to identify any intracellular anchored proteins or proteins whose extracellular domains are very small. In trypanosomes this approach would also isolate the entire VSG coat, which would then dominate the proteome.

To identify the most appropriate technique(s) described above, a simple method for evaluating each technique was proposed. This consisted of developing a *T. brucei* cell line expressing an epitope-tagged *T. brucei* membrane protein and then comparing the relative enrichment of each of these techniques by Western blot analysis. There are a large number of polytopic plasma membrane proteins. However, in view of our aim of quantitatively comparing the membrane proteomes of drug resistant vs. drug sensitive *T. brucei* (as described in chapter 3), we chose to tag TbAT1 – so far the only gene shown to be definitively involved in pentamidine uptake in trypanosomes (Matovu *et al.*, 2003).

One of the challenges of defining the PM sub-proteome, was to exclusively focus on the PM rather than the combined PM and cytoskeleton proteome. To that end, as well as pursuing the above enrichment strategies, we decided to also employ a subtractive approach. In this we wanted to define the cytoskeletal sub-proteome and therefore from it

define which proteins from the PM sub-proteome were actually of cytoskeletal origin. To aid us in this approach, the publication of the flagellar proteome, which has recently been defined was also used (Broadhead *et al.*, 2006).

4.1.2 Membrane Proteomics

As mentioned in the general introduction, there are a number of different proteomic techniques available. To successfully interrogate the membrane proteome we focused our attention on techniques reported to be compatible with polytopic hydrophobic proteins. To maximise membrane sub-proteome coverage, we considered it prudent in view of other reports (Friso *et al.*, 2004) to pursue as many different complementary techniques as possible.

4.1.2.1 Gel-based separations

Gel-based separations have the potential to resolve complex protein samples with high resolution. Almost without exception this has been achieved with 2DGE where proteins are separated based on pI and mass (O'Farrell, 1975). One of the real strengths of 2DGE is its ability to resolve various post-translational modifications i.e. different protein species, by virtue of their different migration on the gel. These subtle modifications are usually not identified by mass spectrometry. In addition, spot patterns can be used to identify changes in gross protein expression levels, although gel-to-gel comparisons are often tricky despite the commercial production of electrophoresis materials e.g. IPG strips. Gels have a limited resolution; however, large format (24 cm) gels are generally able to resolve more than a 1000 of the most abundant spots per gel. By running multiple gels and a combination of zoom gels that focus on a constricted pI range e.g. pH 4-7, excellent resolution and highly reproducible proteome coverage can be achieved. Unfortunately, in terms of defining the membrane proteome, 2DGE is very poor at resolving polytopic proteins, forcing alternative methodologies to be employed (Wilkins *et al.*, 1998).

A two dimensional system that separates proteins on mass alone in both dimensions was developed by Macfarlane *et al* in the late 1990's (Macfarlane, 1989), although its application to membrane proteins was only recognised some time later (Hartinger *et al.*, 1996), when it was realised that 16-BAC rivalled SDS in its ability to solubilise membrane proteins. Despite proteins being separated by mass alone, the different conditions under which the two dimensions are run (cationic 16-BAC at pH 2 for the first dimension compared to anionic SDS at pH 9 for second dimension) ensures that many proteins

deviate to a greater or lesser extent from their theoretical migration pattern in one or more of the dimensions. Although the reasons for such 'aberrant' migration are poorly understood, it is likely that charge and differential detergent binding are important factors. This gel-based system therefore does provide an environment compatible with membrane protein separation, however in comparison to 2DGE the resolution i.e. the total number of spots that can be identified per gel, is greatly reduced. This technique therefore absolutely requires sample pre-fractionation. In addition, protein isoforms of the same M_{wt} are unlikely to be separated. However when examining very low abundance membrane proteins, this could be an advantage, as all species of a single protein are likely to be concentrated together, enabling them to be visualised and identified (Zahedi *et al.*, 2005). Several groups have recently used the 16-BAC system (Coughenour *et al.*, 2004; Dreger *et al.*, 2001), and a modified methodology aimed at providing more efficient stacking and therefore better resolution has also been described (Kramer, 2006).

4.1.2.2 Gel free separations

Gel-free separations utilise high-pressure liquid chromatography (HPLC) coupled to mass spectrometry in a technique termed Multi-Dimensional Protein Identification Technology (MuDPIT) (Washburn *et al.*, 2001). This technique separates peptides rather than whole proteins, although whole protein LC separations are now being developed (Wang and Hanash, 2005). By digesting proteins to peptides prior to separation, solubility issues associated with polytopic or amphipathic proteins can be minimised, as even highly hydrophobic proteins will generate a number of soluble peptides from which a protein identification may be made. However, MuDPIT based approaches can only differentiate between different protein isoforms if the modification is contained within a peptide identified by MS. In the same way when presented with a set of peptides common to two or more closely related proteins, it is impossible to differentiate between them. This is in contrast to gel-based systems that can distinguish different protein isoforms by virtue of their differential gel migration, irrespective of the peptides identified.

MuDPIT does massively increase sample complexity, although a number of developments have made MuDPIT a routinely applicable technique. Micro fluidic HPLC technologies are now able to deliver highly reproducible flow rates and therefore separations. Monolith column technologies are drastically improving throughput (Hart and Gaskell, 2005). Finally, advances in modern computing power and MS instrumentation have enabled MuDPIT approaches to become accessible.

4.1.2.3 Peptide generation

The high cleavage specificity of trypsin and other proteolytic agents is very useful for digesting unfolded proteins. However cleavage of sites located close to or inside TMD α-helices in natively folded proteins could potentially be sterically hindered. This is most likely to be observed in MudPIT experiments, where proteins are often digested in their native state. We therefore decided to pursue trypsin digestion in denaturing conditions e.g. 60% MeOH (Blonder *et al.*, 2004), along with proteolytic small molecules e.g. CNBr in combination with trypsin (van Montfort *et al.*, 2002b). In the same way, to increase solubilisation of hydrophobic peptides and therefore membrane proteome coverage, in-gel digestions were performed in the presence of octyl-β-glucopyranoside (n-OG) (Bierczynska-Krzysik *et al.*, 2006; van Montfort *et al.*, 2002a), a detergent which does not significantly inhibit trypsin activity at a low concentration.

4.1.2.4 Data Analysis

All of our mass spectrometry data was analysed using the MASCOT® search engine (Perkins *et al.*, 1999). This engine assigns protein identifications using the MOWSE (molecular weight search) algorithm (Pappin *et al.*, 1993). Firstly, the experimentally derived peptide masses are used to isolate a list of all peptides from the organism's genome that could encode such a mass. By altering the mass tolerance of this search a larger or smaller list will be created. For peptide mass fingerprint (PMF) searching, the presence of multiple peptides from the same protein is used to assign an ID. If performing tandem MS analysis, empirically defined data is then used to theoretically fragment each peptide in the list and compare it to the experimental tandem MS data using probability based matching to define a score. While Mascot is not the most efficient protocol in terms of its ability to utilise all the information in a mass spectrum, it is well suited to high throughput, fully automated protein identification (Resing *et al.*, 2004). The default significance threshold for a score in Mascot is set at $p = 0.05$ i.e. a false positive would only be expected to occur at random at a frequency of 5%. However it is possible to change this threshold to higher confidence levels (summarised in Table 4.1).

Search Type	Significant score threshold			
	p = < 0.05	p = < 0.01	p = < 0.005	p = < 0.0001
PMF	53	60	63	70
MS/MS	32	39	42	49

Table 4.1 – Significant threshold MOWSE scores for the *T. brucei* proteome as defined by Mascot at various confidence levels for peptide mass fingerprint (PMF) and tandem mass spectrometry (MS/MS) data. These values were generated for standard trypsin digestion (maximum 1 missed cleavage) with variable methionine oxidation and no fixed modifications.

4.2 Materials and Methods

All materials were acquired from Sigma Aldrich unless otherwise stated.

4.2.1 TbAT1-myc expression

4.2.1.1 TbAT1 in pRM481-myc construction

The primer pair MYC F (CGC GTG GCG CCG AAC AAA AAC TTA TTT CTG AAG AAG ATC TGT AGC ATG) and MYC R (CTA CAG ATC TTC TTC AGA AAT AAG TTT TTG TTC GGC GCC A) were annealed to each other (95 °C, 2 min; 95 °C, 20 s - repeat 45x dropping the temperature by 1 °C each round; 50 °C, 5 min; 25 °C, 10 min). This annealed primer pair was then ligated into MluI digested pRM481 (Promega) using T4 DNA Ligase (Promega). 100 µl of competent JM109 cells (Promega) were transformed (as per manufacturers instructions) with the ligation mixture, and then plated out on LB Agar plates containing 100 µg/ml ampicillin. 4 clones were selected (no clones grew on the plate transformed with pRM481 MluI cut plasmid) and grown up in 5 ml LB + Amp O/N. Plasmid mini-preps were made from the 4 clones and following digestion with BglII (Promega) all clones were shown to contain the expected insert and therefore designated pRM481-myc.

TbAT1 was PCR amplified from the genomic DNA isolated from s427 *T. brucei* strain (prepared by Richard Burchmore) using TbAT1 F (CGG ACA CGC GTA TGC TCG GGT TTG ACT CAG CC) and TbAT1 R (CTT GGG AAG CCC CTC ATT GAC AGC C) primers and Pfu polymerase (Promega) as follows: 94 °C, 1 min; [92 °C, 1 min; 50 °C, 1 min; 72 °C, 4 min] for 30 cycles; 72 °C, 7 min. PCR products were run on a gel. The 1.4Kb band was excised, gel purified, digested with MluI (Promega) to give 5' MluI overhangs, and then purified (Qiagen PCR clean-up kit).

The pRM481-myc vector was sequentially digested with SfoI (Promega) and MluI (Promega) to generate a 5' MluI overhang and a 3' blunt end. The MluI treated TbAT1 gene was then ligated into the vector and 50 µl JM109 cells were transformed with the ligation reaction and unligated MluI / SfoI pRM481-myc vector alone. Cells were plated out on LB + Amp and incubated overnight. 5 colonies were selected from the ligation plate and digested with SphI to test for the presence of the insert. All of the clones were positive.

The construction of the pRM481-myc vector and the insertion of TbAT1 are summarised in Figure 4.1.

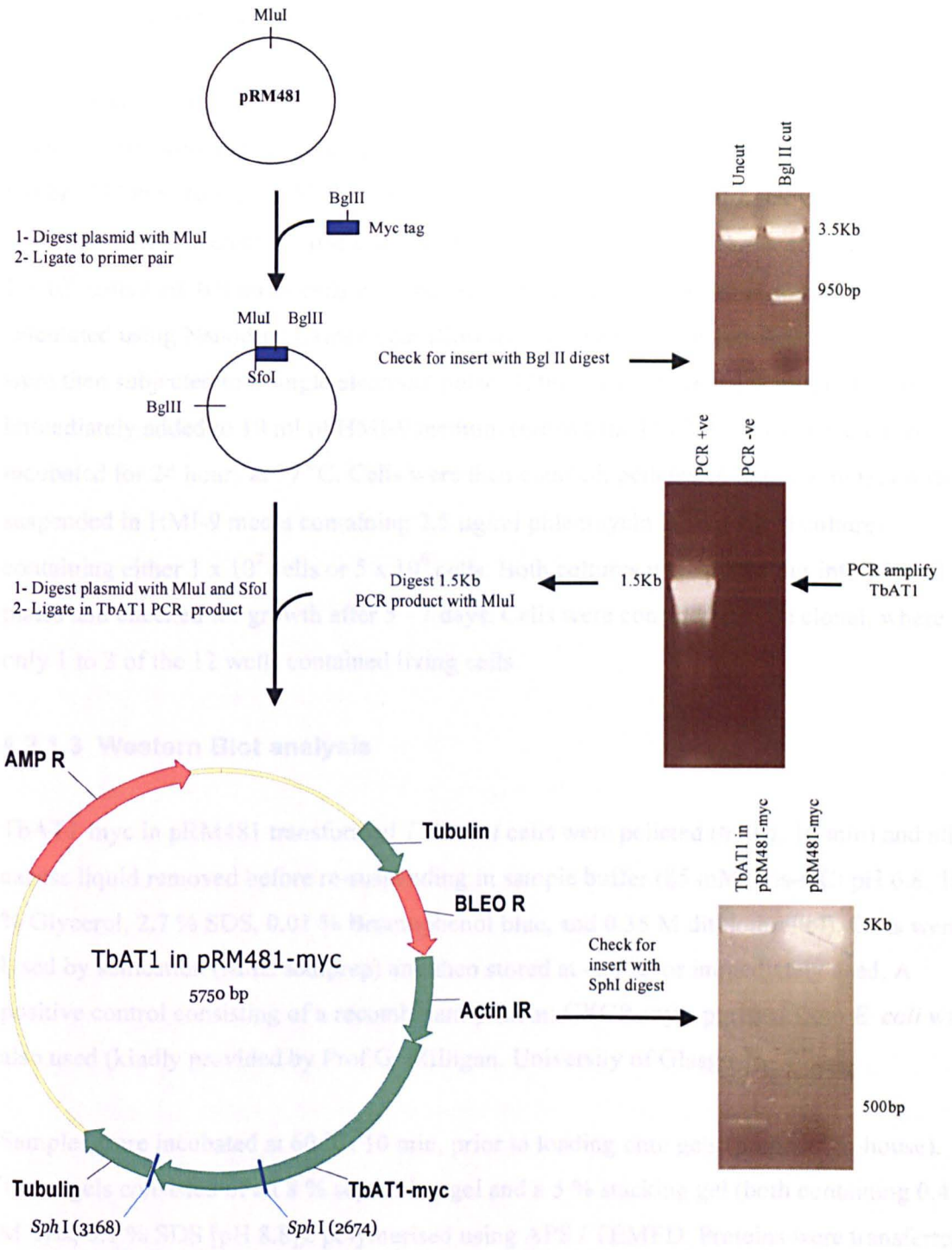


Figure 4.1 - Construction of TbAT1 in pRM481-myc for expression in *T. brucei*. All samples were run on 0.8% acrylamide gels. PCR +ve, complete mix for PCR amplification; PCR -ve, complete mix except for primers.

4.2.1.2 *T. brucei* transformation

Bloodstream form s427 strain *T. brucei* cells were grown to a density of ~5 x 10^5 cells / ml *in vitro*. Cells were then pelleted (600 g, 10 min) before being re-suspended in ZPFMG buffer (132 mM NaCl, 8 mM KCl, 8 mM $Na_2HPO_4.(H_2O)_2$, 1.5 mM $K_2HPO_4.(H_2O)_2$, 1.5 mM Magnesium acetate, 90 μM Calcium acetate, and 1g/L Glucose) at a concentration of 1 x 10^8 cells / ml. 0.5 ml of cells were mixed with 20 μg DNA in sterile water, as calculated using Nanodrop (Amersham Biosciences) in a 0.2 cm electro-cuvette. Cells were then subjected to a single electrical pulse (1500 volts / 25 μF capacity), before being immediately added to 10 ml of HMI-9 medium (pre-warmed to 37 °C). Cells were then incubated for 24 hours at 37 °C. Cells were then counted, pelleted (600 g, 10 min) and re-suspended in HMI-9 media containing 2.5 μg/ml phleomycin in two 18 ml cultures containing either 1 x 10^7 cells or 5 x 10^6 cells. Both cultures were plated out into 24 well plates and checked for growth after 5 - 7 days. Cells were considered to be clonal, where only 1 to 3 of the 12 wells contained living cells.

4.2.1.3 Western Blot analysis

TbAT1-myc in pRM481 transformed *T. brucei* cells were pelleted (600 g, 10 min) and all excess liquid removed before re-suspending in sample buffer (85 mM Tris-HCl pH 6.8, 30 % Glycerol, 2.7 % SDS, 0.01 % Bromophenol blue, and 0.35 M dithiothreitol). Cells were lysed by sonication (MSE soniprep) and then stored at –20 °C or immediately used. A positive control consisting of a recombinant protein; CXCR-myc, purified from *E. coli* was also used (kindly provided by Prof G. Milligan, University of Glasgow).

Samples were incubated at 60 °C, 10 min, prior to loading onto gels (prepared in-house). These gels consisted of an 8 % separating gel and a 5 % stacking gel (both containing 0.4 M Tris, 0.1 % SDS [pH 8.8]), polymerised using APS / TEMED. Proteins were transferred onto nitrocellulose membrane (Amersham Hyband ECL), using the Mini-Trans Blot Cell (Bio-Rad, UK) in transfer buffer (20 % MeOH, 137 mM NaCl, 20 mM Tris-HCl [pH 7.6]) at 4 °C, and a constant voltage (140 V, 45 min). Membranes were then blocked in 20 ml TBS (137 mM NaCl, 20 mM Tris-HCl [pH 7.6]) containing 0.1 % Tween20 and 5 % milk (1 hr, 37 °C). The primary mouse anti c-myc antibody (Oncogene, USA) was applied at a dilution of 1 : 100 in 20 ml TBS buffer containing 0.1 % Tween20 and 1 % milk, and incubated at 4 °C overnight. Membranes were then washed four times (15 min per wash) in 20 ml TBS buffer containing 1 % milk at room temperature. Secondary anti-mouse total IgG (Oncogene USA) was then incubated (2 hr at RT) with the blot at a 1 : 50 dilution in

20 ml CTBS buffer (1.37 M NaCl, 200 mM Tris-HCl [pH 7.6]) containing 1 % milk. The blot was then washed four times (15 min per wash) in 20 ml TBS buffer containing 1 % milk at room temperature, before developing using the ECL kit (Amersham Biosciences).

4.2.2 Preparation of Plasma Membranes

Plasma membranes were prepared in collaboration with P. Voorheis as previously described (Voorheis *et al.*, 1979), with a few alterations. Briefly, BSF trypanosomes were cultured *in vivo* in adult female Wistar rats infected by intraperitoneal injection. Blood was collected by cardiac puncture at peak parasitaemia under terminal anaesthesia. Blood was centrifuged (600 g, 10 min) to generate a parasite-enriched buffy coat. Parasites were separated from the buffy coat on a DEAE-52 (Whatman, Maidstone, United Kingdom) anion-exchange column that had been pre-equilibrated with TSB (44 mM NaCl, 5 mM KCl, 3 mM NaH_2PO_4, 57 mM Na_2HPO_4, 118 mM Sucrose, 10 mM Glucose, 0.2 mM Adenosine [pH 8.0]). Cells were then osmotically stressed by addition of water (at 4 °C), and swelling monitored by phase-contrast microscopy until a specific morphology had been achieved. Cells were then homogenised in an AO Cell Disruptor (Stansted Fluid Power, Stansted, UK), in the presence of protease inhibitors (5 μM Leupeptin, 250 μM TPCK, and 0.1 mM PMSF). The homogenate was then returned to a normal osmotic potential (140 μM) through the addition of 3 M KCl, before pelleting the cells. Cells were then treated with 240 units of DNAse in TES minus buffer (20 mM TES, 150 mM KCl, 5 mM $MgCl_2$, 1 mM 2-mercaptoethanol [pH 7.4]), for 5 min at 20 °C. The reaction was terminated by the addition of 5 volumes of TES buffer (20 mM TES, 150 mM KCl, 1 mM EDTA, 1 mM 2-mercaptoethanol [pH 7.4]). Cells were pelleted and then re-suspended in 40 % sucrose before being layered on a linear 40-60 % sucrose gradient (in TES Buffer). Samples were then centrifuged for 3 hr at 70,000 g on a SW28 rotor (Beckman Coulter). The most prominent dense white band was isolated and washed twice in TES buffer to remove sucrose, before being aliquoted and stored in TES buffer at –80 °C. All procedures were performed at 4 °C unless otherwise stated.

4.2.3 Preparation of Cytoskeleton Preparations

Cytoskeleton samples were kindly prepared and donated by P. Voorheis as follows. *In vivo* derived bloodstream form *T. brucei* isolated from whole blood (as described in section 4.2.2), were suspended in 30mM Tes-sucrose buffer (30 mM TES; 10 % w/v sucrose) containing protease inhibitors (340 μM PMSF; 0.1mM TLCK; 42 μM Leupeptin; 3 μM

Pepstatin). An equal volume of the same buffer containing Triton-X 100 (2%, w/v) was then added and incubated for 5 min at 37 °C. Samples were then layered onto a 70 % (w/v) sucrose cushion (30mM TES), centrifuged (10 min, 2500 x *g*), and the cytoskeletal band retained. This step was then repeated in 10mM Tes-sucrose buffer (10 mM Tes; 10 % w/v sucrose) to further wash the cytoskeletons. The crude cytoskeletons were then re-suspended in 10 ml of TES buffer (10 mM Tes; 110mM KCl; 5mM $MgCl_2$) and warmed to 30°C. 240 units of DNAase was then added per ml of buffer and incubated for 5 min at 37 °C. DNAse activity was terminated by the addition of an equal volume of ice-cold 30mM TES-sucrose buffer, containing 6 mM EDTA. Cytoskeletons were then washed in 30mM Tes-sucrose buffer on a sucrose cushion as above. To remove pellicular microtubules, cytoskeletons were re-suspended in 30mM Tes-sucrose buffer, to which an equal volume of 30mM TES-sucrose buffer containing 200 µM $CaCl_2$ and bovine brain calmodulin (50 µg / ml final) was added and incubated for 30 min at 30 °C. Treated cytoskeletons were then washed in 30mM TES-sucrose buffer containing 100 µM $CaCl_2$ at 0 °C and then twice more in 30mM TES-sucrose at 0 °C. The final pellet contained the cytoskeletons stripped of the pellicular microtubules.

4.2.4 Protein concentration determination

Protein concentrations were calculated using the 2D Quant Kit (Amersham Biosciences) as per manufacturers instructions, or with the Quick Start Bradford protein assay (BioRad) as per manufacturers instructions.

4.2.5 Gel-based approaches

4.2.5.1 1DGE

Samples were either run on NuPAGE 4-12 % gels with MES running buffer and LDS sample preparation buffer (Invitrogen, UK) as per manufacturers instructions, or on gels prepared in-house. These gels consisted of an 8 % separating gel and a 5 % stacking gel (both containing 0.4 M Tris, 0.1 % SDS [pH 8.8]), polymerised using APS / TEMED. Samples were solubilised in sample buffer (85 mM Tris-HCl pH 6.8, 30 % Glycerol, 2.7 % SDS, 0.01 % Bromophenol blue, and 0.35 M dithiothreitol). In both cases samples were incubated to 60 °C for 10 min prior to loading. Samples were then run until the dye had just run off the end of the gel. Protein bands were visualised by staining with colloidal

Coomassie (1.25 M ammonium sulphate, 20 % MeOH, 1.6 % ortho-phosphoric acid, 1.6 % Coomassie blue G250) for 48 hours and then washing in ddH_20.

4.2.5.2 2DGE

Samples were run as follows:

4.2.5.2.1 1st Dimension - Isoelectric focusing

Pelleted PM samples (500 μg) were re-suspended in 450 μl of rehydration buffer (8 M Urea, 2 M Thiourea, 2 % CHAPS, 0.02 % Bromophenol blue, 10 mg DTT, 5 μl IPG buffer), loaded onto plate holders with the relevant IPG strip (Amersham Biosciences), and covered with Drystrip cover fluid. Strips were then run with the following cycle on an IPGPhor machine: 10 – 15 hr rehydration, 20 °C IEF parameters 50 μA / strip; 500 V (500 Vhr); 1000 V (1000 Vhr); 8000 V (3975 Vhr); 8000 V (~5 – 10 hrs).

4.2.5.2.2 2nd dimension - SDS PAGE

After IEF, gel strips were equilibrated for 15 min in 10 ml SDS equilibration buffer (75 mM Tris-HCl pH 8.8, 6 M Urea, 30 % Glycerol, 2 % SDS, 0.002 % Bromophenol blue) plus 100 mg DTT, and then 15 min in 10 ml SDS equilibration buffer plus 250 mg iodoacetamide. Equilibrated gel strips were then placed on top of a vertical slab gel and held in place by the addition of molten agarose. Gels were then loaded in a DALT 12 gel tank filled with SDS electrophoresis buffer (25 mM Tris-HCl, pH 8.3, 192 mM Glycine, 0.2 % SDS), and run at 15 Watts per gel (at 25 °C) until the dye front reached the end of the gel.

4.2.5.2.3 Spot Picking

Gels were scanned and spots selected using the 2D-Elite software (Amersham Biosciences). A pick list was generated from this and used by the Ettan Spot Handling Workstation (Amersham Biosciences) to pick, digest and spot samples.

4.2.5.3 16-BAC

4.2.5.3.1 Gel Preparation

Gels were prepared as with standard SDS-PAGE slab gels, complete with a stacking gel on top of a separating gel. An 8 % separating gel was prepared (3 M urea, 7.3 % acrylamide, 0.3% bis-acrylamide, 3.75 mM ascorbic acid, 7.5 µM ferrous sulphate, 2.4 mM 16-BAC, 70 mM potassium phosphate [pH 2.1]), polymerised by the addition of H_2O_2 (final concentration 28 µM), and overlaid with 75 mM potassium phosphate [pH2.1] overnight to allow complete polymerisation. A 4 % stacking gel (as above, with acrylamide / bis-acrylamide adjusted) was then prepared on top of the separating gel (H_2O_2 final concentration of 60 µM).

4.2.5.3.2 1st Dimension

Plasma membrane pellets were re-suspended in an appropriate volume of sample buffer (3.75 M urea, 125 µM 16-BAC, 0.7 M glycerol, 37.5 mM DTT, 0.025 % Pyronin Y) and incubated at 60 °C for 5 min. Before loading onto the gel, samples were spun down (3 min, 13k rpm) to remove any residual insoluble material. Gels were run towards the cathode in running buffer (2.5 mM 16-BAC, 150 mM Glycine, 50 mM Phosphoric acid) at 30 mA (for 30 min), before increasing to 140 mA until the dye front reaches the end of the gel. Gels were then fixed for 1 hour in 35 % isopropanol, 10 % acetic acid (change solution every 15 min), before being stained for 30 min (0.15 % Brilliant blue R-250, 7.5 % Acetic acid, 50 % MeOH). Gels were then destained (1 x 15 min) in wash solution I (50 % EtOH, 7.5 % Acetic acid, 5 % Glycerol), before being further destained (3 x 20 min) in wash solution II (5 % EtOH, 7.5 % Acetic acid, 5 % Glycerol).

4.2.5.3.3 2nd Dimension

Gels were re-equilibrated (2 x 10 min) in 100 mM Tris-HCl [pH 6.8], the relevant lanes excised (taking care to ensure a straight edge is cut for linear association with the second dimension gel). Gel lanes were then treated as IPG strips and equilibrated / run as above. Spots were either picked using the Ettan Spot Handling Workstation (as above) or by hand, and placed in a 96-well plate.

4.2.5.3.4 Picking and digest

Spots or bands of interest were excised either manually or with the Ettan Spot Handling Workstation (Amersham Bioscience) into 96-well plates. Samples were then digested with trypsin as follows. Briefly, gel pieces were washed five times (3 x 100µl 50 mM AmBic, 50 % MeOH; 2 x 100 µl 75 % ACN), before being dried for 60 min at 37 °C. Samples were then re-hydrated by the addition of 0.5 µg trypsin (in 25 µl 20 mM AmBic), and incubated for 240 min at 37 °C. Peptides were extracted from the gel plugs, by washing twice (2 x 100 µl 50 % ACN / 0.1 % TFA) and transferred in solution to a fresh 96 well plate, where samples were dried down in a SpeedVac (Eppendorf Concentrator 5301) until ready for MS analysis.

4.2.6 MuDPIT based approaches

4.2.6.1 Trypsin Digestion of PM sheets

Plasma membrane pellets were re-suspended in 200 µl 25 mM AmBic, 0.1 % n-OG containing 400 ng trypsin, by repeated vortexing and pipetting. Samples were then incubated for 12 - 16 hours at 37 °C, before being dried down in a SpeedVac (Eppendorf Concentrator 5301) until ready for MS analysis.

4.2.6.2 CNBr and trypsin digestion of PM sheets

Pelleted plasma membrane sheets were dried down in a SpeedVac (Eppendorf Concentrator 5301) before being re-suspended in 50 µl 75 % trifluoroacetic acid (TFA). Two small crystals of CNBr were then added to the sample and incubated for 12 - 16 hours at room temperature (with the lid closed). Samples were then incubated for a further 6 hours at room temperature with the lids open in the fume cupboard to allow complete evaporation of contents. To neutralise the pH, samples were re-suspended in 250 µl 100 mM tri-ethyl ammonium bicarbonate (TEAB), and then dried down in a SpeedVac (Eppendorf Concentrator 5301).

To digest with trypsin, samples were re-suspended in 20 µl 6 M Urea, 1 % n-OG, 100 mM TEAB by vortexing and pipetting, before being diluted 1 : 10 with 25 mM TEAB containing 400 ng trypsin. After incubating at 37 °C for 12 - 16 hours samples were dried down in a SpeedVac (Eppendorf Concentrator 5301) until ready for MS analysis.

4.2.6.3 Trypsin digest in 60% MeOH

PLASMA MEMBRANE pellets were re-suspended in 40 µl 50 mM AmBic and placed in a sonicator bath for 20 min (vortexing regularly), before being incubated at 60 °C for 20 min. Samples were then placed on ice for 3 min, before adding 60 µl MeOH and incubating for a further 5 min in the sonicator bath (with vortexing). Trypsin was then added (16 µl at 20 µg/ml in 25 mM AmBic), before the addition of 24 µl MeOH to give a final concentration of 60 % MeOH. After brief vortexing, samples were then incubated at 37 °C for 12 - 16 hours. The digested sample was dried down to ~5 µl volume in a SpeedVac (Eppendorf Concentrator 5301), adjusted to 85 µl with 1 % Formic acid and then stored at –20 °C until MS analysis.

4.2.7 Mass Spectrometry Analysis

4.2.7.1 MALDI

Samples undergoing MALDI analysis were spotted onto target plates using the Ettan Spot Handling Workstation (Amersham Biosciences). Briefly, dried samples were re-suspended in 3 µl 50 % acetonitrile (ACN), 0.5 % TFA and 0.3 µl spotted onto the target plate. Immediately, 0.3 µl of the matrix; α cyano-4-hydroxycinamic acid (10 mg/ml CHCA in 50 % ACN, 0.5 % TFA) was then added to the sample on the target and mixed. Samples were analysed using the Voyager-DE™ PRO workstation (Applied Biosystems) or the 4700 proteomics analyser (Applied Biosystems). For the Voyager, an average 200 shots per spectrum were collected. Approximately 3,000 laser shots were fired for peaks selected for MS/MS analysis on the 4700 proteomics analyser.

4.2.7.2 ESI-MS/MS

Peptides undergoing electrospray ionisation (ESI) MS (on a QSTAR® XL Hybrid LC/MS/MS System) were separated on an LC system (Famos / Switchos / Ultimate, LC Packings). Simple peptide mixtures i.e. those from gel-resolved samples, were separated on a single dimension using a Pepmap C18 reverse phase column (LC Packings), using a 5 – 85 % acetonitrile gradient (in 0.5 % formic acid) run over 45 min. In the case of more complex samples i.e. MuDPIT based approaches, samples were separated on 2 dimensions by a strong cation exchange column (SCX, 1D 0.5 mm, LC Packings), using 0, 20, 40, 60, 80, 100, 150, 300 and 1000 mM KCl cuts in the first dimension and then on a Pepmap C18

reverse phase column (LC Packings) in the second dimension using a 5 - 85% acetonitrile gradient (in 0.5 % formic acid) run over 90 min. Flow rate was maintained at 0.2 µl / min.

Mass spectrometry analysis was performed using a duty cycle consisting of a 3 second survey MS scan, followed by up to four MS/MS analyses of the most abundant peptides (3 second per peak).

4.2.8 MS Data Analysis

Data generated from the Voyager-DE ™ PRO, and the Q-STAR® XL hybrid mass spectrometers were analysed using Applied Biosystems Data Explorer (v 4.0) and Analyst QS (v1.1) software respectively via the automated Mascot Daemon server (v2.1.06). Data generated from the 4700 mass spectrometer was analysed using GPS Explorer (v3.5). All of these systems define protein identifications using the Mascot search engine (Perkins *et al.*, 1999) to assign probability based MOWSE scores, based on comparisons using a local database containing the *T. brucei* genome (version 4 downloaded from ftp://ftp.sanger.ac.uk/pub/databases/T.brucei_sequences/ T.brucei_genome_v4/). In all cases variable methionine oxidation, and carbamidomethylation as a fixed modification (where iodoacetamide treated), was used for searches. An MS tolerance of 70 ppm (Voyager-DE ™ PRO) or 1.2 Da for MS and 0.4 Da for MS/MS analysis (Q-STAR® XL and 4700) was used. MALDI data was baseline corrected, noise filtered, de-isotoped and mass calibrated using the auto-proteolytic trypsin fragments. Only proteins identified with a significant score ($p = < 0.05$) were included. Where the same protein was identified multiple times, only the highest scoring identification was included in the results.

4.2.9 Protein Analysis

Quality control analysis was performed on the experimental protein lists derived from the cytoskeleton and plasma membrane preparations as follows. Initially all genes were scanned for keywords (see Table 4.2) that localised the protein to a particular cellular compartment or region. Any genes not containing a keyword in their annotation, but being predicted to contain 5 or more TMD's were assigned a 'Membrane' localisation. Each remaining gene's cellular localisation was then manually curated against www.genedb.com. In each case the evidence for the localisation assigned was also recorded (Table 4.3).

Localisation	Keyword
Acidocalcisomes	acido
Cytoplasm	tryparedoxin
Cytoskeleton	actin, tubulin, kinesin, dynein, centrin, microtubule-associated
Endosome	endos
ER	endoplasm
Flagellum	flagell
Glycosomal	glycosom
Golgi	golgi
Intracellular	ribosome, calpain, DNA, RNA, adrenodoxin
Kinetoplast	kineto
Lysosome	lyso
Membrane	adenylate / adenylyl cyclase, cytochrome, GPI
Mitochondria	mitochon, metacaspase
Nucleus	nuclear , nucleolar, nucleoporin, histone, retrotransposon
Peroxisome	perox
Plasma membrane	VSG
Unknown	adenylate kinase, ADP ribosylation factor, chaperone, arginine kinase, calmodulin, heat-shock protein, leucine rich repeat protein

Table 4.2 – Localisations assigned to proteins containing a keyword. Case insensitive searches were performed using an in-house generated PERL script.

Evidence	Explanation
Annotated as	Cellular localisation contained in gene annotation. These genes have not been checked against GeneDB.
GeneDB	Localisation derived from GeneDB
TbFP	Previously identified in the flagellar proteome (Broadhead *et al.*, 2006)
5+ TMD	Contains 5 or more TMD's and therefore are assumed to be membrane proteins
Gene Name	Contains a keyword that suggests it is localised to a particular cellular location e.g. actin assigned to cytoskeleton. These genes have not been checked against GeneDB

Table 4.3 – Explanation of the type of evidence used to assign each gene a cellular localisation.

4.3 Results

4.3.1 Tagged TbAT1

We generated a myc- tagged TbAT1 gene by modifying and then inserting the gene into the pRM481 vector (McCulloch and Barry, 1999) for overexpression in *T. brucei*. This vector targets the α- β- tubulin locus for genomic integration. Unfortunately, despite using a number of different antibodies (data not shown) and analysing multiple transformants, apart from the positive control, we were unable to identify any tagged protein in the cell extracts (Figure 4.2). There are a number of potential reasons for this. Firstly it is possible that the tagged construct was at such a low level that detection was not sensitive enough to generate a signal. However the strong signal generated by the positive control suggests that the TbAT1-myc protein was not expressed. Very little is known about the stability of TbAT1 in the plasma membrane and the -myc tag could have affected protein stability, tertiary folding or insertion into the membrane. Whatever the cause, for a lack of functional TbAT1-myc, we were left without a convenient tool to identify the most effective techniques for enriching for membrane proteins. We therefore had to switch to a mass spectrometry based approach whereby proteomic techniques and subsequently enrichment techniques were evaluated based on the proteins that were identified from them.

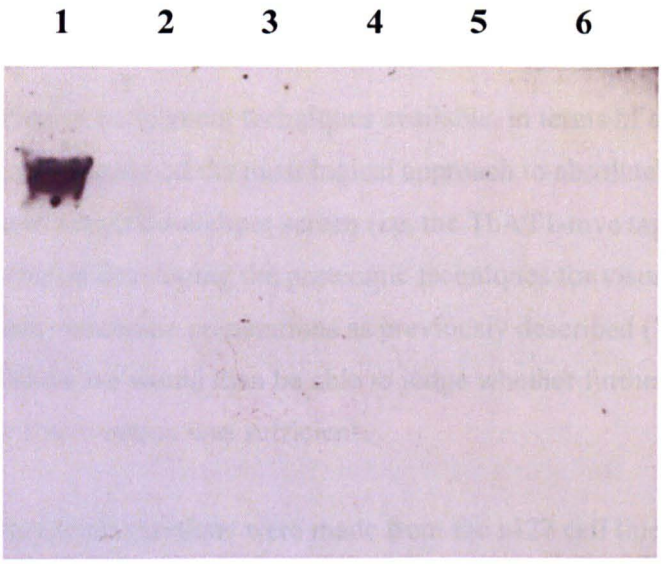


Figure 4.2 - Western blot analysis of cell extracts probed with mouse anti c-myc primary antibody and anti-mouse total IgG secondary antibody, and visualised by ECL. Cell extracts are present in lanes: 1, CXCR-myc tagged protein (positive control); 2 - 5, Clones 1-4 TbAT1 in pRM481-myc; 6, pRM481-myc transformed cells (negative control). Blot was exposed to film for 30min.

4.3.2 Membrane Preparations

Of all the fractionation or enrichment techniques available, in terms of analysing the plasma membrane, we considered the most logical approach to absolutely require fractionation. Without a high throughput screen (i.e. the TbAT1-myc tagged line) we decided to concentrate on developing the proteomic techniques for visualising membrane proteins using plasma membrane preparations as previously described (Voorheis *et al.*, 1979). Once established, we would then be able to judge whether further enrichment was required or whether fractionation was sufficient.

Three plasma membrane preparations were made from the s427 cell line. To examine reproducibility of the three samples, 15 μg of each were run on a 1D SDS PAGE gel (Figure 4.3). From this gel, it would appear that the sample reproducibility, at least at the gross level, is very high. In addition, it can be seen that the vast majority of proteins have a M_{wt} of 50 kDa or more, with a particularly high density of proteins at just over 50 kDa.

4.3.3 Gel-based Approaches

4.3.3.1 2DGE

115 μg or 500 μg of PM prep (Figure 4.4) and the equivalent of 9×10^7 cells of cytoskeleton prep (Figure 4.5) underwent a 2DGE analysis; using pH 3-10NL IPG strips and 12% second dimension gels. A total of 341 and 161 visible spots were picked from the plasma membrane and cytoskeleton prep gels respectively. For the PM gels, where samples yielded good PMF data, but where identifications were either not made or with low confidence, samples were re-run on the ESI-MS/MS Q-STAR instrument.

A total of 139 protein identifications representing 99 different proteins were made from the PM prep gels (see Appendix II). In comparison 356 protein identifications were made from the cytoskeleton prep gel, yet this represented only 83 different proteins (see Appendix III). This suggests that the cytoskeleton samples contain a relatively small number of proteins, but many of which are present as a variety of different protein isoforms i.e. as a result of post-translational modifications or truncations. In a similar manner, 72 out of 161 spots picked from the cytoskeleton prep gel gave one or more proteins identifications, whereas only a minority (84 out of 341) of the spots from the PM prep gels gave positive protein identifications. This was despite additional attempts using tandem mass spectroscopy to improve protein identifications from the PM 2DGE spots.

By comparing the two different protein datasets, we found that 52 proteins were common to both (see Appendix II or III). In view of the methods used to generate the samples and the propensity for 2DGE systems to preferentially resolve soluble hydrophilic proteins, these proteins are most likely to be of cytoskeletal origin. By comparing the spot migration patterns of the two samples a number of similarities are immediately identifiable, suggesting that the PM preps are contaminated with cytoskeleton proteins.

4.3.3.2 16-BAC Gels

A number of gels were run with plasma membrane preparations, both to define the loading capacity of the gel and to perfect the 'art' of running 16-BAC gels. It was noted that particular care needed to be taken to ensure that the polymerisation process had reacted to completion. All of the spots from two gels (334 spots) were excised and analysed by MALDI. Where samples yielded good PMF data, but where identifications were either not made or with low confidence, samples were re-run on the ESI-MS/MS Q-STAR instrument. In addition, 9 gel sections with dense Coomassie staining were excised and analysed by 2D LC-MS/MS. From both analyses, a total of 281 unique proteins were identified (see Appendix IV)

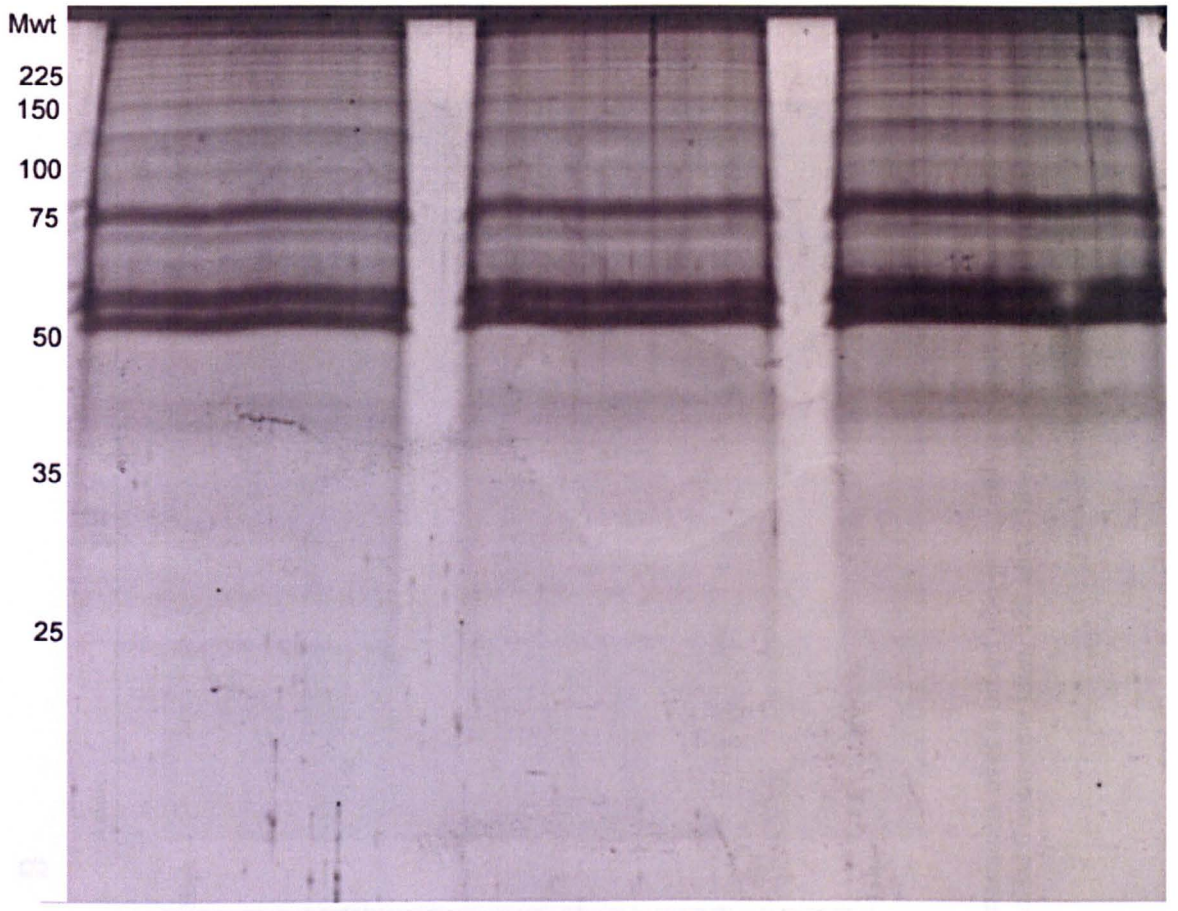


Figure 4.3 - Three samples of fractionated plasma membranes from s427 cells run on a 12% SDS PAGE gel. Each lane was loaded with 15μg of protein and stained with colloidal Coomassie. Molecular weight (M_{wt}) markers (kDa) are shown on the left-hand side.

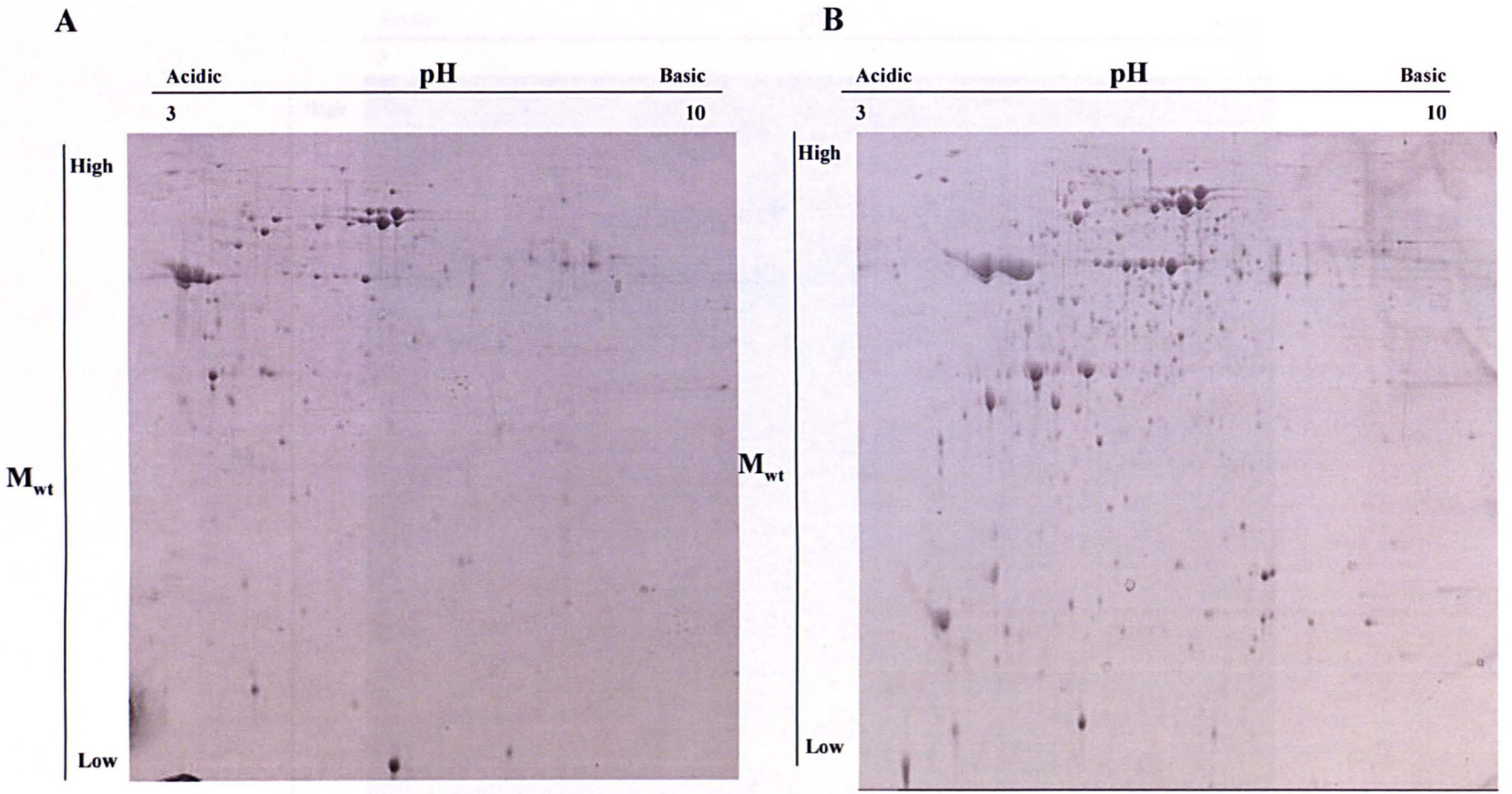


Figure 4.4 - 2DGE of PM preps derived from s427 cell line. Proteins separated in 2 dimensions over a broad non-linear pH range (pH 3-10) and molecular weight. Gels were loaded with 115μg (A) or 500μg (B) of protein and stained with colloidal Coomassie blue. 126 and 215 spots were picked from (A) and (B) respectively for protein identification.

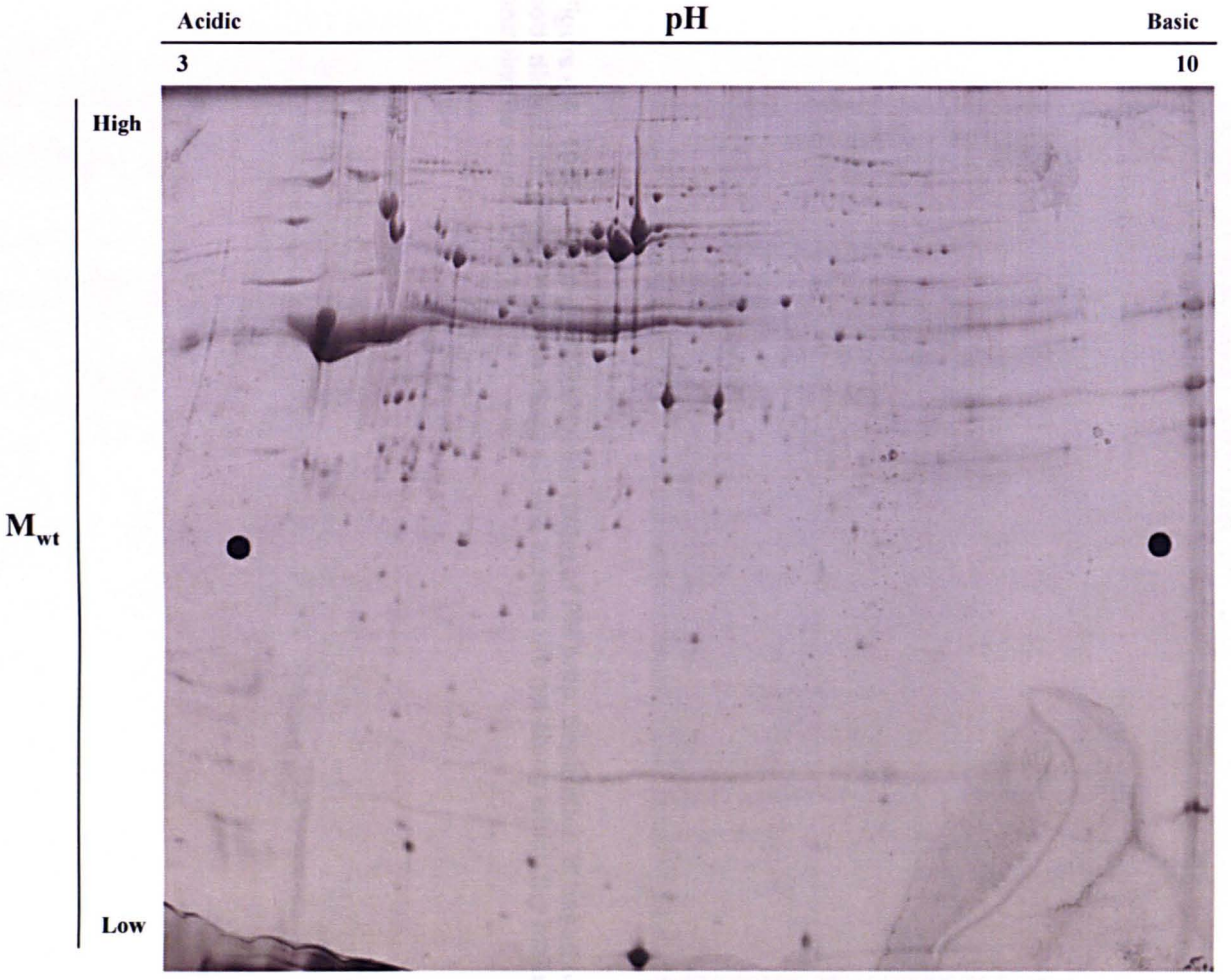


Figure 4.5 - 2DGE of cytoskeleton preparation derived from s427 cell line. Proteins separated in 2 dimensions over a broad non-linear pH range (pH 3-10) and molecular weight. Gel was loaded with the equivalent of 9 x 10^7 cells and stained with colloidal Coomassie blue. 160 spots were picked from the gel for protein identification.

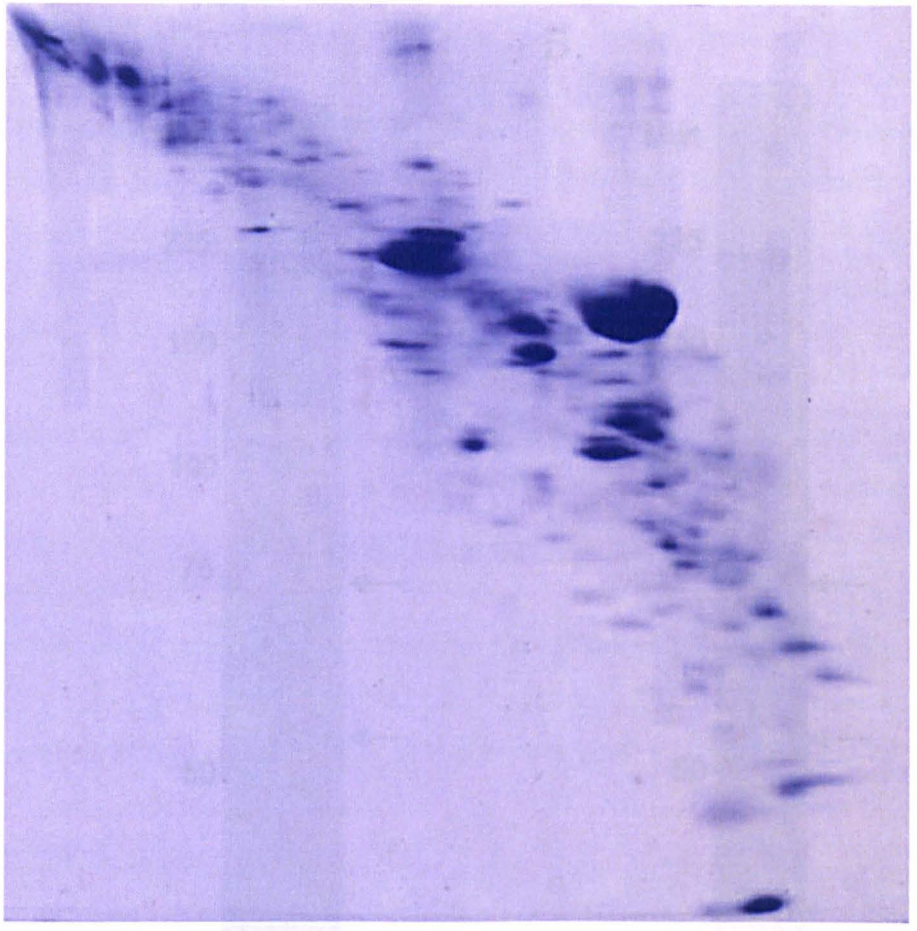


Figure 4.6 - 16-BAC mini-gel loaded with 100μg of PM prep. Separation is based on Mwt in both directions, giving rise to a diagonal line across the gel along which the majority of proteins lie on or close to.

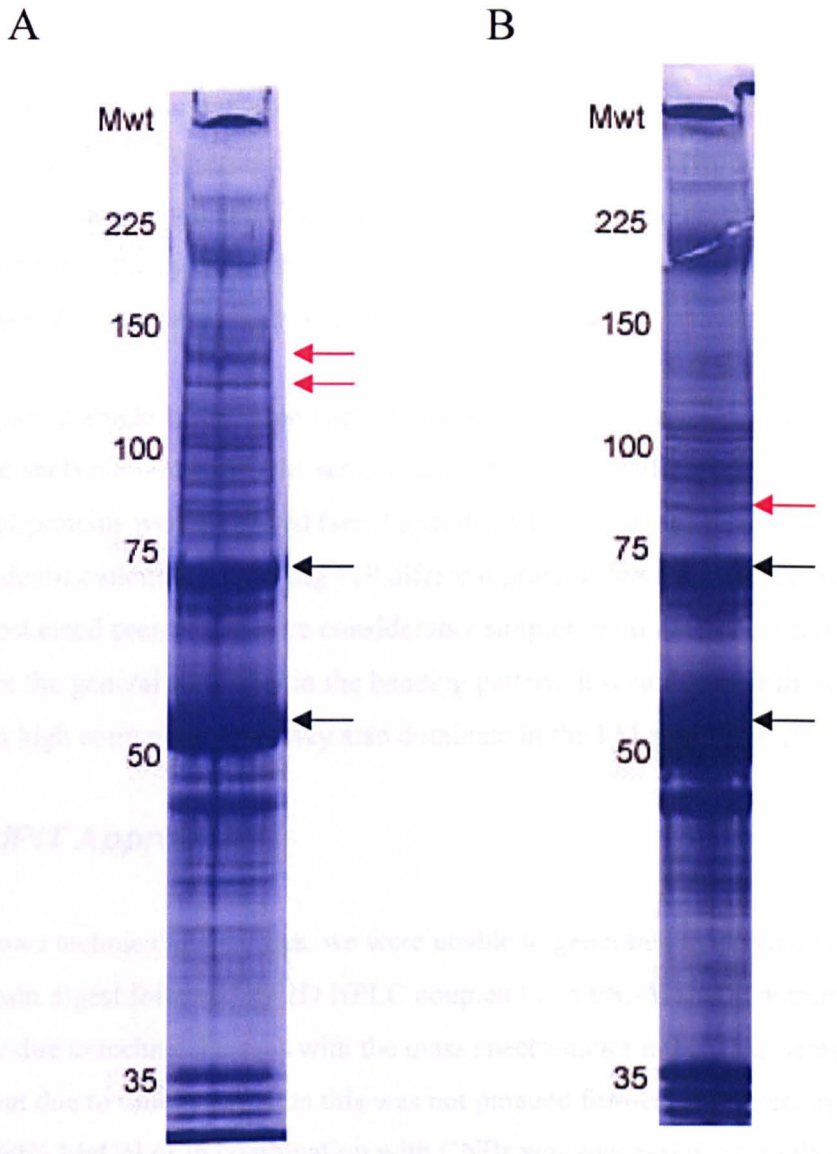


Figure 4.7 –1D PAGE gel loaded with 20μg of plasma membrane sample (A) or cytoskeletal sample (B). Gels were stained with colloidal Coomassie, and molecular weight (kDa) markers (Mwt) are shown on the left hand side. Representative bands common to both gel lanes (←) or uniquely found in a single lane (←) are also shown.

4.3.3.3 1DGE

Both plasma membrane and cytoskeleton-enriched (generated by P. Voorheis) samples were analysed by 1DGE (Figure 4.7). The banding pattern from the two samples is different e.g. there are a number of bands in the 100-150 kDa range present in the PM sample, but absent in the cytoskeletal preparation, although a number of prominent bands e.g. at 50 kDa and 75 kDa are common to both.

For MS analysis, a single lane loaded with 115μg of protein from either sample was sectioned and analysed. From the PM sample, 2721 protein identifications representing 1233 different proteins were observed (see Appendix V), whereas for the cytoskeletal prep, 792 protein identifications representing 719 different proteins were made (see Appendix VI). The cytoskeletal prep is therefore considerably simpler in number of proteins present, however from the general similarity in the banding pattern, it would appear these proteins are present in high copy number as they also dominate in the PM sample prep.

4.3.4 MudPIT Approaches

Due to unknown technical difficulties, we were unable to generate any protein ID's from a standard trypsin digest followed by 2D HPLC coupled to an ESI-MS/MS instrument. This was probably due to technical issues with the mass spectrometer rather than sample generation, but due to time constraints this was not pursued further. However, trypsin digestion in 60% MeOH or in combination with CNBr was successful. From the former analysis, 306 different proteins were made (see Appendix VII), compared to 202 different proteins (see Appendix VIII) in the latter. 147 proteins were common to both datasets.

4.3.5 Technique comparisons

To allow comparisons to be made between the different techniques, only a single experiment e.g. results from one gel or one MuDPIT experiment are used to compare the different approaches, and are summarised in Figure 4.8. From these Venn diagrams, it can be seen that the 1DGE approach far exceeded any other approach in the number of proteins it was able to identify. This may in part have been due to all samples having been analysed by ESI-MS/MS, which is known to generate more information than a MALDI MS/MS system. However, only a few hundred spots were visualised and picked from the 2D gels, while the entire 1DGE lane was sectioned and submitted for analysis. In addition, the limitations of 2DGE in relation to the iso-electric focusing step being unable to resolve

membrane proteins was always likely to limit the ability of this technique to resolve this sample. In contrast the 16-BAC system was able to identify a sub-set of proteins that were not identified by any other approach. In the same way, while the gel-based approaches identified the majority of all proteins, a significant number were uniquely identified by gel-free based systems. From this analysis, for a high-throughput analysis, 1DGE would be the technique of choice, however to maximise proteome coverage (as aimed for in this chapter), as many different techniques as is practical need to be performed.

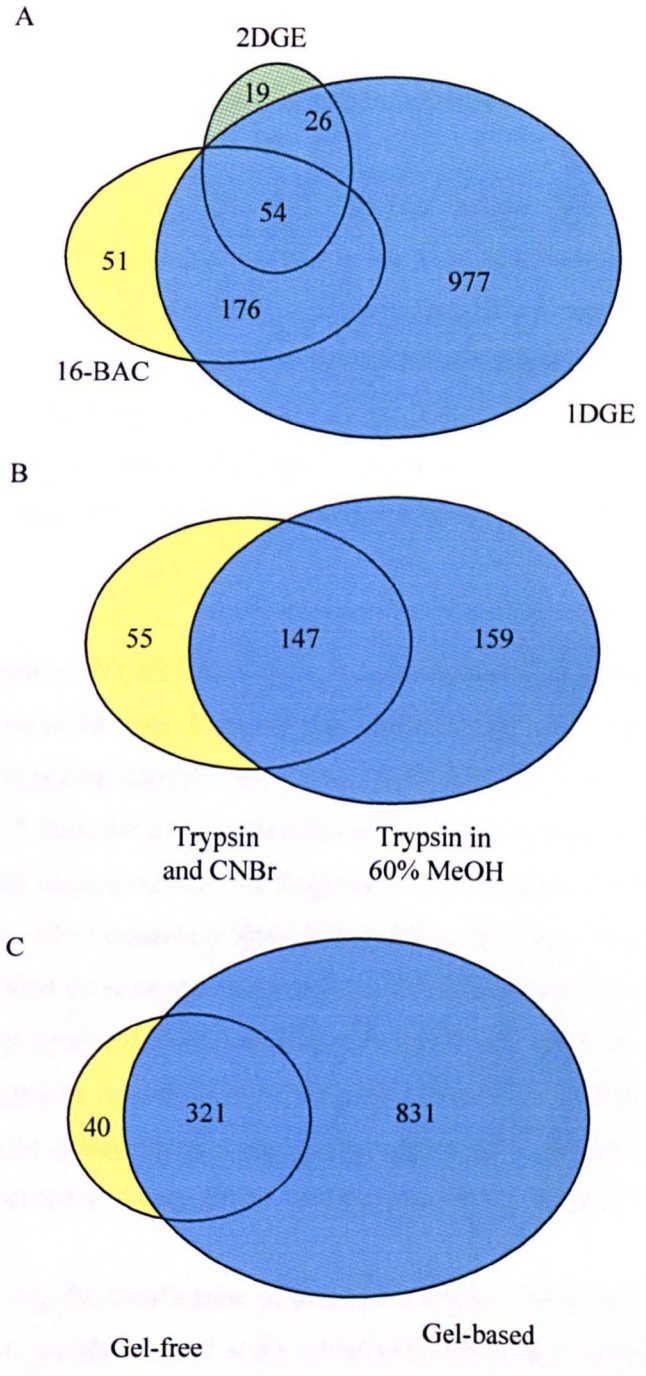


Figure 4.8 - Summary of protein identifications from all proteomic approaches applied to the plasma membrane enriched sample. All gel-based (A), gel-free / MuDPIT analyses (B) and a comparison between the two (C) are shown.

4.3.6 The plasma membrane sub-proteome

All experiments performed were pooled and analysed together. We were aware of the number of potentially contaminating proteins in the plasma membrane fraction, and therefore pursued a subtractive approach to remove potential contaminants. This approach consisted of the analysis of three proteomic datasets derived from the plasma membrane preparations, the cytoskeleton preparations, and the flagellar proteome, as previously reported (Broadhead *et al.*, 2006). The latter two datasets were combined into a single one termed **TbCF** (see Appendix IX), predicted to be primarily composed of cytoskeletal and flagellar proteins.

A total of 1321 proteins were identified from the plasma membrane enriched sample using the various proteomic techniques. Many of these proteins were obvious cytoskeletal contaminants. We therefore applied a subtractive approach where all proteins identified in **TbCF** were removed from the plasma membrane list. While it is possible that a small subset of proteins are integral to both the flagellum / cytoskeleton and plasma membrane, we felt that these would be extremely limited in number. All remaining genes were then analysed (by annotation or keyword searching) for any indication as to their cellular localisation. Proteins localised to intracellular non-membrane location were then also removed from the list (see Appendix X). This gave a final list of 566 proteins constituting the core sub-proteome of long-slender bloodstream forms of *T. brucei* (**TbPM**). The complete list is presented in Appendix XI, and summarised in Table 4.7.

In addition to analysing the localisation of proteins identified from the plasma membrane enriched preparation, we also looked at the cytoskeleton enriched sample (summarised in Table 4.4). This table clearly shows that there is contamination of both preparations with proteins that have been localised to regions other than the plasma membrane or cytoskeleton. However it does show that there is considerable enrichment for membrane proteins in the plasma membrane preparation.

By predicting cytoskeletal / flagellar proteins to be predominately soluble high-copy number proteins in contrast to polytopic low-copy number plasma membrane proteins, we predicted proteins identified in the **TbPM** to be characterised by a high number of trans-membrane domains and a lower average gCAI (see Chapter 2). On the other hand proteins shared between the plasma membrane prep and **TbCF** would comprise relatively few polytopic membrane proteins with a higher average gCAI. Both of these characteristics can

be seen, with an abundance of polytopic membrane proteins found in the plasma membrane prep (Table 4.5) and an overall reduced gCAI value for the same protein set (Table 4.6).

Localisation	Number	
	PM	TbCF
Cytoskeleton & Flagellum	16	123
Intracellular	117	222
Membrane	79	7
Plasma Membrane	46	13
Unknown *	441	425
Total	**699**	**790**

* - Includes all proteins annotated as hypothetical

Table 4.4 – Summary table of all proteins identified from the plasma membrane (PM) enriched sample and cytoskeleton sample combined with the flagellar proteome (TbCF).

Number of TMD's	All Protein ID's	Proteins found in		
		TbPM	TbCF	Shared
0	1289	379	798	596
1	110	75	25	15
2	50	39	4	2
3	18	10	5	3
4	11	9	1	1
5	7	7	-	-
6	6	5	1	1
7	7	6	1	1
8	6	6	-	-
9	7	5	2	2
10	7	7	-	-
11	10	10	-	-
12	2	2	-	-
13	1	1	-	-
14	4	4	-	-
22	1	1	-	-
Total	**1536**	**566**	**837**	**621**

Table 4.5 - Distribution of proteins in the different fractions demonstrating the enrichment of polytopic membrane proteins identified in the plasma membrane sub-proteome (TbPM), the cytoskeletal / flagellar proteome (TbCF), against proteins identified in both the TbPM and TbCF (Shared).

	All Proteins	Proteins found in		
		TbPM	TbCF	Shared
Average	67.5	57.4	73.2	74.8
Min	0.1	0.1	1.5	3.4
Max	100.0	99.2	100.0	100.0
Total	**1536**	**566**	**837**	**622**

Table 4.6 - Average, minimum (Min) and maximum (max) gCAI values for proteins found in the plasma membrane fraction (TbPM), the cytoskeletal / flagellar proteome (TbCF), or proteins identified in both (Shared).

Number of TMD's	Total Proteome	TbPM	
		Number identified	%
0	8311	379	4.6
1	1007	75	7.4
2	394	39	9.9
3	174	10	5.7
4	116	9	7.8
5	65	7	10.8
6	64	5	7.8
7	37	6	16.2
8	42	6	14.3
9	29	5	17.2
10	52	7	13.5
11	62	10	16.1
12	24	2	8.3
13	13	1	7.7
14	11	4	36.4
15	1	-	-
18	2	-	-
20	1	-	-
22	2	1	50.0
25	1	-	-
Total	**10408**	**566**	**5.4**
>4	**406**	**54**	**13.3**

Table 4.7 - Comparison of all ORF's present in the *T. brucei* genome (Total Proteome) against all experimentally identified proteins found in the plasma membrane sub-proteome of bloodtream forms of *T. brucei* (TbPM). Values are expressed as a total number (Number identified) and as a percentage (%) of the Total Proteome.

To examine whether the biophysical properties of membrane proteins make their identification in a sample more difficult, we looked at the effect of increasing the confidence level on protein identifications made (Table 4.8). In fact it appears that increasing the threshold for a score to be considered significant has a fairly linear effect on all proteins, as the confidence level is increased. Interestingly the impact on average gCAI is most pronounced when increasing the confidence threshold from 95 to 99 % confidence (Table 4.9). This suggests that proteins with lower gCAI values (and subsequently those predicted to be expressed at a lower level) are harder to confidently identify by MS. This would most likely be due to only identifying one or two peptides, which fail to give a significant score. This correlates well with other reports, whereby protein abundance correlated directly to number of peptides identified (Liu *et al.*, 2004).

Number of TMD's	Confidence threshold (p = >)			
	0.05	0.01	0.005	0.001
0	379	255	227	191
1	75	68	61	56
2	39	35	34	27
3	10	9	9	5
4	9	8	6	5
5	7	7	6	5
6	5	4	3	3
7	6	5	4	4
8	6	5	5	5
9	5	5	5	5
10	7	5	5	5
11	10	9	9	9
12	2	2	2	2
13	1	0	0	0
14	4	4	4	4
22	1	1	1	1
>4	54	47	44	43
Total	566	422	381	327

Table 4.8 - Effect of increasing the confidence level associated with a significant score on the number of proteins identified in the membrane sub-proteome (TbPM). Proteins in the table have been differentiated by the number of trans-membrane domains (TMD's) that they are predicted to encode.

gCAI	Confidence threshold (p = >)			
	0.05	0.01	0.005	0.001
Average	57.4	60.9	61.5	63.0
Min	0.1	0.4	0.4	5.1
Max	99.2	99.2	99.2	99.2

Table 4.9- Effect of increasing the confidence level associated with a significant score on the gCAI of proteins identified in TbPM. Values are given as average, minimum (Min) and maximum (Max).

4.3.6.1 Hydrophobic peptides

As mentioned above, within the **TbPM** a number of highly hydrophobic proteins were identified. Perhaps the most intriguing finding was the presence of a 22 TMD protein encoding a putative calcium channel (Tb10.70.4750). There are only 5 proteins in the entire genome that are predicted to have more TMD's. Initially we thought it possible that the protein that had been identified was a truncation of the full transcript and that perhaps only half or so (i.e. 11 TMD) was expressed. However closer examination revealed peptides identified along almost the entire length of the protein (Figure 4.9), suggesting that the entire ORF is translated. The sheer number of peptides identified would suggest that this protein is present in the sample at high levels (Liu *et al.*, 2004).

Much has been made of the inability of hydrophobic peptides to be ionised and therefore seen by MS. For this reason Xiang *et al* (Xiang *et al.*, 2004) purposefully aimed to deplete these peptides from their samples. However, other groups have reported the identification of TMD-containing peptides (Fandiño *et al.*, 2005), although they were working with purified membrane proteins. Interestingly two of the identified peptides covered regions predicted to be part of a TMD (Figure 4.9). Looking closely at one of these peptides (Figure 4.10), while there is an almost complete *y* ion series, there does appear to be better ion fragment representation in non-TMD predicted regions e.g. ANDPETG. This does suggest that ionising and therefore identifying these peptides is rare. Interestingly the identification of these peptides from a trypsin digest of 1D SDS-PAGE gel band automatically proves that trypsin has been able to cleave both very close to and within a predicted TMD.

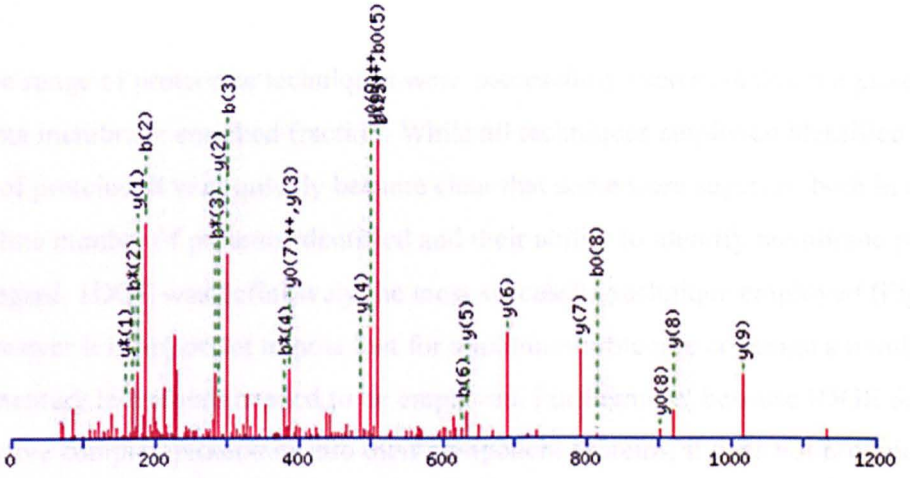
```
   1 MAEPPPPQPL RKGFFRADPH GLRPPESPRD AEEPENHDIV EDPPLPSSRN
  51 LPPPRLRPMG GGPPLAGAEA AATSHSNNRP SPGEVPPLPR PGTTTAVTSA
 101 QYNSLTPRQE LFSSPQTAVE ENIPAGYDNQ MNVTSNDFSH SQDTGIKRIG
 151 DGATGSSAKT RSTIAHAPSG RFSEVVQERP PIESVEQLCQ LQTLAHPVER
 201 FFVEYQPMAT RGCRSTQVPV EEDEALFNHL AEYTDDDSFF SSGSAVLEED
 251 TRRLQNDPLR GRAIRSVFET LSLRSLTKGR IPYELEDQER SSAPWLKQVH
 301 QRAIALHHSS FFIFPAGMKA RVIAYNILHH WLTEMLLMLI ILVYSMMTAA
 351 WSRDTWPTLE KPSFMFFADV FFLCIYAIEF VARLFTSGAV SHSRAYFRSP
 401 WHCLDAAVLL LMILNCTNLQ SMWNFSAFRL IRVLKSSTYV PIPINMKLLA
 451 KSLLRSTSNV VKVSTILFYV LLFFSLVGLQ LFSGVLQHRC VSPTTKNVTN
 501 QLCRFNHSEK NESYYHGATC PSPHLCVADT YGNPHHGYRS FDSVGHSFLS
 551 AFQIMTFQGW TSLLQETSDT TSVAAILYFF LAILICAWII PSLYLGVFIE
 601 KIEKTRRLFV QKQLQLFDGM LLEQRQRLNE AIKLRDFVER DESGKLRRHP
 651 IELIRSASRR IQRSKLSNSQ TSIATESESG EPAVVVKKPI GDTTKGRSRW
 701 TDEQRVQLHL SLTRQRDIAS GGERRRRVVI KNDGEISGGT AGLHSQRKGS
 751 ESIGERAMMV TGDFALGGRV GVVQHHPLTH TIGAAHGTDL PLAVRLDNEE
 801 EQLRFLKDYQ NPIDNDIMRR TNTFEDTNGN LHPSFVLTST QRRGGSFDEA
 851 SNTIRTTTEA AGVIPSRHSA TLNGHNTSRM DSNGSLDEPT SKTQTVSKRG
 901 SPMPRLSSGQ VPEVIIHDPE GGDFRFAETR SQKWGIVRNI LHMFTEGYPR
 951 IITQYIREHR RMQRRFGLTP LNYVNKYEDD VLRKLRQRRV LQVKEPGAPS
1001 QTSRASGDEE LVEVNGNKVI LTDSDDIGDL SPIQMATNIV RNVPVTPFGA
1051 VMLVIVIVNG IFNATRYFQQ PEYWETALFV LGIIFTSFFV LEIVVRVIGL
1101 GLVSFLLDFN NLLDVTVTIL GFVELAYARS NVVTVLNWVR LLRLFRTLPF
1151 APMRRVSRVL LLGFADMLYA LFFFSIYMFM WILIGMSFFG GPNGMVDHTF
1201 QDYYTRGNFD TFSGASFAVS QAFSYTREEW VYLTWNGMQS RGEYTVLYFM
1251 AVVGVAFIAR YFFVAVFAWA WQSEEEEEEN YAAIAKGGSG GRREVTRLRW
1301 FDFTVWRSFK HIHGGFERRD VAPDEVFHLN EDMRKQLRIA EAKERFTKEA
1351 LAQTDLAMSQ RRMGSPMASP SATMGYNTDG YAPAAQVGTA PRYVNVGGQL
1401 QRHINPSVDF VDAQVPPLNA PISQFQAENA QLRFARRYST VPASVYTPVI
1451 QDDGIDRPSP SARSSPSDAG ERSGELGGES QQNGQEEGDG QYSPRGVSPN
1501 GTGGRATSTG LQRKSSVLGR FPRLGYDRAA GKKDGGSRSV SASAMQQTGD
1551 GNELRGDYVN EGEVNGGSMV YEHILYPGPR LRYKHVMRNQ YVRVFERCLD
1601 CNTYQQMPLR APPNVQQRTP EELHAEHCHM AAVRSSRQLV LNAIMGYVRL
1651 QKDINQPPTR DAVETVLGQA WSCGMLLFET IEYLSCSDIE QREYRTWDRT
1701 LEALQLQQWL IGLHVGEEQV GRATLAYTLA HRKREKLAVE HKSFELSWRQ
1751 RSFFFISPSN PVRRLSTRII QSRWFDIFIL TVIFIASFCL CFHIPGKAND
1801 PETGFVVLRA FDGIFTCIFL VEMIMKWISM GVILFRPEAY FWHWWNVFDF
1851 VIVIVSLIGL ADQHSALRSI KVLRCFRILI PMRVSNFNRS LSKISSALLD
1901 CLPTVANILL LFFINYFVWA VLAVRLLNGL THSCSDPSFV DITACEDAKH
1951 EWLPKVRNFD SFFQSLLTMI EVSVGSKWLD VIYTGVNGRT SEHAPMDDHY
2001 LARGFFFIVY YYVSHLILFS LFTASMIYSY LLTKNAAEGV LGITFEHQLW
2051 IRMQRMTLQL KPRVKLVPLC NHVSQFLHNV VIRPIFEVVG ASVLLLNILT
2101 MALHWYGETK SKASVLAAFQ YVWMFYFTVE AAMKIGAHGM RAFSRWAFSF
2151 DFFVLLLSFI GLIVDAASSE GMPFNVNVLR MLRLGRFFSA AKVFKPMRKQ
2201 FSLLHEVLIR SAVSLANVTL ILFLGVFVFT VLGLHLVGGV PVGEGGYFDD
2251 RYTNFNNFGN SLMMTFRLTT LENWSPSLRE GMNVTRKCTE DDCSVNYGSA
2301 FYCLLLLVFL GLIVLSFYMA VIVDHYVTAA RMNTSITRIE DLRRFRDLWS
2351 EFDPNGALVL HTHELPKLLE SLRPPLGLTS RHNRVELLRL LREYDIPNHR
2401 GKVHYHEVLL PLARRVLAMA FSRDTMDYRT TFDTLWRHSE KSLRALPTVL
2451 GKRSHATAAQ HFAASYVQAV CRRKKACREV QRVRSELWHE GRAVCDELGL
2501 PYADYGFGNL LLEGPDPMRD LVPRSASAAS SGGKGTRKGA SEAENAASSP
2551 LWRAGQAESP SSPRSEGETI DGRAPAARLP GAYQPAIEER EKRFGPDVPN
2601 ALRRHETRSE KLRRKDEERM LQSTPDDAVS SPVSNVRSNS QRVNVGEYQP
2651 PLGTDPTSWL GSNVNRGSTV GGPTTESRTS SVMPAPQGPT APE
```

Figure 4.9 – Amino acid sequence of a putative calcium channel (Tb10.70.4750) highlighting all peptides (in red) identified by mass spectrometry with 17% sequence coverage and a score of 1118 and all trans-membrane domains predicted by TMHMM v2.0 (boxed). The peptides ANDPETGFVVLR (see Figure 4.10), and IGAHGMRAFSR both contain regions predicted to form a trans-membrane domain.

A

#	b	b^{++}	b^{*}	b^{*++}	b^{0}	b^{0++}	Seq.	y	y^{++}	y^{*}	y^{*++}	y^{0}	y^{0++}	#
1	72.04	36.53					A							12
2	**186.09**	93.55	**169.06**	85.03			N	1246.64	623.82	1229.62	615.31	1228.63	614.82	11
3	**301.11**	151.06	**284.09**	142.55	283.10	142.06	D	1132.60	566.80	1115.57	558.29	1114.59	557.80	10
4	398.17	199.59	**381.14**	191.07	380.16	190.58	P	**1017.57**	**509.29**	1000.55	500.78	999.56	**500.28**	9
5	527.21	264.11	**510.18**	255.60	**509.20**	255.10	E	**920.52**	460.76	903.49	452.25	**902.51**	451.76	8
6	**628.26**	314.63	611.23	306.12	610.25	305.63	T	**791.48**	396.24	774.45	387.73	773.47	**387.24**	7
7	685.28	343.14	668.25	334.63	667.27	334.14	G	**690.43**	345.72	673.40	337.21			6
8	832.35	416.68	815.32	408.16	**814.34**	407.67	F	**633.41**	317.21	616.38	308.69			5
9	931.42	466.21	914.39	457.70	913.41	457.21	V	**486.34**	243.67	469.31	235.16			4
10	1030.48	515.75	1013.46	507.23	1012.47	506.74	V	**387.27**	194.14	370.24	185.63			3
11	1143.57	572.29	1126.54	563.77	1125.56	563.28	L	**288.20**	144.61	271.18	136.09			2
12							R	**175.12**	88.06	**158.09**	79.55			1

B



C

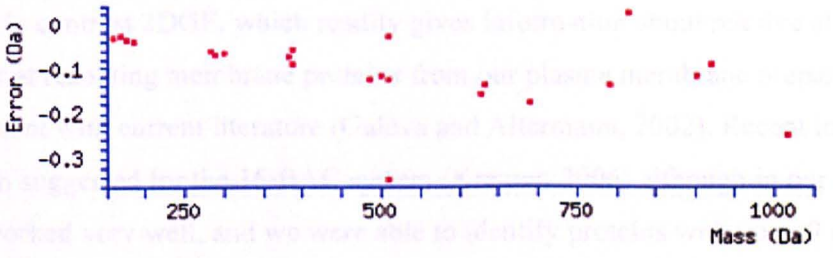


Figure 4.10 – MS/MS fragmentation data from which the peptide ANDPETGFVVLR was identified with a score of 69 from a putative calcium channel protein (Tb10.70.4750; see Figure 4.9). If completely fragmented 114 fragment ions would be generated consisting of a b and y ion series (A). 23 of these potential fragment ions were identified by MS (in red) from their relative ion intensities (B). The error of the calculated Mwt against the experimental Mwt for each of the fragment ions (C) shows very tight association (less then 0.4Da variation over the entire mass spectrum).

4.4 Discussion

Our aim in this chapter was to try and define the plasma membrane proteome of BSF *T. brucei*. Physical fractionation by shearing *T. brucei* cells and isolating the membrane sheets was key to isolating enriched plasma membrane protein samples for mass spectrometry analysis. However we were aware that these samples would contain a range of proteins non-specifically or indirectly associated with the membrane. To identify methods for further enriching the sample a system was devised, using a myc-tagged membrane protein, to characterise additional enrichment techniques. Unfortunately this system was unsuccessful and therefore further enrichment strategies were not pursued.

A diverse range of proteomic techniques were successfully used to define the proteome of the plasma membrane enriched fraction. While all techniques employed identified a large number of proteins, it very quickly became clear that some were superior, both in terms of the absolute number of proteins identified and their ability to identify membrane proteins. In this regard, 1DGE was definitively the most successful technique employed (Figure 4.8). However it is important to note that for maximum proteome coverage a number of complementary techniques needed to be employed. Furthermore, because IDGE does not fully resolve complex proteomes into their component proteins, it does not provide the information on relative protein abundance that is required for a comparative proteomic analysis. In contrast 2DGE, which readily gives information about relative abundance, was very poor at resolving membrane proteins from our plasma membrane preparations. This is in agreement with current literature (Galeva and Altermann, 2002). Recent improvements have been suggested for the 16-BAC system (Kramer, 2006) although in our hands the system worked very well, and we were able to identify proteins with up to 9 predicted TMD's. In comparison to 2DGE, 16-BAC gels have a far inferior resolving power. However by reducing sample complexity (i.e. using plasma membrane enriched rather than whole cell lysates), this was limited as much as possible. Indeed, considering 2DGE's inability to resolve polytopic membrane proteins, for an analysis focusing on polytopic proteins, the 16-BAC approach appears to be far superior.

In tandem with analysing the membrane preparations, we were also able to interrogate a cytoskeletal preparation, and combine these ID's with the previously published flagellar proteome (Broadhead *et al.*, 2006) to define a cytoskeleton / flagellum protein dataset (**TbCF**). This was used in a subtractive approach to identify all proteins that partitioned exclusively in the plasma membrane fraction. Unlike in the flagellar proteome, where an

isoelectric point cut-off was used to remove basic ribosomal protein contaminants (Broadhead *et al.*, 2006), we found very few ribosomal proteins with a pI of >10.2 (6). We then removed all proteins that were localised to non-membrane locations. In summary, 566 proteins were identified in our plasma membrane sub-proteome (**TbPM**) for long slender blood stream form *T. brucei*. While a large number of these proteins contained multiple transmembrane domains (54 with 4 or more TMD's), there were a surprising number of proteins containing 0 to 3 TMD's (503). Some, although not all, of these proteins are probably not true plasma membrane components. Therefore despite removing all potential contaminants through our subtractive approach, our definition of the plasma membrane proteome is likely to require further refinement. It is worth noting however that it is highly unlikely that non-plasma membrane polytopic integral membrane proteins will be present in the PM prep by virtue of the physical separation used to isolate the plasma membrane over other organelle membranes. Four polytopic proteins (5+ TMD's) do appear to be associated with the cytoskeleton (Table 4.5), as they were identified in both the plasma membrane and cytoskeleton preparations (although not in the TbFP). Unfortunately 3 of these proteins are annotated as hypothetical preventing much insight into their function to be made. Intriguingly, though, the fourth (Tb927.8.2380) encodes a putative ABC transporter. It is difficult to imagine a role for such a protein in the cytoskeleton considering its role in membrane transport, suggesting that it is in fact a contaminant of the cytoskeleton preparation. Nevertheless for this protein to have been retained in the cytoskeletal prep it must be intimately associated with some cytoskeleton proteins, perhaps as part of a scaffold.

Genes annotated as hypothetical were very difficult to assess in terms of their function, although hydropathy analysis does permit a prediction to be made about possible integral membrane location. A number of these proteins were investigated by BLAST analysis (http://www.ncbi.nlm.nih.gov/BLAST/). However very little insight into their function was gained and data was therefore not presented here. This lack of gene annotation makes an analytical assessment of the PM proteome coverage very difficult. Fortunately, in budding yeast a global analysis of protein localisation has been performed (Ghaemmaghami *et al.*, 2003; Huh *et al.*, 2003). This analysis indicates that the dominant protein localisations are in the nucleus and cytoplasm, with less than 7% of proteins localised to extracellular membranes i.e. lipid particles, cell periphery, bud or bud neck (Table 4.10). This compares favourably with the **TbPM** where ~6% of all *T. brucei* genes were identified (Table 4.7). Considering that trypanosomes are likely to have a more complex plasma membrane due to cell polarity etc, one might expect this to have been higher, however in terms of total numbers, significantly more proteins were identified in **TbPM** (566 in **TbPM** compared to

354 in the yeast PM). In addition, the yeast analysis is likely to be more comprehensive as each gene was analysed in isolation.

A large number of proteins annotated as 'putative' were unambiguously identified during this analysis. While all MS identifications are made on a statistical basis, for a number of these identifications, the score generated is so high as to essentially exclude the possibility that any other protein could have generated the same MS spectrum e.g. Tb10.70.4750. This knowledge, and in combination with other proteomic approaches, a complete proteomic map can be built up, which would be analogous to the published genome (Berriman *et al.*, 2005), but even more useful.

Approximately 13.3% of all proteins with 4 or more TMDs are identified in **TbPM**. This represents good proteome coverage considering cells have undergone subcellular fractionation and therefore almost all intracellular membranes have been removed. The trypanosome has a complex subcellular architecture that is defined by membrane compartments. The great majority of expressed polytopic membrane proteins will be localised to discrete intracellular membranes of which there are a number in *Trypanosoma* (Figure 4.12) such as the nucleus, mitochondrion, lysosome, glycosome, acidocalcisome, endoplasmic reticulum, Golgi apparatus etc. As indicated above, while soluble proteins might readily contaminate our plasma membrane preparations, it is less likely that integral membrane proteins from organelles might transfer to the sheets of cytoskeleton-associated plasma membrane fragments that we have characterised. While there is limited data on trypanosome membrane proteins, we have not identified known integral membrane proteins from other organelles in our plasma membrane preparations, although we have identified (and removed) a number of soluble factors localised to other cellular compartments. Furthermore, trypanosomes have a digenetic lifecycle, with many proteins known to be expressed in a single lifecycle stage only, or induced under particular environmental conditions e.g. in response to substrate depletion (Mussmann *et al.*, 2004). The current study exclusively used long slender blood-stream form parasites. Identifying a small fraction of the total number of proteins is therefore to be expected and suggests that **TbPM** represents good coverage of the long slender BSF *T. brucei* membrane sub-proteome.

In **TbPM** there is an exceptionally high representation of nucleobase / nucleoside transporters (Figure 4.11). This is perhaps not too surprising in view of the absolute requirement of trypanosomes to scavenge purines from their external environment (De Koning *et al.*, 2005). It is interesting, however, considering that TbAT1 is the only gene

definitively shown to be involved in pentamidine resistance (Matovu *et al.*, 2003), and suggests that a quantitative analysis on strains exhibiting a resistance transport phenotype (as presented in Chapter 3) would be applicable. In contrast to the nucleobase / nucleoside transporters, there is a relative paucity of amino acid transporters identified (one identification which was unable to differentiate between 3 genes). This is surprising considering there are more than 46 amino acid transporters present in the genome (C. Ebikeme, Personal communication). When cultured in HMI-9 media, which is rich in a variety of carbon sources, it might not be surprising to imagine the parasites only expressing a small subset of their potential nucleobase / nucleoside or amino acid transporter repertoire. However these samples were generated from parasites cultured *in vivo* (albeit from an *in vitro* generated innoculum), and the lack of amino acid transporter expression suggests that amino acids are not extensively utilised *in vivo*. This is in contrast to procyclic trypanosomes which can be grown exclusively on proline as their carbon source (Bringaud *et al.*, 2006).

In addition to the enrichment of polytopic membrane proteins in **TbPM**, there is a reduction in the average gCAI, although the maximum and minimum values as outliers are very constant (Table 4.6). The reduction in the average gCAI is statistically significant with a p value of <0.001 (using a Mann-Whitney one-tailed non-parametric test). It is perhaps surprising that the average gCAI does not vary more between the other datasets. This could be due to a sub-set of non-PM proteins skewing the data. Alternatively, it is perhaps naïve to assume that all PM proteins are expressed at low levels, considering many will be PM associated, but not integral to the membrane. In this case the proteins may mediate a variety of functions such as signalling for which high activity may need to be rapidly achieved. gCAI is very useful for monitoring entire datasets, however its application to individual proteins, in the absence of gene annotation is potentially suspect. This is particularly well illustrated, by the average gCAI of all proteins predicted to have 5 or more TMD's as being 81.3%, despite a large proportion of these genes expected to be present at relatively low copy numbers. Nevertheless it can be seen that low-copy / translationally unbiased proteins are being identified (Table 4.9) and that a significant proportion of **TbPM** are weakly expressed. Considering the proportions of the plasma membrane as a 2-dimensional sheet, and therefore the finite limit in the number of integral membrane proteins that can be inserted into the membrane, particularly when taking into account the hyper-abundance of GPI-anchors associated with the VSG coat, this suggests that **TbPM** contains a number of rare polytopic membrane proteins.

In summary, a working version of the *T. brucei* blood-stream form plasma membrane sub-proteome is presented here.

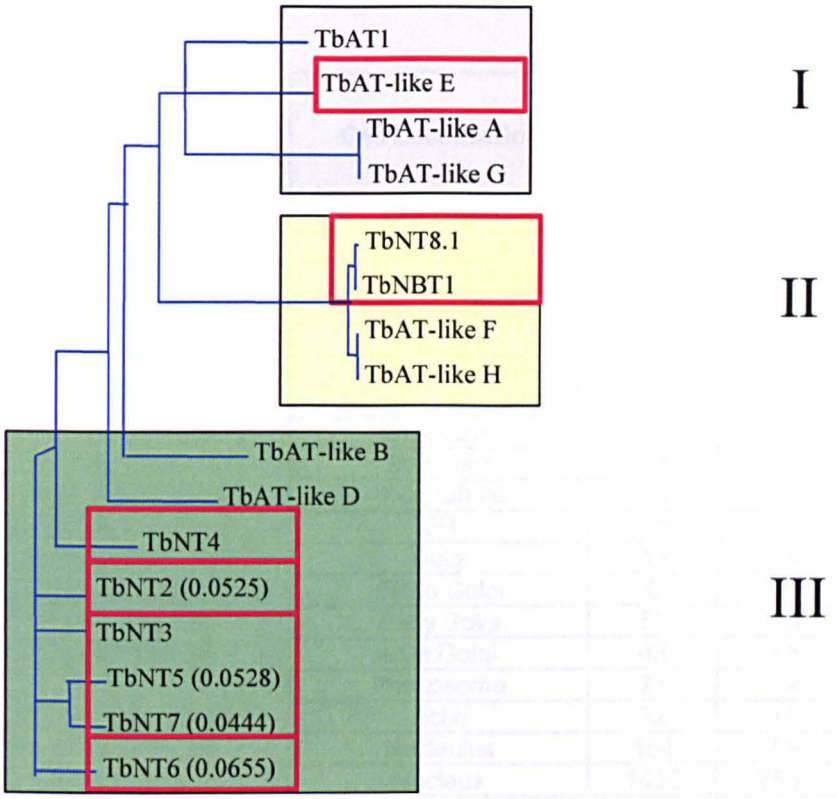


Figure 4.11 - Phylogenetic tree of *T. brucei* nucleobase and nucleoside transporters adapted from (De Koning *et al.*, 2005). Proteins identified by mass spectrometry are highlighted (□). Where data was unable to differentiate between closely related transporters, all potentially identified transporters are included in the same box.

Cell Localisation	Proteins localised	
	Number	%
Cytoplasm	1821	32.4
Spindle pole	66	1.2
Punctate composite	141	2.5
Mitochondrion	527	9.4
Vacuole	163	2.9
Vacuolar membrane	60	1.1
Nuclear periphery	61	1.1
Endosome	49	0.9
Bud neck	98	1.7
Microtubule	20	0.4
ER	296	5.3
Golgi	43	0.8
ER to Golgi	6	0.1
Early Golgi	55	1.0
Late Golgi	46	0.8
Peroxisome	21	0.4
Actin	32	0.6
Nucleolus	164	2.9
Nucleus	1455	25.9
Lipid particle	23	0.4
Cell periphery	160	2.8
Bud	73	1.3
Ambiguous	237	4.2
Total	**5380**	**95.8**

Table 4.10 - Global analysis of protein localisation in budding yeast. Extracted from http://yeastgfp.ucsf.edu/ (Ghaemmaghami *et al.*, 2003; Huh *et al.*, 2003).

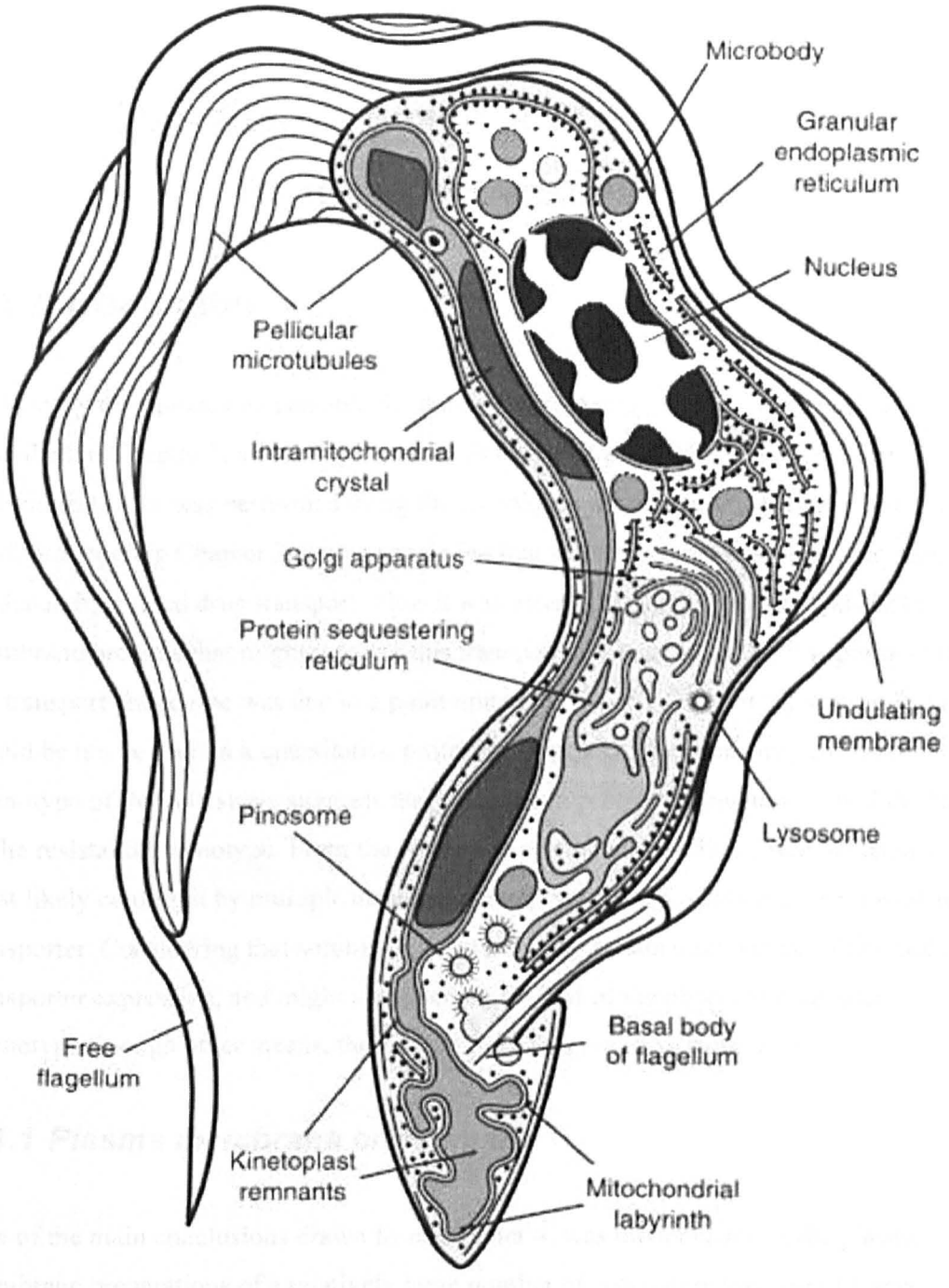


Figure 4.12 – Fine structure of the Trypanosome cell, generated from transmission electron microscopy (Taken from http://www.gsbs.utmb.edu/microbook/ch077.htm (Vickerman and Tetley, 1977))

Chapter 5

Quantitative analysis of the proteome of Pentamidine resistant lines

5.1 Introduction

To identify the proteins responsible for the transport mediated drug resistance phenotype described in Chapter 3, a quantitative analysis of the parental TbAT1 KO and drug resistant B48 line was performed using the techniques developed in Chapter 4. From the work presented in Chapter 3, it was concluded that the phenotype was, at least in part, mediated by altered drug transport. Thus it was essential to include an analysis of the membrane proteins that might mediate this transport component. While it is possible that the transport phenotype was due to a point mutation(s) in the relevant transporter (which would be not be seen in a quantitative proteomic approach) the muti-drug resistant phenotype of the B48 strain suggests that a change in protein expression formed the basis of the resistance phenotype. From the resistance acquisition profile of B48, resistance is most likely conferred by multiple mutations / modifications in addition to the postulated transporter. Considering that soluble proteins might modulate transport activities and/or transporter expression, and might also account for part of the observed resistance phenotype through other means, the soluble proteome was also examined.

5.1.1 Plasma membrane enrichment

One of the main conclusions drawn from Chapter 4, was the inclusion in the plasma membrane preparations of a relatively large number of 'contaminating' non-plasma membrane proteins derived from the cytoskeleton, which is intimately associated with the plasma membrane in these preparations. While robust quantitative techniques have been developed (see below), it is always advisable to reduce sample complexity wherever this is possible without compromising representation of genuine membrane proteins.

We therefore pursued a number of techniques to enrich for 'true' plasma membrane proteins, including Triton X-114 solubilisation, which selectively partitions hydrophilic and hydrophobic proteins into an aqueous and detergent phase respectively (Bordier, 1981; Brusca and Radolf, 1994). Other partitioning systems such as the organic solvent

separations using chloroform / methanol have also been used and in some cases have been shown to be more effective (Zhang *et al.*, 2005). This particular system has even been applied to the enrichment of plasma membrane proteins from crude whole cell extracts (Ferro *et al.*, 2000; Ferro *et al.*, 2002). Despite this techniques success, during the procedure, hydrophobic proteins are precipitated out of solution, the re-suspension of which probably leads to the preferential solubilisation of proteins with lower numbers of TMD's, and potentially a reduction in membrane proteome coverage. Other enrichment strategies include interfering with the non-specific protein-protein / protein-lipid interactions on the membrane sheet. By altering pH e.g. by incubating with sodium carbonate (Fujiki *et al.*, 1982), charge repulsion can be used to disrupt non-specific binding. A final approach that was employed for plasma membrane protein enrichment involved incubating the sample in a calcium / calmodulin buffer (P. Voorheis, Personal Communication). Calmodulin is a highly promiscuous protein that binds to an array of target proteins, predominantly via a hydrophobic patch (Rhoads and Friedberg, 1997; Yamniuk and Vogel, 2004). By incubating the plasma membrane samples with calmodulin it may be possible to disrupt various interactions thereby releasing soluble proteins from the membrane.

5.1.2 Quantitative techniques

Mass spectrometry itself is essentially a quantitative technique, whereby MS peak intensities can be used to measure protein abundance to within a factor of four, if three or more peptides are included in the analysis (Steen and Mann, 2004). However, in biological systems, this level of resolution will fail to identify subtle proteomic changes and as a result a number of enhanced quantitative mass spectrometry-based techniques have been developed, each of which have often been refined or developed for novel applications. For example, protein mass tagging has been developed to target a number of residues e.g. cysteine, and lysine. The frequency of these residues in protein samples is very different; therefore by selecting an appropriate target residue, biological questions can be more readily answered.

There are essentially three quantitative techniques available and these are summarised in Figure 5.1 (although there generally a large number of variations on each, for review see (Beynon and Pratt, 2005; Ong *et al.*, 2003a; Yan and Chen, 2005)). Unfortunately it was not possible to implement some of these, due to financial costs e.g. SILAC (Figure 5.1A) or experimental incompatibilities. For example iTRAQ (Figure 5.1B) labels peptides rather than proteins limiting the subsequent analysis to MuDPIT based experiments. Our *T.*

brucei PM-enriched samples analysed by MuDPIT did not give acceptable proteome coverage (see Chapter 4), although other groups working with different samples have demonstrated how effective this technique can be (Hu *et al.*, 2006). We therefore decided to focus on two quantitative techniques – differential in-gel electrophoresis (DiGE) and isotope-coded affinity tag (I-CAT), both of which are post-isolation quantitation strategies (Figure 5.1B).

5.1.2.1 DiGE

DiGE compares two samples by tagging each sample with different charge and mass matched fluorophores (Cy3 and Cy5 CyDyes). Both samples can then be run on the same gel and visualised by virtue of their different excitation / emission spectra (Ünlü *et al.*, 1997). To increase confidence in changes in expression, multiple samples are run on replicate gels. However as with standard 2DGE, spot migration patterns can vary dramatically from gel-to-gel. To enable easy comparisons of the same spot in multiple gels, a standard sample, tagged with another fluorophore (Cy2), is added to every gel in a multiplexed experiment. The standard sample is composed of a pool of every sample in the analysis, and thus enables normalisation of individual spot volumes on each gel. The spot pattern from the pooled standard can be used to match multiple gel images, so that artifactual changes that are due to technical variability can be excluded from biological changes and spot modulations can be quantified relative to a similar standard that is present on every gel (for an overview of the technique see (Wu, 2006)). Being a gel-based separation technique, there is the added benefit that different species of the same protein arising from post-translational modifications are likely to be resolved on the gel due to a change in mass / pI.

5.1.2.2 I-CAT

Isotope-coded affinity tagging (I-CAT) was the first truly quantitative proteomic MS method developed (Gygi *et al.*, 1999a), and as a result has probably been used more extensively than any other. The technology works by the covalent attachment of a reactive isotopically 'heavy' or 'light' mass tag to free cysteine residues, and is therefore amenable to labelling peptides or whole proteins. Recent developments in this approach include the incorporation of a cleavable affinity moiety for the purification of tagged peptides, and the development of a solid-phase isotope tagging system (Zhou *et al.*, 2002). The main issue when using I-CAT relates to the distribution of cysteine residues in the sample. Considering that only cysteine-containing I-CAT labelled peptides provide the

quantitation, proteins containing no cysteine residues cannot be analysed by this approach. Furthermore, one can either purify the labelled peptides or analyse the labelled and unlabelled peptide mixture. In reality the complexity of analysing all labelled and unlabelled peptides often reduces quantitative data produced, and coverage can be improved by purifying and analysing the labelled peptides alone. However, by purifying only labelled peptides, the likelihood of identifying enough peptides for confident protein assignments is reduced (unless all peptides from a protein contain a cysteine residue). To ensure that the I-CAT procedure would be applicable to the *T. brucei* proteome, the number of proteins devoid of cysteine residues was determined using a bioinformatics approach.

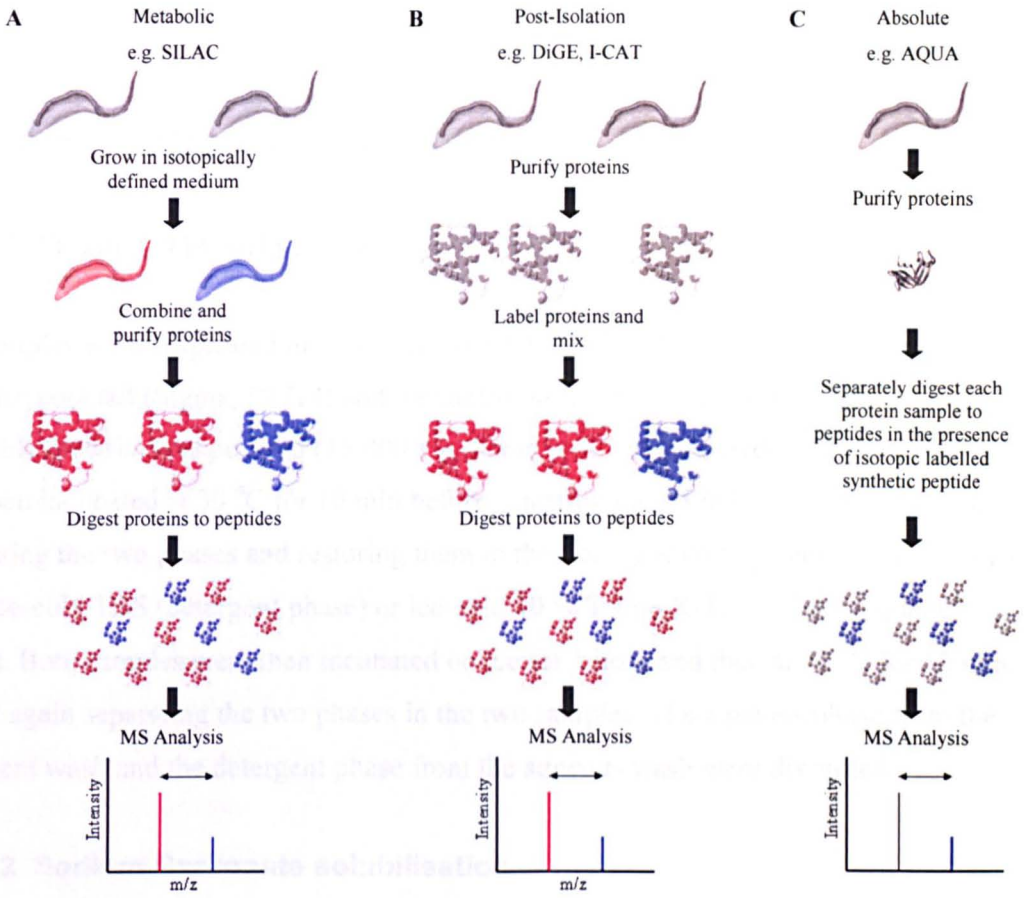


Figure 5.1 - Quantitative mass spectrometry approaches currently described in the literature. Labelled whole cells, proteins and peptides are designated in red (light mass tag) or blue (heavy mass tag).

5.2 Methods

5.2.1 Plasma Membrane enrichment

5.2.1.1 Triton X-114 solubilisation

PM samples were suspended in 2 ml 1 % Triton X-114 in TBS containing a protease inhibitor cocktail (Sigma, P2714) and incubated on ice for 1 hour with gentle stirring. Insoluble material was pelleted (16,000 x *g*, 5 min, 4 °C) and discarded. The supernatant was then incubated at 30 °C for 10 min before centrifuging (16,000 x *g*, 5 min, 30 °C), separating the two phases and restoring them to their original homogenate volume (2 ml) with ice-cold TBS (detergent phase) or ice-cold 10 % Triton X-114 in TBS (aqueous phase). Both samples were then incubated on ice for 1 hour and then at 30 °C for 10 min, before again separating the two phases in the two samples. The aqueous phase from the detergent wash and the detergent phase from the aqueous wash were discarded.

5.2.1.2 Sodium Carbonate solubilisation

PM samples were suspended in 300 µl of 100 mM sodium carbonate (Na_2CO_3) pH 11.5, containing a protease inhibitor cocktail (Sigma, P2714) and stirred on ice for 90 min, before centrifuging (16,000 x *g*, 5 min, 4°C) and removing the supernatant. Any protein in the supernatant was precipitated out of solution by the addition of 4 volumes of ice-cold acetone, incubation for 1 hour at –20 °C, and centrifugation (10 min, 16,000 × *g*, 4 °C). The resultant pellet was then washed with 80 % acetone in distilled water, before air-drying and re-suspending in sample running buffer.

5.2.1.3 Chloroform / Methanol partitioning

PM samples were diluted 1:9 with ice-cold chloroform / methanol (5:4), incubated on ice for 15 min, and centrifuged (12,000 × *g*, 20 min, 4 °C). The pellet (composed of the plasma membrane sheets) was retained, while the supernatant (composed of soluble proteins) was completely dried down in a SpeedVac (Eppendorf Concentrator 5301).

5.2.1.4 Calcium / Calmodulin solubilisation

PM samples were suspended in calcium wash buffer (30 mM TES, 140 mM NaCl, 4 mM KCl, 1 mM calcium acetate) containing 250 units of calmodulin (Sigma, UK) for 30 min at

30 °C, and then centrifuged (16,000 x *g*, 5 min, at room temperature). The supernatant (containing soluble proteins) was retained, while the pellet was washed twice in calcium wash buffer.

5.2.2 DiGE Analyses

5.2.2.1 2DGE sample preparation

Four 150 ml cultures of *T. brucei* TbAT1 KO and B48 cell lines were grown under identical conditions to a cell concentration of ~2 x 10^6 cells / ml and harvested by centrifugation (600 × *g*, 10 min, at room temperature). Cells were then washed twice in ice-cold PBS before being re-suspended in DiGE Lysis buffer (7 M urea, 2 M thiourea, 4 % CHAPS, 25 mM Tris- HCl [pH 8.5]) supplemented with Protease Inhibitor Cocktail (Sigma, P2714). To completely lyse cells, samples were rapidly freeze-thawed by cycling 5 times between liquid nitrogen and a heat block set to 25 °C. Insoluble material was removed by centrifugation (16,000 x *g*, 5 min, at room temperature). Protein was then precipitated out of solution by the addition of 4 volumes of ice-cold acetone, incubation for 1 hour at –20 °C, and centrifugation (10 min, 14000 × *g*). The resultant pellet was then washed with 80 % acetone, before air-drying and re-suspending in 20 µl of DiGE Lysis Buffer. Protein concentration was determined using the 2D Quant Kit (Amersham Biosciences) as per manufacturers instructions. All protein concentrations were adjusted to whichever was lower of either the lowest replicates protein concentration or to 5 mg/ml. Finally all samples were confirmed to be at pH 8-9 using broad range pH indicator paper (Whatman plc, UK) in preparation for CyDye labelling.

5.2.2.2 PM sample preparation

Replicate PM preparations (see Chapter 4 for method) of TbAT1 KO and B48 cell lines were pelleted and re-suspended in DIGE lysis buffer, at 5 mg/ml. Finally all samples were confirmed to be at pH 8-9 in preparation for CyDye labelling.

5.2.2.3 CyDye labelling

For each DiGE experiment a pooled standard (composed of all experimental samples), sufficient to run 50 µg per gel, along with 50 µg aliquots for each sample was generated. Samples were incubated on ice for 10 min before 400 pmol of CyDye was added to each 50 µg aliquot and mixed immediately. Aliquots were then incubated on ice for 30 min in

the dark. To stop the reaction 1 µl of 10 mM Lysine was added and incubated on ice for a further 10 min in the dark.

For the soluble proteome DiGE analysis using whole cell extracts, four gels were run. Each gel contained one TbAT1 KO and one B48, Cy3 / Cy5 labelled sample, as well as the Cy2 labelled pooled standard (50 µg each). A fifth preparative gel containing 500 µg of unlabelled pooled protein as well as 50 µg Cy2 labelled pooled standard was run for picking spots to identify proteins.

For the 16-BAC DiGE analysis, only two PM enriched samples from the TbAT1 KO and B48 lines were able to be prepared. Two comparative gels were therefore run as well as a third preparative gel (as with the soluble proteome DiGE approach).

5.2.2.4 16-BAC and 2DGE

16-BAC or pH 3-10 NL 2DGE gels were run as described in chapter 4.

5.2.2.5 Protein Visualisation

Each gel was scanned at the three excitation / emission wavelengths relevant to Cy2 ($\lambda_{excitation}$= 491 nm, $\lambda_{emission}$= 509 nm) Cy3 ($\lambda_{excitation}$= 553 nm, $\lambda_{emission}$= 569 nm) and Cy5 ($\lambda_{excitation}$= 645 nm, $\lambda_{emission}$= 664 nm) on a Typhoon™ 9400 variable mode imager (GE Healthcare, UK). Preparative gels were stained with Sypro Orange (diluted 1:10,000) after fixing (7 % acetic acid, 10 % methanol, 2 hours) and then scanned ($\lambda_{excitation}$= 553 nm, $\lambda_{emission}$= 569 nm).

5.2.2.6 Gel image analysis

DeCyder™ software v5.01 (Amersham Biosciences) was used for automated gel image analysis. Using the Differential In-gel Analysis (DIA) software module, spots were identified and spot intensity quantified separately for each CyDye gel image. Optimally, 2000 spots were scanned for on the 2DGE images and 1000 on the 16-BAC. Once defined, multiple gel-to-gel comparisons were performed using the Biological Variation Analysis (BVA) module, using a Cy2 image as the master gel to which all others were matched. From this analysis, a statistical analysis was applied to identify all up or down regulated spots. These spots were matched to the preparative gel were identified for MS protein

identification, and a pick list was generated. Minimal user intervention, in the form of matching gel images, was maintained.

5.2.2.7 Digestion and MS analysis

Using the picklist generated by the BVA module, the Ettan Spot Handling Workstation (Amersham Biosciences) picked the relevant spots and loaded the gel plugs into 96-well plates. Samples were then digested with trypsin as follows. Samples were washed five times (3 x 100 µl 50 mM AmBic, 50 % MeOH; 2 x 100 µl 75 % ACN), before being dried for 60 min at 37 °C. Samples were then re-hydrated by the addition of 200 ng trypsin (in 10 µl 20 mM AmBic), and incubated for 240 min at 37 °C. Peptides were extracted from the gel plugs, by washing twice (2 x 100 µl 50 % ACN / 0.1 % TFA) and transferred in solution to a fresh 96 well plate, where samples were dried down until ready for MS analysis.

Samples were initially prepared for MALDI analysis by spotting onto target plates using the Ettan Spot Handling Workstation (Amersham Biosciences). Dried samples were re-suspended in 3 µl 50 % ACN, 0.5 % TFA and 0.3 µl spotted onto the target plate. Immediately, 0.3 µl of the matrix (10 mg/ml CHCA in 50 % ACN, 0.5 % TFA) was then added to the sample on the target and mixed. Samples were analysed using the 4700 proteomics analyser (Applied Biosystems), where approximately 3,000 laser shots were fired for peaks selected for MS/MS analysis.

Where no protein was identified, samples were analysed by MS using electrospray ionisation (on a QSTAR® XL Hybrid LC/MS/MS System). Samples were separated in one dimension on an LC system (Famos / Switchos / Ultimate, LC Packings), using a Pepmap C18 reverse phase column (LC Packings). A 5 – 85 % acetonitrile gradient (in 0.5 % Formic acid) was used to elute the peptides over a period of 45 min. Mass spectrometry analysis was performed using a duty cycle consisting of a 3 second survey MS scan, followed by four MS/MS analyses of the most abundant peptides (3 second per peak, with dynamic peak exclusion).

Protein identifications were made using the MASCOT® search engine described in Chapter 4.

5.2.3 I-CAT

5.2.3.1 Cysteine containing proteins

The *T. brucei* genome (version 4 downloaded from ftp://ftp.sanger.ac.uk/pub/databases/T.brucei_sequences/ T.brucei_genome_v4/) was digested *in silico* using the 'Proteogest' Perl script (Cagney *et al.*, 2003), downloaded from (http://www.utoronto.ca/emililab/proteogest.htm). Cysteine-containing peptides were extracted using self-written Perl scripts and tabulated in Microsoft Excel.

5.2.3.2 Labelling

PM samples were initially washed in 50 mM Tris-HCl pH 8.5 at 4 °C to remove all β-mercaptoethanol, before re-suspending in 80 µl denaturing buffer (50 mM Tris-HCl pH 8.5, 8 M Urea, 2 % Triton X-100, 0.1 % SDS). Proteins were reduced by the addition of 2 µl TCEP and incubating at 20 °C for 30 min. 1 unit each of light I-CAT reagent and heavy I-CAT reagent (Applied Biosystems, UK) were allowed to reach room temperature before being re-suspended in 20 µl acetonitrile. The TbAT1 KO PM sample was then added to the light reagent vial, and the B48 PM sample to the heavy reagent vial. Both tubes were vortexed before incubating in the dark for 2 hours at 37 °C, after which, the tube contents were combined.

5.2.3.3 Gel separation

Samples were run on gels prepared in-house. These gels consisted of an 8 % acrylamide separating gel and a 5 % stacking gel (both containing 0.4 M Tris, 0.1 % SDS [pH 8.8]), polymerised using APS / TEMED. Samples were solubilised in sample buffer (85 mM Tris-HCl pH 6.8, 30 % Glycerol, 2.7 % SDS, 0.01 % Bromophenol blue, and 0.35 M dithiothreitol). Samples were then incubated at 60 °C for 10 min prior to loading onto the gel and then run until the dye front reached the end of the gel. Gels were then fixed for 1 hour in 35 % isopropanol, 10 % acetic acid (change solution every 15min), before being stained for 30 min (0.15 % Brilliant blue R-250, 7.5 % Acetic acid, 50 % MeOH) to visualise protein bands. Gels were then destained (1 x 15min) in wash solution I (50 % EtOH, 7.5 % Acetic acid, 5 % Glycerol), before being further destained (3 x 20 min) in wash solution II (5 % EtOH, 7.5 % Acetic acid, 5 % Glycerol). Each lane was then sectioned into a series of 2 mm slices, each of which was divided into 3 squares.

5.2.3.4 Protein digestion and linker cleavage

I-CAT samples underwent a standard in-gel trypsin digest and subsequent peptide extraction as described in Chapter 4. Extracted peptides were then re-suspended in 200 µl PBS [pH 7.5] and transferred to a 96-well SigmaScreen™ high capacity streptavidin coated plate (Sigma, UK). Samples were incubated on the plate at room temperature for 2 h. Supernatant was removed and dried down in a SpeedVac (Eppendorf Concentrator 5301). To remove excess unbound peptides, wells were washed 3 x with 0.05 % Tween in PBS. Peptides were then eluted from the plate with 70 % ACN, 5 % Formic acid, 1 mM Biotin, transferred to a fresh 96-well plate and dried down in a SpeedVac (Eppendorf Concentrator 5301). The acid-labile linker was then cleaved by re-suspending samples in 5 % TFA and incubating for 4 hours at 37 °C. Finally peptides were dried down in a SpeedVac (Eppendorf Concentrator 5301) and stored at –20 °C.

5.2.3.5 Mass Spectrometry

I-CAT samples were analysed by MS using electrospray ionisation (on a QSTAR® XL Hybrid LC/MS/MS System). Samples were separated in one dimension on an LC system (Famos / Switchos / Ultimate, LC Packings), using a Pepmap C18 reverse phase column (LC Packings). A 5 – 85 % acetonitrile gradient (in 0.5 % Formic acid) was used to elute the peptides over a period of 90 min. Mass spectrometry analysis was performed using a duty cycle consisting of a 3 second survey MS scan, followed by four MS/MS analyses (3 second per peak) of the most abundant peptides, subject to dynamic exclusion.

5.2.3.6 Data analysis

Mass spectrometry data generated from the labelled and unlabelled I-CAT fractions was analysed using the MASCOT® search engine described in Chapter 4 (albeit with I-CAT heavy and light mass tags as fixed modifications for the labelled fraction). Additionally, the labelled I-CAT fraction was quantitatively analysed using Pro I-CAT software v1.1 (Applied Biosytems, UK) and results viewed using ProGroups Viewer v 1.0.5 (Applied Biosytems, UK). Scores generated from the Pro I-CAT software differ from the standard MASCOT® algorithm, and are summarised in Table 5.1.

Confidence		Protein Score threshold
p > 0.01	99%	2.0
p > 0.05	95%	1.3
p > 0.10	90%	1.0
p > 0.33	66%	0.47

Table 5.1- Significant threshold scores for identifications made from the Pro I-CAT software at various confidence levels for tandem mass spectrometry data.

5.3 Results

5.3.1 Enrichment strategies

As stated in Chapter 4, the PM preps contained a number of proteins also identified in the cytoskeletal / flagellar proteome, and are considered to be protein contaminants of the PM preparations. A number of experiments were therefore performed that aimed to 'strip' the membrane of any proteins not intimately associated with or inserted into the plasma membrane. To enable rapid comparisons to be made between different enrichment techniques, 1DGE-banding patterns were compared and where necessary bands were excised and analysed.

5.3.1.1 Triton X-114 solubilisation

One of the major difficulties working with triton-solubilised proteins was its blurring effect on gel bands (see Figure 5.2). Triton's low critical micelle concentration made removal of any excess detergent very difficult (Furth, 1980; Jones, 1999). Attempts using commercial detergent removal kits e.g. Extracti-Gel D Detergent Removing Gel (Pierce, USA), failed to remove sufficient detergent to restore band resolution. Nevertheless 6 bands that displayed differential fractionation in the two phases were excised and analysed (see Figure 5.2), and the proteins identified (Table 5.2). By comparing proteins identified in bands 1-3 against bands 4-6, we can see that there are more polytopic membrane proteins in the latter than the former (2 versus 0). This suggests that hydrophobic proteins are preferentially partitioning in the Triton X-114 phase. However the band blurring is likely to be significantly reducing our ability to identify proteins as it effectively reduces the protein concentration of any one species, reducing the chance of identifying poorly expressed proteins. Comparable bands were not picked from the two lanes for comparison, making it difficult to identify the extent to which partitioning has occurred, although the inherent variability in MS would still only be semi-quantifiable.

5.3.1.2 Sodium Carbonate solubilisation

Washing the membranes with sodium carbonate appears to remove a band around 75 kDa (Figure 5.3). Bands were therefore excised from the gel and analysed by MS (Table 5.3). From this analysis, the most likely identity of the 75 kDa band is paraflagellar rod protein. It would therefore appear that incubating the membrane with the sodium carbonate buffer does remove some hydrophilic proteins, however it seems to be very selective in the

protein species that are removed. In addition, the vast majority of protein is retained in the PM pellet with only a fraction being removed by the enrichment. This effect does seem fairly specific, whereas one would expect the carbonate treatment to have a global effect. It is therefore possible that the proteins lost from the 75 kDa band are merely being diluted out of the sample.

5.3.1.3 Chloroform : Methanol solubilisation

After solubilising the plasma membranes in the chloroform/methanol mix, both the supernatant and the pellet were run on a 1D SDS-PAGE gel. However the supernatant lane did not contain any visible protein bands (data not shown). In addition there was no visible difference in banding pattern or intensity between the untreated membranes and the washed membranes.

5.3.1.4 Calcium / Calmodulin solubilisation

Washing the PM samples with calcium / calmodulin had a fairly large effect on protein banding in a 1D gel (Figure 5.4). In addition, washing in this way removed significantly more protein from the PM samples than the other techniques. To try and identify the proteins that were retained after washing, 9 bands were picked from the gel and analysed by ESI-MS/MS. Unfortunately these bands gave very poor MS data, and only a few identifications, none of which were integral membrane proteins (data not shown), were possible.

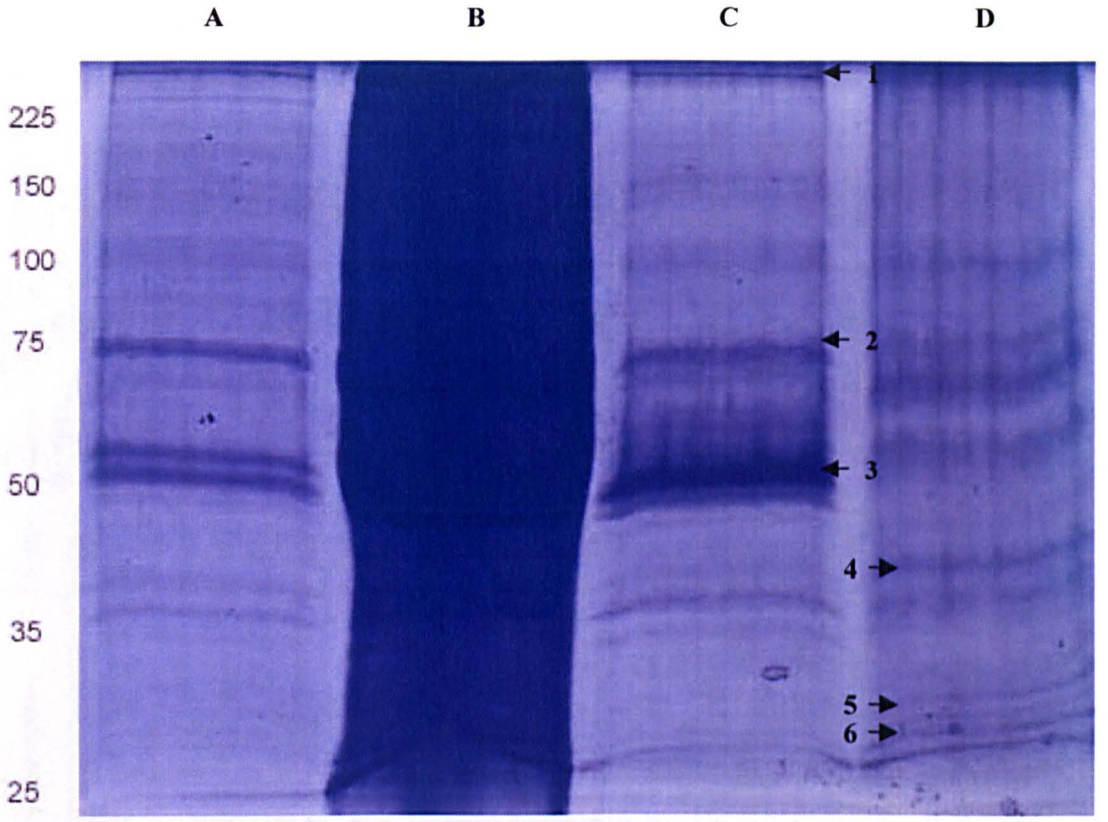


Figure 5.2 - Analysis of Triton X-114 treatment of *T. brucei* plasma membrane preparations run on a 10% 1DGE. Lanes consist of untreated plasma membrane samples (A), proteins insoluble in 1% Triton X-114 (B), proteins solubilised in the hydrophilic aqueous phase (C) and proteins soluble in the hydrophobic detergent phase (D). Protein bands that partitioned differentially in the detergent and aqueous phases were selected for analysis by mass spectrometry (highlighted with an arrow). Molecular weight (kDa) markers (Mwt) are shown on the left hand side.

Band Number	Accession Number	Gene annotation	Score	No. peptides identified	Sequence Coverage (%)	gCAI value (% rank)	TmPred (TMHMM)
2	Tb11.02.5450, Tb11.02.5500	Glucose regulated protein	795	20	35	98.7	0
2	Tb10.6k15.2290	BS2 protein disulfide isomerase	318	10	29	95.5	0
2, 3	Tb927.3.4290, Tb927.3.4300, Tb927.3.4310, Tb927.3.4320, Tb927.3.4330 / Tb927.8.4970, Tb927.8.4980, Tb927.8.4990, Tb927.8.5000, Tb927.8.5010	PFR-1 / PFR-2	146	5	11	99.4	0
2	Tb927.7.2650	Hypothetical	119	5	5	40.3	0
2	Tb10.70.4200	Fatty acyl CoA synthetase	58	2	1	76.3	0
2,3	Tb927.6.3740, Tb927.6.3750, Tb927.6.3800	HSP 70	54	1	3	97.6	0
3	Tb927.1.2340, Tb927.1.2360, Tb927.1.2380, Tb927.1.2400	Alpha tubulin	377	9	32	99.9	0
3	Tb09.211.3550, Tb09.211.3560, Tb09.211.3590, Tb09.211.3540	Glycerol Kinase	334	11	17	98.2	0
3	Tb927.5.4980	VSG	278	13	14	45.6	1
3	Tb927.4.5010, Tb927.8.7410	Calreticulin	150	4	14	97.6	1
3	Tb927.3.3270	PFK	81	4	11	99.3	0
3	Tb11.02.5450, Tb11.02.5500	Glucose regulated protein	80	2	4	98.7	0
3	Tb927.5.1780	Hypothetical	72	2	5	95	0
3	Tb927.6.5070	Hypothetical	63	1	1	47.7	0
3	Tb10.70.5650, Tb10.70.5670	Elongation factor	57	2	4	99.9	0
3	Tb927.8.6660	Hypothetical	56	1	2	96.8	0
3	Tb927.5.2850	Hypothetical	44	2	4	91.9	0
3	Tb10.6k15.2290	BS2 protein disulfide isomerase	39	2	3	95.5	0
4	Tb10.61.0540	Hypothetical	208	4	19	94.7	0
4	Tb09.211.0680	CAAX prenyl protease	96	3	9	94.2	5
4	Tb10.70.1370	Aldolase	79	3	7	99.1	0
4	Tb927.8.5440, Tb927.8.5460, Tb927.8.5470	Calcium binding protein	45	1	6	96.8	0
4	Tb927.2.6150	TbNT2	42	1	1	85.8	11
5	Tb10.6k15.3640	Alternative oxidase	98	3	11	95.9	1
5	Tb927.5.1210	Short-chain dehydrogenase	92	3	15	95.6	2
5	Tb09.211.1750	Mito carrier protein	73	2	4	96.8	1
6	Tb927.8.5440, Tb927.8.5460, Tb927.8.5470	Calcium binding protein	136	3	12	96.8	0
6	Tb11.02.0010	Hypothetical	50	1	4	67.2	0

Table 5.2 - Table of all proteins identified from protein bands that partitioned differentially in the Triton X-114 aqueous and detergent phases (see Figure 5.2). Score, MOWSE score defined by MASCOT® at p = 0.05; gCAI, genomic codon adaptation index (see Chapter 2); TmPred, trans-membrane prediction defined by TMHMM algorithm (see Chapter 2).

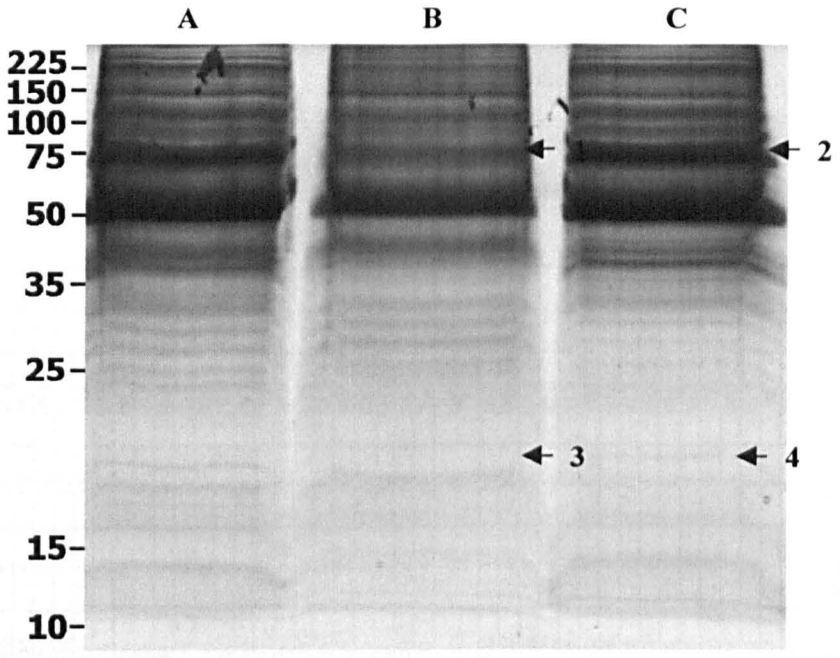


Figure 5.3 - Analysis of sodium carbonate treatment of *T. brucei* plasma membrane preparations run on a 12% 1DGE. Lanes consist of untreated plasma membrane samples (A), and proteins that were insoluble (B) or soluble (C) in aqueous sodium carbonate. Protein bands that displayed differential solubility were selected for analysis by mass spectrometry (highlighted with an arrow). Each lane was loaded with ~35μg of protein. Molecular weight (kDa) markers (Mwt) are shown on the left hand side.

Band Number	Accession Number	Gene annotation	Score	No. peptides identified	Sequence Coverage (%)	gCAI value (% rank)	TmPred (TMHMM)
1	Tb11.02.5450, Tb11.02.5500	Glucose regulated protein	771	24	39	98.7	0
1	Tb927.3.4290, Tb927.3.4300, Tb927.3.4310, Tb927.3.4320, Tb927.3.4330	PFR	307	8	19	99.3	0
1	Tb10.70.4200, Tb09.160.2810	Fatty acyl coA synthetase	240	11	17	76.3	0
1	Tb927.1.2340, Tb927.1.2360, Tb927.1.2380, Tb927.1.2400	Alpha tubulin	205	6	19	99.9	0
1	Tb11.01.2530	Kinesin-like protein	135	7	10	84.1	0
1	Tb11.01.3110	HSP 70	115	6	9	99.3	0
1	Tb11.02.5280	Glycerol 3 phosphate dehydrogenase	114	6	13	95.6	0
1	Tb927.8.4970, Tb927.8.4980, Tb927.8.4990, Tb927.8.5000, Tb927.8.5010	PFR-2	73	6	11	99.4	0
1	Tb927.6.3740, Tb927.6.3750, Tb927.6.3800	HSP 70	72	2	5	97.6	0
1	Tb927.5.890	oligosaccharyl transfer unit	69	4	8	85.1	10
1	Tb10.61.3130	Hypothetical	63	2	5	91.9	0
1	Tb927.8.6660	Hypothetical	61	1	2	96.8	0
1	Tb11.01.3290	Hypothetical	59	4	8	90.7	0
1	Tb11.01.0480	Hypothetical	56	2	6	90.6	9
1	Tb10.6k15.1690	Hypothetical	56	3	4	78.7	7
1	Tb927.7.2650	Hypothetical	53	2	6	40.3	0
1	Tb11.01.1680	Polyubiquitin	51	1	2	98	0
1	Tb927.4.1920	GPI anchor	50	1	2	78.2	2
1	Tb927.7.1400	Hypothetical	46	3	7	33.5	2
1	Tb11.03.0030	ABC transporter	41	1	2	54.7	3
1	Tb11.01.1650	SRP unit	38	2	4	68.7	0
3	Tb927.3.4290, Tb927.3.4300, Tb927.3.4310, Tb927.3.4320, Tb927.3.4330	PFR-1	96	5	5	99.3	0
3	Tb927.6.3840	Reticulon domain protein	80	2	12	95.1	3
3	Tb927.7.3410	Centrin	39	2	14	92.8	0
3	Tb09.244.1980	Leucine rich repeat protein	38	4	2	5.4	0
2	Tb927.3.4290, Tb927.3.4300, Tb927.3.4310, Tb927.3.4320, Tb927.3.4330 / Tb927.8.4970, Tb927.8.4980, Tb927.8.4990, Tb927.8.5000, Tb927.8.5010	PFR-1 / PFR-2	1096	42	43	99.3	0
2	Tb11.02.5450, Tb11.02.5500	Glucose regulated protein	332	9	18	98.7	0
2	Tb05.5K5.130, Tb927.5.4480	PFR Par4	134	6	9	56.3	0
2	Tb927.7.2650	Hypothetical	126	3	7	40.3	0
2	Tb927.6.3740, Tb927.6.3750, Tb927.6.3800	HSP 70	107	3	7	97.6	0
2	Tb927.5.2850	Hypothetical	96	3	7	91.9	0
2	Tb927.8.6660	Hypothetical	88	3	8	96.8	0
2	Tb10.6k15.2290	BS2 protein disulfide isomerase	84	3	6	95.5	0
2	Tb11.02.0170	Hypothetical	83	4	11	42.4	0
2	Tb927.1.2340, Tb927.1.2360, Tb927.1.2380, Tb927.1.2400	Alpha tubulin	74	2	6	99.9	0
2	Tb927.1.4180	Hypothetical	70	1	2	88.7	0
2	Tb11.01.5100	PFR	66	3	7	93.7	0
2	Tb10.6k15.2670	Hypothetical	63	2	3	71.2	0
2	Tb11.01.2530	Kinesin-like protein	61	2	3	84.1	0
2	Tb927.1.2330, Tb927.1.2350, Tb927.1.2370, Tb927.1.2390, Tb927.1.2410	Beta tubulin	52	1	2	99.9	0
2	Tb11.01.1680	Polyubiquitin	47	1	2	98	0
2	Tb11.01.3960	Hypothetical	38	3	8	69.8	0

Table 5.3- Table of all proteins identified from protein bands that were insoluble (Bands 1 and 3) or soluble (Bands 2 and 4) in sodium carbonate solution (see Figure 5.3). Score, MOWSE score defined by MASCOT® at p = 0.05; gCAI, genomic codon adaptation index (see Chapter 2); TmPred, trans-membrane prediction defined by TMHMM algorithm (see Chapter 2).

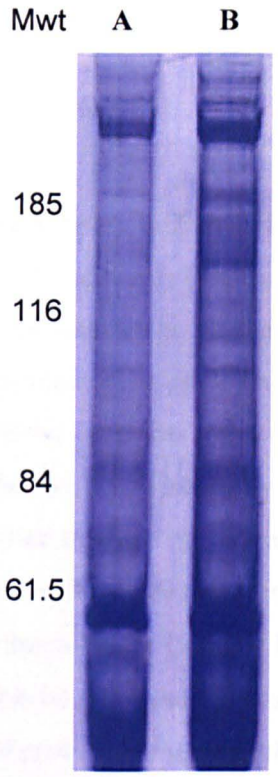


Figure 5.4 - Analysis of calcium calmodulin washed (A) and unwashed (B) *T. brucei* plasma membrane preparations. Samples were run on an 8% 1DGE until the 50kDa marker had reached the end of the gel. Nine corresponding bands from each lane that displayed differential solubility were selected for analysis by mass spectrometry. Molecular weight (kDa) markers (Mwt) are shown on the left hand side.

5.3.2 DiGE Analyses

5.3.2.1 2D PAGE

A total of 111 spots were identified as being differentially regulated between the parental TbAT1 KO and the drug resistant B48 cell lines (Figure 5.5). To enable easy gel-to-gel comparison, BVA and DIA-analysed data can be viewed in a variety of ways. For each spot, these include a 3-dimensional view and a graphical summary of all gels in which the spot was identified (Figure 5.6). Of the 111 spots that were statistically validated as being up or down regulated relative to the two cells lines, only 28 spots (representing a total of 43 proteins) were positively identified by mass spectrometry (Table 5.4). Interestingly, it appears that there are two blocks of co-localised proteins that are essentially exclusively expressed only in the TbAT1 KO line or in the B48 cell line (Figure 5.5). The different masses and iso-electric points of this block of proteins suggest that the blocks represent two differentially expressed sets of proteins rather than the effect being mediated by post-translational modifications to a single group of common proteins. We analysed for any commonality in proteins identified in these two groups, but did not find any duplication.

A major problem with running *T. brucei* whole cell extracts was consistently poor gel resolution in the second dimension i.e. when separating by molecular weight. Despite repeating the experiment several times and adding additional protease inhibitors, we were unable to resolve this issue. When running purified PM samples (see Chapter 4) the resolution was excellent. This suggests that despite supplementing with multiple protease inhibitors, protein degradation was occurring during sample preparation, and that the streaking was not a result of technical errors in running the samples on the gels. This phenomenon was not observed when performing a very similar analysis on procyclic *T. brucei* whole cell extracts (Foucher *et al.*, 2006). This suggests that there could be an additional proteolytic factor associated exclusively with the bloodstream form stage, which is resistant to standard protease inhibitors. Additionally, the fractionation protocol appears to remove this factor thus preventing PM sheet proteolysis.

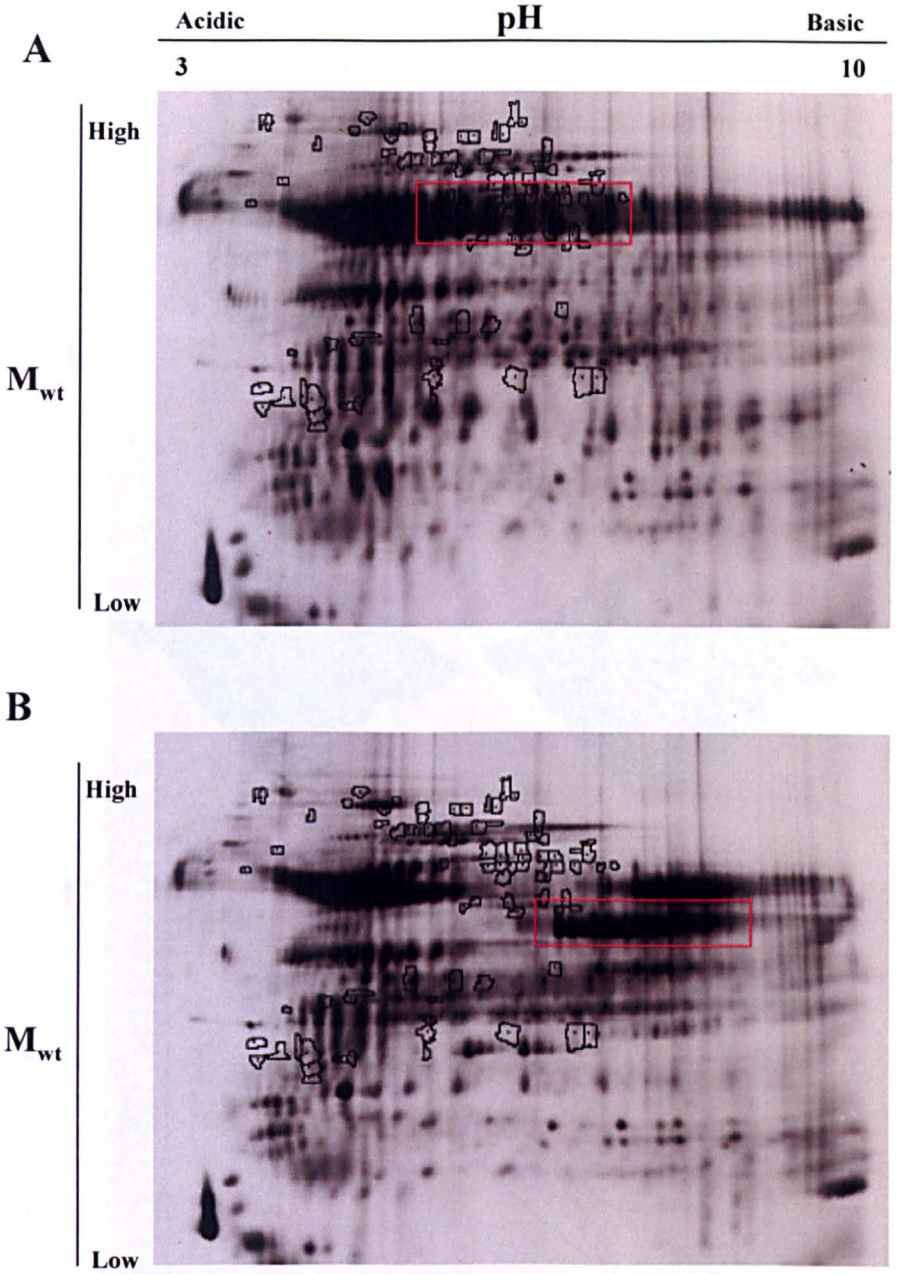


Figure 5.5 - Differential protein expression analysis of *T. brucei* whole cell extracts isolated from TbAT1 KO cells labelled with Cy3 (A) or pentamidine resistant B48 cells labelled with Cy5(B) separated by pI and mass on a pH 3-10 non-linear 2DGE. These images are derived from a single gel, but are representative of the replicate gels. Spots outlined in black were identified as being statistically up or down regulated in the two samples and were selected for MS analysis. A large discordant block of proteins between the two gel images is also highlighted (□).

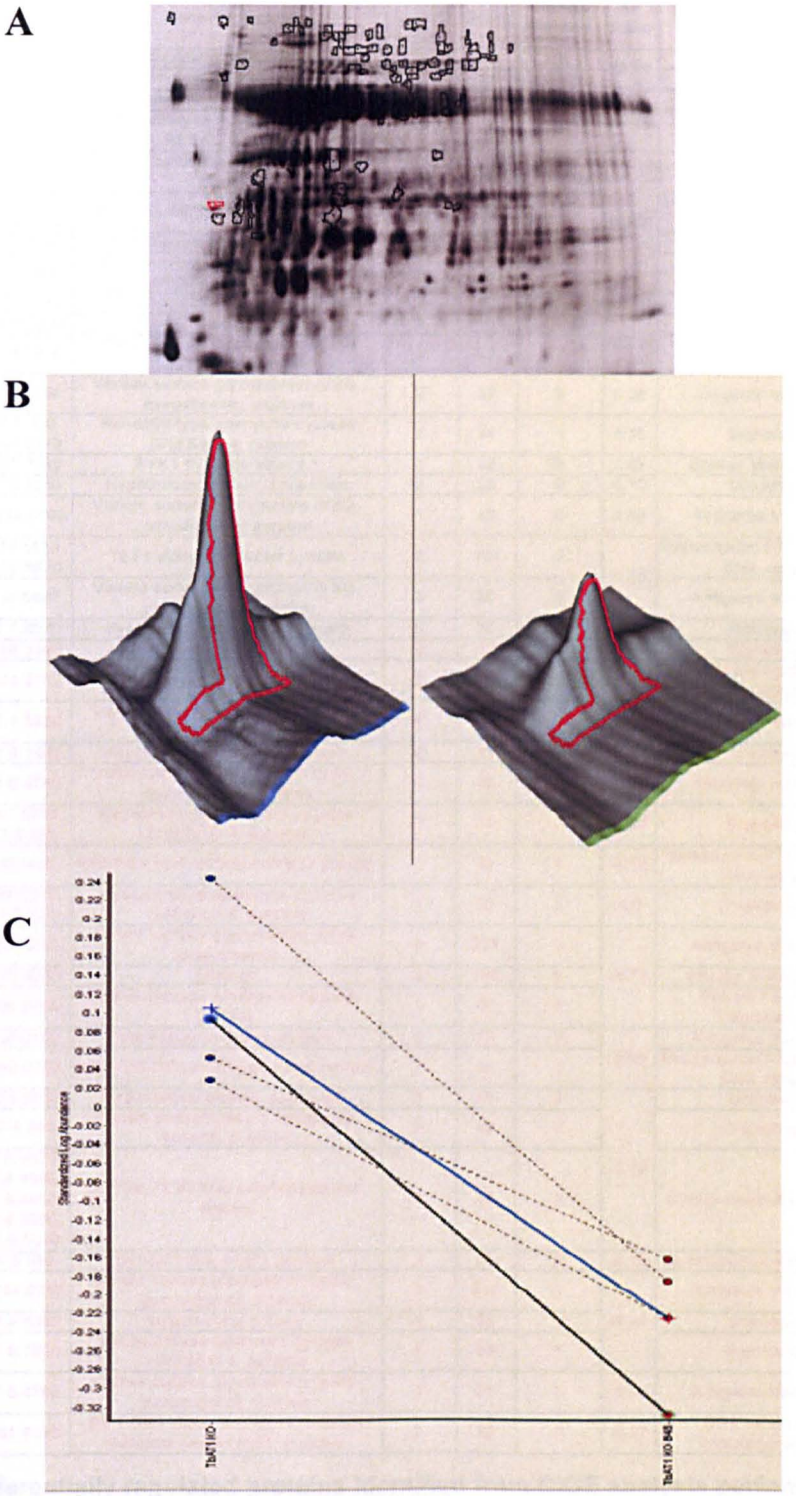


Figure 5.6 – Down regulation of spot 2621 in drug resistant B48 *T. brucei* compared to the parental TbAT1 KO strain. Location of the spot in the B48 gel image (A), its 3D intensity representation (B) and gel-to-gel intensity comparisons (C) are shown. This spot was identified as heat shock protein 70 (Tb11.01.3110) by mass spectrometry.

Spot ID	Accession Number	Gene Annotation	No. peptide ID's	Score	TmPred (TMHMM)	Av. Ratio	Gene Function
1558	Tb11.01.0230	Hypothetical protein, conserved	3	43	0	48.76	Unknown
	Tb927.6.3090	Hypothetical protein, conserved	2	34	0		Unknown
1547	Tb11.22.0007	Hypothetical protein, conserved	3	32	0	26.94	Unknown
	Tb927.8.7590	Receptor-type adenylate cyclase GRESAG 4, putative	5	47	1		Signalling
1626	Tb10.v4.0095	Variant surface glycoprotein (VSG), putative	3	42	0	23.5	Antigenic Variation
	Tb927.4.5020, Tb927.8.7400	RNA polymerase IIA largest subunit	2	38	0		Transcription / Translation / DNA repair
	Tb927.5.2990	Hypothetical protein, conserved	2	32	0		Unknown
	Tb927.8.7590	Receptor-type adenylate cyclase GRESAG 4, putative	4	32	1		Signalling
1545	Tb11.02.1540	Hypothetical protein, conserved	3	37	0	23.44	Unknown
1521	Tb09.160.0930	Protein kinase, putative	2	39	0	22.1	Signalling
	Tb927.8.7590	Receptor-type adenylate cyclase GRESAG 4, putative	3	37	1		Signalling
1515	Tb09.v4.0184	Variant surface glycoprotein (VSG, pseudogene), putative	2	37	0	9.25	Antigenic Variation
1331	Tb927.5.320, Tb11.01.5310	Receptor-type adenylate cyclase GRESAG 4, putative	2	34	1	5.06	Signalling
2669	Tb10.61.2680	PYK1 pyruvate kinase 1	1	32	0	4.42	Energy Metabolism
2353	Tb927.5.3240	Hypothetical protein, conserved	2	32	0	4.13	Unknown
1286	Tb09.244.0780	Variant surface glycoprotein (VSG, pseudogene), putative	1	46	0	3.89	Antigenic Variation
1400	Tb10.70.5650, Tb10.70.5670	TEF1 elongation factor 1-alpha	6	123	0	3.43	Transcription / Translation / DNA repair
	Tb10.v4.0047	Variant surface glycoprotein (VSG, pseudogene), putative	3	33	0		Antigenic Variation
1404	Tb927.7.6670	Hypothetical protein, conserved	1	32	3	2.71	Unknown
1422	Tb11.03.0340	Protein kinase, putative	3	37	0	-2.07	Signalling
2621	Tb11.01.3110	Heat shock protein 70	2	66	0	-2.16	Protein Folding / Degradation
2584	Tb927.1.5320	Variant surface glycoprotein (VSG, pseudogene)	4	35	0	-2.33	Antigenic Variation
1146	Tb927.6.1480	Hypothetical protein, conserved	2	31	1	-2.41	Unknown
1003	Tb927.6.5710	Variant surface glycoprotein (VSG, pseudogene), putative	4	36	0	-3.37	Antigenic Variation
1368	Tb11.01.5310, Tb927.5.320	Receptor-type adenylate cyclase GRESAG 4, putative	1	31	2	-3.47	Signalling
1032	Tb11.02.0490	KREPB4 RNA-editing complex protein	1	33	0	-3.82	Transcription / Translation / DNA repair
2485	Tb11.01.5310, Tb927.5.320	Receptor-type adenylate cyclase GRESAG 4, putative	1	33	2	-4.2	Signalling
2642	Tb09.244.0780	Variant surface glycoprotein (VSG, pseudogene)	6	233	0	-4.71	Antigenic Variation
	Tb10.70.4740	Enolase	5	175	0		Energy Metabolism
	Tb11.01.2000	hslVU complex proteolytic subunit, putative	1	49	0		Protein Folding / Degradation
2644	Tb09.v4.0142	Variant surface glycoprotein	2	33	0	-5.06	Antigenic Variation
	Tb11.02.0490	KREPB4 RNA-editing complex protein	7	36	0		Transcription / Translation / DNA repair
2450	Tb10.26.0090	Hypothetical protein, conserved	3	41	0	-5.13	Unknown
	Tb927.5.590	Protein phosphatase 1, regulatory subunit, putative	3	49	0		Signalling
	Tb927.8.4970, Tb927.8.4980, Tb927.8.4990, Tb927.8.5000, Tb927.8.5010	PFR-B / 2.69 kDa paraflagellar rod protein	3	40	0		Cytoskeleton Associated
2582	Tb927.3.930	Dynein heavy chain, putative	3	37	0	-6.16	Cytoskeleton Associated
2668	Tb09.244.0780	Variant surface glycoprotein (VSG, pseudogene), putative	5	312	0	-6.44	Antigenic Variation
	Tb927.4.5330	Hypothetical protein	1	33	0		Unknown
	Tb927.8.7590	Receptor-type adenylate cyclase GRESAG 4, putative	5	35	1		Signalling
1328	Tb927.5.4780	Variant surface glycoprotein (VSG, pseudogene), putative	3	31	0	-9.04	Antigenic Variation
2422	Tb11.01.6420	PIF1 DNA repair and recombination helicase protein PIF1, putative	2	32	0	-9.57	DNA repair and recombination

Table 5.4 – Differentially regulated proteins identified from DiGE analysis performed on *T. brucei* whole cell extracts. Spots are arranged according to average fold-regulation relative to the parental TbAT1 KO line. Proteins up regulated in B48 are highlighted in green, while those down regulated in B48 are highlighted in red.

5.3.2.2 16-BAC

The enrichment strategies outlined above were not completed in time to be used for the 16-BAC DiGE analysis. The experiment was therefore performed on PM samples generated from the fractionation procedure. To compensate for the reduced resolution of the 16-BAC gels in comparison to standard 2DGE, gels were scanned at two PMT voltages to effectively increase the dynamic range of the scanner.

A total of 16 differentially regulated spots were identified from the analysis, of which 10 spots gave positive protein identifications by MS. 3 of these spots yielded single protein identifications, while two or more protein identifications were made in the other 7 spots. All protein identifications are shown in Table 5.5. The majority of spots showed consistent up or down regulation (Figure 5.8), however one in particular (spot 365) showed very different spot intensities for the replicate samples (Figure 5.9). Initially this was thought to be a contaminant e.g. keratin introduced by the user during sample preparation. However when searched against the NCBI human genome database no significant hits were returned.

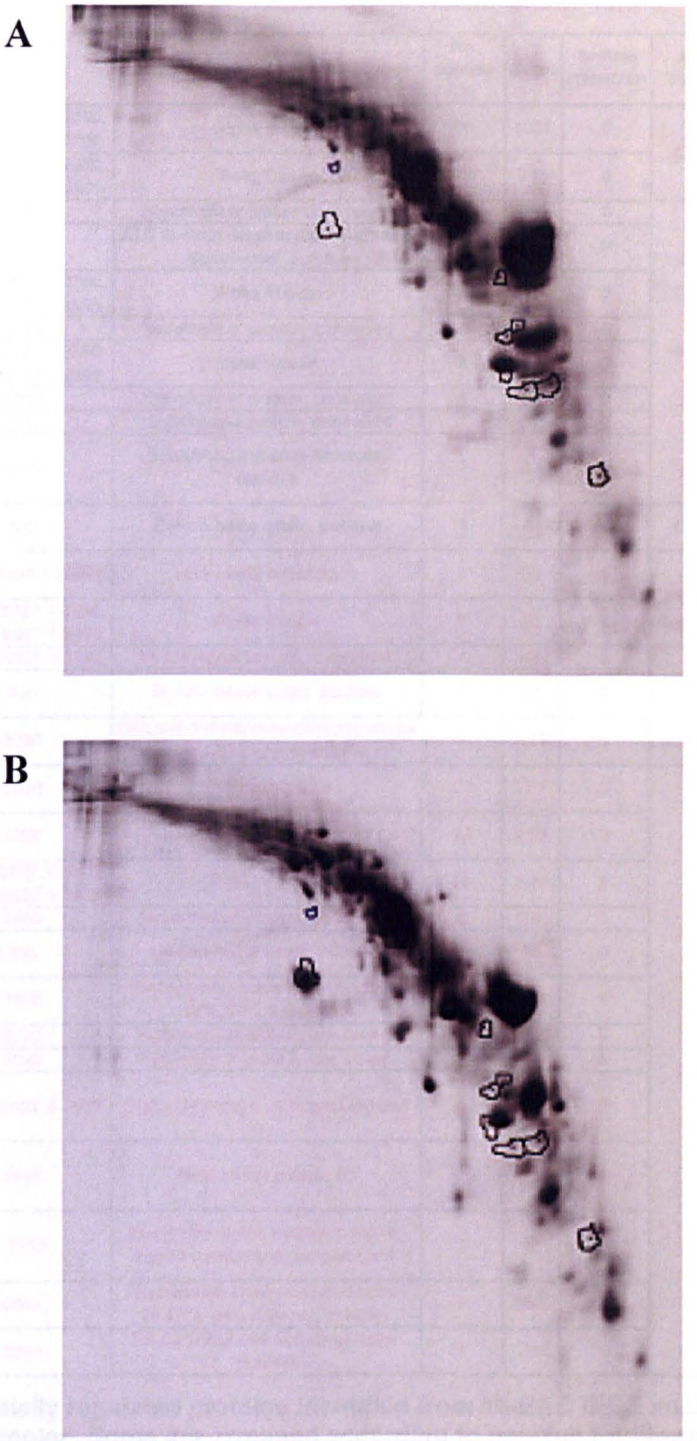


Figure 5.7 - Differential protein expression analysis of *T. brucei* PM samples isolated from TbAT1 KO labelled with Cy3 (A) or pentamidine resistant B48 cells labelled with Cy5 (B) separated by mass on both dimensions on a 16-BAC gel. Outlined spots indicate spots selected for MS analysis as they were identified as being up or down regulated between the two samples.

Spot ID	Accession Number	Gene Annotation	No. peptide ID's	Score	TmPred (TMHMM)	Av. Ratio	Gene Function
733	Tb927.1.2340, Tb927.1.2360, Tb927.1.2380,Tb927.1.2400	Alpha Tubulin	28	1024	0	5.67	Cytoskeleton Associated
	Tb927.1.2330, Tb927.1.2350, Tb927.1.2370, Tb927.1.2390	Beta Tubulin	22	713	0		Cytoskeleton Associated
818	Tb927.2.2160	Hypothetical protein, conserved	16	608	0	4.42	Unknown
	Tb10.70.1370	ALD fructose-bisphosphate aldolase, glycosomal, putative	5	154	0		Energy metabolism
	Tb927.1.2340, Tb927.1.2360, Tb927.1.2380,Tb927.1.2400	Alpha Tubulin	5	151	0		Cytoskeleton Associated
	Tb10.6k15.0140	Hypothetical protein, conserved	3	88	0		Unknown
	Tb927.1.2330, Tb927.1.2350, Tb927.1.2370, Tb927.1.2390	Beta Tubulin	3	73	0		Cytoskeleton Associated
	Tb10.70.4540	Hypothetical protein, conserved	1	32	0		Unknown
	Tb927.1.1330	Hypothetical protein, conserved	1	32	0		Unknown
	Tb09.160.3270	Eukaryotic initiation factor 4a, putative	1	31	0		Transcription / Translation / DNA repair
183	Tb927.3.930	Dynein heavy chain, putative	4	44	0	1.26	Cytoskeleton Associated
483	Tb10.70.5800 / Tb10.70.5820	HK1 / HK2 hexokinase	1	38	0	-1.39	Energy metabolism
	Tb927.1.2340, Tb927.1.2360, Tb927.1.2380, Tb927.1.2400	Alpha tubulin	2	33	0		Cytoskeleton Associated
863	Tb927.3.3770, Tb927.3.3790	Hypothetical protein, conserved	1	46	0	-1.48	Unknown
	Tb927.3.930	Dynein heavy chain, putative	2	32	0		Cytoskeleton Associated
555	Tb927.8.3530	Glycerol-3-phosphate dehydrogenase [NAD+], glycosomal	4	115	0	-2.07	Energy metabolism
432	Tb10.70.5820	HK1 hexokinase	27	787	0	-2.3	Energy metabolism
	Tb10.70.5800	HK2 hexokinase	23	730	0		Energy metabolism
	Tb927.1.2340, Tb927.1.2360, Tb927.1.2380, Tb927.1.2400	Alpha tubulin	24	728	0		Cytoskeleton Associated
	Tb11.01.2460	Hypothetical protein, conserved	3	72	1		Unknown
	Tb927.3.930	Dynein heavy chain, putative	4	55	0		Cytoskeleton Associated
	Tb927.8.7590	Receptor-type adenylate cyclase GRESAG 4, putative	4	42	1		Signalling
	Tb927.5.2270	Hypothetical protein, conserved	2	40	0		Unknown
	Tb11.01.2090	Hypothetical protein, conserved	2	36	0		Unknown
	Tb927.4.5020, Tb927.8.7400	RNA polymerase IIA largest subunit	4	36	0		Transcription / Translation / DNA repair
	Tb11.01.3110	Heat shock protein 70	1	35	0		Protein Folding / Degradation
	Tb09.211.2110	Elongation factor 1-alpha, putative; hsp70 subfamily B suppressor 1	2	33	0		Transcription / Translation / DNA repair
	Tb11.44.0001	Expression site-associated gene (ESAG, pseudogene), putative	5	32	0		Unknown
	Tb927.4.5040	Dihydrolipoamide dehydrogenase, putative	1	31	0		Energy metabolism

Table 5.5 - Differentially regulated proteins identified from 16-BAC DiGE analysis performed on *T. brucei* PM samples. Spots are arranged according to average fold-regulation relative to the parental TbAT1 KO line. Proteins up regulated in B48 are highlighted in green, while those down regulated in B48 are highlighted in red.

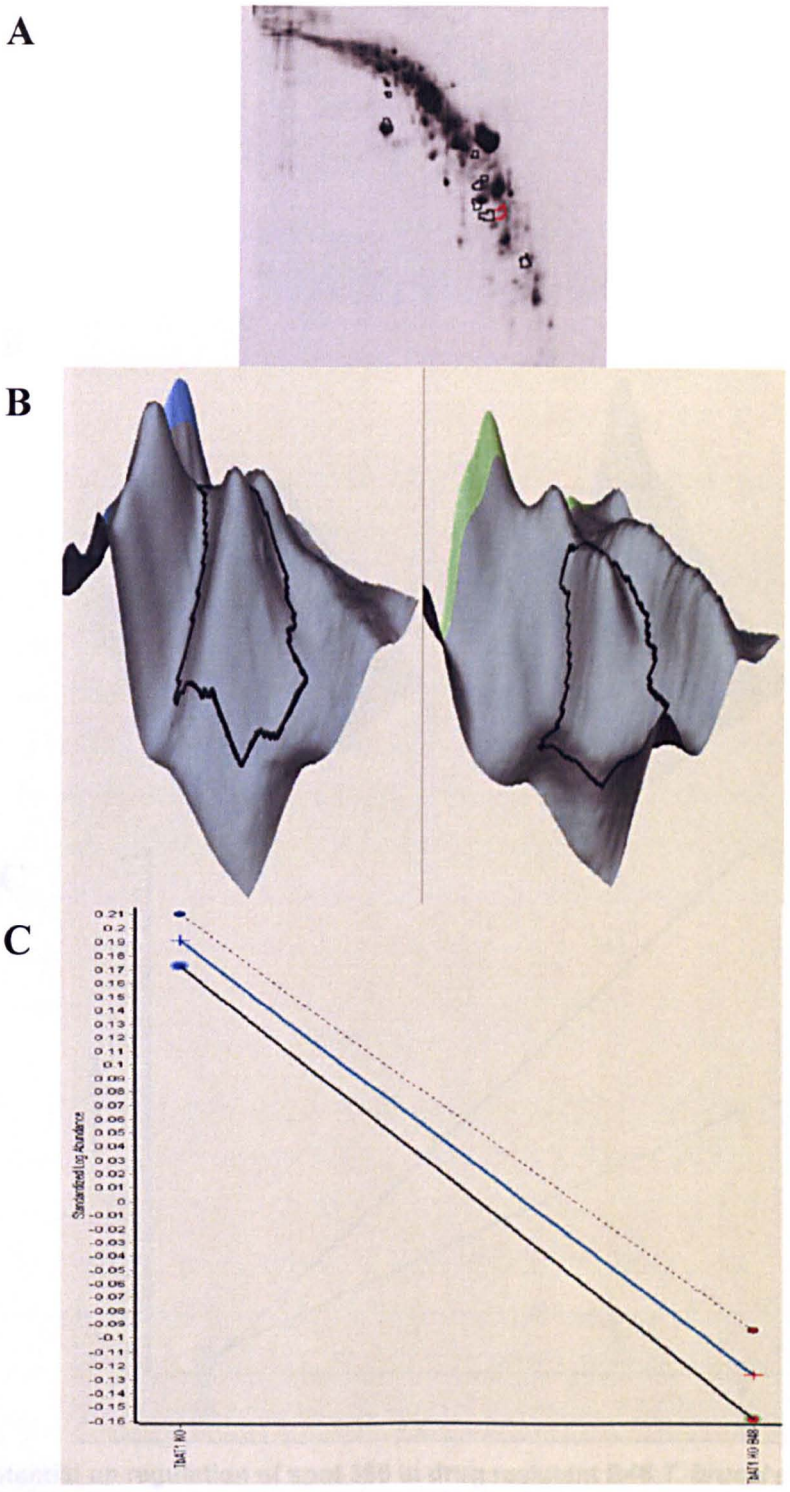


Figure 5.8 – Two-fold down regulation of spot 555 in drug resistant B48 *T. brucei* compared to the parental TbAT1 KO strain. Location of the spot in the TbAT1 KO gel image (A), its 3D intensity representation (B) and gel-to-gel intensity comparisons (C) are shown. This spot was identified as glycerol-3-dehydrogenase.

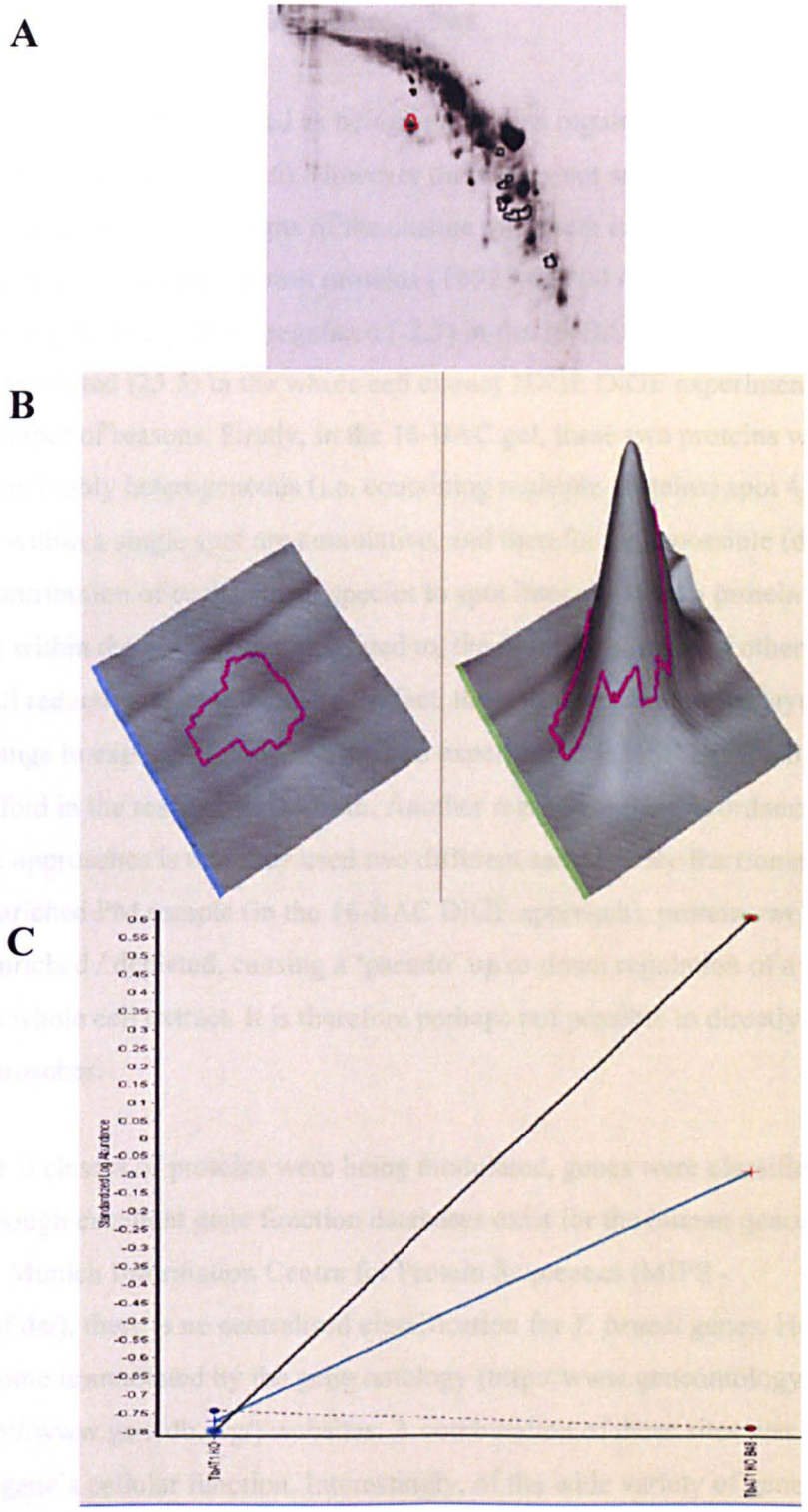


Figure 5.9 – Potential up regulation of spot 365 in drug resistant B48 *T. brucei* compared to the parental TbAT1 KO strain. Location of the spot in the B48 gel image (A), its 3D intensity representation (B) and gel-to-gel intensity comparisons are shown. This spot was not identified by MS analysis.

5.3.2.3 Comparison of the DiGE approaches

A number of proteins were identified as being up or down regulated in both of the DiGE approaches employed (see Table 5.6). However there does not seem to be much harmony between the two experiments in terms of the change in protein expression. For example, the RNA Polymerase IIA large subunit proteins (Tb927.4.5200 / Tb927.8.7400) are identified as being modestly down regulated (-2.3) in the 16-BAC experiment, but massively up regulated (23.5) in the whole cell extract 2DGE DiGE experiment. This may be due to a number of reasons. Firstly, in the 16-BAC gel, these two proteins were identified in the highly heterogeneous (i.e. containing multiple proteins) spot 432. Changes in expression within a single spot are cumulative, and therefore it is possible (depending on the relative contribution of each protein species to spot intensity) that a protein was up regulated, but within the gel region it migrated to, the down regulation of other proteins give an overall reduction in spot intensity. In fact, the only protein that displays a consistent change in expression between the two experiments is HSP 70 which is down regulated ~2-fold in the resistant B48 strain. Another reason for the discordance between the two DiGE approaches is that they used two different samples. By fractionating to generate an enriched PM sample (in the 16-BAC DiGE approach), proteins were specifically enriched / depleted, causing a 'pseudo' up or down regulation of a protein relative to the whole cell extract. It is therefore perhaps not possible to directly compare these two approaches.

To try and see if classes of proteins were being modulated, genes were classified by function. Although excellent gene function databases exist for the human genome e.g. FunCat at the Munich Information Centre for Protein Sequences (MIPS - http://mips.gsf.de/), there is no centralised classification for *T. brucei* genes. However, the *T. brucei* genome is annotated by the gene ontology (http://www.geneontology.org/) and GeneDB (http://www.genedb.org/) websites. A combination of these sites was used to classify each gene's cellular function. Interestingly, of the wide variety of gene functions only 6 functional classes (besides 'unknown') were identified (Table 5.7), perhaps suggesting that despite the different sample preparation procedures, the overall picture is remarkably consistent.

Gene ID	Gene Annotation	Av. Ratio					
		16-BAC DiGE ID's			2DGE DiGE ID's		
Tb927.3.930	Dynein heavy chain, putative	1.26	-1.48	-2.3	-6.16		
Tb927.4.5020 / Tb927.8.7400	RNA polymerase IIA largest subunit	-2.3			23.5		
Tb927.8.7590	Receptor-type adenylate cyclase GRESAG 4, putative	-2.3			26.94	22.1	-6.44
Tb11.01.3110	Heat shock protein 70	-2.3			-2.16		

Table 5.6 – Proteins identified as being up or down regulated in both the standard 2DGE whole cell extract and 16-BAC plasma membrane fractionated DiGE experiments. Where proteins were identified multiple times within an experiment, unique identifications are shown first.

Gene Function	Experimental Approach	
	Soluble DiGE	16-BAC DiGE
Cytoskeleton Associated	2	9
Energy metabolism	2	6
Antigenic Variation	11	-
Protein Folding / Degradation	2	1
Signalling	11	1
Transcription / Translation / DNA repair	5	3
Unknown	10	9
Total	**43**	**29**

Table 5.7 - Functional analysis of the genes identified as being differentially expressed by DiGE analysis in the B48 drug-resistant strain in comparison to the parental TbAT1 KO strain in whole cell extracts (Soluble DiGE) and PM preparations (16-BAC DiGE).

We also noticed that a large number of the regulated proteins were identified as adenylate cyclases. By collating these identifications together (Table 5.8) it immediately becomes obvious that in fact potentially only three AC's are represented (two of which could not be differentiated due to sequence homology). All three of these proteins have been previously observed in proteomic experiments performed on procylic *T. brucei* (Jones *et al.*, 2006), and they are all predicted to contain one (Tb927.5.320 / Tb927.8.7590) or two TMD's (Tb11.01.5310) as well as a signal peptide for trafficking to the plasma membrane. These proteins are therefore likely to be true plasma membrane proteins.

Gene Annotation	Accession Number	TmPred (TMHMM)	Source	Spot ID	No. peptide ID's	Score	Av. Ratio
Receptor-type adenylate cyclase GRESAG 4, putative	Tb927.8.7590	1	2DGE	1547	5	47	26.94
			2DGE	1626	4	32	23.5
			2DGE	1521	3	37	22.1
			16-BAC	432	4	42	-2.3
			2DGE	2668	5	35	-6.44
	Tb927.5.320, Tb11.01.5310	1/2	2DGE	1331	2	34	5.06
			2DGE	1368	1	31	-3.47
			2DGE	2485	1	33	-4.2

Table 5.8 - Regulation of proteins identified as putative adenylate cyclase's identified by standard 2DGE DiGE and 16-BAC DiGE. Av. ratios are expressed relative to KO. TmPred, Number of predicted TMD's.

5.3.3 I-CAT

From the Proteogest analysis (Cagney *et al.*, 2003), 300 proteins i.e. less than 3% of the proteome were identified as lacking cysteine-containing residues. Apart from a lack of polytopic membrane proteins containing 4 or more predicted TMD's, analysis of these proteins did not reveal any obvious associations i.e. they appeared to be a random selection of proteins from the proteome. This suggests that I-CAT should theoretically give good overall proteome coverage. However, it would have been beneficial to extend this analysis to map the number of cysteine containing peptides per protein. By manipulating the theoretical probability of any one peptide being detected by MS, the likely coverage could have been better calculated i.e. if a protein generated 3 cysteine containing peptides and there was a 50% chance that any one peptide would be measured by MS, on average only 1 of these peptides would be observed, which would be unlikely to generate a significant score. In the absence of such an analysis, we felt it would be prudent to analyse both the affinity purified cysteine-derivitised peptides and the unlabelled peptides.

MS data derived from PM fraction / calmodulin washed PM fraction, for both labelled and unlabelled peptides were analysed by MASCOT®. From all of the I-CAT labelled fractions, only two proteins - alpha tubulin (Tb927.1.2340, Tb927.1.2360, Tb927.1.2380, Tb927.1.2400), and paraflagellar rod protein (Tb927.3.4290, Tb927.3.4300, Tb927.3.4310, Tb927.3.4320, Tb927.3.4330) were confidently identified. However when analysed using the Pro I-CAT software, the only significant hit (with a total score of 1.3 i.e. 95% confident) was to Gim5B (Tb09.211.2740), which has been localised to the glycosome. In contrast, 72 proteins were identified from all of the unlabelled fractions.

5.4 Discussion

In order to perform a quantitative analysis on the PM proteome of *T. brucei*, the initial aim was to remove the contaminating cytoskeletal proteins that were identified in the PM preparations in Chapter 4. While a number of techniques have been developed for such an approach, in our hands their ability to selectively enrich for specific proteins (as determined by 1D SDS-PAGE) overall was not very effective. Further complications e.g. removal of the detergent Triton X-114 proved a major problem. It was also difficult to quantify the effect that each enrichment technique had on the proteome, as banding patterns were often difficult to compare. A recombinant line expressing a tagged membrane protein would have been immensely useful in determining the optimum enrichment protocol, although attempts to generate such a line (Chapter 4) were unsuccessful.

With hindsight, it would have been useful to examine the effect of combining multiple enrichment techniques on the proteome. It has been reported that the various techniques differ in the type of protein they select (Zhang *et al.*, 2006). For example chloroform / methanol partitioning was identified as being the most selective in terms of retaining hydrophobic, but not low-abundance proteins. In comparison, Triton X-114 was able to enrich for low-abundance proteins, but was less capable in resolving highly hydrophobic proteins (Zhang *et al.*, 2006). By combining these two procedures, it may be possible to select for low-abundance hydrophobic proteins. An alternative approach would be to try and prevent the non-specific association of proteins to the plasma membrane. Therefore, although cells need to be sheared in a hypo-osmotic buffer, by immediately elevating the pH post-lysis it might have been possible to prevent protein adhesion which has been so difficult to reverse.

It was surprising to see a number of VSGs identified as being differentially regulated in the soluble DiGE approach (i.e. using a whole cell extract), considering trypanosomes exhibit tight regulation on VSG gene expression to ensure that only a single VSG is transcribed at any one time (Donelson, 2003). However, a closer inspection of the MS data reveals that with the exception of Tb09.244.0780 (which was definitively identified twice), all other VSG identifications had scores close to the confidence threshold (although they were unambiguously identified). To confirm the presence of these additional VSG proteins, it would be necessary to repeat the experiment, as it is possible that they are false positives. Another explanation is that a subset of the total population of trypanosomes has switched

expression to one of the other VSGs present in the genome. If so, then these additional isoforms would appear as regulated spots. As expected, no VSG's were identified from the 16-BAC approach.

The 16-BAC DiGE approach employed in this chapter has not previously been published. This new application of the DiGE technology promised to deliver the advantages of a gel-based system i.e. resolution of different protein isoforms (although most likely only molecular weight variants e.g. truncations etc), while removing traditional 2DGE-based impediments e.g. hydrophobic protein incompatibilities. In summary, 16-BAC DiGE is able to identify changes in the proteome, including polytopic membrane proteins (although none were identified as being regulated in this analysis). There are limitations in terms of this technique's ability to resolve highly complex mixtures, and this approach is therefore envisaged to be only applicable for highly fractionated / simple samples. However, it is important to consider that most polytopic integral membrane proteins (by virtue of the limited space that is available for them to occupy) are expressed at very low levels, and would therefore only be identified if highly enriched, for example by subcellular fractionation. One of the issues that may be resolvable is where multiple proteins are identified in a single spot, and therefore where the relative contribution of each to the total fluorescence is unknown. While it would be impossible to determine the fluorescence contribution of each protein, the relative abundance of the different protein species may be estimated following an association being identified between the number of peptides identified and the relative abundance of that protein (Liu *et al.*, 2004). Taking this to its ultimate conclusion, it should be possible to run gels with different sample ratios and use changes in the number of peptide identifications to determine protein abundance.

It would have been interesting to analyse the physicochemical properties of proteins in relation to their deviation from the theoretical diagonal expected when separating proteins by MW in two dimensions on the 16-BAC gels, to test whether particular protein classes were specifically affected. Time permitting, applying the same approach to the newly developed 16-BAC multiphasic buffer system (m16-BAC) (Kramer, 2006) could also have enabled the development of protocols to selectively isolate particular protein classes by virtue of their gel migration, as a result of the different conditions employed within the two dimensions.

In comparison to the soluble DiGE analysis, a far greater proportion of the spots selected for MS analysis were identified in the 16-BAC DiGE approach. This could be due to the reduced resolution in the 16-BAC gels that results in proteins being located far closer to

one another. In addition, the permissive 16-BAC conditions enable many more (particularly hydrophobic) proteins to enter the gel phase, which would have been lost during the IEF stage in 2DGE. Any spot picked at random would therefore be more likely to contain more than one protein, making a protein identification of one (or more) of the proteins also more likely. Indeed, the relatively large number of proteins that are identified from certain spots supports this hypothesis. For example 14 proteins were identified from spot 432. Interestingly, when searching against the NCBI database, this spot was the only one found to contain keratin contamination. On the other hand it is likely that a number of spots do in fact only contain a single protein. However, unlike 2DGE spots, all species of a specific protein possessing the same or a similar molecular weight will co-migrate. This would lead to more intense spots i.e. higher local protein concentrations, which in turn would improve the chance of identifying that specific protein.

The dynamic range of protein regulation is far smaller in the 16-BAC over the 2DGE DiGE analysis (~6 fold versus ~48 fold respectively). This could be due to two reasons. Firstly, if a regulated protein co-migrates with a number of other (unregulated) proteins into a single spot, only the cumulative change in spot intensity will be measured. Therefore the other proteins present in the same spot will mask even large changes in expression of a single protein species. Another possibility is that instead of regulating protein copy number, the ratio of different protein isoforms is regulated. If these protein species essentially do not vary in molecular weight e.g. phosphorylation, they are unlikely to be separated on the 16-BAC gel, and therefore as long as total protein concentration remains constant, no change in protein levels will be identified.

The I-CAT analysis was extremely disappointing in terms of its ability to provide quantitative data. Originally heralded for its ability as a technique to reduce sample complexity (by virtue of selective retention of cysteine containing residues), in fact this very reduction has reduced sample complexity to such an extent that only three proteins (α-tubulin, paraflagellar rod, and Gim 5B) were identified, and none with quantitative data. Compared with the plasma membrane sub-proteome (as defined in Chapter 4), this number of identifications very obviously does not reflect the true sample complexity. While some peptides have no doubt been lost during the separation of labelled and unlabelled peptides, the majority of the labelled peptides should have been retained for analysis. A large number of I-CAT labelled peptides were identified, although virtually none were useful for protein identification, most probably due to a lack of corroborating peptides. In addition, the results from the two search engines; Pro I-CAT and MASCOT® did not correlate. This was despite both being given the same search parameters (e.g. MS tolerance, cleavage

specificity etc) and raw MS data files. This is perhaps not surprising considering that MASCOT® and Pro I-CAT use different algorithms: MOWSE and INTERROGATOR™ respectively. However, one would expect a certain amount of consistency between the two engines.

Attention therefore focused on the unlabelled peptide set, from which a considerable but by no means exhaustive set of proteins were identified (72 compared to 1233 proteins in Chapter 4). This reduced number must, at least in part, be due to the removal of cysteine-containing peptides, although judging by the average peptide score it does appear that the MS data quality was slightly inferior. It is possible that specifically labelling membrane proteins was poor in this experiment, and if repeated it would be advisable to label (in the presence of SDS and Urea) denatured proteins (Ramus *et al.*, 2006). However, although MASCOT® is able to identify heavy and light I-CAT tags on cysteine containing peptides, it is unable to quantify changes in expression. Frustratingly, there therefore appears to be enough information in the two data sets to both identify proteins (unlabelled) and quantify (labelled) their expression levels, however there is no software capable of combining the two. It would seem then that the greatest limitation to I-CAT is the inability to integrate these two complementary data sets, without which (particularly when focusing on low abundance polytopic membrane proteins) success would seem far from guaranteed.

The DiGE approaches identified a number of proteins that were expressed at different levels in the B48 pentamidine resistant strain compared to the parental TbAT1 KO line. Considering the resistance phenotype outlined in Chapter 3, it was difficult to imagine a direct role for any of the proteins identified in pentamidine uptake, considering none were predicted to contain more than 3 TMD's, and were therefore unlikely to represent pentamidine transporters or channels.

Despite the lack of polytopic integral membrane proteins identified, it is possible that one or more could modulate the transporter(s) responsible for drug transport, or its expression. Transport activity could be altered by protein-protein interactions or protein modifications, e.g. phosphorylation to induce protein conformational change. While more than 150 kinases have been identified in the *T. brucei* kinome (Berriman *et al.*, 2005) most have only been identified by homology searching, and therefore their roles remain unknown. Interestingly, the *T. brucei* kinome lack a number of kinase families, e.g. there are no receptor-linked kinases, but other kinase families are expanded, possibly to cope with environmental stress factors (Naula *et al.*, 2005). Two putative protein kinases were identified as being differentially regulated in the 2DGE DiGE approach. One in particular

(Tb09.160.0930) was massively up regulated in the B48 strain. Kinases, as with most enzymes are able to catalyse reactions extremely rapidly. Assuming that this kinase phosphorylates HAPT1to either reduce its activity, substrate specificity or propensity to be internalised, all of which would result in a reduction in pentamidine accumulation, up regulating this kinase would provide one route to down regulation of the transporter activity. A kinase-activated cascade to changes in gene expression is another possibility. What is clear from the 16-BAC analysis is that there have been extensive changes in the pellicular cytoskeleton associated with the plasma membrane. For example a number of tubulins have been up regulated in the drug resistant strain. It is difficult to see that reinforcing the structural integrity of the cells in this way assists the parasite in adapting to the drug. Alternatively this response may be a common stress response to the effect of the pentamidine.

A large group of proteins identified as being differentially regulated are the receptor-type adenylate cyclases (AC's). This class of proteins synthesise a second messenger molecule - cyclic adenosine monophosphate (cAMP), which is found in almost all living organisms, including trypanosomes. Interestingly the structure of trypanosomal AC's differ markedly from their mammalian hosts, and consist of a highly variable extracellular N-terminal domain and an intracellular catalytic domain linked by a single trans-membrane domain (Naula and Seebeck, 2000). In addition, while mammalian AC's are stimulated by G-protein coupled receptors (GPCR), no GPCRs have yet been detected in any of the trypanosomatids. This suggests that trypanosomal AC's, of which more than 60 genes are annotated as AC's in the published genome (Berriman *et al.*, 2005), may be enzyme linked receptors, and therefore represent novel drug targets (Seebeck *et al.*, 2004). From Table 5.8, a maximum of three AC's are regulated (two ORF's are identified from the same set of peptides). These genes are identified as being differentially regulated in multiple spots i.e. there are multiple protein isoforms. Indeed the vast majority of identifications are made from the soluble proteome (separated by 2DGE), whereas only one identification is made from the PM preps separated on a 16-BAC gel. Furthermore, the level of regulation in the 2DGE approach is far greater e.g. >20 fold, than in the 16-BAC approach e.g. -2.3 fold. All this suggests that total protein abundance of each of AC might be approximately constant, and that regulation of these AC's therefore alters the abundance of specific AC isoforms. This explains why there is both up and down regulation of the same protein. For both proteins these isoforms consist of either a relatively small truncation product (<50 kDa), or a protein series with slight variations in molecular weight and / or isoelectric point. Unfortunately due to the low sequence coverage the exact modifications could not be

identified. However for both proteins, the heavily truncated product (2668 and 2485) was down regulated in B48 and with the exception of spot 1368, all other isoforms up regulated. From the size of the truncation, this isoform obviously can not retain full AC function, therefore this would give an overall increase in the functional full-length transcript. Assuming that the full-length protein isoforms are differentially active, we postulate that this specific AC, Tb927.8.7590, and maybe one or two others, could in some way regulate HAPT1activity. This could be via a number of mechanisms including altering protein expression, sequestering HAPT1from the plasma membrane to intracellular vesicles, or directly modulating HAPT1activity. It is not very plausible that these AC isoforms are regulated as a result of directly sensing extracellular levels of pentamidine, as one would expect the resistance phenotype to be reversible over time and the PM preparations were derived from cells grown in the absence of pentamidine. We consider that these signalling proteins are very important and are likely to be able to elicit a number of cellular and/or metabolic effects, many of which could ablate the activity of HAPT, the physiological function of which is still unknown.

A number of the regulated proteins are involved in transcription, translation or DNA repair. This is interesting considering pentamidine's minor groove DNA binding affinity (Shapiro and Englund, 1990), and the subsequent implication that its mode of action involves DNA damage (Wilson *et al.*, 2005). This mode of action is supported by the finding that pentamidine is excluded from the mitochondria in a pentamidine resistant *L. mexicana* strain (Basselin *et al.*, 2002). We therefore looked at the localisation of all these proteins and found that the majority are localised to the nucleus or cytoplasm. Interestingly though, the helicase - Pif1 (Tb11.01.6420), which was shown to be ~10-fold down regulated in our drug resistant strain is predicted to be present in both the nucleus and mitochondrion. Pif1-like proteins have been studied in many different organisms, but were first identified in *Saccharomyces cerevisiae* (Lahaye *et al.*, 1991). Although not essential, Pif1 helps mediate nuclear genomic stability by the inhibition (Zhou *et al.*, 2000) and removal (Boule *et al.*, 2005) of telomerase to reduce telomere elongation and *de novo* telomere formation (Schulz and Zakian, 1994). In addition to these roles, Pif1 helicases mediate other events in nuclear DNA replication, although its function in these processes is less well characterised (for review see (Boule and Zakian, 2006)). In the mitochondria, Pif1 is heavily involved in maintaining mitochondrial genome stability (O'Rourke *et al.*, 2005), and if absent in yeast, the mitochondrial DNA is lost (Boule and Zakian, 2006). From the fluorescence work in Chapter 3, it is clear that down regulation of Pif1 does not give rise to an obvious loss of mitochondrial DNA. However it possible that pentamidine-

bound DNA promotes Pif1 helicase activity in the kinetoplast to generate mini-circle linearisation. Indeed this outcome has previously been observed in *T. equiperdum* (Shapiro and Englund, 1990) and can lead to dyskinetoplastidy (i.e. the loss of the kinetoplast / mitochondrial DNA). We therefore hypothesise that Pif1 detects, unwinds, and maintains pentamidine-bound DNA in a linear conformation, creating a barrier to replication. By down-regulating expression of Pif1, mitochondrial DNA is not linearised and this replication block can be bypassed. This mechanism is likely to be associated with a fitness cost in terms of the parasite's ability to withstand oxidative stress (Doudican *et al.*, 2005). This hypothesis could be tested by assessment of the sensitivity of the parental and drug resistant strains to hydrogen peroxide.

The lack of in-depth functional annotation of many of the genes identified means that approximately 30% (21 / 75) of all the regulated proteins are simply annotated as hypothetical. This group of proteins represent the hardest to investigate, after all, where does one start characterising a protein without any indication of its role or function? Fortunately, a huge variety of techniques are available to interrogate gene function in trypanosomes e.g. RNAi, gene knockout, over-expression, as well as good assays for characterisation of these manipulated strains. Nevertheless there are many challenges in analysing this class of genes, but perhaps these genes represent the most exciting findings considering their potential for mediating a range of novel functions, possibly even including completely new biochemical pathways.

In summary then it can be seen that a number of proteins have been identified as being differentially regulated in the B48 strain. The potential role of many of these in the drug resistance phenotype remains unclear and it is unlikely that all are directly linked to the phenotype. Many may be differentially regulated to compensate for those that mediate the resistance phenotype e.g. it is difficult to envisage structural proteins such as tubulin having a direct role in resistance. Alternatively, regulation of these proteins may simply reflect proteomic drift over time. Considering the dynamic nature of the proteome, and the fact that each sample analysis effectively represents a proteomic 'snapshot', a component of the differential regulation observed may fall within the homeostatic thresholds of the cell and despite the experimental replicates may represent natural variation within the population (e.g. with VSG switching, or differences in the cell cycle phase between populations). Despite these caveats, there are a number of interesting proteins e.g. Pif1, whose identification hints at their role in pentamidine resistance. Further investigation of these proteins is now required.

Chapter 6 -

General Discussion

A number of different strands of research have culminated in the production of this thesis. This chapter aims to combine the different strands to integrate the research into the current models of drug resistance in trypanosomes and highlight areas of further interest and where attention needs to be focused to ultimately elucidate the mechanism(s) of pentamidine resistance in our lab derived B48 *T. brucei* strain.

6.1 The Problem

HAT is a disease that affects the rural poor throughout Africa. While the World Health Organisation has recently revised the number of people infected with trypanosomiasis(WHO, 2006), prompting new hopes to eradicate this disease (Barrett, 2006), it is likely to be with us for the foreseeable future and drug resistance is a major factor in this. In addition, trypanosomiasis in livestock is still an acute problem holding back agricultural and, hence, economic development in the tsetse belt. In the interests of eradicating African trypanosomiasis, new tools are required to combat it. While the impact of active surveillance and the establishment of continent-wide treatment is no doubt the key to success, drug resistance threatens to undermine any control strategy. Understanding resistance with a view to implementing better diagnostics for the application of appropriate treatments (Stewart *et al.*, 2005) as well as developing resistance reversers and new therapies must therefore be a priority (Pink *et al.*, 2005).

6.2 The approach

Our approach of developing resistance by exposure to sub-curative levels of drug is by no means a new idea, and has previously been applied a number of times to trypanosomes (Barrett *et al.*, 1995; Berger *et al.*, 1995; Fairlamb *et al.*, 1992; Frommel and Balber, 1987; Rollo and Williamson, 1951; Scott *et al.*, 1996). However with the exception of the P2 (TbAT1) transporter (Matovu *et al.*, 2001b; Matovu *et al.*, 2001a), no other proposed resistance mechanism has been validated in the field. For example over expression of TbMRPA has been shown to mediate resistance (Shahi *et al.*, 2002), but this has not yet been observed in the field (P. Mäser, Personal Communication). Identifying resistance determinants from field isolates is possible if a candidate resistance marker has previously

been identified. However, in the current study we chose to induce resistance in a controlled environment as no candidate genes for pentamidine resistance other than P2 have been suggested and our comparative approach necessitates the use of a parental reference strain. In view of the resistance phenotype based on reduced intracellular accumulation, observed with TbAT1 and TbMRPA (which do not act synergistically (Lüscher *et al.*, 2006)), and the identification of additional pentamidine routes of entry (HAPT and LAPT (De Koning, 2001)) we hypothesised that further resistance might well be achieved in a similar fashion. Consistent with this model, we show here that the drug resistant trypanosomes thus generated had massively reduced pentamidine transport, and crucially, appeared to lack the HAPT transporter. The loss of this transporter allowed us to investigate transporter selectivity, and show that DAPI is transported by HAPT. This finding could easily be combined with other fluorescence tests (Stewart *et al.*, 2005) for resistance analysis in the field. The increase in pentamidine resistance coincided (albeit to a lesser extent) with increased resistance to other drugs e.g. melarsen oxide. Considering the structural diversity of these compounds, it is difficult to imagine how cross-resistance arises, except if transport-related (as seen with TbAT1). We therefore propose that another transporter, i.e. HAPT, contributes the largest component of resistance. Analysing the status of this transporter in drug refractory strains would be of very considerable value in determining the level of drug resistance in a population.

Considering the relative ease with which resistance was generated makes the absence of pentamidine resistance in the field rather perplexing, considering the long-term use of the drug (Bray *et al.*, 2003). This would suggest that the loss of HAPT is detrimental to the parasite, although its endogenous substrate remains undetermined (De Koning, 2001). Indeed, difficulties in establishing infections *in vivo* (Chapter 3) strongly suggest that B48 has reduced virulence. Such a reduction in fitness would conceivably prevent the spread of such a strain in the wild.

The proteomic approach that we applied to identifying resistance determinants such as HAPT focused intensively on the plasma membrane. This previously uncharacterized proteome required the development of pre-fractionation and proteomic separation techniques as well as the implementation of various bioinformatic approaches. Our characterisation of the plasma membrane sub-proteome (TbPM) is a working definition and as such will no doubt be revised as improvements in HPLC separations, MS instrumentation and novel techniques continue to impact proteomics (Tyers and Mann, 2003). Considering the number of techniques and repeated experiments that were performed, and the proportion of the total predicted trypanosome membrane proteome that

we have observed, we feel that most of the core proteins have been identified. The TbPM therefore represents a great resource for further analysis of the *T. brucei* plasma membrane. A large number of proteins that were annotated as putative were identified in the analysis, enabling more questions to be asked about their function and role. Even without functional data, a brief analysis of the TbPM raises a number of intriguing questions from which new insights into the biology of trypanosomes will no doubt be gained. For example: why were so few amino acid transporters identified? Why were so many nucleoside / nucleobase transporters expressed? On the other hand, there is a huge gap in our understanding, with so many genes completely lacking any annotation. This is a problem common to almost all shotgun proteomic investigations, and at the moment it is the bottleneck in fully utilizing the power of proteomics. Unfortunately these genes often represent dead ends for gaining further insights into an organism's biology, and in the absence of data compelling enough to warrant functional analysis (with far from guaranteed success), will probably remain so for the time being. This is particularly likely in such a divergent organism like *T. brucei*, where so many novel processes have been elucidated and yet where potentially so many more remain.

In addition to the integral membrane proteins identified, there are obviously a large number of accessory / membrane associated proteins. One of the challenges in this approach was defining whether a protein was truly or non-specifically associated with the membrane as a result of the fractionation procedure. Even where proteins have been localised to a specific cellular compartment e.g. the flagellum / cytoskeleton (TbCF), it is always possible that a small proportion of the protein is distributed in other locations. Our definition of TbPM is therefore conservative in its definition as it excludes proteins previously identified in TbCF, however we felt that the TbPM would predominantly define integral or membrane-anchored proteins and therefore decided to apply this subtractive approach.

Having identified HAPT as being absent in the B48 drug resistant strain, it was disappointing not to find any differentially regulated proteins that could be recognized as a drug channel or transporter (as judged by the number of predicted TMDs and/or annotation). However there are a number of reasons that could explain this result. One of the most obvious is simply that the approach failed to identify a change in expression in the spot in which the transporter was present. This could have been due to sample variation, obscurement by juxtaposed spots or even software errors in matching spots across multiple gels. There were also a relatively large number of regulated spots that were not identified by MS, any of which could have contained HAPT. Unfortunately there wasn't enough time to run additional preparative gels to increase protein identification, although this would

almost definitely have improved coverage (Liu *et al.*, 2004). Another possibility is that HAPT is present at such low levels that it would be essentially impossible to identify by MS from these samples unless further enrichment was performed. Indeed this may be crucial to the success of a quantitative membrane approach such as performed here, and combining multiple enrichment techniques would no doubt improve coverage of integral membrane proteins (Zhang *et al.*, 2006). However it is difficult to quantify a technique's ability to enrich for certain proteins as MS runs can vary quite considerably (even with the same sample) making comparative analyses challenging. It was therefore frustrating that our attempts to tag a 'prototypic' membrane transporter e.g. TbAT1, and analyse its selective enrichment by Western blot, which would circumvent MS variability, was unsuccessful.

This result may also have been due to the 'Achilles heel' of proteomics – the general inability to distinguish between subtly different protein species. It is entirely possible that mutations to one or a few key residues that alter substrate affinity in HAPT account for the dramatic reduction in pentamidine accumulation in strain B48. If so, HAPT protein abundance could have remained constant, while any minor protein modifications / mutations, that confer the phenotype would not have affected protein migration, especially in the 16-BAC system. Obviously, no change in protein expression would be identified in this case unless peptide(s) containing the mutations were unambiguously identified by MS or if the mutations affect the stability or location of the protein. Additionally, both the mutated and the original sequence would need to be identified otherwise divergence from the genome (as sequenced from strain 927) would probably be attributed to strain variation.

It is also possible that HAPT is modified into an inactive form e.g. by covalent modifications such as phosphorylation. It is possible that such an event could occur as a result of dramatic changes in the expression of regulatory proteins and the identification of multiple regulated kinases in the DiGE analyses merits their further analysis. In the same way the removal of an accessory protein could also disable the transporter, although the existence of such a protein remains hypothetical.

Despite these potential problems, we feel that because accumulation of the structurally distinct DAPI and pentamidine (albeit both are diamidines) was so markedly reduced, and combined with the highly stable drug resistance phenotype, this strongly suggests the basis for drug resistance is a genetic lesion, resulting in the functional loss of the HAPT transporter activity.

6.3 Future directions

How then can we identify HAPT? DiGE analyses on further enriched plasma membranes i.e. using a combination of techniques, could potentially solve many of the problems highlighted above. In contrast, until I-CAT can utilize all of the MS information in the two peptide sets (i.e. labelled and unlabelled fractions) the likelihood of identifying quantitative differences in membrane protein expression using this technique is poor. Concerns over artefactual differences introduced post-sample isolation can best be answered by the use of a metabolic labelling technique (for review see (Beynon and Pratt, 2005)). For example, SILAC as a form of metabolic labelling has been used for many unicellular organisms, however while trypanosomes are unicellular, to generate a sufficient quantity of plasma membrane protein, they need to be cultured *in vivo*. This is a far harder task and was not feasible to attempt in the time available, although recent reports suggest it may be possible (Wu *et al.*, 2004).

Another approach that could help to identify HAPT is photoaffinity labelling (Ji, 1977), which has most successfully been applied to the analysis of glucose transporters (Holman and Cushman, 1994; Klip *et al.*, 1984). Synthesis of these photoactive (and radiolabelled) compounds can be difficult; however a number of polyamine photoreactive compounds (Felschow *et al.*, 1997) have been developed suggesting that this would be an achievable goal.

If time had allowed we would have wanted to confirm some of the quantitative findings. To do this there are a whole range of techniques that could be employed. Raising antibodies to the target proteins would enable Western blots to be performed to confirm the regulation of protein expression in the two cell lines. Antibodies could also be used to analyse protein cellular localisation. However this approach is very low-throughput and can only really be considered when targeting a few proteins. Considering the number of proteins that were identified in this project, and in the absence of a 'prime' candidate, the time taken to pursue this approach would be prohibitive. Additionally, while it would be useful to confirm the quantitative proteomic findings, the up- or down-regulation of a protein does not prove its involvement in the phenotype. The key test in validating any of the regulated proteins would be to 'knock-down' e.g. by RNAi or conventional gene knockout strategies (e.g. as taken with TbAT1 (Matovu *et al.*, 2003)) to confer drug resistance or 'knock-in' to reverse the resistant phenotype. Again, due to the relatively low-throughput of gene 'knock out' and 'knock-in' strategies, we would have liked to

pursue an RNAi approach to selectively target each of the regulated proteins. This technique has been used extensively in *T. brucei* (Beverley, 2003), and in this case could be used to confer resistance in the parental line (for genes identified as down-regulated in B48) or restore drug sensitivity in B48 (for genes identified as up-regulated in B48). It is highly likely that the drug resistance involves a number of genes and therefore RNAi performed on any one gene would not be expected to confer total resistance or conversely to restore a parental sensitivity phenotype entirely. Where necessary, overexpression of a protein implicated in the resistance phenotype could also be performed, as with TbMRPA (Lüscher *et al.*, 2006; Shahi *et al.*, 2002).

In respect to the potential for some of the regulated proteins to elicit the resistant phenotype through modulation of another protein e.g. phosphorylation, we would have liked to investigate functional protein-protein interactions. For example, running proteins in their native confirmation on 2D blue-native gels, which has been applied to membrane proteins (Brookes *et al.*, 2002; Fandiño *et al.*, 2005), would no doubt reveal some of the integral membrane protein binding partners.

Another more directed approach to further analysing the drug resistance phenotype would be to investigate proteins closely related to TbAT1, i.e. TbAT-like E, A and G (see Figure 4.11), the substrates for which remain unknown (De Koning *et al.*, 2005). Intriguingly, initial experiments involving the expression of TbAT-A in *Xenopus laevis* oocytes or *Saccharomyces cerevisiae* suggested that pentamidine was a substrate for TbAT-A, however a formal identification of HAPT or LAPT through kinetic analysis was not possible in these expression systems (De Koning, unpublished).

6.4 Summary

Within this thesis a working draft of the plasma membrane sub-proteome of bloodstream form *T. brucei* is presented. In addition, a transport-mediated drug resistant strain of *T. brucei* was developed and a number of proteins in this strain were identified as being up- or down-regulated, using conventional techniques and the novel 16-BAC DiGE approach. Despite the clear link with loss of HAPT activity, we were unable to identify the exact mechanism of drug resistance, although a number of interesting regulated proteins were identified. The work that is described in this thesis will no doubt pave the way for such a finding.

Appendices

Appendix I

Gene ID	Gene Annotation
Tb10.70.2650	Elongation factor 2
Tb10.70.2660	Elongation factor 2
Tb10.70.5650	TEF1 elongation factor 1-alpha
Tb10.70.5670	TEF1 elongation factor 1-alpha
Tb10.70.5680	TEF1 elongation factor 1-alpha
Tb10.70.1370	ALD fructose-bisphosphate aldolase, glycosomal, putative
Tb11.02.3210	TIM triosephosphate isomerase
Tb927.3.3270	TbPFK ATP-dependent phosphofructokinase
Tb927.6.4280	GAPDH glyceraldehyde 3-phosphate dehydrogenase, glycosomal
Tb927.6.4300	GAPDH glyceraldehyde 3-phosphate dehydrogenase, glycosomal
Tb10.406.0330	Histone H2B, putative
Tb10.406.0340	Histone H2B, putative
Tb10.406.0350	Histone H2B, putative
Tb10.406.0360	Histone H2B, putative
Tb10.406.0370	Histone H2B, putative
Tb10.406.0380	Histone H2B, putative
Tb10.406.0390	Histone H2B, putative
Tb10.406.0400	Histone H2B, putative
Tb10.406.0410	Histone H2B, putative
Tb10.406.0420	Histone H2B, putative
Tb10.406.0430	Histone H2B, putative
Tb10.406.0440	Histone H2B, putative
Tb10.406.0450	Histone H2B, putative
Tb10.406.0460	Histone H2B, putative
Tb927.5.4170	Histone H4, putative
Tb927.5.4180	Histone H4, putative
Tb927.5.4190	Histone H4, putative
Tb927.5.4200	Histone H4, putative
Tb927.5.4210	Histone H4, putative
Tb927.5.4220	Histone H4, putative
Tb927.5.4230	Histone H4, putative
Tb927.5.4240	Histone H4, putative
Tb927.5.4250	Histone H4, putative
Tb927.5.4260	Histone H4, putative
Tb927.7.2820	Histone H2A, putative
Tb927.7.2830	Histone H2A, putative
Tb927.7.2840	Histone H2A, putative
Tb927.7.2850	Histone H2A, putative
Tb927.7.2860	Histone H2A, putative
Tb927.7.2870	Histone H2A, putative
Tb927.7.2880	Histone H2A, putative
Tb927.7.2890	Histone H2A, putative
Tb927.7.2900	Histone H2A, putative
Tb927.7.2910	Histone H2A, putative
Tb927.7.2920	Histone H2A, putative
Tb927.7.2930	Histone H2A, putative
Tb927.7.2940	Histone H2A, putative
Tb09.211.4511	Kinetoplastid membrane protein KMP-11
Tb09.211.4512	Kinetoplastid membrane protein KMP-11
Tb09.211.4513	Kinetoplastid membrane protein KMP-11

Tb927.3.4290	PFR1 PFRC 73 kDa paraflagellar rod protein
Tb927.3.4300	PFR1 PFRC 73 kDa paraflagellar rod protein
Tb927.3.4310	PFR1 PFRC 73 kDa paraflagellar rod protein
Tb927.3.4320	PFR1 PFRC 73 kDa paraflagellar rod protein
Tb927.3.4330	PFR1 PFRC 73 kDa paraflagellar rod protein
Tb927.8.4970	PFR2 PFR 69 kDa paraflagellar rod protein
Tb927.8.4980	PFR-B PFR-A 69 kDa paraflagellar rod protein
Tb927.8.4990	PFR-B PFR-A 69 kDa paraflagellar rod protein
Tb927.8.5000	PFR-B PFR-A 69 kDa paraflagellar rod protein
Tb927.8.5010	PFR2 PFR 69 kDa paraflagellar rod protein
Tb09.160.0815	60S ribosomal protein L38, putative
Tb09.211.2650	RPL27A; RPL28; RPL29 60S ribosomal protein L27a
Tb09.244.2730	60S ribosomal protein L5, putative
Tb09.244.2740	60S ribosomal protein L5, putative
Tb09.v1.0640	RPL27A; RPL28; RPL29 60S ribosomal protein L27a
Tb10.05.0220	60S ribosomal protein L10a
Tb10.26.0560	60S ribosomal protein L6, putative
Tb10.61.1960	RPS2 40S ribosomal protein S2, putative
Tb10.61.2070	RPS2 40S ribosomal protein S2, putative
Tb10.61.2090	60S ribosomal protein L17, putative
Tb10.6k15.0510	60S ribosomal protein L22
Tb10.6k15.3340	40S ribosomal protein S24E, putative
Tb10.6k15.3350	40S ribosomal protein S24E, putative
Tb10.70.2170	Ubiquitin/ribosomal protein S27a, putative
Tb10.70.3360	40S ribosomal protein S3a, putative
Tb10.70.3370	40S ribosomal protein S3a, putative
Tb10.70.7020	RPS23 40S ribosomal protein S23, putative
Tb10.70.7030	RPS23 40S ribosomal protein S23, putative
Tb10.70.7695	40S ribosomal proteins S11, putative
Tb11.01.1920	60S ribosomal protein L22, putative
Tb11.01.3675	40S ribosomal protein S17, putative
Tb11.01.3676	40S ribosomal protein S17, putative
Tb11.02.0740	60S ribosomal protein L44
Tb11.0290	40s ribosomal protein S14, putative
Tb11.0390	40s ribosomal protein S14, putative
Tb11.46.0001	60S acidic ribosomal subunit protein, putative
Tb11.46.0002	60S acidic ribosomal subunit protein, putative
Tb927.1.3180	40S ribosomal protein S11, putative
Tb927.6.2100	40S ribosomal protein S30, putative
Tb927.6.2110	40S ribosomal protein S30, putative
Tb927.7.1730	60S ribosomal protein L7, putative
Tb927.7.5000	60S ribosomal protein L19, putative
Tb927.7.5020	60S ribosomal protein L19, putative
Tb927.7.5170	60S ribosomal protein L23a
Tb927.8.6030	60S ribosomal protein L12, putative
Tb927.8.6150	40S ribosomal protein S8, putative
Tb927.8.6160	40S ribosomal protein S8, putative
Tb11.01.1350	S-adenosylhomocysteine hydrolase, putative
Tb927.1.2330	Beta tubulin
Tb927.1.2340	Alpha tubulin
Tb927.1.2350	Beta tubulin
Tb927.1.2360	Alpha tubulin
Tb927.1.2370	Beta tubulin
Tb927.1.2380	Alpha tubulin
Tb927.1.2390	Beta tubulin
Tb927.1.2400	Alpha tubulin
Tb927.1.2410	Beta tubulin
Tb10.70.0800	ZFP universal minicircle sequence binding protein (UMSBP), putative

Appendix II to IIX

See attached CD-ROM

Appendix IX

Localisation	Accession Number	Gene Annotation	Identified in		Protein score	gCAI Value (% Rank)	Number of predicted TMD's	Protein Mwt	Sequence Coverage (%)	Number of peptides identified	Evidence
			PM	FP							
Cytoskeleton	Tb927.1.2360	Alpha tubulin	+	+	2985	99.9	0	49756	63	154	TbFP
Flagellum	Tb927.2.5660	Adenylate kinase, putative	+	+	513	85	0	29492	43	12	TbFP
Flagellum	Tb10.70.7330	Adenylate kinase, putative	+	+	341	34.8	0	29860	36	7	TbFP
Flagellum	Tb927.5.3970	Adenylate kinase, putative	+	+	102	54.9	0	30082	20	4	TbFP
Flagellum	Tb10.70.1880	ADP-ribosylation factor, putative	+	+	42	81.3	0	30594	8	2	TbFP
Flagellum	Tb927.6.4990	ATP synthase, epsilon chain, putative	+	+	163	82	0	20362	28	4	TbFP
Flagellum	Tb927.1.2330	Beta tubulin	+	+	3983	99.9	0	49672	54	289	TbFP
Flagellum	Tb927.6.2720	Calcium-binding protein, putative	+	+	168	58.3	0	42087	12	4	TbFP
Flagellum	Tb11.01.4621	Calmodulin	+	+	555	98.5	0	16828	81	13	TbFP
Flagellum	Tb09.160.4520	Calmodulin, putative	+	+	194	93.4	0	17888	33	8	TbFP
Flagellum	Tb11.02.5800	Calmodulin, putative	+	+	95	72.7	0	17946	12	2	TbFP
Flagellum	Tb09.211.2540	Calmodulin-like protein, putative; EF hand containing protein	+	+	125	83.1	0	15907	24	3	TbFP
Flagellum	Tb927.1.2100	Calpain-like cysteine peptidase, putative; cysteine peptidase, Clan CA, f	+	+	36	91.5	0	127046	3	3	TbFP
Flagellum	Tb927.1.2230	Calpain-like protein fragment, putative	+	+	224	28.6	0	13698	32	5	TbFP
Flagellum	Tb927.1.2260	Calpain-like protein fragment, putative	+	+	212	73.3	0	14192	31	6	TbFP
Flagellum	Tb927.4.3950	CAP5.5 cytoskeleton-associated protein CAP5.5, putative	+	+	116	88.6	0	95734	6	5	TbFP
Flagellum	Tb927.4.2260	Centrin, putative	+	+	81	72.6	0	21162	15	3	TbFP
Flagellum	Tb11.01.5470	Centrin, putative	+	+	91	38.1	0	20812	16	2	TbFP
Flagellum	Tb11.01.6780	Chaperone protein DNAJ, putative; heat shock protein-like protein, putati	+	+	408	89.8	0	35526	35	9	TbFP
Flagellum	Tb10.6k15.3150	Chromatin binding protein, putative	+	+	40	69.2	0	52254	14	5	TbFP
Flagellum	Tb927.8.6910	cyclophilin, putative	+	+	87	30	0	21668	12	2	TbFP
Flagellum	Tb927.3.4760	Dynamin, putative	+	+	46	86.9	0	73103	8	4	TbFP
Flagellum	Tb927.7.820	Dynein arm light chain, axonemal, putative	+	+	58	63.7	0	34629	22	6	TbFP
Flagellum	Tb11.02.3200	Dynein arm light chain, axonemal, putative	+	+	157	85.8	0	27582	16	4	TbFP
Flagellum	Tb11.01.0390	Dynein heavy chain, putative	+	+	1224	90.5	0	477699	15	54	TbFP
Flagellum	Tb927.2.5270	Dynein heavy chain, putative	+	+	481	89.3	0	489199	7	29	TbFP
Flagellum	Tb927.3.930	Dynein heavy chain, putative	+	+	1640	95.2	0	534556	19	75	TbFP
Flagellum	Tb927.7.920	Dynein heavy chain, putative	+	+	1152	75.7	0	471780	16	48	TbFP
Flagellum	Tb927.8.3250	Dynein heavy chain, putative	+	+	471	93.8	0	541071	6	26	TbFP
Flagellum	Tb10.70.1720	Dynein heavy chain, putative	+	+	796	86.8	0	478621	11	38	TbFP
Flagellum	Tb11.01.3010	Dynein heavy chain, putative	+	+	1017	90.6	0	483604	13	45	TbFP
Flagellum	Tb11.02.0760	Dynein heavy chain, putative	+	+	2025	92.4	0	535352	12	22	TbFP
Flagellum	Tb927.2.4060	Dynein intermediate chain, putative	+	+	141	84.5	0	95205	7	6	TbFP
Flagellum	Tb11.02.2640	Dynein intermediate chain, putative	+	+	284	89.5	0	74103	14	8	TbFP
Flagellum	Tb927.4.5370	Dynein light chain 2B, cytoplasmic, putative; predicted dynein modulator;	+	+	134	25.7	0	13184	43	4	TbFP
Flagellum	Tb10.70.0090	Dynein light chain, putative	+	+	60	93.5	0	14149	28	2	TbFP
Flagellum	Tb11.02.3390	Dynein light chain, putative	+	+	73	67.9	0	24116	13	2	TbFP
Flagellum	Tb11.50.0007	Dynein light chain, putative	+	+	222	94	0	10479	38	4	TbFP
Flagellum	Tb11.01.5910	Dynein, putative	+	+	114	89.4	0	68758	5	2	TbFP
Flagellum	Tb927.5.1730	Ecotin, putative	+	+	85	43.3	0	19615	13	3	TbFP
Flagellum	Tb927.5.1880	Ecotin, putative	+	+	203	92.9	0	17913	29	3	TbFP
Flagellum	Tb09.211.1370	Glyceraldehyde-3-phosphate dehydrogenase, putative	+	+	301	62.9	0	39234	27	6	TbFP
Flagellum	Tb11.01.3110	Heat shock protein 70	+	+	488	99.3	0	75719	20	11	TbFP
Flagellum	Tb927.7.3440	I/6 autoantigen	+	+	365	16.1	0	27261	35	9	TbFP
Flagellum	Tb927.7.3450	I/6 autoantigen	+	+	89	59.5	0	21571	22	4	TbFP
Flagellum	Tb927.7.6290	Kinesin, putative	+	+	321	62.3	0	97348	17	13	TbFP
Flagellum	Tb927.8.2630	Kinesin, putative	+	+	754	72.2	0	85480	28	20	TbFP
Flagellum	Tb10.61.1020	Kinesin, putative	+	+	67	35.2	0	124515	3	3	TbFP
Flagellum	Tb927.7.1920	Leucine-rich repeat protein (LRRP), putative	+	+	173	46.8	0	60773	11	5	TbFP
Flagellum	Tb10.406.0560	Microtubule-associated protein, putative	+	+	109	93.4	0	237349	3	8	TbFP
Flagellum	Tb10.406.0650	Microtubule-associated protein, putative	+	+	109	93.3	0	254488	3	8	TbFP
Flagellum	Tb09.211.2560	Nucleoside diphosphate kinase, putative	+	+	91	27.9	0	37767	11	2	TbFP
Flagellum	Tb927.4.1720	Nucleoside diphosphate kinase, putative	+	+	160	80.8	0	39380	15	4	TbFP
Flagellum	Tb927.2.4230	NUP-1 protein, putative	+	+	620	45.4	0	407284	9	31	TbFP
Flagellum	Tb09.160.2360	Poly(A) export protein, putative	+	+	43	24.8	0	38327	9	4	TbFP
Flagellum	Tb927.7.4020	PP2C protein phosphatase 2C, putative	+	+	208	53.8	0	78830	11	6	TbFP

Flagellum	Tb927.8.5230	PPIase cyclophilin-type peptidyl-prolyl cis-trans isomerase, putative	+	+	85	83.2	0	26784	9	2	TbFP
Flagellum	Tb11.47.0034	Radial spoke protein 3, putative; radial spoke 3 protein, putative	+	+	177	77.3	0	39353	16	5	TbFP
Flagellum	Tb927.2.340	Retrotransposon hot spot (RHS) protein, putative	+	+	40	21.4	0	98350	2	2	TbFP
Flagellum	Tb927.5.520	Stomatin-like protein, putative	+	+	59	81.1	0	56401	2	1	TbFP
Flagellum	Tb10.70.5670	TEF1 elongation factor 1-alpha; EF-1-alpha	+	+	291	99.9	0	49474	18	6	TbFP
Flagellum	Tb09.160.4280	TRYP1 tryparedoxin peroxidase	+	+	249	97.2	0	22752	43	8	TbFP
Flagellum	Tb927.8.1990	TRYP2 tryparedoxin peroxidase	+	+	356	93.4	0	25786	31	8	TbFP
Flagellum	Tb10.70.0480	Trypanin	+	+	106	60.2	0	54119	6	3	TbFP
Flagellum	Tb09.244.2800	Trypanin-related protein, putative	+	+	258	80.8	0	54303	14	6	TbFP
Flagellum	Tb11.02.1380	TRYPARP actin, putative	+	+	58	72.4	0	47529	12	4	TbFP
Cytoplasm	Tb09.211.0540	FBPase fructose-1,6-bisphosphate, cytosolic	+	-	101	90.9	0	38681	11	2	GeneDB
Cytoplasm	Tb10.70.6660	Hypoxanthine-guanine phosphoribosyltransferase, putative	+	-	199	76.8	0	26515	34	6	GeneDB
Cytoplasm	Tb09.160.3630	PDE cAMP-specific phosphodiesterase	+	+	330	84.6	0	104065	12	13	GeneDB
Cytoplasm	Tb10.61.2680	PYK1 pyruvate kinase 1	+	-	40	98.9	0	39170	3	1	GeneDB
Cytoplasm	Tb10.70.5650	TEF1 elongation factor 1-alpha	+	-	291	99.9	0	49474	18	6	GeneDB
Cytoplasm	Tb10.70.5680	TEF1 elongation factor 1-alpha; EF-1-alpha	+	-	122	100	0	38126	7	3	GeneDB
Cytoplasm	Tb09.160.4250	TRYP1 tryparedoxin peroxidase	+	-	249	97.2	0	22752	43	8	GeneDB
Cytoskeleton	Tb09.211.0620	Actin A	+	-	209	96.2	0	42154	18	4	GeneDB
Cytoskeleton	Tb927.1.2340	Alpha tubulin	+	-	2985	99.9	0	49756	63	154	GeneDB
Cytoskeleton	Tb927.7.3410	Centrin, putative	+	-	133	92.8	0	16542	49	4	GeneDB
Flagellum	Tb927.1.2670	Axoneme central apparatus protein, putative; importin alpha-1 subunit, pu	+	+	324	72.5	0	56091	28	9	GeneDB
Flagellum	Tb11.01.4622	Calmodulin	+	-	555	98.5	0	16828	81	13	GeneDB
Flagellum	Tb11.01.4623	Calmodulin	+	-	555	98.5	0	16828	81	13	GeneDB
Flagellum	Tb11.01.4624	Calmodulin	+	-	555	98.6	0	16828	81	13	GeneDB
Flagellum	Tb927.8.6070	TBBC basal body component	+	-	390	57.3	0	164308	13	17	GeneDB
Glycosome	Tb927.3.3270	TbPFK ATP-dependent phosphofructokinase	+	-	1072	99.3	0	53997	42	28	GeneDB
Glycosome	Tb11.02.3210	TIM triosephosphate isomerase	+	-	318	99	0	26973	23	5	GeneDB
Intracellular	Tb927.8.8330	Calpain, putative	+	-	419	87.1	0	99535	15	9	GeneDB
Intracellular	Tb11.02.2210	PKA-R protein kinase A regulatory subunit	+	-	494	83.3	0	57212	24	12	GeneDB
Mitochondrion	Tb09.211.1320	AMP deaminase, putative	+	-	192	73.6	0	164323	5	7	GeneDB
Nucleus	Tb10.61.1920	Fibrillarin, putative	+	-	33	82.7	0	31653	7	2	GeneDB
Nucleus	Tb11.02.5250	H2BVAR histone H2B variant, putative	+	-	62	93.2	0	15834	14	2	GeneDB
Nucleus	Tb10.61.1090	h3vaR histone H3 variant	+	-	97	35.2	0	16126	12	2	GeneDB
Nucleus	Tb927.2.450	Retrotransposon hot spot (RHS) protein, putative	+	-	67	20.2	0	98336	5	4	GeneDB
Nucleus	Tb09.160.4090	TOP2 DNA topoisomerase II	+	-	255	89.2	0	138218	9	9	GeneDB
Unknown	Tb927.1.4490	Acetyltransferase, putative	+	-	170	89.5	0	19726	20	4	GeneDB
Unknown	Tb09.160.5250	Adenosine monophosphate deaminase, putative; AMP deaminase, putative	+	-	199	76.3	0	191075	5	8	GeneDB
Unknown	Tb09.211.4460	ADP-ribosylation factor, putative	+	-	147	97.6	0	20753	15	3	GeneDB
Unknown	Tb11.01.7390	AMP deaminase, putative	+	-	165	68.4	0	164131	4	6	GeneDB
Unknown	Tb11.01.1550	Calmodulin, putative	+	-	124	28.5	0	18513	22	3	GeneDB
Unknown	Tb927.5.800	Casein kinase, putative	+	-	113	74.5	0	38522	12	3	GeneDB
Unknown	Tb09.211.2150	Poly(A)-binding protein 1; PABP2	+	-	218	98.5	0	62335	20	9	GeneDB
Unknown	Tb11.01.1680	Polyubiquitin, putative	+	-	146	98	0	76556	6	7	GeneDB
Unknown	Tb927.8.890	Small GTP-binding protein Rab1, putative	+	-	315	95.7	0	22929	37	7	GeneDB
Unknown	Tb927.4.2070	Antigenic protein, putative	+	-	413	24.9	0	513177	3	15	GeneDB
Unknown	Tb10.70.2650	Elongation factor 2	+	-	198	99.4	0	95300	11	7	GeneDB
Unknown	Tb10.70.2660	Elongation factor 2	+	-	198	99.4	0	95300	11	7	GeneDB
Unknown	Tb11.01.4390	Leucine-rich repeat protein (LRRP), putative	+	-	220	67.1	0	70844	12	6	GeneDB
Unknown	Tb927.7.4570	Nucleoside hydrolase, putative	+	-	133	80.3	0	39708	14	3	GeneDB
Unknown	Tb10.61.1900	Protein kinase, putative	+	-	197	76	0	71001	13	7	GeneDB
Unknown	Tb11.01.3320	Trichohyalin, putative	+	-	56	44.9	0	78920	11	8	GeneDB
Cytoskeleton	Tb09.211.0630	Actin A	+	+	209	96.3	0	42154	18	4	Gene name
Cytoskeleton	Tb09.160.3960	Actin, putative	+	+	109	78.5	0	47347	5	2	Gene name
Cytoskeleton	Tb11.01.1870	Actin-like protein, putative	+	+	60	90.9	0	43969	2	1	Gene name
Cytoskeleton	Tb927.1.2380	Alpha tubulin	+	-	2985	99.9	0	49756	63	154	Gene name
Cytoskeleton	Tb927.1.2400	Alpha tubulin	+	-	2985	99.9	0	49756	63	154	Gene name
Cytoskeleton	Tb927.1.2350	Beta tubulin	+	-	3983	99.9	0	49672	54	289	Gene name
Cytoskeleton	Tb927.1.2370	Beta tubulin	+	-	3983	99.9	0	49672	54	289	Gene name
Cytoskeleton	Tb927.1.2390	Beta tubulin	+	-	3983	99.9	0	49672	54	289	Gene name
Cytoskeleton	Tb927.1.2410	Beta tubulin, pseudogene	+	-	1785	99.9	0	33050	53	144	Gene name
Cytoskeleton	Tb10.6k15.1830	Centrin, putative	+	-	191	83.3	0	18790	46	5	Gene name
Cytoskeleton	Tb927.4.870	Dynein heavy chain, putative	+	-	1212	91.4	0	513145	14	49	Gene name
Cytoskeleton	Tb927.8.6950	Dynein light chain 2B, cytoplasmic, putative	+	-	134	25.6	0	13183	43	4	Gene name

Cytoskeleton	Tb10.389.0350	Dynein light chain 2B, cytoplasmic, putative; predicted dynein modulator; roadblock/LC7 family member	+	-	136	47	0	12818	36	4	Gene name
Cytoskeleton	Tb10.70.2520	Dynein light chain, putative	+	-	87	89.1	0	13191	18	2	Gene name
Cytoskeleton	Tb11.02.5620	Dynein light chain, putative	+	-	57	20.5	0	13183	13	1	Gene name
Cytoskeleton	Tb11.03.0815	Dynein light chain, putative	+	-	89	95.2	0	10477	32	2	Gene name
Cytoskeleton	Tb11.0845	Dynein light chain, putative	+	-	222	94	0	10479	38	4	Gene name
Cytoskeleton	Tb09.211.4920	Dynein-associated protein, putative	+	-	84	8.5	0	11237	43	2	Gene name
Cytoskeleton	Tb09.160.1200	GB4 mitotubule-associated protein Gb4, putative; dynein heavy chain, cytosolic, putative	+	-	60	55.8	0	927802	0	3	Gene name
Cytoskeleton	Tb11.02.0790	Kinesin, putative; MCAK-like kinesin, putative	+	-	231	76.3	0	80648	16	9	Gene name
Cytoskeleton	Tb10.v4.0052	Microtubule-associated protein 2	+	-	298	94.7	0	561328	1	11	Gene name
Intracellular	Tb10.70.1670	40S ribosomal protein S10, putative	+	-	50	98	0	19331	14	2	Gene name
Intracellular	Tb927.2.5910	40S ribosomal protein S13, putative	+	-	41	91.8	0	17412	5	1	Gene name
Intracellular	Tb10.61.1390	40S ribosomal protein S13, putative	+	-	41	94.1	0	17412	5	1	Gene name
Intracellular	Tb927.7.2340	40S ribosomal protein S15, putative	+	-	57	98.8	0	17463	10	1	Gene name
Intracellular	Tb927.7.2370	40S ribosomal protein S15, putative	+	-	57	97.5	0	20083	9	1	Gene name
Intracellular	Tb927.7.1040	40S ribosomal protein S16, putative	+	-	201	98.9	0	17101	39	6	Gene name
Intracellular	Tb927.7.1050	40S ribosomal protein S16, putative	+	-	201	98.9	0	17101	39	6	Gene name
Intracellular	Tb11.01.3675	40S ribosomal protein S17, putative	+	-	55	99.7	0	16212	9	1	Gene name
Intracellular	Tb11.01.3676	40S ribosomal protein S17, putative	+	-	55	99.7	0	16212	9	1	Gene name
Intracellular	Tb10.70.1740	40S ribosomal protein S18, putative	+	-	178	98.3	0	17614	27	4	Gene name
Intracellular	Tb11.01.1475	40S ribosomal protein S27, putative	+	-	178	98.8	0	9966	46	3	Gene name
Intracellular	Tb10.70.3360	40S ribosomal protein S3a, putative	+	-	135	99.2	0	29632	9	3	Gene name
Intracellular	Tb11.02.1085	40s ribosomal protein S4, putative	+	-	62	97.7	0	30739	10	3	Gene name
Intracellular	Tb11.02.1090	40S ribosomal protein S4, putative	+	-	62	97.5	0	30739	10	3	Gene name
Intracellular	Tb09.244.2630	40S ribosomal protein S6, putative	+	-	40	98	0	28516	8	2	Gene name
Intracellular	Tb10.100.0080	40S ribosomal protein S6, putative	+	-	40	96.9	0	28480	8	2	Gene name
Intracellular	Tb927.8.1110	40S ribosomal protein S9, putative	+	-	93	98.9	0	22209	14	3	Gene name
Intracellular	Tb10.70.1380	40S ribosomal protein S9, putative	+	-	93	98.8	0	22209	14	3	Gene name
Intracellular	Tb11.01.2560	40S ribosomal protein SA, putative	+	-	112	98.6	0	27820	13	3	Gene name
Intracellular	Tb11.01.2680	40S ribosomal protein SA, putative	+	-	112	98.3	0	31562	11	3	Gene name
Intracellular	Tb10.70.7695	40S ribosomal proteins S11, putative	+	-	139	99.4	0	20303	17	3	Gene name
Intracellular	Tb10.05.0220	60S ribosomal protein L10a	+	-	39	99	0	25037	7	2	Gene name
Intracellular	Tb927.3.3320	60S ribosomal protein L13, putative	+	-	71	98.7	0	25412	18	3	Gene name
Intracellular	Tb10.61.2090	60S ribosomal protein L17, putative	+	-	38	99.3	0	19093	14	2	Gene name
Intracellular	Tb11.02.2430	60S ribosomal protein L17, putative	+	-	38	98.8	0	19093	14	2	Gene name
Intracellular	Tb09.211.2630	60S ribosomal protein L23, putative	+	-	82	98.8	0	15067	23	3	Gene name
Intracellular	Tb09.211.2640	60S ribosomal protein L23, putative	+	-	82	98.7	0	15067	23	3	Gene name
Intracellular	Tb927.7.5170	60S ribosomal protein L23a	+	-	87	99.2	0	18159	23	3	Gene name
Intracellular	Tb927.7.5180	60S ribosomal protein L23a, putative	+	-	85	97.5	0	24754	17	3	Gene name
Intracellular	Tb09.211.4850	60S ribosomal protein L26, putative	+	-	40	98.1	0	16615	6	1	Gene name
Intracellular	Tb927.8.6180	60S ribosomal protein L26, putative	+	-	40	97.1	0	16585	6	1	Gene name
Intracellular	Tb927.3.5050	60S ribosomal protein L4	+	-	60	98.5	0	41944	8	2	Gene name
Intracellular	Tb927.8.1330	60S ribosomal protein L7a, putative	+	-	41	98.5	0	30860	16	6	Gene name
Intracellular	Tb927.8.1340	60S ribosomal protein L7a, putative	+	-	41	98.5	0	30860	16	6	Gene name
Intracellular	Tb10.70.5950	Calpain, putative; cysteine peptidase, Clan CA, family C2, putative	+	-	437	82.9	0	180312	11	15	Gene name
Intracellular	Tb11.47.0036	Calpain, putative; cysteine peptidase, Clan CA, family C2, putative	+	-	453	48.2	0	160387	19	34	Gene name
Intracellular	Tb927.7.4060	Calpain-like cysteine peptidase, putative	+	-	140	96.3	0	12979	36	5	Gene name
Intracellular	Tb11.01.5800	Calpain-like cysteine peptidase, putative; cysteine peptidase, Clan CA, f	+	-	350	44.8	0	166949	10	14	Gene name

Intracellular	Tb11.57.0008	Calpain-like protein, putative; cytoskeleton assocociated protein, putati	+	-	546	89.5	0	659211	4	35	Gene name
Intracellular	Tb09.160.3670	NHP2 ribosomal protein S6, putative; NHP2/RS6-like protein	+	-	102	95.8	0	13620	19	3	Gene name
Intracellular	Tb09.160.2550	Ribosomal protein S7, putative	+	-	40	94.6	0	23883	9	2	Gene name
Intracellular	Tb11.01.1470	RPL10A 60S ribosomal protein L10a, putative	+	-	39	98.9	0	25037	7	2	Gene name
Intracellular	Tb10.70.3160	RPL30 60S ribosomal protein L30	+	-	56	98	0	12569	8	1	Gene name
Intracellular	Tb10.70.3170	RPL30 60S ribosomal protein L30	+	-	56	97.9	0	11733	9	1	Gene name
Intracellular	Tb927.6.4980	RPS14 40S ribosomal protein S14	+	-	234	98.9	0	15674	19	4	Gene name
Intracellular	Tb11.0290	RPS14 40s ribosomal protein S14, putative	+	-	234	99.1	0	15674	19	4	Gene name
Intracellular	Tb11.0390	RPS14 40s ribosomal protein S14, putative	+	-	234	99.1	0	15674	19	4	Gene name
Intracellular	Tb10.70.1730	RPS18 40S ribosomal protein S18, putative	+	-	178	98.1	0	17614	27	4	Gene name
Intracellular	Tb10.61.1960	RPS2 40S ribosomal protein S2, putative	+	-	168	99.4	0	28795	26	5	Gene name
Intracellular	Tb10.61.2070	RPS2 40S ribosomal protein S2, putative	+	-	168	99.4	0	28795	26	5	Gene name
Intracellular	Tb09.160.4450	RPS3 40S ribosomal protein S3, putative	+	-	144	97.3	0	30724	21	4	Gene name
Intracellular	Tb10.26.0370	RPS3 40S ribosomal protein S3, putative	+	-	144	98.4	0	24372	26	4	Gene name
Plasma Membrane	Tb11.v4.0035	Variant surface glycoprotein (VSG), putative	+	-	38	54	0	51581	1	6	Gene name
Plasma Membrane	Tb10.v4.0185	Variant surface glycoprotein (VSG, pseudogene), putative	+	-	42	20.1	0	55391	4	2	Gene name
Plasma Membrane	Tb11.1820	Variant surface glycoprotein (VSG, pseudogene), putative	+	-	48	13.8	0	48453	6	3	Gene name
Plasma Membrane	Tb927.3.300	Variant surface glycoprotein (VSG, pseudogene), putative	+	-	58	27.9	0	58431	15	6	Gene name
Plasma Membrane	Tb927.5.4900	Variant surface glycoprotein (VSG, pseudogene), putative	+	-	36	6.4	0	56202	6	5	Gene name
Plasma Membrane	Tb09.244.0790	Variant surface glycoprotein (VSG, pseudogene), putative	+	-	50	17	0	51570	5	2	Gene name
Nucleus	Tb927.4.3810	DNA-directed RNA polymerase II subunit 2, putative	+	-	36	87.5	0	136102	4	5	Gene name
Nucleus	Tb927.7.6360	Histone H2A, putative	+	-	101	97.3	0	18769	17	4	Gene name
Nucleus	Tb927.7.2820	Histone H2A, putative	+	-	335	99.7	0	14203	43	8	Gene name
Nucleus	Tb927.7.2830	Histone H2A, putative	+	-	335	99.6	0	14203	43	8	Gene name
Nucleus	Tb927.7.2840	Histone H2A, putative	+	-	335	99.6	0	14203	43	8	Gene name
Nucleus	Tb927.7.2850	Histone H2A, putative	+	-	335	99.6	0	14203	43	8	Gene name
Nucleus	Tb927.7.2860	Histone H2A, putative	+	-	335	99.7	0	14203	43	8	Gene name
Nucleus	Tb927.7.2870	Histone H2A, putative	+	-	335	99.7	0	14203	43	8	Gene name
Nucleus	Tb927.7.2880	Histone H2A, putative	+	-	335	99.6	0	14203	43	8	Gene name
Nucleus	Tb927.7.2890	Histone H2A, putative	+	-	335	99.6	0	14203	43	8	Gene name
Nucleus	Tb927.7.2900	Histone H2A, putative	+	-	335	99.6	0	14203	43	8	Gene name
Nucleus	Tb927.7.2910	Histone H2A, putative	+	-	335	99.6	0	14203	43	8	Gene name
Nucleus	Tb927.7.2920	Histone H2A, putative	+	-	335	99.6	0	14203	43	8	Gene name
Nucleus	Tb927.7.2930	Histone H2A, putative	+	-	335	99.6	0	14203	43	8	Gene name
Nucleus	Tb927.7.2940	Histone H2A, putative	+	-	335	99.7	0	14203	43	8	Gene name
Nucleus	Tb10.406.0330	Histone H2B, putative	+	-	291	99.8	0	12562	36	11	Gene name
Nucleus	Tb10.406.0340	Histone H2B, putative	+	-	291	99.7	0	12570	36	11	Gene name
Nucleus	Tb10.406.0350	Histone H2B, putative	+	-	291	99.8	0	12562	36	11	Gene name
Nucleus	Tb10.406.0360	Histone H2B, putative	+	-	291	99.7	0	12562	36	11	Gene name
Nucleus	Tb10.406.0370	Histone H2B, putative	+	-	291	99.8	0	12562	36	11	Gene name
Nucleus	Tb10.406.0380	Histone H2B, putative	+	-	291	99.8	0	12562	36	11	Gene name
Nucleus	Tb10.406.0390	Histone H2B, putative	+	-	291	99.8	0	12589	36	11	Gene name
Nucleus	Tb10.406.0400	Histone H2B, putative	+	-	291	99.8	0	12562	36	11	Gene name
Nucleus	Tb10.406.0410	Histone H2B, putative	+	-	291	99.8	0	12562	36	11	Gene name
Nucleus	Tb10.406.0420	Histone H2B, putative	+	-	291	99.8	0	12562	36	11	Gene name
Nucleus	Tb10.406.0430	Histone H2B, putative	+	-	291	99.8	0	12562	36	11	Gene name
Nucleus	Tb10.406.0440	Histone H2B, putative	+	-	291	99.7	0	12562	36	11	Gene name
Nucleus	Tb10.406.0450	Histone H2B, putative	+	-	291	99.8	0	12562	36	11	Gene name
Nucleus	Tb10.406.0460	Histone H2B, putative	+	-	291	99.8	0	12562	36	11	Gene name
Nucleus	Tb927.1.2430	Histone H3, putative	+	-	130	97.4	0	14855	25	3	Gene name
Nucleus	Tb927.1.2450	Histone H3, putative	+	-	130	97.4	0	14855	25	3	Gene name
Nucleus	Tb927.1.2470	Histone H3, putative	+	-	130	97.4	0	14855	25	3	Gene name
Nucleus	Tb927.1.2490	Histone H3, putative	+	-	130	97.1	0	14855	25	3	Gene name
Nucleus	Tb927.1.2510	Histone H3, putative	+	-	130	97.4	0	14855	25	3	Gene name
Nucleus	Tb927.1.2530	Histone H3, putative	+	-	130	97.4	0	14855	25	3	Gene name
Nucleus	Tb927.1.2550	Histone H3, putative	+	-	130	97.4	0	14855	25	3	Gene name
Nucleus	Tb927.2.2670	Histone H4, putative	+	-	280	95.3	0	11246	45	6	Gene name
Nucleus	Tb927.5.4170	Histone H4, putative	+	-	390	99.1	0	11135	47	10	Gene name
Nucleus	Tb927.5.4180	Histone H4, putative	+	-	390	99.2	0	11135	47	10	Gene name
Nucleus	Tb927.5.4190	Histone H4, putative	+	-	390	99.2	0	11135	47	10	Gene name
Nucleus	Tb927.5.4200	Histone H4, putative	+	-	390	99.1	0	11135	47	10	Gene name
Nucleus	Tb927.5.4210	Histone H4, putative	+	-	390	99.2	0	11135	47	10	Gene name
Nucleus	Tb927.5.4220	Histone H4, putative	+	-	390	99.2	0	11135	47	10	Gene name
Nucleus	Tb927.5.4230	Histone H4, putative	+	-	390	99.2	0	11135	47	10	Gene name
Nucleus	Tb927.5.4240	Histone H4, putative	+	-	390	99.1	0	11135	47	10	Gene name
Nucleus	Tb927.5.4250	Histone H4, putative	+	-	390	99.1	0	11135	47	10	Gene name
Nucleus	Tb927.5.4260	Histone H4, putative	+	-	390	99.1	0	11135	47	10	Gene name
Nucleus	Tb927.2.470	Retrotransposon hot spot (RHS) protein, putative	+	-	76	16.7	0	98769	4	4	Gene name

Nucleus	Tb927.1.120	Retrotransposon hot spot (RHS) protein, putative; retrotransposon hot spot	+	-	60	12.8	0	98534	4	3	Gene name
Nucleus	Tb927.2.1120	Retrotransposon hot spot protein (RHS, pseudogene), putative	+	-	31	19.4	0	81543	4	2	Gene name
Unknown	Tb09.211.4470	ADP-ribosylation factor, putative	+	-	147	97.4	0	20725	15	3	Gene name
Unknown	Tb09.211.4480	ADP-ribosylation factor, putative	+	-	147	97.2	0	20725	15	3	Gene name
Unknown	Tb09.211.4490	ADP-ribosylation factor, putative	+	-	147	97.4	0	20725	15	3	Gene name
Unknown	Tb927.1.4300	Chaperone protein DNAJ, putative	+	-	33	58.1	0	55142	7	2	Gene name
Unknown	Tb927.3.5310	Hypothetical protein, conserved	+	-	31	79.2	0	243828	1	2	Gene name
Unknown	Tb09.211.3140	Hypothetical protein, conserved	+	-	32	71.7	0	110655	1	2	Gene name
Unknown	Tb927.3.3220	Hypothetical protein, conserved	+	-	32	49.6	0	102479	2	3	Gene name
Unknown	Tb927.4.2880	Hypothetical protein, conserved	+	-	32	80.1	0	225449	0	4	Gene name
Unknown	Tb10.70.4860	Hypothetical protein, conserved	+	-	33	65.7	0	115410	2	2	Gene name
Unknown	Tb10.70.3460	Hypothetical protein, conserved	+	-	34	56.4	0	236985	0	2	Gene name
Unknown	Tb927.4.2820	Hypothetical protein, conserved	+	+	34	58.6	0	109656	2	2	Gene name
Unknown	Tb927.8.6250	Hypothetical protein, conserved	+	-	34	24.1	0	76739	4	2	Gene name
Unknown	Tb927.2.2650	Hypothetical protein, conserved	+	-	35	82.3	0	367468	0	3	Gene name
Unknown	Tb09.160.2860	Hypothetical protein, conserved	+	-	37	48.6	0	39900	3	1	Gene name
Unknown	Tb10.61.0620	Hypothetical protein, conserved	+	-	37	50.4	0	377010	0	1	Gene name
Unknown	Tb11.01.5240	Hypothetical protein, conserved	+	-	37	58.2	0	183500	0	1	Gene name
Unknown	Tb927.3.5290	Hypothetical protein, conserved	+	+	37	64	0	105146	4	5	Gene name
Unknown	Tb09.211.1800	Hypothetical protein, conserved	+	-	38	88.2	0	53601	4	2	Gene name
Unknown	Tb927.4.4040	Hypothetical protein, conserved	+	-	38	48.9	0	30483	15	3	Gene name
Unknown	Tb927.7.5690	Hypothetical protein, conserved	+	+	38	33.5	0	68791	4	5	Gene name
Unknown	Tb11.01.3520	Hypothetical protein, conserved	+	+	39	73.6	0	111275	1	1	Gene name
Unknown	Tb10.61.2190	Hypothetical protein, conserved	+	-	41	46.3	0	53016	3	1	Gene name
Unknown	Tb11.02.4540	Hypothetical protein, conserved; predicted tetratricopeptide repeat (TPR)	+	-	41	31.4	0	45804	7	2	Gene name
Unknown	Tb927.7.2350	Hypothetical protein, conserved	+	-	41	54.7	0	116245	2	3	Gene name
Unknown	Tb927.8.2440	Hypothetical protein, conserved	+	-	42	60.5	0	23484	4	1	Gene name
Unknown	Tb11.02.0460	Hypothetical protein, conserved	+	-	44	87.3	0	89897	5	3	Gene name
Unknown	Tb10.70.1570	Hypothetical protein, conserved	+	-	45	57.8	0	107854	1	1	Gene name
Unknown	Tb09.211.3240	Hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative	+	+	49	48.2	0	32738	5	1	Gene name
Unknown	Tb10.61.1550	Hypothetical protein, conserved	+	+	49	86.8	0	47972	11	3	Gene name
Unknown	Tb10.6k15.0480	Hypothetical protein, conserved	+	+	50	88.2	0	17297	13	2	Gene name
Unknown	Tb927.4.620	Hypothetical protein, conserved	+	-	51	53.3	0	279443	1	5	Gene name
Unknown	Tb11.52.0004	Hypothetical protein, conserved	+	+	52	84.3	0	35207	5	2	Gene name
Unknown	Tb927.4.4600	Hypothetical protein, conserved	+	-	53	69.8	0	31457	20	4	Gene name
Unknown	Tb927.8.3990	Hypothetical protein, conserved	+	-	53	32.4	0	33042	3	1	Gene name
Unknown	Tb11.01.1850	Hypothetical protein, conserved	+	+	54	71.1	0	46965	4	2	Gene name
Unknown	Tb11.02.4120	Hypothetical protein, conserved	+	+	55	94.5	0	27640	8	1	Gene name
Unknown	Tb11.01.5260	Hypothetical protein, conserved	+	+	56	26.3	0	22306	11	2	Gene name
Unknown	Tb927.8.8000	Hypothetical protein, conserved	+	-	57	52.1	0	61767	5	2	Gene name
Unknown	Tb10.70.3090	Hypothetical protein, conserved	+	-	58	41.5	0	293469	2	7	Gene name
Unknown	Tb927.5.2330	Hypothetical protein, conserved	+	-	58	42.4	0	492356	1	4	Gene name
Unknown	Tb927.7.4220	Hypothetical protein, conserved	+	-	59	83	0	122027	7	8	Gene name
Unknown	Tb11.02.2350	Hypothetical protein, conserved	+	-	60	75.2	0	58547	3	2	Gene name
Unknown	Tb927.5.4150	Hypothetical protein, conserved	+	+	60	89.3	0	42563	18	5	Gene name
Unknown	Tb09.211.2700	Hypothetical protein, conserved	+	+	61	94.2	0	50755	2	1	Gene name

Unknown	Tb927.4.2060	Hypothetical protein, conserved	+	-	61	91.9	0	67023	10	6	Gene name
Unknown	Tb927.7.760	Hypothetical protein, conserved	+	+	61	71.6	0	64370	7	3	Gene name
Unknown	Tb11.02.3680	Hypothetical protein, conserved	+	+	62	33.5	0	24736	7	2	Gene name
Unknown	Tb927.4.670	Hypothetical protein, conserved	+	+	63	41.7	0	74519	2	2	Gene name
Unknown	Tb927.7.4870	Hypothetical protein, conserved	+	+	63	80.2	0	30666	10	2	Gene name
Unknown	Tb09.160.0600	Hypothetical protein, conserved	+	-	64	62.6	0	68594	10	3	Gene name
Unknown	Tb927.4.3330	Hypothetical protein, conserved	+	+	64	81.5	0	20089	7	1	Gene name
Unknown	Tb09.211.0600	Hypothetical protein, conserved	+	-	66	61.5	0	20184	11	2	Gene name
Unknown	Tb11.02.4380	Hypothetical protein, conserved	+	+	67	72.5	0	46179	2	1	Gene name
Unknown	Tb10.70.1360	Hypothetical protein, conserved	+	+	68	70.1	0	21867	10	2	Gene name
Unknown	Tb927.4.3130	Hypothetical protein, conserved	+	+	68	75.2	0	30733	8	2	Gene name
Unknown	Tb927.6.570	Hypothetical protein, conserved	+	+	68	79.8	0	11954	45	4	Gene name
Unknown	Tb927.5.1540	Hypothetical protein, conserved	+	-	70	57.7	0	83052	3	3	Gene name
Unknown	Tb10.70.7980	Hypothetical protein, conserved	+	+	75	66.6	0	34112	7	2	Gene name
Unknown	Tb11.02.2130	Hypothetical protein, conserved	+	+	75	79.6	0	35836	10	4	Gene name
Unknown	Tb927.8.610	Hypothetical protein, conserved	+	+	77	45.7	0	96706	3	3	Gene name
Unknown	Tb10.70.0310	Hypothetical protein, conserved	+	-	80	13.2	0	19632	11	2	Gene name
Unknown	Tb10.70.4990	Hypothetical protein, conserved	+	+	81	31.8	0	13757	19	2	Gene name
Unknown	Tb927.3.4980	Hypothetical protein, conserved (pseudogene)	+	-	81	57	0	63339	5	3	Gene name
Unknown	Tb927.4.3060	Hypothetical protein, conserved	+	-	82	60.1	0	17722	12	2	Gene name
Unknown	Tb927.3.3040	Hypothetical protein, conserved	+	+	83	40.4	0	29200	14	3	Gene name
Unknown	Tb927.6.4610	Hypothetical protein, conserved	+	+	83	47.7	0	46410	7	3	Gene name
Unknown	Tb10.70.2730	Hypothetical protein, conserved	+	-	87	32.4	0	116673	4	4	Gene name
Unknown	Tb10.v4.0039	Hypothetical protein, conserved	+	-	87	32.5	0	116686	4	4	Gene name
Unknown	Tb927.3.1990	Hypothetical protein, conserved	+	+	88	62.7	0	54090	3	2	Gene name
Unknown	Tb11.01.0840	Hypothetical protein, conserved	+	-	89	44.1	0	33533	18	3	Gene name
Unknown	Tb11.02.0352	Hypothetical protein, conserved	+	-	89	63.6	0	31806	12	2	Gene name
Unknown	Tb10.61.2110	Hypothetical protein, conserved	+	+	90	83.4	0	75764	11	4	Gene name
Unknown	Tb11.01.5400	Hypothetical protein, conserved	+	-	91	46.3	0	83066	7	3	Gene name
Unknown	Tb11.02.3900	Hypothetical protein, conserved	+	+	91	84.7	0	49667	17	5	Gene name
Unknown	Tb11.02.4760	Hypothetical protein, conserved	+	-	91	61.5	0	187905	2	4	Gene name
Unknown	Tb927.6.5030	Hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative	+	+	91	33.9	0	65091	9	4	Gene name
Unknown	Tb09.160.1690	Hypothetical protein, conserved	+	-	92	44.6	0	8390	28	2	Gene name
Unknown	Tb10.70.1810	Hypothetical protein, conserved	+	+	92	33.5	0	126413	6	4	Gene name
Unknown	Tb09.211.0690	Hypothetical protein, conserved	+	-	94	85.4	0	39231	10	2	Gene name
Unknown	Tb11.01.6050	Hypothetical protein, conserved	+	-	95	64.3	0	85971	3	2	Gene name
Unknown	Tb11.50.0001	Hypothetical protein, conserved	+	+	95	31.2	0	30219	19	4	Gene name
Unknown	Tb927.5.2930	Hypothetical protein, conserved	+	+	96	71.3	0	43396	11	4	Gene name
Unknown	Tb927.7.4910	Hypothetical protein, conserved	+	+	96	95.4	0	41028	9	3	Gene name
Unknown	Tb927.8.8280	Hypothetical protein, conserved	+	-	96	67.5	0	30364	7	2	Gene name
Unknown	Tb09.160.1180	Hypothetical protein, conserved	+	-	98	71.5	0	66216	26	20	Gene name
Unknown	Tb10.406.0270	Hypothetical protein, conserved	+	+	98	52.2	0	29637	21	5	Gene name
Unknown	Tb11.03.0470	Hypothetical protein, conserved	+	+	98	57.5	0	40745	8	3	Gene name
Unknown	Tb10.70.6570	Hypothetical protein, conserved	+	-	100	41.4	0	333607	1	4	Gene name
Unknown	Tb927.2.2770	Hypothetical protein, conserved	+	-	100	87.4	0	13665	13	2	Gene name
Unknown	Tb09.211.3470	Hypothetical protein, conserved	+	-	104	90.7	0	17556	16	3	Gene name
Unknown	Tb927.7.4840	Hypothetical protein, conserved	+	-	104	56.7	0	38486	12	3	Gene name
Unknown	Tb11.01.3000	Hypothetical protein, conserved	+	-	106	58.1	0	42313	9	3	Gene name

Unknown	Tb11.01.4850	Hypothetical protein, conserved	+	-	107	21	0	30257	14	3	Gene name
Unknown	Tb927.2.2510	Hypothetical protein, conserved	+	-	109	97.1	0	29684	11	2	Gene name
Unknown	Tb927.2.2520	Hypothetical protein, conserved	+	-	109	97.1	0	29684	11	2	Gene name
Unknown	Tb09.160.5060	Hypothetical protein, conserved	+	-	110	41.5	0	41365	31	13	Gene name
Unknown	Tb927.4.1890	Hypothetical protein, conserved	+	+	110	35.7	0	26588	8	2	Gene name
Unknown	Tb927.7.7240	Hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative	+	+	111	62.9	0	79780	5	3	Gene name
Unknown	Tb927.6.3140	Hypothetical protein, conserved	+	-	112	37.3	0	10843	29	2	Gene name
Unknown	Tb10.70.1080	Hypothetical protein, conserved	+	+	113	44.8	0	16356	38	4	Gene name
Unknown	Tb927.8.6920	Hypothetical protein, conserved	+	+	114	43.9	0	29698	19	3	Gene name
Unknown	Tb927.8.2030	Hypothetical protein, conserved	+	+	115	85.7	0	48209	8	4	Gene name
Unknown	Tb10.70.0600	Hypothetical protein, conserved	+	-	116	74.1	0	78175	4	3	Gene name
Unknown	Tb927.7.3920	Hypothetical protein, conserved	+	+	116	35.5	0	36347	15	3	Gene name
Unknown	Tb927.8.1560	Hypothetical protein, conserved	+	+	118	84.8	0	69782	12	5	Gene name
Unknown	Tb05.5K5.120	Hypothetical protein, conserved	+	-	119	64.6	0	97130	7	5	Gene name
Unknown	Tb927.5.4470	Hypothetical protein, conserved	+	+	119	64.9	0	97130	7	5	Gene name
Unknown	Tb09.160.3930	Hypothetical protein, conserved	+	+	120	90.9	0	34427	13	3	Gene name
Unknown	Tb11.01.5060	Hypothetical protein, conserved	+	-	122	88.1	0	56006	8	3	Gene name
Unknown	Tb927.4.1300	Hypothetical protein, conserved	+	+	122	98	0	42444	13	4	Gene name
Unknown	Tb927.7.6890	Hypothetical protein, conserved	+	+	122	96.5	0	29061	20	3	Gene name
Unknown	Tb10.406.0550	Hypothetical protein, conserved	+	+	123	38.2	0	20336	27	4	Gene name
Unknown	Tb10.406.0640	Hypothetical protein, conserved	+	-	123	39.6	0	20336	27	4	Gene name
Unknown	Tb927.3.2890	Hypothetical protein, conserved	+	-	123	63.9	0	37387	35	9	Gene name
Unknown	Tb09.211.1790	Hypothetical protein, conserved	+	+	124	66.7	0	40830	8	2	Gene name
Unknown	Tb927.1.4280	Hypothetical protein, conserved	+	-	124	28.4	0	76031	5	3	Gene name
Unknown	Tb09.211.2250	Hypothetical protein, conserved	+	+	126	52.9	0	31663	16	4	Gene name
Unknown	Tb927.7.5050	Hypothetical protein, conserved	+	+	126	83.3	0	105771	7	6	Gene name
Unknown	Tb10.6k15.3330	Hypothetical protein, conserved	+	-	129	86.2	0	200550	4	8	Gene name
Unknown	Tb09.211.0170	Hypothetical protein, conserved	+	+	131	80.9	0	29240	22	4	Gene name
Unknown	Tb927.8.6980	Hypothetical protein, conserved	+	-	131	52	0	95570	7	6	Gene name
Unknown	Tb09.211.0890	Hypothetical protein, conserved	+	+	132	74.2	0	20148	14	3	Gene name
Unknown	Tb10.61.2920	Hypothetical protein, conserved; PF20	+	+	133	72.7	0	66326	6	3	Gene name
Unknown	Tb11.18.0012	Hypothetical protein, conserved	+	+	134	60.5	0	123971	5	4	Gene name
Unknown	Tb11.47.0026	Hypothetical protein, conserved	+	-	136	25.4	0	53506	11	6	Gene name
Unknown	Tb10.70.4840	Hypothetical protein, conserved	+	-	137	90.3	0	39021	9	3	Gene name
Unknown	Tb10.70.5560	Hypothetical protein, conserved	+	+	137	90.9	0	34759	18	4	Gene name
Unknown	Tb927.3.2050	Hypothetical protein, conserved	+	-	137	74.8	0	173107	4	7	Gene name
Unknown	Tb09.160.0720	Hypothetical protein, conserved	+	+	138	71.9	0	63763	7	3	Gene name
Unknown	Tb09.211.4780	Hypothetical protein, conserved	+	+	138	90.5	0	83476	6	4	Gene name
Unknown	Tb10.6k15.0710	Hypothetical protein, conserved	+	+	138	23.9	0	39551	14	3	Gene name
Unknown	Tb10.70.4780	Hypothetical protein, conserved	+	+	140	69.7	0	85239	8	4	Gene name
Unknown	Tb927.7.4100	Hypothetical protein, conserved	+	-	140	62.6	0	56971	12	5	Gene name
Unknown	Tb10.61.2870	Hypothetical protein, conserved	+	+	142	89.8	0	51170	18	5	Gene name
Unknown	Tb10.70.7560	Hypothetical protein, conserved	+	+	143	83.1	0	25744	20	5	Gene name
Unknown	Tb927.3.5060	Hypothetical protein, conserved	+	-	143	60.4	0	39501	10	3	Gene name
Unknown	Tb927.8.1540	Hypothetical protein, conserved	+	-	144	92	0	98630	4	4	Gene name
Unknown	Tb11.01.7750	Hypothetical protein, conserved	+	+	145	73.1	0	72444	10	5	Gene name
Unknown	Tb927.4.4690	Hypothetical protein, conserved	+	+	145	74	0	31831	15	5	Gene name
Unknown	Tb09.211.0775	Hypothetical protein, conserved	+	-	146	97.8	0	9640	34	4	Gene name

Unknown	Tb927.3.2490	Hypothetical protein, conserved	+	-	146	77.1	0	110689	7	6	Gene name
Unknown	Tb927.7.2300	Hypothetical protein, conserved	+	+	147	59	0	133931	5	5	Gene name
Unknown	Tb10.70.4130	Hypothetical protein, conserved	+	+	148	42.5	0	110685	7	6	Gene name
Unknown	Tb927.4.2840	Hypothetical protein, conserved	+	+	148	95	0	28469	18	3	Gene name
Unknown	Tb09.160.1660	Hypothetical protein, conserved	+	-	150	84.1	0	9322	33	3	Gene name
Unknown	Tb927.2.3020	Hypothetical protein, conserved	+	+	150	62.9	0	119330	8	7	Gene name
Unknown	Tb927.5.2250	Hypothetical protein, conserved	+	+	154	90.1	0	49690	13	4	Gene name
Unknown	Tb927.2.4780	Hypothetical protein, conserved	+	-	155	50.4	0	107624	8	5	Gene name
Unknown	Tb11.01.8640	Hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative	+	+	157	71.6	0	71271	7	6	Gene name
Unknown	Tb927.7.3310	Hypothetical protein, conserved	+	-	157	83.1	0	39081	11	4	Gene name
Unknown	Tb927.4.1610	Hypothetical protein, conserved	+	-	161	85.4	0	39892	11	4	Gene name
Unknown	Tb09.211.0160	Hypothetical protein, conserved	+	-	165	22.2	0	71770	8	4	Gene name
Unknown	Tb927.4.1680	Hypothetical protein, conserved	+	+	166	63.1	0	52555	10	4	Gene name
Unknown	Tb927.3.1740	Hypothetical protein, conserved	+	+	168	50.8	0	77549	12	6	Gene name
Unknown	Tb927.8.2290	Hypothetical protein, conserved	+	+	170	78.9	0	23641	24	4	Gene name
Unknown	Tb10.6k15.0690	Hypothetical protein, conserved	+	+	172	76	0	155268	6	7	Gene name
Unknown	Tb11.01.2390	Hypothetical protein, conserved	+	+	175	16.4	0	22550	27	5	Gene name
Unknown	Tb11.01.6510	Hypothetical protein, conserved	+	+	176	94.7	0	41175	11	4	Gene name
Unknown	Tb10.389.0860	Hypothetical protein, conserved	+	+	177	30.7	0	105205	7	6	Gene name
Unknown	Tb927.5.1230	Hypothetical protein, conserved	+	+	177	45.6	0	58643	9	5	Gene name
Unknown	Tb927.7.2390	Hypothetical protein, conserved	+	-	177	76.1	0	102796	8	9	Gene name
Unknown	Tb11.01.4030	Hypothetical protein, conserved	+	-	181	89.9	0	32752	30	7	Gene name
Unknown	Tb927.3.2950	Hypothetical protein, conserved	+	-	182	67.8	0	84641	9	6	Gene name
Unknown	Tb11.02.0140	Hypothetical protein, conserved	+	+	183	78.8	0	33941	24	6	Gene name
Unknown	Tb11.02.1500	Hypothetical protein, conserved	+	-	183	43.2	0	63324	12	5	Gene name
Unknown	Tb11.01.1210	Hypothetical protein, conserved	+	+	184	78.4	0	62353	21	9	Gene name
Unknown	Tb927.4.360	Hypothetical protein, conserved	+	-	184	53	0	33621	18	5	Gene name
Unknown	Tb927.6.4140	Hypothetical protein, conserved	+	+	185	93.1	0	13229	42	6	Gene name
Unknown	Tb927.4.1740	Hypothetical protein, conserved	+	+	191	58.4	0	47989	17	6	Gene name
Unknown	Tb09.211.3955	Hypothetical protein, conserved; predicted heat shock factor binding protein	+	-	193	95.1	0	9244	60	3	Gene name
Unknown	Tb11.02.5550	Hypothetical protein, conserved; predicted WD40 repeat protein	+	+	193	94.6	0	67653	19	11	Gene name
Unknown	Tb927.4.2140	Hypothetical protein, conserved	+	+	193	69.1	0	115150	11	9	Gene name
Unknown	Tb11.01.2670	Hypothetical protein, conserved	+	+	195	73.5	0	100301	11	8	Gene name
Unknown	Tb927.7.4460	Hypothetical protein, conserved	+	+	198	56.4	0	92510	10	8	Gene name
Unknown	Tb927.3.5010	Hypothetical protein, conserved	+	+	201	84.8	0	23466	22	7	Gene name
Unknown	Tb927.7.5660	Hypothetical protein, conserved	+	+	201	58.8	0	65613	15	7	Gene name
Unknown	Tb927.8.3790	Hypothetical protein, conserved	+	+	203	32.7	0	25503	34	5	Gene name
Unknown	Tb10.61.0940	Hypothetical protein, conserved	+	+	207	30.1	0	32822	22	5	Gene name
Unknown	Tb927.5.2850	Hypothetical protein, conserved	+	+	208	91.9	0	57836	21	8	Gene name
Unknown	Tb11.18.0003	Hypothetical protein, conserved	+	+	209	70.8	0	91091	9	6	Gene name
Unknown	Tb11.02.2530	Hypothetical protein, conserved	+	+	211	89.9	0	31088	20	4	Gene name
Unknown	Tb927.1.3250	Hypothetical protein, conserved	+	-	211	69.1	0	198546	7	12	Gene name
Unknown	Tb10.61.2060	Hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative	+	+	212	64.7	0	122609	11	8	Gene name
Unknown	Tb927.7.6610	Hypothetical protein, conserved	+	-	216	50.7	0	98325	8	7	Gene name
Unknown	Tb11.01.3070	Hypothetical protein, conserved	+	-	217	61.8	0	121921	8	10	Gene name
Unknown	Tb927.6.3820	Hypothetical protein, conserved	+	+	218	67.8	0	94228	8	5	Gene name

Unknown	Tb927.6.3220	Hypothetical protein, conserved	+	+	222	59.8	0	117260	10	10	Gene name
Unknown	Tb927.8.3870	Hypothetical protein, conserved	+	-	227	86.3	0	127828	4	6	Gene name
Unknown	Tb927.3.5020	Hypothetical protein, conserved	+	-	229	76.9	0	156845	11	9	Gene name
Unknown	Tb927.4.660	Hypothetical protein, conserved	+	+	230	75.6	0	79501	16	8	Gene name
Unknown	Tb927.6.4670	Hypothetical protein, conserved	+	+	230	90.3	0	41003	15	6	Gene name
Unknown	Tb10.61.0560	Hypothetical protein, conserved	+	-	231	32.3	0	197618	4	7	Gene name
Unknown	Tb11.02.2490	Hypothetical protein, conserved	+	+	236	91.6	0	30253	30	5	Gene name
Unknown	Tb11.02.5460	Hypothetical protein, conserved	+	+	236	47.7	0	26354	31	6	Gene name
Unknown	Tb05.5K5.40	Hypothetical protein, conserved	+	-	239	48.1	0	15168	40	4	Gene name
Unknown	Tb927.5.4390	Hypothetical protein, conserved	+	-	239	48.1	0	15168	40	4	Gene name
Unknown	Tb927.4.4700	Hypothetical protein, conserved	+	+	241	91.7	0	31151	30	8	Gene name
Unknown	Tb927.6.1660	Hypothetical protein, conserved	+	+	241	75.8	0	65495	23	9	Gene name
Unknown	Tb927.3.1900	Hypothetical protein, conserved	+	+	245	89.8	0	95871	9	7	Gene name
Unknown	Tb10.61.1200	Hypothetical protein, conserved	+	+	247	55.8	0	132758	11	10	Gene name
Unknown	Tb10.61.2100	Hypothetical protein, conserved	+	+	248	66	0	81886	13	8	Gene name
Unknown	Tb927.7.7250	Hypothetical protein, conserved	+	+	255	66.7	0	144193	8	10	Gene name
Unknown	Tb927.8.4460	Hypothetical protein, conserved	+	+	255	49.3	0	110896	11	8	Gene name
Unknown	Tb927.3.5370	Hypothetical protein, conserved	+	-	258	24.9	0	34290	23	6	Gene name
Unknown	Tb927.8.810	Hypothetical protein, conserved	+	+	263	79.6	0	37894	25	7	Gene name
Unknown	Tb10.6k15.1450	Hypothetical protein, conserved	+	-	268	90.1	0	17587	38	6	Gene name
Unknown	Tb927.6.5070	Hypothetical protein, conserved	+	+	268	47.7	0	52188	13	6	Gene name
Unknown	Tb11.02.3880	Hypothetical protein, conserved	+	-	270	67.5	0	191666	10	14	Gene name
Unknown	Tb927.8.6230	Hypothetical protein, conserved	+	+	270	82.4	0	37530	28	11	Gene name
Unknown	Tb927.5.500	Hypothetical protein, conserved	+	+	272	83.6	0	154811	10	11	Gene name
Unknown	Tb927.6.4520	Hypothetical protein, conserved	+	+	275	61.1	0	52672	23	9	Gene name
Unknown	Tb10.389.1320	Hypothetical protein, conserved	+	+	276	83.4	0	24929	44	7	Gene name
Unknown	Tb10.70.7320	Hypothetical protein, conserved	+	-	282	23.9	0	196727	7	12	Gene name
Unknown	Tb11.02.5150	Hypothetical protein, conserved	+	-	284	84.3	0	58750	19	9	Gene name
Unknown	Tb11.02.0810	Hypothetical protein, conserved	+	+	285	69.8	0	108024	15	13	Gene name
Unknown	Tb927.2.4520	Hypothetical protein, conserved	+	+	287	55.5	0	144760	10	12	Gene name
Unknown	Tb927.7.3330	Hypothetical protein, conserved	+	-	293	65.6	0	504918	3	16	Gene name
Unknown	Tb10.6k15.0640	Hypothetical protein, conserved; predicted C2 domain protein	+	-	308	52.6	0	133954	16	15	Gene name
Unknown	Tb10.6k15.2670	Hypothetical protein, conserved	+	+	308	71.2	0	77166	9	7	Gene name
Unknown	Tb11.02.0210	Hypothetical protein, conserved	+	+	309	87.3	0	51028	28	9	Gene name
Unknown	Tb927.4.1040	Hypothetical protein, conserved	+	+	312	22.5	0	19358	57	9	Gene name
Unknown	Tb927.8.4870	Hypothetical protein, conserved	+	+	313	56.4	0	137720	9	10	Gene name
Unknown	Tb927.3.5620	Hypothetical protein, conserved	+	-	323	74.3	0	113215	17	13	Gene name
Unknown	Tb927.3.1040	Hypothetical protein, conserved (pseudogene)	+	+	325	27.5	0	46407	22	8	Gene name
Unknown	Tb11.01.6790	Hypothetical protein, conserved	+	-	328	61.3	0	62564	19	11	Gene name
Unknown	Tb10.6k15.2630	Hypothetical protein, conserved	+	+	330	54.6	0	26306	30	6	Gene name
Unknown	Tb11.01.2200	Hypothetical protein, conserved	+	-	330	49.7	0	63808	20	7	Gene name
Unknown	Tb927.3.1670	Hypothetical protein, conserved	+	+	332	71.6	0	113170	16	12	Gene name
Unknown	Tb11.02.0990	Hypothetical protein, conserved	+	+	335	87.9	0	116028	21	14	Gene name
Unknown	Tb10.6k15.0140	Hypothetical protein, conserved	+	-	337	89.3	0	37447	24	7	Gene name
Unknown	Tb10.26.0090	Hypothetical protein, conserved	+	-	338	56.5	0	131876	15	14	Gene name
Unknown	Tb11.01.1960	Hypothetical protein, conserved	+	-	343	34.5	0	53514	22	9	Gene name
Unknown	Tb11.02.0170	Hypothetical protein, conserved	+	-	343	42.4	0	63817	24	10	Gene name
Unknown	Tb927.3.3770	Hypothetical protein, conserved	+	+	344	94.3	0	31927	27	9	Gene name

Unknown	Tb927.3.3790	Hypothetical protein, conserved	+	-	344	94.3	0	31927	27	9	Gene name
Unknown	Tb927.7.2650	Hypothetical protein, conserved	+	-	344	40.3	0	62248	10	8	Gene name
Unknown	Tb927.5.1680	Hypothetical protein, conserved	+	+	346	60.8	0	196569	7	13	Gene name
Unknown	Tb927.7.2790	Hypothetical protein, conserved	+	+	352	54.9	0	79596	14	8	Gene name
Unknown	Tb10.6k15.1500	Hypothetical protein, conserved	+	+	364	95.1	0	45371	44	28	Gene name
Unknown	Tb927.3.2310	Hypothetical protein, conserved	+	+	369	69.3	0	33935	47	12	Gene name
Unknown	Tb09.211.1910	Hypothetical protein, conserved	+	-	373	42.1	0	119499	15	14	Gene name
Unknown	Tb927.3.1010	Hypothetical protein, conserved	+	-	375	61.8	0	64472	19	10	Gene name
Unknown	Tb11.02.0354	Hypothetical protein, conserved	+	+	380	61.8	0	80555	22	11	Gene name
Unknown	Tb927.3.1060	Hypothetical protein, conserved (pseudogene)	+	+	386	22.1	0	44081	32	13	Gene name
Unknown	Tb927.3.3750	Hypothetical protein, conserved	+	+	386	86.6	0	20018	46	9	Gene name
Unknown	Tb10.26.0680	Hypothetical protein, conserved	+	+	392	77.2	0	14433	57	9	Gene name
Unknown	Tb11.02.0860	Hypothetical protein, conserved	+	+	395	56.7	0	103474	13	11	Gene name
Unknown	Tb927.3.1200	Hypothetical protein, conserved	+	-	396	50.7	0	84794	13	8	Gene name
Unknown	Tb11.01.4370	Hypothetical protein, conserved	+	-	398	87.7	0	90998	16	11	Gene name
Unknown	Tb09.211.1470	Hypothetical protein, conserved	+	+	399	60.7	0	35560	46	12	Gene name
Unknown	Tb11.02.1190	Hypothetical protein, conserved	+	+	404	81.7	0	61204	25	12	Gene name
Unknown	Tb09.211.4280	Hypothetical protein, conserved	+	-	405	65	0	179535	10	15	Gene name
Unknown	Tb09.211.3890	Hypothetical protein, conserved	+	-	406	37.7	0	145586	11	13	Gene name
Unknown	Tb11.01.0680	Hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative	+	+	410	44.2	0	79596	25	16	Gene name
Unknown	Tb11.01.1170	Hypothetical protein, conserved	+	+	415	81.3	0	96703	16	11	Gene name
Unknown	Tb10.61.2210	Hypothetical protein, conserved	+	+	419	92.4	0	37689	33	8	Gene name
Unknown	Tb10.6k15.1510	Hypothetical protein, conserved	+	+	422	46.9	0	81160	20	13	Gene name
Unknown	Tb10.61.0540	Hypothetical protein, conserved	+	+	431	94.7	0	36731	30	10	Gene name
Unknown	Tb10.6k15.2540	Hypothetical protein, conserved	+	+	433	67.6	0	152230	12	13	Gene name
Unknown	Tb10.61.2450	Hypothetical protein, conserved	+	+	435	77.9	0	122868	23	18	Gene name
Unknown	Tb11.39.0004	Hypothetical protein, conserved	+	+	448	91.5	0	45606	27	11	Gene name
Unknown	Tb11.01.6740	Hypothetical protein, conserved	+	+	453	70	0	128361	9	14	Gene name
Unknown	Tb927.7.6280	Hypothetical protein, conserved	+	+	453	75.1	0	72692	26	12	Gene name
Unknown	Tb11.01.6840	Hypothetical protein, conserved	+	+	454	75.3	0	77410	24	13	Gene name
Unknown	Tb10.61.2220	Hypothetical protein, conserved	+	+	455	89.4	0	46782	19	9	Gene name
Unknown	Tb927.8.6660	Hypothetical protein, conserved	+	+	458	96.8	0	69012	23	12	Gene name
Unknown	Tb10.26.0760	Hypothetical protein, conserved	+	+	465	39.3	0	101359	20	14	Gene name
Unknown	Tb927.1.4180	Hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative	+	-	466	88.7	0	82679	16	10	Gene name
Unknown	Tb927.4.2600	Hypothetical protein, conserved	+	+	474	64.6	0	183147	12	17	Gene name
Unknown	Tb10.70.5350	Hypothetical protein, conserved	+	+	481	59.7	0	120351	16	16	Gene name
Unknown	Tb927.4.5380	Hypothetical protein, conserved	+	+	533	40.5	0	167413	10	14	Gene name
Unknown	Tb927.8.6940	Hypothetical protein, conserved	+	-	533	40.8	0	167477	10	14	Gene name
Unknown	Tb11.01.2800	Hypothetical protein, conserved	+	+	535	94.5	0	41627	41	15	Gene name
Unknown	Tb927.8.8200	Hypothetical protein, conserved	+	-	537	86.4	0	100090	20	14	Gene name
Unknown	Tb927.3.3300	Hypothetical protein, conserved	+	-	538	53.9	0	89487	20	12	Gene name
Unknown	Tb927.3.4970	Hypothetical protein, conserved	+	+	551	56.9	0	116142	22	19	Gene name
Unknown	Tb927.8.6240	Hypothetical protein, conserved	+	+	554	95.5	0	30365	60	16	Gene name
Unknown	Tb09.160.0790	Hypothetical protein, conserved	+	-	563	89.9	0	77653	21	14	Gene name
Unknown	Tb11.01.1625	Hypothetical protein, conserved	+	-	564	87.6	0	11972	72	17	Gene name
Unknown	Tb927.4.4130	Hypothetical protein, conserved	+	+	568	85.3	0	100263	25	17	Gene name

Unknown	Tb11.01.3960	Hypothetical protein, conserved; calmodulin-like protein containing EF hand domain	+	+	570	69.8	0	67399	23	14	Gene name
Unknown	Tb927.8.4580	Hypothetical protein, conserved	+	+	577	82.6	0	58220	28	19	Gene name
Unknown	Tb927.2.2160	Hypothetical protein, conserved	+	-	581	75	0	38221	40	15	Gene name
Unknown	Tb10.6k15.2920	hypothetical protein, conserved; rib72 protein- like protein	+	+	585	74.8	0	83470	28	16	Gene name
Unknown	Tb09.160.0650	Hypothetical protein, conserved	+	-	586	77.4	0	508129	8	26	Gene name
Unknown	Tb927.4.3740	Hypothetical protein, conserved	+	-	587	86.9	0	193592	10	22	Gene name
Unknown	Tb927.5.2950	Hypothetical protein, conserved	+	+	589	58	0	89227	24	16	Gene name
Unknown	Tb11.01.2310	Hypothetical protein, conserved	+	+	604	75.5	0	99725	23	19	Gene name
Unknown	Tb927.6.620	Hypothetical protein, conserved	+	-	613	86	0	304434	13	27	Gene name
Unknown	Tb11.02.4230	Hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative	+	+	615	77.4	0	96506	20	14	Gene name
Unknown	Tb11.02.1260	Hypothetical protein, conserved	+	+	635	87.3	0	29138	53	11	Gene name
Unknown	Tb11.01.8770	Hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative	+	+	675	95.4	0	111078	21	19	Gene name
Unknown	Tb927.6.3150	Hypothetical protein, conserved	+	+	691	47.1	0	505384	11	40	Gene name
Unknown	Tb927.7.3560	Hypothetical protein, conserved	+	+	705	55.1	0	201249	15	25	Gene name
Unknown	Tb927.8.4400	Hypothetical protein, conserved	+	+	705	52.9	0	76286	25	13	Gene name
Unknown	Tb927.7.6910	Hypothetical protein, conserved	+	+	720	87.8	0	92484	24	17	Gene name
Unknown	Tb10.6k15.1760	Hypothetical protein, conserved	+	-	761	79.2	0	247538	18	37	Gene name
Unknown	Tb927.2.4810	Hypothetical protein, conserved	+	-	778	88.7	0	151320	22	24	Gene name
Unknown	Tb927.7.1310	Hypothetical protein, conserved	+	+	876	67.7	0	153566	24	27	Gene name
Unknown	Tb10.6k15.3460	Hypothetical protein, conserved	+	+	889	80.3	0	287544	16	33	Gene name
Unknown	Tb927.7.3550	Hypothetical protein, conserved	+	+	919	83.1	0	138765	18	23	Gene name
Unknown	Tb927.4.2080	Hypothetical protein, conserved	+	+	1012	14.8	0	105048	26	27	Gene name
Unknown	Tb927.7.3740	Hypothetical protein, conserved	+	+	1057	82.4	0	93536	28	21	Gene name
Unknown	Tb10.6k15.0810	Hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative	+	+	1084	33.5	0	120018	34	28	Gene name
Unknown	Tb927.1.4310	Hypothetical protein, conserved	+	+	1091	78.8	0	184336	24	40	Gene name
Unknown	Tb11.02.4320	Hypothetical protein, conserved	+	+	1184	92.2	0	107744	37	29	Gene name
Unknown	Tb11.47.0006	Hypothetical protein, conserved	+	+	1190	86.8	0	87200	49	36	Gene name
Unknown	Tb09.160.1160	Hypothetical protein, conserved	+	+	1235	80.1	0	86488	38	31	Gene name
Unknown	Tb927.8.4780	Hypothetical protein, conserved	+	-	1337	74.1	0	471038	13	50	Gene name
Unknown	Tb927.8.1550	Hypothetical protein, conserved	+	+	1637	95.8	0	88466	51	43	Gene name
Unknown	Tb09.211.2160	Hypothetical protein	+	-	56	43.8	0	21856	9	2	Gene name
Unknown	Tb11.0840	Hypothetical protein	+	-	95	29.9	0	29846	19	4	Gene name
Unknown	Tb11.01.4380	Hypothetical protein	+	+	115	34.7	0	44161	17	5	Gene name
Unknown	Tb10.v4.0053	Hypothetical protein	+	-	317	94.3	0	483570	2	11	Gene name
Unknown	Tb11.1380	Hypothetical protein	+	-	448	91.5	0	45606	27	11	Gene name
Unknown	Tb11.1220	Hypothetical protein	+	-	561	75.2	0	58219	28	12	Gene name
Cytoplasm	Tb927.8.3060	Cytosolic leucyl aminopeptidase, putative	+	-	471	89.5	0	71726	22	13	Annotated as
Cytoskeleton	Tb11.01.3805	CAP15 microtubule-associated protein; corset-associated protein 15	+	-	237	55	0	14082	30	6	Annotated as
Flagellum	Tb927.8.4640	Flagellar protofilament ribbon protein, putative	+	+	237	92.1	0	46837	19	6	Annotated as
Flagellum	Tb11.02.2060	Flagellar radial spoke component, putative; flagellar radial spoke protei	+	+	427	82.5	0	67377	18	10	Annotated as
Flagellum	Tb927.3.3690	Flagellar radial spoke protein-like, putative	+	+	287	91.2	0	61411	19	9	Annotated as
Flagellum	Tb05.5K5.130	Paraflagellar rod component Par4, putative	+	-	805	56.3	0	68443	35	17	Annotated as
Flagellum	Tb927.5.4480	Paraflagellar rod component Par4, putative	+	+	805	55.4	0	68443	35	17	Annotated as
Flagellum	Tb11.01.5100	Paraflagellar rod component, putative	+	+	1453	93.7	0	68934	52	51	Annotated as
Flagellum	Tb927.2.4330	Paraflagellar rod protein, putative	+	+	193	63.6	0	87717	10	6	Annotated as
Flagellum	Tb927.3.4320	PFR1 73 kDa paraflagellar rod protein; PFR1	+	+	1560	99.3	0	69096	48	75	Annotated as
Flagellum	Tb927.3.4290	PFR1 73 kDa paraflagellar rod protein; PFR1	+	-	1560	99.3	0	69096	48	75	Annotated as

Flagellum	Tb927.3.4300	PFR1 73 kDa paraflagellar rod protein; PFR1	+	-	1560	99.3	0	69096	48	75	Annotated as
Flagellum	Tb927.3.4310	PFR1 73 kDa paraflagellar rod protein; PFR1	+	-	1560	99.3	0	69096	48	75	Annotated as
Flagellum	Tb927.3.4330	PFR1 73 kDa paraflagellar rod protein; PFR1	+	-	1560	99.4	0	69096	48	75	Annotated as
Flagellum	Tb927.8.5000	PFR2 69 kDa paraflagellar rod protein; PFR1	+	+	1710	99.4	0	69953	46	99	Annotated as
Flagellum	Tb927.8.4970	PFR2 69 kDa paraflagellar rod protein; PFR1	+	-	1710	99.4	0	69953	46	99	Annotated as
Flagellum	Tb927.8.4980	PFR2 69 kDa paraflagellar rod protein; PFR1	+	-	1710	99.4	0	69953	46	99	Annotated as
Flagellum	Tb927.8.4990	PFR2 69 kDa paraflagellar rod protein; PFR1	+	-	1710	99.4	0	69953	46	99	Annotated as
Flagellum	Tb927.8.5010	PFR2 69 kDa paraflagellar rod protein; PFR1	+	-	1710	99.4	0	69953	46	99	Annotated as
Flagellum	Tb927.8.5470	TB-17 flagellar calcium-binding protein	+	-	342	96.2	0	25728	20	8	Annotated as
Flagellum	Tb927.8.5440	TB-24 flagellar calcium-binding protein TB-24	+	+	342	96.8	0	24580	21	8	Annotated as
Flagellum	Tb927.8.5460	TB-44A flagellar calcium-binding protein TB-44A	+	-	336	96.4	0	46317	11	8	Annotated as
Glycosome	Tb10.70.1370	ALD fructose-bisphosphate aldolase, glycosomal, putative	+	-	929	99.1	0	41045	38	22	Annotated as
Glycosome	Tb927.6.4280	GAPDH glyceraldehyde 3-phosphate dehydrogenase, glycosomal	+	+	1824	99	0	39023	64	146	Annotated as
Glycosome	Tb927.6.4300	GAPDH glyceraldehyde 3-phosphate dehydrogenase, glycosomal	+	-	1824	99	0	39023	64	146	Annotated as
Glycosome	Tb09.211.3550	glk1 glycerol kinase, glycosomal	+	-	681	98.2	0	57071	28	14	Annotated as
Glycosome	Tb09.211.3560	glk1 glycerol kinase, glycosomal	+	-	681	98.2	0	57131	28	14	Annotated as
Glycosome	Tb09.211.3570	glk1 glycerol kinase, glycosomal	+	-	681	98.2	0	57117	28	14	Annotated as
Glycosome	Tb09.211.3590	glk1 glycerol kinase, glycosomal	+	-	681	98.1	0	57131	28	14	Annotated as
Glycosome	Tb09.211.3540	Glk1 glycerol kinase, glycosomal	+	-	337	98.3	0	57041	17	9	Annotated as
Glycosome	Tb927.8.3530	Glycerol-3-phosphate dehydrogenase [NAD+], glycosomal	+	+	359	98.9	0	38408	19	8	Annotated as
Glycosome	Tb09.160.3590	PDE2C cAMP-specific phosphodiesterase	+	+	660	88.1	0	104726	22	20	Annotated as
Kinetoplast	Tb11.01.8090	Kinetoplast DNA-associated protein, putative	+	-	45	51.5	0	36586	25	13	Annotated as
Membrane	Tb09.211.4511	Kinetoplastid membrane protein KMP-11	+	-	312	99.4	0	11069	64	16	Annotated as
Membrane	Tb09.211.4512	Kinetoplastid membrane protein KMP-11	+	-	312	99.5	0	11069	64	16	Annotated as
Membrane	Tb09.211.4513	Kinetoplastid membrane protein KMP-11	+	-	312	99.5	0	11069	64	16	Annotated as
Mitochondrion	Tb10.70.0280	HSP60 chaperonin Hsp60, mitochondrial precursor	+	+	876	98.1	0	59751	49	21	Annotated as
Mitochondrion	Tb927.5.2790	Pol beta-PAK mitochondrial DNA polymerase beta-PAK	+	-	216	68.3	0	87289	9	6	Annotated as
Nucleus	Tb10.6k15.2350	Nuclear pore complex protein (NUP155), putative; nucleoporin	+	+	53	69.2	0	146215	2	3	Annotated as
Nucleus	Tb927.8.3750	Nucleolar protein, putative	+	-	105	95.5	0	54723	6	2	Annotated as
Nucleus	Tb10.6k15.3670	Nucleoporin interacting component (NUP93), putative	+	-	174	80.6	0	97306	8	5	Annotated as
Nucleus	Tb927.4.5020	RNA polymerase IIA largest subunit	+	-	35	86.1	0	197793	1	3	Annotated as
Nucleus	Tb927.8.7400	RNA polymerase IIA largest subunit	+	-	35	86.6	0	197802	1	3	Annotated as
Nucleus	Tb927.3.1120	rtb2 GTP-binding nuclear protein rtb2, putative	+	-	168	98.4	0	24732	29	3	Annotated as
Flagellum	Tb10.70.2920	Prohibitin, putative	+	+	200	91.8	1	32366	31	7	TbFP
Mitochondrion	Tb10.6k15.3640	AOX alternative oxidase	+	-	469	95.9	1	37738	36	14	GeneDB
Unknown	Tb927.4.2450	Thioredoxin, putative	+	-	118	94.7	1	44748	5	4	GeneDB
Intracellular	Tb10.70.1690	40S ribosomal protein S10, putative	+	-	50	95.9	1	24864	11	2	Gene name
Nucleus	Tb927.2.1100	Retrotransposon hot spot protein (RHS, pseudogene), putative	+	-	44	3.4	1	69046	5	4	Gene name
Unknown	Tb10.61.3100	Chaperone protein DNAJ, putative	+	-	33	68.3	1	27217	3	1	Gene name
Unknown	Tb927.4.1150	Hypothetical protein, conserved	+	-	75	52.1	1	83784	6	4	Gene name
Unknown	Tb927.6.4500	Hypothetical protein, conserved	+	-	234	96.3	1	22748	28	6	Gene name
Unknown	Tb927.3.3200	Hypothetical protein, conserved	+	+	303	41.4	1	45491	17	8	Gene name
Unknown	Tb10.61.2200	Hypothetical protein, conserved	+	-	315	70.9	1	95788	17	10	Gene name
Unknown	Tb11.01.4810	Hypothetical protein, conserved	+	+	318	88.8	1	24236	29	5	Gene name
Unknown	Tb10.70.4610	Hypothetical protein, conserved	+	+	322	41.4	1	209350	11	16	Gene name
Glycosome	Tb09.211.2730	Gim5A Gim5A protein; glycosomal membrane protein	+	-	322	90.8	1	26790	34	11	Annotated as

Glycosome	Tb09.211.2740	Gim5B Gim5B protein; glycosomal membrane protein	+	-	273	92.9	1	26529	30	10	Annotated as
Mitochondrion	Tb09.211.1750	Mitochondrial carrier protein, putative; mitochondrial phosphate transporter	+	-	476	96.8	1	34729	35	10	Annotated as
Unknown	Tb927.5.1900	Hypothetical protein, conserved	+	+	64	90.3	2	78116	4	2	Gene name
Glycosome	Tb11.01.3370	PEX11 glycosomal membrane protein, putative	+	-	156	97.4	2	24240	14	4	Annotated as
Unknown	Tb09.211.1550	Chaperone protein DNAJ, putative	+	-	43	69.9	3	56676	1	1	GeneDB
Mitochondrion	Tb10.61.1810	Mitochondrial carrier protein, putative; ADP/ATP translocase 1, putative	+	-	188	96.7	3	34053	30	19	Annotated as
Mitochondrion	Tb10.61.1820	Mitochondrial carrier protein, putative; ADP/ATP translocase 1, putative	+	-	188	96.7	3	34053	30	19	Annotated as
Mitochondrion	Tb10.61.1830	Mitochondrial carrier protein, putative; ADP/ATP translocase 1, putative	+	-	188	96.7	3	34053	30	19	Annotated as
Membrane	Tb927.3.1840	3-oxo-5-alpha-steroid 4-dehydrogenase, putative	+	-	143	94.2	4	33619	8	4	GeneDB
Unknown	Tb10.70.3750	Hypothetical protein, conserved	+	-	117	81.4	6	31571	9	2	Gene name
Unknown	Tb11.02.3050	Hypothetical protein, conserved	+	-	114	39.7	7	77727	7	4	Gene name
Unknown	Tb11.01.0480	Hypothetical protein, conserved	+	-	120	90.6	9	64353	7	4	Gene name
Membrane	Tb927.8.2380	ABC transporter, putative	+	-	36	75.2	9	128959	0	1	5+ TMD
Flagellum	Tb11.01.4180	Ankyrin repeat protein, putative	-	+	37	75.8	0	17846	13	2	TbFP
Flagellum	Tb09.160.2070	Cyclophilin type peptidyl-prolyl cis-trans isomerase, putative	-	+	34	81.1	0	36327	2	1	TbFP
Flagellum	Tb10.61.2630	Protein transport protein Sec13, putative; cytosolic coat protein, putative	-	+	54	78.9	0	42048	8	2	TbFP
TbFP	Tb10.6k15.2590	Hypothetical protein, conserved	-	+	-	93.1	0	-	-	-	TbFP
TbFP	Tb10.6k15.3580	Hypothetical protein, conserved	-	+	-	86.3	0	-	-	-	TbFP
TbFP	Tb10.70.0920	Hypothetical protein, conserved	-	+	-	36.4	0	-	-	-	TbFP
TbFP	Tb10.70.3720	NADH-dependent fumarate reductase, putative	-	+	-	84.9	0	-	-	-	TbFP
TbFP	Tb10.70.4340	Hypothetical protein, conserved	-	+	-	50.5	0	-	-	-	TbFP
TbFP	Tb10.70.4370	Hypothetical protein, conserved	-	+	-	45.4	0	-	-	-	TbFP
TbFP	Tb10.70.4530	Hypothetical protein, conserved	-	+	-	53.3	0	-	-	-	TbFP
TbFP	Tb10.70.6960	Dynein arm light chain, axonemal, putative	-	+	-	86.6	0	-	-	-	TbFP
TbFP	Tb11.01.0160	Hypothetical protein, conserved	-	+	-	36	0	-	-	-	TbFP
TbFP	Tb11.01.4210	Hypothetical protein, conserved	-	+	-	59.8	0	-	-	-	TbFP
TbFP	Tb11.01.4400	Hypothetical protein	-	+	-	73	0	-	-	-	TbFP
TbFP	Tb11.01.5770	Hypothetical protein, conserved	-	+	-	60	0	-	-	-	TbFP
TbFP	Tb11.01.6040	Hypothetical protein, conserved	-	+	-	56.8	0	-	-	-	TbFP
TbFP	Tb11.01.6870	Calpain-like cysteine peptidase, putative	-	+	-	92.1	0	-	-	-	TbFP
TbFP	Tb11.01.7370	Hypothetical protein, conserved	-	+	-	31.8	0	-	-	-	TbFP
TbFP	Tb11.01.7610	Hypothetical protein, conserved	-	+	-	67	0	-	-	-	TbFP
TbFP	Tb11.01.8370	Hypothetical protein, conserved	-	+	-	52.8	0	-	-	-	TbFP
TbFP	Tb11.02.1135	Dynein light chain, putative	-	+	-	93.9	0	-	-	-	TbFP
TbFP	Tb11.02.4150	PPDK pyruvate phosphate dikinase	-	+	-	93.8	0	-	-	-	TbFP
TbFP	Tb11.02.4620	Hypothetical protein, conserved	-	+	-	82.4	0	-	-	-	TbFP
TbFP	Tb11.02.4660	Hypothetical protein, conserved	-	+	-	72.2	0	-	-	-	TbFP
TbFP	Tb11.02.5590	Hypothetical protein, conserved	-	+	-	61	0	-	-	-	TbFP
TbFP	Tb11.03.0810	Hypothetical protein, conserved	-	+	-	57.1	0	-	-	-	TbFP
TbFP	Tb11.0330	Hypothetical protein	-	+	-	45.7	0	-	-	-	TbFP
TbFP	Tb927.3.3540	Hypothetical protein, conserved	-	+	-	90.2	0	-	-	-	TbFP
TbFP	Tb927.3.5140	Hypothetical protein, conserved	-	+	-	63.5	0	-	-	-	TbFP
TbFP	Tb927.5.2270	Hypothetical protein, conserved	-	+	-	38.1	0	-	-	-	TbFP
TbFP	Tb927.5.2320	Hypothetical protein, conserved	-	+	-	78.8	0	-	-	-	TbFP
TbFP	Tb927.5.2530	Hypothetical protein, conserved	-	+	-	64	0	-	-	-	TbFP
TbFP	Tb927.5.3630	Hypothetical protein, conserved	-	+	-	38.9	0	-	-	-	TbFP
TbFP	Tb927.6.1720	Hypothetical protein, conserved	-	+	-	57.8	0	-	-	-	TbFP
TbFP	Tb927.6.3920	Hypothetical protein, conserved	-	+	-	51.3	0	-	-	-	TbFP

TbFP	Tb927.7.1000	Hypothetical protein, conserved	-	+	-	74.4	0	-	-	-	TbFP
TbFP	Tb927.7.4510	Hypothetical protein, conserved	-	+	-	61.5	0	-	-	-	TbFP
TbFP	Tb927.7.5650	Kinesin, putative	-	+	-	48.1	0	-	-	-	TbFP
TbFP	Tb927.7.6830	Trans-sialidase, putative	-	+	-	91.6	0	-	-	-	TbFP
TbFP	Tb927.7.6850	TbTS trans-sialidase	-	+	-	87.2	0	-	-	-	TbFP
TbFP	Tb927.7.6950	Hypothetical protein, conserved	-	+	-	73.4	0	-	-	-	TbFP
TbFP	Tb927.8.3780	Hypothetical protein, conserved	-	+	-	60.9	0	-	-	-	TbFP
TbFP	Tb927.8.4810	Prohibitin	-	+	-	71.8	0	-	-	-	TbFP
TbFP	Tb927.8.6260	Hypothetical protein, conserved	-	+	-	40.4	0	-	-	-	TbFP
TbFP	Tb927.8.640	Hypothetical protein, conserved	-	+	-	34.3	0	-	-	-	TbFP
TbFP	Tb927.8.6740	DNA repair protein, putative	-	+	-	49.4	0	-	-	-	TbFP
TbFP	Tb927.8.780	Hypothetical protein, conserved	-	+	-	32	0	-	-	-	TbFP
Cytoplasm	Tb927.8.5880	Eukaryotic translation initiation factor 1A, putative	-	-	47	96.8	0	19434	13	2	GeneDB
Cytoplasm	Tb10.6k15.3850	GAP glyceraldehyde 3-phosphate dehydrogenase, cytosolic	-	-	39	97.3	0	35760	2	1	GeneDB
Golgi	Tb10.6k15.2500	BAD1 adaptin complex 1 subunit, putative; beta-adaptin, fragment	-	-	33	94.8	0	76688	1	1	GeneDB
Intracellular	Tb927.4.3570	Translation elongation factor 1-beta, putative	-	-	73	97.8	0	28614	8	2	GeneDB
Intracellular	Tb927.4.3590	Translation elongation factor 1-beta, putative	-	-	73	98	0	28586	8	2	GeneDB
Intracellular / Mitochondria	Tb927.8.1590	upl3 ubiquitin-protein ligase, putative	-	-	36	78.1	0	475321	0	3	GeneDB
Mitochondrion	Tb09.160.3520	Radical SAM domain protein, putative	-	-	57	81.3	0	45780	13	4	GeneDB
Nucleus	Tb927.3.5100	DNA repair helicase and transcription factor protein, putative	-	-	36	75.3	0	106855	2	3	GeneDB
Nucleus	Tb10.6k15.3160	Fibrillarin	-	-	109	54.7	0	31794	8	2	GeneDB
Nucleus	Tb11.01.5570	NRBD1 RNA-binding protein; NRBD1	-	-	44	97.7	0	28759	9	2	GeneDB
Nucleus	Tb11.01.5590	NRBD2 RNA-binding protein; NRBD2	-	-	44	97.9	0	30256	9	2	GeneDB
Nucleus	Tb09.244.2790	rRNA processing protein, putative	-	-	39	69.7	0	28466	3	1	GeneDB
Nucleus	Tb927.4.470	snoRNP protein gar1, putative	-	-	55	69.3	0	21770	11	1	GeneDB
Nucleus	Tb927.3.3490	TDP1 high mobility group protein, putative	-	-	49	83	0	30847	2	1	GeneDB
Unknown	Tb927.8.2020	Arginase, putative	-	-	53	60.7	0	36722	15	4	GeneDB
Unknown	Tb10.6k15.3780	Caltractin, putative	-	-	79	29.3	0	20725	10	2	GeneDB
Unknown	Tb11.01.3180	Guanine nucleotide-binding protein beta subunit-like protein; activated p	-	-	90	98.4	0	35181	22	7	GeneDB
Unknown	Tb927.6.1400	PPIase cyclophilin-type peptidyl-prolyl cis-trans isomerase, putative	-	-	248	34.1	0	28989	20	6	GeneDB
Unknown	Tb11.01.3170	TRACK guanine nucleotide-binding protein beta subunit- like protein; activated protein	-	-	90	98.4	0	35181	22	7	GeneDB
Membrane	Tb11.01.2420	Beta-adaptin 3, putative; adaptin complex 3 subunit, putative	-	-	34	67	0	101210	0	1	GeneDB
Unknown	Tb09.v4.0172	Expression site-associated gene (ESAG, pseudogene), putative	-	-	37	1.8	0	38223	2	3	GeneDB
Unknown	Tb927.7.3420	PPIase peptidyl-prolyl cis-trans isomerase, putative	-	-	67	77.9	0	12388	21	3	GeneDB
Unknown	Tb927.2.5230	Protein kinase, putative	-	-	46	71.2	0	31174	4	1	GeneDB
Unknown	Tb927.7.2440	Pyrroline-5-carboxylate reductase, putative	-	-	49	64.5	0	28652	8	3	GeneDB
Cytoskeleton	Tb05.5K5.90	Dynein light chain, putative	-	-	55	96	0	10596	23	1	Gene name
Cytoskeleton	Tb927.5.4440	Dynein light chain, putative	-	-	55	95.7	0	10596	23	1	Gene name
Intracellular	Tb11.01.7960	60S ribosomal protein L2, putative; 60S ribosomal protein L8, putative	-	-	59	98.8	0	28695	11	3	Gene name
Intracellular	Tb927.5.1110	60S ribosomal protein L2, putative; 60S ribosomal protein L8, putative	-	-	59	98.3	0	28695	11	3	Gene name
Intracellular	Tb927.6.720	40S ribosomal protein L14, putative	-	-	87	98.7	0	21601	15	3	Gene name
Intracellular	Tb11.01.3020	40S ribosomal protein L14, putative	-	-	87	97.9	0	21627	15	3	Gene name
Intracellular	Tb927.1.3180	40S ribosomal protein S11, putative	-	-	65	99.6	0	20275	14	3	Gene name
Intracellular	Tb10.6k15.3340	40S ribosomal protein S24E, putative	-	-	170	99.5	0	15731	21	5	Gene name
Intracellular	Tb10.6k15.3350	40S ribosomal protein S24E, putative	-	-	170	99.3	0	15731	21	5	Gene name
Intracellular	Tb10.70.3370	40S ribosomal protein S3a, putative	-	-	139	99.3	0	29618	21	5	Gene name
Intracellular	Tb11.02.4170	40S ribosomal protein S5, putative	-	-	138	98.8	0	21487	27	5	Gene name
Intracellular	Tb927.8.6150	40S ribosomal protein S8, putative	-	-	138	99.2	0	24979	14	2	Gene name

Intracellular	Tb927.8.6160	40S ribosomal protein S8, putative	-	-	138	99.1	0	24979	14	2	Gene name
Intracellular	Tb927.4.750	50S ribosomal protein L7Ae, putative	-	-	33	86.6	0	16424	7	1	Gene name
Intracellular	Tb10.70.4060	60S acidic ribosomal protein P2, putative	-	-	33	96.1	0	11642	7	1	Gene name
Intracellular	Tb10.70.4070	60S acidic ribosomal protein P2, putative	-	-	33	96.1	0	11642	7	1	Gene name
Intracellular	Tb09.160.4200	60S acidic ribosomal protein, putative	-	-	40	98.9	0	11140	24	2	Gene name
Intracellular	Tb927.5.1820	60S acidic ribosomal protein, putative	-	-	96	97.1	0	10183	13	1	Gene name
Intracellular	Tb11.46.0001	60S acidic ribosomal subunit protein, putative	-	-	192	99.5	0	34891	13	3	Gene name
Intracellular	Tb11.46.0002	60S acidic ribosomal subunit protein, putative	-	-	192	99.5	0	34891	13	3	Gene name
Intracellular	Tb09.211.4550	60S ribosomal protein L12, putative	-	-	68	96.7	0	24287	6	2	Gene name
Intracellular	Tb927.8.6030	60S ribosomal protein L12, putative	-	-	68	99.3	0	17668	9	2	Gene name
Intracellular	Tb927.3.3310	60S ribosomal protein L13, putative	-	-	195	97.9	0	26737	20	5	Gene name
Intracellular	Tb927.4.3550	60S ribosomal protein L13a, putative	-	-	107	96	0	39358	7	2	Gene name
Intracellular	Tb927.5.1610	60S ribosomal protein L13a, putative	-	-	107	98	0	25867	11	2	Gene name
Intracellular	Tb10.70.3510	60S ribosomal protein L18a, putative	-	-	186	96.8	0	21119	32	6	Gene name
Intracellular	Tb927.7.5000	60S ribosomal protein L19, putative	-	-	72	99.5	0	29448	3	1	Gene name
Intracellular	Tb927.7.5020	60S ribosomal protein L19, putative	-	-	72	99.5	0	29448	3	1	Gene name
Intracellular	Tb11.0880	60S ribosomal protein L21E, putative	-	-	107	97	0	18219	17	2	Gene name
Intracellular	Tb10.70.1540	60S ribosomal protein L24, putative	-	-	42	98	0	14700	11	1	Gene name
Intracellular	Tb10.70.1560	60S ribosomal protein L24, putative	-	-	42	98.1	0	14730	11	1	Gene name
Intracellular	Tb09.244.2590	60S ribosomal protein L32	-	-	90	96.3	0	15400	34	6	Gene name
Intracellular	Tb10.100.0155	60S ribosomal protein L32, putative	-	-	90	93.5	0	15443	34	6	Gene name
Intracellular	Tb927.4.2180	60S ribosomal protein L35A, putative	-	-	60	98.1	0	17124	16	2	Gene name
Intracellular	Tb09.244.2730	60S ribosomal protein L5, putative	-	-	54	99.5	0	34672	2	1	Gene name
Intracellular	Tb09.244.2740	60S ribosomal protein L5, putative	-	-	54	99.7	0	34672	2	1	Gene name
Intracellular	Tb10.26.0560	60S ribosomal protein L6, putative	-	-	64	99.6	0	21229	12	2	Gene name
Intracellular	Tb11.01.5720	Ribosomal protein L18, putative	-	-	186	97.4	0	21119	32	6	Gene name
Intracellular	Tb927.4.1100	Ribosomal protein L21E (60S), putative	-	-	107	96.7	0	18232	17	2	Gene name
Intracellular	Tb11.50.0005	Ribosomal protein L21E (60S), putative	-	-	107	97	0	18219	17	2	Gene name
Intracellular	Tb927.7.1730	Ribosomal protein L7, putative	-	-	177	99.2	0	27810	28	6	Gene name
Intracellular	Tb927.7.1740	Ribosomal protein L7, putative	-	-	177	98.3	0	29456	27	6	Gene name
Intracellular	Tb927.7.1750	Ribosomal protein L7, putative	-	-	177	98.3	0	29456	27	6	Gene name
Intracellular	Tb927.4.1860	Ribosomal protein S19, putative	-	-	86	98.7	0	19176	13	2	Gene name
Intracellular	Tb10.6k15.2050	RPS12 40S ribosomal protein S12, putative	-	-	74	95.6	0	16279	16	1	Gene name
Plasma Membrane	Tb11.v4.0029	Variant surface glycoprotein (VSG), putative	-	-	34	36.9	0	52580	3	2	Gene name
Plasma Membrane	Tb927.5.5330	Variant surface glycoprotein (VSG, pseudogene), putative	-	-	41	15.2	0	57564	3	3	Gene name
Plasma Membrane	Tb11.0950	Variant surface glycoprotein (VSG, pseudogene), putative	-	-	34	26.1	0	50801	5	2	Gene name
Plasma Membrane	Tb10.v4.0027	Variant surface glycoprotein (VSG, pseudogene), putative	-	-	35	21.9	0	45472	7	4	Gene name
Plasma Membrane	Tb11.43.0005	Variant surface glycoprotein (VSG, pseudogene), putative	-	-	35	17	0	64811	2	2	Gene name
Plasma Membrane	Tb10.v4.0169	Variant surface glycoprotein (VSG, pseudogene), putative	-	-	36	29.7	0	55164	4	2	Gene name
Nucleus	Tb927.4.1330	DNA topoisomerase IB, large subunit	-	-	61	70.2	0	79608	3	2	Gene name
Nucleus	Tb10.389.0110	DNA-directed RNA polymerases II subunit, putative	-	-	60	79.8	0	25479	6	1	Gene name
Nucleus	Tb927.5.1150	Pre-mRNA splicing factor ATP-dependent RNA helicase, putative	-	-	42	78.2	0	83673	2	1	Gene name
Nucleus	Tb09.v4.0067	Retrotransposon hot spot protein, pseudogene	-	-	65	14.7	0	63417	3	2	Gene name
Nucleus	Tb09.211.2970	SMC1 structural maintenance of chromosome 1, putative	-	-	78	54.1	0	145762	2	3	Gene name
Nucleus	Tb927.5.3510	SMC3 structural maintenance of chromosome 3 , putative	-	+	228	23.5	0	137200	7	7	Gene name
Nucleus	Tb927.2.1810	Transcription activator, putative	-	-	149	68.8	0	133531	4	5	Gene name
Intracellular	Tb09.211.4540	RNA-binding protein, putative; DRBD2	-	-	83	70.9	0	34711	10	3	Gene name
Intracellular	Tb11.03.0620	UBP1 RNA-binding protein, putative	-	-	49	72.9	0	24394	7	1	Gene name
Intracellular	Tb11.03.0580	UBP2 RNA-binding protein, putative	-	-	49	79.9	0	19604	9	1	Gene name
Intracellular	Tb927.4.4980	Adrenodoxin precursor, putative	-	-	31	86.5	0	18007	8	2	Gene name

Unknown	Tb09.160.2400	Hypothetical protein, conserved	-	-	31	58.1	0	67666	3	2	Gene name
Unknown	Tb10.389.1590	Hypothetical protein, conserved	-	-	31	60.3	0	50423	1	1	Gene name
Unknown	Tb10.70.1420	Hypothetical protein, conserved	-	-	31	19	0	39111	5	2	Gene name
Unknown	Tb11.02.4270	Hypothetical protein, conserved	-	-	31	69.3	0	163938	1	2	Gene name
Unknown	Tb927.6.4700	Hypothetical protein, conserved	-	-	31	71.7	0	38110	8	2	Gene name
Unknown	Tb927.7.800	Hypothetical protein, conserved	-	-	31	76.3	0	61130	4	3	Gene name
Unknown	Tb09.160.3020	Hypothetical protein, conserved	-	-	32	18.5	0	46917	6	1	Gene name
Unknown	Tb11.49.0006	Hypothetical protein, conserved (pseudogene)	-	-	32	1.5	0	35404	7	2	Gene name
Unknown	Tb927.7.7460	Hypothetical protein, conserved	-	-	32	53.1	0	22686	6	1	Gene name
Unknown	Tb11.01.2885	Hypothetical protein, conserved	-	-	33	83.9	0	140956	1	1	Gene name
Unknown	Tb927.8.4210	Hypothetical protein, conserved	-	-	33	74.9	0	46908	3	2	Gene name
Unknown	Tb927.2.2550	Hypothetical protein, conserved	-	-	34	51.6	0	58062	5	2	Gene name
Unknown	Tb927.6.2120	Hypothetical protein, conserved	-	-	35	29.2	0	27711	8	2	Gene name
Unknown	Tb10.70.1660	Hypothetical protein, conserved; predicted WD40 protein	-	-	36	81.9	0	141732	1	2	Gene name
Unknown	Tb10.70.2950	Hypothetical protein, conserved	-	-	36	56.9	0	31926	3	1	Gene name
Unknown	Tb10.70.5410	Hypothetical protein, conserved	-	-	36	82.4	0	41723	4	1	Gene name
Unknown	Tb927.8.2870	Hypothetical protein, conserved	-	-	37	18.9	0	108086	2	3	Gene name
Unknown	Tb927.8.5100	Hypothetical protein, conserved	-	-	38	62.7	0	316944	1	5	Gene name
Unknown	Tb10.70.3580	Hypothetical protein, conserved	-	-	39	29.3	0	112233	1	1	Gene name
Unknown	Tb10.70.7350	Hypothetical protein, conserved	-	-	39	41.3	0	37806	2	1	Gene name
Unknown	Tb927.2.5870	Hypothetical protein, conserved	-	-	40	28.4	0	217912	1	3	Gene name
Unknown	Tb09.160.1740	Hypothetical protein, conserved	-	-	41	74.1	0	47929	4	2	Gene name
Unknown	Tb09.160.4710	Hypothetical protein, conserved	-	-	41	56.1	0	171553	2	3	Gene name
Unknown	Tb11.01.2100	Hypothetical protein, conserved	-	-	41	54.1	0	70919	3	1	Gene name
Unknown	Tb927.4.790	Hypothetical protein, conserved	-	-	41	21	0	25489	7	1	Gene name
Unknown	Tb927.7.1830	Hypothetical protein, conserved	-	-	42	32.7	0	41238	3	1	Gene name
Unknown	Tb927.7.1870	Hypothetical protein, conserved	-	-	42	32.5	0	41238	3	1	Gene name
Unknown	Tb10.70.1600	Hypothetical protein, conserved	-	+	44	59.3	0	32881	6	2	Gene name
Unknown	Tb09.211.0180	Hypothetical protein, conserved	-	-	48	89.1	0	76657	2	2	Gene name
Unknown	Tb09.211.1770	Hypothetical protein, conserved	-	-	50	66	0	23568	9	2	Gene name
Unknown	Tb927.8.3160	Hypothetical protein, conserved	-	-	50	36.6	0	18692	10	1	Gene name
Unknown	Tb10.389.0680	Hypothetical protein, conserved	-	-	52	67.3	0	41886	4	1	Gene name
Unknown	Tb11.02.3670	Hypothetical protein, conserved	-	-	54	39	0	64501	12	5	Gene name
Unknown	Tb927.3.1820	Hypothetical protein, conserved	-	-	54	12.7	0	25417	9	2	Gene name
Unknown	Tb10.389.1890	Hypothetical protein, conserved	-	-	55	77	0	37044	19	5	Gene name
Unknown	Tb927.1.690	Hypothetical protein, conserved	-	-	55	15	0	27608	11	2	Gene name
Unknown	Tb11.03.0530	Hypothetical protein, conserved	-	-	56	95.1	0	31330	6	1	Gene name
Unknown	Tb05.5K5.60	Hypothetical protein, conserved	-	-	57	92	0	23693	11	2	Gene name
Unknown	Tb11.01.7630	Hypothetical protein, conserved	-	+	57	73.7	0	110646	1	1	Gene name
Unknown	Tb927.5.4410	Hypothetical protein, conserved	-	-	57	92	0	23693	11	2	Gene name
Unknown	Tb10.389.0280	Hypothetical protein, conserved	-	-	58	69.1	0	137667	1	2	Gene name
Unknown	Tb11.02.5340	Hypothetical protein, conserved	-	-	59	83.6	0	23826	4	1	Gene name
Unknown	Tb927.2.5810	Hypothetical protein, conserved	-	-	60	51.7	0	175790	2	2	Gene name
Unknown	Tb09.211.3340	Hypothetical protein, conserved	-	-	62	90.3	0	18107	9	2	Gene name
Unknown	Tb11.01.2960	Hypothetical protein, conserved	-	-	64	92.5	0	19583	13	2	Gene name
Unknown	Tb11.46.0011	Hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative	-	+	64	49	0	46041	6	3	Gene name
Unknown	Tb927.7.5040	Hypothetical protein, conserved	-	-	66	72.5	0	70316	15	7	Gene name

Unknown	Tb927.3.1770	Hypothetical protein, conserved; leucine-rich repeat protein (LRRP), putative	-	+	67	21.1	0	101807	4	3	Gene name
Unknown	Tb927.4.5340	Hypothetical protein, conserved	-	+	68	35.1	0	95154	4	4	Gene name
Unknown	Tb10.05.0270	Hypothetical protein, conserved	-	+	76	39.8	0	117310	3	2	Gene name
Unknown	Tb10.6k15.3810	Hypothetical protein, conserved	-	+	82	90.5	0	36090	27	7	Gene name
Unknown	Tb11.52.0002	Hypothetical protein, conserved	-	-	83	70.3	0	439620	1	4	Gene name
Unknown	Tb927.5.1940	Hypothetical protein, conserved	-	+	86	74.8	0	21614	23	4	Gene name
Unknown	Tb11.01.6600	Hypothetical protein, conserved	-	-	88	48.7	0	79242	6	3	Gene name
Unknown	Tb927.6.1470	Hypothetical protein, conserved	-	-	90	29.9	0	27505	13	3	Gene name
Unknown	Tb927.6.4330	Hypothetical protein, conserved	-	-	94	35.6	0	24276	11	2	Gene name
Unknown	Tb11.01.2570	Hypothetical protein, conserved	-	+	99	29.6	0	30398	18	4	Gene name
Unknown	Tb10.61.1340	Hypothetical protein, conserved	-	-	105	64.7	0	77927	4	2	Gene name
Unknown	Tb10.70.6610	Hypothetical protein, conserved	-	-	110	93.1	0	27493	14	3	Gene name
Unknown	Tb10.6k15.1470	Hypothetical protein, conserved	-	-	112	22	0	26911	8	2	Gene name
Unknown	Tb927.1.3310	Hypothetical protein, conserved	-	+	161	35.1	0	28714	24	6	Gene name
Unknown	Tb10.26.0640	Hypothetical protein, conserved	-	+	171	71.4	0	25182	23	4	Gene name
Unknown	Tb927.4.2030	Hypothetical protein, conserved	-	-	179	97.3	0	22691	14	6	Gene name
Unknown	Tb927.4.2040	Hypothetical protein, conserved	-	-	221	97.7	0	20796	28	4	Gene name
Unknown	Tb927.7.2170	Hypothetical protein, conserved	-	-	363	66.4	0	169597	10	13	Gene name
Unknown	Tb09.142.0260	Hypothetical protein (pseudogene)	-	-	38	1.8	0	40061	5	3	Gene name
Unknown	Tb10.v4.0248	Hypothetical protein	-	-	73	89.9	0	35988	23	6	Gene name
Kinetoplast	Tb10.6k15.1600	Kinetoplast DNA-associated protein, putative	-	-	125	25.9	0	23603	17	3	Annotated as
Nucleus	Tb11.02.0620	NOG1 nucleolar GTP-binding protein 1	-	-	41	92.4	0	75220	4	3	Annotated as
Flagellum	Tb927.5.940	NADH-dependent fumarate reductase, putative	-	+	40	87.6	1	95546	0	1	TbFP
TbFP	Tb927.3.5510	Hypothetical protein, conserved	-	+	-	50.2	1	-	-	-	TbFP
TbFP	Tb927.7.6100	Hypothetical protein	-	+	-	83.5	1	-	-	-	TbFP
Membrane	Tb11.02.0530	PRS phosphoribosylpyrophosphate synthetase, putative	-	-	41	93.5	1	50237	2	1	GeneDB
Unknown	Tb05.5K5.510	Phosphoglycan beta 1,3 galactosyltransferase, degenerate	-	-	32	2.3	1	44695	5	2	GeneDB
Intracellular	Tb927.7.270	Ribosome biogenesis protein, putative	-	-	40	63.3	1	49841	1	1	Gene name
Plasma Membrane	Tb10.v4.0122	Variant surface glycoprotein (VSG, atypical), putative	-	-	33	17.5	1	57037	6	3	Gene name
Nucleus	Tb11.0820	Retrotransposon hot spot protein (RHS, pseudogene), putative; retrotransposon h	-	-	91	13.7	1	92596	1	1	Gene name
Unknown	Tb927.5.920	Hypothetical protein, conserved	-	-	31	81.5	1	57409	1	1	Gene name
Nucleus	Tb09.160.3820	Nucleolar RNA binding protein, putative	-	-	85	94.5	1	55540	8	2	Annotated as
TbFP	Tb927.8.4010	Fla1 flagellum-adhesion glycoprotein	-	+	-	75.1	2	-	-	-	TbFP
Unknown	Tb10.389.1340	Hypothetical protein, conserved	-	+	121	27	2	42089	15	4	Gene name
Unknown	Tb927.7.6670	Hypothetical protein, conserved	-	-	38	76.6	3	335263	0	2	Gene name

Appendix X

Localisation	Accession Number	Gene Annotation	Protein Mwt	Sequence Coverage (%)	Number of peptides identified	Protein score	gCAI Value (% Rank)	Number of predicted TMD's	Evidence
Centrosome	Tb927.8.1080	Centrin, putative	21778	10	1	64	96	0	GeneDB
Cytoplasm	Tb10.6k15.2290	BS2 protein disulfide isomerase; bloodstream-specific protein 2 precursor	55887	26	13	423	95.5	0	GeneDB
Cytoplasm	Tb11.01.6880	Cytosolic coat protein, putative	25663	13	3	121	50.3	2	GeneDB
Cytoplasm	Tb11.01.8470	Dihydrolipoyl dehydrogenase	51085	5	2	69	98.6	0	GeneDB
Cytoplasm	Tb11.01.5710	Phenylalanyl-tRNA synthetase alpha chain, putative	57305	5	3	32	83.5	0	GeneDB
Cytoplasm	Tb11.01.8510	TCP-1-alpha t-complex protein 1, alpha subunit, putative	54875	1	1	38	97.5	0	GeneDB
Cytoplasm	Tb11.42.0003	TCP-1-beta t-complex protein 1, beta subunit, putative	58363	3	1	39	90.6	0	GeneDB
Cytoplasm	Tb11.01.5860	TCP-1-epsilon t-complex protein 1, epsilon subunit, putative	60199	1	1	37	96.1	0	GeneDB
Cytoplasm	Tb927.3.3760	TRYP1 tryparedoxin	16025	24	3	123	97.4	0	GeneDB
Cytoplasm	Tb927.3.3780	Tryparedoxin	15995	24	3	123	97.5	0	GeneDB
Cytoplasm	Tb09.160.3270	Eukaryotic initiation factor 4a, putative	45447	8	2	114	98.6	0	GeneDB
Cytoplasm	Tb927.8.5710	Recombination initiation protein NBS1, putative	101283	3	3	57	36.5	0	GeneDB
Cytoplasm / ER	Tb11.01.1650	Signal recognition particle receptor alpha subunit, putative	64057	4	2	77	68.7	0	GeneDB
Cytoskeleton	Tb927.7.3160	Dynein heavy chain, cytosolic, putative	596281	0	4	55	64.3	0	Gene name
Cytoskeleton	Tb11.02.0030	Dynein heavy chain, putative	484662	1	5	53	92.5	0	Gene name
Cytoskeleton	Tb10.406.0615	Dynein light chain, putative	12438	28	3	65	80	0	Gene name
Cytoskeleton	Tb927.3.2020	Kinesin, putative	83184	2	2	30	76.2	0	Gene name
Cytoskeleton	Tb927.3.3390	Kinesin, putative	66375	3	2	32	61.9	0	Gene name
Cytoskeleton	Tb927.3.4960	Kinesin, putative	182380	2	3	32	52.1	0	Gene name
Cytoskeleton	Tb11.02.2260	Kinesin, putative; MCAK-like kinesin, putative; kinesin family member 6, putative	79050	4	3	54	39	0	Gene name
Cytoskeleton	Tb11.01.2530	Kinesin-like protein, putative	70582	18	12	430	84.1	0	Gene name
Cytoskeleton	Tb927.8.7840	Dynein heavy chain, cytosolic, putative	90695	1	1	44	53.4	0	GeneDB
Cytoskeleton	Tb10.61.0990	Kinesin, putative	189641	1	2	66	50.6	0	GeneDB
Cytoskeleton	Tb927.5.2090	OSM3-like kinesin, putative	123105	2	2	64	91.4	0	GeneDB
Cytoskeleton	Tb10.61.1750	TBKIFC1 C-terminal motor kinesin, putative	91485	8	5	165	95.7	0	GeneDB
Cytoskeleton	Tb927.5.2300	Formin, putative	101988	2	2	30	42.5	0	GeneDB
Endosomes	Tb927.8.4330	TbRAB11 small GTP-binding protein Rab11	23741	11	2	122	97.7	0	GeneDB
Endosomes	Tb927.6.3500	Endosomal trafficking protein RME-8, putative	256904	2	4	54	74.6	0	Annotated as
Endosomes / Lysosomes	Tb927.5.1810	Lysosomal/endosomal membrane protein p67; lysosomal membrane glycoprotein	73028	3	2	73	95.2	2	Annotated as
Endosomes / Lysosomes	Tb927.5.1830	Lysosomal/endosomal membrane protein p67; lysosomal membrane glycoprotein	73387	3	2	73	94.1	2	Annotated as
ER	Tb927.8.4890	Endoplasmic reticulum oxidoreductin, putative	49912	10	4	235	43.6	1	Annotated as
ER	Tb10.70.1190	VCP valosin-containing protein homolog; Transitional endoplasmic reticulum ATPase, putative	86570	7	4	38	94	0	Annotated as
ER	Tb11.02.5450	Glucose-regulated protein 78, putative; luminal binding protein 1 (BiP),	71505	44	35	1149	98.7	0	GeneDB
ER	Tb11.02.5500	Glucose-regulated protein 78, putative; luminal binding protein 1 (BiP),	71505	44	35	1149	98.7	0	GeneDB
ER	Tb927.6.3840	Reticulon domain protein	21285	12	3	133	95.1	3	GeneDB

ER / Golgi	Tb10.70.0120	COP-coated vesicle membrane protein erv25 precursor, putative; ER--golgi	28285	13	3	54	88	2	Annotated as
ER / Golgi	Tb11.55.0012	Vesicular-fusion protein SEC18, putative	98304	1	2	65	48.8	1	GeneDB
Flagellum	Tb927.8.4060	Flagellum-adhesion glycoprotein, putative	64947	10	6	161	73.4	2	Annotated as
Flagellum	Tb927.8.4110	Flagellum-adhesion glycoprotein, putative	64787	2	1	74	69.6	2	Annotated as
Flagellum	Tb10.6k15.1950	FTZC flagellum transition zone component	135112	3	4	62	39.5	0	Annotated as
Glycosome	Tb927.2.4210	Glycosomal phosphoenolpyruvate carboxykinase	58927	17	7	277	97.9	0	Annotated as
Glycosome	Tb927.1.3830	PGI glucose-6-phosphate isomerase, glycosomal	67744	32	16	474	74.9	0	Annotated as
Glycosome	Tb10.70.5820	HK1 hexokinase	51776	35	18	718	97.5	0	GeneDB
Glycosome	Tb10.70.5800	HK2 hexokinase	51630	28	16	783	97.7	0	GeneDB
Glycosome	Tb927.1.720	PGKA phosphoglycerate kinase	56054	15	7	184	62.8	0	GeneDB
Glycosome	Tb927.1.710	PGKB phosphoglycerate kinase	45560	33	12	433	93	0	GeneDB
Glycosome	Tb927.1.700	PGKC phosphoglycerate kinase	47216	46	13	526	95.2	0	GeneDB
Golgi / Lysosomes	Tb927.8.1870	tGLP1 Golgi/lysosome glycoprotein 1	68353	11	6	155	92.9	1	Annotated as
Intracellular	Tb927.3.5090	Tryparedoxin, putative	22394	32	8	315	59.1	1	Gene name
Intracellular	Tb11.02.4000	40S ribosomal protein S15a, putative	14864	20	3	98	97.9	0	Gene name
Intracellular	Tb927.3.1370	40S ribosomal protein S25, putative	12666	23	2	70	98.4	0	Gene name
Intracellular	Tb927.6.4690	60S ribosomal protein L9, putative	21705	12	2	54	97	0	Gene name
Intracellular	Tb10.70.7010	60S ribosomal protein L9, putative	21901	12	2	54	98.4	0	Gene name
Intracellular	Tb11.47.0035	Calpain-like cysteine peptidase, putative; antigen, putative; cysteine peptidase	665259	2	11	225	49.4	0	Gene name
Intracellular	Tb09.244.2720	Ribosomal protein L15, putative	26380	7	2	33	98.6	0	Gene name
Intracellular	Tb927.6.5040	Ribosomal protein L15, putative	24418	7	2	33	98.7	0	Gene name
Intracellular	Tb10.70.4800	Ribosomal protein S25, putative	12666	23	2	70	97.7	0	Gene name
Intracellular	Tb09.160.2490	Ribosomal protein S7, putative	23853	9	2	40	94.6	0	Gene name
Intracellular	Tb927.2.460	DNA-directed RNA polymerase, pseudogene, putative; DNA-directed RNA polymerase	74902	3	3	35	2.7	0	Gene name
Intracellular	Tb11.02.2930	SNF2 DNA repair protein, putative; SNF2 family protein	107104	0	1	31	52.2	0	Gene name
Intracellular	Tb927.8.5090	TRP11 DNA-directed RNA polymerase I largest subunit	197249	7	8	58	80	0	Gene name
Intracellular	Tb927.2.540	DNA-directed RNA polymerase, pseudogene, putative;	72024	6	4	32	2.5	2	Gene name
Intracellular	Tb927.1.2120	Calpain, putative; cysteine peptidase, Clan CA, family C2, putative	83537	21	13	555	69.2	0	GeneDB
Intracellular	Tb927.6.2370	Ubiquitin-protein ligase, putative	129734	1	2	35	70.7	1	GeneDB
Intracellular	Tb927.4.760	Gamma-adaptin 1, putative	88088	7	5	84	94.4	0	GeneDB
Intracellular	Tb927.5.4500	ADP-ribosylation factor, putative	21814	36	4	197	81.9	0	GeneDB
Intracellular	Tb927.7.3620	Tyrosyl-tRNA synthetase, putative	76609	1	2	88	93.9	0	Gene name
Intracellular	Tb11.01.3550	2-oxoglutarate dehydrogenase, E2 component, dihydrolipoamide succinyltransferase, putative	41117	9	4	31	98.4	0	GeneDB
Intracellular	Tb10.70.4880	Eukaryotic translation initiation factor 5, putative	42918	2	1	31	95	0	Gene name
Intracellular	Tb11.39.0006	Translation initiation factor eIF2B delta subunit, putative;	68539	1	1	32	82.1	0	Gene name
Intracellular	Tb10.70.0830	CHC clathrin heavy chain	192158	10	13	266	97.5	0	GeneDB
Kinetoplast	Tb927.8.7260	Kinetoplast-associated protein, putative	114493	8	7	177	62.4	0	Annotated as

Lysosomes	Tb927.8.6390	TbLysoPLA lysophospholipase, putative	30245	10	3	67	88.9	0	Annotated as
Mitochondrion	Tb10.70.5250	MCA4 metacaspase MCA4; cysteine peptidase, Clan CD, family C13, putative	39628	29	10	347	66	0	Gene name
Mitochondrion	Tb927.6.940	Metacaspase MCA2	38496	5	1	69	63.3	0	Gene name
Mitochondrion	Tb10.6k15.0960	54 NDH2 NADH dehydrogenase	54774	7	2	99	76	0	GeneDB
Mitochondrion	Tb927.2.3030	ATP-dependent Clp protease subunit, heat shock protein 78 (HSP78), putative	90951	9	6	163	90.7	0	GeneDB
Mitochondrion	Tb10.70.4280	Delta-1-pyrroline-5-carboxylate dehydrogenase, putative	62624	4	3	78	95.5	0	GeneDB
Mitochondrion	Tb11.02.5280	Glycerol-3-phosphate dehydrogenase, putative	67747	44	32	1113	95.6	0	GeneDB
Mitochondrion	Tb11.01.2000	hslVU complex proteolytic subunit, putative; hslVU complex proteolytic subunit	22898	10	2	126	96.2	0	GeneDB
Mitochondrion	Tb10.70.1750	KREPB2 RNA-editing complex protein; KREPB2	66597	1	1	34	37.4	0	GeneDB
Mitochondrion	Tb11.02.3130	Malic enzyme, putative	63650	11	4	55	89.6	0	GeneDB
Mitochondrion	Tb11.02.0730	mca1 metacaspase; cysteine peptidase, Clan CD, family C13	40895	5	2	133	90	1	GeneDB
Mitochondrion	Tb927.6.2420	p22 protein precursor	25391	10	2	143	87.4	0	GeneDB
Mitochondrion	Tb10.v4.0045	Prohibitin, putative	32366	31	7	200	92	1	GeneDB
Mitochondrion	Tb10.6k15.3080	Dihydrolipoamide acetyltransferase precursor, putative	48077	15	5	48	96.5	0	GeneDB
Mitochondrion	Tb927.4.4910	3,2-trans-enoyl-CoA isomerase, mitochondrial precursor, putative	45476	7	2	54	94.8	0	Annotated as
Mitochondrion	Tb927.3.860	Acyl carrier protein, mitochondrial precursor, putative	16577	21	2	92	85.8	0	Annotated as
Mitochondrion	Tb927.8.1420	acyl-CoA dehydrogenase, mitochondrial precursor, putative	56915	5	2	75	28.8	0	Annotated as
Mitochondrion	Tb927.7.7420	ATP synthase alpha chain, mitochondrial precursor	63862	15	9	307	97.8	0	Annotated as
Mitochondrion	Tb927.7.7430	ATP synthase alpha chain, mitochondrial precursor	63862	15	9	307	97.8	0	Annotated as
Mitochondrion	Tb927.3.1380	ATP synthase beta chain, mitochondrial precursor	55969	13	7	295	98.7	0	Annotated as
Mitochondrion	Tb11.55.0009	GBP21 mitochondrial RNA binding protein 1; gBP21, MRP1	23353	6	1	45	83.7	0	Annotated as
Mitochondrion	Tb927.6.3740	heat shock 70 kDa protein, mitochondrial precursor, putative	72000	22	13	550	97.6	0	Annotated as
Mitochondrion	Tb927.6.3750	heat shock 70 kDa protein, mitochondrial precursor, putative	72000	22	13	550	97.7	0	Annotated as
Mitochondrion	Tb927.6.3800	heat shock 70 kDa protein, mitochondrial precursor, putative	72000	22	13	550	97.7	0	Annotated as
Mitochondrion	Tb11.02.0250	Heat shock protein, mitochondrial precursor, putative; TNFR-associated pr	84832	15	8	364	97.2	0	Annotated as
Mitochondrion	Tb10.70.0430	HSP60 chaperonin Hsp60, mitochondrial precursor	59751	49	21	876	98.2	0	Annotated as
Mitochondrion	Tb927.4.3300	Mitochondrial ATP-dependent zinc metallopeptidase, putative	79451	1	1	50	65.6	1	Annotated as
Mitochondrion	Tb927.2.2970	Mitochondrial carrier protein, putative	34283	8	2	79	40.9	0	Annotated as
Mitochondrion	Tb927.7.3940	Mitochondrial carrier protein, putative; ADP/ATP mitochondrial translocase	36603	5	1	56	81.9	0	Annotated as
Mitochondrion	Tb10.389.0690	Mitochondrial carrier protein, putative; mitochondrial 2-oxoglutarate/malate carrier protein	33395	12	3	96	94.5	4	Annotated as
Mitochondrion	Tb927.8.2740	TbRBP38 mitochondrial RNA binding protein	38474	31	10	123	89.8	0	Annotated as
Mitochondrion	Tb11.02.2070	Long-chain-fatty acid-CoA ligase protein, putative; acyl-CoA synthetase,	73959	3	1	40	75.5	0	GeneDB
Mitochondrion	Tb09.160.0680	SLY1 sec1 family transport protein, putative	69321	13	5	102	71	0	GeneDB
Mitochondrion / Nucleus	Tb11.02.4730	PIF1 DNA repair and recombination helicase protein PIF1, putative; mitochondrial precursor	102443	0	1	34	86.5	0	GeneDB

Nucleus	Tb927.8.760	Nopp44/46 nucleolar RNA-binding protein	35022	5	2	43	78.1	0	Annotated as
Nucleus	Tb927.4.3840	Nucleolar protein, putative	59000	4	2	34	81.3	0	Annotated as
Nucleus	Tb927.8.740	Nucleolar RNA-binding protein, truncated	14875	12	2	45	24.4	0	Annotated as
Nucleus	Tb11.03.0140	Nucleoporin, putative; serine peptidase, Clan SP, family S59, putative 20	159537	1	2	41	91	0	Annotated as
Nucleus	Tb09.211.1510	SNAP50 small nuclear RNA gene activation protein (SNAP) 50, putative; small nuclear R	51958	1	1	39	55.8	0	Annotated as
Nucleus	Tb11.0050	Retrotransposon hot spot protein (RHS, pseudogene), putative	-	3	2	37	38.9	0	Gene name
Nucleus	Tb927.1.300	Retrotransposon hot spot protein (RHS, pseudogene), putative	252873	3	7	31	8.4	3	Gene name
Nucleus	Tb927.1.450	Retrotransposon hot spot protein (RHS, pseudogene), putative	240302	2	5	32	13.5	0	Gene name
Nucleus	Tb927.1.500	Retrotransposon hot spot protein (RHS, pseudogene), putative	213634	2	5	31	13.1	0	Gene name
Nucleus	Tb927.2.1050	Retrotransposon hot spot protein (RHS, pseudogene), putative	55552	19	10	32	3.1	1	Gene name
Nucleus	Tb927.2.480	Retrotransposon hot spot protein (RHS, pseudogene), putative	94992	5	4	32	3.9	0	Gene name
Nucleus	Tb927.6.5160	Retrotransposon hot spot protein (RHS, pseudogene), putative	67354	4	3	31	10.6	0	Gene name
Nucleus	Tb927.6.5170	Retrotransposon hot spot protein (RHS, pseudogene), putative	38994	6	2	34	12	0	Gene name
Nucleus	Tb11.02.1930	ATP-dependent RNA helicase, putative; DEAD/DEAH box RNA helicase, putativ	248230	2	4	32	80.1	0	GeneDB
Nucleus	Tb10.70.6220	DAC1 histone deacetylase 1; histone deacetylase-like 1 protein	43067	2	1	32	25.3	0	GeneDB
Nucleus	Tb927.8.3290	DNA polymerase zeta catalytic subunit, putative	220112	1	3	34	53.1	0	GeneDB
Nucleus	Tb10.61.2040	Fibrillarin, putative	24982	10	2	33	88.9	0	GeneDB
Nucleus	Tb927.3.3090	Helicase, putative	106100	3	3	31	55.6	0	GeneDB
Nucleus	Tb11.01.7810	minichromosome maintenance (MCM) complex subunit, putative	82124	0	1	31	54.8	0	GeneDB
Nucleus	Tb927.8.6840	MLH1 mismatch repair protein MLH1	97105	3	3	30	88.7	0	GeneDB
Nucleus	Tb09.211.2260	Protein kinase, putative	119500	2	2	34	28.5	0	GeneDB
Nucleus	Tb11.01.6260	RNA helicase, putative; DEAD/DEAH box helicase, putative	157905	0	1	39	73.8	1	GeneDB
Nucleus	Tb927.2.4710	RNA-binding protein, putative	50466	6	2	66	79.5	0	GeneDB
Nucleus	Tb10.406.0600	SMC2 structural maintenance of chromosome 2, putative	134550	4	7	39	80	0	GeneDB
Nucleus	Tb927.6.930	TbMCA4 metacaspase MCA3	39454	5	1	69	72.6	0	GeneDB
Nucleus / Cytoplasm	Tb927.7.2500	Proteasome regulatory ATPase subunit 1	48730	1	1	39	94	0	GeneDB
Peroxisome	Tb09.160.0620	Peroxisomal membrane protein 4, putative	34973	6	2	103	67.5	3	Annotated as

Appendix XI

Localisation	Accession Number	Gene Annotation	Protein Mwt	Sequence Coverage (%)	Number of peptides identified	Protein score	gCAI Value (% Rank)	Number of predicted TMD's	Evidence
Membrane	Tb10.70.4750	Calcium channel protein, putative	305238	17	39	1118	63.8	22	5+ TMD
Membrane	Tb927.4.4380	PPase1 vacuolar-type proton translocating pyrophosphatase 1, putative	85880	1	3	76	97.6	14	5+ TMD
Membrane	Tb927.8.7980	TVP1 vacuolar-type proton translocating pyrophosphatase 1	85948	1	3	76	97.3	14	5+ TMD
Unknown	Tb927.3.4080	Hypothetical protein, conserved	63290	9	4	131	93.4	14	Gene name
Unknown	Tb927.3.4100	Hypothetical protein, conserved	63951	9	4	131	92.2	14	Gene name
Unknown	Tb927.3.4090	Hypothetical protein, conserved	63594	18	11	37	93.7	13	Gene name
Unknown	Tb927.3.4070	Hypothetical protein, conserved	64968	6	3	167	91.8	12	Gene name
Unknown	Tb927.8.4360	Hypothetical protein, conserved	153116	6	7	199	37.9	12	Gene name
Membrane	Tb927.3.590	Adenosine transporter, putative	50720	6	5	31	92	11	5+ TMD
Membrane	Tb927.4.4830	Amino acid transporter, putative	48495	5	2	71	91.1	11	5+ TMD
Membrane	Tb927.4.4850	Amino acid transporter, putative	48495	5	2	71	91.1	11	5+ TMD
Membrane	Tb927.4.4870	Amino acid transporter, putative	48625	5	2	71	91.7	11	5+ TMD
Membrane	Tb927.5.900	Oligosaccharyl transferase subunit, putative	93085	7	5	146	86.1	11	5+ TMD
Membrane	Tb927.2.6150	TbNT2/927 adenosine transporter 2	51198	7	5	165	85.8	11	5+ TMD
Membrane	Tb927.2.6200	TbNT3 adenosine transporter 2, putative	51584	8	5	182	80	11	5+ TMD
Membrane	Tb927.2.6240	TbNT5 adenosine transporter 2	51302	8	5	182	71.5	11	5+ TMD
Membrane	Tb927.2.6320	TbNT6 adenosine transporter 2, putative	50823	11	7	108	77.2	11	5+ TMD
Membrane	Tb927.2.6280	TbNT7 adenosine transporter 2, putative	51795	8	5	182	82.2	11	5+ TMD
Membrane	Tb927.5.3400	Calcium-translocating P-type ATPase; calcium pump	111331	37	34	1374	94.8	10	5+ TMD
Membrane	Tb927.5.890	Oligosaccharyl transferase subunit, putative	90031	8	6	225	85.1	10	5+ TMD
Membrane	Tb927.5.910	Oligosaccharyl transferase subunit, putative	93206	4	4	133	85.4	10	5+ TMD
Membrane	Tb927.8.1200	TbA2 vacuolar-type Ca2+-ATPase 2	119892	16	16	745	78.9	10	5+ TMD
Membrane	Tb927.2.6220	TbNT4 adenosine transporter 2, putative	51648	7	4	132	61.7	10	5+ TMD
Membrane	Tb927.8.1160	Vacuolar-type Ca2+-ATPase, putative	122134	0	1	31	67.8	10	5+ TMD
Unknown	Tb927.8.1460	Hypothetical protein, conserved	45038	10	5	71	73	10	Gene name
Membrane	Tb11.02.1100	NT8.1 nucleobase/nucleoside transporter 8.1; nucleobase transporter	48234	12	3	147	93.4	9	5+ TMD
Membrane	Tb11.02.1105	NT8.1 nucleobase/nucleoside transporter 8.1; nucleobase transporter	48248	12	3	147	93.6	9	5+ TMD
Membrane	Tb11.02.1106	Nucleobase transporter, putative	48884	7	2	77	92.1	9	5+ TMD
Membrane	Tb11.02.4100	Pretranslocation protein, alpha subunit, putative; SEC61-like	54069	12	7	293	96	9	5+ TMD
Unknown	Tb11.01.6990	Hypothetical protein, conserved	71335	8	4	56	52.9	9	Gene name
Membrane	Tb09.244.2570	Calcium motive p-type ATPase, putative	115271	17	12	361	96.1	8	5+ TMD
Membrane	Tb927.8.650	Cation-transporting ATPase, putative	142132	7	6	132	78.6	8	5+ TMD
Membrane	Tb927.4.4490	MRPE; PGPA multidrug resistance protein E; P-glycoprotein	195164	1	2	32	82.9	8	5+ TMD
Membrane	Tb927.8.1180	TbA1 vacuolar-type Ca2+-ATPase 1	122578	16	16	717	78.8	8	5+ TMD
Unknown	Tb09.211.1830	Hypothetical protein, conserved	147436	5	5	121	37.8	8	Gene name
Unknown	Tb10.389.0140	Hypothetical protein, conserved	46166	2	1	51	78.7	8	Gene name
Membrane	Tb927.6.3550	Phospholipid-translocating P-type ATPase (flippase), putative	134073	3	3	37	59	7	5+ TMD
Membrane	Tb10.406.0290	Protein tyrosine phosphatase, putative	29826	9	3	142	88.8	7	5+ TMD
Membrane	Tb10.389.1170	P-type H+-ATPase, putative	101597	31	29	1138	95.9	7	5+ TMD
Membrane	Tb10.389.1180	P-type H+-ATPase, putative	101075	23	19	816	95.8	7	5+ TMD
Unknown	Tb10.6k15.1690	Hypothetical protein, conserved	67074	7	3	96	78.7	7	Gene name
Unknown	Tb10.70.5220	Hypothetical protein, conserved	64592	8	3	41	28.3	7	Gene name
Membrane	Tb10.61.2650	Aquaglyceroporin (small solute channel), putative	33016	3	1	75	91.8	6	5+ TMD
Membrane	Tb10.61.2640	Aquaporin 9, putative	33572	4	1	41	89.6	6	5+ TMD
Membrane	Tb927.5.1300	Vacuolar proton translocating ATPase subunit A, putative	90372	5	2	69	90.5	6	5+ TMD
Unknown	Tb11.02.3570	Hypothetical protein, conserved	32246	3	1	34	91.6	6	Gene name
Unknown	Tb927.1.4500	Hypothetical protein, conserved	32960	4	1	67	57.7	6	Gene name
Membrane	Tb09.211.0680	CAAX prenyl protease 1, putative; metallo-peptidase, Clan M- Family M4	49045	20	8	369	94.2	5	5+ TMD
Membrane	Tb927.7.4180	Fatty acid elongase, putative	34077	11	4	109	94	5	5+ TMD
Membrane	Tb10.100.0090	Vacuolar ATP synthase, putative	19538	7	1	51	94.6	5	5+ TMD
Unknown	Tb10.70.6830	Hypothetical protein, conserved	37944	9	3	113	94.2	5	Gene name
Unknown	Tb927.3.3820	Hypothetical protein, conserved	45772	2	1	42	75.5	5	Gene name
Unknown	Tb927.7.7320	Hypothetical protein, conserved	81430	3	2	41	70	5	Gene name

Unknown	Tb927.8.2460	Hypothetical protein, conserved	31167	16	3	110	83.1	5	Gene name
Membrane	Tb927.4.4420	Adenylyl cyclase, pseudogene, putative; adenylyl cyclase, frameshift	143846	3	5	73	10.7	4	Gene name
Membrane	Tb927.6.270	Receptor-type adenylate cyclase GRESAG 4, pseudogene, putative	139855	3	5	34	3.9	4	Gene name
Membrane	Tb927.8.7860	Receptor-type adenylate cyclase GRESAG 4, putative	145860	4	5	44	36	4	Gene name
Membrane	Tb927.8.2100	Vacuolar ATP synthase 16 kDa proteolipid subunit, putative	17085	9	1	52	96	4	GeneDB
Unknown	Tb09.211.0220	Rhomboid-like protein; serine peptidase, Clan S- , family S54, putative	37048	4	1	41	55.3	4	GeneDB
Unknown	Tb11.01.0110	Hypothetical protein, conserved	32630	4	1	41	45.5	4	Gene name
Unknown	Tb927.4.2530	Hypothetical protein, conserved	16881	24	6	213	98.4	4	Gene name
Unknown	Tb927.7.1420	Hypothetical protein, conserved	176218	6	8	222	70.2	4	Gene name
Unknown	Tb927.7.4470	Hypothetical protein, conserved	22441	12	2	105	28.1	4	Gene name
Membrane	Tb927.6.390	Adenylyl cyclase, pseudogene, putative; adenylyl cyclase	139296	7	6	44	4.5	3	Gene name
Plasma Membrane	Tb11.03.0030	ABCD3 ABC transporter, putative	76261	9	5	170	54.7	3	GeneDB
Unknown	Tb10.61.3180	CaLB calcium-dependent lipid binding protein, putative; synaptotagmin, putative	67665	4	2	54	84.9	3	GeneDB
Unknown	Tb09.211.1220	Hypothetical protein, conserved	38895	5	2	46	45.3	3	Gene name
Unknown	Tb10.6k15.2410	Hypothetical protein, conserved	30619	3	1	44	39.4	3	Gene name
Unknown	Tb10.6k15.3930	Hypothetical protein, conserved	34575	13	3	52	90	3	Gene name
Unknown	Tb10.70.1640	Hypothetical protein, conserved	68331	12	5	146	37.9	3	Gene name
Unknown	Tb10.70.2450	Hypothetical protein, conserved	27369	9	2	54	95.2	3	Gene name
Unknown	Tb927.4.5080	Hypothetical protein, conserved	14743	13	1	44	25.1	3	Gene name
Unknown	Tb927.5.2860	Hypothetical protein, conserved	46936	2	1	31	55.8	3	Gene name
Membrane	Tb11.01.4701	MBAP1 membrane-bound acid phosphatase 1 precursor	60479	7	4	81	67.6	2	Annotated as
Membrane	Tb927.7.7520	Receptor-type adenylate cyclase GRESAG 4, putative	138933	7	6	136	51.9	2	Gene name
Membrane	Tb927.7.7530	Receptor-type adenylate cyclase GRESAG 4, putative	139003	7	6	136	51.8	2	Gene name
Membrane	Tb927.8.7870	Receptor-type adenylate cyclase GRESAG 4, putative	140838	7	8	276	43.6	2	Gene name
Membrane	Tb927.8.7900	Receptor-type adenylate cyclase GRESAG 4, putative	142111	12	12	385	46.6	2	Gene name
Membrane	Tb927.8.7920	Receptor-type adenylate cyclase GRESAG 4, putative	142118	8	10	308	47.8	2	Gene name
Membrane	Tb927.8.7930	Receptor-type adenylate cyclase GRESAG 4, putative	142391	9	10	358	45.8	2	Gene name
Membrane	Tb927.8.7940	Receptor-type adenylate cyclase GRESAG 4, putative	139830	9	11	382	40.1	2	Gene name
Membrane	Tb09.v4.0009	Receptor-type adenylate cyclase, putative; adenylyl cyclase, putative	138301	2	2	80	91	2	Gene name
Membrane	Tb11.01.7400	GPI transamidase component Tta1	42288	6	2	71	87.1	2	GeneDB
Membrane	Tb11.01.1780	Short-chain dehydrogenase, putative	34279	5	1	53	36.2	2	GeneDB
Plasma Membrane	Tb927.5.390	75 kDa invariant surface glycoprotein, putative	58157	9	5	134	53.6	2	GeneDB
Plasma Membrane	Tb927.4.1920	Glycosylphosphatidylinositol (GPI) anchor, putative	76557	7	4	101	78.2	2	Gene name
Unknown	Tb09.211.2020	Synaptojanin (N-terminal domain), putative	83905	1	1	43	50.2	2	GeneDB
Unknown	Tb927.5.610	Acidic phosphatase, putative	47653	2	1	45	51.7	2	GeneDB
Unknown	Tb927.4.4210	ATP-dependent zinc metallopeptidase, putative	96448	7	4	66	86.9	2	GeneDB
Unknown	Tb927.7.160	Expression site-associated gene (ESAG, pseudogene), putative	58916	3	2	44	0.4	2	GeneDB
Unknown	Tb09.244.2380	Expression site-associated gene 4 (ESAG4) protein, putative	138326	2	2	80	91.3	2	GeneDB
Unknown	Tb927.5.1210	Short-chain dehydrogenase, putative	34083	34	8	353	95.6	2	GeneDB
Unknown	Tb927.5.3710	Sphingomyelin phosphodiesterase, putative	64978	5	3	39	42.4	2	GeneDB
Unknown	Tb11.01.4960	Regulator of nonsense transcripts 1, putative (pseudogene); nonsense mRNA	237335	2	7	44	8.8	2	GeneDB
Unknown / Membrane	Tb09.211.3650	Phospholipase A2-like protein, putative	49879	8	4	32	56	2	GeneDB
Unknown	Tb05.5K5.210	Hypothetical protein	89801	9	7	232	67.7	2	Gene name
Unknown	Tb05.5K5.220	Hypothetical protein	89464	10	9	365	76.6	2	Gene name
Unknown	Tb09.160.2370	Hypothetical protein, conserved	24334	18	3	68	58.7	2	Gene name
Unknown	Tb09.v1.0650	Hypothetical protein, conserved	62906	8	3	109	69.4	2	Gene name
Unknown	Tb10.61.1070	Hypothetical protein, conserved	205801	3	5	42	17.8	2	Gene name
Unknown	Tb11.01.1330	Hypothetical protein, conserved	33340	12	3	33	85.2	2	Gene name
Unknown	Tb11.02.3760	Hypothetical protein, conserved	11271	21	2	45	97.8	2	Gene name
Unknown	Tb11.02.3770	Hypothetical protein, conserved	15558	16	2	45	92.9	2	Gene name
Unknown	Tb927.2.1700	Hypothetical protein, conserved	85062	6	3	59	47.7	2	Gene name
Unknown	Tb927.3.2420	Hypothetical protein, conserved	14252	19	5	34	88.4	2	Gene name

Unknown	Tb927.3.4950	Hypothetical protein, conserved	52749	20	7	57	53.7	2	Gene name
Unknown	Tb927.5.4570	Hypothetical protein, conserved	89667	9	7	232	72.2	2	Gene name
Unknown	Tb927.5.4580	Hypothetical protein, conserved	89434	10	9	365	77.1	2	Gene name
Unknown	Tb927.6.2600	Hypothetical protein, conserved	16099	12	2	31	93.1	2	Gene name
Unknown	Tb927.6.4180	Hypothetical protein, conserved	16317	24	2	106	83.9	2	Gene name
Unknown	Tb927.7.1400	Hypothetical protein, conserved	60830	7	3	91	33.5	2	Gene name
Unknown	Tb927.7.2190	Hypothetical protein, conserved	29338	16	5	170	83	2	Gene name
Membrane	Tb11.02.5420	CPR NADPH--cytochrome p450 reductase, putative	70570	4	3	53	68	1	Gene name
Membrane	Tb927.4.4460	GRESAG 4.4B receptor-type adenylate cyclase GRESAG 4, putative; receptor-type adenylat	145252	4	4	61	35.4	1	Gene name
Membrane	Tb927.8.7910	Receptor-type adenylate cyclase GRESAG 4 (pseudogene), putative	134802	7	8	269	21.9	1	Gene name
Membrane	Tb927.7.6050	Receptor-type adenylate cyclase GRESAG 4, putative	140200	2	2	32	76.9	1	Gene name
Membrane	Tb927.4.4410	Receptor-type adenylate cyclase GRESAG 4, putative	139307	10	9	169	49	1	Gene name
Membrane	Tb927.4.4430	Receptor-type adenylate cyclase GRESAG 4, putative	139320	4	4	61	46.7	1	Gene name
Membrane	Tb927.4.4440	Receptor-type adenylate cyclase GRESAG 4, putative	139753	4	4	61	47.5	1	Gene name
Membrane	Tb927.4.4450	Receptor-type adenylate cyclase GRESAG 4, putative	139040	4	4	61	43.2	1	Gene name
Membrane	Tb927.4.4470	Receptor-type adenylate cyclase GRESAG 4, putative	143017	4	4	60	31.3	1	Gene name
Membrane	Tb927.7.6070	Receptor-type adenylate cyclase GRESAG 4, putative	140342	7	6	76	78.3	1	Gene name
Membrane	Tb927.8.7590	Receptor-type adenylate cyclase GRESAG 4, putative	140691	3	15	197	64	1	Gene name
Membrane	Tb927.5.3560	VAMP vesicle-associated membrane protein, putative; synaptobrevin, putative	24672	5	1	41	82.1	1	Gene name
Membrane	Tb11.01.6800	1-acyl-sn-glycerol-3-phosphate acyltransferase protein, putative	30682	10	2	58	84.3	1	GeneDB
Membrane	Tb11.01.0170	CPR NADPH--cytochrome P450 reductase, putative	71586	5	4	64	75.7	1	GeneDB
Membrane	Tb11.01.6360	Metalloprotease, putative; cell division protein FtsH homologue, putative	74771	4	2	46	72.1	1	GeneDB
Membrane	Tb927.5.3320	Protein kinase, putative	136785	4	4	43	76.1	1	GeneDB
Membrane	Tb10.61.1980	Putative syntaxin	36500	4	2	53	37.5	1	GeneDB
Plasma Membrane	Tb09.v4.0145	Variant surface glycoprotein (VSG, pseudogene), putative	50412	16	6	54	19.9	1	Gene name
Plasma Membrane	Tb10.v4.0026	Variant surface glycoprotein (VSG, pseudogene), putative	45907	5	3	40	15.1	1	Gene name
Plasma Membrane	Tb927.5.4980	Variant surface glycoprotein (VSG, pseudogene), putative	53967	14	11	457	45.6	1	Gene name
Plasma Membrane	Tb927.2.3270	65 kDa invariant surface glycoprotein	48696	14	6	159	25.1	1	GeneDB
Plasma Membrane	Tb927.2.3280	65 kDa invariant surface glycoprotein	48594	14	6	159	23.3	1	GeneDB
Plasma Membrane	Tb927.2.3290	65 kDa invariant surface glycoprotein	48498	14	6	159	24.5	1	GeneDB
Plasma Membrane	Tb927.2.3300	65 kDa invariant surface glycoprotein	48512	14	6	159	24	1	GeneDB
Plasma Membrane	Tb927.2.3310	65 kDa invariant surface glycoprotein	48468	14	6	146	23.5	1	GeneDB
Plasma Membrane	Tb927.2.3320	65 kDa invariant surface glycoprotein	48466	3	1	52	38.3	1	GeneDB
Plasma Membrane	Tb927.5.1390	65 kDa invariant surface glycoprotein, putative	47213	8	4	94	22.2	1	GeneDB
Plasma Membrane	Tb927.5.360	ISG75 75 kDa invariant surface glycoprotein	58646	21	9	175	42.1	1	GeneDB
Plasma Membrane	Tb927.5.350	ISG75 75 kDa invariant surface glycoprotein, putative	59128	4	3	77	42.5	1	GeneDB
Plasma Membrane	Tb927.5.370	ISG75 75 kDa invariant surface glycoprotein, putative	59069	7	5	67	51.6	1	GeneDB
Unknown	Tb09.160.3090	Heat shock protein, putative; HSP70-like protein	91357	13	9	276	81.3	1	Gene name
Unknown	Tb10.70.0170	Chaperone protein DNAJ, putative	31350	14	3	77	60.2	1	GeneDB
Unknown	Tb10.70.3240	Short-chain dehydrogenase, putative	38591	8	2	73	78.5	1	GeneDB
Unknown	Tb927.5.630	Acidic phosphatase, putative	44264	13	7	266	74	1	GeneDB
Unknown	Tb927.8.7410	Calreticulin, putative	44994	57	27	167	97.8	1	GeneDB
Unknown	Tb927.4.5010	Calreticulin, putative	45242	14	4	150	97.6	1	GeneDB
Unknown	Tb11.01.0290	Carbonic anhydrase-like protein	46148	5	2	41	90.6	1	GeneDB
Unknown	Tb927.7.6860	Expression site-associated gene (ESAG) protein, putative	52337	5	3	61	58.5	1	GeneDB
Unknown	Tb09.244.1010	Expression site-associated gene 3 (ESAG3, pseudogene), putative	36731	11	3	32	1	1	GeneDB
Unknown	Tb927.4.2230	Glycosyltransferase ALG2, putative	57662	5	2	68	67.4	1	GeneDB
Unknown	Tb927.3.3580	LPG3 lipophosphoglycan biosynthetic protein, putative	87712	24	16	512	96.4	1	GeneDB

Unknown	Tb09.211.2460	Hypothetical protein, conserved	101940	4	3	66	17.3	1	Gene name
Unknown	Tb09.211.2530	Hypothetical protein, conserved	41696	10	2	79	47.2	1	Gene name
Unknown	Tb09.v1.0510	Hypothetical protein, conserved	32616	2	1	35	11.6	1	Gene name
Unknown	Tb09.v1.0540	Hypothetical protein, conserved	32616	2	1	35	11.6	1	Gene name
Unknown	Tb10.05.0040	Hypothetical protein, conserved	113901	0	1	34	58.9	1	Gene name
Unknown	Tb10.05.0050	Hypothetical protein, conserved	18500	4	1	39	91.8	1	Gene name
Unknown	Tb10.389.0310	Hypothetical protein, conserved	97290	1	2	39	19.2	1	Gene name
Unknown	Tb10.61.2720	Hypothetical protein, conserved	22373	5	1	38	22.3	1	Gene name
Unknown	Tb10.6k15.2270	Hypothetical protein, conserved	30993	8	2	33	23	1	Gene name
Unknown	Tb11.01.1980	Hypothetical protein, conserved	12901	16	2	95	14.7	1	Gene name
Unknown	Tb11.01.2460	Hypothetical protein, conserved	45430	14	5	245	70	1	Gene name
Unknown	Tb11.01.4740	Hypothetical protein, conserved	61457	12	5	85	87.7	1	Gene name
Unknown	Tb11.01.5120	Hypothetical protein, conserved	105831	3	4	40	50	1	Gene name
Unknown	Tb11.02.0800	Hypothetical protein, conserved	166840	13	15	357	60.7	1	Gene name
Unknown	Tb11.02.5660	Hypothetical protein, conserved	46550	21	5	257	85.2	1	Gene name
Unknown	Tb927.3.3130	Hypothetical protein, conserved	179339	1	3	55	69.9	1	Gene name
Unknown	Tb927.3.3650	Hypothetical protein, conserved	69219	6	3	58	60.1	1	Gene name
Unknown	Tb927.3.5350	Hypothetical protein, conserved	11649	18	2	99	36.4	1	Gene name
Unknown	Tb927.4.1390	Hypothetical protein, conserved	103631	1	1	41	88	1	Gene name
Unknown	Tb927.4.1540	Hypothetical protein, conserved	55776	10	3	126	81.9	1	Gene name
Unknown	Tb927.4.1600	Hypothetical protein, conserved	27046	21	7	46	38	1	Gene name
Unknown	Tb927.4.590	Hypothetical protein, conserved	88425	11	8	232	79.8	1	Gene name
Unknown	Tb927.5.1930	Hypothetical protein, conserved	24400	6	2	46	91.4	1	Gene name
Unknown	Tb927.5.310	Hypothetical protein	45019	10	4	71	53.3	1	Gene name
Unknown	Tb927.5.860	Hypothetical protein, conserved	12370	7	1	56	90.7	1	Gene name
Unknown	Tb927.6.1850	Hypothetical protein, conserved	59744	4	2	45	43.3	1	Gene name
Unknown	Tb927.6.4320	Hypothetical protein, conserved	44510	9	3	107	86.9	1	Gene name
Unknown	Tb927.7.5700	Hypothetical protein, conserved	45807	11	5	205	32.1	1	Gene name
Unknown	Tb927.7.900	Hypothetical protein, conserved	64872	12	5	148	82	1	Gene name
Unknown	Tb927.8.1290	Hypothetical protein, conserved	121121	3	4	67	49.1	1	Gene name
Unknown	Tb927.8.3050	Hypothetical protein, conserved	62728	5	2	68	94.6	1	Gene name
Unknown	Tb927.8.3540	Hypothetical protein, conserved	107424	5	5	99	58.7	1	Gene name
Unknown	Tb927.8.5760	Hypothetical protein, conserved	54459	6	3	75	89.4	1	Gene name
Unknown	Tb927.8.7720	Hypothetical protein, conserved	25114	12	2	121	19	1	Gene name
Membrane	Tb10.61.3060	GPI-anchor transamidase subunit 8 (GPI8); cysteine peptidase, Clan CD, family C13	37231	4	2	59	87.7	0	Gene name
Membrane	Tb927.2.6000	GPI-PLC glycosylphosphatidylinositol-specific phospholipase C	41093	13	4	181	15.9	0	Gene name
Membrane	Tb927.4.130	Receptor-type adenylate cyclase GRESAG 4, pseudogene, putative;	109850	5	5	38	14	0	Gene name
Membrane	Tb10.70.6790	Mismatch repair protein, putative	128951	2	3	31	76.1	0	GeneDB
Membrane	Tb09.160.2770	ACS1 fatty acyl CoA synthetase 1	79765	6	5	140	74.3	0	GeneDB
Membrane	Tb09.160.2810	ACS3 fatty acyl CoA synthetase 3	78603	15	13	353	93.4	0	GeneDB
Membrane	Tb09.160.2840	ACS4 fatty acyl CoA synthetase 4	78507	10	8	186	93.7	0	GeneDB
Membrane	Tb10.100.0070	ATP synthase F1 subunit gamma protein, putative	34463	13	2	74	72	0	GeneDB
Membrane	Tb927.5.3220	Signal peptidase type I, putative	23862	18	3	147	53.2	0	GeneDB
Plasma Membrane	Tb09.354.0090	Variant surface glycoprotein (VSG), putative	52791	1	1	32	57.7	0	Gene name
Plasma Membrane	Tb927.1.5300	Variant surface glycoprotein (VSG), putative	50179	3	2	39	37.5	0	Gene name
Plasma Membrane	Tb09.244.0020	Variant surface glycoprotein (VSG, pseudogene), putative	58551	3	4	36	17.1	0	Gene name
Plasma Membrane	Tb09.244.0610	Variant surface glycoprotein (VSG, pseudogene), putative	59803	14	13	35	11.8	0	Gene name
Plasma Membrane	Tb09.244.0630	Variant surface glycoprotein (VSG, pseudogene), putative	55980	5	4	43	15.7	0	Gene name
Plasma Membrane	Tb09.244.0960	Variant surface glycoprotein (VSG, pseudogene), putative	57403	7	5	42	13	0	Gene name
Plasma Membrane	Tb09.244.1310	Variant surface glycoprotein (VSG, pseudogene), putative	54302	1	1	32	13.7	0	Gene name
Plasma Membrane	Tb09.354.0240	Variant surface glycoprotein (VSG, pseudogene), putative	56892	3	2	34	23.1	0	Gene name
Plasma Membrane	Tb09.v4.0068	Variant surface glycoprotein (VSG, pseudogene), putative	34621	6	2	32	9.4	0	Gene name
Plasma Membrane	Tb09.v4.0081	Variant surface glycoprotein (VSG, pseudogene), putative	56244	3	2	35	49.6	0	Gene name
Plasma Membrane	Tb09.v4.0190	Variant surface glycoprotein (VSG, pseudogene), putative	50064	8	5	37	33.3	0	Gene name
Plasma Membrane	Tb10.v4.0020	Variant surface glycoprotein (VSG, pseudogene), putative	56149	3	3	33	34.2	0	Gene name
Plasma Membrane	Tb11.57.0030	Variant surface glycoprotein (VSG, pseudogene), putative	54072	8	6	32	12.6	0	Gene name
Plasma Membrane	Tb11.v4.0026	Variant surface glycoprotein (VSG, pseudogene), putative	58823	3	3	39	7.7	0	Gene name

Plasma Membrane	Tb11.v4.0039	Variant surface glycoprotein (VSG, pseudogene), putative	-	5	3	32	6.6	0	Gene name
Plasma Membrane	Tb11.v4.0044	Variant surface glycoprotein (VSG, pseudogene), putative	53653	4	3	32	13.1	0	Gene name
Plasma Membrane	Tb11.v4.0046	Variant surface glycoprotein (VSG, pseudogene), putative	52590	5	4	46	15.5	0	Gene name
Plasma Membrane	Tb927.3.5850	Variant surface glycoprotein (VSG, pseudogene), putative	51906	4	2	38	6.5	0	Gene name
Plasma Membrane	Tb927.4.5620	Variant surface glycoprotein (VSG, pseudogene), putative	57897	9	5	50	7.2	0	Gene name
Plasma Membrane	Tb927.5.4850	Variant surface glycoprotein (VSG, pseudogene), putative	57973	4	2	34	3.6	0	Gene name
Plasma Membrane	Tb927.5.5290	Variant surface glycoprotein (VSG, pseudogene), putative	59767	5	3	30	1.6	0	Gene name
Plasma Membrane	Tb927.5.5300	Variant surface glycoprotein (VSG, pseudogene), putative	54286	4	2	30	14.6	0	Gene name
Plasma Membrane	Tb927.5.5320	Variant surface glycoprotein (VSG, pseudogene), putative	55607	3	5	34	23.1	0	Gene name
Plasma Membrane	Tb927.6.5390	Variant surface glycoprotein (VSG, pseudogene), putative	57910	4	3	32	71	0	Gene name
Plasma Membrane	Tb927.7.120	Variant surface glycoprotein (VSG, pseudogene), putative	57796	3	2	45	13.4	0	Gene name
Plasma Membrane	Tb927.7.140	Variant surface glycoprotein (VSG, pseudogene), putative	59082	7	7	52	5.1	0	Gene name
Plasma Membrane	Tb927.3.310	Variant surface glycoprotein (VSG, pseudogene), putative	59176	5	3	34	10.2	0	Gene name
Plasma Membrane	Tb927.3.500	Variant surface glycoprotein (VSG, pseudogene), putative	58631	2	4	40	5.5	0	Gene name
Plasma Membrane	Tb927.5.180	Variant surface glycoprotein (VSG, pseudogene), putative	57462	2	1	33	9.3	0	Gene name
Plasma Membrane	Tb927.5.4800	Variant surface glycoprotein (VSG, pseudogene), putative	21639	11	5	32	16.1	0	Gene name
Unknown	Tb09.160.4570	AK arginine kinase	41856	18	5	199	96.9	0	Gene name
Unknown	Tb09.160.4590	AK arginine kinase	40457	18	5	198	97.1	0	Gene name
Unknown	Tb927.6.4710	Calmodulin, putative	74990	12	7	303	76.4	0	Gene name
Unknown	Tb10.26.1080	Heat shock protein 83	81169	9	5	218	99.2	0	Gene name
Unknown	Tb927.5.1520	Heat shock protein HslVU, ATPase subunit HslU, putative	52612	7	3	78	93.2	0	Gene name
Unknown	Tb927.7.1320	HSP10 10 kDa heat shock protein, putative	10664	70	6	240	97.1	0	Gene name
Unknown	Tb927.7.1340	HSP10 10 kDa heat shock protein, putative	10664	70	6	240	97.1	0	Gene name
Unknown	Tb927.2.5980	HSP100 ATP-dependent Clp protease subunit, heat shock protein 100 (HSP100), putative	97263	3	3	40	92.8	0	Gene name
Unknown	Tb927.7.710	HSP70 heat shock 70 kDa protein, putative	70168	4	3	37	90.8	0	Gene name
Unknown	Tb927.1.480	Leucine-rich repeat protein (LRRP), putative	163922	3	5	36	17.2	0	Gene name
Unknown	Tb927.1.5030	Leucine-rich repeat protein (LRRP), putative	82206	8	5	172	23.7	0	Gene name
Unknown	Tb927.7.7110	Leucine-rich repeat protein (LRRP), putative	84162	4	3	32	54.8	0	Gene name
Unknown	Tb09.244.1980	Leucine-rich repeat protein (LRRP, pseudogene), putative	-	2	4	38	5.4	0	Gene name
Unknown	Tb927.2.1310	Leucine-rich repeat protein (LRRP, pseudogene), putative	164434	2	4	42	8.4	0	Gene name
Unknown	Tb927.4.160	Leucine-rich repeat protein (LRRP, pseudogene), putative	167614	3	5	44	5.5	0	Gene name
Unknown	Tb09.211.4960	Leucine-rich repeat protein (LRRP, pseudogene), putative; point mutation	164550	2	3	32	10.1	0	Gene name
Unknown	Tb927.1.4830	Phospholipase A1, putative	32782	4	1	66	26.5	0	GeneDB
Unknown	Tb10.70.4200	Fatty acyl CoA synthetase, putative; Long-chain- fatty-acid--CoA ligase 4	80017	17	11	270	76.3	0	GeneDB
Unknown	Tb10.61.0150	Inosine-5'-monophosphate dehydrogenase; IMP dehydrogenase	48964	12	6	276	95.6	0	GeneDB
Unknown	Tb927.8.7170	Inositol polyphosphate 1-phosphatase, putative	42055	2	1	39	91.8	0	GeneDB
Unknown	Tb10.6k15.3990	VPS45 vacuolar protein sorting-associated protein 45, putative	65675	3	2	38	85.9	0	GeneDB
Unknown	Tb11.02.2550	AAA ATPase, putative	59658	3	2	33	45.6	0	GeneDB
Unknown	Tb927.8.3310	Acetyltransferase, putative	76566	1	1	31	80.7	0	GeneDB
Unknown	Tb09.160.4560	AK arginine kinase	44973	16	5	199	96.4	0	GeneDB
Unknown	Tb11.02.1160	Calmodulin, putative	17721	16	2	73	74.6	0	GeneDB
Unknown	Tb927.7.540	Chaperone protein DNAJ, putative	51126	2	1	32	68.7	0	GeneDB
Unknown	Tb927.7.990	Chaperone protein DNAJ, putative	86833	2	2	38	37.9	0	GeneDB
Unknown	Tb927.6.2170	Co-chaperone GrpE, putative	23977	16	3	92	79.7	0	GeneDB
Unknown	Tb09.211.0560	DRBD3 RNA-binding protein, putative; DRBD3	36961	6	2	37	96.5	0	GeneDB
Unknown	Tb10.70.6300	Dual specificity protein phosphatase, putative; serine/threonine protein	60420	4	2	58	57.9	0	GeneDB
Unknown	Tb11.02.4030	Eukaryotic release factor 3, putative	76900	3	3	31	95.6	0	GeneDB

Unknown	Tb10.05.0080	Glucosidase, putative	91887	7	7	87	83.6	0	GeneDB
Unknown	Tb10.6k15.2930	GTPase activating protein, putative	60127	7	3	41	88.9	0	GeneDB
Unknown	Tb927.3.3330	Heat shock protein 20, putative	15946	22	2	52	86.1	0	GeneDB
Unknown	Tb11.01.3080	Heat shock protein 70, putative	73586	9	6	39	89.2	0	GeneDB
Unknown	Tb927.1.370	Leucine-rich repeat protein (LRRP), putative	163515	2	3	46	19.3	0	GeneDB
Unknown	Tb927.8.3550	Mitogen-activated protein kinase 3, putative	42600	7	2	36	42.6	0	GeneDB
Unknown	Tb10.70.1320	NAT1 N-acetyltransferase subunit Nat1, putative	82138	1	1	31	82.6	0	GeneDB
Unknown	Tb11.47.0002	Phosphatidylinositol (3,5) kinase, putative	162330	1	2	32	64.1	0	GeneDB
Unknown	Tb10.70.2440	Phosphatidylinositol-4-phosphate 5-kinase, putative	51824	2	1	43	68.8	0	GeneDB
Unknown	Tb10.70.5520	piwi-like protein 1	124197	9	9	58	53.4	0	GeneDB
Unknown	Tb09.211.2410	PKAC1 protein kinase A catalytic subunit isoform 1; protein kinase A catalytic subunit	38015	3	1	37	79.9	0	GeneDB
Unknown	Tb09.211.2360	PKAC2 protein kinase A catalytic subunit isoform 2; protein kinase A catalytic subunit	38545	3	1	37	72.8	0	GeneDB
Unknown	Tb927.7.4770	PPIase cyclophilin-type peptidyl-prolyl cis-trans isomerase, putative	18635	32	5	157	96.9	0	GeneDB
Unknown	Tb927.7.1300	Protein disulfide isomerase, putative	42256	18	7	327	96.1	0	GeneDB
Unknown	Tb927.7.5790	Protein disulfide isomerase, putative	15505	23	4	114	70.1	0	GeneDB
Unknown	Tb927.7.3210	Protein kinase, putative	70896	7	3	76	41.2	0	GeneDB
Unknown	Tb927.7.3580	Protein kinase, putative	66785	13	6	53	65.6	0	GeneDB
Unknown	Tb10.70.2070	Protein kinase, putative; mitogen-activated protein kinase 2, putative	49710	12	4	32	76.5	0	GeneDB
Unknown	Tb11.02.4830	Protein kinase, putative; serine/threonine protein kinase, putative	67076	7	4	43	42.2	0	GeneDB
Unknown	Tb10.389.0550	RAB5A ras-related protein rab-5; small GTPase, putative	24641	28	5	193	87.5	0	GeneDB
Unknown	Tb09.211.2330	RAB7 small GTPase, putative; GTP-binding protein, putative	24103	27	5	191	51.3	0	GeneDB
Unknown	Tb11.01.8310	RNA-binding protein, putative	49800	3	2	40	74.3	0	GeneDB
Unknown	Tb927.5.4380	Serine/threonine protein phosphatase, putative	72009	1	1	50	72.2	0	GeneDB
Unknown	Tb927.8.7110	Serine/threonine-protein kinase A, putative	49587	7	3	42	77.3	0	GeneDB
Unknown	Tb927.8.8140	Small GTP-binding rab protein, putative	68928	5	3	46	50.8	0	GeneDB
Unknown	Tb09.160.0780	Syntaxin binding protein 1, putative	72297	7	4	148	48.8	0	GeneDB
Unknown	Tb09.160.2020	Thioredoxin (trx)	12298	53	5	57	15.7	0	GeneDB
Unknown	Tb11.01.2280	Ubiquinone biosynthesis methyltransferase, putative	32302	3	1	49	82.6	0	GeneDB
Unknown	Tb927.3.4720	Dynamin, putative	73276	6	4	46	84.5	0	GeneDB
Unknown	Tb10.26.0070	33 kDa inner dynein arm light chain, axonemal, putative	42408	10	3	51	26.3	0	GeneDB
Unknown	Tb927.6.3050	Aldehyde dehydrogenase family, putative	59937	17	6	124	41	0	GeneDB
Unknown	Tb927.6.4210	ALDH aldehyde dehydrogenase, putative	65266	6	4	106	69.3	0	GeneDB
Unknown	Tb927.4.410	Cell differentiation protein, putative	35300	3	1	32	58.8	0	GeneDB
Unknown	Tb927.2.1260	Expression site-associated gene (ESAG, pseudogene), putative	116146	5	6	39	16.7	0	GeneDB
Unknown	Tb11.57.0010	Expression site-associated gene (ESAG, pseudogene), putative;	47636	4	2	42	1.3	0	GeneDB
Unknown	Tb11.57.0050	Expression site-associated gene (ESAG, pseudogene), putative;	43030	9	4	32	0.1	0	GeneDB
Unknown	Tb11.14.0030	Expression site-associated gene (ESAG, pseudogene), putative; (ESAG3), frameshifted and degenerate	38633	5	2	30	1.4	0	GeneDB
Unknown	Tb927.7.3500	Glutathione-S-transferase/glutaredoxin,putative	35812	5	1	45	87.4	0	GeneDB
Unknown	Tb927.5.2080	Inosine-5'-monophosphate dehydrogenase, putative	52776	10	2	117	94.5	0	GeneDB
Unknown	Tb927.6.2790	L-threonine 3-dehydrogenase, putative	37333	9	2	56	95.5	0	GeneDB
Unknown	Tb10.70.6450	NOT1	261017	1	3	30	92.7	0	GeneDB
Unknown	Tb927.3.5520	RPN1 26S proteasome regulatory non-ATPase subunit	100519	4	2	36	87.8	0	GeneDB
Unknown	Tb927.1.4050	Ser/Thr protein phosphatase, putative	107225	5	4	93	90.9	0	GeneDB
Unknown	Tb05.5K5.30	Serine/threonine protein phosphatase, putative	72009	1	1	50	71.6	0	GeneDB
Unknown	Tb05.5K5.150	Small GTP-binding protein, putative	21814	36	4	197	82	0	GeneDB
Unknown	Tb927.6.3630	Sphingosine phosphate lyase-like protein, putative	60343	3	2	100	43.6	0	GeneDB
Unknown	Tb927.8.920	Ubiquitin-conjugating enzyme E2, putative	19379	6	1	60	18.7	0	GeneDB
Unknown	Tb927.2.6404	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase (pseudogene), putative	-	2	2	35	1.9	0	GeneDB
Unknown	Tb05.5K5.160	Hypothetical protein, conserved	108136	3	4	36	24.1	0	Gene name
Unknown	Tb05.5K5.50	Hypothetical protein, conserved	47543	6	2	79	46.5	0	Gene name

Unknown	Tb07.11L3.100	Hypothetical protein, conserved	139292	2	3	33	62.9	0	Gene name
Unknown	Tb09.142.0350	Hypothetical protein	20751	34	4	58	10.3	0	Gene name
Unknown	Tb09.160.0350	Hypothetical protein, conserved	60062	11	6	254	30.5	0	Gene name
Unknown	Tb09.160.0520	Hypothetical protein, conserved	183684	3	5	82	30.2	0	Gene name
Unknown	Tb09.160.0610	Hypothetical protein, conserved	52918	9	3	47	45.1	0	Gene name
Unknown	Tb09.160.1110	Hypothetical protein, conserved	21647	6	1	31	26.2	0	Gene name
Unknown	Tb09.160.1250	Hypothetical protein, conserved	122569	3	2	39	84.9	0	Gene name
Unknown	Tb09.160.1610	Hypothetical protein, conserved	55277	9	3	120	61.9	0	Gene name
Unknown	Tb09.160.1710	Hypothetical protein, conserved	43418	3	1	38	72.2	0	Gene name
Unknown	Tb09.160.2180	Hypothetical protein, conserved	30255	4	1	52	36.1	0	Gene name
Unknown	Tb09.160.2470	Hypothetical protein, conserved	80550	7	3	71	55.2	0	Gene name
Unknown	Tb09.160.3740	Hypothetical protein, conserved	12364	36	3	118	57.3	0	Gene name
Unknown	Tb09.211.0080	Hypothetical protein, conserved		3	3	31	46.5	0	Gene name
Unknown	Tb09.211.0290	Hypothetical protein, conserved	67382	15	6	204	62.2	0	Gene name
Unknown	Tb09.211.0380	Hypothetical protein	38508	5	2	36	27.9	0	Gene name
Unknown	Tb09.211.1150	Hypothetical protein, conserved	20679	52	9	242	59	0	Gene name
Unknown	Tb09.211.1230	Hypothetical protein, conserved	115220	4	4	31	59.3	0	Gene name
Unknown	Tb09.211.1240	Hypothetical protein, conserved	38499	7	2	46	92.2	0	Gene name
Unknown	Tb09.211.1390	Hypothetical protein, conserved	80094	9	4	146	18.4	0	Gene name
Unknown	Tb09.211.1630	Hypothetical protein, conserved	91764	6	4	33	57.8	0	Gene name
Unknown	Tb09.211.2450	Hypothetical protein, conserved	52548	5	2	61	33.3	0	Gene name
Unknown	Tb09.211.3490	Hypothetical protein, conserved	144056	2	2	34	29	0	Gene name
Unknown	Tb09.211.3690	Hypothetical protein, conserved	46605	20	5	61	21.2	0	Gene name
Unknown	Tb09.211.3750	Hypothetical protein, conserved	121004	4	3	37	47.5	0	Gene name
Unknown	Tb09.211.4100	Hypothetical protein, conserved	18541	6	1	41	20.3	0	Gene name
Unknown	Tb09.211.4380	Hypothetical protein, conserved	41521	25	6	56	63.9	0	Gene name
Unknown	Tb09.244.2680	Hypothetical protein, conserved	36848	5	2	37	57.1	0	Gene name
Unknown	Tb10.26.0630	Hypothetical protein, conserved	156828	3	4	35	47.2	0	Gene name
Unknown	Tb10.26.0960	Hypothetical protein, conserved	184001	5	7	90	87.8	0	Gene name
Unknown	Tb10.26.1020	Hypothetical protein, conserved	11254	11	1	42	93.6	0	Gene name
Unknown	Tb10.389.0080	Hypothetical protein, conserved	121421	0	1	32	38.8	0	Gene name
Unknown	Tb10.389.0150	Hypothetical protein, conserved	293841	2	4	41	68.9	0	Gene name
Unknown	Tb10.389.0360	Hypothetical protein, conserved	51112	3	1	49	50.5	0	Gene name
Unknown	Tb10.389.0570	Hypothetical protein, conserved	35139	11	3	66	76.7	0	Gene name
Unknown	Tb10.389.1030	Hypothetical protein, conserved	25236	20	4	161	57.9	0	Gene name
Unknown	Tb10.389.1780	Hypothetical protein, conserved	14697	13	1	62	96.6	0	Gene name
Unknown	Tb10.61.0370	Hypothetical protein, conserved	19159	10	2	36	23	0	Gene name
Unknown	Tb10.61.0680	Hypothetical protein, conserved	86126	5	3	78	48.7	0	Gene name
Unknown	Tb10.61.0800	Hypothetical protein, conserved	132341	1	2	38	40.8	0	Gene name
Unknown	Tb10.61.1260	Hypothetical protein, conserved	58233	5	2	103	74.4	0	Gene name
Unknown	Tb10.61.1790	Hypothetical protein, conserved	83611	1	1	38	23.6	0	Gene name
Unknown	Tb10.61.1930	Hypothetical protein, conserved	44321	11	3	137	62.5	0	Gene name
Unknown	Tb10.61.1970	Hypothetical protein, conserved	37678	7	2	34	74.7	0	Gene name
Unknown	Tb10.61.2370	Hypothetical protein, conserved	61065	3	2	42	69.3	0	Gene name
Unknown	Tb10.61.2430	Hypothetical protein, conserved	127716	9	9	200	35.3	0	Gene name
Unknown	Tb10.61.2670	Hypothetical protein, conserved	55505	12	5	164	93	0	Gene name
Unknown	Tb10.61.2850	Hypothetical protein, conserved	50899	16	5	127	18.4	0	Gene name
Unknown	Tb10.61.3130	Hypothetical protein, conserved	75312	7	3	82	91.9	0	Gene name
Unknown	Tb10.6k15.0310	Hypothetical protein, conserved	109928	2	2	30	87.1	0	Gene name
Unknown	Tb10.6k15.0400	Hypothetical protein, conserved	97986	8	5	91	54	0	Gene name
Unknown	Tb10.6k15.0670	Hypothetical protein, conserved	111364	1	2	30	55.4	0	Gene name
Unknown	Tb10.6k15.1040	Hypothetical protein, conserved	57952	5	1	69	62.7	0	Gene name
Unknown	Tb10.6k15.1720	Hypothetical protein, conserved	85447	9	5	50	58.7	0	Gene name
Unknown	Tb10.6k15.1900	Hypothetical protein, conserved	39746	5	1	63	76.8	0	Gene name
Unknown	Tb10.6k15.2090	Hypothetical protein, conserved	415900	0	3	34	64.2	0	Gene name
Unknown	Tb10.6k15.2280	Hypothetical protein, conserved	43161	1	1	30	15.6	0	Gene name
Unknown	Tb10.6k15.2340	Hypothetical protein, conserved	129183	3	3	31	66.3	0	Gene name
Unknown	Tb10.6k15.2450	Hypothetical protein, conserved	82957	9	4	103	88	0	Gene name
Unknown	Tb10.6k15.2510	Hypothetical protein, conserved	28816	8	2	39	97	0	Gene name
Unknown	Tb10.6k15.3050	Hypothetical protein, conserved	41451	7	4	46	37.5	0	Gene name
Unknown	Tb10.6k15.3110	Hypothetical protein, conserved		3	3	32	46	0	Gene name
Unknown	Tb10.6k15.3900	Hypothetical protein, conserved	36953	25	5	62	84.6	0	Gene name
Unknown	Tb10.70.0470	Hypothetical protein, conserved	55490	4	1	41	60.2	0	Gene name

Unknown	Tb10.70.0760	Hypothetical protein, conserved	82202	2	2	59	48.6	0	Gene name
Unknown	Tb10.70.0870	Hypothetical protein, conserved	129127	1	2	32	69.9	0	Gene name
Unknown	Tb10.70.0890	Hypothetical protein, conserved	173623	1	3	31	47.4	0	Gene name
Unknown	Tb10.70.1030	Hypothetical protein	67133	12	6	121	41.9	0	Gene name
Unknown	Tb10.70.1070	Hypothetical protein, conserved	52245	11	5	132	71.5	0	Gene name
Unknown	Tb10.70.1150	Hypothetical protein, conserved	59726	9	4	161	71	0	Gene name
Unknown	Tb10.70.2380	Hypothetical protein, conserved	38368	3	1	32	25.9	0	Gene name
Unknown	Tb10.70.3380	Hypothetical protein, conserved	53278	7	2	89	60	0	Gene name
Unknown	Tb10.70.3540	Hypothetical protein, conserved	235841	7	14	295	46.2	0	Gene name
Unknown	Tb10.70.3550	Hypothetical protein, conserved	65476	1	1	38	58.1	0	Gene name
Unknown	Tb10.70.4540	Hypothetical protein, conserved	134333	10	10	376	45.7	0	Gene name
Unknown	Tb10.70.5010	Hypothetical protein, conserved	133827	2	3	53	69.5	0	Gene name
Unknown	Tb10.70.5860	Hypothetical protein, conserved	158114	9	11	236	88.7	0	Gene name
Unknown	Tb10.70.7180	Hypothetical protein, conserved	205417	3	7	128	40.4	0	Gene name
Unknown	Tb10.70.7490	Hypothetical protein, conserved	118813	2	3	63	54.6	0	Gene name
Unknown	Tb10.70.7530	Hypothetical protein, conserved	33445	8	6	35	47	0	Gene name
Unknown	Tb10.70.7600	Hypothetical protein, conserved	61218	2	2	45	24.5	0	Gene name
Unknown	Tb10.70.7770	Hypothetical protein, conserved	37550	11	2	92	21.8	0	Gene name
Unknown	Tb10.70.7950	Hypothetical protein, conserved	16561	33	3	156	29.5	0	Gene name
Unknown	Tb11.01.0250	Hypothetical protein, conserved	73826	6	3	122	49.5	0	Gene name
Unknown	Tb11.01.0510	Hypothetical protein, conserved	46446	16	6	72	45.1	0	Gene name
Unknown	Tb11.01.0980	Hypothetical protein, conserved	206529	2	3	63	38.6	0	Gene name
Unknown	Tb11.01.1020	Hypothetical protein, conserved	46311	3	1	38	67.4	0	Gene name
Unknown	Tb11.01.1140	Hypothetical protein, conserved	68830	1	1	39	52.8	0	Gene name
Unknown	Tb11.01.1300	Hypothetical protein, conserved	63938	3	2	31	83.2	0	Gene name
Unknown	Tb11.01.1880	Hypothetical protein, conserved	100303	8	7	249	37	0	Gene name
Unknown	Tb11.01.2150	Hypothetical protein, conserved	43150	13	4	37	36.7	0	Gene name
Unknown	Tb11.01.2550	Hypothetical protein, conserved	31225	7	3	59	54.4	0	Gene name
Unknown	Tb11.01.2730	Hypothetical protein, conserved	314478	2	9	125	55.7	0	Gene name
Unknown	Tb11.01.3290	Hypothetical protein, conserved	69583	9	5	167	90.7	0	Gene name
Unknown	Tb11.01.3410	Hypothetical protein, conserved	47126	2	1	42	87.2	0	Gene name
Unknown	Tb11.01.3720	Hypothetical protein	53654	8	4	66	55.2	0	Gene name
Unknown	Tb11.01.5270	Hypothetical protein, conserved	86591	4	3	40	35.5	0	Gene name
Unknown	Tb11.01.6280	Hypothetical protein, conserved	122855	2	3	34	19.5	0	Gene name
Unknown	Tb11.01.6610	Hypothetical protein, conserved	94359	2	2	46	34	0	Gene name
Unknown	Tb11.01.6770	Hypothetical protein, conserved	78807	7	5	74	28.8	0	Gene name
Unknown	Tb11.01.6900	Hypothetical protein, conserved	67653	5	3	58	37.5	0	Gene name
Unknown	Tb11.01.6960	Hypothetical protein, conserved	81767	6	5	91	45.6	0	Gene name
Unknown	Tb11.01.7450	Hypothetical protein, conserved	90387	6	4	71	43.1	0	Gene name
Unknown	Tb11.01.8600	Hypothetical protein, conserved	102378	1	2	34	82.1	0	Gene name
Unknown	Tb11.01.8650	Hypothetical protein, conserved	230864	9	15	377	34.3	0	Gene name
Unknown	Tb11.02.0010	Hypothetical protein, conserved	24706	31	4	255	67.2	0	Gene name
Unknown	Tb11.02.0110	Hypothetical protein, conserved	146279	1	3	34	78	0	Gene name
Unknown	Tb11.02.0190	Hypothetical protein, conserved	122437	3	3	37	28.2	0	Gene name
Unknown	Tb11.02.0350	Hypothetical protein, conserved	53529	8	4	97	47.3	0	Gene name
Unknown	Tb11.02.0358	Hypothetical protein, conserved	116140	6	5	89	31.2	0	Gene name
Unknown	Tb11.02.0610	Hypothetical protein, conserved	32275	2	1	34	62.8	0	Gene name
Unknown	Tb11.02.1490	Hypothetical protein, conserved	35201	5	2	71	63.9	0	Gene name
Unknown	Tb11.02.1670	Hypothetical protein, conserved	59131	2	9	40	83.1	0	Gene name
Unknown	Tb11.02.1900	Hypothetical protein, conserved	65124	5	3	51	55	0	Gene name
Unknown	Tb11.02.1980	Hypothetical protein, conserved	59824	6	3	82	81.6	0	Gene name
Unknown	Tb11.02.2840	Hypothetical protein, conserved	28017	8	3	39	75.4	0	Gene name
Unknown	Tb11.02.3065	Hypothetical protein, conserved	10017	16	1	51	92.6	0	Gene name
Unknown	Tb11.02.3310	Hypothetical protein, conserved	13569	15	3	92	74.1	0	Gene name
Unknown	Tb11.02.3500	Hypothetical protein, conserved	104532	8	7	104	31.3	0	Gene name
Unknown	Tb11.02.4790	Hypothetical protein, conserved	61016	5	4	33	72.8	0	Gene name
Unknown	Tb11.02.5030	Hypothetical protein, conserved	73312	3	2	44	68.6	0	Gene name
Unknown	Tb11.03.0210	Hypothetical protein, conserved	136688	3	5	36	48.4	0	Gene name
Unknown	Tb11.03.0475	Hypothetical protein, conserved	12130	21	1	57	95.9	0	Gene name
Unknown	Tb11.03.0520	Hypothetical protein, conserved	57991	9	4	48	56	0	Gene name
Unknown	Tb11.03.0690	Hypothetical protein, conserved	90151	1	1	38	25.8	0	Gene name
Unknown	Tb11.03.0830	Hypothetical protein, conserved	76098	3	2	34	22.8	0	Gene name
Unknown	Tb11.1400	Hypothetical protein	68539	1	1	32	81.6	0	Gene name
Unknown	Tb11.1410	Hypothetical protein		2	1	31	41.6	0	Gene name
Unknown	Tb11.22.0004	Hypothetical protein, conserved	56654	4	2	44	75.6	0	Gene name
Unknown	Tb11.39.0007	Hypothetical protein, conserved		2	1	31	41.6	0	Gene name
Unknown	Tb11.47.0011	Hypothetical protein, conserved	330695	3	10	119	53.4	0	Gene name
Unknown	Tb11.52.0006	Hypothetical protein, conserved	139679	9	10	229	62	0	Gene name
Unknown	Tb11.55.0010	Hypothetical protein, conserved	61708	7	2	47	85	0	Gene name
Unknown	Tb11.55.0017	Hypothetical protein, conserved	21642	14	2	135	94.6	0	Gene name
Unknown	Tb11.55.0021	Hypothetical protein, conserved		3	1	37	28.9	0	Gene name
Unknown	Tb927.1.3450	Hypothetical protein, conserved	85915	1	1	46	60.7	0	Gene name
Unknown	Tb927.1.990	Hypothetical protein, conserved	42181	11	4	57	56.6	0	Gene name

Unknown	Tb927.2.1740	Hypothetical protein, conserved		3	1	31	59.3	0	Gene name
Unknown	Tb927.2.2090	Hypothetical protein	110449	3	3	32	22.7	0	Gene name
Unknown	Tb927.2.2370	Hypothetical protein, conserved	152536	8	9	53	68.6	0	Gene name
Unknown	Tb927.2.2530	Hypothetical protein, conserved	88891	2	2	39	41.2	0	Gene name
Unknown	Tb927.2.3000	Hypothetical protein, conserved	47454	4	1	52	69	0	Gene name
Unknown	Tb927.2.3180	Hypothetical protein, conserved		3	3	32	88	0	Gene name
Unknown	Tb927.2.4050	Hypothetical protein, conserved	129529	4	3	80	89.6	0	Gene name
Unknown	Tb927.2.4160	Hypothetical protein, conserved	93560	5	4	52	25.2	0	Gene name
Unknown	Tb927.2.5010	Hypothetical protein, conserved	55064	3	1	32	40.9	0	Gene name
Unknown	Tb927.2.5760	Hypothetical protein, conserved	332716	5	17	425	66.5	0	Gene name
Unknown	Tb927.2.6100	Hypothetical protein, conserved	53247	7	4	55	16.8	0	Gene name
Unknown	Tb927.3.1190	Hypothetical protein, conserved	259181	2	5	67	53.6	0	Gene name
Unknown	Tb927.3.1480	Hypothetical protein, conserved	33722	15	3	30	8.4	0	Gene name
Unknown	Tb927.3.2070	Hypothetical protein, conserved	151744	3	5	115	68.6	0	Gene name
Unknown	Tb927.3.2200	Hypothetical protein, conserved	39377	31	7	71	72	0	Gene name
Unknown	Tb927.3.3000	Hypothetical protein, conserved	65592	7	3	42	96.1	0	Gene name
Unknown	Tb927.3.3120	Hypothetical protein, conserved	48063	9	3	95	81.7	0	Gene name
Unknown	Tb927.3.3180	Hypothetical protein, conserved	98314	3	2	37	85.3	0	Gene name
Unknown	Tb927.3.4580	Hypothetical protein, conserved	96523	2	1	49	56.6	0	Gene name
Unknown	Tb927.3.4620	Hypothetical protein, conserved	160066	0	1	34	70.9	0	Gene name
Unknown	Tb927.3.4710	Hypothetical protein, conserved	59491	5	2	40	39	0	Gene name
Unknown	Tb927.3.880	Hypothetical protein, conserved	27457	11	2	83	19.5	0	Gene name
Unknown	Tb927.4.1240	Hypothetical protein	13306	24	3	34	10.4	0	Gene name
Unknown	Tb927.4.2300	Hypothetical protein, conserved	81650	5	3	75	44	0	Gene name
Unknown	Tb927.4.2400	Hypothetical protein, conserved	158021	6	7	165	52	0	Gene name
Unknown	Tb927.4.2640	Hypothetical protein, conserved	146478	3	4	40	37.3	0	Gene name
Unknown	Tb927.4.2850	Hypothetical protein, conserved	72764	2	2	55	73.5	0	Gene name
Unknown	Tb927.4.2920	Hypothetical protein, conserved	107526	6	6	176	58.4	0	Gene name
Unknown	Tb927.4.2990	Hypothetical protein, conserved	144613	3	4	49	44.3	0	Gene name
Unknown	Tb927.4.310	Hypothetical protein, conserved	743666	0	5	33	39.9	0	Gene name
Unknown	Tb927.4.3360	Hypothetical protein, conserved	49350	2	1	33	91.6	0	Gene name
Unknown	Tb927.4.4540	Hypothetical protein, conserved	13774	10	1	59	43.1	0	Gene name
Unknown	Tb927.4.5000	Hypothetical protein, conserved	97016	18	11	358	73.9	0	Gene name
Unknown	Tb927.4.5130	Hypothetical protein, conserved	28794	9	2	32	80.5	0	Gene name
Unknown	Tb927.4.580	Hypothetical protein, conserved	47444	7	3	36	45.1	0	Gene name
Unknown	Tb927.5.1690	Hypothetical protein, conserved	51987	18	8	305	87.8	0	Gene name
Unknown	Tb927.5.1780	Hypothetical protein, conserved	49188	20	7	441	95	0	Gene name
Unknown	Tb927.5.1920	Hypothetical protein, conserved	70682	2	9	33	76.9	0	Gene name
Unknown	Tb927.5.2190	Hypothetical protein, conserved	49690	13	4	154	90.1	0	Gene name
Unknown	Tb927.5.2220	Hypothetical protein, conserved	49690	13	4	154	90.1	0	Gene name
Unknown	Tb927.5.2500	Hypothetical protein, conserved	75939	5	3	41	23.2	0	Gene name
Unknown	Tb927.5.3290	Hypothetical protein, conserved	129925	1	2	34	40.9	0	Gene name
Unknown	Tb927.5.3330	Hypothetical protein, conserved	521434	0	3	33	40.7	0	Gene name
Unknown	Tb927.5.3660	Hypothetical protein, conserved	56029	2	1	50	20.1	0	Gene name
Unknown	Tb927.5.4090	Hypothetical protein, conserved	26551	8	2	101	28.7	0	Gene name
Unknown	Tb927.5.4400	Hypothetical protein, conserved	47543	6	2	79	46.5	0	Gene name
Unknown	Tb927.5.4520	Hypothetical protein, conserved	108136	3	4	36	24.1	0	Gene name
Unknown	Tb927.5.570	Hypothetical protein, conserved	235296	7	13	218	49.6	0	Gene name
Unknown	Tb927.5.870	Hypothetical protein, conserved	6729	44	2	96	90.3	0	Gene name
Unknown	Tb927.5.980	Hypothetical protein, conserved	35043	2	1	32	36.9	0	Gene name
Unknown	Tb927.6.1180	Hypothetical protein, conserved	141903	20	20	60	70.3	0	Gene name
Unknown	Tb927.6.1220	Hypothetical protein, conserved	65870	3	2	39	18.7	0	Gene name
Unknown	Tb927.6.1920	Hypothetical protein, conserved	42395	5	3	43	44.6	0	Gene name
Unknown	Tb927.6.1940	Hypothetical protein, conserved	73827	4	2	69	39	0	Gene name
Unknown	Tb927.6.2090	Hypothetical protein, conserved	78931	6	3	82	58.6	0	Gene name
Unknown	Tb927.6.2220	Hypothetical protein, conserved	15032	12	1	73	68.2	0	Gene name
Unknown	Tb927.6.2230	Hypothetical protein, conserved	87159	1	1	82	83.1	0	Gene name
Unknown	Tb927.6.2760	Hypothetical protein, conserved	166791	0	1	50	43.1	0	Gene name
Unknown	Tb927.6.2930	Hypothetical protein, conserved	59855	6	3	92	25.4	0	Gene name
Unknown	Tb927.6.3640	Hypothetical protein, conserved	64572	10	4	88	13.3	0	Gene name
Unknown	Tb927.6.3670	Hypothetical protein, conserved	364916	12	29	660	38.2	0	Gene name
Unknown	Tb927.6.4230	Hypothetical protein, conserved	55178	4	2	37	25.7	0	Gene name
Unknown	Tb927.6.4770	Hypothetical protein, conserved	83881	6	4	103	81.8	0	Gene name
Unknown	Tb927.6.590	Hypothetical protein, conserved	12269	8	1	42	85.9	0	Gene name
Unknown	Tb927.7.1430	Hypothetical protein, conserved	48300	4	2	99	74.8	0	Gene name
Unknown	Tb927.7.2630	Hypothetical protein, conserved	100617	0	1	33	62	0	Gene name
Unknown	Tb927.7.3060	Hypothetical protein, conserved		5	2	118	33.9	0	Gene name
Unknown	Tb927.7.3130	Hypothetical protein, conserved	102765	1	2	34	68.6	0	Gene name
Unknown	Tb927.7.3540	Hypothetical protein, conserved	29782	10	2	59	28.5	0	Gene name
Unknown	Tb927.7.510	Hypothetical protein, conserved	39971	11	4	37	33.3	0	Gene name
Unknown	Tb927.7.5300	Hypothetical protein, conserved	73234	3	2	39	50.5	0	Gene name
Unknown	Tb927.7.5320	Hypothetical protein, conserved	83949	5	3	40	61	0	Gene name
Unknown	Tb927.7.5340	Hypothetical protein, conserved	56959	25	12	285	91.1	0	Gene name

Unknown	Tb927.7.5410	Hypothetical protein, conserved	100337	2	3	59	68	0	Gene name
Unknown	Tb927.7.5430	Hypothetical protein, conserved	100220	2	3	59	67.6	0	Gene name
Unknown	Tb927.7.6250	Hypothetical protein, conserved	49011	3	2	39	32.7	0	Gene name
Unknown	Tb927.7.6260	Hypothetical protein, conserved	36774	4	1	52	89.7	0	Gene name
Unknown	Tb927.7.650	Hypothetical protein, conserved	105866	0	1	31	64.4	0	Gene name
Unknown	Tb927.7.6640	Hypothetical protein, conserved	173130	8	11	201	72.4	0	Gene name
Unknown	Tb927.7.7000	Hypothetical protein, conserved	177219	7	9	136	56.7	0	Gene name
Unknown	Tb927.7.7400	Hypothetical protein, conserved	249044	0	2	40	25.5	0	Gene name
Unknown	Tb927.7.860	Hypothetical protein, conserved	117832	1	2	38	25.1	0	Gene name
Unknown	Tb927.8.1050	Hypothetical protein, conserved	295070	5	12	128	67	0	Gene name
Unknown	Tb927.8.1360	Hypothetical protein, conserved	472876	1	4	34	39.5	0	Gene name
Unknown	Tb927.8.1740	Hypothetical protein, conserved	62618	7	3	41	89.8	0	Gene name
Unknown	Tb927.8.1790	Hypothetical protein, conserved	17429	12	5	46	90.3	0	Gene name
Unknown	Tb927.8.2060	Hypothetical protein, conserved	328056	1	4	34	52.7	0	Gene name
Unknown	Tb927.8.2070	Hypothetical protein, conserved	22184	33	5	167	65.6	0	Gene name
Unknown	Tb927.8.2080	Hypothetical protein, conserved	20362	30	5	168	62.7	0	Gene name
Unknown	Tb927.8.2130	Hypothetical protein, conserved	57640	6	3	73	56.9	0	Gene name
Unknown	Tb927.8.2260	Hypothetical protein, conserved	22154	33	5	167	64.5	0	Gene name
Unknown	Tb927.8.2270	Hypothetical protein, conserved	22152	33	5	167	66.3	0	Gene name
Unknown	Tb927.8.2280	Hypothetical protein, conserved	22154	33	5	167	64.5	0	Gene name
Unknown	Tb927.8.2430	Hypothetical protein, conserved	131586	2	2	45	47.6	0	Gene name
Unknown	Tb927.8.2550	Hypothetical protein, conserved	70218	1	1	49	85.4	0	Gene name
Unknown	Tb927.8.2620	Hypothetical protein, conserved	93611	6	4	175	40.9	0	Gene name
Unknown	Tb927.8.3000	Hypothetical protein, conserved	57787	6	3	72	38.3	0	Gene name
Unknown	Tb927.8.3590	Hypothetical protein, conserved	72481	4	2	53	58.1	0	Gene name
Unknown	Tb927.8.3950	Hypothetical protein, conserved	103041	1	1	30	59	0	Gene name
Unknown	Tb927.8.4320	Hypothetical protein, conserved	168858	2	5	69	28.6	0	Gene name
Unknown	Tb927.8.4510	Hypothetical protein, conserved	70319	5	3	116	69.7	0	Gene name
Unknown	Tb927.8.4900	Hypothetical protein, conserved	115096	1	1	33	19	0	Gene name
Unknown	Tb927.8.4960	Hypothetical protein, conserved	163037	0	1	33	48.8	0	Gene name
Unknown	Tb927.8.5020	Hypothetical protein, conserved	50820	2	1	46	73.8	0	Gene name
Unknown	Tb927.8.5140	Hypothetical protein, conserved	63515	4	3	57	72.9	0	Gene name
Unknown	Tb927.8.5170	Hypothetical protein, conserved	136596	4	6	111	60.3	0	Gene name
Unknown	Tb927.8.540	Hypothetical protein, conserved	270011	0	2	35	34.9	0	Gene name
Unknown	Tb927.8.5830	Hypothetical protein, conserved	36881	11	2	79	85.8	0	Gene name
Unknown	Tb927.8.5940	Hypothetical protein, conserved	54844	6	2	51	11.7	0	Gene name
Unknown	Tb927.8.6270	Hypothetical protein, conserved	69923	5	3	71	35.3	0	Gene name
Unknown	Tb927.8.6640	Hypothetical protein, conserved	68141	1	1	38	64.2	0	Gene name
Unknown	Tb927.8.7040	Hypothetical protein, conserved	21523	23	3	154	58.1	0	Gene name
Unknown	Tb927.8.7060	Hypothetical protein, conserved	171163	1	2	42	23.4	0	Gene name
Unknown	Tb927.8.7080	Hypothetical protein, conserved	176605	4	6	35	48	0	Gene name
Unknown	Tb927.8.7420	Hypothetical protein, conserved	97015	18	11	358	74.7	0	Gene name
Unknown	Tb927.8.7790	Hypothetical protein, conserved	13055	10	1	35	52.3	0	Gene name
Unknown	Tb927.8.7800	Hypothetical protein, conserved	131530	1	3	45	64.5	0	Gene name
Unknown	Tb927.8.790	Hypothetical protein, conserved	146362	3	5	35	46.4	0	Gene name
Unknown	Tb927.8.7960	Hypothetical protein	48374	6	2	56	27.1	0	Gene name
Unknown	Tb927.8.7970	Hypothetical protein	53692	9	4	142	67.8	0	Gene name
Unknown	Tb927.8.8120	Hypothetical protein, conserved	30408	3	1	49	85.4	0	Gene name
Unknown	Tb927.8.8150	Hypothetical protein, conserved	19734	7	2	45	76.5	0	Gene name
Unknown	Tb927.8.8160	Hypothetical protein, conserved	88946	2	2	82	44	0	Gene name

References

Aebersold, R. 2003, "Constellations in a cellular universe", *Nature*, vol. 422, no. 6928, pp. 115-116.

Aebersold, R. & Mann, M. 2003, "Mass spectrometry-based proteomics", *Nature*, vol. 422, no. 6928, pp. 198-207.

Agranoff, D., Stich, A., Abel, P., & Krishna, S. 2005, "Proteomic fingerprinting for the diagnosis of human African trypanosomiasis", *Trends Parasitol.*, vol. 21, no. 4, pp. 154-157.

Albert, M. A., Haanstra, J. R., Hannaert, V., Van Roy, J., Opperdoes, F. R., Bakker, B. M., & Michels, P. A. 2005, "Experimental and *in silico* analyses of glycolytic flux control in bloodstream form *Trypanosoma brucei*", *J.Biol.Chem.*, vol. 280, no. 31, pp. 28306-28315.

Añez-Rojas, N., García-Lugo, P., Crisante, G., Rojas, A., & Añez, N. 2005, "Isolation, purification and characterization of GPI-anchored membrane proteins from *Trypanosoma rangeli* and *Trypanosoma cruzi*", *Acta Trop.*, vol. 97, no. 2, pp. 140-145.

Apted, F. I. C. 1970, "Treatment of Human Trypanosomiasis," in *The African Trypanosomiases*, H. M. Mulligan, ed., Allen & Unwin, pp. 685-687.

Bacchi, C. J. 1993, "Resistance to clinical drugs in African trypanosomes", *Parasitol.Today*, vol. 9, no. 5, pp. 190-193.

Bailey, J. W. & Smith, D. H. 1994, "The quantitative buffy coat for the diagnosis of trypanosomes", *Trop.Doct.*, vol. 24, no. 2, pp. 54-56.

Baldwin, M. A. 2005, "Mass spectrometers for the analysis of biomolecules", *Methods Enzymol.*, vol. 402, pp. 3-48.

Bales, J. D., Jr., Harrison, S. M., Mbwabi, D. L., & Schecter, P. J. 1989, "Treatment of arsenical refractory Rhodesian sleeping sickness in Kenya", *Ann.Trop.Med.Parasitol.*, vol. 83 Suppl 1, pp. 111-114.

Barrett, M. P. 1999, "The fall and rise of sleeping sickness", *Lancet*, vol. 353, no. 9159, pp. 1113-1114.

Barrett, M. P. 2006, "The rise and fall of sleeping sickness", *Lancet*, vol. 367, no. 9520, pp. 1377-1378.

Barrett, M. P., Burchmore, R. J., Stich, A., Lazzari, J. O., Frasch, A. C., Cazzulo, J. J., & Krishna, S. 2003, "The trypanosomiases", *Lancet*, vol. 362, no. 9394, pp. 1469-1480.

Barrett, M. P. & Fairlamb, A. H. 1999, "The biochemical basis of arsenical-diamidine crossresistance in African trypanosomes", *Parasitol.Today*, vol. 15, no. 4, pp. 136-140.

Barrett, M. P., Zhang, Z. Q., Denise, H., Giroud, C., & Baltz, T. 1995, "A diamidine-resistant *Trypanosoma equiperdum* clone contains a P2 purine transporter with reduced substrate affinity", *Mol.Biochem.Parasitol.*, vol. 73, no. 1-2, pp. 223-229.

Barrett, S. V. & Barrett, M. P. 2000, "Anti-sleeping sickness drugs and cancer chemotherapy", *Parasitol.Today*, vol. 16, no. 1, pp. 7-9.

Basselin, M., Denise, H., Coombs, G. H., & Barrett, M. P. 2002, "Resistance to pentamidine in *Leishmania mexicana* involves exclusion of the drug from the mitochondrion", *Antimicrob.Agents Chemother.*, vol. 46, no. 12, pp. 3731-3738.

Berger, B. J., Carter, N. S., & Fairlamb, A. H. 1993, "Polyamine and pentamidine metabolism in African trypanosomes", *Acta Trop.*, vol. 54, no. 3-4, pp. 215-224.

Berger, B. J., Carter, N. S., & Fairlamb, A. H. 1995, "Characterisation of pentamidine-resistant *Trypanosoma brucei brucei*", *Mol.Biochem.Parasitol.*, vol. 69, no. 2, pp. 289-298.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Bohme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T. J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P. J., Moule, S., Martin, D. M., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Tallon, L., Turner, C. M., Tait, A., Tivey, A. R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B. G., Donelson, J. E., Hall, N., Fraser, C. M., Melville, S. E., & El Sayed, N. M. 2005, "The genome of the African trypanosome *Trypanosoma brucei*", *Science*, vol. 309, no. 5733, pp. 416-422.

Beverley, S. M. 2003, "Protozomics: trypanosomatid parasite genetics comes of age", *Nat.Rev.Genet.*, vol. 4, no. 1, pp. 11-19.

Beynon, R. J. & Pratt, J. M. 2005, "Metabolic labeling of proteins for proteomics", *Mol.Cell Proteomics*, vol. 4, no. 7, pp. 857-872.

Biagini, G. A., Pasini, E. M., Hughes, R., De Koning, H. P., Vial, H. J., O'Neill, P. M., Ward, S. A., & Bray, P. G. 2004, "Characterization of the choline carrier of *Plasmodium falciparum*: a route for the selective delivery of novel antimalarial drugs", *Blood*, vol. 104, no. 10, pp. 3372-3377.

Bierczynska-Krzysik, A., Kang, S. U., Silberrring, J., & Lubec, G. 2006, "Mass spectrometrical identification of brain proteins including highly insoluble and transmembrane proteins", *Neurochem.Int.*, vol. 49, no. 3, pp. 245-255.

Bigelow, H. R., Petrey, D. S., Liu, J., Przybylski, D., & Rost, B. 2004, "Predicting transmembrane beta-barrels in proteomes", *Nucleic Acids Res.*, vol. 32, no. 8, pp. 2566-2577.

Bitonti, A. J., Cross-Doersen, D. E., & McCann, P. P. 1988, "Effects of α-difluoromethylornithine on protein synthesis and synthesis of the variant-specific glycoprotein (VSG) in *Trypanosoma brucei brucei*", *Biochem.J.*, vol. 250, no. 1, pp. 295-298.

Blonder, J., Conrads, T. P., Yu, L. R., Terunuma, A., Janini, G. M., Issaq, H. J., Vogel, J. C., & Veenstra, T. D. 2004, "A detergent- and cyanogen bromide-free method for integral membrane proteomics: application to Halobacterium purple membranes and the human epidermal membrane proteome", *Proteomics*, vol. 4, no. 1, pp. 31-45.

Blonder, J., Goshe, M. B., Moore, R. J., Pasa-Tolic, L., Masselon, C. D., Lipton, M. S., & Smith, R. D. 2002, "Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography-tandem mass spectrometry", *J.Proteome.Res.*, vol. 1, no. 4, pp. 351-360.

Blum, M. L., Down, J. A., Gurnett, A. M., Carrington, M., Turner, M. J., & Wiley, D. C. 1993, "A structural motif in the variant surface glycoproteins of *Trypanosoma brucei*", *Nature*, vol. 362, no. 6421, pp. 603-609.

Bordier, C. 1981, "Phase separation of integral membrane proteins in Triton X-114 solution", *J.Biol.Chem.*, vol. 256, no. 4, pp. 1604-1607.

Borst, P., Evers, R., Kool, M., & Wijnholds, J. 1999, "The multidrug resistance protein family", *Biochim.Biophys.Acta*, vol. 1461, no. 2, pp. 347-357.

Borst, P., Evers, R., Kool, M., & Wijnholds, J. 2000, "A family of drug transporters: the multidrug resistance-associated proteins", *J.Natl.Cancer Inst.*, vol. 92, no. 16, pp. 1295-1302.

Boule, J. B., Vega, L. R., & Zakian, V. A. 2005, "The yeast Pif1p helicase removes telomerase from telomeric DNA", *Nature*, vol. 438, no. 7064, pp. 57-61.

Boule, J. B. & Zakian, V. A. 2006, "Roles of Pif1-like helicases in the maintenance of genomic stability", *Nucleic Acids Res.*, vol. 34, no. 15, pp. 4147-4153.

Bouteille, B., Oukem, O., Bisser, S., & Dumas, M. 2003, "Treatment perspectives for human African trypanosomiasis", *Fundam.Clin.Pharmacol.*, vol. 17, no. 2, pp. 171-181.

Bray, P. G., Barrett, M. P., Ward, S. A., & De Koning, H. P. 2003, "Pentamidine uptake and resistance in pathogenic protozoa: past, present and future", *Trends Parasitol.*, vol. 19, no. 5, pp. 232-239.

Bringaud, F., Riviere, L., & Coustou, V. 2006, "Energy metabolism of trypanosomatids: adaptation to available carbon sources", *Mol.Biochem.Parasitol.*, vol. 149, no. 1, pp. 1-9.

Broadhead, R., Dawe, H. R., Farr, H., Griffiths, S., Hart, S. R., Portman, N., Shaw, M. K., Ginger, M. L., Gaskell, S. J., McKean, P. G., & Gull, K. 2006, "Flagellar motility is required for the viability of the bloodstream trypanosome", *Nature*, vol. 440, no. 7081, pp. 224-227.

Brookes, P. S., Pinner, A., Ramachandran, A., Coward, L., Barnes, S., Kim, H., & Darley-Usmar, V. M. 2002, "High throughput two-dimensional blue-native electrophoresis: a tool for functional proteomics of mitochondria and signaling complexes", *Proteomics*, vol. 2, no. 8, pp. 969-977.

Brun, R., Schumacher, R., Schmid, C., Kunz, C., & Burri, C. 2001, "The phenomenon of treatment failures in Human African Trypanosomiasis", *Trop.Med.Int.Health*, vol. 6, no. 11, pp. 906-914.

Brusca, J. S. & Radolf, J. D. 1994, "Isolation of integral membrane proteins by phase partitioning with Triton X-114", *Methods Enzymol.*, vol. 228:182-93., pp. 182-193.

Burri, C., Nkunku, S., Merolle, A., Smith, T., Blum, J., & Brun, R. 2000, "Efficacy of new, concise schedule for melarsoprol in treatment of sleeping sickness caused by *Trypanosoma brucei gambiense*: a randomised trial", *Lancet*, vol. 355, no. 9213, pp. 1419-1425.

Cagney, G., Amiri, S., Premawaradena, T., Lindo, M., & Emili, A. 2003, "*In silico* proteome analysis to facilitate proteomics experiments using mass spectrometry", *Proteome.Sci.*, vol. 1, no. 1, p. 5.

Carbone, A., Zinovyev, A., & Képès, F. 2003, "Codon adaptation index as a measure of dominating codon bias", *Bioinformatics.*, vol. 19, no. 16, pp. 2005-2015.

Carrington, M., Carnall, N., Crow, M. S., Gaud, A., Redpath, M. B., Wasunna, C. L., & Webb, H. 1998, "The properties and function of the glycosylphosphatidylinositol-phospholipase C in *Trypanosoma brucei*", *Mol.Biochem.Parasitol.*, vol. 91, no. 1, pp. 153-164.

Carter, N. S. & Fairlamb, A. H. 1993, "Arsenical-resistant trypanosomes lack an unusual adenosine transporter", *Nature*, vol. 361, no. 6408, pp. 173-176.

Casadio, R., Fariselli, P., Taroni, C., & Compiani, M. 1996, "A predictor of transmembrane alpha-helix domains of proteins based on neural networks", *Eur.Biophys.J.*, vol. 24, no. 3, pp. 165-178.

Cattand, P., Jannin, J., & Lucas, P. 2001, "Sleeping sickness surveillance: an essential step towards elimination", *Trop.Med.Int.Health*, vol. 6, no. 5, pp. 348-361.

Chappuis, F., Loutan, L., Simarro, P., Lejon, V., & Büscher, P. 2005, "Options for field diagnosis of human african trypanosomiasis", *Clin.Microbiol.Rev.*, vol. 18, no. 1, pp. 133-146.

Chattopadhyay, A., Jones, N. G., Nietlispach, D., Nielsen, P. R., Voorheis, H. P., Mott, H. R., & Carrington, M. 2005, "Structure of the C-terminal domain from *Trypanosoma brucei* variant surface glycoprotein MITat1.2", *J.Biol.Chem.*, vol. 280, no. 8, pp. 7228-7235.

Chen, C. P., Kernytsky, A., & Rost, B. 2002, "Transmembrane helix predictions revisited", *Protein Sci.*, vol. 11, no. 12, pp. 2774-2791.

Chen, C. P. & Rost, B. 2002, "Long membrane helices and short loops predicted less accurately", *Protein Sci.*, vol. 11, no. 12, pp. 2766-2773.

Clayton, C. E. 2002, "Life without transcriptional control? From fly to man and back again", *EMBO J.*, vol. 21, no. 8, pp. 1881-1888.

Coelho, A. C., Beverley, S. M., & Cotrim, P. C. 2003, "Functional genetic identification of PRP1, an ABC transporter superfamily member conferring pentamidine resistance in *Leishmania major*", *Mol.Biochem.Parasitol.*, vol. 130, no. 2, pp. 83-90.

Cooks, R. G., Ouyang, Z., Takats, Z., & Wiseman, J. M. 2006, "Detection Technologies. Ambient mass spectrometry", *Science*, vol. 311, no. 5767, pp. 1566-1570.

Coughenour, H. D., Spaulding, R. S., & Thompson, C. M. 2004, "The synaptic vesicle proteome: A comparative study in membrane protein identification", *Proteomics*, vol. 4, no. 10, pp. 3141-3155.

Cowman, A. F. & Crabb, B. S. 2003, "Functional genomics: identifying drug targets for parasitic diseases", *Trends Parasitol.*, vol. 19, no. 11, pp. 538-543.

Croft, S. L., Barrett, M. P., & Urbina, J. A. 2005, "Chemotherapy of trypanosomiases and leishmaniasis", *Trends Parasitol.*, vol. 21, no. 11, pp. 508-512.

Cuthbertson, J. M., Doyle, D. A., & Sansom, M. S. 2005, "Transmembrane helix prediction: a comparative evaluation and analysis", *Protein Eng Des Sel*, vol. 18, no. 6, pp. 295-308.

D'Ieteren, G. D., Authie, E., Wissocq, N., & Murray, M. 1998, "Trypanotolerance, an option for sustainable livestock production in areas at risk from trypanosomosis", *Rev.Sci.Tech.*, vol. 17, no. 1, pp. 154-175.

Damper, D. & Patton, C. L. 1976a, "Pentamidine transport and sensitivity in *brucei*-group trypanosomes", *J.Protozool.*, vol. 23, no. 2, pp. 349-356.

Damper, D. & Patton, C. L. 1976b, "Pentamidine transport in *Trypanosoma brucei* - kinetics and specificity", *Biochem.Pharmacol.*, vol. 25, no. 3, pp. 271-276.

De Koning, H. P. 2001, "Uptake of pentamidine in *Trypanosoma brucei brucei* is mediated by three distinct transporters: implications for cross-resistance with arsenicals", *Mol.Pharmacol.*, vol. 59, no. 3, pp. 586-592.

De Koning, H. P., Anderson, L. F., Stewart, M., Burchmore, R. J., Wallace, L. J., & Barrett, M. P. 2004, "The trypanocide diminazene aceturate is accumulated predominantly through the TbAT1 purine transporter: additional insights on diamidine resistance in African trypanosomes", *Antimicrob.Agents Chemother.*, vol. 48, no. 5, pp. 1515-1519.

De Koning, H. P., Bridges, D. J., & Burchmore, R. J. 2005, "Purine and pyrimidine transport in pathogenic protozoa: from biology to therapy", *FEMS Microbiol.Rev.*, vol. 29, no. 5, pp. 987-1020.

De Koning, H. P. & Jarvis, S. M. 1999, "Adenosine transporters in bloodstream forms of *Trypanosoma brucei brucei*: substrate recognition motifs and affinity for trypanocidal drugs", *Mol.Pharmacol.*, vol. 56, no. 6, pp. 1162-1170.

De Koning, H. P. & Jarvis, S. M. 2001, "Uptake of pentamidine in *Trypanosoma brucei brucei* is mediated by the P2 adenosine transporter and at least one novel, unrelated transporter", *Acta Trop.*, vol. 80, no. 3, pp. 245-250.

Delespaux, V. & De Koning, H. P. Drugs and drug resistance in African trypanosomiasis . Drug Resistance Updates . 2006.
Ref Type: In Press

Deshusses, J. M., Burgess, J. A., Scherl, A., Wenger, Y., Walter, N., Converset, V., Paesano, S., Corthals, G. L., Hochstrasser, D. F., & Sanchez, J. C. 2003, "Exploitation of specific properties of trifluoroethanol for extraction and separation of membrane proteins", *Proteomics*, vol. 3, no. 8, pp. 1418-1424.

Docampo, R. & Moreno, S. N. 2003, "Current chemotherapy of human African trypanosomiasis", *Parasitol.Res.*, vol. 90 Supp 1:S10-3., p. S10-S13.

Domon, B. & Aebersold, R. 2006, "Mass spectrometry and protein analysis", *Science*, vol. 312, no. 5771, pp. 212-217.

Donelson, J. E. 2003, "Antigenic variation and the African trypanosome genome", *Acta Trop.*, vol. 85, no. 3, pp. 391-404.

Doudican, N. A., Song, B., Shadel, G. S., & Doetsch, P. W. 2005, "Oxidative DNA damage causes mitochondrial genomic instability in *Saccharomyces cerevisiae*", *Mol.Cell Biol.*, vol. 25, no. 12, pp. 5196-5204.

Dreger, M., Bengtsson, L., Schoneberg, T., Otto, H., & Hucho, F. 2001, "Nuclear envelope proteomics: novel integral membrane proteins of the inner nuclear membrane", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 98, no. 21, pp. 11943-11948.

Edelman, J. 1993, "Quadratic minimization of predictors for protein secondary structure. Application to transmembrane alpha-helices", *J.Mol.Biol.*, vol. 232, no. 1, pp. 165-191.

Egbe-Nwiyi, T. N., Igbokwe, I. O., & Onyeyili, P. A. 2005, "Diminazene aceturate resistance on the virulence of *Trypanosoma brucei* for rats", *J.Comp Pathol.*, vol. 133, no. 4, pp. 286-288.

Ekwanzala, M., Pepin, J., Khonde, N., Molisho, S., Bruneel, H., & De Wals, P. 1996, "In the heart of darkness: sleeping sickness in Zaire", *Lancet*, vol. 348, no. 9039, pp. 1427-1430.

Enanga, B., Burchmore, R. J., Stewart, M. L., & Barrett, M. P. 2002, "Sleeping sickness and the brain", *Cell Mol.Life Sci.*, vol. 59, no. 5, pp. 845-858.

Ens, W. & Standing, K. G. 2005, "Hybrid quadrupole/time-of-flight mass spectrometers for analysis of biomolecules", *Methods Enzymol.*, vol. 402, pp. 49-78.

Fairlamb, A. H. 2003, "Chemotherapy of human African trypanosomiasis: current and future prospects", *Trends Parasitol.*, vol. 19, no. 11, pp. 488-494.

Fairlamb, A. H., Carter, N. S., Cunningham, M., & Smith, K. 1992, "Characterisation of melarsen-resistant *Trypanosoma brucei brucei* with respect to cross-resistance to other drugs and trypanothione metabolism", *Mol.Biochem.Parasitol.*, vol. 53, no. 1-2, pp. 213-222.

Fairlamb, A. H., Henderson, G. B., Bacchi, C. J., & Cerami, A. 1987, "*In vivo* effects of difluoromethylornithine on trypanothione and polyamine levels in bloodstream forms of *Trypanosoma brucei*", *Mol.Biochem.Parasitol.*, vol. 24, no. 2, pp. 185-191.

Fandiño, A. S., Rais, I., Vollmer, M., Elgass, H., Schagger, H., & Karas, M. 2005, "LC-nanospray-MS/MS analysis of hydrophobic proteins from membrane protein complexes isolated by blue-native electrophoresis", *J.Mass Spectrom.*, vol. 40, no. 9, pp. 1223-1231.

Felschow, D. M., Mi, Z., Stanek, J., Frei, J., & Porter, C. W. 1997, "Selective labelling of cell-surface polyamine-binding proteins on leukaemic and solid-tumour cell types using a new polyamine photoprobe", *Biochem.J.*, vol. 328, no. Pt 3, pp. 889-895.

Felsenfeld, G. 1978, "Chromatin", *Nature*, vol. 271, no. 5641, pp. 115-122.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. 1989, "Electrospray ionization for mass spectrometry of large biomolecules", *Science*, vol. 246, no. 4926, pp. 64-71.

Ferguson, M. A., Brimacombe, J. S., Brown, J. R., Crossman, A., Dix, A., Field, R. A., Guther, M. L., Milne, K. G., Sharma, D. K., & Smith, T. K. 1999, "The GPI biosynthetic pathway as a therapeutic target for African sleeping sickness", *Biochim.Biophys.Acta*, vol. 1455, no. 2-3, pp. 327-340.

Ferro, M., Salvi, D., Riviere-Rolland, H., Vermat, T., Seigneurin-Berny, D., Grunwald, D., Garin, J., Joyard, J., & Rolland, N. 2002, "Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 99, no. 17, pp. 11487-11492.

Ferro, M., Seigneurin-Berny, D., Rolland, N., Chapel, A., Salvi, D., Garin, J., & Joyard, J. 2000, "Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins", *Electrophoresis*, vol. 21, no. 16, pp. 3517-3526.

Fey, S. J. & Larsen, P. M. 2001, "2D or not 2D. Two-dimensional gel electrophoresis", *Curr.Opin.Chem.Biol.*, vol. 5, no. 1, pp. 26-33.

Fojo, T. & Bates, S. 2003, "Strategies for reversing drug resistance", *Oncogene*, vol. 22, no. 47, pp. 7512-7523.

Foucher, A. L., McIntosh, A., Douce, G., Wastling, J., Tait, A., & Turner, C. M. 2006, "A proteomic analysis of arsenical drug resistance in *Trypanosoma brucei*", *Proteomics*, vol. 6, no. 9, pp. 2726-2732.

Friso, G., Giacomelli, L., Ytterberg, A. J., Peltier, J. B., Rudella, A., Sun, Q., & Wijk, K. J. 2004, "In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database", *Plant Cell*, vol. 16, no. 2, pp. 478-499.

Frommel, T. O. & Balber, A. E. 1987, "Flow cytofluorimetric analysis of drug accumulation by multidrug-resistant *Trypanosoma brucei brucei* and *T. b. rhodesiense*", *Mol.Biochem.Parasitol.*, vol. 26, no. 1-2, pp. 183-191.

Fujiki, Y., Hubbard, A. L., Fowler, S., & Lazarow, P. B. 1982, "Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum", *J.Cell Biol.*, vol. 93, no. 1, pp. 97-102.

Furth, A. J. 1980, "Removing unbound detergent from hydrophobic proteins", *Anal.Biochem.*, vol. 109, no. 2, pp. 207-215.

Galeva, N. & Altermann, M. 2002, "Comparison of one-dimensional and two-dimensional gel electrophoresis as a separation tool for proteomic analysis of rat liver microsomes: cytochromes P450 and other membrane proteins", *Proteomics*, vol. 2, no. 6, pp. 713-722.

Gallo, J. M. & Precigout, E. 1988, "Tubulin expression in trypanosomes", *Biol.Cell*, vol. 64, no. 2, pp. 137-143.

Geerts, S., Holmes, P. H., Eisler, M. C., & Diall, O. 2001, "African bovine trypanosomiasis: the problem of drug resistance", *Trends Parasitol.*, vol. 17, no. 1, pp. 25-28.

Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., & Gygi, S. P. 2003, "Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 100, no. 12, pp. 6940-6945.

Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., & Weissman, J. S. 2003, "Global analysis of protein expression in yeast", *Nature*, vol. 425, no. 6959, pp. 737-741.

Gruhler, A., Schulze, W. X., Matthiesen, R., Mann, M., & Jensen, O. N. 2005, "Stable isotope labeling of *Arabidopsis thaliana* cells and quantitative proteomics by mass spectrometry", *Mol.Cell Proteomics*, vol. 4, no. 11, pp. 1697-1709.

Guerrera, I. C. & Kleiner, O. 2005, "Application of mass spectrometry in proteomics", *Biosci.Rep.*, vol. 25, no. 1-2, pp. 71-93.

Gull, K. 1999, "The cytoskeleton of trypanosomatid parasites", *Annu.Rev.Microbiol.*, vol. 53, pp. 629-655.

Gutteridge, W. E. 1985, "Existing chemotherapy and its limitations", *Br.Med.Bull.*, vol. 41, no. 2, pp. 162-168.

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., & Aebersold, R. 1999a, "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags", *Nat.Biotechnol.*, vol. 17, no. 10, pp. 994-999.

Gygi, S. P., Rochon, Y., Franza, B. R., & Aebersold, R. 1999b, "Correlation between protein and mRNA abundance in yeast", *Mol.Cell Biol.*, vol. 19, no. 3, pp. 1720-1730.

Hager, K. M. & Hajduk, S. L. 1997, "Mechanism of resistance of African trypanosomes to cytotoxic human HDL", *Nature*, vol. 385, no. 6619, pp. 823-826.

Hart, S. R. & Gaskell, S. J. 2005, "LC-tandem MS in proteome characterization", *Trends in Analytical Chemistry*, vol. 24, no. 7, pp. 566-575.

Hartinger, J., Stenius, K., Högemann, D., & Jahn, R. 1996, "16-BAC/SDS-PAGE: a two-dimensional gel electrophoresis system suitable for the separation of integral membrane proteins", *Anal.Biochem.*, vol. 240, no. 1, pp. 126-133.

Helenius, A. & Simons, K. 1975, "Solubilization of membranes by detergents", *Biochim.Biophys.Acta*, vol. 415, no. 1, pp. 29-79.

Herbert, W. J. & Lumsden, W. H. 1976, "*Trypanosoma brucei*: a rapid "matching" method for estimating the host's parasitemia", *Exp.Parasitol.*, vol. 40, no. 3, pp. 427-431.

Hillenkamp, F. & Karas, M. 1990, "Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization", *Methods Enzymol.*, vol. 193:280-95., pp. 280-295.

Hirumi, H. & 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.*, vol. 75, no. 6, pp. 985-989.

Holman, G. D. & Cushman, S. W. 1994, "Subcellular localization and trafficking of the GLUT4 glucose transporter isoform in insulin-responsive cells", *Bioessays.*, vol. 16, no. 10, pp. 753-759.

Hopfgartner, G., Varesio, E., Tschappat, V., Grivet, C., Bourgogne, E., & Leuthold, L. A. 2004, "Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules", *J.Mass Spectrom.*, vol. 39, no. 8, pp. 845-855.

Howard, E. M., Zhang, H., & Roepe, P. D. 2002, "A novel transporter, Pfcrt, confers antimalarial drug resistance", *J.Membr.Biol.*, vol. 190, no. 1, pp. 1-8.

Hu, J., Qian, J., Borisov, O., Pan, S., Li, Y., Liu, T., Deng, L., Wannemacher, K., Kurnellas, M., Patterson, C., Elkabes, S., & Li, H. 2006, "Optimized proteomic analysis of a mouse model of cerebellar dysfunction using amine-specific isobaric tags", *Proteomics*, vol. 6, no. 15, pp. 4321-4334.

Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., & O'Shea, E. K. 2003, "Global analysis of protein localization in budding yeast", *Nature*, vol. 425, no. 6959, pp. 686-691.

Ikeda, M., Arai, M., Lao, D. M., & Shimizu, T. 2002, "Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies", *In Silico.Biol.*, vol. 2, no. 1, pp. 19-33.

Iten, M., Matovu, E., Brun, R., & Kaminsky, R. 1995, "Innate lack of susceptibility of Ugandan *Trypanosoma brucei rhodesiense* to DL-alpha-difluoromethylornithine (DFMO)", *Trop.Med.Parasitol.*, vol. 46, no. 3, pp. 190-194.

Ji, T. H. 1977, "A novel approach to the identification of surface receptors. The use of photosensitive hetero-bifunctional cross-linking reagent", *J.Biol.Chem.*, vol. 252, no. 5, pp. 1566-1570.

Jones, A., Faldas, A., Foucher, A., Hunt, E., Tait, A., Wastling, J. M., & Turner, C. M. 2006, "Visualisation and analysis of proteomic data from the procyclic form of *Trypanosoma brucei*", *Proteomics*, vol. 6, no. 1, pp. 259-267.

Jones, M. N. 1999, "Surfactants in membrane solubilisation", *Int.J.Pharm.*, vol. 177, no. 2, pp. 137-159.

Jones, O. T., Earnest, J. P., & McNamee, M. G. 1990, "Solubilization and reconstitution of membrane proteins," in *Biological membranes: a practical approach*, J. B. C. Findlay & W. H. Evans, eds., Oxford University Press, Oxford, pp. 139-178.

Joshi, P. P., Shegokar, V. R., Powar, R. M., Herder, S., Katti, R., Salkar, H. R., Dani, V. S., Bhargava, A., Jannin, J., & Truc, P. 2005, "Human trypanosomiasis caused by *Trypanosoma evansi* in India: the first case report", *Am.J.Trop.Med.Hyg.*, vol. 73, no. 3, pp. 491-495.

Julka, S. & Regnier, F. E. 2005, "Recent advancements in differential proteomics based on stable isotope coding", *Brief.Funct.Genomic.Proteomic.*, vol. 4, no. 2, pp. 158-177.

Kaminsky, R. & Mäser, P. 2000, "Drug Resistance in African Trypanosomes", *Current Opinion in Anti-infective Investigational Drugs*, vol. 2, no. 1, pp. 76-82.

Kapuscinski, J. 1995, "DAPI: a DNA-specific fluorescent probe", *Biotech.Histochem.*, vol. 70, no. 5, pp. 220-233.

Karas, M., Gluckmann, M., & Schafer, J. 2000, "Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors", *J.Mass Spectrom.*, vol. 35, no. 1, pp. 1-12.

Karas, M. & Kruger, R. 2003, "Ion formation in MALDI: the cluster ionization mechanism", *Chem.Rev.*, vol. 103, no. 2, pp. 427-440.

Kaur, J. & Dey, C. S. 2000, "Putative P-glycoprotein expression in arsenite-resistant *Leishmania donovani* down-regulated by verapamil", *Biochem.Biophys.Res.Commun.*, vol. 271, no. 3, pp. 615-619.

Keller, A., Purvine, S., Nesvizhskii, A. I., Stolyar, S., Goodlett, D. R., & Kolker, E. 2002, "Experimental protein mixture for validating tandem mass spectral analysis", *OMICS.*, vol. 6, no. 2, pp. 207-212.

Klein, I., Sarkadi, B., & Varadi, A. 1999, "An inventory of the human ABC proteins", *Biochim.Biophys.Acta*, vol. 1461, no. 2, pp. 237-262.

Klip, A., Deziel, M., & Walkert, D. 1984, "Chemical identity of the glucose transporter with the [^{3}H]cytochalasin B-photolabelled component of human erythrocyte membranes. Equal sensitivity to trypsin and endoglycosidase F", *Biochem.Biophys.Res.Commun.*, vol. 122, no. 1, pp. 218-224.

Klokouzas, A., Shahi, S., Hladky, S. B., Barrand, M. A., & van Veen, H. W. 2003, "ABC transporters and drug resistance in parasitic protozoa", *Int.J.Antimicrob.Agents*, vol. 22, no. 3, pp. 301-317.

Kramer, M. L. 2006, "A new multiphasic buffer system for benzyldimethyl-n-hexadecylammonium chloride polyacrylamide gel electrophoresis of proteins providing efficient stacking", *Electrophoresis*, vol. 27, no. 2, pp. 347-356.

Krijgsveld, J., Ketting, R. F., Mahmoudi, T., Johansen, J., Artal-Sanz, M., Verrijzer, C. P., Plasterk, R. H., & Heck, A. J. 2003, "Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics", *Nat.Biotechnol.*, vol. 21, no. 8, pp. 927-931.

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. 2001, "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes", *J.Mol.Biol.*, vol. 305, no. 3, pp. 567-580.

Kuhn, K., Prinz, T., Schafer, J., Baumann, C., Scharfke, M., Kienle, S., Schwarz, J., Steiner, S., & Hamon, C. 2005, "Protein sequence tags: a novel solution for comparative proteomics", *Proteomics*, vol. 5, no. 9, pp. 2364-2368.

Kyte, J. & Doolittle, R. F. 1982, "A simple method for displaying the hydropathic character of a protein", *J.Mol.Biol.*, vol. 157, no. 1, pp. 105-132.

Lahaye, A., Stahl, H., Thines-Sempoux, D., & Foury, F. 1991, "PIF1: a DNA helicase in yeast mitochondria", *EMBO J.*, vol. 10, no. 4, pp. 997-1007.

Lane, C. S. 2005, "Mass spectrometry-based proteomics in the life sciences", *Cell Mol.Life Sci.*, vol. 62, no. 7-8, pp. 848-869.

Lanteri, C., Stewart, M., Brock, J., Alibu, V. P., Meshnick, S., Tidwell, R. R., & Barrett, M. P. 2006, "Roles for the *Trypanosoma brucei* P2 transporter in DB75 uptake and resistance", *Mol.Pharmacol.*, vol. 70, no. 5, pp. 1585-1592.

Legros, D., Evans, S., Maiso, F., Enyaru, J. C., & Mbulamberi, D. 1999, "Risk factors for treatment failure after melarsoprol for *Trypanosoma brucei gambiense* trypanosomiasis in Uganda", *Trans.R.Soc.Trop.Med.Hyg.*, vol. 93, no. 4, pp. 439-442.

Legros, D., Ollivier, G., Gastellu-Etchegorry, M., Paquet, C., Burri, C., Jannin, J., & Buscher, P. 2002, "Treatment of human African trypanosomiasis--present situation and needs for research and development", *Lancet Infect.Dis.*, vol. 2, no. 7, pp. 437-440.

Li, F., Hua, S. B., Wang, C. C., & Gottesdiener, K. M. 1998, "*Trypanosoma brucei brucei*: characterization of an ODC null bloodstream form mutant and the action of alpha-difluoromethylornithine", *Exp.Parasitol.*, vol. 88, no. 3, pp. 255-257.

Liu, H., Lin, D., & Yates, J. R., III 2002, "Multidimensional separations for protein/peptide analysis in the post-genomic era", *Biotechniques*, vol. 32, no. 4, pp. 898, 900, 902.

Liu, H., Sadygov, R. G., & Yates, J. R., III 2004, "A model for random sampling and estimation of relative protein abundance in shotgun proteomics", *Anal.Chem.*, vol. 76, no. 14, pp. 4193-4201.

Loo, J. A., Edmonds, C. G., & Smith, R. D. 1990, "Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry", *Science*, vol. 248, no. 4952, pp. 201-204.

Ludewig, G. & Staben, C. 1994, "Characterization of the PNT1 pentamidine resistance gene of *Saccharomyces cerevisiae*", *Antimicrob.Agents Chemother.*, vol. 38, no. 12, pp. 2850-2856.

Ludewig, G., Williams, J. M., Li, Y., & Staben, C. 1994, "Effects of pentamidine isethionate on *Saccharomyces cerevisiae*", *Antimicrob.Agents Chemother.*, vol. 38, no. 5, pp. 1123-1128.

Lüscher, A., Nerima, B., & Mäser, P. 2006, "Combined contribution of TbAT1 and TbMRPA to drug resistance in *Trypanosoma brucei*", *Mol.Biochem.Parasitol.*, vol. 150, no. 2, pp. 364-366.

Macfarlane, D. E. 1989, "Two dimensional benzyldimethyl-n-hexadecylammonium chloride----sodium dodecyl sulfate preparative polyacrylamide gel electrophoresis: a high capacity high resolution technique for the purification of proteins from complex mixtures", *Anal.Biochem.*, vol. 176, no. 2, pp. 457-463.

Mann, M., Hendrickson, R. C., & Pandey, A. 2001, "Analysis of proteins and proteomes by mass spectrometry", *Annu.Rev.Biochem.*, vol. 70:437-73., pp. 437-473.

Mann, M. & Jensen, O. N. 2003, "Proteomic analysis of post-translational modifications", *Nat.Biotechnol.*, vol. 21, no. 3, pp. 255-261.

Marmagne, A., Rouet, M. A., Ferro, M., Rolland, N., Alcon, C., Joyard, J., Garin, J., Barbier-Brygoo, H., & Ephritikhine, G. 2004, "Identification of new intrinsic proteins in Arabidopsis plasma membrane proteome", *Mol.Cell Proteomics*, vol. 3, no. 7, pp. 675-691.

Mäser, P. & Kaminsky, R. 1998, "Identification of three ABC transporter genes in *Trypanosoma brucei spp*", *Parasitol.Res.*, vol. 84, no. 2, pp. 106-111.

Mäser, P., Sutterlin, C., Kralli, A., & Kaminsky, R. 1999, "A nucleoside transporter from *Trypanosoma brucei* involved in drug resistance", *Science*, vol. 285, no. 5425, pp. 242-244.

Maslov, D. A., Podlipaev, S. A., & Lukes, J. 2001, "Phylogeny of the kinetoplastida: taxonomic problems and insights into the evolution of parasitism", *Mem.Inst.Oswaldo Cruz*, vol. 96, no. 3, pp. 397-402.

Mathis, A. M., Holman, J. L., Sturk, L. M., Ismail, M. A., Boykin, D. W., Tidwell, R. R., & Hall, J. E. 2006, "Accumulation and intracellular distribution of antitrypanosomal diamidine compounds DB75 and DB820 in African trypanosomes", *Antimicrob.Agents Chemother.*, vol. 50, no. 6, pp. 2185-2191.

Matovu, E., Enyaru, J. C., Legros, D., Schmid, C., Seebeck, T., & Kaminsky, R. 2001a, "Melarsoprol refractory *T. b. gambiense* from Omugo, north-western Uganda", *Trop.Med.Int.Health*, vol. 6, no. 5, pp. 407-411.

Matovu, E., Geiser, F., Schneider, V., Mäser, P., Enyaru, J. C., Kaminsky, R., Gallati, S., & Seebeck, T. 2001b, "Genetic variants of the TbAT1 adenosine transporter from African trypanosomes in relapse infections following melarsoprol therapy", *Mol.Biochem.Parasitol.*, vol. 117, no. 1, pp. 73-81.

Matovu, E., Seebeck, T., Enyaru, J. C., & Kaminsky, R. 2001c, "Drug resistance in *Trypanosoma brucei spp.*, the causative agents of sleeping sickness in man and nagana in cattle", *Microbes.Infect.*, vol. 3, no. 9, pp. 763-770.

Matovu, E., Stewart, M. L., Geiser, F., Brun, R., Mäser, P., Wallace, L. J., Burchmore, R. J., Enyaru, J. C., Barrett, M. P., Kaminsky, R., Seebeck, T., & De Koning, H. P. 2003, "Mechanisms of Arsenical and Diamidine Uptake and Resistance in *Trypanosoma brucei*", *Eukaryot.Cell*, vol. 2, no. 5, pp. 1003-1008.

McCulloch, R. & Barry, J. D. 1999, "A role for RAD51 and homologous recombination in *Trypanosoma brucei* antigenic variation", *Genes Dev.*, vol. 13, no. 21, pp. 2875-2888.

Mehlitz, D., Zillmann, U., Scott, C. M., & Godfrey, D. G. 1982, "Epidemiological studies on the animal reservoir of Gambiense sleeping sickness. Part III. Characterization of trypanozoon stocks by isoenzymes and sensitivity to human serum", *Tropenmed.Parasitol.*, vol. 33, no. 2, pp. 113-118.

Metcalf, B. W., Bey, P., Danzin, C., Jung, M., Casara, P., & Vevert, J. P. 1978, "Catalytic irreversible inhibition of mammalian ornithine decarboxylase (E.C.4.1.1.17) by substrate and product analogs", *J.Am.Chem.Soc*, vol. 100, no. 8, pp. 2551-2553.

Möller, S., Croning, M. D., & Apweiler, R. 2001, "Evaluation of methods for the prediction of membrane spanning regions", *Bioinformatics.*, vol. 17, no. 7, pp. 646-653.

Molloy, M. P., Herbert, B. R., Slade, M. B., Rabilloud, T., Nouwens, A. S., Williams, K. L., & Gooley, A. A. 2000, "Proteomic analysis of the *Escherichia coli* outer membrane", *Eur.J.Biochem.*, vol. 267, no. 10, pp. 2871-2881.

Molyneux, D. H. 2001, "Sterile insect release and trypanosomiasis control: a plea for realism", *Trends Parasitol.*, vol. 17, no. 9, pp. 413-414.

Murray, M., Trail, J. C., & D'Ieteren, G. D. 1990, "Trypanotolerance in cattle and prospects for the control of trypanosomiasis by selective breeding", *Rev.Sci.Tech.*, vol. 9, no. 2, pp. 369-386.

Mussmann, R., Engstler, M., Gerrits, H., Kieft, R., Toaldo, C. B., Onderwater, J., Koerten, H., van Luenen, H. G., & Borst, P. 2004, "Factors affecting the level and localization of the transferrin receptor in *Trypanosoma brucei*", *J.Biol.Chem.*, vol. 279, no. 39, pp. 40690-40698.

Naula, C. & Burchmore, R. J. 2003, "A plethora of targets, a paucity of drugs: progress towards the development of novel chemotherapies for human African trypanosomiasis", *Expert.Rev.Anti.Infect.Ther.*, vol. 1, no. 1, pp. 157-165.

Naula, C., Parsons, M., & Mottram, J. C. 2005, "Protein kinases as drug targets in trypanosomes and *Leishmania*", *Biochim.Biophys.Acta*, vol. 1754, no. 1-2, pp. 151-159.

Naula, C. & Seebeck, T. 2000, "Cyclic AMP signaling in trypanosomatids", *Parasitol.Today*, vol. 16, no. 1, pp. 35-38.

Njiokou, F., Laveissiere, C., Simo, G., Nkinin, S., Grebaut, P., Cuny, G., & Herder, S. 2006, "Wild fauna as a probable animal reservoir for *Trypanosoma brucei gambiense* in Cameroon", *Infect.Genet.Evol.*, vol. 6, no. 2, pp. 147-153.

Njiru, Z. K., Ndung'u, K., Matete, G., Ndungu, J. M., & Gibson, W. C. 2004, "Detection of *Trypanosoma brucei rhodesiense* in animals from sleeping sickness foci in East Africa using the serum resistance associated (SRA) gene", *Acta Trop.*, vol. 90, no. 3, pp. 249-254.

O'Farrell, P. H. 1975, "High resolution two-dimensional electrophoresis of proteins", *J.Biol.Chem.*, vol. 250, no. 10, pp. 4007-4021.

O'Rourke, T. W., Doudican, N. A., Zhang, H., Eaton, J. S., Doetsch, P. W., & Shadel, G. S. 2005, "Differential involvement of the related DNA helicases Pif1p and Rrm3p in mtDNA point mutagenesis and stability", *Gene*, vol. 354, pp. 86-92.

Ogada, T. 1974, "Clinical Mel B, resistance in Rhodesian sleeping sickness", *East Afr.Med.J.*, vol. 51, no. 1, pp. 56-59.

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. 2002, "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics", *Mol.Cell Proteomics*, vol. 1, no. 5, pp. 376-386.

Ong, S. E., Foster, L. J., & Mann, M. 2003a, "Mass spectrometric-based approaches in quantitative proteomics", *Methods*, vol. 29, no. 2, pp. 124-130.

Ong, S. E., Kratchmarova, I., & Mann, M. 2003b, "Properties of ^{13}C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC)", *J.Proteome.Res.*, vol. 2, no. 2, pp. 173-181.

Opperdoes, F. R. & Borst, P. 1977, "Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei*: the glycosome", *FEBS Lett.*, vol. 80, no. 2, pp. 360-364.

Osman, A. S., Jennings, F. W., & Holmes, P. H. 1992, "The rapid development of drug-resistance by *Trypanosoma evansi* in immunosuppressed mice", *Acta Trop.*, vol. 50, no. 3, pp. 249-257.

Pandey, A. & Mann, M. 2000, "Proteomics to study genes and genomes", *Nature*, vol. 405, no. 6788, pp. 837-846.

Papadopoulos, M. C., Abel, P. M., Agranoff, D., Stich, A., Tarelli, E., Bell, B. A., Planche, T., Loosemore, A., Saadoun, S., Wilkins, P., & Krishna, S. 2004, "A novel and accurate diagnostic test for human African trypanosomiasis", *Lancet*, vol. 363, no. 9418, pp. 1358-1363.

Pappin, D. J., Hojrup, P., & Bleasby, A. J. 1993, "Rapid identification of proteins by peptide-mass fingerprinting", *Curr.Biol.*, vol. 3, no. 6, pp. 327-332.

Pedersen, S. K., Harry, J. L., Sebastian, L., Baker, J., Traini, M. D., McCarthy, J. T., Manoharan, A., Wilkins, M. R., Gooley, A. A., Righetti, P. G., Packer, N. H., Williams, K. L., & Herbert, B. R. 2003, "Unseen proteome: mining below the tip of the iceberg to find low abundance and membrane proteins", *J.Proteome.Res.*, vol. 2, no. 3, pp. 303-311.

Penchenier, L., Grebaut, P., Njokou, F., Eboo, E., V, & Buscher, P. 2003, "Evaluation of LATEX/*T.b.gambiense* for mass screening of *Trypanosoma brucei gambiense* sleeping sickness in Central Africa", *Acta Trop.*, vol. 85, no. 1, pp. 31-37.

Penchenier, L., Simo, G., Grebaut, P., Nkinin, S., Laveissiere, C., & Herder, S. 2000, "Diagnosis of human trypanosomiasis, due to *Trypanosoma brucei gambiense* in central

Africa, by the polymerase chain reaction", *Trans.R.Soc.Trop.Med.Hyg.*, vol. 94, no. 4, pp. 392-394.

Pepin, J. & Milord, F. 1994, "The treatment of human African trypanosomiasis", *Adv.Parasitol.*, vol. 33:1-47., pp. 1-47.

Perkins, D. N., Pappin, D. J., Creasy, D. M., & Cottrell, J. S. 1999, "Probability-based protein identification by searching sequence databases using mass spectrometry data", *Electrophoresis*, vol. 20, no. 18, pp. 3551-3567.

Phillips, M. A. & Wang, C. C. 1987, "A *Trypanosoma brucei* mutant resistant to alpha-difluoromethylornithine", *Mol.Biochem.Parasitol.*, vol. 22, no. 1, pp. 9-17.

Picozzi, K., Fevre, E. M., Odiit, M., Carrington, M., Eisler, M. C., Maudlin, I., & Welburn, S. C. 2005, "Sleeping sickness in Uganda: a thin line between two fatal diseases", *BMJ*, vol. 331, no. 7527, pp. 1238-1241.

Pink, R., Hudson, A., Mouries, M. A., & Bendig, M. 2005, "Opportunities and challenges in antiparasitic drug discovery", *Nat.Rev.Drug Discov.*, vol. 4, no. 9, pp. 727-740.

Pradines, B., Pages, J. M., & Barbe, J. 2005, "Chemosensitizers in drug transport mechanisms involved in protozoan resistance", *Curr.Drug Targets.Infect.Disord.*, vol. 5, no. 4, pp. 411-431.

Ramus, C., de Peredo, A. G., Dahout, C., Gallagher, M., & Garin, J. 2006, "An optimized strategy for ICAT quantification of membrane proteins", *Mol.Cell Proteomics*, vol. 5, no. 1, pp. 68-78.

Räz, B., Iten, M., Grether-Buhler, Y., Kaminsky, R., & Brun, R. 1997, "The Alamar Blue assay to determine drug sensitivity of African trypanosomes (*T.b. rhodesiense* and *T.b. gambiense*) in vitro", *Acta Trop.*, vol. 68, no. 2, pp. 139-147.

Reichard, R. E. 2002, "Area-wide biological control of disease vectors and agents affecting wildlife", *Rev.Sci.Tech.*, vol. 21, no. 1, pp. 179-185.

Resing, K. A., Meyer-Arendt, K., Mendoza, A. M., Aveline-Wolf, L. D., Jonscher, K. R., Pierce, K. G., Old, W. M., Cheung, H. T., Russell, S., Wattawa, J. L., Goehle, G. R., Knight, R. D., & Ahn, N. G. 2004, "Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics", *Anal.Chem.*, vol. 76, no. 13, pp. 3556-3568.

Rhoads, A. R. & Friedberg, F. 1997, "Sequence motifs for calmodulin recognition", *FASEB J.*, vol. 11, no. 5, pp. 331-340.

Righetti, P. G., Castagna, A., Antonioli, P., & Boschetti, E. 2005, "Prefractionation techniques in proteome analysis: the mining tools of the third millennium", *Electrophoresis*, vol. 26, no. 2, pp. 297-319.

Rogers, D. J. & Randolph, S. E. 2002, "A response to the aim of eradicating tsetse from Africa", *Trends Parasitol.*, vol. 18, no. 12, pp. 534-536.

Rollo, I. M. & Williamson, J. 1951, "Acquired resistance to 'Melarsen', tryparsamide and amidines in pathogenic trypanosomes after treatment with 'Melarsen' alone", *Nature*, vol. 167, no. 4239, pp. 147-148.

Romijn, E. P., Krijgsveld, J., & Heck, A. J. 2003, "Recent liquid chromatographic-(tandem) mass spectrometric applications in proteomics", *J.Chromatogr.A*, vol. 1000, no. 1-2, pp. 589-608.

Sadygov, R. G., Cociorva, D., & Yates, J. R., III 2004, "Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book", *Nat.Methods*, vol. 1, no. 3, pp. 195-202.

Sanchez, C. P., McLean, J. E., Rohrbach, P., Fidock, D. A., Stein, W. D., & Lanzer, M. 2005, "Evidence for a pfcrt-associated chloroquine efflux system in the human malarial parasite *Plasmodium falciparum*", *Biochemistry*, vol. 44, no. 29, pp. 9862-9870.

Sanchez, C. P., Stein, W., & Lanzer, M. 2003, "Trans stimulation provides evidence for a drug efflux carrier as the mechanism of chloroquine resistance in *Plasmodium falciparum*", *Biochemistry*, vol. 42, no. 31, pp. 9383-9394.

Schirmer, E. C., Florens, L., Guan, T., Yates, J. R., III, & Gerace, L. 2003, "Nuclear membrane proteins with potential disease links found by subtractive proteomics", *Science*, vol. 301, no. 5638, pp. 1380-1382.

Schofield, C. J. & Dias, J. C. 1999, "The Southern Cone Initiative against Chagas disease", *Adv.Parasitol.*, vol. 42, pp. 1-27.

Schofield, C. J. & Maudlin, I. 2001, "Trypanosomiasis control", *Int.J.Parasitol.*, vol. 31, no. 5-6, pp. 614-619.

Schulz, G. E. 2002, "The structure of bacterial outer membrane proteins", *Biochim.Biophys.Acta*, vol. 1565, no. 2, pp. 308-317.

Schulz, V. P. & Zakian, V. A. 1994, "The *Saccharomyces* PIF1 DNA helicase inhibits telomere elongation and *de novo* telomere formation", *Cell*, vol. 76, no. 1, pp. 145-155.

Scott, A. G., Tait, A., & Turner, C. M. 1996, "Characterisation of cloned lines of *Trypanosoma brucei* expressing stable resistance to MelCy and suramin", *Acta Trop.*, vol. 60, no. 4, pp. 251-262.

Seebeck, T., Schaub, R., & Johner, A. 2004, "cAMP signalling in the kinetoplastid protozoa", *Curr.Mol.Med.*, vol. 4, no. 6, pp. 585-599.

Seed, J. R. 2000, "Current status of African Trypanosomiasis", *ASM News*, vol. 66, pp. 395-402.

Seed, J. R. 2001, "African trypanosomiasis research: 100 years of progress, but questions and problems still remain", *Int.J.Parasitol.*, vol. 31, no. 5-6, pp. 434-442.

Seed, J. R. & Wenck, M. A. 2003, "Role of the long slender to short stumpy transition in the life cycle of the African trypanosomes", *Kinetoplastid.Biol.Dis.*, vol. 2, no. 1, p. 3.

Shahi, S. K., Krauth-Siegel, R. L., & Clayton, C. E. 2002, "Overexpression of the putative thiol conjugate transporter TbMRPA causes melarsoprol resistance in *Trypanosoma brucei*", *Mol.Microbiol.*, vol. 43, no. 5, pp. 1129-1138.

Shaked-Mishan, P., Suter-Grotemeyer, M., Yoel-Almagor, T., Holland, N., Zilberstein, D., & Rentsch, D. 2006, "A novel high-affinity arginine transporter from the human parasitic protozoan *Leishmania donovani*", *Mol.Microbiol.*, vol. 60, no. 1, pp. 30-38.

Shapiro, T. A. & Englund, P. T. 1990, "Selective cleavage of kinetoplast DNA minicircles promoted by antitrypanosomal drugs", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 87, no. 3, pp. 950-954.

Shapiro, T. A. & Englund, P. T. 1995, "The structure and replication of kinetoplast DNA", *Annu.Rev.Microbiol.*, vol. 49:117-43., pp. 117-143.

Sharp, P. M. & Li, W. H. 1987, "The Codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications", *Nucleic Acids Res.*, vol. 15, no. 3, pp. 1281-1295.

Sharp, P. M., Stenico, M., Peden, J. F., & Lloyd, A. T. 1993, "Codon usage: mutational bias, translational selection, or both?", *Biochem.Soc.Trans.*, vol. 21, no. 4, pp. 835-841.

Sharp, P. M., Tuohy, T. M., & Mosurski, K. R. 1986, "Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes", *Nucleic Acids Res.*, vol. 14, no. 13, pp. 5125-5143.

Singh, N. 2006, "Drug resistance mechanisms in clinical isolates of *Leishmania donovani*", *Indian J.Med.Res.*, vol. 123, no. 3, pp. 411-422.

Smith, T. K., Crossman, A., Brimacombe, J. S., & Ferguson, M. A. 2004, "Chemical validation of GPI biosynthesis as a drug target against African sleeping sickness", *EMBO J.*, vol. 23, no. 23, pp. 4701-4708.

Stebeck, C. E., Beecroft, R. P., Singh, B. N., Jardim, A., Olafson, R. W., Tuckey, C., Prenevost, K. D., & Pearson, T. W. 1995, "Kinetoplastid membrane protein-11 (KMP-11) is differentially expressed during the life cycle of African trypanosomes and is found in a wide variety of kinetoplastid parasites", *Mol.Biochem.Parasitol.*, vol. 71, no. 1, pp. 1-13.

Steen, H. & Mann, M. 2004, "The ABC's (and XYZ's) of peptide sequencing", *Nat.Rev.Mol.Cell Biol.*, vol. 5, no. 9, pp. 699-711.

Stenoien, H. K. 2005, "Adaptive basis of codon usage in the haploid moss *Physcomitrella patens*", *Heredity*, vol. 94, no. 1, pp. 87-93.

Stewart, M. L., Krishna, S., Burchmore, R. J., Brun, R., De Koning, H. P., Boykin, D. W., Tidwell, R. R., Hall, J. E., & Barrett, M. P. 2005, "Detection of arsenical drug resistance in *Trypanosoma brucei* with a simple fluorescence test", *Lancet*, vol. 366, no. 9484, pp. 486-487.

Stich, A., Abel, P. M., & Krishna, S. 2002, "Human African trypanosomiasis", *BMJ*, vol. 325, no. 7357, pp. 203-206.

Stich, A., Barrett, M. P., & Krishna, S. 2003, "Waking up to sleeping sickness", *Trends Parasitol.*, vol. 19, no. 5, pp. 195-197.

Taylor, S. W., Fahy, E., & Ghosh, S. S. 2003, "Global organellar proteomics", *Trends Biotechnol.*, vol. 21, no. 2, pp. 82-88.

TDR Strategic Direction for Research 2002, *African trypanosomiasis.*

Tomlinson, S. & Raper, J. 1996, "The lysis of *Trypanosoma brucei brucei* by human serum", *Nat.Biotechnol.*, vol. 14, no. 6, pp. 717-721.

Torr, S. J., Hargrove, J. W., & Vale, G. A. 2005, "Towards a rational policy for dealing with tsetse", *Trends Parasitol.*, vol. 21, no. 11, pp. 537-541.

Tusnády, G. E. & Simon, I. 1998, "Principles governing amino acid composition of integral membrane proteins: application to topology prediction", *J.Mol.Biol.*, vol. 283, no. 2, pp. 489-506.

Tusnády, G. E. & Simon, I. 2001, "The HMMTOP transmembrane topology prediction server", *Bioinformatics.*, vol. 17, no. 9, pp. 849-850.

Tyers, M. & Mann, M. 2003, "From genomics to proteomics", *Nature*, vol. 422, no. 6928, pp. 193-197.

Ullman, B. 1995, "Multidrug resistance and P-glycoproteins in parasitic protozoa", *J.Bioenerg.Biomembr.*, vol. 27, no. 1, pp. 77-84.

Ulmschneider, M. B. & Sansom, M. S. 2001, "Amino acid distributions in integral membrane protein structures", *Biochim.Biophys.Acta*, vol. 1512, no. 1, pp. 1-14.

Ünlü, M., Morgan, M. E., & Minden, J. S. 1997, "Difference gel electrophoresis: a single gel method for detecting changes in protein extracts", *Electrophoresis*, vol. 18, no. 11, pp. 2071-2077.

van Montfort, B. A., Canas, B., Duurkens, R., Godovac-Zimmermann, J., & Robillard, G. T. 2002a, "Improved in-gel approaches to generate peptide maps of integral membrane proteins with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry", *J.Mass Spectrom.*, vol. 37, no. 3, pp. 322-330.

van Montfort, B. A., Doeven, M. K., Canas, B., Veenhoff, L. M., Poolman, B., & Robillard, G. T. 2002b, "Combined in-gel tryptic digestion and CNBr cleavage for the generation of peptide maps of an integral membrane protein with MALDI-TOF mass spectrometry", *Biochim.Biophys.Acta*, vol. 1555, no. 1-3, pp. 111-115.

Van Nieuwenhove, S. 1992, "Advances in sleeping sickness therapy", *Ann.Soc.Belg.Med.Trop.*, vol. 72 Suppl 1, pp. 39-51.

Van Nieuwenhove, S. 1999, "Present strategies in the treatment of human African trypanosomiasis," in *Progress in human African trypanosomiasis, sleeping sickness*, M. Dumas, B. Bouteille, & A. Buguet, eds., Springer-Verlag, Paris, France, pp. 253-280.

Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D. P., Lins, L., Van Den, A. J., Pays, A., Tebabi, P., Van Xong, H., Jacquet, A., Moguilevsky, N., Dieu, M., Kane, J. P., De Baetselier, P., Brasseur, R., & Pays, E. 2003, "Apolipoprotein L-I is the trypanosome lytic factor of human serum", *Nature*, vol. 422, no. 6927, pp. 83-87.

Vestal, M. L. & Campbell, J. M. 2005, "Tandem time-of-flight mass spectrometry", *Methods Enzymol.*, vol. 402, pp. 79-108.

Vickerman, K. & Tetley, L. 1977, "Recent ultrastructural studies on trypanosomes", *Ann.Soc.Belg.Med.Trop.*, vol. 57, no. 4-5, pp. 441-457.

Vinhaes, M. C. & Schofield, C. J. 2003, "Trypanosomiasis control: surmounting diminishing returns", *Trends Parasitol.*, vol. 19, no. 3, pp. 112-113.

von Heijne, G. 1994, "Membrane proteins: from sequence to structure", *Annu.Rev.Biophys.Biomol.Struct.*, vol. 23:167-92., pp. 167-192.

Voorheis, H. P., Gale, J. S., Owen, M. J., & Edwards, W. 1979, "The isolation and partial characterization of the plasma membrane from *Trypanosoma brucei*", *Biochem.J.*, vol. 180, no. 1, pp. 11-24.

Vreysen, M. J. 2001, "Principles of area-wide integrated tsetse fly control using the sterile insect technique", *Med.Trop.(Mars.).*, vol. 61, no. 4-5, pp. 397-411.

Wang, C. C. 1995, "Molecular mechanisms and therapeutic approaches to the treatment of African trypanosomiasis", *Annu.Rev.Pharmacol.Toxicol.*, vol. 35:93-127., pp. 93-127.

Wang, H. & Hanash, S. 2005, "Intact-protein based sample preparation strategies for proteome analysis in combination with mass spectrometry", *Mass Spectrom.Rev.*, vol. 24, no. 3, pp. 413-426.

Wang, Y. K., Ma, Z., Quinn, D. F., & Fu, E. W. 2001, "Inverse ^{18}O labeling mass spectrometry for the rapid identification of marker/target proteins", *Anal.Chem.*, vol. 73, no. 15, pp. 3742-3750.

Washburn, M. P., Wolters, D., & Yates, J. R., III 2001, "Large-scale analysis of the yeast proteome by multidimensional protein identification technology", *Nat.Biotechnol.*, vol. 19, no. 3, pp. 242-247.

Wei, J., Sun, J., Yu, W., Jones, A., Oeller, P., Keller, M., Woodnutt, G., & Short, J. M. 2005, "Global proteome discovery using an online three-dimensional LC-MS/MS", *J.Proteome.Res.*, vol. 4, no. 3, pp. 801-808.

WHO 1998, "Control and surveillance of African trypanosomiasis", *Technical Report Series*, vol. 881, no. I-VI, p. 1114.

WHO 2000a, "African Trypanosomiasis," in ***WHO report on global surveillance of epidemic-prone infectious diseases*** , pp. 95-106.

WHO 2000b, "Health Systems Improving Performance", *World Health Report 2000.*

WHO 2006, "Human African trypanosomiasis (sleeping sickness): epidemiological update", *Wkly.Epidemiol.Rec.*, vol. 81, no. 8, pp. 71-80.

Wickware, P. 2002, "Resurrecting the resurrection drug", *Nat.Med.*, vol. 8, no. 9, pp. 908-909.

Wilkins, M. R., Gasteiger, E., Sanchez, J. C., Bairoch, A., & Hochstrasser, D. F. 1998, "Two-dimensional gel electrophoresis for proteome projects: the effects of protein hydrophobicity and copy number", *Electrophoresis*, vol. 19, no. 8-9, pp. 1501-1505.

Wilson, W. D., Nguyen, B., Tanious, F. A., Mathis, A., Hall, J. E., Stephens, C. E., & Boykin, D. W. 2005, "Dications that target the DNA minor groove: compound design and preparation, DNA interactions, cellular distribution and biological activity", *Curr.Med.Chem.Anticancer Agents*, vol. 5, no. 4, pp. 389-408.

Wirth, D. 2001, *A Harvest Not Yet Reaped: Genomics to New Drugs in Leishmania and Trypanosomes.*

Wolters, D. A., Washburn, M. P., & Yates, J. R., III 2001, "An automated multidimensional protein identification technology for shotgun proteomics", *Anal.Chem.*, vol. 73, no. 23, pp. 5683-5690.

Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E., & Yates, J. R., III 2004, "Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis", *Anal.Chem.*, vol. 76, no. 17, pp. 4951-4959.

Wu, C. C. & Yates, J. R., III 2003, "The application of mass spectrometry to membrane proteomics", *Nat.Biotechnol.*, vol. 21, no. 3, pp. 262-267.

Wu, T. L. 2006, "Two-dimensional difference gel electrophoresis", *Methods Mol.Biol.*, vol. 328, pp. 71-95.

Wysocki, V. H., Resing, K. A., Zhang, Q., & Cheng, G. 2005, "Mass spectrometry of peptides and proteins", *Methods*, vol. 35, no. 3, pp. 211-222.

Xiang, R., Shi, Y., Dillon, D. A., Negin, B., Horvath, C., & Wilkins, J. A. 2004, "2D LC/MS analysis of membrane proteins from breast cancer cell lines MCF7 and BT474", *J.Proteome.Res.*, vol. 3, no. 6, pp. 1278-1283.

Yamniuk, A. P. & Vogel, H. J. 2004, "Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides", *Mol.Biotechnol.*, vol. 27, no. 1, pp. 33-57.

Yan, W. & Chen, S. S. 2005, "Mass spectrometry-based quantitative proteomic profiling", *Brief.Funct.Genomic.Proteomic.*, vol. 4, no. 1, pp. 27-38.

Yeates, C. 2003, "DB-289 Immtech International", *IDrugs.*, vol. 6, no. 11, pp. 1086-1093.

Zahedi, R. P., Meisinger, C., & Sickmann, A. 2005, "Two-dimensional benzyldimethyl-n-hexadecylammonium chloride/SDS-PAGE for membrane proteomics", *Proteomics*, vol. 5, no. 14, pp. 3581-3588.

Zhang, L., Xie, J., Wang, X., Liu, X., Tang, X., Cao, R., Hu, W., Nie, S., Fan, C., & Liang, S. 2005, "Proteomic analysis of mouse liver plasma membrane: use of differential extraction to enrich hydrophobic membrane proteins", *Proteomics*, vol. 5, no. 17, pp. 4510-4524.

Zhang, L. J., Wang, X. E., Peng, X., Wei, Y. J., Cao, R., Liu, Z., Xiong, J. X., Yin, X. F., Ping, C., & Liang, S. 2006, "Proteomic analysis of low-abundant integral plasma membrane proteins based on gels", *Cell Mol.Life Sci.*, vol. 63, no. 15, pp. 1790-1804.

Zhang, W., Zhou, G., Zhao, Y., White, M. A., & Zhao, Y. 2003, "Affinity enrichment of plasma membrane for proteomics analysis", *Electrophoresis*, vol. 24, no. 16, pp. 2855-2863.

Zhou, H., Ranish, J. A., Watts, J. D., & Aebersold, R. 2002, "Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry", *Nat.Biotechnol.*, vol. 20, no. 5, pp. 512-515.

Zhou, J., Monson, E. K., Teng, S. C., Schulz, V. P., & Zakian, V. A. 2000, "Pif1p helicase, a catalytic inhibitor of telomerase in yeast", *Science*, vol. 289, no. 5480, pp. 771-774.

Zieske, L. R. 2006, "A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies", *J.Exp.Bot.*, vol. 57, no. 7, pp. 1501-1508.