Metabolism and drug resistance in trypanosomatids

David Wildridge, BSc (Hons) MRes

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

The School of Life Sciences
The College of Medical, Veterinary, and Life Sciences

University of Glasgow

June 2012

Author's declaration

The results stated in this thesis are my own work, except where stated otherwise.

David Wildridge

Abstract

The principle aim of this project is the investigation of metabolism and mechanisms of pentamidine resistance in trypanosomatids. An understanding of these mechanisms may allow the development of novel drugs to treat Leishmaniasis and human African trypanosomiasis (HAT), caused by the protozoan parasites *Leishmania spp* and *Trypanosoma brucei*. In this study a pentamidine resistance *L. mexicana* promastigote cell line was generated *in vitro*. This cell line was 20-fold resistant to pentamidine when compared to the parental wild type cells. Furthermore, these lines were cross resistant to other diamidine compounds. A proteomic analysis of these cell lines revealed numerous changes to the proteome, with the down regulation of several flagellar proteins. A hypothesis to investigate a role of the voltage dependent anion channel (VDAC) in pentamidine resistance was also explored. The metabolomic approach involved the investigation of transketolase and the pentose phosphate pawthway. A previous study involving a transketolase knockout *T. brucei* cell line indicated that an increased sensitivity to pentamidine and methylene blue. A transketolase deficient *L. mexicana* cell line was generated to test this hypothesis in *Leishmania*, however the differences were minimal. A metabolomic analysis of the *L. mexicana tkt* null cell line (*lmtkt*-/-) revealed an increase in ribose 5-phosphate, a key substrate of transketolase. Erythrose 4-phosphate also increased in the *lmtkt*-/- cells, indicating a source of this metabolite independent of TKT. It appears that the deletion of TKT prevents any flux through the oxidative branch of the PPP returning to the glycolytic pathway. Interestingly, the *lmtkt*-/- cells do not acidify the medium to the same extent as the wild type cells; however a glucose assay indicated that both cell lines used similar quantities of glucose. This would suggest that there is a change in the metabolites excreted by the *lmtkt*-/- cell line. Finally, a global metabolomics approach was investigated using high resolution mass spectrometry. Metabolomics is a rapidly developing field in systems biology, and whilst significant improvements have been made in mass spectrometry; the ability to analyse and interpret raw metabolomic datasets on a global scale has been largely neglected. Consequently, a database program to query these complex datasets was constructed.

Table of contents

List of figures

List of tables

Definitions

ASCT	acetate:succinate CoA transferase
ADP	adenosine diphosphate
ATP	adenosine triphosphate
AVGM	Average Mass
AD	aromatic diamidines
AIA	arylimidamine
amu	atomic mass unit
bp	base pair
BLAST	Basic Local Alignment Search Tool
BBB	blood brain barrier
BSF	bloodstream form
CNS	central nervous system
CSF	cerebro-spinal fluid
Da	Dalton
°C	degrees Celsius
DB75	2,5 bis (4-amidinophenyl) furan
2-DOG	2-deoxy-D-glucose
DAPI	4', 6-Diamidino-2-phenylindole
DNA	deoxyribonucleic acid
DFMO	difluoromethylornithine
DMSO	dimethylsulfoxide
2DE	Two dimensional gel electrophoresis
DHAP	dihydroxyacetone phosphate
DiGE	Differential Gel Electrophoresis
DDT	dichlorodiphenyltrichloroethane
ES	electo spray
ELISA	Enzyme linked immuno serum assay
EDTA	ethylenediamine tretraacetic acid
EtOH	Ethanol
ESI	Electrospray Surface Ionization
ExPASy	Expert Protein Analysis System
FCS	foetal calf serum
FTP	File Transfer Protocol
FT-MS	Fourier Transform Mass Spectrometry
g	gram
gDNA	genomic DNA
hsDNA	herring sperm DNA
h	hour
HAT	human African trypanosomiasis
HAPT	High Affinity Pentamidine Transporter
HPLC	High Pressure Liquid Chromatography
HILIC	Hydrophilic Interaction Chromatography
IC50	50% Inhibitory Concentration
IAEA	International Atomic Energy Agency
IEF	isoelectric focussing
IPTG	isopropylthio-ß-D-galactoside
ITS1	internal transcribed spacer 1
InChI	IUPAC International Chemical Identifier
ITNs	insecticide treated bednets
kb	kilobase
kDa	Kilodalton
kg	kilogram
kDNA	kinetoplast DNA
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Luria-Bertani medium
LAPT	Low Affinity Pentamidine Transporter
LIT	linear ion traps
LC-MS	Liquid Chromatography-Mass Spectrometry
MS	Mass spectrometry
MALDI	Matrix assisted laser desorption ionisation

Tm	melting temperature
mRNA	messenger RNA
Km	Michaelis constant
µg	microgram
µl	microlitre
µM	micromolar
mg	milligram
ml	millilitre
mM	millimolar
min	minutes
M	molar
MF	Molecular Formula
MOWSE	MOlecular Weight SEarch
MIM	Mono-Isotopic Mass
MelB	Melarsoprol
MelOx	Melarsen Oxide
mtDNA	mitochondrial DNA
MCS	multiple cloning site
MMP	mitochondrial membrane potential
nm	nanometre
nM	nanomolar
NCBI	National Centre for Biotechnology Information
NAD+	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate (reduced form)
nt	nucleotide
ORF	open reading frame
OD	Optical Density
ODC	Ornithine decarboxylase
OP	oxidative phosphorylation
pI	Isoelectric point
ppm	parts per million
PMF	Peptide Mass Fingerprinting
PBS	phosphate buffered saline
PEPCK	phosphoenol pyruvate carboxylase
pmol	picomole
PAGE	Poly-Acrylamide Gel Electrophoresis
PCR	polymerase chain reaction
PPP	Pentose Phosphate Pathway
PEP	phosphoenolpyruvate
PFR	paraflagellar rod
PSG	promastigote secretion gel
RFLP	Restriction enzyme Fragment Length Polymorphism
RNA	ribonucleic acid
RNAi	RNA interference
SHAM	salicylhydroxamic acid
SSC	saline sodium citrate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
s	second
SDM	semi-defined medium
SDS	sodium dodecylsulphate
SEM	standard error around the mean
SILAC	Stable isotope labelling by amino acids in cell culture
SLP	substrate level phosphorylation
SQL	Structured Query Language
SMILES	simplified molecular-input line-entry system
SIT	sterile insect technique (SIT)
TEMED	N,N,N',N'-tetramethyl ethylenediaminide
Taq	Thermus aquaticus
TOF	Time of flight
TM	transmembrane
TbAATP	Trypanosoma brucei amino acid transporter
TbAT1	T. brucei adenosine transporter 1
TKT	Transketolase gene (functional)
TKT	Transketolase protein
TMD	Transmembrane domain

UV	ultraviolet
UTR	Untranslated Region
VSG	variable surface glycoprotein
V	volt
VBA	visual basic application
WHO	World Health Organisation
wt	wild type
X-Gal	5-bromo-4-chloro-3-indol-ß-D-galactopyranoside

Acknowledgements

First and foremost, I would like to thank my supervisors, Prof. Mike Barrett and Dr. Richard Burchmore. I will never be able to repay them for their patience, guidance, advice and support throughout my PhD. Thank you. I would also like to thank my assessors, Prof. Sylke Müller and Prof. Andy Waters, for their support and advice.

I would also like to acknowledge the assistance of Pius Alibu, Gordon Campbell, Charles Ebikeme, and Christina Naula who were invaluable at the start of my project, and to everyone else in the Barrett and Burchmore lab groups, both past and present, for their knowledge and help over the years – it was much appreciated. An enormous thank you to the Mottram Group (in particular Jim Hilley, Lesley Morrison), and to Ryan Bissett, who took the time to provide help with various aspects of molecular biology and culturing *Leishmania* parasites. Additionally, I would like to thank Fabien Jourdan, Richard Scheltema, Ludovic Cottret, Richard Orton, and David Leader for their invaluable guidance and advice on the computational elements of this project.

This project was funded by the Medical Research Council. I would like to thank them for providing me with the opportunity to do a PhD.

As well as those metioned above, I would also like to thank the following people; Aga, Alli, Amy C, Amy G, Anne, Anu, Ben, Caroline, Cat, Corinna, Daniela, Darren, Ed, Elmarie, Federica, Fed, Flo, Georgie, Jana A, Jana R, Jane, Janet, Jo, Joe, Kat, Liam, Marta, Matt, Mhairi, Mona, Nadja, Nath, Neil, Nick, Nicola, Nisha, Raquel, Richard, Rod, Sam, Seb, Sonia, and Will (and any I have accidentally missed). Their friendship over the years has been important to me, and hopefully will continue in the years to come.

The numerous Scotland football trips I've endured over the last few years have been a welcome break from my studies. Thanks to all that I've travelled with, and to those I've met along the way. From the impressive Macedonian cuisine, to the breathtaking nature of Iceland, and from Colditz Castle in Saxony, to the glacial Swiss Alps, it's been an experience!

I am very grateful to Cristina Costa, who generously gave a lot of her time and energy over the last 4 years (irrespective of geographical location!), and has been a fantastic friend throughout. Thank you.

Finally, I would like to thank my parents and family. Without their encouragement, support and love throughout my many years of education, I would not be where I am today. Thank you.

Quia quamdiu Centum ex nobis viui remanserint, nuncquam Anglorum dominio aliquatenus volumus subiugari. Non enim propter gloriam, diuicias aut honores pugnamus set propter libertatem solummodo quam Nemo bonus nisi simul cum vita amittit.

Declaration of Freedom,
An extract from the Declaration of Arbroath, 1320 AD

1 General Introduction

1.1 Leishmania and trypanosomes

Leishmania and trypanosomes are obligate protozoan parasites of the family Trypanosomatidae and order Kinetoplastida. The Kinetoplastida are defined by the presence of a kinetoplast, an intercatenated network of circular mitochondrial DNA molecules located within the single mitochondrion. The kinetoplast can be distinguished as a clearly distinct DNA containing structure, separate from the nucleus, using fluorescent DNA binding compounds such as DAPI.

Both *Leishmania* and *Trypanosoma* have digenetic life cycles with a vertebrate host and an insect vector. *Leishmania* is transmitted by sandflies of the genera *Lutzomyia* (in the New World) and *Phlebotomus* (in the Old World). African trypanosomes are transmitted by the tsetse flies of the genus *Glossina*.

1.2 Leishmaniasis

Leishmaniasis, caused by a variety of *Leishmania* species, describes a wide range of zoonotic diseases with a spectrum of clinical manifestations (Herwaldt, BL 1999). There are 21 species of *Leishmania* that are known to be infectious to humans, and are transmitted by approximately 30 species of phlebotomine sandflies (Bates, PA 2007).

There are several clinical manifestations of the disease;

Visceral Leishmaniasis (VL), also known as kala-azar, was independently described by Leishman and Donovan in India in 1903. Today, the main causative agent of VL is known as *Leishmania donovani*; prevalent in the Indian sub-continent and eastern Africa. However, New World and Mediteranean VL is caused by *L. infantum* and *L. chagasi*. Symptoms include fever, weight loss, anaemia, hepatosplenomegaly, pancytopenia, and hypergammaglobulinaemia. Failure to treat VL results in death.

Cutaneous Leishmaniasis (CL), representing 50-75 % of new cases (WHO, 2001), is the most common form of Leishmaniasis causing skin lesions at the site of the phlebotomine sandfly bite. Whilst the lesions usually self heal, the patient can be left with disfiguring scars. The causative agents are *Leishmania* species found in both the Old World and New World.

Diffuse Cutaneous leishmaniasis (DCL) is a chronic and progressive form of the disease characterised by disseminated skin lesions which resemble those of lepromatous leprosy. Unlike CL, the lesions do not heal spontaneously, and are not easily treated. The causative agents are *L. aethiopica* in the Old World, or the *L. mexicana* species complex in the New World.

Mucocutaneous leishmaniasis (ML), also known as espundia, is a metastatic complication of CL, primarily caused by *Leishmania braziliensis* involving the infection, and subsequent destruction of the naso-oropharyngeal mucosal membranes. Manifestation of the disease may arise concurrently with the cutaneous lesions or several years later. Early symptoms include chronic nasal problems. This form of the disease is extremely disfiguring, and many untreated patients die as a consequence of secondary bacterial infections of the open wounds.

The global distribution of leishmaniasis is shown in Figure 1-1. Currently, the disease is endemic in 88 tropical and sub-tropical countries with twelve million people infected, and a further 200 million at risk (Singh, VP *et al*. 2012). However, it is likely that there is considerable under-reporting of the disease with an estimated 500,000 new cases of VL, and 1.5 million new cases of CL annually. Over 90 % of VL cases are found in Bangladesh, India, Nepal, Sudan, and Brazil; whereas 90 % of CL cases are found in Afghanistan, Algeria, Iran, Iraq, Saudi Arabia, Syria, Brazil, and Peru. *Leishmania mexicana*, the organism used in this study, is responsible for CL in Mexico and Argentina.

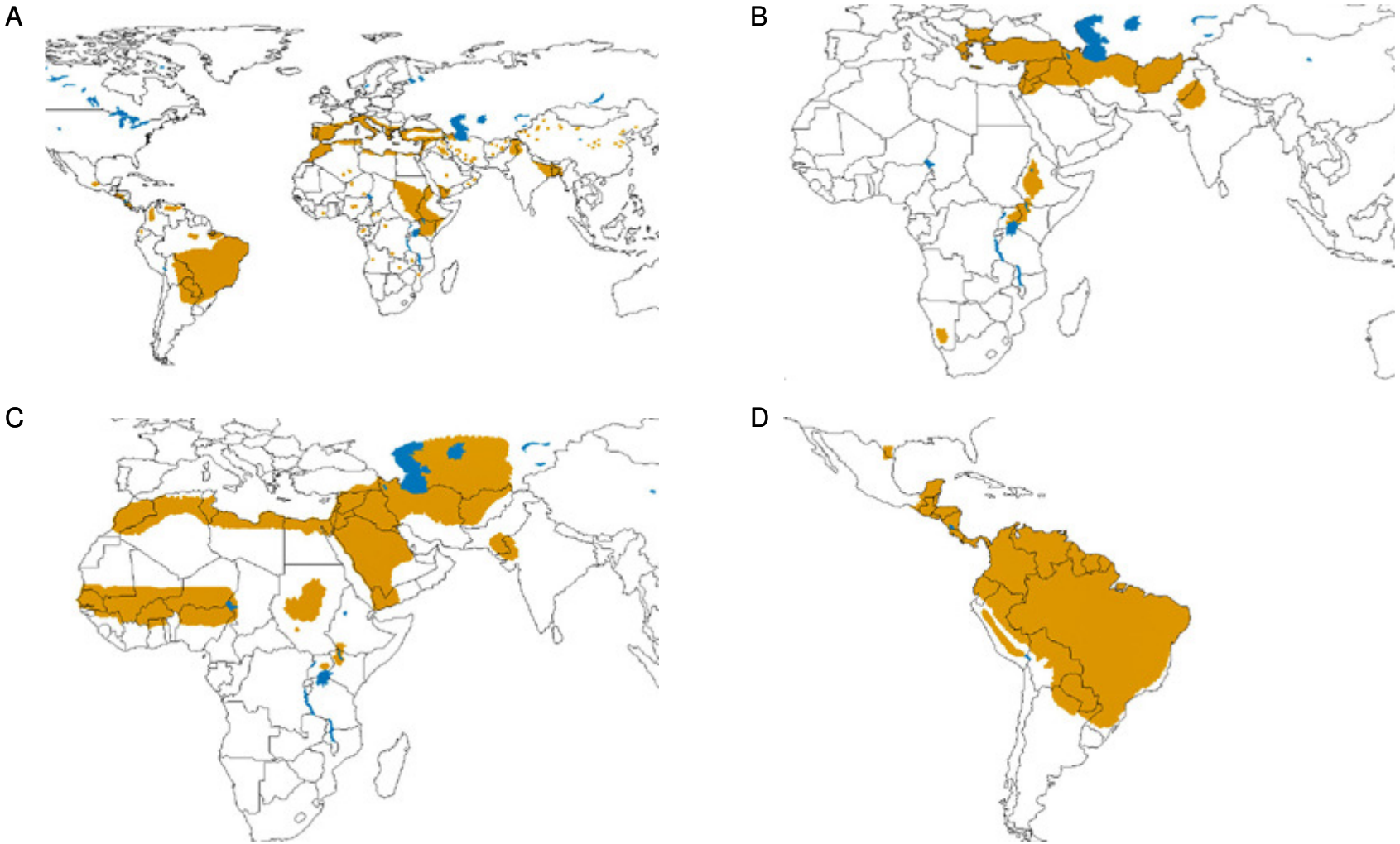


Figure 1-1. Global distribution of Leishmaniasis in 2010.

Figure 1-1, continued from Page 17. A – The geographical distribution of visceral leishmaniasis in the Old and New world. B – The geographical distribution of Old World cutaneous leishmaniasis caused by *L. major*. C – The geographical distribution of Old World cutaneous leishmaniasis caused by *L. tropica* and related species, and *L. aethiopica*. D – Geographical distribution of cutaneous and mucocutaneous leishmaniasis in the New World. WHO[1]

1.3 Leishmania Life Cycle

The life cycle of *Leishmania* involves an invertebrate host, the female phlebotomine sandfly, and a vertebrate host, often a warm blooded mammal (Figure 1-2). In the midgut of the sandfly *Leishmania* exist as motile dividing promastigotes, and as they migrate to the foregut they differentiate into metacyclic promastigotes, the mammalian infective form.

When the sandfly has a bloodmeal, these metacyclic promastigotes are introduced into the bloodstream, and within 24 hours they are phagocytosed by neutrophils (Ribeiro-Gomes, FL and Sacks, D 2010). The *Leishmania* promastigotes inhibit apoptosis of the PMN cells, and within two days, macrophages are recruited to the site of infection, where they phagocytose the PMNs (Aga, E *et al*. 2002). At this point, promastigotes differentiate into non-motile amastigotes where they undergo proliferation at the acidic pH of the parasitophorus vacuole (PV). The size of the PV depends on the infecting species, and thus harbours single or multiple amastigotes. For example *L. donovani* produce small vacuoles which contain few amastigotes, whereas *L. mexciana* produce larger vacuoles containing many amastigotes (Rittig, MG and Bogdan, C 2000; Handman, E and Bullen, DV 2002). When the amastigotes are released from the macrophage, they infect other macrophages.

The life cycle is complete when the sandfly takes a bloodmeal from an infected mammal; amastigotes are released from the macrophages, and differentiate into promastigotes. There are four main stages of promastigote development as the parasites move from the abdominal midgut through the thoracic midgut towards the foregut and stromodeal valve. These stages are as follows; procyclic promastigotes, nectomonad promastigotes, leptomonad promastigotes, and metacyclic promastigotes (Gossage, SM *et al*. 2003). Depending on the species of *Leishmania*, the process takes between six and nine days. Accumulation of the

[1] http://www.who.int/entity/leishmaniasis/leishmaniasis_maps/en

infective metacyclic promastigotes at the stromdeal valve allows transmission to occur when the sandfly takes a bloodmeal (Rogers, ME *et al*. 2002). A combination of high promastigote infection and the presence of the promastigote secretory gel (PSG) plug reduce the efficiency of blood uptake during a bloodmeal. Consequently, the sandfly must make more bites, thus increasing the chance of transmission (Rogers, ME *et al*. 2004).

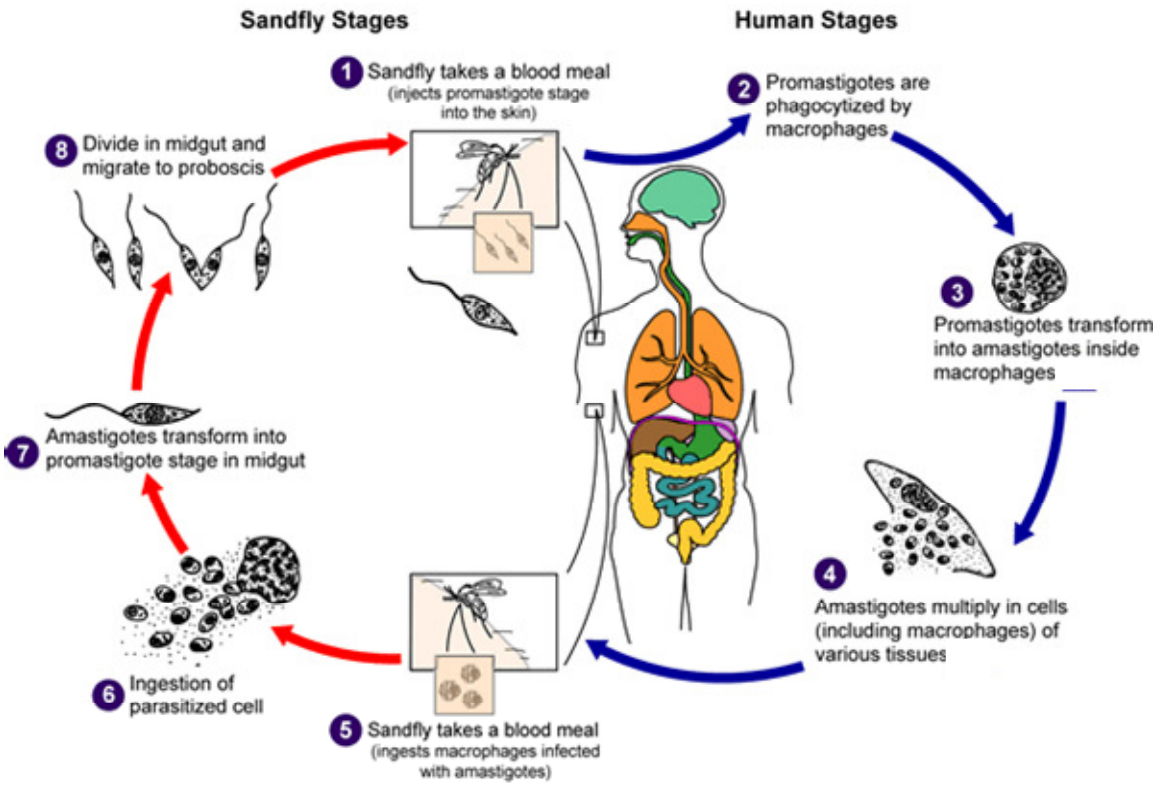


Figure 1-2. The life cycle of *Leishmania*.

Female phlebotomine sandflies inject infective promastigotes from their proboscis during a blood meal. The immune system of the mammalian host responds, primarily through macrophage phagocytosis, although other phagocytic cells are involved. Residing in the macrophage, the promastigotes transform into infective amastigotes, which are capable of infecting other phagocytic cells. The clinical manifestation of the disease depends on the infecting species. If an infected macrophage is ingested during a blood meal, amastigotes released from the macrophage will develop into promastigotes in the midgut, which then migrate to the proboscis. CDC[2].

[2] http://www.cdc.gov/parasites/leishmaniasis/biology.html

1.4 African Trypanosomiasis

The prevalence of African trypanosomiasis, a disease affecting mammals, is strongly correlated with the distribution of the tsetse fly, *Glossina spp*, which acts as the insect vector, and is restricted to sub-Saharan Africa. This discovery was made by David Bruce in 1895 when he discovered that trypanosomes were the causative agent of Nagana, trypanosomiasis of livestock (Connor, RJ 1994).

Human African trypanosomiasis (HAT) is caused by *Trypanosoma brucei gambiense* and *Trypanosoma brucei rhodesiense*. It is estimated that HAT threatens 60 million people in 36 sub-Saharan African countries, with around 300-500 thousand people infected at the end of the twentieth century; although recent data indicates that numbers are falling (Barrett, MP 2006). In addition to the species that cause HAT, a number of trypanosomes species are also infectious to domestic livestock including *Trypanosoma brucei brucei*, *Trypanosoma congolense*, and *Trypanosoma vivax*. The economic consequences and impact on food security caused by these parasites is very serious running into billions of dollars (Connor, RJ 1994).

As previously mentioned there are two clinically significant species which infect humans; consequently HAT occurs in two forms. Chronic HAT, caused by *T. b. gambiense*, generally located in the tropical regions of Central and West Africa. Whilst the disease is inevitably fatal, patients can survive with chronic HAT for many years before symptoms are apparent. However, acute HAT is caused by *T. b. rhodesiense*, and is restricted to East and Southern Africa. Disease progression is rapid in acute HAT, with death occurring in a matter of weeks or months. The distribution of these two species is depicted in Figure 1-3.

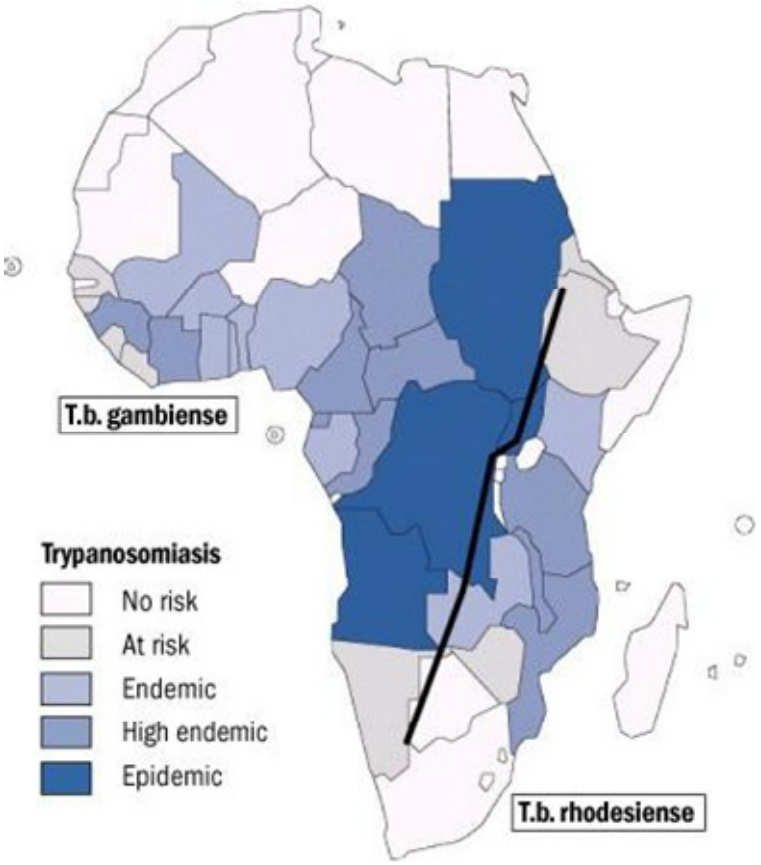


Figure 1-3. Distribution of Human African Trypanosomiasis in 1999.

Countries at high risk (endemic) of trypanosomiasis are shown in dark blue, whereas countries at no risk are shown in white. The black line represents the divide between *T. b. gambiense* and *T. b. rhodesiense* as the causative species. WHO[3]

HAT has two stages. In stage 1, or the haemolyphatic phase, the trypanosomes are present in the bloodstream. In stage 2, or the neurological phase, the parasites invade the central nervous system, with profound neurological consequences. Symptoms in the early stage of both chronic and acute HAT are non-specific, including headaches and general malaise. Therefore, an early diagnosis is imperative as progression to the late stage is inevitably fatal in the absence of chemotherapy (Barrett MP *et al* 2003). As the disease progresses to the late stage, disruption of the circadian rhythm is evident with many patients suffering from sleep disorders. This observation gave rise to the colloquial name 'sleeping sickness'.

[3] http://www.who.int/csr/resources/publications/CSR_ISR_2000_1tryps/en/index.html

1.5 Trypanosome Life Cycle

An infected tsetse fly bites a mammalian host, introducing metacyclic form trypanosomes into the blood stream. The metacyclic form is pre-adapted for survival in the mammalian host in that it already expresses the dense glycoprotein coat, which can be systematically altered through the process of antigenic variation, to evade the mammalian host's immune system. This is known as the variant surface glycoprotein (VSG) coat.

Metacylic trypanosomes transform into the long slender bloodstream form, which proliferate in the blood and lymph (Pays, E *et al*. 2006). As these parasites reach their maximum density, they are thought to secrete Stumpy Induction Factor (SIF) which, when accumulated, promotes differentiation into the short stumpy form, which do not divide as the cell cycle is arrested (Vassella, E *et al*. 1997). Physiological changes pre-adapt these short stumpy cells for life within the tsetse fly midgut. For example, the mitochondrion is more developed to allow the metabolism of amino acids (Jensen, BC *et al*. 2009). This differentiation mechanism also limits parasitaemia (Tyler, KM *et al*. 2001).

Transmission to the tsetse fly occurs during a blood meal. The change in physiological conditions, such as temperature and pH, and also the presence of cisaconitate (CCA; C00417) or citrate (C00158) triggers the rapid differentiation from the short stumpy form to the procyclic form (Szoor, B *et al*. 2010). It is at this point that the VSG coat is lost, and replaced with an invariant GPEET or EP procyclin coat (Butikofer, P *et al*. 1997; Vassella, E *et al*. 2004). The procyclic cells proliferate in the tsetse midgut; undergo differentiation to the non-dividing proventricular mesocyclic form, a process which takes three to four weeks (Aksoy, S *et al*. 2003). The proventricular mesocyclic forms differentiate to the epimastigote form whilst migrating to the salivary glands. The epimastigote attaches itself to the microvilli of the gland; this proliferative form is believed to undergo sexual genetic change (Jenni, L *et al*. 1986; Peacock, L *et al*. 2011). Further differentiation steps occur, culminating in the mature metacyclic which is the only stage infective to mammals. These metacyclics are adapted for life in the mammal, expressing a VSG coat, and are transferred during a blood meal. A detailed overview of all trypanosome differentiation stages is provided in (Vickerman, K 1985).

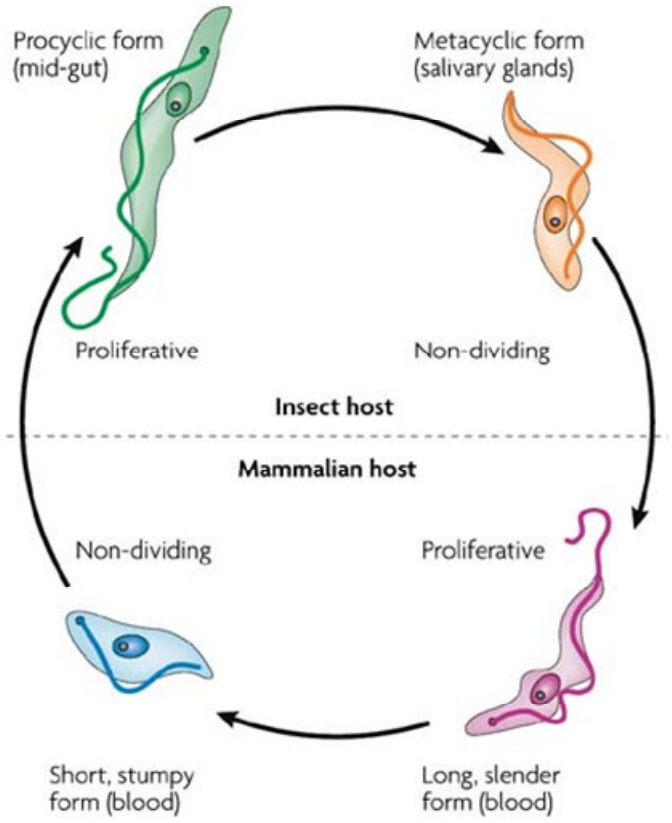


Figure 1-4. The life cycle of *Trypanosoma brucei*

The life cycle of *T. brucei* alternates between the insect host and a mammalian host. Procyclic form cells, residing in the tsetse fly midgut, transform to a metacyclic form, pre-programmed for life in the mammalian bloodstream. These cells then migrate to the salivary gland where they are passed to the mammalian host. The proliferative long slender blood stream form is capable of evading the immune system through changes in the VSG. Differentiation to the non-diving short stumpy bloodstream form adapts the cells for the mid-gut of the tsetse fly. Pays, E *et al.* 2006. Permission to reproduce this image has been granted by Nature Publishing Group.

1.6 Kinetoplastid morphology and ultrastructure

The Kinetoplastida are defined by the presence of a kinetoplast, an intercatenated network of circular mitochondrial DNA molecules located within the single mitochondrion. *Leishmania* and trypanosomes contain many organelles found in higher eukaryotic cells, although also contain a number of specialised organelles, such as glycosome and acidicalcisome. A schematic representation of the principle structures of trypanosomatids is depicted in Figure 1-5.

The kinetoplast can be distinguished as a clearly distinct DNA containing disc-shaped structure, separate from the nucleus, using fluorescent DNA binding compounds such as DAPI. This mitochondrial DNA, known as kinetoplast DNA

(kDNA), accounts for 10-20 % of the total cell DNA, and is arranged in 30-50 maxi-circles and 5,000-10,000 mini-circles (Stuart, K 1983; Liu, B *et al*. 2005). Like higher eukaryotes, the maxi-circles encode genes for rRNA and subunits of the respiratory complex (Simpson, L and Shaw, J 1989; Shaw, J *et al*. 2003). The mini-circles maintain the organised kDNA structure, and also encode guide RNAs, which facilitate the RNA editing of the maxi-circle transcripts through insertion of uridylate residues at specific locations (Liu, B *et al*. 2005). The kinetoplast is located at the posterior end of the single mitochondrion.

Kinetoplastid organisms possess a single mitochondrion, with the degree of development dependent on the life cycle stage. In trypanosomes, the procyclic form derives energy from oxidative phosphorylation, and as a consequence in the mitochondrion is well developed with many cristae. In contrast, the bloodstream form has a poorly developed mitochondrion, as energy is derived from substrate level phosphorylation, which occurs in the glycosomal and cytosolic compartments (Michels, PA *et al*. 2000; Coustou, V *et al*. 2006).

Glycosomes are small spherical organelles, closely related to perixosomes. In the bloodstream form of *T. brucei* the glycosome contains the first seven enzymes of the glycolytic pathway, converting D-glucose (C00009) to 1,3 bisphosphoglycerate (C00236) or 3-phosphoglycerate (C00197) (Coustou, V *et al*. 2005). Compartmentalisation of these enzymes increases the concentration of these enzymes, and consequently, flux through the glycolytic pathway (Michels, PA *et al*. 2000). The remaining steps occur in the cytosol, where ATP for cellular processes is generated. In the bloodstream form trypanosome over 90% of the total protein content of the glycosome is glycolytic enzymes (Aman, RA *et al*. 1985; Misset, O *et al*. 1986), compared to 40-50% in procyclic trypanosomes (Hart, DT *et al*. 1984).

Acidicalcisomes, spherical organelles containing electron dense granules, located at the periphery of the cell. They are believed to be storage compartments for several cations, dications, and phosphate, important for homeostatic regulation (Docampo, R and Moreno, SN 1999; Rodrigues, CO *et al*. 1999).

Kinetoplastid organisms possess a single flagellum; the most obvious function of this organelle is for motility (Santrich, C *et al*. 1997). However, the flagellum also facilitates the attachment of *Leishmania* promastigotes to the epithelium of the sandfly gut (Killick-Kendrick, R *et al*. 1974; Killick-Kendrick, R *et al*. 1974), and is thought to act as an environmental sensor (Bastin, P *et al*. 2000). The *Leishmania* flagellum begins at the basal body, physically linked to the mitochondrion, located at the posterior of the cell, and emerges from the flagellar pocket. In promastigotes the flagellum is not attached to the cell body, whereas in the amastigote stage, the flagellum is much shorter and not visible externally. Amastigotes are non-motile. However, in trypanosomes, the flagellum arises from the basal body at the anterior of the cell, extending along the entire length of the cell body, attached by an undulating membrane.

The surface of *Leishmania* is covered with a glycocalyx consisting of glycolipids, which varies in thickness depending on the life cycle stage. The promastigote has a relatively thick glycocalyx allowing survival in the sand fly (Turco, SJ *et al*. 2001). In contrast, the amastigote has a thin glycocalyx, as the intracellular environment of the macrophage provides the parasite with protection from the immune system. Multiple phosphoglycans constitute the glycocalyx, however the predominant surface molecule is the lipophosphoglycan (LPG), a hyperglycosylated form of glycoinositol-phospholipid (GPI). Many of these phosphoglycans (PGs) are essential for survival in the sandfly (Sacks, D and Sher, A 2002). Furthermore, a lack of PGs results in a reduced ability of promastigotes to infect macrophages, although there is no effect on amastigote infections (Capul, AA *et al*. 2007). As previously mentioned, the amastigote glycocalyx is thinner, consisting of free GPIs and host derived glycophospholipids (McConville, MJ *et al*. 2002).

Bloodstream form trypanosomes express Variant Surface Glycoproteins (VSGs), which form a dense surface coat that covers the entire cell body (Englund, PT *et al*. 1982). These proteins are synthesised at the Golgi apparatus, transported through the cytosol, and are excreted at the flagellar pocket (Vickerman, K *et al*. 1993). The VSG functions to protect surface proteins and receptors from the host immune system (Vickerman, K and Luckins, AG 1969). The host immune system will produce antibodies to recognise the VSG, thus reducing the level of parasitaemia. However, the process of antigenic variation allows the

trypanosome to express an alternative VSG (Barry, JD *et al*. 1979), allowing some parasites to evade the immune system and proliferate (Sternberg, JM 1998), thus increasing parasitaemia. In procyclic trypanosomes, the VSG coat is replaced with and invariant coat of EP or GPEET procyclins (Roditi, I *et al*. 1989; Butikofer, P *et al*. 1997; Roditi, I *et al*. 1998). Silencing hexokinase, phosphofructokinase, and the hexose transporter resulted in a switch from EP to GPEET, indicating a link between glucose availability and expression of procyclin (Morris, JC *et al*. 2002; Vassella, E *et al*. 2004). Whilst EP is expressed throughout the procyclic life cycle, GPEET is down-regulated in late procyclics (Vassella, E *et al*. 2000; Acosta-Serrano, A *et al*. 2001), as the parasite prepares for the amino acid to glycolytic metabolic shift.

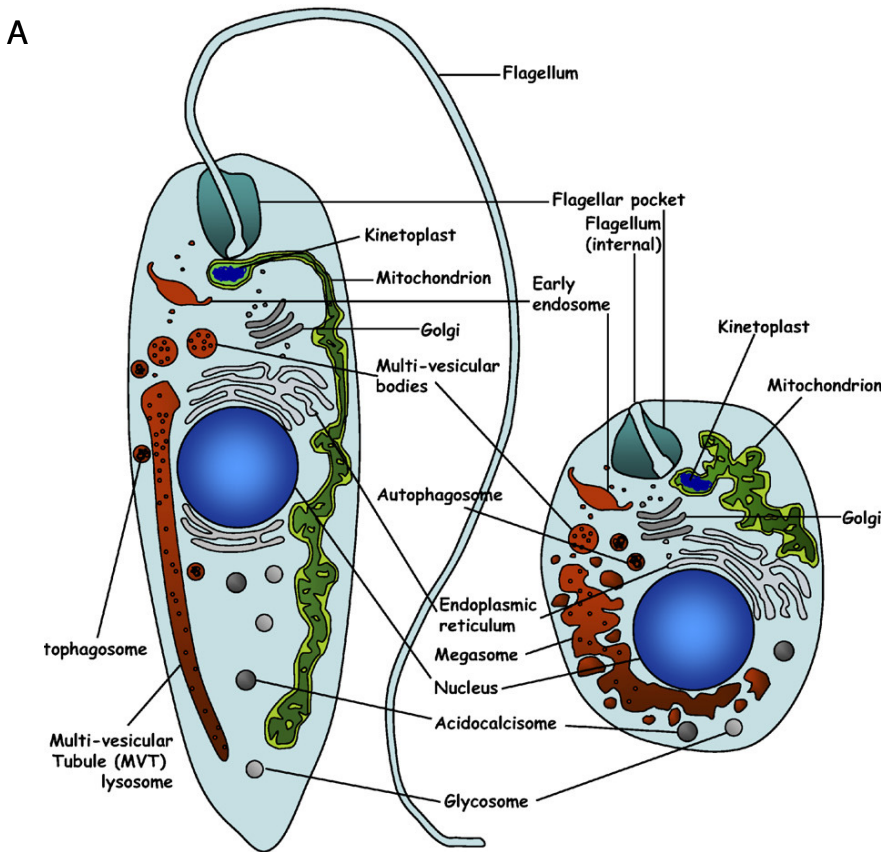


Figure 1-5. Schematic representation of the principle structures of trypanosomatids.

A – The long thin *Leishmania* promastigote possesses a visible flagellum, arising from the flagellar pocket at the anterior end of the cell, is shown on the left. The round amastigote form with an internal flagellum is shown on the right. Besteiro, S *et al.* 2007. Permission to reproduce this image has been granted by Elsevier. B – The *T. brucei* trypomastigotes Vickerman, K *et al.* 1993

1.7 Diagnosis

1.7.1 Visceral Leishmania

VL is a fatal condition, with untreated patients acting as a reservoir for the parasites, thus contributing to the spread of the disease. Consequently, an early diagnosis of VL is essential (Chappuis, F *et al*. 2007). As the drugs used to treat VL are toxic, an accurate and sensitive diagnosis is necessary. The classical confirmatory test of VL is the use of microscopy to identify the presence of amastigotes in spleen, bone marrow, or lymph node aspirates, although a high level of technical expertise and care are required. Alternatively, the presence of *Leishmania* DNA can be detected with a diagnostic PCR test (Schonian, G *et al*. 2003; Reithinger, R and Dujardin, JC 2007), although this can prove impractical in the field.

Other diagnostic approaches, which vary in sensitivity and specificity, include the identification of pancytopenia, the formol-gel test (FGT), the indirect fluorescence antibody test (IFAT), the direct agglutination test (DAT), and the rK39 dipstick test (Boelaert, M *et al*. 2004). The DAT and rK39 dipstick test were specifically developed for use in the field (Chappuis, F *et al*. 2007).

1.7.2 HAT

An accurate diagnosis of HAT is essential for treatment and control of the disease (Chappuis, F *et al*. 2005). There are two methods in which HAT may be suspected; patient history or population screening. Local knowledge of endemic diseases, pathology, and epidemiology are extremely useful. Patient history can be difficult to interpret as the clinical presentations of early stage HAT are variable, can be non-specific, and thus are often confused with other tropical diseases and ailments. These symptoms include headache, relapsing fever, general malaise, anaemia, and cutaneous lesions (Kennedy, PG 2006). Alternatively, populations can be screened for HAT. The Card Agglutination Test for Trypanosomiasis (CATT), developed at the Institute of Tropical Medicine in Antwerp, is the preferred option for cheap, reliable diagnosis in the field, although is specific for *T. b. gambiense* infections (Chappuis, F *et al*. 2005).

If HAT is suspected, either through the patient history or screening, then it is necessary to detect trypanosomes in the lymphatic system. This can prove

difficult as parasitaemia will fluctuate in relation to VSG expression, thus there may be very few parasites present in any sample. Consequently, multiple samples taken at different time points are necessary. Parasitaemia is higher in *T. b. rhodesiense* infections when compared to *T. b. gambiense* infections, which can make the diagnosis procedure easier. Given the toxicity and stage specific nature of the drugs currently available, determining the stage of the disease is critical for the patient to receive the appropriate treatment. The identification of live trypanosomes in the CSF after an invasive lumbar puncture is the definitive diagnosis of late stage trypanosomiasis (Enanga, B *et al.* 2002). However, the WHO also recognises a CSF protein content of greater than 370 mg/l or a CSF white blood cell count greater than 5 cells/µl as defining parametes of late stage trypanosomiasis (WHO); although neither of these are specific to trypanosomiasis. The methods of stage determination are identical for both *T. b. gambiense* and *T. b. rhodesiense* infections.

1.8 Vaccine prospects

Trypanosomatids introduced to the mammalian host through the bite of an infected tsetse fly or sandfly are immediately exposed to the host immune system. The surface proteins present on these parasites provide an abundance of epitopes which could potentially be exploited in the development of a vaccine.

The discovery of the VSG coat provided a focus for vaccine research in trypanosomes (Magez, S *et al.* 2009). Unfortunately, the process of antigenic variation (Van der Ploeg, LH *et al.* 1992), combined with a larger diversity of possible expressed VSGs than initially thought (McCulloch, R and Horn, D 2009), hindered research using VSG epitopes for the development of a vaccine.

It has been documented that patients with CL maintain residual parasites, however are also resistant to re-infection (Handman, E 2001). This obervation has not only stimulated interest in the development of a vaccine against Leishmaniasis, but also indicated that vaccination is realistic option. Similar to the use of the cowpox virus or vaccinia virus to prevent smallpox, the deliberate and controlled *Leishmania* infection on unexposed areas of the skin to prevent lesions of visible parts of the skin were previously used in some parts of the Middle East and former USSR (Handman, E 1997). In recent years, attention has focused on killed *Leishmania* vaccines (Handman, E 1997; Chappuis, F *et al.*

2007), genetically manipulated parasites for the creation of live attenuated vaccines (Alexander, J *et al*. 1998; Uzonna, JE *et al*. 2004), purified *Leishmania* proteins or peptides (Handman, E 2001), recombinant DNA (Handman, E 2001; Kedzierski, L *et al*. 2006), and the Leish-111f multi-epitope subunit vaccine (Skeiky, YA *et al*. 2002; Coler, RN and Reed, SG 2005; Coler, RN *et al*. 2007).

The development of suitable vaccines would greatly enhance the ability to control both trypanosomiasis and Leishmaniasis, with research into a potential *Leishmania* vaccine at a more advanced stage than any trypanosome vaccine.

1.9 Vector Control

Several vector control schemes have been developed over the years with a view to decreasing the tsetse fly population. Edward Knipling, an American entomologist, first proposed the idea of controlling or eradicating a species by altering its reproduction (Knipling, EF 1946). Initiated in 1994, a program was conducted by the International Atomic Energy Agency, which targeted tsetse flies on the island of Unguja, Zanzibar, 35km from mainland Africa. The sterile insect technique (SIT) was used successfully to eradicate tsetse flies over a period of four years, during which 8.5 million sterile males were released into the environment (Vreysen, MJ *et al*. 2000; Vreysen, MJ 2001). Whilst this program was expensive, however the area is now free from trypanosomiasis. Yet, great care must be taken to ensure that tsetse flies are not re-introduced.

Whilst the expensive SIT program was effective, there are cheaper alternatives. These include aerial and ground spraying of insecticides, and the deployment of tsetse traps. These traps are use black or blue cloth, colours which are visually attractive to the fly. Chemicals such as acetone, which mimics cows' breath, can be used to further increase the chance of attracting the flies (Barrett, MP *et al*. 2003). Treating the traps with insecticides, or applying insecticides to animals, in an attempt to kill the flies are alternative options of reducing transmission. Visceral Leishmaniasis almost disappeared from the Indian subcontinent in the 1950's following large scale antimalarial insecticide campaigns (sandflies are also susceptible to DDT), although the disease re-emerged when these campaigns were stopped (Chappuis, F *et al*. 2007). The use of insecticide treated bednets (ITNs) can prevent VL and other vector borne diseases such as malaria (Ritmeijer, K *et al*. 2007). The identification of other host animals is

also necessary; for example, dogs are the main reservoir of *L. infantum* in zoonotic VL (Chappuis, F *et al*. 2007). In Iran, the use of deltramethrine treated collars reduced the risk of infection (Gavgani, AS *et al*. 2002).

There are several factors which make vector control programs difficult to achieve; notably poor infrastructure and a lack of funding. However, a co-ordinated effort from multiple governments and organisations is required, although can be difficult to achieve given the political instability in these regions.

1.10 Chemotherapy

Chemotherapy remains the most effective method of controlling trypanosomiasis and leishmaniasis, as the transmission of the parasite is interrupted. There are a limited number of licensed drugs available to treat HAT, Chagas' disease (caused by American trypanosomes) and leishmaniasis (

Table 1-1). Unfortunately, many of these drugs tend to be disease, or stage specific, thus further narrowing the available treatments. Furthermore, the available drugs are expensive, difficult to administer, and often induce detrimental side effects. Most drugs are administered by multiple intramuscular injections, although oral therapy for leishmaniasis is now available (Croft, SL *et al*. 2006).

The general lack of progress in drug development is disappointing to say the least. Three out of the five licensed drugs for treating trypanosomiasis have been in use for more than 50 years, and only one has been introduced in the last 20 years. However, there have been encouraging advances in the treatment of leishmaniasis with introduction of miltefosine, although this is still far from ideal. Recently, there has been a focus on reducing the toxicity of existing compounds. Clinical interventions aim to reduce mortality from VL, and morbidity from CL.

Surprisingly, drug resistance in the field is a widespread problem, despite the length of time drugs have been used for. This would indicate that these drugs have multiple targets, and multiple routes of entry.

Table 1-1. Drugs currently used to treat leishmaniasis and trypanosomiasis

Drug Information	Species and disease	Dosage	Route	References
Amphotericin B AmBisome®, Fujisawa Healthcare and Gilead Sciences, US [CID 5280965]	*Leishmania* (VL, ML)	5 mg/kg, single dose	i.v. injection	Khoo, SH *et al.* 1994, Sundar, S *et al.* 2003
Antimony gluconate Solustibosan® BCM Corporation, India [CID 16683012]	*Leishmania* (VL)	15-20 mg/kg/day for 21-28 days	i.m. injection i.v. injection	Croft SL, & Yardley, V 2002
Eflornithine Ornidyl® Sanofi-Aventis, France [CID 3009]	*T. b. gambiense* (late stage)	100 mg/kg, 4 times per daily, for 7 - 14 days	i.v. injection	Burri, C & Brun, R 2003
Meglumine antimoniate Glucantime®, Sanofi-Aventis, France [CID 64953]	*Leishmania* (VL, CL)	15-20 mg/kg/day for 21-28 days	i.m. injection i.v. injection	Croft SL, & Yardley, V 2002, Croft, SL *et al.* 2006,
Melarsoprol Arsobal® Sanofi-Aventis, France [CID 10311]	*T. b. gambiense* (late stage) *T. b. rhodesiense* (late stage)	2.2 mg/kg/day for 10 days	i.v. injection	Fairlamb, AH 2003
Miltefosine Impavido® and Miltex® Zentaris GmbH, Germany [CID 3599]	*Leishmania* (VL, CL)	100 mg/kg/day for 28 days	orally	Sundar, S *et al.* 1998, Jha, TK *et al.* 1999
Nifurtimox Lampit® Bayer, Germany [CID 6842999]	*T. b. gambiense* (late stage)	10 mg/kg, 3-4 doses per day for 60-120 days	orally	Wilkinson, SR *et al.* 2011
Paromomycin sulfate Humatin® [CID 165580]	*Leishmania* (VL, CL)	11 mg/kg/day for 21 days	i.m. injection (VL) topical (CL)	Croft, SL & Yardley, V 2002, Sundar, S *et al.* 2009
Pentamidine isethionate Pentacarinat® Sanofi-Aventis, France [CID 4735]	*T. b. gambiense* (early stage) *Leishmania* (VL)	4 mg/kg, daily or alternate days (7-10 doses) 2-4 mg/kg on alternate days for 30 days	i.m. injection i.m. injection	Sands, M *et al.* 1985, Sundar, S & Rai, M 2002
Sodium stibogluconate Pentostam® GlaxoSmithKline, UK [CID 16683881]	*Leishmania* (VL, CL)	15-20 mg/kg/day for 21-28 days	i.m. injection i.v. injection	Berman, JD *et al.* 1997, Croft, SL *et al.* 2006
Suramin Germanin® Bayer, Germany [CID 5361]	*T. b. gambiense* (early stage) *T. b. rhodesiense* (early stage)	20 mg/kg/day on days 3, 10, 17, 24, and 31	i.v. injection	Jannin, J & Cattand, P 2004

i.v. = intravenous
i.m. = intramuscular
CID = PubChem Compound Identifier

1.10.1 Pentamidine

Pentamidine isethionate, an aromatic diamidine, is used to treat the early stages of *T. b. gambiense infection*, and is used as a second line drug against antimony resistant *Leishmania* (Sundar, S and Rai, M 2002; Bray, PG *et al*. 2003).

Treatment failures are rare for trypanosomiasis; the likely causes being misdiagnosis of late stage HAT, for which pentamidine is known to be ineffective. The treatment involves a dose of 4 mg/kg for 7-10 days by intramuscular injection (Sands, M *et al*. 1985). For the treatment of visceral leishmaniasis in India, treatment consists of 2-4 mg/kg on alternate days for 30 days by intramuscular injection (Sundar, S and Rai, M 2002). However, since the 1980s, cure rates have dropped from 100 % to 70 %, despite an increase in dosage. This has caused problems with toxicity, cost, and availability of treatment.

A detailed overview of pentamidine is provided in Chapter 5.

1.10.2 Melarsoprol

Melarosoprol, an organic arsenical, was introduced in 1949 and can be used to treat the late stage infection of both *T. b. gambiense* and *T. b. rhodesiense*. The lipophilic properties allow the drug to cross the blood brain barrier into the CNS. The major drawback of melarosoprol is the potential risk of a reactive encephalopathy that occurs in around 5-10 % of patients who are treated; approximately half of these are fatal (Pepin, J and Milord, F 1994). For this reason, along with increasing incidence of treatment failures with melarsoprol (Brun, R *et al*. 2001), eflornithine has recently emerged the drug of choice for *T. b. gambiense* infections.

Currently, melarsoprol treatment involves the intravenous administration of 2.2 mg/kg for 10 days. This superseded the previous treatment regime which involved increasing the melarsoprol dosage from 1.2 – 3.6 mg/kg for 4 days, followed by 7 - 10 days no treatment (Fairlamb, AH 2003). This is repeated a further two or three times. Unfortunately, it has been reported that the differences in the treatment schedules do not alter the incidence of reactive encephalopathy (Burri, C *et al*. 2000).

The mode of action for melarsoprol has yet to be determined. In the patient, melarsoprol is rapidly converted to melarsen oxide, the active compound, which in turn in selectively taken up by the parasite by the P2 transporter (Carter, NS and Fairlamb, AH 1993; de Koning, HP and Jarvis, SM 1999). Several mutations, leading to six amino acid changes, were identified in the P2 transporter isolated from an arsenical resistant *T. brucei* cell line, which consequently led to an inability to transport adenosine (Maser, P *et al*. 1999). Analysis of field isolates from the north west of Uganda, derived from patients who had relapsed after melarsoprol treatment, indicated that 38 out of 65 patients had the same six amino acid changes (Matovu, E *et al*. 2003). Whilst mutations in the P2 transporter suggest that this transporter is responsible for a decrease in the accumulation of arsenical compounds, resistance to melarsen oxide in the P2 knockout cell line is less pronounced (Matovu, E *et al*. 2001). This indicates that an additional transporter carries the drug across the membrane, with the HAPT a likely candidate (Bridges, DJ *et al*. 2007).

1.10.3 Eflornithine

Eflornithine, or difluoromethylornithine (DFMO), is a suicide inhibitor of ornithine decarboxylase (ODC; E.C. 4.1.1.17), therefore inhibiting polyamine biosynthesis. Originally developed as an anti-cancer drug, it was introduced in 1990 to treat both early and late stages of *T. b. gambiense* infections, primarily due to the high toxicity of melarsoprol. Unfortunately the drug is ineffective against Rhodesian sleeping sickness (Iten, M *et al*. 1995).

Eflornithine has relatively few side effects in comparison to melarsoprol, but can include seizures, myelosuppression, and gastrointestinal problems (Priotto, G *et al*. 2006). Treatment involves intravenous doses of 100 mg/kg at six hour intervals, for 14 days (Burri, C and Brun, R 2003). When given orally, eflornithine has poor bioavailability, and the high EC_{50} determines that large doses are required (Vincent, IM *et al*. 2012) which is a logistical problem in rural communities. In recent years, nifurtimox has been given in combination with eflornithine for the treatment of HAT (Priotto, G *et al*. 2009). These drugs are thought to act in synergy, thus permitting the administration of lower doses of both drugs, thus reducing any adverse side-effects and costs.

Eflornithine has similar affinity for both *T.b. gambiense* and mammalian ODC; however the latter is replenished at a faster rate, thus giving a selective advantage (Iten, M *et al.* 1997). Irreversible inhibition of ODC occurs when eflornithine binds at the active site (Poulin, R *et al.* 1992). This inhibition leads to changes in the polyamine levels. Ornithine (C00077), S-adenosyl methionine (C00019) and decarboxylated S-adenosyl methionine (C01137) all increase, whereas putrescine (C00134), spermidine (C00315), and trypanothione (C02090) all decrease (Bacchi, CJ *et al.* 1983; Fairlamb, AH *et al.* 1987; Xiao, Y *et al.* 2009). Polyamines are crucial for a number of cellular processes that include the production of trypanothione, which protects the trypanosome from oxidative stress (Krauth-Siegel, RL *et al.* 2003), and G-S phase transition in the cell cycle (Li, F *et al.* 1996). Eflornithine is a trypanostatic compound, thus preventing the long slender blood stream forms from proliferating. Instead, they are forced to differentiate to the non-proliferative stumpy forms which are auxotrophic for polyamines (Barrett, MP *et al.* 2007). The immune system is then able to kill the parasites. It was suggested that a transport route for eflornithine uptake was likely (Phillips, MA and Wang, CC 1987). This was confirmed when TbAAT6 (Tb927.8.5450), an amino acid transporter, was lost in eflornithine resistant cells, indicating that this transporter is capable of transporting eflornithine (Vincent, IM *et al.* 2010).

1.10.4 Suramin

Suramin, introduced in 1922, remains the drug of choice for early stage *T. b. rhodesiense* infections as the drug does not cross the blood brain barrier. Treatment with suramin involves an initial test dose of 4-5 mg/kg, followed by doses of 20 mg/kg (maximum 1 g) on days 3, 10, 17, 24, and 31 (Jannin, J and Cattand, P 2004). Treatment failures are not uncommon; however, similar to pentamidine treatment failure, the likely cause is due to the misdiagnosis of late stage disease. At therapeutic levels of suramin, approximately 1.1 mM, it is estimated that 85% of the total drug concentration is bound to plasma proteins, 15% of which is low density lipoprotein (LDL). It has been reported that suramin bound to LDL enters trypanosomes via receptor mediated endocytosis (Vansterkenburg, EL *et al.* 1993), however there is evidence that in procylic form trypanosomes uptake is via LDL independent receptor mediated endocytosis (Pal, A *et al.* 2002).

Whilst the mode of action remains unclear, suramin is a large polyanion, and consequently has inhibitory effects on a large number of enzymes including; 6-phosphogluconate dehydrogenase (E.C. 1.1.1.49) (Barrett, MP *et al*. 2007); dihydrofolate reductase (E.C. 1.5.1.3); fumarate hydratase (E.C. 4.2.1.2); glycerol-3-phosphate dehydrogenase (E.C. 1.1.99.5); hexokinase (E.C. 2.7.1.1); thymidine kinase (E.C. 2.7.1.21) (Delespaux, V and de Koning, HP 2007). The large number of targets provides a possible reason as to why no resistance to suramin has emerged for almost a century (Fairlamb, AH 2003), however there have been reports of resistance in field isolates of *T. b. rhodesiense* (Bacchi, CJ *et al*. 1990) and *T. evansi* (El Rayah, IE *et al*. 1999). A recent article revealed that suramin resistance can be selected by knockdown of expression of a series of genes associated with an endocytic pathway related to binding of ISG75 (Alsford, S *et al*. 2012).

1.10.5 Nifurtimox

Nifurtimox has been used to treat Chagas disease, American trypanosomiasis, since the 1960's (Hall, BS *et al*. 2011). The drug is not registered to treat HAT, however in cases where eflornithine and melarsoprol have been unsuccessful in the treatment of late stage infections; it may be given on compassionate grounds (Bouteille, B *et al*. 2003).

Nifurtimox is a pro-drug, which requires reduction of the nitro group to activate the drug. In trypanosomes, this process is enzymatic through NADH-dependent type 1 nitroreductases. The precise mode of action is unknown, however it is hypothesised that reduced nifurtimox induces the production of hydroxylamine derivative, which in turn generates cytotoxic metabolites and free radicals, which are capable of interacting with numerous biomolecules (Fairlamb, AH 2003; Wilkinson, SR and Kelly, JM 2009; Wilkinson, SR *et al*. 2011).

Treatment regimes require doses of 10 mg/kg in three or four doses per day, over a 60- to 120-day period. Side effects include CNS toxicity, seizures, muscles weakness (Wilkinson, SR *et al*. 2011).

1.10.6 Antimony

Trivalent antimonial compounds were first used to treat leishmaniasis in 1913, and pentavalent antimonials were introduced in the 1920s, leading to synthesis

of several drugs which are currently used today (Croft, SL and Yardley, V 2002). Antimony gluconate (Solustibosan®) was introduced in 1937, with sodium stibogluconate (Pentostam®) and meglumine antimoniate (Glucantime®) in the 1940s (Croft, SL and Yardley, V 2002).

The exact treatment regime with antimonial compounds differs depending on the infecting species and the health of the patient, although it is likely to be in the region of 15-20 mg Sb^V/kg per day. Treatment usually lasts for a period of 21-28 days (Berman, JD 1997; Croft, SL and Yardley, V 2002), allowing the drug to accumulate to levels which are toxic to the parasite. However, this long treatment course can result in side effects including hepatitis, pancreatitis and cardiac arrhythmia.

The use of pentavalent antimonial compounds in the Indian sub-continent is no longer recommended are due to widespread resistance (Sundar, S *et al*. 2000); however these drugs are used elsewhere in the world (Croft, SL *et al*. 2005).

The cellular targets of antimonial compounds remain unknown; however inhibition of ADP phosphorylation, DNA I topoisomerase, trypanothione, and beta-oxidation of fatty acids have all been proposed as a mode of action (Croft, SL *et al*. 2005).

1.10.7 Paromomycin

Paromomycin, an aminoglycoside antibiotic, isolated from *Streptomyces* spp was introduced in the early 1960s and shown to exhibit anti-leishmanial (Kellina, OI 1963) anti-protozoan activity (Croft, SL and Yardley, V 2002). Renewed interest in the early 1980s led to the development of various topical formulations which were shown to be effective in several clinical studies (Croft, SL and Yardley, V 2002). There were however several problems with this approach. First, *Leishmania* amastigotes reside deep in the dermal layer and are subsequently not exposed to therapeutic concentrations. Second, amastigotes for some *Leishmania* spp are not confined to the site of the lesion due to dissemination throughout the lymphatic system (Croft, SL and Yardley, V 2002).

Consequently, intramuscular formulations were investigated for the treatment of VL, and phase III clinical trials were completed in India in 2005. The current

treatment with paromomycin is a single daily intramuscular injection of 11 mg/kg/day for 21 days (Sundar, S *et al*. 2009). Despite being the cheapest anti-leishmanial compound available, a shorter treatment regime or combination therapy would be advantageous.

Aminoglycosides are known to bind to the ribosome small subunit in bacteria, thus inhibiting protein synthesis, although the mode of action for *Leishmania* has not been fully determined. Dissociation of ribosomal sub-units and effects on mitochondrial function has been suggested as a possible mechanism (Croft, SL and Yardley, V 2002).

1.10.8 Amphotericin B

Ampotericin B, isolated from *Streptomyces nodosus*, was shown to have antileishmanial activity in the 1960s (Yardley, V and Croft, SL 2000). Amphotericin B binds to sterols present in the plasma membrane leading to the formation of aqueous pores. Ergosterol is present in plasma membrane of *Leishmania*, whereas the mammalian plasma membrane contains cholesterol. Amphotericin B has a higher affinity for ergosterol than cholesterol, and therefore the mammalian plasma membrane is less susceptible to the drug (Croft, SL and Yardley, V 2002). At concentrations greater than 0.1 μM, the influx of cations and anions promote osmotic changes resulting in cell lysis, whereas a collapse in the membrane potential due to cation influx occurs at concentrations below 0.1 μM (Ramos, H *et al*. 1996).

Despite the selective advantage of targeting the *Leishmania* plasma membrane, the drug exhibits delayed side effects, thus reducing the attractiveness of this drug and was consequently regarded as an alternative treatment for MCL and VL (Khoo, SH *et al*. 1994). However, subsequent problems in treating MCL, and the emergence of resistance to antimonial compounds for VL, led to renewed interest in the drug and a desire to improve the formulation.

These improvements include using lipid formulations for drug delivery, thus reducing the toxicity to enable higher concentrations of active compound to be used (Croft, SL *et al*. 2005). For example, AmBiosome is registered for the treatment of VL, where a single does of 5 mg/kg has been reported to cure 90 % of patients (Sundar, S *et al*. 2003). There have also been reports of curing cases

of VL where the patient no longer responds to antimonial compounds, with Amphotericin B now used as first line treatment in India (Chappuis, F *et al*. 2007).

1.10.9 Miltefosine

Miltefosine, an alkylphosphocholine, is effective against both CL and VL (Sundar, S *et al*. 2002; Soto, J *et al*. 2004). Anti-leishmanial activity was first reported in 1987 against *L. donovani* promastigotes (Achterberg, V and Gercken, G 1987) and *L. donovani* amastigotes in BALB/c mice models (Croft, SL *et al*. 1987). Miltefosine has been registered for treatment in India since 2002 (Davies, CR *et al*. 2003).

Treatment involves a dose of 100 mg/kg/day for 28 days, and has been reported to cure 95 % of patients with mild-moderate VL during clinical trials in the Bihar region (Sundar, S *et al*. 1998; Jha, TK *et al*. 1999).

Importantly, miltefosine has several advantages when compared to other antileishmanial compounds. These include, the availability of the drug as an oral formulation, and that miltefosine has been used to successfully treat cases of antimonial resistance VL (Sundar, S *et al*. 1999). Additionally, side effects are minimal, although due to teratogenic properties it should not be administered to pregnant women (Sundar, S *et al*. 2002). The mode of action remains unknown, although lipid metabolism, cell signalling, and the induction of apoptosis have been suggested (Croft, SL and Yardley, V 2002).

1.10.10 Drug summary

In summary, the drugs currently used in treatment of HAT and leishmaniasis all suffer from similar problems. Many are toxic; most have to be given parenterally, and for long treatment durations. Costs can be high and distribution difficult to the areas in which they are given. Only recently is insight into modes of action becoming clear. The recent emergence of pipelines of new compounds, particularly through the Drugs for Negleced Diseases initiative (DNDi), does offer hope,although as yet the pipelines are not sufficiently robust to offer optimism that the recently announced WHO goals of HAT elimination and leishmaniasis control by 2020 will be met.

1.11 Metabolism

The protozoan parasite *Trypanosoma brucei* must adapt to different host organisms throughout its complex lifecycle. The contrasting environments of these host organisms present the parasite with many challenges that include host defence mechanisms and energy metabolism.

1.11.1 Glucose metabolism

Glycolysis is the conversion of one mole of D-glucose (C00031) to two moles pyruvate (C00022), with the net production of 2 moles of ATP (C00002). Energy metabolism in the bloodstream form of the parasite is simple; the glycolytic pathway is the sole mechanism of deriving ATP. The mammalian infective form of the parasite resides in the bloodstream, an environment abundant in D-glucose. Glucose enters the parasite through a high capacity facilitative diffusion hexose transporter (Tb10.6k15.2040) (Bakker, BM *et al*. 1999).

In bloodstream form *T. brucei*, the first seven steps of the gycolytic pathway occur in the glycosome, producing two moles of 3-phosphoglycerate (C00197) per mole of D-glucose. The remaining three steps occur in the cytosol; producing pyruvate, which is not metabolised further, and consequently excreted by the parasite.

The glycosomal ATP/ADP balance is maintained as the two moles of ATP consumed by hexokinase (EC 2.7.1.1) and phosphofructokinase (EC 2.7.1.11) are replaced by two moles of ATP produced via the phosphoglycerate kinase (EC 2.7.2.3) reaction. The net ATP gain occurs in the cytosol through the conversion of phosphoenolpyruvate (PEP; C00074) to pyruvate (C00022) by pyruvate kinase (EC 2.7.1.40).

Similarly, the glycosomal redox balance is maintained as the NADH produced by glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) is re-oxidised by glycerol-3-phosphate dehydrogenase (EC 1.1.1.8). Consequently, glycerol 3-phosphate (G3P; C00093) is transported to the mitochondrion and converted to dihyroxyacetone phosphate (DHAP; C00111) via a FAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) and alternative oxidase; with oxygen acting as the final acceptor (Pollakis, G *et al*. 1995). The DHAP produced in the mitochondrion is transported back to the glycosome.

In the absence of oxygen the G3P/DHAP mitochondrial shuttle does not operate. However, bloodstream form trypanosomes can survive by producing ATP through the conversion of glycerol 3-phosphate (G3P; C00093) to glycerol (C00116) by glycerol kinase (EC 2.7.1.30). Under anoxic conditions the reduction in NADH levels reduces the flux of glyceraldehyde 3-phosphate through the glycolytic pathway. Whilst D-glucose consumption remains constant, the production of pyruvate and glycerol is equimolar. Consequently, cytosolic ATP production is reduced (Opperdoes, FR 1995).

Glycolytic activity is ten times higher in the bloodstream form trypanosomes than in typical cells of the mammalian host (Cazzulo, JJ 1992). This observation can be explained as the net gain of ATP in the trypanosome is only two moles per mole of glucose, and thus the parasite must rely on a high glycolytic activity to meet energy demands.

Bloodstream form trypanosomes have reduced mitochondrial metabolism, reflected in the underdeveloped mitochondrian representative of this stage of the parasite. In the procyclic form trypanosomes the situation is more complex; glucose is generally less abundant in the tsetse fly (with the exception of immediately after a blood meal).

The procyclic form trypanosomes have a fully developed mitochondrion with a functional electron transport chain capable of generating ATP via the F_0/F_1-ATP synthase. However, data suggest that when D-glucose is present the electron transport chain is not essential in procylic form trypanosomes and ATP is derived via substrate level phosphorylation (Coustou, V *et al*. 2006).

Phosphoglycerate kinase (EC 2.7.2.3) is cytosolic, and thus 3-phosphoglycerate (C00197) is produced in the cytosol rather than in the glycosome. This would lead to an imbalance in the glycosomal ATP-ADP balance; however, procylic trypanosomes express several additional enzymes to address this.

Succinate is the major excreted end product, and is produced in both the mitochondrion and the glycosome by the succinate fermentation pathway (Coustou, V *et al*. 2008; Ebikeme, C *et al*. 2010).

Approximately 50% of the cytosolic PEP re-enters the glycosome (Coustou, V *et al*. 2006), where it is converted to oxaloacetate (C00036) by phosphoenolpyruvate carboxykinase (PEPCK; EC 4.1.1.49), or pyruvate by pyruvate phosphate dikinase (PPDK; EC 2.7.9.1); both reactions produce ATP. Oxaloacetate is subsequently converted to glycosomal succinate (C00042) by malate dehydrogenase (EC 1.1.1.37), fumarate hydratase (FH; EC4.2.1.2), and fumarate reductase (FRD; EC 1.3.1.6). However, there is conflicting evidence as to whether fumarate (C00122) is produced in the glycosome. Genome predictions for *T. brucei* indicated the existence of two class I fumarate hydratase genes encoding mitochondrial (FHm) and cytosolic (FHc) isoforms of FH. Both of these enzymes account for total cellular FH activity. Sub-cellular localisation experiments and proteomic analysis indicated that neither FH enzyme localised to the glycosome (Colasante, C *et al*. 2006; Coustou, V *et al*. 2006). However, there is a cryptic PTS1 glycosomal targeting motif that may allow dual localisation of the enzyme depending on environmental conditions (Coustou, V *et al*. 2006); this may explain why glycosomal FH activity was previously detected (Besteiro, S *et al*. 2002). The glycosomal redox balance is maintained by the re-oxidisation of NADH by malate dehydrogenase and fumarate reductase.

An alternative fate for glucose is conversion to glucose 6-phosphate (G6P), which then enters the pentose phosphate pathway (PPP). The main functions of the PPP in trypanosomatids is the regeneration of NADPH and the production of ribose 5-phosphate (R5P) for nucleotide biosynthesis (Barrett, MP 1997). A detailed overview of the PPP is provided in Chapter 4.

1.11.2 Amino-Acid Metabolism

In addition to utilising glucose as an energy source, procyclic trypanosomes and *Leishmania* promastigotes can metabolise L-proline (C00148) (Balogun, RA 1974), although under standard experimental conditions used by many laboratories, procyclic trypanosomes principally prefer glucose (Lamour, N *et al.* 2005).

Proline metabolism occurs solely in the mitochondrion and is degraded to three end products; succinate (C00042), carbon dioxide (CO2, C00011), and acetate (C00033) (van Weelden, SW *et al.* 2005). Succinate and CO2 are end products produced by the TCA cycle, and are subsequently secreted by the cell (Coustou, V *et al.* 2003). However, succinate can also be converted to malate (C00149) in a two step reaction by succinate dehydrogenase (EC 1.3.99.1) and mitochondrial fumarase (EC 4.2.1.2). It has been proposed that malate is then exported and converted to PEP by malate dehydrogenase (EC 1.1.1.37) and phosphoenol pyruvate carboxykinase (PEPCK, EC 4.1.1.49). PEP would then be converted to acetate as normal (van Weelden, SW *et al.* 2005).

Energy, in the form of ATP, is produced by the conversion of succinyl-CoA (C00091) to succinate, and also as a result of the electron transport chain. The TCA cycle consumes NAD^+, although this is replenished by the first stage of the electron transport chain (Figure 1-7). The cycle is directly linked to the electron transport chain complex II by succinate dehydrogenase.

Both proline and glutamine (C00064) are converted to glutamate (C00025) before being further metabolised to 2-ketoglutarate (C00026) (Bringaud, F *et al.* 2006). Threonine (C00188) is converted into amino oxobutyrate (AOB, C03508) by L-threonine dehydrogenase (TDH; EC 1.1.1.103), which is then converted into acetyl coenzyme A (Acetyl CoA, C00024) by coenzyme A (EC 2.3.1.29) (Lamour, N *et al.* 2005). As previously mentioned, acetyl CoA is used as a co-factor in the conversion of oxaloacetate (KEGG) to citrate by citrate synthase (EC 4.1.3.7), or is converted to acetate by acetate:succinate CoA-transferase (ASCT, EC 2.8.3.5) (Coustou, V *et al.* 2008).

1.11.3 Metabolic differences between trypanosomatids

As trypanosomes and *Leishmania* originated from a common ancestor, it is not surprising that the genome sequencing projects of *L. major*, *T. brucei*, and *T. cruzi* revealed metabolic similarities at the genomic level. However, there are also notable differences related to central carbon metabolism (Opperdoes, FR and Coombs, GH 2007). Many of these differences can be attributed to the environmental conditions in which these parasites reside. The parasitophorous vacuole within mononuclear cells of the mammalian host provides an acidic environment, in which macromolecules are degraded to simple metabolites such as amino acids, lipids, and monosaccharides (Burchmore, RJ and Barrett, MP 2001).

Research has tended to focus on promastigotes, which are easier to work with than the amastigotes. The promastigotes are grown in a nutrient rich medium, for example SDM-79 or HOMEM, which may differ substantially from the nutrients available in the sandfly midgut. However, there are broad similarities with procylic trypanosome metabolism, as amino acids and glucose can be used as energy sources (Lamour, N *et al.* 2005; Bringaud, F *et al.* 2006). Sandflies are able to feed on nectar, which is comprised of a numerous sugars. Consequently, and in contrast to trypanosomes, *Leishmania* express enzymes, such as amylase or a sucrase-like protein, to metabolise these substrates (Blum, JJ and Opperdoes, FR 1994; Opperdoes, FR and Coombs, GH 2007). Genes to encode ribulokinase and xylulokinase are present in *Leishmania* genome; additionally, these enzymes have PTS targeting signal, indicating these enzymes localise to the glycosome.

In *Leishmania*, methionine is fully oxidised, feeding into the TCA cycle, whereas in *Trypanosoma*, the end point is 2-ketobutyrate (C00109) due to the absence of the downstream enzymes. There are also differences in threonine metabolism. In *T. brucei*, threonine is converted to AOB and acetyl CoA (section 1.11.2), whereas in *Leishmania*, threonine is converted to glycine by the THF-dependent pathway, or alternatively, via 2-ketobutyrate (C00109) to succinyl CoA (C00091) (Opperdoes, FR and Coombs, GH 2007). Ascorbate is required for the ascorbate dependent peroxidise (*L. major* and *T. cruzi*) and the iron-ascorbate peroxidise (*T. brucei*), suggesting a role in the defence against oxidative stress. Neither ascorbate nor erythroascorbate have been detected in *Leishmania*, and the

source of D-arabinose, the precursor for the pathway, is not known (Opperdoes, FR and Coombs, GH 2007).

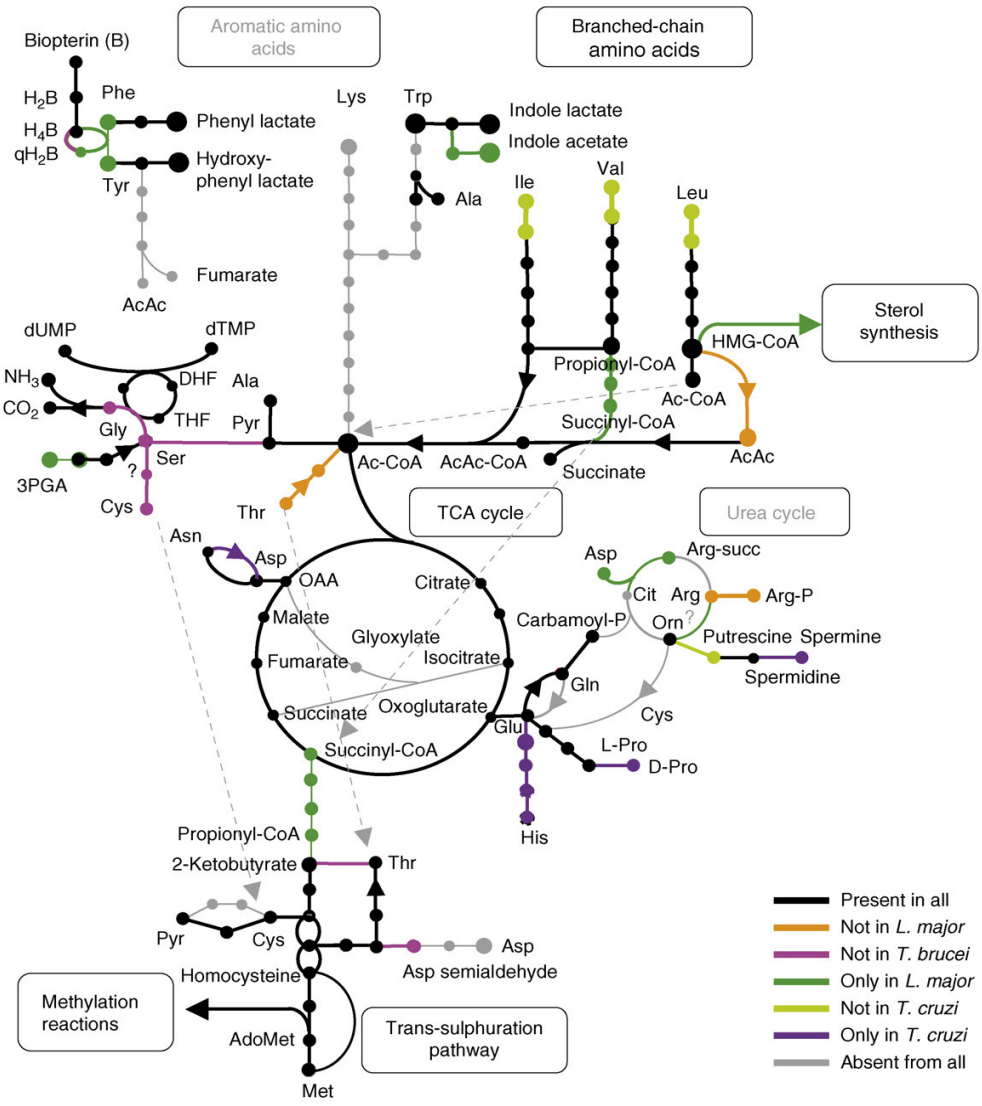


Figure 1-6. Differences in Trypanosomatid amino acid metabolism.

The amino acid metabolism in *L. major* compared with *T. brucei* and *T. cruzi*. Common biochemical reactions are shown in black, whereas reactions not present in trypanosomatids, but present in other eukaryotes are shown in grey. The representation of other differences is depicted in the key. Enzyme-catalysed reactions, for which no unambiguous gene identification could be made, are represented by a question mark. Abbreviations are as follows: AcAc, acetoacetate; AdoMet, adenosylmethionine; B, biopterin; Cit, citrulline; DHF, dihydrofolate; HMGCoA, hydroxymethylglutaryl CoA; OAA, oxaloacetic acid; Orn, ornithine; PGA, phosphoglyceric acid; qH2B, quinoid form of dihydrobiopterin; THF, tetrahydrofolate. Taken from Opperdoes, FR and Coombs, GH 2007. Permission to reproduce this image has been granted by Elsevier.

A fully functional urea cycle is not thought to occur in *L. major*, *T. brucei*, or *T. cruzi*. Arginine, ornithine and urea can only be produced by *L. major*, although

both *L. major* and *T. brucei* express ODC (Opperdoes, FR and Coombs, GH 2007). Amino Acid metabolism is depicted in Figure 1-6. Other metabolic differences include folate metabolism, haem synthesis, and amino acid synthesis (Opperdoes, FR and Coombs, GH 2007).

1.11.4 Electron transport chain

The electron transport chain consists of membrane protein complexes and redox reactions to transfer electrons from an electron donor, for example NADH, to an electron acceptor, for example molecular oxygen (O_2) (Figure 1-7). These reactions occur in sequence, with the electron donor passing electrons to an acceptor of greater electro negativity. Electrons are shuttled between the immobile macromolecular protein complexes by the lipid soluble quinones, or the water soluble cytochromes. An electrochemical proton gradient, generated by the translocation of protons (derived from the Krebs cycle) from the mitochondrial matrix to the intermembrane space, allows ATP synthase to convert mechanical energy to chemical energy by utilising the proton flow back to the mitochondrial matrix to generate ATP (Boyer, PD 1997).

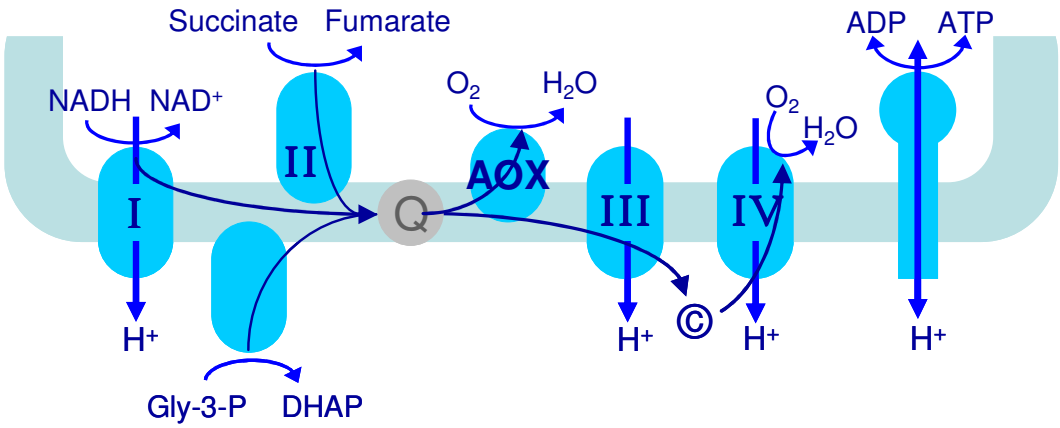


Figure 1-7. Electron Transport Chain

Complex I, NADH dehydrogenase, EC 1.6.5.3; complex II, succinate dehydrogenase, EC 1.3.5.1; complex III, cytochrome bc1 complex, EC 1.10.2.2; complex IV, cytochrome c oxidase EC 1.9.3.1; Gly-3-P, glycerol 3-phosphate; DHAP, dihydroxyacetone phosphate; AOX, alternative oxidase; ©, cytochrome *c*; Q, ubiquinone/ubiquinol.

A major product of the Krebs cycle is NADH, which can be oxidised by Complex I (NADH dehydrogenase; EC 1.6.5.3) to generate NAD^+ (which is then consumed by

enzymatic processes in the mitochondrion). During this process two electrons are removed from NADH, transferred to ubiquinone (Q), forming ubiquinol (QH2), and four protons are translocated from the mitochondrial matrix across the membrane to the intermembrane space, producing a proton gradient. Alternatively, electrons can be transferred, via FAD, to Q either by Complex II (succinate dehydrogenase; EC 1.3.5.1), through the conversion of succinate to fumarate, or by the conversion of Gly3P to DHAP. In Complex III (cytochrome bc1 complex; EC 1.10.2.2), two electrons are removed from QH2, and transferred to two molecules of cytochrome *c*, located in the intermembrane space. Two protons are translocated from the mitochondrial matrix to the intermembrane space, and an additional two protons from cytochrome *c* are released to the intermembrane space, contributing to the proton gradient. Four electrons provided by cytochrome *c* are passed to Complex IV (cytochrome *c* oxidase; EC 1.9.3.1), where molecular oxygen (O_2), the most electro negative acceptor, acts as the terminal acceptor for the electrons. Two molecules of water (H_2O) are produced during this reaction. The translocation of four protons across the membrane contributes to the proton gradient.

The efflux of protons from the mitochondrial matrix to the intermembrane space generates an electrochemical (proton) gradient. The F_1F_O-ATP synthase (EC 3.6.3.14), consisting of a F_O transmembrane component and a F_1 water soluble component, utilises this proton gradient with the F_O component acting as an ion channel to allow the flux of protons to flow back to the mitochondrial matrix. Indirect conformational changes of the two components allows the generation of ATP from ADP and inorganic phosphate (Boyer, PD 1997).

The alternative oxidase (AOX), present in some organisms including protists, provides a cytochrome *c* independent method for the reduction of molecular oxygen. However, as electrons are obtained from QH2, this bypasses the proton translocation of complex III and complex IV, thus reducing ATP generation. Interestingly, *T. brucei* is entirely dependent on the AOX for respiration, thus the enzyme, trypanosome alternative oxidase (TAO), is an attractive drug target (Chaudhuri, M *et al*. 2006). Ascofuranone and salicylhydroxamic acid (SHAM) inhibit TAO. No functional TAO homolog has been detected in *T. cruzi* or *L. major*, consequently neither of these organisms are sensitive to compounds which inhibit the AOX (Opperdoes, FR and Coombs, GH 2007).

1.12 Proteomics

The post-genomic era has seen the introduction of another 'omics' approach - proteomics (Mann, M *et al*. 2001; Aebersold, R and Mann, M 2003). Whilst genome sequence projects for many organisms have been completed, there is only so much one can infer from this information alone. Transcriptomics (the study of mRNA transcripts) can also yield useful information, although an incease in mRNA transcript does not necessarily mean an increase in protein level. However, for meaningful biological information, it is necessary to look beyond the genome, and investigate changes at the protein level. The proteome can be defined as the all proteins, including protein isoforms, polymorphisms and modifications, protein-protein interactions, that occur in a biological system under specific conditions (Pandey, A and Mann, M 2000). Technological advances allow for the identification and quantification of proteins from biological samples.

1.12.1 DIGE

Difference gel electrophoresis (DIGE) is a two-dimensional polyacrylamide gel electrophoresis (2-DE) technique to resolve complex mixtures of proteins from different samples on a single gel. In a 2-DE approach, proteins are separated by two properties in two dimensions; typically by charge in the first dimension using ioselectric focusing (IEF), and by molecular mass in the second dimension (O'Farrell, PH 1975). This approach has several drawbacks. Initially, a large quantity of protein is required as visualisation of the proteins is performed by staining the gel with colloidal coomassie blue (or silver stain), requiring approximately 500 µg or 50 µg respectively. Furthermore, technical variances between gels can make gel to gel comparisons difficult as many gels are required to perform a robust pairwise comparison of the protein samples. In DIGE, the protein samples are labelled with cyanine dyes prior to electrophoresis, which is more sensitive for protein detection than using a non-fluorescent post-stain (Shaw, J *et al*. 2003). One protein sample, derived from a wild type cell line or healthy tissue for example, can be labelled with either Cy3 or Cy5 fluorophore, and another protein sample, derived from a drug resistant cell line or diseased tissue, can be labelled with the other fluorophore (Unlu, M *et al*. 1997). These two protein samples can then be mixed together and run on a single gel, resulting in protein 'spot map'. The cyanine dyes have low unspecific binding

(Enzo Life Sciences website), are structurally similar, and have distinct spectral properties. Therefore, proteins originating from different samples can be efficiently labelled and visualised using different excitation and emission filters (Lilley, KS and Friedman, DB 2004). Furthermore, an internal pooled standard consisting of equal quantities of protein from the two samples can be labelled with Cy2 to enable an accurate comparison of protein quantities between replicate gels (Alban, A *et al*. 2003). DIGE overcomes the problems of poor reproducibility and technical variance as two protein samples are run on the same gel; therefore there is co-migration of protein spots. Consequently, this reduces the number of gels required in order to generate robust and reproducible data. The use of cyanine fluorophores allows increased sensitivity as proteins of low expression levels can be labelled and accurately compared using computer algorithms (Shaw, J *et al*. 2003). Whilst identification of low abundance proteins by mass spectrometry can be difficult (Unwin, RD *et al*. 2006), the recent and continual technological advances in mass spectrometry make this less of an issue.

1.12.2 Labelling proteins

The most common labelling reaction, known as DIGE minimal labelling, typically involves labelling only 1-5% of the total lysine residues of a protein (Unlu, M *et al*. 1997; Greengauz-Roberts, O *et al*. 2005; McNamara, LE *et al*. 2010). This approach, using N-Hydroxysiccinimide (NHS) ester linkage, ensures that a minority of proteins typically have one labelled lysine residue (Figure 1-8). If more than one lysine residue is labelled, multiple protein spots may appear on the gel in the form of a 'charge train', thus reducing the resolution.

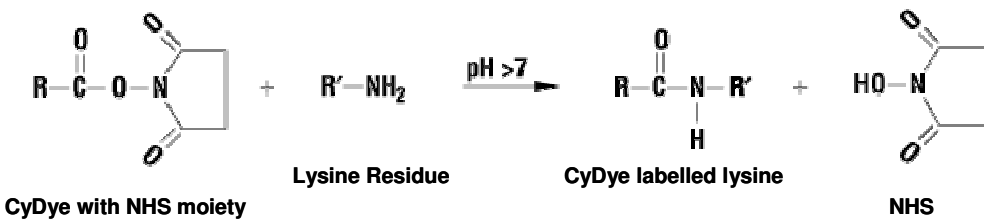


Figure 1-8. CyDye labelling reaction of lysine residues.

The cyanine fluorophores are synthesised with reactive groups, such as the highly reactive NHS moiety, to allow conjugation with proteins or nucleic acids. The reaction between the NHS ester and the primary amine forms an amide bond between the cyanine fluorophore and the side chain of

the lysine residue; NHS is released in the process. Abbreviations - R1, CyDye; R2, Lysine residue; NHS, N-Hydroxysuccinimide.

In contrast, saturation labelling involves labelling all of the cysteine residues with excess dye. Saturation cannot be achieved by labelling all lysine residues as the hydrophobicity of the cyanine dyes conjugated to the polar lysine residues causes proteins to precipitate from solution (Unlu, M *et al.* 1997; Greengauz-Roberts, O *et al.* 2005). The main advantage of saturation labelling is an increased sensitivity in comparison to minimal labelling. Comparisons with as little as 5 µg of protein can be performed, in contrast to 50 µg for minimal labelling (Shaw, J *et al.* 2003). However, there are several limitations to saturation labelling of proteins. The most significant limitation is that cysteine is less common than lysine, thus restricting the number of proteins that can be detected. Random analysis of approximately 65,000 human proteins revealed 3.7% of them did not contain a cysteine residue (Sitek, B *et al.* 2005). Whilst the majority of proteins are labelled, it is important to consider that some proteins will not be labelled. Furthermore, only Cy3 and Cy5 are available with maleimide chemistry for labelling cysteine residues, and therefore an internal standard labelled with the Cy2 fluorophore cannot be performed. Whilst increased sensitivity can be achieved by saturation labelling, allowing detection of reproducible changes in samples with low protein abundance, there can still be a problem in identifying proteins by mass spectrometry from an analytical gel. As saturation labelling is more applicable for the analysis of samples of low protein abundance, there can also be problems in identifying proteins if the protein concentration is too low. Experiments determined that loading between 1 pmol and 10 pmol of each protein was necessary to identify a protein in a gel, despite mass spectrometry sensitivity in the femtomol range. Consequently, a second preparative gel is required to allow successful protein identification, although improvements in mass spectrometry, as previously mentioned, sample preparation, and handling will improve analysis and allow direct identification of proteins from the analytical gel (Greengauz-Roberts, O *et al.* 2005).

1.12.3 Trypsin

The differentially expressed protein spots are excised from the gel, enzymatically digested with trypsin, and the resulting peptide fragments are analysed by mass spectrometry. Trypsin cleaves peptide chains at the carboxyl

side of either lysine or arginine (except when either is followed by proline). Whilst it is possible to measure the exact mass of an intact protein, it is favourable to analyse peptide fragments for a number of reasons. (1) The physico-chemical properties of some proteins can lead to problems of insolubility, however digestion of a protein into peptide fragments can overcome this, and therefore the structure of the native protein is irrelevant (Steen, H and Mann, M 2004). This is particularly important when investigating membrane proteins, which require the use of detergents for solubilisation. As detergents are in excess in relation to proteins, and are easily ionised they can interfere with mass spectrometry. Therefore digestion of a protein into peptide fragments enables easier identification of membrane proteins. (2) MS instruments are more sensitive when measuring the masses of peptides than proteins. (3) Due to protein modifications it is difficult to identify a protein by mass spectrometry based on the exact mass, and conversely, the exact mass of an intact mature protein from its amino acid sequence (Steen, H and Mann, M 2004). (4) Specificity of trypsin is well defined, hydrolysing peptide bonds where the carbonyl group is supplied by an arginine or lysine residue, unless the carbonyl group is bound to an N-terminal proline residue. This makes it easy to predict peptide fragments by an in silico trypsin digest of protein databases. The use of a non-specific enzyme generates a complex subset of overlapping peptides which in turn requires improved chromatographic resolution and greater computational processing power (Wu, CC and Yates, JR, 3rd 2003).

Using computer algorithms, it is possible to determine which proteins could give rise to the masses of specific peptides; although this approach is less reliable for complex protein samples. Consequently, tandem MS is preferred as the primary structure amino acid sequence can be obtained.

1.13 Mass Spectrometry

MS can be described as an analytical technique that accurately measures the mass, or more accurately, the mass to charge ratio (m/z) of ionised molecules (Walther, TC and Mann, M 2010), and has an important role in systems biology. In proteomics, MS provides the ability to measure the m/z ratio of specific peptide fragments (bottom-up approach), or intact mature proteins (top-down approach), although the later is more challenging. In metabolomics, MS can be used to measure the m/z ratio of ionised metabolites (section 1.15.2).

There are three main components to MS instruments. These are (i) ionisation of the molecule, or analyte, (ii) the mass analyser which allows the selection and detection of the desired ions, and finally (iii) the detection system which measures the m/z ratio.

1.13.1 Ionisation

The 'soft' ionisation technique used for biological samples ionises and vaporises a sample, without fragmenting it. With regards to a protein sample digested with trypsin, only the masses of the intact peptides are measured. The sample, in the liquid or solid phase, is ionised to the gas phase and becomes a charged molecule. The two methods of ionisation for proteomic approaches are Matrix-Assisted Laser Desorption/Ionization (MALDI) and Electrospray Surface Ionization (ESI).

In MALDI, the peptide sample, or analyte, is mixed with excess matrix solution, consisting of crystallised molecules in both aqueous and organic solvents to allow both hydrophilic and hydrophobic molecules to dissolve. The solution is applied to the MALDI plate to allow the solvents and water to evaporate, allowing the molecules to co-crystallise. A UV laser pulse is then directed at the co-crystallised sample matrix. Consequently, a portion of the sample is vaporised and ionised as protons are transferred from the matrix to the analytes (Hillenkamp, F and Karas, M 1990).

In ESI, the analyte is disolved in a volatile solvent which is then passed through a stainless steel capillary to a hypodermic needle. An electric field at the tip of the needle charges the surface of the solution as it is aersolised. The solvent evaporates until the charge density on the surface of the molecule reaches a critical point (Raleigh limit) and the droplet undergoes a 'Coulombic explosion' creating smaller droplets. This process is repeated until discrete quasi-molecular ions are produced, which are in turn suitable for mass spectrometry analysis (Fenn, JB *et al.* 1989). As molecules are prepared in the liquid phase, ESI is often combined with High Pressure Liquid Chromatography (HPLC) to separate molecules based on their polarity, prior to ionisation. Ionised analytes typically have multiple charges when generated by ESI, in contrast to single charges when generated by MALDI.

1.13.2 Mass to charge analysis

Ions are subjected to electrostatic or magnetic fields and the resulting motion is a function of the mass to charge ratio. There are numerous ways to calculate this ratio, and consequently multiple different types of mass analysers have been developed; these include time of flight (ToF), quadrupole, ion-trap, orbitrap, and Fourier-transform ion cyclotron resonance (FT-ICR). The strengths and weaknesses (such as resolution, mass accuracy, sensitivity, dynamic range) of these various mass analysers are reviewed in (Domon, B and Aebersold, R 2006). As there are a number of different methods of analysing molecules, there is the capability to combine these to utilise the advantages of multiple approaches.

1.13.3 TOF

In ToF mass spectrometry the m/z ratio is calculated by measuring the time taken for peptide ions to travel a fixed distance inside a vacuum tube. Peptides with a high m/z ratio take longer to travel down this tube than peptides of a low m/z ratio.

1.13.4 Ion-trap

The ion trap is capable of trapping ions in a quadrupolar field, thus accumulating ions within a specific range of interest. Ions, from a low m/z to a high m/z, are then ejected from the trap by changing the electrode voltages, which makes the ion trajectories unstable (Jonscher, KR and Yates, JR, 3rd 1997). The major advantage of this method is that ions are accumulated over time, thus increasing the signal to noise ratio (sensitivity) for analytes of low abundance. However, ion traps generally have poor resolution and poor mass accuracy. The introduction of linear ion traps (LITs) allows improved ion trapping,

1.13.5 Orbitrap

The Orbitrap is the most recent MS instrument, offering extremely high mass accuracy, resolution and sensitivity (Hu, Q *et al*. 2005). With the Orbitrap, the ions are injected into the instrument where they are confined by an electrostatic field. The ions oscillate around a central spindal due to the attraction towards the central electrode, counteracted by centrifugal forces. The oscillation frequency of these ions is inversely proportional to the square

root of the m/z ratio. The Orbitrap offers mass accuracy from 2-5 ppm (Hu, Q *et al*. 2005), although sub ppm accuracies can be obtained with an internal calibration (Olsen, JV *et al*. 2005). The LTQ orbitrap is the commercially available instrument, featuring ESI, a linear ion trap, and RF only quadrupole C-trap.

1.13.6 HPLC

High pressure liquid chromatography is used to separate analytes by chromatography, and is appropriate for complex samples (Jandera, P 2011). Analytes, in the mobile phase, are forced through a column (stationary phase), and a detector determines the the time taken for this to occur. This is known as the retention time, and will vary depending on the the composition and ratio of the solvents used for the mobile phase, the flow rate of the mobile phase, and the interactions between analytes and the stationary phase.

In normal phase (NP) chromatography, a non-polar, predominantly non aqueous solvent and a polar stationary phase are used to separate analytes (Jandera, P 2011). Non-polar compounds are diluted first, whereas polar compounds are separated dute to slow elution. However, polar compounds have poor solubility in these solvents, which can be problematic. Conversely, in RP chromatography, a polar solvent and non-polar stationary phase are used to separate the analytes. Consequently, polar compounds are diluted first, with the non polar compounds separated over time.

The stationary phase consists of silica peptides with specific polar or non-polar functional groups depending on the type of chromatography required. Chromatographic resolution improves as particle size decreases. The mobile phase may consist of a gradient where the compositions of solvents are altered over time, thus enabling compounds to elute sooner than normal.

1.13.7 HILIC

Hydrophilic Interaction Chromatography (HILIC) is an HPLC method used in the separate the polar / hydrophilic compounds (Alpert, AJ 1990). The liquid phase consists of water, typically between 5-40%, and a water-miscible polar organic solvent, usually acetonitrile (ACN), which is not only compatible with MS, but offers higher sensitivity. Elution is achieved by gradually increasing the quantity

of water (Naidong, W 2003; Hemstrom, P and Irgum, K 2006). The technique is similar to normal phase (NP) chromatography, and orthogonal to reverse phase (RP) chromatography as polar compounds are retained longer than non-polar compounds (Boersema, PJ *et al*. 2008).

Reverse phase chromatography is not suitable for the separation of polar compounds, as these compounds fail to bind to the column, and consequently, are eluted rapidly. A ZIC-HILIC column combined with an Orbitrap FT-MS has shown to be advantageous in separating polar metabolites from whole cell lysates (Kamleh, A *et al*. 2008). Separation of metabolites based on ploarity is achieved using acetonitrile and water. Metabolites which are less polar are eluted from the column first, whereas polar metabolites are retained. Consequently, metabolites are separated by hydrophobicity by chromatography, and then by mass using mass spectrometry.

1.13.8 MS/MS

In tandem mass spec (MS/MS) individual peptides can be targeted with sufficient energy to achieve fragmentation, and the subsequent m/z values of these ionised fragments are recorded. Fragmentation is mainly achieved through collision induced dissociation (CID) and electron capture dissociation (EID), in series with two TOF instruments. In CID the peptide will undergo multiple collisions with inert gas molecules until the ion accumulates enough vibrational energy to break the weakest chemical bonds, usually the amide bond of the peptide, generating b- (N-terminus) and y- (C-terminus) fragments. This approach is more appropriate for smaller peptides with no unstable modifications, as these bonds are weaker than the peptide bonds. In EID the positively charged peptide reacts with an electron donor generating a peptide fragment with an unpaired electron that is highly unstable. The peptide backbone splits at the Cα-C’ bond to generate c- (N terminus) and z- (C-terminus) fragments. However charge reduction of the peptides is more common than fragmentation leading to a decrease in sensitivity (Mallick, P and Kuster, B 2010). Other methods of fragmentation include electron capture dissociation (ECD) and post source decay (PSD).

1.14 The identification of proteins

The identification of proteins is performed by peptide mass fingerprinting (PMF), sequence comparisons or MS/MS ion queries using the online MASCOT web application. The MASCOT search engine incorporates code from Molecular Weight Search (MOWSE), an early program for detecting proteins by mass fingerprinting from a mass list (Pappin, DJ *et al*. 1993), although includes probability based scoring (Perkins, DN *et al*. 1999). Searching the sequence of any FASTA format file is also possible; including protein sequences directly, or the translation of nucleic acid sequences.

An in silico digest of a proteome database will yield a theoretical spectrum for each predicted peptide, which can then be compared to the experimental spectrum to gain peptide and protein identifications (Mallick, P and Kuster, B 2010). Alternatively, de novo sequencing of peptide fragments can be used to determine the amino acid sequence as the mass difference between adjacent signals often differ by that of an amino acid. If no genome sequence is available then de novo sequencing is certainly more applicable, however the extensive number of completed genome projects (Kyrpides, NC 1999) favours the in silico approach.

Careful consideration should be taken when preparing the samples so as to minimise unwanted protein modifications and proteolytic degradation. Consequently, harvested cells are lysed in a DIGE compatible lysis buffer containing chaotropic agents and CHAPS which aid protein solubility; protease inhibitors and keeping the samples on ice help prevent proteoloytic degradation. Lysis can be further aided by mechanical methods such as sonication or repeated freeze-thaw if necessary. As it is often necessary to precipitate protein samples, it is important to have enough starting material as this process leads to loss of proteins. Precipitation of protein samples is important for the concentration of proteins, and the removal of contaminants, although these can be reduced by using high quality reagents.

1.15 Metabolomics

Metabolomics can be defined as “the comprehensive (qualitative and quantitative) analysis of the complete set of all low molecular weight

metabolites present in and around growing cells at a given time during their growth or production cycle" (Mashego, MR *et al*. 2007). Low molecular weight metabolites are considered to be less than 1500 Da (Dunn, WB 2008). When we consider the other –omic disciplines in systems biology, the genome, proteome, and transcriptome are relatively static in comparison to the dynamic fluidity of the metabolome. Consequently, the study of metabolomics allows us to gain an instantaneous picture of the biological system at a particular time. This can provide a clearer insight into biological changes at the genomic, transcriptomic, or proteomic level. Furthermore, when we consider the metabolome of an organism, there are fewer metabolites than there are genes, transcripts or proteins (Kell, DB 2006). In theory, this should make the datasets more concise and easier to interperate; however, in reality, the datasets contain background ions, adducts, contaminants.

Metabolomics can be either targeted or untargeted. A targeted approach involves the quantitative identification of structurally known and biochemically annotated metabolites, in relation to known metabolic pathways. In contrast, the untargeted approach is more difficult to interpret, with some of the peaks belonging to unknown metabolites for which no structural data may be available. However, both approaches can be aided through the use of isotopic labelling, which allows the detection of any products derived from the labelled substrate.

Broadly speaking, there are three main areas of metabolomics; sample preparation, data aquistion (mass spectrometry or NMR), data analysis (the computational or bioinformatic approaches used to interperate data).

1.15.1 Sample preparation

The metabolomics approach varies depending on the biological system and the metabolites of interest. Consequently, careful experimental design is required to deal with numerous factors that include; metabolism quenching, metabolite extraction, sample storage, and the biological system being investigated.

Metabolomics provides an insight into the metabolome at a precise point in time; therefore, both the time of metabolite extraction and metabolism quenching are extremely important to reflect the current state of the biological system of interest (Faijes, M *et al*. 2007). The metabolome of the organism may

change due factors such as age, gender, diet, or disease state. Alternatively, if a cell culture system is being investigated, then change to the temperature, cell cycle stage, culture medium, or atmospheric conditions may play a role. As a consequence, it is important to standardise these conditions as much as possible, or at least consider them during analysis.

The quenching of metabolism should be rapid and performed at an extreme temperature to inhibit all possible enzymatic activity to prevent degradation of metabolites. A common method used to quench metabolism is treating the cells with an aqueous methanol solution at a temperature of -40 °C (de Koning, W and van Dam, K 1992). There have been numerous metabolite extraction protocols published, which vary depending on the organism. As the metabolome represents a wide range of compounds (sugars, amino acids, organic acids, fatty acids, nucleotides), which in turn provides a large chemical and physical variability, the solvent chosen will extract different metabolites. For example, acetonitrile:methanol:water (2:2:1) and water:isopropanol:methanol (2:2:5) were found to have sufficient coverage of the *Saccharophagus degradans* metabolome, in contrast to pure methanol or acetonitrile:water (1:1, v/v) (Shin, MH *et al*. 2010). Another study, using *Lactobacillus plantum*, found that cold methanol, boiling ethanol, and perchloric acid provided suitable coverage of the metabolome; however, both chloroform:methanol (1:1) and chloroform:water (1:1) were not as efficient (Faijes, M *et al*. 2007). The optimisation of metabolite extraction has also been investigated in *Leishmania* (Saunders, EC *et al*. 2010; t'Kindt, R *et al*. 2010), *Saccharomyces cerevisiae* (Villas-Boas, SG *et al*. 2005), and *E. coli* (Chassagnole, C *et al*. 2002).

If quenching and extraction are performed as two independent steps, then metabolite leakage during quenching must be considered (Mashego, MR *et al*. 2007), although any method used for either metabolite extraction or metabolism quenching will result in metabolite loss (Dettmer, K *et al*. 2007).

The optimal storage conditions vary depending on biological sample, and should be investigated prior to analysis to prevent, or minimize, degradation of metabolites. Blood, plasma and urine samples can be stored at -80 C for up to 6 months, whereas plant materials can be stored at -80 C for up to 4 weeks with no changes to metabolome (t'Kindt R, personal communication).

These results suggest that it will not be possible to develop a standardized protocol for metabolome quenching, metabolite extraction, and sample storage, with experimental optimisation specific to the biological system and metabolites of interest.

1.15.2 Data acquisition

Data is aquired using NMR or mass spectrometry (section 1.13). Whilst NMR is not the most sensitive technique, it is both an accurate and reproducible method of detecting many metabolites (Gupta, N *et al*. 1999). One MS approach involves using gas chromatography (GC) coupled to MS (GC-MS) to quantity metabolites from *Arabidopsis thaliana* (Fiehn, O *et al*. 2000); however the main limitation of GC-MS is organic phosphates are not detected. Consequently, attention has focussed to an alternative chromatographic technique; HILIC (section 1.13.7) coupled to an Orbitrap Mass Spectrometer (section 1.13.5) in an attempt to resolve the global trypanosomatid metabolome (Kamleh, A *et al*. 2008; t'Kindt, R *et al*. 2010; Vincent, IM *et al*. 2010). The sample preparation method influences the metabolites which are detected (section 1.15.1), although it is possible to resolve more metabolites. The Orbitrap has a high mass accuracy, which whilst in the region of 2 ppm (Hu, Q *et al*. 2005), can be reduced to 0.21 ppm using internal background ions for provide additional calibration (Scheltema, RA *et al*. 2008). The methods of data acquisition most likely outweigh the ability to interpret the data, therefore research should concentrate on optimising data analysis.

1.15.3 Data analysis

The data analysis step for the interpretation of metabolomic data is the weakest aspect of metabolomic experiments, although has improved in recent years. Data analysis can be further broken down into two distinct areas; the deconvolution of raw data files obtained from the mass spectrometer, and providing a biological context for the data.

The successful processing of raw data relies on numerous factors; statistical analsysis to obtain reproducible data ignoring background noise, the ability to handle any related peaks derived from one particular parent metabolite peak (adducts, fragments, isotopes and multiply charged species), and the ability to differentiate between isomers (using retention time). The Xcalibur software

(Thermo) is a basic tool to identify peaks isolated from the FT-MS Orbitrap; however, is unable to identify the subtle differences between peaks of low intensity. Therefore, more advanced software for metabolomic data processing has been developed by the metabolomics community, and currently includes MZMatch (Scheltema, RA *et al*. 2011), XCMS (Smith, CA *et al*. 2006; Tautenhahn, R *et al*. 2010), IDEOM (Creek, DJ *et al*. 2012). Common feature of these approaches include retention time alignment, peak detection, peak matching, and assigning an identification with information derived from online metabolite databases. It is necessary to find the correct balance between filtering the false positives, whilst identifying and retaining reprocible peaks; MZMatch attempts to identify and combine related peaks.

While the identification of metabolites is important, the graphical representation of the data is equally important. The MetaNetter plug-in for Cytoscape allowed the creation of an ab initio network using metabolomic data and a list of biological transformations (Jourdan, F *et al*. 2008). Similar to the exact mass of a metabolite, the biological transformations (for example dehydrogenation), also have an exact mass. Consequently, if two unknown exact masses differ by the exact mass of a particular biological transformation, then it is possible that these masses correspond to metabolites in the same pathway. From Metabolite to Metabolite (FMM) is a similar approach, which utilises all the enzymes in the KEGG database to reconstruct the metabolic pathways between metabolites (Chou, CH *et al*. 2009). The techniques used in these two methods attempt to identify relationships and links between metabolites that are not predicted to occur. However, to provide a biological context, the data needs to be compared to a reference metabolome, such as KEGG (Kanehisa, M and Goto, S 2000) and BioCyc (Caspi, R *et al*. 2010). The construction of these metabolic databases was made possible due to the availability of complete genome sequences which allowed the prediction of metabolites based on the presence of the genes which encode enzymes identified by classical biochemical techniques (Cottret, L *et al*. 2010). The MassTRIX webserver (Suhre, K and Schmitt-Kopplin, P 2008) and Pathos webserver (Leader, DP *et al*. 2011) present the opportunity to overlay, or compare experimental data with the KEGG pathways. Alternatively, the MetExplore web server offers the possibility to link metabolites identified in untargeted metabolomics experiments, within the

context of genome-scale reconstructed metabolic networks (Cottret, L *et al.* 2010).

1.15.4 Applications of metabolomics

There are several useful applications of metabolomics. When we consider that the mode of action of many drugs used to treat trypanosomiasis or leishmaniasis is unknown, metabolomics could be a play a pivotal role in elucidating the mode of action. A statistical comparison of metabolomic fingerprints (metabolic snapshot) derived from biological samples treated with drugs of known and unknown modes of action can provide information useful information (Yi, ZB *et al.* 2007). Alternatively, a metabolic profiling approach (identify as many metabolic changes as possible) can be used to detect pertubations (both direct and indirect) to metabolome, induced by the addition of a particular drug (Le Roch, KG *et al.* 2008). Subsequent experimental approaches would then focus on genetic manipulation of the genome to knockout or knockdown the gene which encodes the protein of interest, in an attempt to confirm the interaction between the drug and protein.

Another topical area of research is the identification of biomarkers associated with specific diseases. A biomarker can be defined as "biological molecules that represent health and disease states" (Lyons, TJ and Basu, A 2012). The identification of metabolites associated with infectious disease or non-infectious disease would allow the possibility to develop cheap, robust, specific medical tests for various conditions or diseases. One example of a biomarker is the detection of PSA (prostate specific antigen), which is used as an early indication of prostate cancer (Kuriyama, M *et al.* 1980). A metabolomics approach could be used to identify potential biomarkers.

2 Materials and methods

2.1 Cell culture

2.1.1 Leishmania

Leishmania mexicana promastigotes (MNYC/BZ/62/M379) were maintained in HOMEM medium (Invitrogen, Paisley, Scotland), supplemented with 10% (v/v) heat-inactivated foetal calf serum (Biosera, Ringmer, UK) (Appendix 7.1.1). The starting density was approximately 1 x 10^5 cells/ml, and cells were sub-passaged when the reached stationary phase, approximately 2 x 10^7 cells/ml.

Where necessary, antibiotics were added to the transgenic lines used and generated in this study at the following concentrations; Hygromycin B (Roche, Burgess Hill, UK) at 50 µg/ml, Nouseothricin (Lexsy NTC, Jena Bioscience, Germany) at 100 µg/ml, G418 (Sigma-Aldrich, Poole, UK) at 50 µg/ml, and puromycin (Calbiochem, Nottingham, UK) at 50 µg/ml.

2.1.2 Trypanosomes

Trypanosoma brucei procyclic form (PCF) strain 427 were cultivated in SDM79 medium (Brun, R and Schonenberger 1979) or SDM80 (Lamour, N *et al*. 2005) at 27°C supplemented with 10% (v/v) heat-inactivated foetal calf serum (Biosera, Ringmer, UK) (Appendix 7.1.1). For SDM80, the FCS was dialysed by ultrafiltration in a volume ratio of 10 parts 1x PBS to 1 part FCS. Dialysis was performed a total of four times.

Dialysis tubing, with a molecular mass cut off of 10 kDa, was boiled in 10 mM $NaHCO_3$, 1 mM EDTA pH 8.0 for 10 minutes. This was repeated twice, before being washed in ddH_2O, and stored in 70 % ethanol at 4°C. The dialysis tubing was washed in 1x PBS prior to use.

2.1.3 Determining cell density

Cells were mixed with an equal volume of 1% formaldehyde in 1x PBS counted using an improved Neubauer haemocytometer. Four of the large boxes were counted in triplicate and the average multiplied by 2 x 10^4 to determine the number of cells per ml. For very dense cultures, cells were diluted 1 in 5 with 1% formaldehyde in 1x PBS, and the average count multiplied by 5 x 10^4 to determine the number of cells per ml.

2.1.4 Preparation of stabilates

Leishmania promastigotes in the late log phase of growth were mixed in 1:1 (v/v) with the stabilate freezing solution (70% FCS / 30% glycerol). This mixture was transferred to 1.5 ml CryoVials in 1 ml aliquots; the samples were then stored at -80°C overnight, before transferring to liquid nitrogen for long term storage. *T. brucei* procyclic cells were prepared as stated above; except the freezing solution was 70% cell culture medium / 30% glycerol. For the revival of stabilates, aliquots were defrosted at room temperature and immediately transferred to 9 ml culture medium at the appropriate temperature for growth. A 1 ml aliquot of this mixture was transferred to 9 ml to reduce the concentration of glycerol and any toxic waste products present.

2.2 Isolation of genomic DNA

2.2.1 Phenol-Choloroform Method

Approximately 10 ml of late-log promastigotes were pelleted at 1500 g, washed in PBS, resuspended in 0.5 ml cell lysis buffer A (Appendix 7.2), and incubated overnight at 50°C. The solution was mixed with 1 volume of phenol, and incubated at room temperature for 5 minutes, before the upper layer was isolated by centrifugation at 855 g for 10 minutes. This was repeated twice using chloroform instead of phenol. The upper aqueous layer was mixed with 2 volumes of ethanol, and 0.1 volume 3M NaCl. The DNA precipitated from the solution, was carefully removed, and washed with 70% ethanol. The gDNA was pelleted, and resuspended in 100 µl dH_2O.

2.2.2 DNeasy Blood and Tissue Kit (QIAGEN)

Alternatively, genomic DNA was isolated from a late-log culture of promastigotes using the DNeasy® Blood and Tissue Kit (QIAGEN, Crawley, UK) according to the manufacturer's instructions.

2.3 Molecular Cloning Techniques

2.3.1 Polymerase Chain Reaction

PCR was used to amplify specific regions of DNA that would be used for molecular cloning, species identification, and to confirm integration in transfected cell lines. All oligonucleotides were synthesised by Eurofins MWG Operon (Ebersberg, Germany).

The standard PCR reaction used is as follows; 1x reaction buffer (Promega, Southampton, UK), 1.5 mM $MgCl_2$ (Promega, Southampton, UK), 0.3 µM of each primer, 0.3 µM dNTPs (Invitrogen, Paisley, Scotland), 1.25 U Go*Taq* (Promega, Southampton, UK), and between 1 and 5 ng of plasmid DNA, or 200 ng of gDNA.

The amplification was performed using a thermal cycler (PTC200 DNA Engine Thermal Cycler, MJ Research). An initial denaturation step at 94°C for 1 minute was followed by between 25 and 35 cycles of 94°C for 30 seconds, an annealing temperature (T_m) specific for each set of primers for 30 seconds, and an extension step of 1 minute per kb of DNA. A final extension step at 72°C for 5 minutes was followed by a holding temperature of 4°C. Failure to amplify specific products was often rectified by increasing the $MgCl_2$ concentration gradually, or adding DMSO (Riedel-de Haën, Germany) to a final concentration of 5%.

For the amplification of transketolase from gDNA, KOD polymerase (Calbiochem, Nottingham, UK) was used. This high fidelity enzyme has an extremely quick extension time of 1 kb every 20 seconds. The reaction components used is as follows; 1x reaction buffer, 2 mM MgSO4, 5% DMSO, 0.8 mM dNTP, 0.3 µM MB250, 0.3 µM MB251, 200 ng of template gDNA, 1 U KOD. The reaction volume was 50 µl. The cycling conditions were as follows; an initial denaturation step at 95°C for 1 minute was followed by between 30 cycles of 95°C for 15 seconds, an annealing temperature of 63.4°C for 15 seconds, and an extension temperature of 68°C for 40 seconds. A final extension step at 68°C for 2 minutes was followed by a holding temperature of 4°C. The addition of DMSO was necessary.

2.3.2 Cloning PCR products

The enzyme *Taq* polymerase adds a single 'A' residue at the 3' end of each strand of the PCR product. This could be sub-cloned into the linear pGEM-T Easy vector (Promega, Southampton, UK) which has a single 'T' residue at the 3' ends of each strand. This process is known as TA cloning. For DNA polymerases which do not add an 'A' residue at the 3' end strand, such as KOD polymerase (Calbiochem, Nottingham, UK), it was necessary to incubate the PCR product with *Taq* polymerase at 72°C for 15 minutes immediately after the final extension temperature of the PCR reaction. This allowed ligation into the TA cloning vector.

2.3.3 Olignucleotides

Oligonucleotides were designed using Vector NTI (Invitrogen, Paisley, Scotland) and ordered from Eurofins MWG Operon (Ebersberg, Germany) (Table 2-1).

2.3.4 DNA gel electrophoresis and gel extraction

PCR products were fractionated on a 1% (w/v) agarose gel (Invitrogen, Paisley, Scotland), in 1x TAE (Appendix 7.2.2). SybrSafe™ DNA gel stain (Invitrogen, Paisley, Scotland) was added to the agarose at a 1x concentration to allow visualisation of the DNA fragments. The desired DNA fragments were excised from an agarose gel using a scalpel, and extracted using the QIAquick® Gel Extraction Kit (Qiagen, Crawley, UK) according to the manufacturer's instructions. Alternatively, PCR products were purified using the QIAquick® PCR Purification Kit (Qiagen, Crawley, UK) according to the manufacturer's instructions.

2.3.5 Ligations

The 20 µl PCR product ligation reaction was prepared as follows; 50 ng of pGEM-T easy vector, between 3 and 8 µl of eluted PCR product, 1 U T4 ligase (Promega, Southampton, UK), and a final concentration of 1x rapid ligation Buffer (Promega, Southampton, UK). The reaction was carried out overnight at 4°C. DNA fragments from the pGEM-T easy intermediate vectors could be sub-cloned into the destination vector by restriction digest, and ligation using the T4 ligase as previously described.

Table 2-1. Oligonucleotide Sequences used in this study.

Oligonucleotide ID	Sequence	Description	Length (nt)	Restriction Site
SP 6	CATTTAGGTGACACTATAG	sequencing oligo	19	n/a
T3	AATTAACCCTCACTAAAGGG	sequencing oligo	20	n/a
T7	TAATACGACTCACTATAGGG	sequencing oligo	20	n/a
M13 uni (-21)	TGTAAAACGACGGCCAGT	sequencing oligo	18	n/a
MB0037	ATGATTGAACAAGATGGATTGC	Neomycin (S)	22	n/a
MB0038	TCAGAAGAACTCGTCAAGAAG	Neomycin (AS)	21	n/a
MB0039	ATGAAAAAGCCTGAACTCAC	Hygromycin (S)	20	n/a
MB0040	ACTCTATTCCTTTGCCCTCG	Hygromycin (AS)	20	n/a
MB0079	TAGCGTCGACTGTGCTTGTGGGTGAGGGCG	*L. mexicana* Transketolase 5'flank (S)	30	SalI
MB0080	TAGCAAGCTTGGCCGCTTCGCACCACACGA	*L. mexicana* Transketolase 5'flank (AS)	30	HindIII
MB0091	GCAGCTCAGCCGGCCCTCAC	outer primer for Transketolase 5' flank (S)	20	n/a
MB0094	CGCTGTCACGGGCACGATAG	outer primer for Transketolase 3' flank (AS)	20	n/a
MB0095	CTGGATCATTTTCCGATG	*Leishmania* ITS1 (S)	18	n/a
MB0096	TGATACCACTTATCGCACTT	*Leishmania* ITS1 (AS)	20	n/a
MB0210	ACTAGTGATGAAGATTTCGGTG	Streptothricin acetyltransferase (S)	22	n/a
MB0211	GGATCCTTAGGCGTCATCCTGT	Streptothricin acetyltransferase (AS)	22	n/a
MB0212	ATGATTACGCTTGCTGTGGT	antisense primer located in 5' DHFR to check integration	20	n/a
MB0213	TGTCTCTTGTCGGTGCTCAC	sequencing oligo for pXG (5'FR)	20	n/a
MB0217	TAGCCCCGGGTGCTCCGAAACGTGAGGAAT	*L. mexicana* Transketolase 3'flank (S)	30	SmaI
MB0218	TAGCAGATCTACTTCCTTGCCCTTCCGATA	*L. mexicana* Transketolase 3'flank (AS)	30	BglII
MB0245	GTGGGTGGAGGGTTTGAGGCCGAC	TKT 5' flanking region (use with MB212)	24	n/a
MB0250	TAGACTCGAGCTTCGCCTCTCTTCGTCGCCCT	*L. mexicana* Transketolase ORF (S)	32	XhoI
MB0251	TAGAGCGGCCGCCGCCTCTTCCGGTGTCATTC	*L. mexicana* Transketolase ORF (AS)	32	NotI
MB0333	AGCAAGGTGAGATGACAGGAGATCC	within Neomycin	25	n/a
MB0334	CTGCGTGCCATGCCAAACCT	within TKT	20	n/a
MB0337	TGTGCGCCTCAGTAGACCTTG	Transketolase OUTER (AS)	21	n/a
MB0340	ATGACCGAGTACAAGCCCACG	Puromycin (S)	21	n/a
MB0341	TCGTAGAAGGGGAGGTTGCG	Puromycin (AS)	20	n/a
MB0342	GCGGCATTTTGCCTTCCTGT	B-Lactamase (S)	20	n/a
MB0343	CCAATGCTTAATCAGTGAGGCACC	B-Lactamase (AS)	24	n/a
MB0492	CTGATCGATTTTCGACGGCT	*L. mexicana* VDAC 1 (S)	20	n/a
MB0493	TACTGATCGCCCACCTTCGT	*L. mexicana* VDAC 1 (AS)	20	n/a
MB0494	TCACCGCCACGTCTAGGAAC	*L. mexicana* VDAC 2 (S)	20	n/a
MB0495	GTGGACATCACGCACTCACG	*L. mexicana* VDAC 2 (AS)	20	n/a

Restriction sites are underlined in the sequence and the restriction enzymes are noted; n/a = not applicable; nt = nucleotides

2.3.6 Transformations

The chemically competent *E. coli* cell line JM109 (Promega, Southampton, UK) were transformed with the desired plasmid. 50 µl of cells were thawed on ice, mixed with 20 µl of ligation product (or 1 µl plasmid DNA), and left on ice for 30 minutes. The cells were then heat-shocked at 42°C for 90 seconds, before being placed immediately on ice for 2 minutes. 200 µl of SOC medium or LB medium was added and the cells were incubated at 37°C for between 30 and 60 minutes, before being plated on LB agar plates supplemented with 100 µg/ml ampicillin (Sigma-Aldrich, Poole, UK).

Single colonies were isolated from the plates using a sterile 200 µl pipette tip and grown in 5 ml of LB medium supplemented with 100 µg/ml ampicillin (Sigma-Aldrich, Poole, UK) for 16 hours in a shaking incubator. If necessary, colonies were screened by PCR, in a *Taq* DNA polymerase mediated reaction mix to determine whether the correct insert was present. PCR products were fractionated on an agarose gel as described in section 0. The following day, cells were pelleted at 4500 g for 10 minutes at 4°C, and the DNA was isolated using the QIAprep® Spin Miniprep Kit (Qiagen, Crawley, UK) according to the manufacturer's instructions. The yield and quality of the plasmid DNA was determined using the NanoDrop® ND-1000 UV-Vis Spectrophotometer.

Presence of the specific DNA fragment was determined by PCR or restriction digest, and confirmed by sequencing (Eurofins MWG Operon, Ebersberg, Germany).

2.3.7 Plasmid generation

DNA fragments were transferred between vectors by digesting with restriction endonucleases. Therefore, it was necessary to include restriction sites at the 5' end of the oligonucleotides used in the PCR of certain DNA fragments. The plasmids used in this study are outlined in Table 2-2 and discussed in (Chapter 4 and Chapter 5).

Table 2-2. Plasmid constructs used in this study

Plasmid ID	Details
pMB-G39	*L. mexicana* transketolase knockout construct conferring resistance to Hygromycin B
pMB-G40	*L. mexicana* transketolase knockout construct conferring resistance to Nouseothricin.
pMB-G49	*L. mexicana* transketolase ribosomal re-expression construct conferring resistance to Puromycin
pMB-G51	pGEM-T Easy (Promega) construct containing the *L. mexicana* transketolase open reading frame
pMB-G52	*L. mexicana* transketolase episomal expression construct conferring resistance to G418
pMB-G53	Empty episomal expression construct conferring resistance to G418
pMB-G105	pGEM-T Easy (Promega) construct containing the *L. mexicana* VDAC1 open reading frame
pMB-G111	*L. mexicana* VDAC1 episomal expression construct conferring resistance to G418

2.3.8 Preparation of glycerol stocks

Glycerol stocks of transformed *E. coli* cell lines were prepared by mixing 0.5 ml of dense overnight culture, with 0.5 ml of 30% glycerol in LB medium. Glycerol stocks were stored at -80°C for future use.

2.4 Southern Blotting

Genomic DNA, at a concentration of between 1 and 10 µg, was digested overnight using between 5 and 50 U of restriction enzyme to ensure complete digest of the sample. The DNA fragments were fractionated on a 0.7% agarose gel at 1 volt cm^{-1}. The gel was subjected to two depurination washes for 15 minutes (Appendix 7.2.3), two denaturation washes for 30 minutes (Appendix 7.2.4), and two neutralisation washes for 30 minutes (Appendix 7.2.5), before

the DNA was transferred to a Hybond-N membrane (Amersham, Chalfont St Giles, UK) overnight in the presence of 20x SSC transfer buffer (Appendix 7.2.6).

The following day, the membrane was washed in 6x SSC, and the DNA was cross-linked to the membrane by exposure to UV using the SpectroLinker XL-1000 UV Crosslinker (Advanced Engineering Ltd, Basingstoke, UK). The membrane was stained in 0.002% methylene blue (Sigma-Aldrich, Poole, UK) to visualise the DNA.

Prehybridisation of the membrane was carried out for a minimum of two hours at 42°C in 50% formamide (Promega, Southampton, UK), 5 x SSC, 10 x Denhardt's solution (Appendix 7.2.7), 1 % SDS, 20 mM NaH_2PO_4, 5 mM EDTA pH 8, and 0.2 mg/ml denatured sonicated salmon sperm DNA (Stratagene, La Jolla, CA, USA).

The probe was prepared using purified plasmid DNA fragments. The desired fragment (tkt ORF or tkt 3' flank) were obtained from the appropriate plasmid construct (section 2.3.7), quantified using a 2-log DNA ladder (NEB, Hitchin, UK), and labelled with $\alpha^{32}P$ (Perkin Elmer, Beaconfield, UK) by random priming using the Prime It II Kit (Stratagene, La Jolla, CA, USA) according to the manufacturer's instructions. The radio-labeled probe was denatured at 95°C for 5 minutes, and added to the membrane. Hybridisation was performed overnight at 42°C.

The following day the solution was discarded, and the membrane incubated with preheated 0.1x SSC, 0.1% SDS and incubated at 56°C for 45 minutes to remove non-specific binding. Two additional wash stops at 58°C for 30 minutes were required. The southern blot was heat sealed, placed in a developing cassette with a sheet of KODAK Medical X-Ray Film General Purpose Blue (MXB), and exposed overnight at -80°C.

2.4.1 Eluting the probe from a membrane

The radiolabelled probe was eluted from the membrane by washing the membrane three times in a boiling 0.1% SDS solution. The membrane was then exposed to a film for 48 hours to ensure the probe had been removed. The membrane could then re-probed using a different probe as previously described in section 2.4.

2.5 Species Identification

In order to confirm that the parasites used in this study were *Leishmania mexicana*, a simple PCR based approach was used (Schonian, G *et al*. 2003). A repeat region from the 18S ribosomal locus was amplified by PCR using oligonucleotides MB95 and MB96, and purified using the QIAquick® PCR Purification Kit (Qiagen, Crawley, UK). The eluted DNA was digested with *Hae*III (Promega, Southampton, UK) and the products fractionated on 2% LE Agarose (NuSieve) + 1% regular agarose gel (Invitrogen, Paisley, Scotland) in 0.5x TAE. The gel was stained in a 100 ml 1x TAE with ethidium bromide (Sigma-Aldrich, Poole, UK) at a final concentration of 0.2 µg / ml. The restriction digest profiles were compared to those of known standards, thus enabling identification of the species.

2.6 Transfection of Leishmania promastigotes

2.6.1 Preparation of DNA

Plasmid DNA was prepared and quantified as previously described (section 2.3.6). A total of 50 µg of plasmid DNA was digested overnight with the appropriate restriction endonuclease enzymes. The DNA was then fractionated by electrophoresis and purified as previously described (section 2.3.4). A total of 10 µg was required for each transfection event.

2.6.2 Transfection and cloning

L. mexicana promastigotes were transfected using the Amaxa system (Amaxa AG, Köln, Germany). Approximately 3 mL of a culture in the stationary phase of growth were pelleted at 1,250 g, and resuspended in the Human T-cell transfection buffer as described in the manufacturer's instructions. Approximately 10 µg of DNA was added to this solution, and the mixture electroporated using program U-033. The electroporated cells were added to 10 ml of pre-warmed HOMEM media, immediately split into 2 flasks to generate clones derived from independent transfection events, and incubated at 25 °C overnight in the absence of antibiotic selection.

The following day, 4 ml of cell suspension is added to 20 ml of HOMEM supplemented with the appropriate antibiotic at 50 µg/ml to select parasites expressing the drug resistance marker. This was followed by two 12-fold serial

dilutions, all of which were plated out in 96-well plates at 200 µl per well to select clones. Plates were incubated at 25°C for three to four weeks; clonal cell lines were then analysed for correct integration of the DNA cassette.

For *Leishmania* promastigotes expressing ectopic copies of a gene, the appropriate antibiotic was added to the medium 24 hours after transfection. These promastigote populations were maintained in the presence of antibiotic to prevent loss of the ectopic plasmid.

2.7 Alamar Blue Assay

The alamarBlue® assay (Raz, B *et al*. 1997) was used to determine the EC^{50} values for a variety of trypanocidal and leishmaniacidal compounds. Briefly, 100 µl of medium was added to each well in a 96-well plate. To generate serial dilutions of a drug, 100 µl was added to the first column, carefully mixed by pipetting, and transferred to the next column. This was repeated across the row, with the last column remaining drug free. *Leishmania* promastigotes at a concentration of 1 x 10^6 cells per ml were added each well, and incubated at 25ºC for 72 hours, at which point 20 µl of resazurin (0.049 mM) was added to each well, and incubated for a further 24 hours. The plates were read using fluorometer (emission 530, excitation 595) (FLUOstar Optima, BMG Labtech). The fluorescence values could be plotted using GraphPad Prism5 to determine the EC^{50} values and the 95 % confidence interval values.

2.8 Proteomic Analysis

2.8.1 Protein Preparation and Precipitation

The density of *L. mexicana* promastigotes in the mid-late log phase was determined as described in section 2.1.3. Approximately 1 x 10^8 cells were harvested and washed twice in 1x PBS, and then pelleted at 1,250 g for ten minutes. The pellet was then resuspended by vigorous pipetting in 1 ml of ice cold lysis buffer (Appendix 7.3.1) supplemented with 10 µg/ml DNase, 5 mM $MgCl_2$, and a protease inhibitor cocktail. The cells were lysed by sonication (1-2 seconds probe sonication followed by one minute on ice for three repetitions). The cell lysates were then incubated at room temperature for ten minutes, with gentle mixing, before being centrifuged at 16,000 g for 10 minutes to remove

the insoluble material. The supernatant was recovered and stored at -20°C for further analysis.

The protein samples were then added to 4x sample volume of -20°C acetone, vortexed, and incubated overnight at -20°C. Samples were centrifuged at 16,000 g for 10 minutes, and the pellets washed in 80 % acetone at -20°C. Samples were again centrifuged at 16,000 g for 10 minutes and the supernatant was discarded. The pellet was air dried for 5 minutes at room temperature, before being re-suspended in lysis buffer (Appendix 7.3.1) containing a protease inhibitor cocktail (Appendix 7.3.2). Complete re-suspension involved vortexing, and vigorous mixing. The samples were centrifuged at 16,000 g for 10 minutes to remove any insoluble material and the supernatant was recovered and stored at -20°C.

2.8.2 Bradford Assay

The Bradford Assay was used to determine protein concentration. Bradford Reagent was diluted 5-fold in distilled water, filtered and stored at 4°C. A series of BSA (Promega, Southampton, UK) standards were prepared; the linear range of the assay is between X µg/ml and Y µg/ml. For the unknown protein sample, 10 µl was mixed with 790 µl dH_2O and 200 µl diluted Bradford Reagent. The protein samples were incubated for 30 minutes, and absorbance at OD_{595} was measured using a spectrophotometer. The absorbance values of the protein standards were plotted on a scatter graph using Microsoft Excel; unknown protein samples could be determined using the equation: $y = mx + c$. The protein concentration of the wild type and resistance samples were adjusted to 5 mg/ml.

2.8.3 Cyanine Dye Labelling

Three protein samples are labelled with cyanine dyes at an alkaline pH; these samples are condition 1, condition 2, and a pooled standard. The pooled standard has equal quantities of protein from condition 1 and condition 2. 50 µg of each protein sample in a 10 µl volume is mixed with 400 pmol CyDye and incubated on ice in the dark for 30 minutes. The labelling reaction is stopped by the addition of 1 µl of 10 mM lysine, and incubated for a further ten minutes. Labelled samples can be stored at -80°C.

2.8.4 DiGE

The cyanine labelled protein samples were mixed together and made up to 460 µl with DiGE Rehydration Buffer (Appendix 7.3.3). This solution was then added to the Iso-Electric Focusing (IEF) ceramic tray, and an IEF strip (4-7) was laid gel side down, covered with 1 ml of immersion oil, and left overnight at conditions stated in Table 2-3.

Table 2-3. Iso-Electric Focusing conditions.

Step	Voltage (V)	Time (hrs)	Description
1	30	10	step and hold
2	300	2	step and hold
3	600	2	gradient
4	1000	2	gradient
5	8000	3	gradient
6	8000	8.5	step and hold

The IEF strips were removed from the ceramic dish with care, and washed in 10 ml DiGE Equilibration Buffer 1 for 15 minutes (Appendix 7.3.4), 10 ml DiGE Equilibration Buffer 2 for 15 minutes (Appendix 7.3.5), and then briefly with 1x SDS running buffer, before being placed at the top of the large SDS gel. Hot agarose was used to hold the IEF strip in place.

The proteins were separated overnight at 120 volts, with 1x SDS running buffer in the bottom chamber, and 2x SDS running buffer in the top chamber. Electrophoresis was stopped around one hour after the bromophenol blue dye had reached the bottom of the gel. Gels were scanned using the typhoon scanner, then fixed for two hours in DiGE fixing solution (Appendix 7.3.6).

2.8.5 Colloidal Coomassie Staining

Colloidal Coomassie blue was initially used to stain the protein gels. Colloidal Coomassie dye stock was prepared as follows; 50 g Ammonium sulphate, 6ml 85% phosphoric acid, dH_2O to 500 ml, 10 ml Coomassie stock (5% Coomassie Brilliant Blue G-250 in dH_2O). Initially, the gel was fixed for one hour (40% ethanol, 10% acetic acid), washed twice with dH_2O for 10 minutes, and stained overnight in four parts Colloidal Coomassie dye stock to one part methanol. The staining step could be extended to improve sensitivity.

2.8.6 SYPRO Orange Staining

SYPRO® Orange was used as an alternative to Colloidal Coomassie stain due to increased sensitivity. Gels are fixed for a period of 1-2 hours in 7% acetic acid and 10% methanol. The gel was then stained in freshly prepared Sypro Orange Protein Gel Stain (Sigma-Aldrich, Poole, UK) (diluted 1:10,000) in 7% acetic acid and immediately scanned on Typhoon using the green laser (emission 580nm).

2.8.7 Spot Picking and protein identification

The gel images generated by the Typhoon scanner were processed by the DeCyder software, version 6 (GE Healthcare) to generate a pick list. To obtain peptide fragments from the gel, it was necessary to perform an in-gel trypsin digest. The desired gel spots were cut from the gel and initially washed in 100 mM ammonium bicarbonate whilst shaking at room temperature for 1 hour, followed by a second wash step of 50% acetonitrile/100 mM ammonium bicarbonate. The proteins were then reduced by treatment with 3 mM DTT in 100 mM ammonium bicarbonate at 60°C for 30 minutes. This was followed by alkylation step, involving treatment with 10 mM iodoacetamide at room temperature in the dark for 30 minutes. The gel pieces were washed with 50% acetonitrile/100 mM ammonium bicarbonate whilst shaking at room temperature for 1 hour, before dehydration with 100 µl acetonitrile at room temperature for 10 minutes. Samples were then dried to completion using a speed vac, and rehydrated in an appropriate volume of trypsin (2 mg/ml in 25 mM ammonium bicarbonate). Digestion was performed overnight at 37°C. The supernatant was recovered, and the gel pieces washed with 50% acetonitrile for 10 minutes. This wash was pooled with supernatant initially recovered, and the sample dried to completion. The tryptic peptide fragments were solubilised in 0.5% formic acid, and fractionated on a nanoflow HPLC system (Dionex, Camberley, UK), before being analysed by ESI mass spectrometry on a Q-STAR® Pulsar I hybrid MS/MS System. Separation of peptides was achieved using a Pepmap C18 reversed phase column (Dionex, Camberley, UK), using a 5 – 85% acetonitrile gradient (in 0.5% v/v formic acid), over 45 minutes, with a flow rate was 0.2 µl / min. The mass spectrometric analysis involved a 3 second survey MS scan, followed by MS/MS analyses on the most abundant peptides (3 second per peak) in Information Dependent Acquisition (IDA) mode, choosing 2+ to 4+ ions above the threshold of 30 counts, with dynamic exclusion for 120 seconds. The Applied Biosystems Analyst QS (v1.1.) software and the automated Matrix Science Mascot Daemon

server (v2.1.06) were used to analyse data generated using the Q-STAR® Pulsar I hybrid mass spectrometer. The Mascot search engine[4] was used to assign probability based MOWSE scores to identify potential *Leishmania* protein matches. A tolerance of 1.2 Da was used for MS analysis, whereas 0.4 Da was used for MS/MS. Variable methionine oxidation was allowed in searches, and the carbanidomethylation of cysteines was selected as a fixed modification.

2.9 Metabolomics

2.9.1 Determining Glucose Concentration

The levels of glucose present in the medium were measured over a 90 hour period. *Leishmania* promastigotes were counted, and approximately 5 x10^7 cells were pelleted and resuspended in 5 ml fresh medium. The promastigotes were incubated at 25°C, with samples taken at 0 hours (T_0), 30 hours (T_{30}), and 90 hours (T_{90}). An accurate cell count was obtained in triplicate, and a 500 µl aliquot was centrifuged at 1500 g for 10 minutes. The supernatant was recovered and stored at -20°C.

Quantification of glucose concentration was determined using the Glucose Oxidase (GO) assay kit (Sigma-Aldrich, Poole, UK) according to the manufacturer's instructions.

The assay is accurate at determining between 20 and 80 µg glucose/ml. The concentration of glucose in HOMEM is approximately 15 mM; therefore 30 µl of HOMEM would contain 80 µg glucose. Therefore, 30 µl of sample would contain the maximum quantity of glucose in the present in the supernatant. As a control, HOMEM with no cells was used.

The sample was diluted in a total volume of 1 ml ddH_2O, and the reaction started by adding 2 ml of Assay reagent. The reaction was allowed to occur for 30 minutes, before being stopped by the addition of 2 ml 12 N H_2SO_4. Absorbance at 540 nm was recorded, and the quantity of glucose present in each sample was derived from the glucose standards.

[4] www.matrixscience.com

2.9.2 Preparation of metabolite extracts for FT-MS analysis

Trypanosoma brucei procyclic form (PCF) strain 427 were cultivated in SDM80 as described in section 2.1.2. To generate a glucose-rich environment, the SDM80 was supplemented with 10 mM D-glucose, and 0.1 mM proline. A glucose deficient medium was produced by supplementing SDM80 with 10 mM L-proline, and 0.1 mM D-glucose.

Cells were cultured in 175 cm^2 flasks (Corning®) at a volume of 100 ml per flask, where they typically reached densities of between 1 x 10^6 cells/ml (mid-log phase) and 1 x 10^7 cells/ml (late-log phase). The cells were pooled and then pelleted at 1,250 g for 10 minutes at 4°C. The low temperature would slow the rate of metabolism, minimising the degradation of intermediate metabolites. The cell pellet was re-suspended in 1 ml of serum free glucose-rich SDM80 or serum free glucose deficient SDM80, and kept of ice whilst an accurate cell count was determined (described in section 2.1.3). The volume was adjusted to reflect a final concentration of 1 x 10^9 cells/ml; and the cells were incubated at 28°C for 30 minutes to allow steady state metabolism to re-establish after centrifugation.

To obtain total metabolite extracts (both extracellular and intracellular), 0.2 mL aliquots of concentrated cell culture were applied into 0.8 mL 100% hot ethanol and incubated at 80°C for 2 minutes to lyse the cells and quench metabolism. The lysate chilled on ice for five minutes, briefly vortexed, and then centrifuged at 16,000 g for 2 minutes to remove any insoluble material. The supernatant was recovered, flash frozen in liquid nitrogen, and stored at -80°C prior to analysis.

To obtain the extracellular metabolite extracts, 0.2 mL aliquots of concentrated cell culture were centrifuged at 4°C for 5 minutes to separate the cells and spent media. The supernatant was re-suspended in 0.8 mL 100% hot ethanol and incubated at 80°C. The cell pellet, containing the intracellular metabolites, was re-suspended in 1 mL 80% hot ethanol and incubated at 80°C. Cell lysates were processed as previously described for obtaining the total metabolites.

2.9.3 Liquid Chromatography Mass Spectrometry

Liquid Chromatography-Mass Spectrometry (LC-MS) data were acquired with a ZIC-HILIC column (5mm, 150 x 4.6mm; HiChrom, Reading, UK) coupled to a

Finnigan LTQ Orbitrap instrument (Thermo Fisher, Hemel Hempstead, UK). The analysis was performed using both positive and negative ionisation centroid modes of detection. Parameters included a mass scanning range of 50–1200 m/z, 25,000 resolution, spray voltages +4.5 kV and -2.6 kV, capillary temperature 200°C, the sheath gas flow rates 30 units, and the auxiliary gas flow rates 10 units (Kamleh, A *et al*. 2008). The apparatus was operated by Muhammed Anas Kamleh at the University of Strathclyde.

The LC/MS system (controlled by Xcalibur version 2.0, Thermo Fisher Corporation) was run in binary gradient mode. Solvent A was 0.1% v/v formic acid/water and solvent B was acetonitrile containing 0.1% v/v formic acid; the flow rate was 0.3 mL/min. A ZIC-HILIC column (5mm, 150 x 4.6mm; HiChrom, Reading, UK) was used for all analyses. The gradient was as follows: 80% B (0 min) to 60% B at 8 min to 60% B at 24 min to 80% B at 32 min.

2.9.4 Metabolomic data processing

Initially, raw data files were supplied as Microsoft Excel files containing the m/z ratios of the 65,000 peaks of greatest intensity. This data was subsequently analysed using MetaNetter, a plug-in developed by Fabien Jourdan (INRA UMR1331, Toulouse, France), in an attempt to identify biochemical pathways.

Alternatively, each raw data file was converted to mzXML using msconvert, and the metabolite signal peaks eluted from LC-MS apparatus were isolated, identified, and quanitated using the mzMatch software developed by Richard Scheltema (Max Planck Institute for Biochemistry, München, Germany) (Scheltema, RA *et al*. 2008; Scheltema, RA *et al*. 2011). The mzMatch software features a RSD filter which can be used to filter the data, thus reducing noise and non-reproducible peaks. The XML files contained a maximum of 460,000 individual mass measurements in 800 scans, were converted to a human readable tab-delimited text file which could be interpreted using Microsoft Excel or Microsoft Access.

2.9.5 Database development

A significant proportion of this thesis focused on improving metabolomic data analysis methods. Initially, metabolite information was stored using Microsoft Excel, although as the quantity and complexity of data increased, it was

necessary to transfer these tables to Microsoft Access, which has a greater capacity to deal with large volumes of data.

Microsoft Access uses SQL commands and macros to sort and query the data. There was no prior SQL programming knowledge involved in the development of the database; consequently, all information was obtained from Microsoft Office Access 2003 QuickSteps (John Cronan, Virginia Andersen, and Brenda Bryant Andersen), and an 'Introduction to Access' course run by IT Services, University of Glasgow. A significant extent of the process was trial and error, although advice was provided by Richard Scheltema (Max Planck Institute for Biochemistry, München, Germany), Richard Orton (School of Veterinary Medicine, University of Glasgow), and Fabien Jourdan (INRA UMR1331, Toulouse, France).

3 Developing a database to analyse Kinetoplastid metabolomic experiments

3.1 Introduction

Technological advances over the last 20 years have propelled biological sciences into the genomics era. By the end of 2011, complete genome sequences existed for 3080 organisms (Kyrpides, NC 1999)[5], and with the introduction of next generation sequencing this looks set to rise exponentially. However, there is a limit to what one can infer from the genetic sequence of an organism. The study of mRNA (transcriptomics) and proteins (proteomics) have become established over the last decade, and in turn, have generated a plethora of information about gene and protein regulation. Even then, these techniques focus on quantifying transcript or protein levels in a specific system, and ultimately confer no information on the activity of proteins within a system.

FT-MS technology enables the mass measurement of low molecular weight compounds from a complex mixture, to sub PPM accuracies. The identification of metabolites, and their relative intensities, can provide a 'snap shot' of the metabolome that reflects the physiological state of a cell at a given time point.

'Omic' technologies, by their global nature, generate large quantities of data, and metabolomics is no exception to this. The datasets are extremely data rich, and information poor, presenting a problem in the identification of metabolites. A raw data file obtained from the LTQ-Orbitrap yields approximately half a million masses. The greatest challenge is to identify the masses which are representative of genuine compounds, whilst eliminating those which are background noise. It is of course advantageous to know which metabolites are expected to change between two biological conditions, provided it is a manageable number. However, the global metabolomics approach provides an opportunity to investigate biological systems with no preconception as to which metabolites are fluctuating, thus reducing experimental bias.

[5] http://genomesonline.org

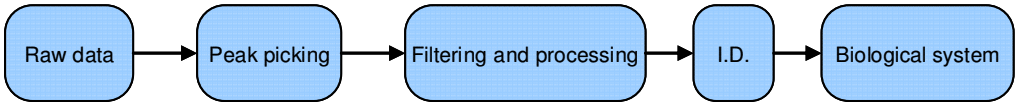


Figure 3-1. Metabolics pipeline.

The main stages of data acquisition and data analysis for a metabolomic experiment. Raw data is acquired by mass spectrometry. Peak picking and the filtering and processing steps are performed by software such as Xcalibur or Sieve, Thermo Fisher Scientific. Identification of metabolites and their context within a biological system can be performed by MassTrix (Suhre, K and Schmitt-Kopplin, P 2008), although alternative software is now available (Scheltema, RA *et al.* 2011).

The metabolomic pipeline illustrates the main stages of data acquisition and data analysis (Figure 3-1). The LTQ-Orbitrap, or similar FT-MS instrument, is responsible for acquiring the raw data. It is common to use a chromatographic technique, such as hydrophilic interaction chromatography (HILIC), to separate metabolites prior to mass measurements by the LTQ-Orbitrap (Kamleh, A *et al.* 2008). The HILIC approach is excellent in the separation of polar compounds, whereas lipophilic compounds will elute rapidly from the column (Kamleh, A *et al.* 2008). Altering the chromatographic separation technique or conditions may therefore affect the downstream data analysis.

Software available from Thermo Fisher Scientific is capable of 'peak picking' and 'filtering and processing'; however, there are limitations to these software. The Xcalibur (version 2.0) software supplied with the LTQ-Orbitrap (Thermo Fisher Scientific) is responsible for peak picking. The user can explore a mass spectrum generated from a single sample, and derive a list of potential formulae for any measured mass. Even at a precision of 1 PPM it is not always possible to assign unambiguously a formula to an experimental mass given the number of possible elements (t'Kindt, R *et al.* 2010); although more possible if the biological context is taken into consideration (Suhre, K and Schmitt-Kopplin, P 2008). The Sieve software, also provided by Thermo Fisher Scientific, is a reasonably powerful tool that allows filtering and processing of the data. Sieve uses principle component analysis to distinguish which masses are statistically different in abundance and peak height between different samples; although this does ignore signals of low intensity, and focus on the most abundant signals. To date, the commercially available software does not focus on identifying metabolites in these data files, or attempt to put this data in a biological

context. Therefore, it is imperative that the metabolomics community develop computer software tools to analyse these datasets.

MZMatch (Scheltema, RA *et al*. 2011) and MassTRIX (Suhre, K and Schmitt-Kopplin, P 2008) are two approaches developed by research groups in the metabolomics community in the last few years. Whilst both approaches deal with the identification of experimental masses, MZMatch is capable of filtering and processing raw data, whereas MassTRIX provides a biological context for the identified metabolites.

The MZMatch software, developed by Richard Scheltema (Max Planck Institute for Biochemistry, München, Germany),is capable of extracting and automatically aligning data from multiple measurements (Scheltema, RA *et al*. 2011), a task which the Xcalibur is unable to perform. This approach reduces the background noise and compiles a manageable list of reproducible masses with valid chromatograms. MZMatch compound identifications are provided through communication with the ChemSpider server, an online search engine which scans multiple biological compound databases that include both KEGG and HMDB. The retention time can be used to make an approximate distinction between isomeric compounds, although an orthogonal approach is required for a definitive identification. Whilst this is software is extremely powerful in the processing of raw data, there are deficiencies in the graphical representation of the results in a biological context.

MassTRIX provides graphical representation of FT-MS data using the KEGG website (Suhre, K and Schmitt-Kopplin, P 2008). A peak list can be uploaded to the MassTRIX web server, and mapped against the KEGG metabolic networks of a chosen organism to provide metabolite identification and a biological context. There are two main concerns with this approach; firstly, the raw FT-MS data file requires processing prior to analysis by MassTRIX, and secondly, the analysis is limited to compounds only present in the KEGG database. Details of the MassTRIX web server were published after work in this chapter had started; therefore, there are notable similarities between the two approaches, however there are subtle differences.

3.2 Experimental approach

Due to the change in environmental conditions that PCF and BSF trypanosomes experience during their lifecycle, energy metabolism has been extensively studied (van Weelden, SW *et al*. 2003; Besteiro, S *et al*. 2005; van Weelden, SW *et al*. 2005; Bringaud, F *et al*. 2006), however our knowledge is by no means complete. As procyclic trypanosomes are able to metabolise both D-glucose and L-proline, this provides a suitable model to study metabolic adaptations in response to changes in the principal carbon source. Whilst this particular experimental approach may not address metabolic changes in the human infective form or a clinically relevant species, the global metabolomics approach has the capability of detecting changes to the metabolome that are not predicted to occur. Traditionally, trypanosomatid enzymatic reactions have been identified experimentally using biochemical assays. However, in recent years, genome projects have identification of enzymes through inference from orthologous genes.

3.3 Aim

The initial aim of the experiments in this chapter was to take a global metabolomics approach by extracting metabolites from procyclic trypanosomes, using liquid chromatography coupled to an ultra-high resolution LTQ-Orbitrap mass spectrometer, and confirm metabolic changes within central carbon metabolism that had been previously characterised. However, it became quickly apparent that the tools to analyse the data were insufficient to handle complex datasets. Consequently, the focus of the project shifted from trying to interpret this particular dataset, to constructing a database to address the data analysis issues encountered by the metabolomics community with the following aims:

1. Produce a simple list of metabolites associated with trypanosomatids. Journal articles, biochemistry textbooks and online resources such as KEGG, provide an appropriate foundation for metabolites known to occur in these organisms, and are therefore expected to be present in datasets.

2. Identify a suitable database program and construct a simple relational database with extensive details on metabolites. Whilst a simple list of metabolites in a Microsoft Excel table is a useful reference, it is necessary to associate metabolites with specific biochemical pathways and biochemical reactions. Furthermore, a database system provides an organised method of storing and relating data

3. Develop a query to allow a user to analyse FT-MS derived experimental datasets. The raw data acquired during a metabolomic experiment is data rich and information poor. The current tools, such as Xcalibur (Thermo Fisher Scientific), support a targeted metabolomics approach where there are few metabolites of interest. However, if a global metabolomics approach is desired, there are insufficient tools to identify which metabolites are present in any given sample, and how those metabolites relate to a biological system.

4. Improve the efficiency of searching for metabolites. Whilst the search engine on the KEGG website allows a user to search for metabolites by name, or synonym, it does not for instance, provide the ability to search for metabolite if only the molecular formula or exact mass are known.

3.4 Simple Metabolite Table

There are many chemical compound and metabolite databases available online including KEGG (Kanehisa, M and Goto, S 2000)[6] and LeishCyc [7], which provide a range of metabolite data for trypanosomes and *Leishmania*. Using this information, with some editing based on knowledge of the scientific literature, a table of trypanosomatid specific metabolites was constructed. This table included information such as metabolite names, molecular formula, exact mass, links to original source, and what metabolic pathways the metabolite was involved in[8].

Composing a table of metabolites associated with trypanosomatids allows the user to browse for metabolites of interest, and search for the respective mass or formula in the mass chromatograms using the Xcalibur software. Alternatively, a list of the 65,000 most intense peaks could be imported into Microsoft Excel, and the 'find' function could be used to search for a mass. However, these two methods rely on a hypothesis as to which metabolites or pathways are likely to change. Furthermore, the masses or formulae of the metabolites of interest can only be searched on an individual basis, which is often time consuming, and consequently impractical. Therefore, the ability to automatically search a large number of masses against a list of experimentally derived masses would greatly enhance the understanding of analysis of metabolomic datasets.

Finally, the main advantage of an untargeted metabolomics approach to study global metabolism is that no prior hypothesis is required, as the technique has the potential to resolve the entire metabolome. It is of no real benefit to restrict the analysis to any one particular organism if it is possible to search a larger number of metabolites.

---

[6] http://www.genome.jp/kegg/

[7] http://leishcyc.bio21.unimelb.edu.au/

[8] Refer to CD: simple_metabolite_table.xls

3.4.1.1 Mono-Isotopic Mass

The exact mass, or mono-isotopic mass (MIM), of a compound is derived from the sum of the exact atomic mass of the most commonly occurring isotope. This differs from the average mass which takes into account the mass and abundance of all significant naturally occurring isotopes. For example, the exact mass of carbon is 12 amu, whereas the average mass of carbon is 12.0107 amu, as this takes into account the mass and abundance of both ^{12}C and ^{13}C (Table 3-1). As the metabolite data was acquired from different sources, the MIM for each metabolite had to be standardised to ensure consistency. Whilst database websites provide an exact mass, they do not provide a reference to the method used to calculate it.

Table 3-1. Naturally occurring isotopes of carbon.

	Isotope Atomic Mass Unit	Abundance
^{12}C	12	0.9893
^{13}C	13.00335484	0.0107

The exact mass, or isotope atomic mass unit, and abundances of the significant naturally occurring isotopes of carbon

An online exact mass calculator[9] derives the MIM for any compound or metabolite if the user provides the molecular mass. The process is quite laborious as each molecular formula had to be entered on an individual basis. One method to improve this would be to have a function capable of doing this in Microsoft Excel; unfortunately, no Add-In capable of calculating the MIM was freely available. Consequently, a Microsoft Excel Add-In was developed in collaboration with Richard Orton (School of Veterinary Medicine, University of Glasgow). The exact mass of each individual element was obtained via the source code from the exact mass calculator[10] and, using Microsoft Visual Basic Application, a script to calculate the MIM of a molecular formula from the sum of the individual elements was written. The MIM Add-In described here provides an efficient method of calculating the MIM using Microsoft Excel for any number of metabolites in a table.

[9] http://www.sisweb.com/referenc/tools/exactmass.htm

[10] view-source:http://www.sisweb.com/referenc/tools/exactmass.htm

3.5 A suitable database program

The initial step for creating a database was to identify a suitable database program from the several that are available; these include specialist programs such as Oracle or MySQL, or more simple applications such as Microsoft Access.

Microsoft Access is a relational database management system with a graphical user interface. A database file is created using several different types of database objects that include tables, queries, macros and forms. Importantly, all information is managed from this single file and consequently, is easy to keep track of.

In any database system the data are stored in tables; this is known as the back end of the database. A table is a collection of data arranged in rows and columns. Each column, or field, contains information about a specific type of information, such as the molecular formula of a metabolite; whereas each row, or record, contains information about a specific metabolite, such as metabolite name, molecular formula, and exact mass. A database that contains data in one table is known as a flat file system, whereas a database that contains more than one table is known as a relational database system. To develop a relational database, it is necessary to establish links between tables using common fields.

Considered as the front end of a database, queries are used to view, modify, and analyse data in a number of different ways. Queries are constructed using Structured Query Language (SQL), although Access features a 'query designer' which is simple to use, and thus negates prior knowledge and understanding of SQL. When creating a query using this method, Access automatically generates the SQL statement in the background, which can then be viewed or edited in the 'SQL view'. Microsoft Access supports numerous different types of queries, although not all can created using the 'query designer' and have to be created directly in the 'SQL view'.

A select query can be are used to extract data from one or more tables, creating a dynaset. This information is dynamically linked back to the original data, and will therefore change if the original data is altered in any way. Parameters can be introduced into any query to request specific data; for example, to display any metabolites which have a molecular formula of '$C_6H_{12}O_6$'.

There are a number of action queries that can be used to change data in specific tables. An update query can be used to make global changes to multiple records; for example, to add 22.98932 to every exact mass value to reflect the exact mass of a metabolite as a sodium adduct. It also is possible to delete specific records; for example, it is accepted that the Orbitrap cannot accurately measure the mass of ions less than 80 Da. Therefore, a delete query could be used to delete any metabolites which have a mono-isotopic mass of less than 80 Da. An append query can be used to add a group of records from one table to another; for example, a table of recently discovered metabolites could be appended to an existing metabolite table. A final action query allows the user to make a new table from all or part of the data in one or more tables; this is similar to the select query, however instead a producing a dynaset, a new table is created in which the data is not dynamically linked to the original data. Whilst action queries as useful, they should be used cautiously, as the data in tables can be physically altered.

Macroinstructions, or simply macros, can be used to perform a series of commands in one step, thus reducing the time spent executing each command individually. These are usually limited to simple commands such as opening or closing a table, or displaying a message box.

An Access form provides a graphical and interactive front end to a database. A form is a database object and can be used to enter or view data, operate as a custom dialog box, or function as a switchboard to open other forms in the database.

A form can be linked to one or more tables in the database, in which the form's record source refers to the any fields in the underlying table(s). Any tabular data is viewed in the same style as a traditional record card. It is also possible to create a new record and enter the data on the form, which is then automatically added to original table(s). For a form to function as a custom dialog box, it must accept user input and carry out an action based on that input. Command buttons can be used on a form to activate pre-defined events such as a query or a macro. For example, a user can enter '$C_6H_{12}0_6$' in a text box on a form, and use a command button to activate a query to select all metabolites with that specific molecular formula. These commands are controlled by the Visual Basic

(VB) code, an event-driven programming language. Through modification of this code, a single command button can be programmed to activate multiple queries and macros in a specific order, creating a one-step event-driven procedure.

There are several advantages to using Access as an alternative to MySQL or Oracle. As Access is part of Microsoft Office, it is available on most desktop computers and is relatively user-friendly; any tabular data can be easily imported from Microsoft Excel or text files; and finally, whilst knowledge of both SQL and VB is useful, especially for more complex commands, it is not essential for the creation of a simple database. However, there are some disadvantages of using Access instead of any specialist database programs. A single database file with large volumes of data requires a lot of RAM and processing power to run queries efficiently; therefore, a database is often split into the back end and the front end and stored on web servers. This approach is used by Suhre and Schmitt-Koplin in the development of MassTRIX, in which the front end web server communicates with the back end data stored on the KEGG API directory (Suhre, K and Schmitt-Kopplin, P 2008). The ability for other applications to communicate with the back end of a database is extremely useful for academic projects which ultimately benefit the scientific community.

Utilising the various features and functions of Access allows the dynamic development of a relational database file for the interrogation of experimental datasets and improved searching of metabolites.

The simple metabolite table described in section 3.4 only contained around one thousand metabolites, most of which are found in trypanosomatids, compared with approximately twelve thousand metabolites, some of which not known to occur in trypanosomatid biochemistry, in the KEGG database. It was therefore advantageous to retrieve the metabolite data from the KEGG website to provide a comprehensive and structured back end to the database.

3.5.1 Creating a flat file system

KEGG provides detailed information on metabolites, enzymes, and metabolic pathways for a number of organisms. Furthermore, the KEGG website provides a user friendly GUI allowing the user to navigate between metabolic pathways, and search for metabolites and enzymes of interest.

All the KEGG information is stored in a tabular form in the FTP directory, and can be viewed at the following link[11]. This information can be extracted using a Java script and saved as text file, or simply copied and pasted into a text file. These text files can then be opened in Microsoft Excel, an appropriate program for editing and organising the information prior to using Microsoft Access. All Java scripts were written by Richard Scheltema (Max Planck Institute for Biochemistry, Munich, Germany).

The following FTP directory link provides information on all the metabolites currently held within the KEGG database[12]. The information extracted from this directory included the metabolite name, molecular formula, exact mass, and the KEGG compound identifier – these attributes were identified as the most essential for the development of the database (Table 3-2).

Table 3-2. Basic metabolite information extracted from KEGG

Field	Description
Metabolite	The common name for the metabolite
MolecularFormula	The chemical formula of the metabolite
CompCode	The unique identifier for the metabolite
ExactMass	The exact mass stated in KEGG

The KEGG compound identifier, designated as the 'CompCode', is a unique value associated with each metabolite and is designated as the primary key. Furthermore, it is possible to generate a URL to the KEGG website using the CompCode as every metabolite in the KEGG database has its own web page. For example, trypanothione, a metabolite found in exclusively in trypanosomatids, has the following URL; http://www.genome.jp/dbget-bin/www_bget?compound+C02090. The URL can be divided into two parts; a fixed section and a variable section. The fixed section 'http://www.genome.jp/dbget-bin/www_bget?compound+' is the same regardless of the metabolite, whereas the variable section 'C02090' corresponds to the CompCode. Microsoft Excel has the ability to join the text strings of two or more cells using the 'concatenate' function. Therefore, we can generate the

[11] ftp://ftp.genome.jp/pub/kegg/

[12] ftp://ftp.genome.jp/pub/kegg/ligand/compound/compound

URL to every metabolite in KEGG provided we know the CompCode. Similarly, there is also a webpage for an image of the structure of each metabolite, and a URL for this can be constructed using the same process. These URLs provide the user with a direct link from the metabolite table to KEGG.

Table 3-3. Total KEGG Database

Field	Description	Section
Metabolite	The common name for the metabolite	this section
MolecularFormula	The chemical formula of the metabolite	this section
CompCode	The unique identifier for the metabolite	this section
Exact_Mass_(MIM)	Mono-Isotopic Mass of the metabolite	3.4.1.1.
KEGG_Link	URL for the metabolite specific web page	this section
Image_Link	URL for the image of structure	this section
InChI_ID	IUPAC International Chemical Identifer	3.5.2.2.
Leishmania_major	annotated in *L. major*	3.5.2.3.
Plasmodium_falciparum	annotated in *P. falciparum*	3.5.2.3.
Trypanosoma_brucei	annotated in *T. brucei*	3.5.2.3.
Trypanosoma_cruzi	annotated in *T. cruzi*	3.5.2.3.
ALL_KEGG	all metabolites in KEGG	this section
Shortlist	metabolites chosen by the user	3.7.5.
Measured_RT	retention time added by the user	1.5.1.2.

A list of the fields present in the Total KEGG database, and the relevant section they are discussed in this chapter.

As the method of calculating the KEGG exact mass values was not apparent, these were recalculated for each metabolite from the molecular formula using the MIM calculator (section 3.4.1.1). To ensure there were no major discrepancies, the difference between these two exact mass values was calculated.

Extracting some basic information on metabolites from the KEGG FTP directory, and expanding this information using some simple Excel functions has created a table of metabolites that is representative of a diverse number of metabolic pathways and organisms. Unlike the simple metabolite table previously discussed (section 3.4), the information that is displayed in this table is correctly structured with a unique identifier, and can thus be used in the development of the back end of a relational database.

3.5.2 Creating a relational database

Whilst the metabolite table (section 3.4) provides the user with basic information on metabolites, it lacks the detailed information on which organisms, metabolic pathways and reactions that the metabolites are involved in. To incorporate this information, the application must evolve from the flat file system containing one table, to a relational database system in which multiple tables are linked to one anther using a common entity.

It was necessary to identify and retrieve the useful information from the KEGG FTP directory, and create distinct tables in Microsoft Excel which could then be imported into Microsoft Access. Furthermore, it was important to consider how the information in these tables would relate to one another, and identify suitable primary and foreign keys.

3.5.2.1 Synonyms

Some metabolites may have more than one name and therefore it was logical to include this information to improve the ability to search for metabolites. The KEGG FTP directory contains a table which lists synonyms ('Name') and their corresponding CompCode ('Entry')[13]; a Java script was then used to retrieve this information to create Table 3-4.

Table 3-4. Synonyms – (metaSearch 'Table_147')

Field	Description
Metabolite_Synonym	A given name for a metabolite
CompCode	The unique identifier for the metabolite

In this instance, the CompCode cannot act as the primary key as it can be present as a duplicate value; however, the CompCode does act as a foreign key as it refers to the primary key of the metabolite table (section 3.5.1).

An example of a metabolite with numerous synonyms is trypanothione, which is also known as N^1,N^8-Bis(γ-L-glutamyl-L-cysteinyl-glycyl)spermidine; N^1,N^8-Bis(glutathionyl)spermidine; TSH; or reduced trypanothione (Table 3-5). All of these names are correct, and should therefore be included in the database.

[13] ftp://ftp.genome.jp/pub/kegg/ligand/compound/compound

Table 3-5. The synonyms for trypanothione listed in the KEGG database.

CompCode	Metabolite_Synonym
C02090	Trypanothione
C02090	N^1,N^8-Bis(γ-L-glutamyl-L-cysteinyl-glycyl)spermidine
C02090	N^1,N^8-Bis(glutathionyl)spermidine
C02090	TSH
C02090	Reduced trypanothione

3.5.2.2 IUPAC International Chemical Identifier

The IUPAC International Chemical Identifier, abbreviated to InChI, is a unique text string with multiple layers of information that may include the molecular formula, bond connectivity of atoms, stereochemistry, electronic charge, and isotope information[14]. Unlike the simplified molecular-input line-entry specification (SMILES) (Weininger, D 1988), every molecular structure has a unique InChI; an important attribute in database applications. Therefore, the InChI can be used to distinguish between isomeric compounds. Whilst the InChI is considered human readable, there are many computational algorithms available to convert the text string to the graphical representation of the compound.

A list of KEGG compound ID's and their respective InChI are stored on the KEGG FTP directory[15]; this information was extracted and imported into Microsoft Access (Table 3-6). To ensure that every metabolite in the total KEGG database table had an InChI, a query was used to update the 'InChI_ID' column using the 'CompCode' as a common entity (section 3.5.1).

Table 3-6. InChI (metaSearch 'Table_112')

Field	Description
InChI	The IUPAC International Chemical Identifier
CompCode	The unique identifier for the metabolite

[14] http://www.iupac.org/inchi/index.html

[15] ftp://ftp.genome.jp/pub/kegg/ligand/compound/compound.InChI

3.5.2.3 Metabolites specific for organisms

One of the disadvantages for considering the entire KEGG metabolite database is that many of these metabolites are known not to be present in trypanosomatids.

The KEGG FTP directory[16] contains a list of KEGG compounds IDs for 1222 organisms (129 eukaryotes, 1014 bacteria, and 79 archaea). Metabolite lists specific for *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania major*, and *Plasmodium falciparum* were incorporated into Microsoft Access to provide the user with greater flexibility when analysing datasets. This approach is particularly useful as the database can be adapted to any organism of interest, provided it has been annotated in KEGG. As a column for each organism had been created in the total KEGG database (section 3.5.1), a query was used to update the value in these columns to 'True' if the metabolite was known to be present. Several simple queries were written to filter the metabolites based on organism; these can be selected using the appropriate command button (Figure 3-2).

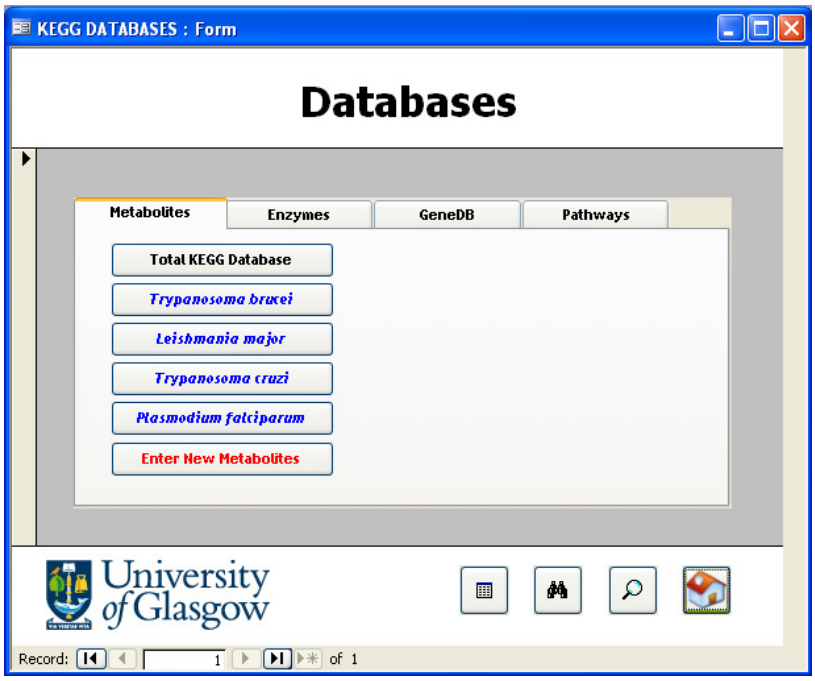


Figure 3-2. Databases.

The user can view metabolites from the Total KEGG database, or alternatively select one of the four organisms.

[16] ftp://ftp.genome.jp/pub/kegg/pathway/organisms/

3.5.2.4 Metabolites specific for reactions

The ability to provide a link between the metabolites and biochemical reactions is extremely useful as it allows the user to select a metabolite and determine which biochemical reactions that metabolite is involved in.

Information on biochemical reactions is available on the KEGG FTP directory[17] where 'definition' and 'entry' were identified as the essential attributes. The 'definition' field contains a description of the reaction, and is thus a useful reference for the user. The 'entry' field contains a unique identifier for the reaction, denoted 'KEGG_Reaction_ID', and can be used to generate a URL to the KEGG website; this was performed in Microsoft Excel by the concatenation function as described in section 3.5.1. This table was imported into Microsoft Access and saved as 'Table 142_KEGG_Reactions' (Table 3-7).

Table 3-7. KEGG biochemical reactions

Field	Description
KEGG_Reaction_ID	The unique identifier for the reaction
Reaction_Name	Description of the reaction
KEGG_Reaction_Hyperlink	URL for the reaction specific web page
KEGG_Reaction_Image_Hyperlink	URL for the image of reaction

Whilst this table provides a description of the metabolites in the Reaction_Name field, it does not provide a list of the metabolites with any relevant attributes such as mass or molecular formula.

However, the KEGG FTP directory contains a list of 'KEGG_Reaction_IDs' and 'CompCodes'[18] which was copied and then imported into Microsoft Access. The resulting table was saved as '135_Reactions_CompCodes' (Table 3-8). Duplicate values were permitted as one metabolite may be involved in many reactions, and one reaction must contain at least two metabolites. At a glance, this table contains no meaningful information; however is does act as an intermediate table providing a link between the metabolite and reaction tables.

[17] ftp://ftp.genome.jp/pub/kegg/ligand/reaction/reaction

[18] ftp://ftp.genome.jp/pub/kegg/ligand/reaction/reaction.lst

Table 3-8. KEGG biochemical reactions and metabolites

Field	Description
KEGG_Reaction_ID	The unique identifier for the reaction
CompCode	The unique identifier for the metabolite

A significant improvement to the reaction information includes the introduction of EC numbers and linking them to the reaction information. Whilst this information is stored in the same location of the KEGG FTP directory, it is logical to separate the EC numbers from the reaction details to eliminate duplicate values. One enzyme may catalyse more than one reaction, and conversely, one reaction may be catalysed by more than one enzyme. Therefore, the most practical solution is to store the 'KEGG_Reaction_ID' and 'EC numbers' in a separate table, saved as 'Table_136', and allow duplicate values for both fields (Table 3-9). This information can be retrieved from the FTP directory[19] using a Java script, or in this case, the table can be created manually to demonstrate the concept.

Table 3-9. KEGG biochemical reactions and EC numbers

Field	Description
KEGG_Reaction_ID	The unique identifier for the reaction
EC	Enzyme Classification number

The inclusion of EC numbers is meaningless, unless another table containing information on the enzyme names and EC numbers is introduced. The Expert Protein Analysis System (ExPASy) bioinformatic resource provides information on enzymes on the FTP site[20], which can be accessed using a Java Script. Unfortunately the information on reactions and enzymes in metaSearch is incomplete due to time constraints.

[19] ftp://ftp.genome.jp/pub/kegg/ligand/enzyme/enzyme

[20] ftp://ftp.expasy.org/databases/enzyme/enzyme.dat

3.5.2.5 Metabolites specific for pathways

Metabolic pathways are a convenient method of grouping together related metabolites. The metabolic pathways described in KEGG are one example of how metabolites may be grouped together. A list of pathway names and pathway IDs were obtained from the KEGG FTP directory. An additional field designated 'show' was created retrospectively to allow certain pathways to be ignored as many KEGG pathways, such as 'Insect Hormone Biosynthesis', are not relevant to many biological systems (Table 3-10).

Table 3-10. Pathway Names

Field	Description
Path_ID	The unique identifier for the pathway
Pathway_Name	The pathway name
Show	can be used to toggle pathways on and off

Also available from the KEGG FTP directory is a list of compound IDs and pathway IDs (Table 3-11); these can be used in an intermediate table to relate a metabolite to a metabolic pathway. Duplicate values may appear in both columns as there are many metabolites in one pathway, and any one metabolite may be present in more than one pathway. The 'CC_hyperlink' is a query generated attribute providing a hyperlink to the metabolite in the specific KEGG metabolic map. This is a concatenation of a fixed text string with the pathway code and compound ID to generate a URL.

Table 3-11. Intermediate Compound-Pathway table

Field	Description
CompCode	The unique identifier for the metabolite
Path_ID	The unique identifier for the pathway
CC_Hyperlink	hyperlink to a particular metabolite in a specific metabolic map

3.5.3 The database Schema

The flat file system is a relatively simple method of storing specific information about metabolites such as molecular formulae and MIM. However, the introduction of the additional tables allows the development of a relational database system providing the ability to retrieve detailed information on each metabolite.

For example, more detailed information on whether the metabolite is known to occur in biochemical reactions, metabolic pathways, or particular organisms can be linked table containing the basic metabolite information. Common fields in the various tables can be joined to one another, yielding a relational schema, as depicted in Figure 3-3.

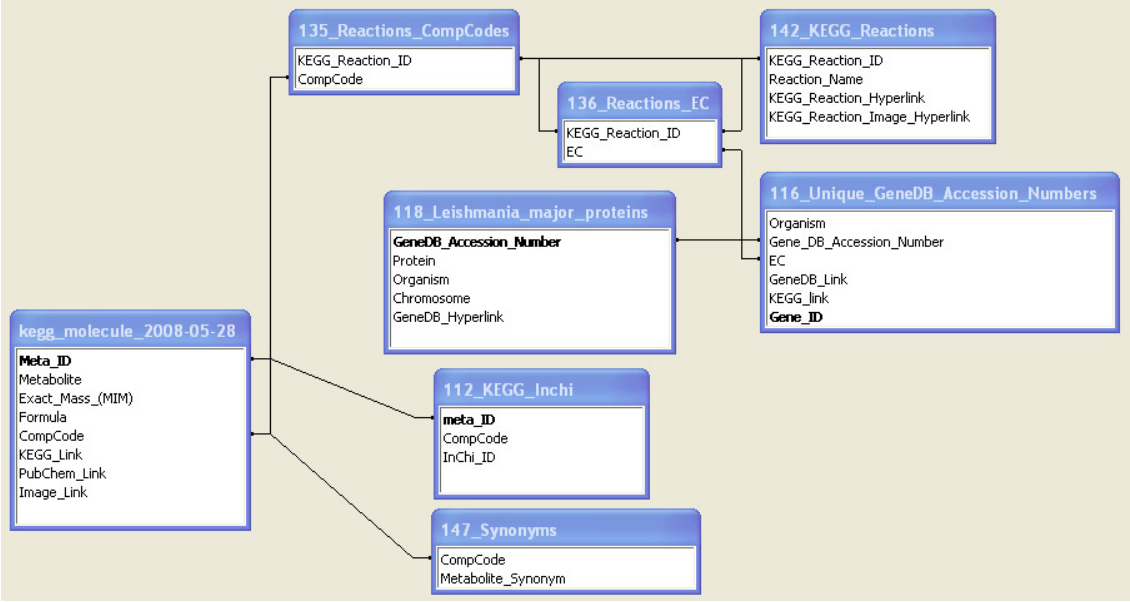


Figure 3-3. metaSearch relational schema.

The schema illustrates the relationships between metabolites and higher order functional information including synonyms, biochemical reactions, associated enzymes, and the *L. major* gene accession number.

The total KEGG database table was expanded with columns for InChI_ID, *T. brucei*, *T. cruzi*, *P. falciparum*, and *L. major*. A series of queries updated the information in these tables.

3.6 MetaSearch for analysis

The most challenging aspect of this particular project is to extrapolate the relational database to enable a user to enter a list of experimental masses and to match these to the MIM of known metabolites. As this is a relational database, it is possible to retrieve the detailed information on these metabolites as described in section 3.5.2. The primary goal of metaSearch was the development of a simple program that allowed a user to compare a list of experimental masses derived from kinetoplastid metabolomic experiments with a list of masses from known metabolites. Whilst this is relatively straight forward to implement using an SQL statement, it was more challenging to make the process user friendly. Furthermore, metaSearch is a multi functional database file which allows the storage and analysis of metabolomic datasets. Additionally, it would be advantageous to be able to upload results directly into the database instead of pasting data into a table, potentially increasing the chances of a mistake.

Two methods of analysis are described; the first involves selecting a particular dataset and finding out which potential metabolites are present, whereas the second focuses on finding a particular metabolite or mass irrespective of the experiment name.

3.6.1.1 Table of reference metabolites

Ideally, the experimental masses are compared to the MIM's of all the metabolites in the KEGG database; however, this requires substantial RAM and processing power. This problem was solved by creating a user defined intermediate table which would store the metabolites of interest (Table 3-12). It was necessary to include the MIM, molecular formula, CompCode, and the metabolite name in this table. As the purpose of this table is the temporary storage of data, which is at the discretion of the user, metabolite information can be selected by organism or metabolic pathway.

Table 3-12. Information stored in the user defined 'metabolite' table

Field	Description
Exact_Mass_(MIM)	Mono-Isotopic Mass of the metabolite
Formula	The molecular formula of the metabolite
CompCode	The unique identifier for the metabolite
Metabolite	The metabolite name

3.6.1.2 Table of experimental masses

All the experimental data is stored in the masses table (Table 3-13). The important fields for demonstrating the analysis query are mass and experiment. The retention time and intensity can be obtained from the PeakML output and included in this table for future reference. Currently, PeakML does not provide a P-value, although the ability to perform a statistical test has been taken into consideration when writing the SQL statements. Each mass in this table is associated with an experiment name; therefore a simple SQL statement could retrieve all the experimental masses associated with a particular experiment.

Table 3-13. Experimental masses (metaSearch 'masses')

Field	Description
MASS	experimentally measured mass of the ion
retention_time	time in which the ion elutes from the column
Intensity	the measured intensity of the ion
PValue	value for determining statistical significance
Experiment	the name of the experiment

3.6.2 User friendly

In Microsoft Access, the use of 'queries' and 'forms' provide a front end to the database; Queries allow the user to retrieve specific information from one or more tables, whereas forms greatly enhance the ability to generate an environment in which the user has both freedom to select the desired parameters and a controlled series of commands that can be followed. Command buttons can be implemented to allow the user to navigate between forms in a process similar to web pages.

Figure 3-4. MetaSearch main switchboard.

The main panel provides access to other forms and tables in the database through the use of command buttons and event procedures. The five options shown are as follows; databases, presents a list of tables which hold information of metabolites and proteins; input data (masses), opens a table to allow data to be uploaded to the software; analysing a dataset, allows the user to search their own data against metabolites from KEGG according to the desired parameters; search, allows the user to search the KEGG data for specific masses or metabolite names; quit, closure of the software.

3.6.3 Analysing a dataset

As previously mentioned a simple SQL statement could retrieve all the masses associated with a particular experiment. The aim of this query was to take an experimentally measured mass, subtract or add 'x' PPM to yield two masses (mass – 'x' PPM and mass + 'x' PPM), and determine if there is a MIM of a known metabolite within this range. As the value of 'x' increases, the greater the chance of matching an experimental mass to a known MIM, however the rate of error also increases since an incorrect metabolite name may be matched to a mass. Under optimal conditions, the LTQ-Orbitrap has an accuracy of greater than 2 PPM therefore an acceptable value for 'x' is 2; however this is at the discretion of the user.

To increase the flexibility of the search, the user must be able to select the appropriate parameters for the search which they wish to perform. Such parameters include the PPM tolerance, that organism that they wish to search against, and which experiment they wish to search.

3.6.3.1 Analysis Command

The analysis SQL statement must display all the experimental masses from a particular dataset that correspond to the MIM's of known metabolites within a given PPM tolerance. A series of SQL statements were required to achieve this. It is possible to run each SQL statement individually; however, it was more user friendly to have a single command button to activate the SQL statements in the correct sequence. This event procedure was programmed using visual basic application (VBA). All of the SQL statements performed during the 'analysis' command sequence involve intentionally deleting information from tables, or appending information to tables. Microsoft Access has several warning messages alerting the user to prevent any accidental modification of data or tables. Therefore, the initial step of the sequence is to turn these warning messages off.

A summary of these individual commands and the corresponding function are displayed in Table 3-14.

Table 3-14. Event procedure for the 'Analysis' command button.

Order	Command	Function
1	249_Warnings_Off	Turns off warning commands
2	071_Delete_all_records_in_131	Deletes all the records in table_131
3	065_Analysis	Compares the values in the dataset to those in the known metabolite table within the user chosen tolerance and adds the each formula to table_131
4	070_Delete_all_records_in_132	Deletes all the records in table_132
5	069_Add_Unique_Masses_To_New_Table_132	Filters duplicate formula from table_131 and adds them to as single records to table_132
6	250_Warnings_On	Turns on warning commands
7	254_Open_Table_132	Opens the table_132 which contains the Molecular Formula, Mono Isotopic Mass and KEGG isomers

A series of SQL statements (0##) and macros (2##) activated by the 'Analysis' command button.

The most important SQL statement is outlined in supplementary figure 8-9, although a simplied version is shown below in Figure 3-5. The MIM, formula, CompCode and retention time fields are selected from the 'metabolites' table and inserted into '131_Results' when the MIM of the metabolite is between the experimental mass + two PPM, and experimental mass – two PPM, and if that experimental mass is found in the experiment titled '*Leishmania* Dataset'.

```
---------------------------------------------------------------------------------------------------------------------------
INSERT INTO 131_Results ( [Exact_Mass_(MIM)], Formula, CompCode, RT_minutes )
SELECT DISTINCT Metabolites.[Exact_Mass_(MIM)], Metabolites.Formula, [109_Total KEGG database].CompCode,
Masses.Retention_Time
FROM Masses, [109_Total KEGG database] INNER JOIN Metabolites ON [109_Total KEGG
database].Formula=Metabolites.Formula
WHERE (((metabolites.[Exact_Mass_(MIM)]) Between Masses.Mass-(Masses.Mass/'1000000'*'2') And
Masses.Mass+(Masses.Mass/'1000000'*'2')) And ((masses.Experiment)='Leishmania Dataset');
---------------------------------------------------------------------------------------------------------------------------
```

Figure 3-5. 'Analysis' SQL statement.

The detail of any metabolite that corresponds to experimental mass within 2 PPM, provided that mass belongs to the '*Leishmania* dataset', is appended to a table named '131_Results'.

The analysis does not involve matching experimental masses to all the KEGG metabolites due to memory and processing constraints; therefore, the INNER JOIN statement was deemed necessary to link the results back to '109_Total KEGG database'. This would indicate any isomers in the KEGG database regardless of whether they were present in the 'metabolites' table initially used in the analysis. For example, if a mass within two PPM of 260.029723 was detected in a dataset, the metabolite could be glucose 6-phosphate (G6P, C00092). Whilst G6P is known to be present in trypanosomatids, the user should not assume that the experimental mass is definitely that of G6P, as it could be any metabolite with this MIM. There are a total of 38 metabolites in the KEGG database with a MIM of 260.029723. One possible method of distinguishing isomers from one another is the point at which they elute from the chromatographic column. Therefore, the inclusion of measured retention time data in the 'masses' table, and an option to assign retention times to the KEGG metabolite table, provide the user with an extremely valuable resource allowing an educated decision as to which isomer is present.

As previously stated, the results are appended to '131_Results', an intermediate table containing information on MIM, molecular formula, CompCode and retention time (Table 3-15). There are no metabolite names present in this table as the KEGG compound ID is used to identify the compounds. However, prior to inserting any new records in this intermediate table, it is necessary to delete all previous records using a 'delete' query (Supplementary Figure 8-10).

Table 3-15. Analysis 131_Results

Field	Description
Exact_Mass_(MIM)	Mono-Isotopic Mass of the metabolite
Formula	The molecular formula of the metabolite
CompCode	The unique identifier for the metabolite
RT_minutes	Elution time of the compound

Table 131 provides a list of compounds which may be present in the experimental dataset. It is important to remember that the identification of an experimental mass is not indicative of a particular metabolite; it is indicative of one or more specific mono-isotopic elemental compositions. Therefore, it was necessary to refine the data in Table 131 by grouping all the results by 'formula'

and append these to Table 132 using an ‘append’ query (Supplementary Figure 8-12). All previous data had to be deleted from this table by a delete ‘query’ (Supplementary Figure 8-11). This process removed isomeric entries, providing a list of distinct molecular formulae, which may be present in the experimental dataset (Table 3-16). The ‘KEGG_Isomers’ field is a count reflecting the number of isomers for each distinct molecular formula in the total KEGG database. A sub-datasheet linking the molecular formula attribute to the molecular formula attribute in the total KEGG database provides the user with the detailed metabolite information (section 3.5.1).

Table 3-16. Analysis 132_Unique_Formula_RESULTS

Field	Description
Formula	The molecular formula derived from the mass match
Exact_Mass_(MIM)	Mono-Isotopic Mass
KEGG_Isomers	The number of isomers in the KEGG database
RT_minutes	Elution time of the compound

The final steps of the visual basic program sequence used macros to turn the warning messages back on, and open the ‘132_Unique_Formula_RESULTS’ table to automatically present the user with the outcome of the analysis (section 3.6.3.4).

3.6.3.2 Input data

Whilst the user is able to enter their data directly into the 'masses' table, it is more convenient to paste the data into the intermediate table embedded in the 'Input Data' form (Figure 3-6). An experiment name is entered into the 'Experiment Name' text box. The 'Add' command button activates a series of SQL statements and macros to append the information to the 'masses' table, and then delete all the records in the intermediate table ready for the next data input. Navigation to the 'Analysing a Dataset' form is performed by clicking on the binoculars icon.

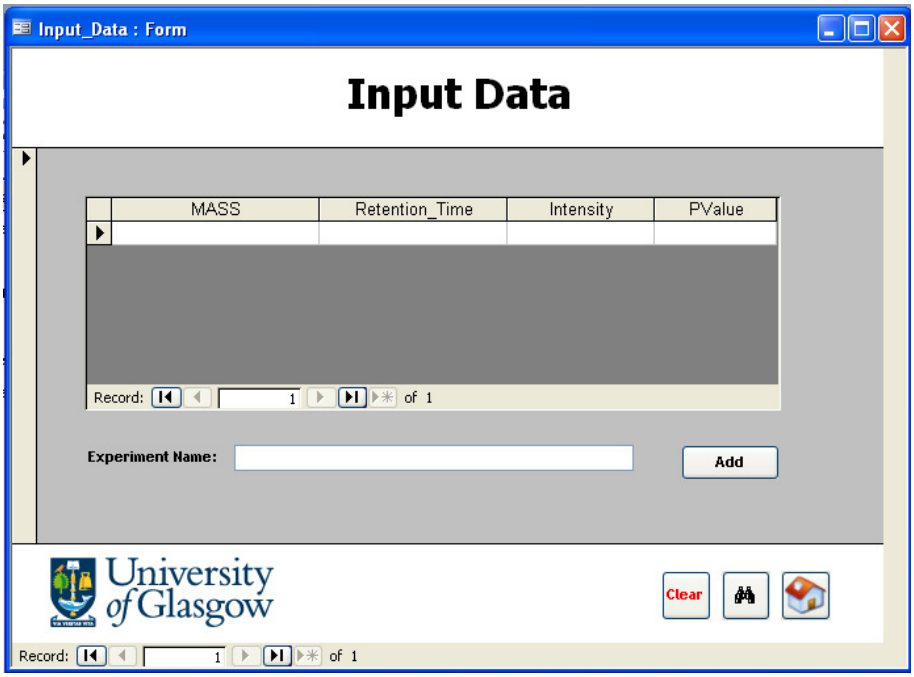


Figure 3-6. MetaSearch Input Data Form.

Experimental data can be pasted into the table embedded in the form. An Experiment name is entered in the 'Experiment Name' text box, and the data added to the 'masses' table using the 'Add' command button. There are navigation links to 'analysis' and 'home' at the bottom of the form.

3.6.3.3 The analysis tab

Use of the 'analysis' tab for analysing a dataset (Figure 3-7) does not require the user to have prior knowledge of SQL, thus maintaining the user-friendly aspect of metaSearch. The options available to the user include a series of command buttons, combo boxes and text boxes to input information and analyse experimental datasets.

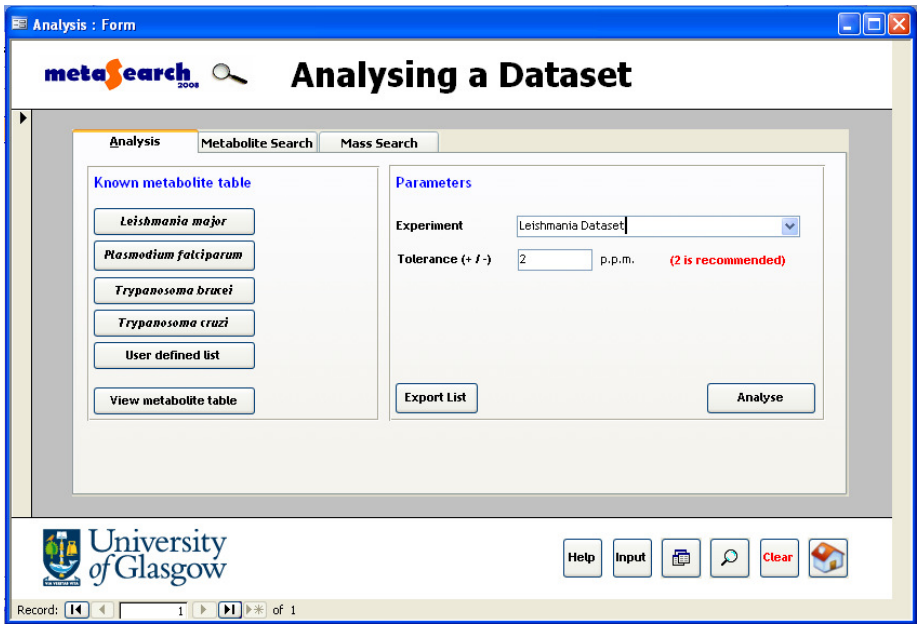


Figure 3-7. MetaSearch 'Analysis' Tab – analysing a dataset.

After the input of experimental masses this form displays command buttons that allow the user to select which experimental dataset they wish to analyse, the organism to compare the experimental masses to, and the desired PPM tolerance.

The left hand side of the form provides options for which metabolite table is used to match the experimental data to; currently there are four command buttons representing *L. major*, *T. brucei*, *T. cruzi*, and *P. falciparum*. There is also an option to choose a 'user defined list' that allows the user to create a custom reference table (described in section 3.7.5). Clicking any of these command buttons creates a table called 'metabolites'. The user may view this table by clicking on the 'view metabolite table' command button.

The right hand side of the form deals with the parameters for the SQL statement; these include the experimental dataset to be analysed and the PPM tolerance value. The 'Experimental' option is a combo box named 'ExpList' via which a SQL statement selects distinct experiments from the masses table

(QUERY_064). The PPM tolerance value is entered into a text box named 'PPMrange'. Links to the 'ExpList' and 'PPMrange' are used in the analysis query instead of '*Leishmania* Dataset' and '2' respectively. This allows the user to choose the desired experiment and PPM accuracy without having to directly alter the SQL statement. The 'analyse' command button activates the event sequence of SQL statements and macros (described in section 3.6.3.1). The 'export list' option presents the results in a simple table which can be copied to Microsoft Excel.

At the top of the form there are links to the two other analysis tabs currently available; these are 'metabolite search' and 'mass search' (described in section 3.6.4). There are several options at the bottom of the form to allow navigation to other forms in metaSearch; 'Input' (section 3.6.3.2); 'Tables' (section 3.5.2.3); 'Search' (section 3.7); 'Home' (section 3.6.2). Additionally, there is a 'clear' command button which deletes any value in the 'PPMrange' text box.

3.6.3.4 Output

The output of the analysis query is multi-layered, exploiting the relational database. The initial view, 'Table_132', displays the molecular formula, MIM, number of isomers in KEGG, and retention time (Figure 3-8A). These attributes were chosen as successfully matching an experimental mass to a MIM is indicative of a molecular formula rather than a particular metabolite. The number of isomers in KEGG is achieved by grouping the results in 'Table_131' by 'molecular formula' and counting the number of the compounds in each group.

The multi-layered aspect is achieved using sub-datasheets to relate the information in Table_132 to the main metabolite table by the molecular formula; clicking the '+' symbol next to the formula will reveal all metabolites with that particular molecular formula. Additionally, there is another sub-datasheet which links the metabolites to any biochemical reaction that the particular metabolite is involved in. For example, there is only one metabolite in the KEGG database with the molecular formula $C_{27}H_{47}N_9O_{10}S_2$. This metabolite is trypanothione disulphide, and is known to occur in five biochemical reactions (Figure 3-8B).

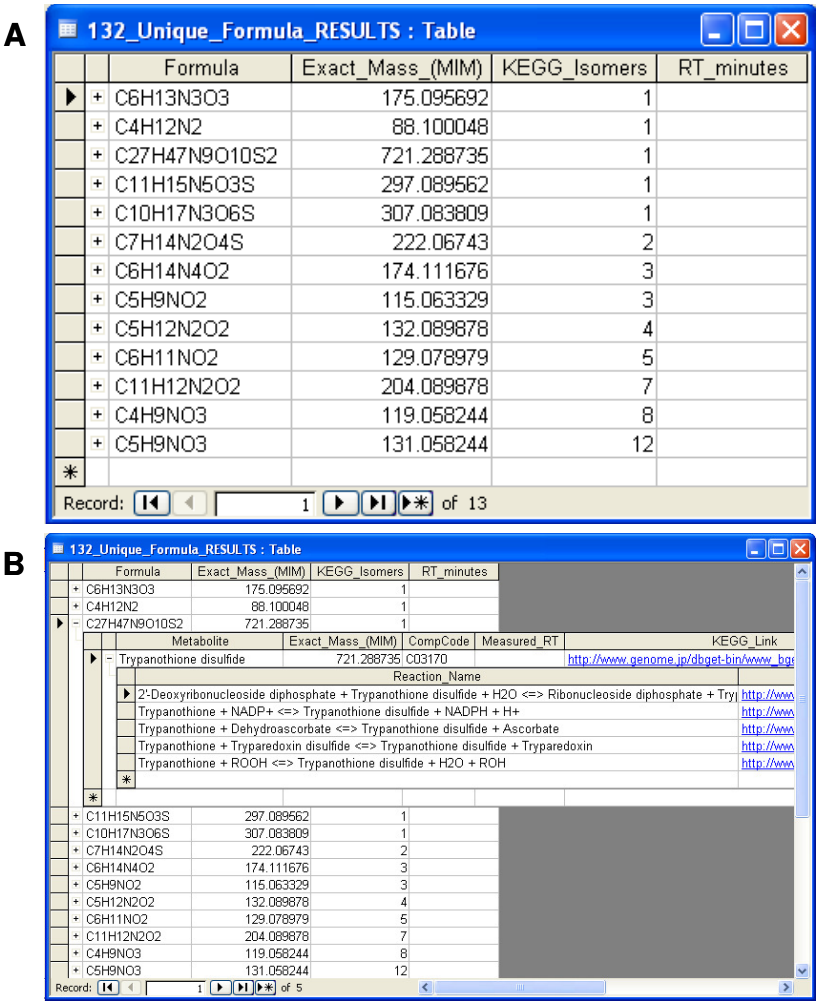


Figure 3-8. The output of the analysis query.

A – The initial view provides a list of molecular formulae, exact masses, the number of isomers, and retention times. B – Accessing the sub-datasheets utilises the relational aspect of the database to provide the user with more detailed information on the potential metabolite and any biochemical reactions the metabolite is involved in.

The initial view of the results is of limited use as no metabolite names are mentioned; however, the introduction of sub-datasheets to reveal additional levels of information is particularly valuable. If this level of detail is not required, there is an option to view a simple list of potential metabolites (‘Export List’ on the ‘Analysis’ tab - section 3.6.3.3) using a SQL statement.

3.6.4 Searching datasets for a single metabolite or mass

The current software used for metabolomic analysis tends to focus on one particular dataset. However, it is useful to search quickly for a metabolite of interest from all datasets, rather than searching every dataset individually.

Table 3-17. Event procedure for the 'Show metabolites' command button.

Order	Command	Function
1	249_Warnings_Off	Turns off warning commands
2	078_delete_values_from_139	Deletes all the values in table_139
3	077_Specific_Metabolite_Search	Identifies any metabolite names which matches the text string entered by the user and stores them in table_139
6	250_Warnings_On	Turns on warning commands
7	255_close_analysis_open_form_2	Opens the table_139 which contains the list of metabolites with MIM and molecular formula

A series of SQL statements (0##) and macros (2##) activated by the 'Show metabolites' command button.

In this two-step process, the user must initially search for the desired metabolite based on the partial or complete name (Figure 3-9A). Executing the 'Show metabolites' command button executes a series of SQL statements (Supplementary figures 8-13 and 8-14) and macros (Table 3-17), displaying all the possible metabolite names, in addition to their respective MIM and the molecular formula. The user then selects the desired metabolite, and enters a mass tolerance (Figure 3-9B). Finally, clicking the 'Search' command button activates a SQL statement (Supplementary figure 8-15) that will return a list of all experiments in which a mass corresponding to the MIM of that metabolite is present (Figure 3-9C). Both the experimentally measured mass and the MIM are displayed.

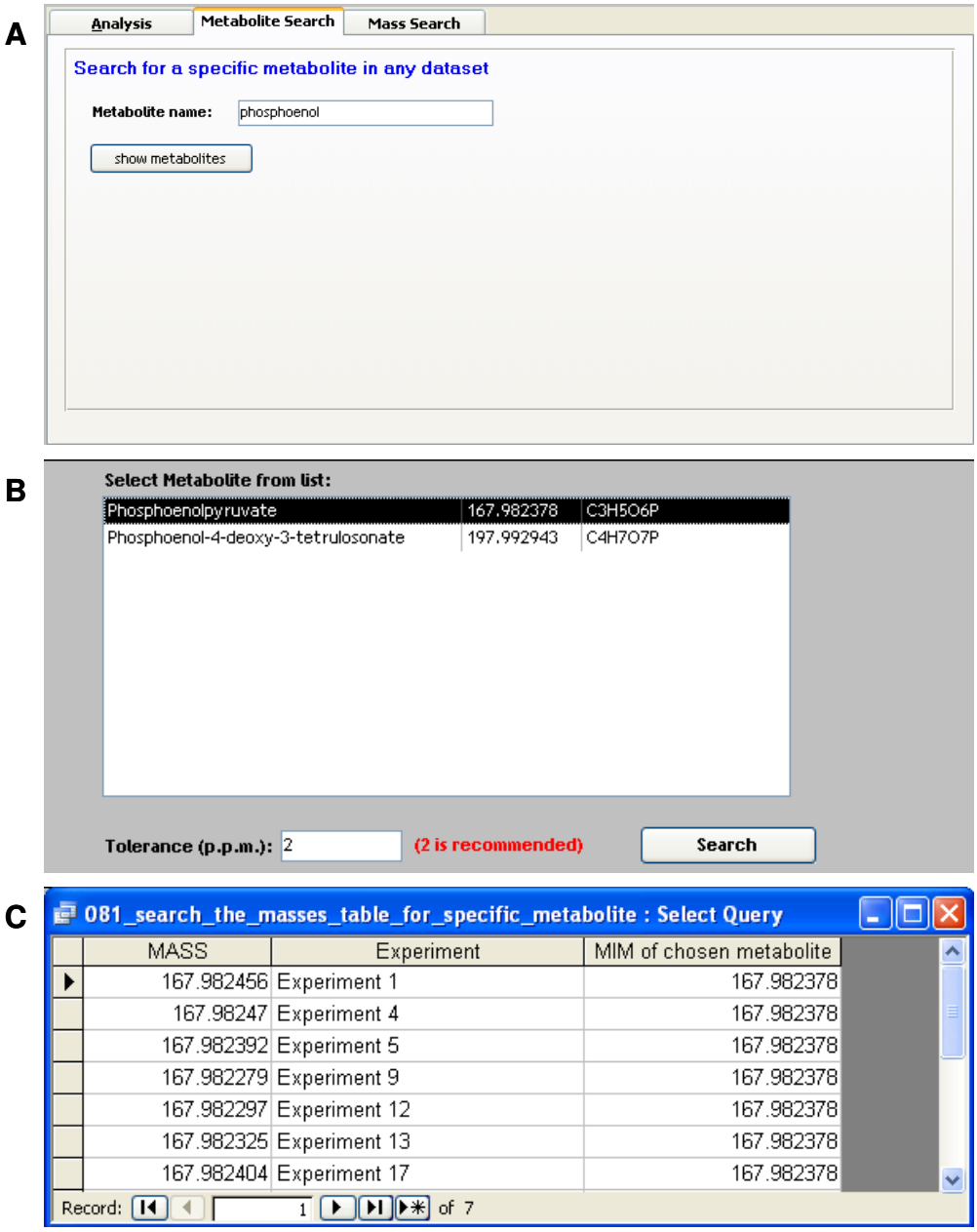


Figure 3-9. Searching for a specific metabolite.

A – Text box to enter complete or partial metabolite name. B – Selection of potential metabolite names and a text box for entering the PPM tolerance. C – List of experiments in which an experimental mass corresponding to the MIM of a selected metabolite.

However, as the mass used in the query is obtained by selecting for metabolite name, this query will only work if the metabolite of interest is present in the KEGG database. Therefore, an additional query was written to allow a user to search for a mass within a given PPM tolerance in all datasets. To test this, the mass for phosphoenolpyruvate (C00074) was used with a PPM tolerance of 2 (Figure 3-10A). The experiments returned were identical to those returned when selecting 'phosphoenolpyruvate' from the metabolite list, indicating that the query was working correctly (Figure 3-10B).

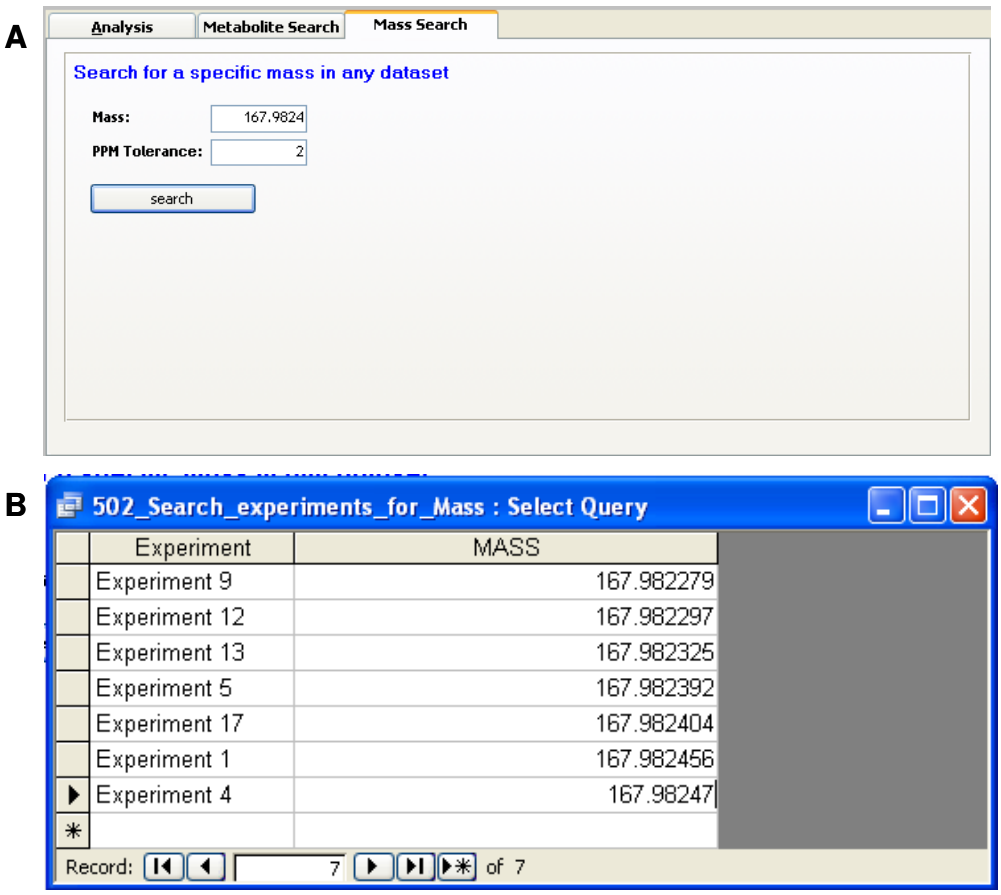


Figure 3-10. Searching for a specific mass.

A – Text boxes to enter a mass and PPM tolerance. B – A list of experiments which contain an experimental mass within the PPM tolerance to the mass chosen.

The ability to search all experimental datasets for a specific metabolite is incredibly useful. This feature is not known to have been included in any other metabolomic analysis software, thus providing metaSearch with a novel edge for metabolomic analysis.

3.7 Searching the KEGG Database

As previously mentioned, the KEGG online database is an extremely useful resource; however, the ability to search for metabolites based on specific criteria is relatively inefficient.

The Xcalibur software, supplied by Thermo Fisher Scientific, can only provide an experimental mass and molecular formula for an ionised compound. It is essential that the user can quickly determine which metabolite a particular mass or molecular formula may correspond to. Whilst the KEGG website search engine allows a user to search for metabolites using a metabolite name or KEGG compound ID, it does not allow the user to search for a mass or molecular formula. Therefore, it was necessary to provide the metabolomics community with a link between analytical software, such as Xcalibur, and a comprehensive metabolite database.

A 'Search' form (Figure 3-11) and several SQL statements were introduced providing a front-end to the database. The form contains several text boxes and command buttons to allow the user to retrieve information on metabolites when only the mass, molecular formula, partial metabolite name, or KEGG compound ID are known.

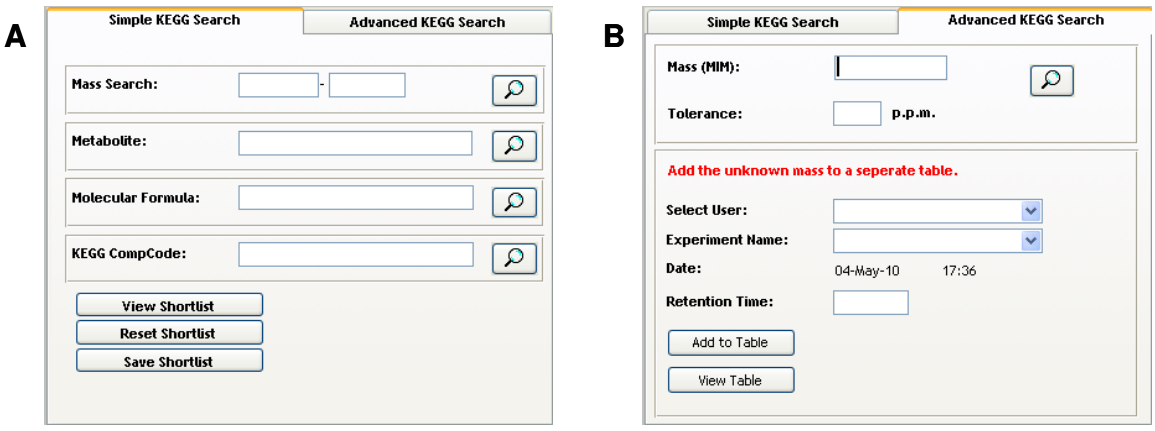


Figure 3-11. Searching the KEGG database.

The search options are displayed using two tabs. A – The 'Simple KEGG search' tab. The user can enter various terms in the appropriate text boxes to search for metabolites within given mass ranges, particular names, molecular formulae, or KEGG compound ID's. B – The 'Advanced KEGG Search' tab. The user can search the KEGG database for a specific mass within a chosen PPM tolerance.

3.7.1 Mass Searches

To provide the user with greater flexibility when searching for a specific mass, two separate SQL commands were implemented. The first SQL command (Supplementary Figure 8-16) utilises 'greater than' and 'less than' logic statements to display any metabolite(s) between two the mass values entered by the user (Figure 3-11A). The second SQL command (Supplementary Figure 8-17) is slightly more complex, displaying any metabolite(s) within a specific mass with a chosen PPM tolerance value (Figure 3-11B). This is useful when the user wishes to search a single experimental mass against all metabolites in the KEGG database. A similar query, described in section 3.6.4, which involves searching for a specific mass in experimental datasets, instead of the KEGG metabolite database. Once the user has entered appropriate values in the 'mass (MIM)' and 'PPM tolerance' text boxes, the command button will return any metabolite(s) within the PPM tolerance of the mass entered. If there is no mass match, the user has the option to provide the details of which experiment the mass came from and add the information to a separate table. Compiling a list of unidentified masses from a variety of experiments may allow identification of masses which recur in numerous experimental datasets. If more than one metabolite is returned in either mass search, then the metabolites are displayed in ascending order of the MIM.

3.7.2 Molecular Formula Search

Similar to searching for a mass, the ability to search for a specific molecular formula is also advantageous, particularly as the KEGG website does not facilitate this operation. The user must enter the complete molecular formula in the appropriate text box, with the respective command button activating a SQL statement (Supplementary Figure 8-18) displaying any metabolite(s) with that molecular formula.

3.7.3 Metabolite Name Search

Using the KEGG website it is possible to search all the synonyms for a partial metabolite name. As this is an appropriate method to search for a metabolite name, a SQL command was written that would perform the same function (Supplementary Figure 8-19). The user is required to enter a partial metabolite name in the appropriate text box on the search tab (Figure 3-11A). The

associated command button activates a query to retrieve that metabolite names from a table of synonyms, rather than the total KEGG metabolite table. The numbers of hits from a variety of text strings derived from the KEGG website and metaSearch synonym table were compared (Table 3-18).

Table 3-18. Comparing the number of metabolite hits from KEGG website and Microsoft Access searches for metabolite names. Unique hits are shown in brackets.

Search string	KEGG website	MetaSearch
'phospho'	351	531 (350)
'phosphoenolpyruvate'	1	1 (1)
'enolpyruvate'	3	3 (3)
'sphoenolpy'	1	2 (1)
'py*te'	0	220 (173)
'phosp*ate'	0	956 (613)

The metaSearch SQL statement returns the same number of unique hits as the KEGG website search for the examples tested. The only exceptions to this are the additional hit for the text string 'phospho', as the KEGG website is updated on a daily basis, and any text string that contains a wildcard symbol. Interestingly, the KEGG website search, unlike Microsoft Access, does not support the use of wild cards. The wild card symbol '*' can be used to replace any character, or sequence of characters, in a text string. This will generally result in a significant number of hits, although is useful approach if the user is unsure of the spelling for a specific metabolite.

The metaSearch approach is more efficient than the KEGG website search as wildcard symbols are accepted. Aside from this, the two approaches produce the same results.

3.7.4 KEGG Compound ID Search

It is possible to search for the KEGG compound ID (CompCode) on the KEGG website. The KEGG compound ID is one of two unique identifiers that can be used to distinguish between isomers, the other being the InChI string. Therefore, an SQL statement (Supplementary Figure 8-20) and command button were implemented in metaSearch so the user can obtain information on a specific compound if they know the KEGG compound code. This is the least useful search

as the principal role of the CompCode is to function as a primary key in relating information from one table to another.

3.7.5 Defining the shortlist

The shortlist option provides a link between searching the database for KEGG metabolites and analysing experimental data. As previously discussed, the user is required to select one of four parasitic organisms for analysis of an experimental dataset (section 3.6.3.3). This is an extremely powerful function providing the user with the option to assemble a metabolite table of interest, and compare these masses to a list of experimental masses (Figure 3-7).

When performing any search described in section 3.7, a column labelled ‘shortlist’ is displayed on the left hand side of the results. The shortlist attribute is a true/false value, and is displayed as a check box, allowing the user to simply toggle metabolites ‘true’ or ‘false’. For example, selecting various metabolites during searches for a specific metabolite name (Figure 3-12A), molecular formula (Figure 3-12B), or KEGG compound ID (Figure 3-12C) may result in a list of metabolites which need not be specific to any particular pathway or organism (Figure 3-12D). This list can be viewed using the ‘View Shortlist’ command button on the ‘search tab’ (Figure 3-11A). A SQL command will return any metabolites which have the value ‘true’ in the shortlist column. The shortlist can be cleared by updating all the values to ‘false’ using the ‘Reset Shortlist’ command button. Alternatively, the information can be saved to a separate table for future use by clicking the ‘Save Shortlist’ command button.

A

507_SYNONYM_KEGG_SEARCH : Select Query

Shortlist	Metabolite_Synonym	Exact_Mass_(MIM)	Formula	CompCode
☑	Trypanothione disulfide	721.288735	C27H47N9O10S2	C03170
☐	Oxidized trypanothione	721.288735	C27H47N9O10S2	C03170
☐	Reduced trypanothione	723.304385	C27H49N9O10S2	C02090
☑	Trypanothione	723.304385	C27H49N9O10S2	C02090
☐	Homotrypanothione	737.320035	C28H51N9O10S2	C16567
☐	Homotrypanothione disulfide	735.304385	C28H49N9O10S2	C16568
☐				

Record: 4 of 6

B

019_MF_KEGG_SEARCH : Select Query

Shortlist	Metabolite	Exact_Mass_(MIM)	Formula	CompCode
☑	L-Proline	115.063329	C5H9NO2	C00148
☐	D-Proline	115.063329	C5H9NO2	C00763
☐	Proline	115.063329	C5H9NO2	C16435
☐				

Record: 1 of 3

C

031_COMPCODE_KEGG_SEARCH : Select Query

Shortlist	Metabolite	Exact_Mass_(MIM)	Formula	CompCode
☑	L-Ornithine	132.089878	C5H12N2O2	C00077
☐				

Record: 1 of 1

D

047_View_Shortlist : Select Query

Shortlist	Metabolite	Exact_Mass_(MIM)	Formula	CompCode
☑	L-Ornithine	132.089878	C5H12N2O2	C00077
☑	L-Proline	115.063329	C5H9NO2	C00148
☑	Trypanothione	723.304385	C27H49N9O10S2	C02090
☑	Trypanothione disulfide	721.288735	C27H47N9O10S2	C03170
☐				

Record: 1 of 4

Figure 3-12. Compiling a shortlist.

The shortlist column contains a checkbox which can be toggled 'true' or 'false' depending on whether the metabolite is of interested to the user. A – Metabolite name search for 'trypanothione'. B – Molecular formula search for 'C5H9NO2'. C – KEGG compound ID search for 'C00077'. D – The shortlist of metabolites checked during the three searches.

3.8 The metaSearch Database

The metaSearch database can be found on the CD at the back of this thesis[21].

Please follow the instructions below;

1. When the database is opened, a 'Security Warning' will appear, click 'Open'.
2. View > Database Objects > Forms
3. Forms are listed in alphabetical order, double click on the 'Main Panel (Start Here)'.
4. A small example dataset 'Leishmania Dataset' is provided for analysis.

[21] Refer to CD: metaSearch 3.0.mdb

3.9 Pathos

Pathos, produced in collaboration with David Leader and Hani Ajahdali (Bioinformatics Research Centre, University of Glasgow), is a data-analysis tool from the Scottish Metabolomics Facility that allows a user to upload and analyse a processed data file from a metabolomic experiment (Leader, DP *et al*. 2011)[22]. Whilst the concept of this web facility is based on metaSearch, the ability to communicate with the KEGG metabolic maps provides a potential for improved analysis.

Once the user has uploaded a file for analysis an organism, ionisation mode, adduct status, and ppm tolerance must be chosen from the drop down menu. The processing of results is quicker than that of MassTrix, as the web facility does not directly interact with the KEGG website, instead using the table of metabolites from the metaSearch access database as a reference. The output is a summary of potential identifications divided into the broad areas of metabolism used by KEGG (Figure 3-13). The potential metabolite identifications can be viewed as a list of metabolites for each area, for example 'arginine and praline metabolism', or alternatively, through graphical representation of an annotated KEGG map. Furthermore, if a list of masses with paired intensity values from two conditions are supplied, the metabolites on the KEGG map will be annotated as blue or red if more intense in one condition, or as yellow if the intensities in the two conditions are roughly equal. Finally, if a specific organism has been selected, the enzymes identified as being present in that organism are also highlighted on the KEGG map. Enzyme annotation is determined by KEGG; not by the user or the Metabolomics Web Facility.

[22] http://motif.gla.ac.uk/Pathos/index.html

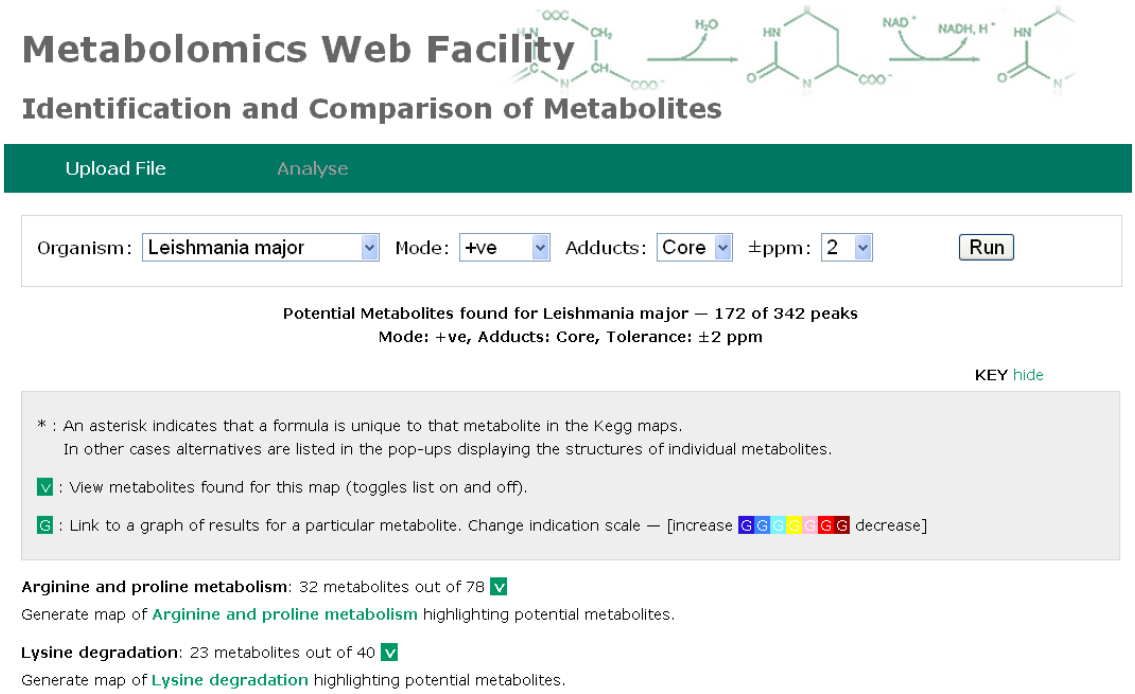


Figure 3-13. Metabolomics Web Facility - screenshot.

The parameters chosen by the user are displayed in the drop down menus at the top of the page. From the sample dataset, 172 of the 342 have a potential identification; 32 of these are related to 'arginine and proline metabolism', although this does include isomeric compounds. These can be viewed as a list by clicking the 'V' in the green box, or an annotated KEGG map for 'arginine and proline metabolism' can be generated.

The ability to obtain annotated KEGG maps based on the analysis of an experimental dataset is extremely useful, and a feature that metaSearch could not perform. Whilst MassTRIX is capable of providing a similar analysis, it cannot generate KEGG maps annotated to reflect changes in intensity between two conditions.

3.10 Discussion

Metabolomics is a rapidly developing field in systems biology. Significant improvements have been made in the area of high resolution mass spectrometry; the ability to analyse and interpret raw metabolomic datasets on a global scale has been largely neglected.

The metabolome can be defined as "the quantitative complement of low-molecular-weight metabolites present in a cell under a given set of physiological conditions" (Kell, DB *et al*. 2005). Consequently, the identification of compounds, and an understanding of how they relate to one another provides an insight on cellular function.

Metabolites are interconnected through biochemical reactions. In biochemistry textbooks, and online resources such as BioCyc (Caspi, R *et al*. 2010) or KEGG (Kanehisa, M and Goto, S 2000), the metabolites are typically grouped into classical biochemical pathways that are easy to understand. The global analysis of metabolomic datasets can be performed using MassTrix (Suhre, K and Schmitt-Kopplin, P 2008), which compares the experimental data with the KEGG metabolites. KEGG maps are also annotated during the analysis.

In a cell, metabolites are not conveniently grouped into pathways. Therefore, why should metabolomic analysis focus be restricted to comparing experimental data to a reference pathway? Convenience is probably the most compelling reason. The relationships between biochemical pathways can often be difficult to visualise, primarily due to some metabolites occurring in more than one pathway; thus the pathways can be rather rigid or inflexible. However, the information stored within databases such as KEGG remains the same, irrespective of how the information is displayed. Utilising this information it is possible to breakdown the non-existent physical barriers between biochemical pathways, instead concentrating on how these metabolites relate to one another in a global sense. This approach has been used a number of applications. MetaNetter allows the construction of *ab initio* metabolic networks based on biochemical transformations (Jourdan, F *et al*. 2008), whereas MetExplore includes all metabolic pathways, irrespective of organism, creating a metabolic network (Cottret, L *et al*. 2010).

A similar approach was used in metaSearch. Whilst metaSearch lacks visualisation of metabolomic networks, primarily due to computational limitations, an attempt has been made to remove the boundaries between pathways. The results from the analysis of experimental data are initially displayed as molecular formulae, with sub-datasheets providing information on the metabolites, and the reactions they are involved in (refer to Figure 3-8B). MetaSearch is a relational database produced using Microsoft Access, with metabolite information retrieved from the KEGG FTP site[23].

Microsoft Access is 'well suited to small, department applications' [24], and consequently was an ideal choice for this project. The single database file can operate on any computer providing Microsoft Access has been installed, and does not require internet access, unless the user wishes to utilise any hyperlinks provided in the tables. The processed metabolomic data can be stored in the database, or alternatively, kept as Microsoft Excel files and imported into the database when necessary. Furthermore, it is relatively straightforward to implement additional queries or macros, import and export data, and design forms to provide a user friendly GUI. Whilst the database described here is able to compare smaller datasets derived from PeakML, comparisons involving larger experimental datasets and larger lists of known metabolites require a large quantity of processing power to efficiently complete any analysis.

MetaSearch, like MassTRIX, uses mass matching to identify compounds where each mass in the dataset is compared to the MIMs of a selection of metabolites from the KEGG database within a chosen PPM tolerance. It is advantageous to match an experimental dataset to as large a compound or metabolite database as possible. However, this will generate many false positives in the context of a biological system. It is important to remember that mass matching does not identify a potential metabolite; it identifies a potential molecular formula. Therefore, mass matching alone cannot distinguish between isomeric compounds, although a likely identification can often be proposed when considering the organism in question. The final identification of a metabolite must include further identification steps such as retention time comparisons or a tandem mass spectrometry approach.

---

[23] http://www.kegg.jp/kegg/download/

[24] http://technet.microsoft.com/en-gb/library/cc917612.aspx

MetaSearch responds to these challenges in several ways. Firstly, the user can analyse data that has already been processed by PeakML; therefore, peaks should be reproducible and dataset should contain less noise. Secondly, the user may assign retention times for a series of standards to the KEGG metabolites. This allows a manual comparison of retention times between an experimental mass and possible compound. Finally, the user is able to view how many isomeric compounds are present in the KEGG database for each mass match, and which of these have been annotated as occurring in *T. brucei* (or any other organism of interest). It is important to recognise any biologically occurring isomers of the most likely identification, rather than overlook or ignore these possible alternatives.

If a targeted metabolomic analysis is required, it is possible to search against a specific metabolic pathway, or any list of metabolites defined by the user. Even using a targeted approach it is essential that the user knows of any isomeric compounds that are present in a different pathway of the same organism, as well as any naturally occurring metabolite.

The diversity of the cell lines investigated, and the diversity of experimental techniques utilised in performing metabolomic experiments, creates diverse sets of metabolites. A single dataset is not diverse. Consequently, it is advantageous to have the ability to ignore the boundaries between datasets, and take a holistic view of the data. As multiple metabolomic datasets can be stored in metaSearch, it is possible to search all the experimental datasets for a KEGG metabolite by partial name, or search for any experimental mass within a given PPM tolerance. This is not possible using MassTRIX as each experimental dataset is uploaded and analysed individually, rather than stored in a specific location, and analysed when necessary. These particular approaches are extremely important, as the user can rapidly search for a metabolite or mass of interest, without confining the search to a particular dataset. Whilst these queries are extremely useful, the necessity to store all experimental data in a single file and the processing power required to perform the queries, make this an unrealistic approach using Microsoft Access. However, if the data were to be stored on a web server, then the processing power would not be an issue.

A limitation of the KEGG website search engine was the inability to search for a specific mass or molecular formula, both of which are important in metabolomic analysis. This issue was addressed in metaSearch with queries to return any metabolites within a given PPM tolerance of a mass, return any metabolites between two masses, or return any metabolites of a specific molecular formula. Searching for a metabolite by a name or partial name is possible on the KEGG website; this feature is retained in metaSearch.

Interpreting a metabolomic dataset using MassTRIX will result in the identification of the metabolites present in that dataset in relation to a specific organism, and map these as differentially coloured objects on KEGG pathways. Whilst it is easy to navigate between the KEGG metabolic pathways, they are viewed as distinct entities which, conceptually, is a reductionist approach not suitable for global metabolomic analysis. This can be demonstrated when considering the fate of glucose 6-phosphate, which, depending on cellular requirements can supply both the glycolytic pathway and the PPP. Additionally, F6P and Ga3P, produced by the oxidative branch of the PPP, can supply the glycolytic pathway. Considering the dynamic relationship between these pathways it would be confusing to interpret differential data on two separate KEGG maps. Therefore, it would be advantageous to adopt a holistic approach and view both pathways as a single entity. This is a limitation of KEGG, rather than MassTRIX, which ultimately is capable of annotating any KEGG map.

The From Metabolite to Metabolite (FMM) server is capable of investigating which biochemical reactions are required to convert one metabolite to another (Chou, CH *et al*. 2009). Whilst this is limited to the KEGG reaction data, the approach does consider that the two metabolites may be present in different KEGG maps. Although this particular approach is not particularly useful in the analysis of experimental data, the concept of linking metabolites on different KEGG maps is convenient.

Unfortunately, the KEGG metabolic networks and biochemical reaction data make no differentiation between mammalian and insect stages of the parasites. This is an important consideration due to the changes in nutrient availability and gene expression during the parasite lifecycle; consequently, there will be changes to the metabolome and proteome. The classical biochemical

approaches, typical of the pre-genomics era, have provided the trypanosomatid community with a plethora of detailed biochemical and enzymatic data for various strains and life cycle stages. The trypanosomatid community should make a greater effort to collate and integrate this data with bioinformatic approaches such as KEGG. An accurate representation of the enzymes expressed and metabolites present during various life cycle stages on interactive metabolic pathways would provide a true biological context that is both robust and meaningful to any experimental 'omic' data.

A number of methods for analysing metabolomic data exist. The introduction of approaches such as MetaNetter, MetExplore, and the database described in this chapter demonstrates that looking at metabolism in a global sense can provide alternative information to simply restricting the analysis to a particular pathway or organism. These complementary techniques will further enhance our understanding of global metabolism.

4 Genetic manipulation of *L. mexicana* transketolase

4.1 The Pentose Phosphate Pathway

The pentose phosphate pathway (PPP) provides an alternative route to glycolysis for the oxidation of glucose (Barrett, MP 1997). The main functions of the PPP are the generation of ribose 5-phosphate (R5P; C00117), essential for nucleotide biosynthesis, and the regeneration of NADPH, important in the defence against oxidative stress and in producing reducing equivalents for anabolic processes such as lipid biosynthesis (Barrett, MP 1997). Futhermore, the pathway produces several phosphorylated sugars which can feed back into the glycolytic pathway, in the case of fructose 6-phosphate (F6P; C00085) and glyceraldehyde 3-phosphate (GA3P; C00118), or alternatively are used as precursors in biosynthetic reactions, in the case of erythrose 4-phosphate (E4P; C00279) which is used in the biosynthesis of aromatic amino acids in some organisms. The pathway can be divided into two sections; the oxidative section, and the non-oxidative section.

4.1.1 Oxidative Branch

The oxidative section consists of three reactions converting one mole of glucose-6-phosphate (G6P; C01172) to one mole of D-ribulose 5-phosphate (Ru5P; C00199), 2 moles of NADPH (C00005), and one mole of CO_2. NADPH (C00005) is a hydride donor, and is used in many biosynthetic reactions, or in the defence against oxidative stress. The fate of G6P is governed by the cellular requirements for ATP, R5P and GA3P. NADPH inhibits glucose-6-phosphate dehydrogenase (G6PDH; EC 1.1.1.49), thus inhibiting the PPP. The flux through the PPP increases when NADPH is oxidised. The oxidative section of the pathway is often regarded as irreversible under normal physiological conditions. Ru5P is the end product of the oxidative section of the pathway, and the precursor for the non-oxidative section.

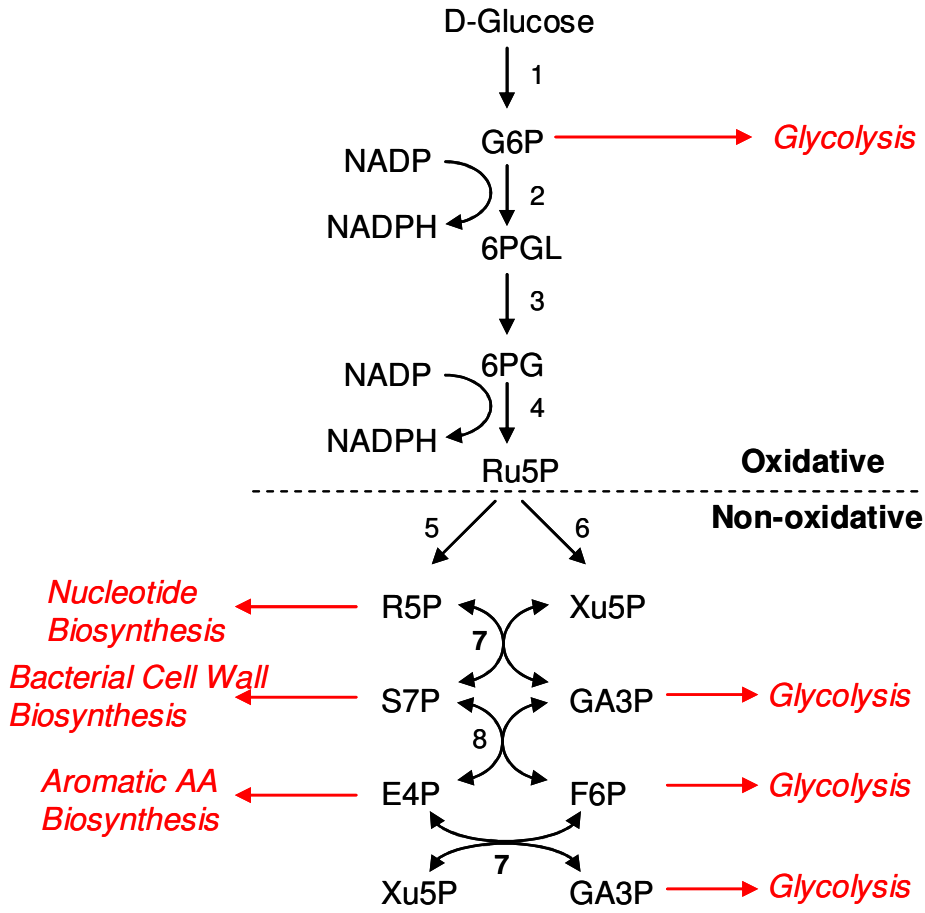


Figure 4-1. The Pentose Phosphate Pathway.

The pathway is divided into an oxidative branch and a non-oxidative branch. Metabolites: - G6P, glucose 6-phosphate; 6PGL, 6-phosphogluconolactone; 6PG, 6-phosphogluconate; Ru5P, ribulose 5-phosphate; R5P, ribose 5-phosphate; Xu5P, xylulose 5-phosphate; S7P, sedoheptulose 7-phosphate; GA3P, glyceraldehyde 3-phosphate; E4P, erythrose 4-phosphate; F6P, fructose 6-phosphate. Enzymes: - 1, hexokinase; 2, glucose-6-phosphate dehydrogenase; 3, 6-phosphogluconolactonase; 4, 6-phosphogluconate-dehydrogenase; 5, pentosephosphate isomerase; 6, ribulose-5-phosphate epimerase; 7, transketolase; 8, transaldolase. Pathways linked to the PPP are shown in red.

4.1.2 Non-oxidative Branch

The non-oxidative section of the PPP is responsible for generating a variety of sugar phosphates. Both F6P and GA3P can feed back into the glycolytic pathway. Also generated are sedoheptulose 7-phosphate (S7P; C05382), a component of some bacterial cell walls, and E4P, a precursor for many vitamins and aromatic amino acid biosynthesis in some organisms.

4.1.3 PPP activity in Trypanosomatids

All the enzymes of the PPP have been identified in *T. brucei* procyclic form cells (Cronin, CN *et al*. 1989) *L. mexicana* promastigotes (Maugeri, DA *et al*. 2003), although transketolase (EC 2.2.1.1) and ribulose 5-phosphate epimerase (EC 5.1.3.1) were absent in the bloodstream form trypanosome (Cronin, CN *et al*. 1989). The quantity of glucose available to the bloodstream form trypanosome may indicate non-oxidative branch of the PPP is not required; however the presence of transaldolase (EC 2.2.1.2) may signify the involvement of additional enzymes, which are not part of the classical pathway. In *T. cruzi*, all of the classical enzymes were identified from all four developmental stages, and that glucose consumption through the PPP increased by 2-fold (Maugeri, DA and Cazzulo, JJ 2004). Specific activities for each of the *L. mexicana* PPP enzymes (Maugeri, DA *et al*. 2003) and *T. cruzi* PPP enzymes (Maugeri, DA and Cazzulo, JJ 2004) were also determined. The *L. mexicana* promastigote enzymes of the oxidative branch were found to have greater activity than those from *T. brucei* procyclic form; a possible hypothesis was the greater demand for NADPH required when the promastigotes infect macrophages (Maugeri, DA *et al*. 2003).

4.1.4 Roles in Biology

The primary functions of the pentose phosphate pathway are the regeneration of NADPH, production of ribose 5-phosphate for nucleotide biosynthesis, and the production of erythrose 4-phosphate for aromatic amino acid biosynthesis.

4.1.4.1 NADPH regeneration

The oxidative section of the PPP is responsible for the regeneration of NADPH, a co-factor used in many enzymatic reactions including the maintenance of reduced thiols and fatty acid biosynthesis. Glutathione (GSH; C00051) is the major thiol produced in cells and plays an important role in defending cells against oxidative stress by the detoxification of reactive oxygen species and free radicals. However, this process requires glutathione to donate an electron, causing glutathione to become reactive. Two reactive gluathione molecules can react with one another forming oxidised glutathione (GSSG; C00127), which is then converted back to GSH by gluathione reductase, an enzyme requiring NADPH as a co-factor.

Whilst glutathione is responsible for the neutralisation of free radicals and reactive oxygen species in the majority of cell types, trypanosomatids possess a unique thiol named trypanothione to perform this role. Trypanothione, an amphiprotic compound, is an unusual derived form of glutathione consisting of two molecules of GSH that are linked by one molecule of spermidine (C00315) (Fairlamb, AH *et al.* 1985). In a similar fashion to glutathione, trypanothione can exist in an oxidised (T(SH)2, C03170) or reduced (T(S)2, C02090) state. Trypanothione-disulfide reductase (TryR, EC 1.8.1.12) is specific to trypanosomatids, essential for parasite survival, and as a consequence is considered as a possible target for drug development. The enzyme requires FAD as a cofactor, and catalyses the following reaction; Trypanothione + $NADP^+$ <=> Trypanothione disulfide + NADPH + H^+. The NADPH regenerated by the oxidative branch of the PPP is essential to maintain trypanothione in the reduced form, thereby protecting the cells from oxidative stress and free radicals.

4.1.4.2 Nucleotide Biosynthesis

Ribose 5-phosphate (C00117) is necessary for the synthesis of ribonucleotides which are subsequently incorporated into nucleic acids. There are numerous methods for the synthesis of R5P; from the Ru5P generated by the oxidative section using pentosephosphate isomerase, the non-oxidative section from GA3P and S7P acting as the ketose donor, or finally by the phosphorylation of free ribose (C00121) by ribokinase (EC: 2.7.1.15).

4.2 Transketolase

The phosphorylated sugars present in the non-oxidative branch of the PPP are substrates for a number of metabolic pathways (Figure 4-1), and as a result transketolase (TKT; EC 2.2.1.1) and transaldolase (TAL; EC 2.2.1.2) are responsible for the interconversion of these phosphorylated sugars depending on cellular requirements.

Transketolase is classically believed to catalyse two reversible reactions in the non-oxidative branch of the PPP (Figure 4-1) with the transfer of a 2-carbon glycoaldehyde subunit from a ketose donor, to an aldose acceptor. Typical biochemical reactions performed by transketolase include the transfer of a 2-carbon subunit from Xu5P to either R5P or E4P, resulting in the production of

GA3P and S7P or F6P respectively. However, some tkt enzymes utilise a broad range of substrates; both phosphorylated and non-phosphorylated sugars. This has led to tkt having industrial and commercial value, for example, in the synthesis of aromatic amino acids.

Transketolase has been identified in a wide range of organisms, including trypanosomatids (Cronin, CN *et al*. 1989; Maugeri, DA and Cazzulo, JJ 2004; Veitch, NJ *et al*. 2004). Whilst activity has been identified in *T.brucei* procyclics (Stoffel, SA *et al*. 2011) and *L. mexicana* promastigotes (Veitch, NJ *et al*. 2004), the enzyme was found to be absent in both *T.brucei* bloodstream form (Cronin, CN *et al*. 1989). It was inconclusive as to whether tkt was present in *L. mexicana* amastigotes due to a lack of controls, although the enzyme was present in *L. mexicana* axenic amastigotes (Veitch NJ, PhD Thesis). However, the enzyme has been identified in all 4 major development stages of *T. cruzi* (Maugeri, DA and Cazzulo, JJ 2004).

4.2.1 Structure

The crystal structures of transketolase have been determined for several species; including *Saccharomyces cerevisiae* (Lindqvist, Y *et al*. 1992), *Homo sapiens* (Mitschke, L *et al*. 2010), and *L. mexicana* (Veitch, NJ *et al*. 2004). The *L. mexicana* tkt protein was shown to be structurally similar to *S. cerevisiae* tkt, with an amino acid sequence identity of 50.4% (Veitch, NJ *et al*. 2004).

Transketolase is functional as a dimer, requiring one calcium ion and one thiamine pyrophosphate (TPP; C00068) molecule per subunit. TPP is a co-factor in many enzymatic reactions, and is required by transketolase to catalyse the reaction. The three-dimensional structure of *S. cerevisiae* tkt identified several residues thought to be important in the binding of the TPP, and also in the binding and stabilisation of the substrate (Lindqvist, Y *et al*. 1992; Nilsson, U *et al*. 1997). The interactions between the TPP and *L. mexicana* transketolase enzyme are illustrated in Figure 4-2.

The diphosphate group of the TPP molecule interacts with the Ca^{2+} ion through two oxygen atoms of the diphosphate group, whereas the side chains of Asp155 and Asn185, and the oxygen atom present in the main chain of Ile187 interact with the Ca^{2+} ion. Additionally, hydrogen bonding occurs between the diphosphate group and several conserved residues including His66, His261, Asp183 and Gly156. The remaining portion of the TPP interacts with several amino acids from both residues, and as a result is bound within the two subunits. The pyrimidine ring is located within a pocket of conserved residues from the second subunit. Several conserved residues from the second subunit (denoted by an *) that include Phe432*, Phe435*, and Tyr438* form a pocket in which the pyrimidine ring binds through hydrophobic interactions. Furthermore, hydrogen bonding occurs between the pyrimidine ring and Gly114, His472*, Leu116, and Glu409*. The C4 methyl group of the thiazolium ring forms hydrophobic interactions with the side chains of Val114* and Leu377*, both of which are from the second subunit. The negative charge present on the Asp376* is thought to compensate for the positive charge of the thiazolium ring (Lindqvist, Y *et al*. 1992). Consequently, the only atom of the TPP molecule that is accessible from solution is the carbon atom at the C-2 position of thiazolium ring.

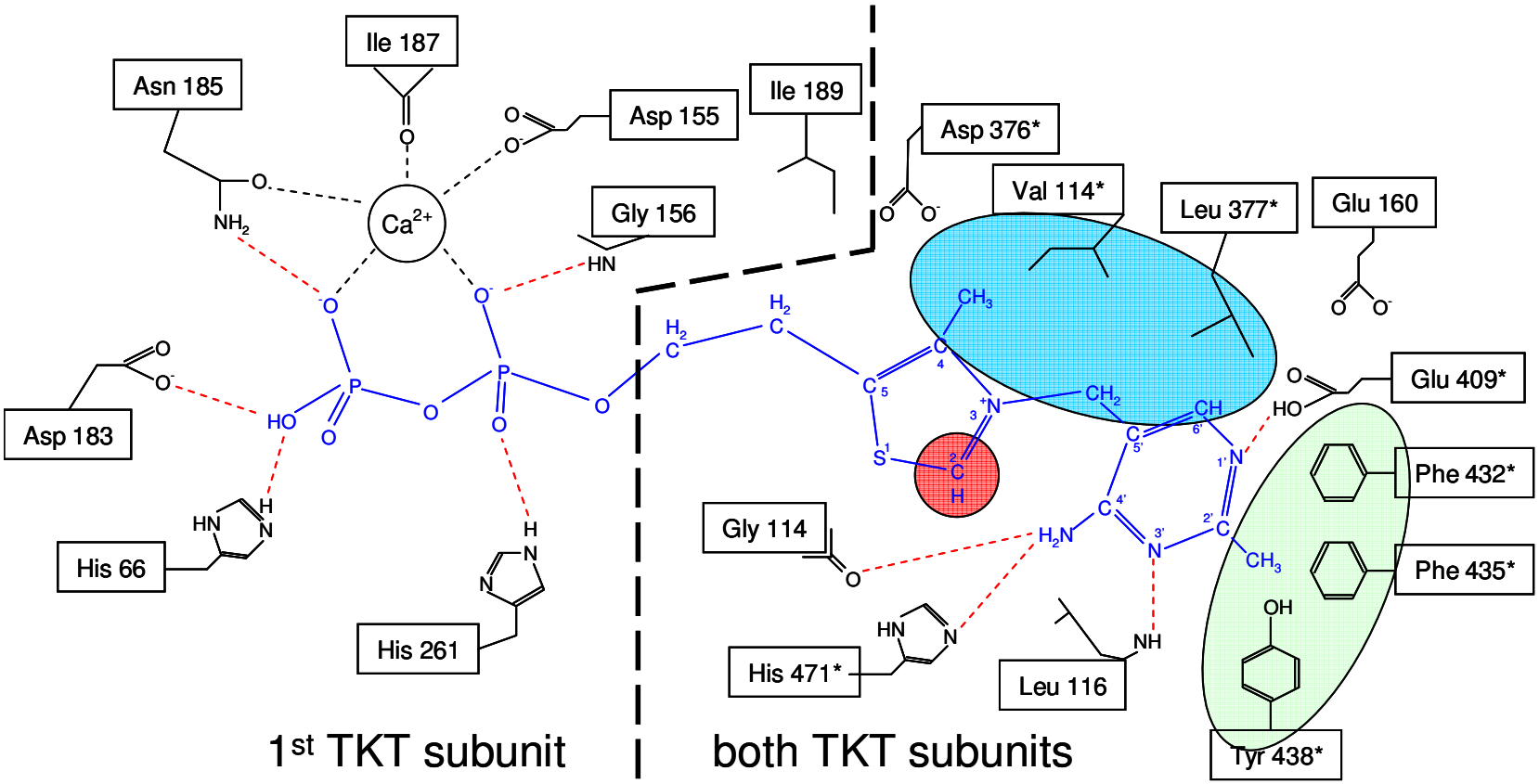


Figure 4-2. Schematic diagram of the interactions of TPP at the cofactor binding site of *L. mexicana* transketolase.

The TPP molecule is indicated in blue lines and blue text, whereas the amino acid residues are black lines and black text. Asterisk, amino acid residues from the second sub-unit; dashed red lines, possible hydrogen bonds; dashed black lines, ligands between the protein and calcium ion; green circle, the hydrophobic pocked binding the pyramidine ring; blue circle, hydrophobic interactions between the C-4 methyl group of the thaizolium ring; red circle, C-2 atom of the thaizolium ring. Reproduced from Lindqvist, Y *et al.* 1992.

A deep cleft between the two subunits allows transketolase to interact with a broad range of keto and aldo sugars. An example involving E4P is shown in Figure 4-3. The aldehyde oxygen atom of the acceptor substrate (or the ketol oxygen atom of the ketose donor) can form hydrogen bonds with the side chains of His26* and His 261*, and the C-2 hydroxyl group interacts with the side chain of Asp467. These interactions not only stabilise the substrate, but also bring the C-1 carbon of the substrate within 4 Å of the C-2 carbon atom of the thiazolium ring of TPP or the α,β-dihydroxyethyl TPP complex (C13378) (Lindqvist, Y *et al*. 1992; Nilsson, U *et al*. 1997). If the donor or acceptor substrate is phosphorylated, then the oxygen atoms of the phosphate group form hydrogen bonds with His459, Arg518, and Ser380. An additional interaction between the side chain of Arg353 and the phosphate group forms a salt bridge (Nilsson, U *et al*. 1997). A non-phosphorylated substrate does not bind as tightly to transketolase as a phosphorylated substrate; subsequently, the efficiency of the reaction decreases when substrates are not phosphorylated (Nilsson, U *et al*. 1997).

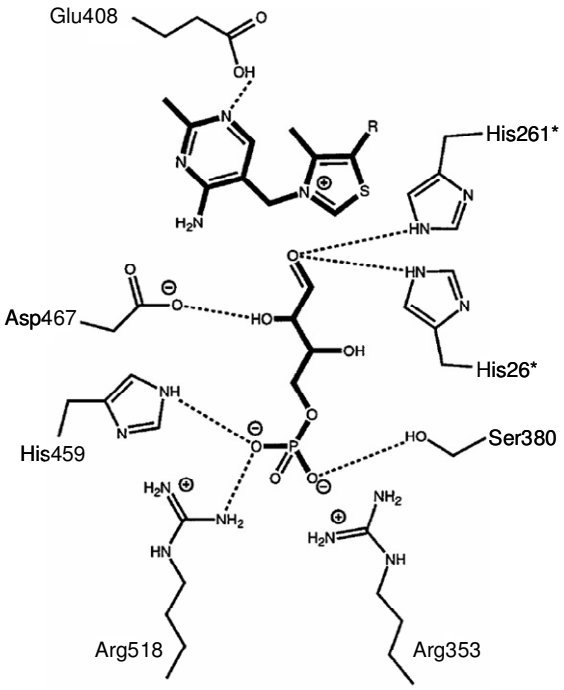


Figure 4-3. Schematic view of the interactions of erythrose 4-phosphate with *L. mexicana* holo-transketolase.

Possible hydrogen bonds are indicated by dashed lines. R = -CH2-CH2-P2O7H3. Asterisks denote residues from the second subunit. Adapted from Nilsson, U *et al.* 1997.

Activity of the *S. cerevisiae* transketolase enzyme could be impaired by targeting these conserved residues by site-directed mutagenesis, thus confirming their importance in enzyme activity (Nilsson, U *et al*. 1997). Unsurprisingly, the majority of residues involved in the binding of TPP and the substrate are conserved between transketolase enzymes isolated from different species (Veitch, NJ *et al*. 2004). The *L. mexicana* transketolase enzyme has 671 amino acid residues per subunit, with a molecular weight of approximately 74 kDa per monomer (Veitch, NJ *et al*. 2004).

4.2.2 Reaction mechanism

The biochemical reaction performed by transketolase follows a ping-pong mechanism in which a 2-carbon glycoaldehyde subunit is transferred from a ketose donor to an aldose acceptor (Figure 4-4). The ketose donor substrate, for example F6P, is bound within a deep cleft between the two transketolase subunits (Section 4.2.1). Catalysis is initiated by the deprotonation of the C-2 carbon atom of the thiazolium ring of TPP. This is followed by the nucleophilic attack of the carbonyl carbon atom of the F6P, resulting in the cleavage of the bond between the C-2 and C-3 carbon atoms. The ketol group remains bound to the C-2 carbon atom of the thiazolium ring, forming α,ß-dihydroxyethyl thiamine diphosphate (C13378), whilst the resulting aldose, in this case E4P, is released. GA3P then binds within the active site, reacts with α,ß-dihydroxyethyl thiamine diphosphate and forms Xu5P.

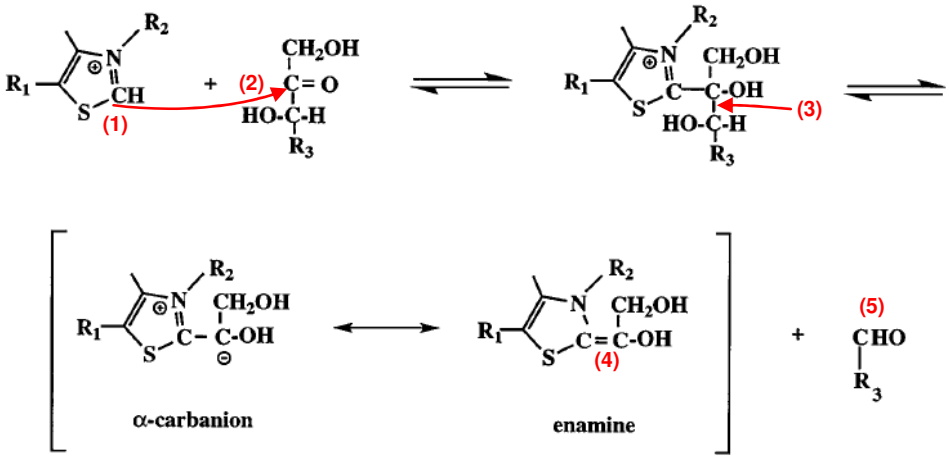


Figure 4-4. Transketolase reaction mechanism.

First part of the ping-pong reaction mechanism of transketolase. (1) Deprotonation of the C-2 carbon atom of the thiazolium ring; (2), the resulting carbanion attacks the carbonyl carbon of the ketose donor substrate; (3), the carbon-carbon bond between the C-2 and C-3 carbon atoms of the ketose is cleaved; (4), the ketol group remains covantently bound to the C-2 carbon atom of the thiazolium ring; (5), the aldose is released from the enzyme. In the second part of the reaction an aldosugar acceptor binds in the substrate cleft, and the order of chemical steps is then reversed. Adapted from Nilsson U et al 1997.

4.2.3 Localisation

In *Leishmania* all of the PPP enzymes currently annotated, with the exception of transaldolase, have a perixisomal targeting signal (PTS); these are divided into C-terminal PTS-1 or N-terminally located PTS-2 glycosomal targeting signal (Opperdoes, FR and Szikora, JP 2006). However, some proteins do lack a PTS1 or PTS2 targeting signal and may enter the glycosome by 'piggybacking' on proteins that does posses a PTS targeting signal (Opperdoes, FR and Szikora, JP 2006; Galland, N *et al*. 2007). There are two ORFs which code for ribulose 5-phosphate 3-epimerase (Ru5PE, EC: 5.1.3.1), one with a glycosomal targeting signal, and one without. In *L. mexicana* promastigotes, transketolase is localised to both the glycosome and the cytosol; although the majority of the enzyme is located to the cytosol (Veitch, NJ *et al*. 2004). In *T.brucei*, transketolase was also shown to have a PTS1 targeting signal. Interestingly, phosphoglucose isomerise (PGI; EC 5.3.1.9) is 90% cytosolic in *L. mexicana*, whereas it is almost exclusively glycosomal in *T.brucei*. PGI is required for the conversion of F6P to G6P for the regeneration of NADPH via the oxidative branch of the PPP. *Leishmania* spp have high levels of cytosolic trypanothione, which is maintained in the reduced form (C02090) by NADPH and TryR. The cytosolic localisation of PGI indicates that a substantial percentage of the oxidative section of the PPP occurs in the cytosol in order to reduce the cytosolic $NADP^+$ (C00006) to maintain the high levels of NADPH. In *T. cruzi* the PPP enzymes were identified as being cytosolic, however it is possible that all are present in the glycosome at low levels (Maugeri, DA and Cazzulo, JJ 2004). The evidence suggests that in trypanosomatids the PPP occurs in both the cytosol and the glycosome.

4.3 Aim

Preliminary data indicated that transketolase knockouts in procyclic form trypanosomes were found to be hypersensitive to pentamidine and compounds that induced oxidative stress (Stoffel, SA *et al*. 2011).

To determine the biological role of transketolase (LmjF24.2060, EC 2.2.1.1) in *L. mexicana* promastigotes the gene was deleted by targeted gene disruption, and the phenotype of the transketolase-null mutants was investigated.

4.4 Targetted Gene Disruption of the L. mexicana TKT gene

4.4.1 Constructs

In order to target both tkt alleles in *L. mexicana*, plasmids were generated containing either a hygromycin B phosphotransferase (*HYG*) gene or a streptothricin acetyltransferase (*SAT*) gene, flanked by DHFR-TS untranslated regions for constitutive expression of the drug resistance markers. This region was flanked with DNA sequences immediately upstream and downstream of the tkt ORF. These constructs are derivatives of those used for the deletion of the *L. mexicana* cysteine peptidase B (*lmcpb*) gene array (Mottram, JC *et al*. 1996).

To generate knockout construct pMB-G40, a 396-bp region immediately upstream (5') of the tkt gene was amplified using primers MB79 and MB80. The amplified product was digested with HindIII/SalI, and cloned into the pGL520 vector (Mottram Laboratory, University of Glasgow). A 600-bp region downstream (3') of the tkt gene was amplified using MB217 and MB218. The amplified product was digested with SmaI/BglII, and cloned into the knockout vector to give pMB-G40 (Figure 4-5). The knockout vector was sequenced by Eurofins MWG Operon (Ebersberg, Germany) to ensure that the ligation of both the 5' *TKT* and 3' *TKT* flanks had occurred correctly.

Leishmania are diploid organisms and therefore a second round transfection using a different antibiotic and selection marker is required to delete both alleles. The streptothricin acetyltransferase gene from pMB-G40 was replaced with the hygromycin B phosphotransferase from pGL345 (Mottram Laboratory, University of Glasgow) following a BamHI/SpeI digestion to give pMB-G39. The HYG gene was sequenced by Eurofins MWG Operon (Ebersberg, Germany) to ensure that ligation had occurred correctly.

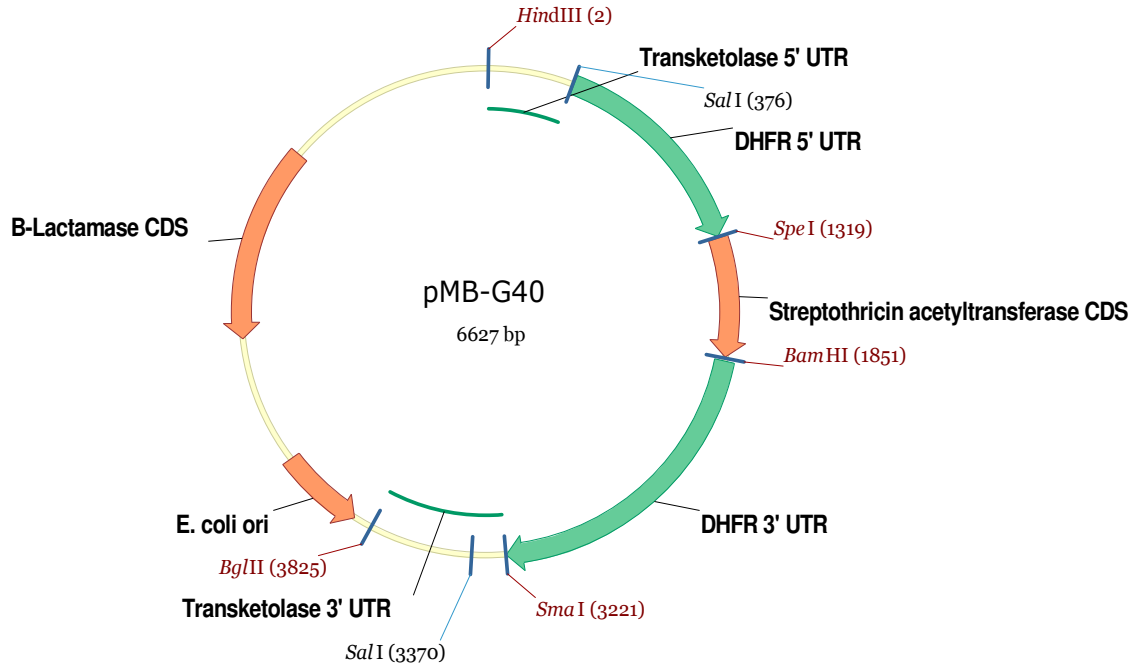


Figure 4-5. pMB-G40 – Transketolase Knockout Vector.

The features are as follows; horizontal green line, *TKT* flanks; solid green arrow, DHFR flanks; solid orange arrows, ORFs for streptothricin acetyltransferase and β-lactamase. The major restriction sites used for cloning are also indicated; unique restriction sites are shown in red, and the numbers in parenthesis indicate the position on the plasmid.

4.4.2 First round transfection

The TKT knockout cassette conferring resistance to hygromycin B (Figure 4-6) was released from the vector by restriction digest with HindIII/BglII, and upon transfection (section 2.6.2) of this DNA, homologous recombination of the knockout cassette should occur at the tkt locus.

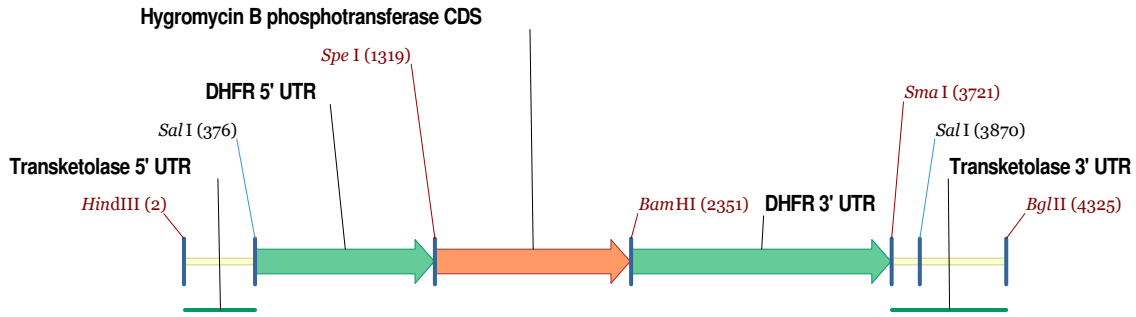


Figure 4-6. Transketolase knockout cassette with the Hygromycin B phosphotransferase drug marker.

The features are as follows; horizontal green line, TKT flanks; solid green arrow, DHFR flanks; solid orange arrow, hygromycin B phosphotransferase CDS. The major restriction sites used for cloning are also indicated; unique restriction sites are shown in red.

Limiting dilutions in 96-well plates were used to clone the cells. Poisson distribution states that if the population in each well was derived from a single cell, then 37% of the wells would have no growth. The average number of cells per well can be calculated using the following equation; m = -LN(F_0), where 'm' is the average number of cells per well and 'F_0' is the frequency of wells receiving no cells.

After the first round transfection the percentage of negative wells in the first two plates for both populations was less than 37%, indicating that the cells may not be clonal. However, more than 37% of the wells in the 3rd plate for both populations were negative, indicating that the populations in the positive wells were likely to have been derived from a single cell. The average numbers of cells in the 3rd plate were calculated to be 0.065 and 0.076 per well for populations A and B respectively. These results are summarised in Table 4-1.

Table 4-1. Analysis of the first round transfection.

	% negative wells		F_0		m	
	Pop. A	**Pop. B**	**Pop. A**	**Pop. B**	**Pop. A**	**Pop. B**
Plate 1 (6-fold)	0	0	0.00	0.00	n.c	n.c
Plate 2 (72-fold)	29	24	0.29	0.24	1.23	1.43
Plate 3 (864-fold)	94	93	0.94	0.93	0.06	0.08

The percentage of negative wells, the frequency of wells receiving no cells (F_0), and the calculated average number of cells per well (m) are shown for the three dilutions of both populations of potential single allele knockout cells. n.c. = not calculated.

A total of 13 clones were recovered from both plates. To ascertain whether one allele for the tkt gene had been deleted from any of these clones, a series of PCR reactions were performed.

4.4.3 PCR Screen

The genomic DNA isolated from the thirteen potential single allele knockout clones was screened by PCR for the internal transcribed spacer 1 (ITS1) region to confirm the presence of *Leishmania* gDNA (Schonian, G *et al*. 2003). One clone was negative for this marker and was therefore excluded from any further screens. The remaining twelve clones were screened for the presence of the *HYG* gene, episomes, and successful integration at the *TKT* locus. These findings are displayed in Figure 4-7, and summarised in Table 4-2.

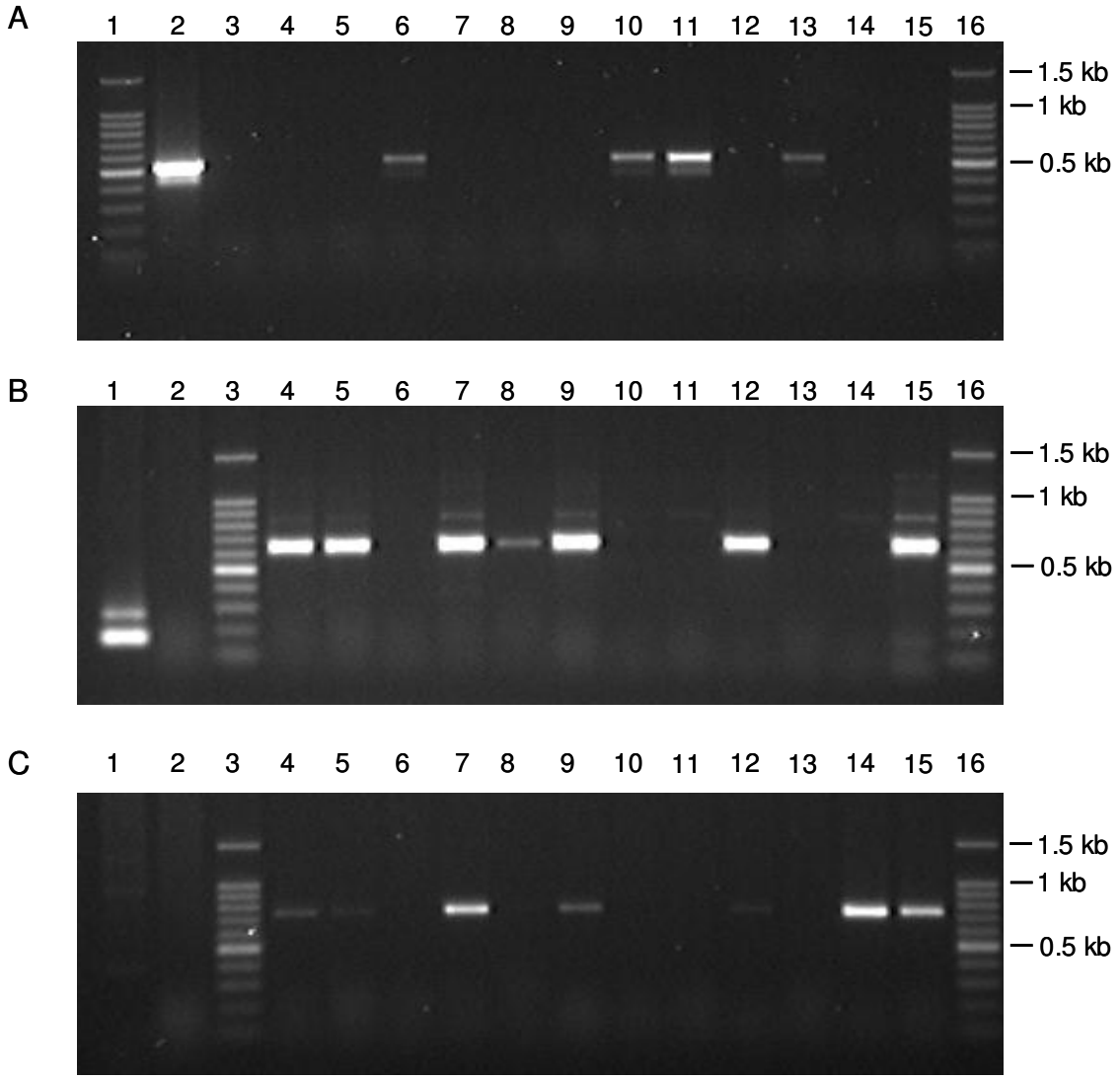


Figure 4-7. PCR analysis of *L. mexicana* single allele transketolase knockout clones.

A - Screen for episomal copies using Sp6 and MB212 to generate an expected product of 535 bp; B - Screen for 5' integration using MB91 and MB212 to generate an expected product of 651 bp; C - Screen for 3' integration using MB94 and MB213 to generate an expected product of 749 bp. The lanes on each of the images are as follows, A; 1, 100 bp ladder; 2, plasmid DNA; 3, negative control; 4, clone A1; 5, clone A2; 6, clone A3; 7, clone A4; 8, clone A6; 9, clone B1; 10, clone B2; 11, clone B3; 12, clone B4; 13, clone B5; 14, clone B6; 15, clone B7; 16, 100 bp ladder. B and C; 1, WT DNA; 2, negative control; 3, 100 bp ladder; 4, clone A1 ;5, clone A2; 6, clone A3; 7, clone A4; 8, clone A6; 9, clone B1; 10, clone B2; 11, clone B3; 12, clone B4; 13, clone B5; 14, clone B6; 15, clone B7; 16, 100 bp ladder. The 100 bp ladder is supplied by Promega.

Presence of the *HYG* gene was detected using MB39 and MB40, and thus any sample positive for this marker would indicate a successful electroporation event. All clones tested were found positive for this marker.

The restriction digest of the plasmid DNA to release the knockout cassette is not always successful with uncut plasmid contaminating the sample. This can be reduced by fractionating the restriction digest on an agarose gel, and extracting the DNA from the appropriate fragment using a commercial kit. However, uncut plasmid may remain at low levels, with *Leishmania* parasites retaining the ability to express the antibiotic selection marker from the plasmid. Therefore, it is extremely important that cell lines are tested for presence of episomes using oligonucleotides Sp6 and MB212. These primers amplify across the HindIII restriction site; no PCR product will be generated if the plasmid has been digested efficiently. Clones which are positive for an episomal copy of the knockout cassette can be disregarded from future analysis.

Similarly the clones were tested for integration of knockout cassette at both the 5' and 3' flanks. Primers were selected that would anneal outside the recombination sites; therefore any product generated would be specific to a successful integration. Oligonucleotides MB91 and MB212 would generate a PCR product of approximately 650 bp if the cassette had integrated correctly at the 5' flank. Similarly, oligonucleotides MB94 and Mb213 would generate a PCR product of approximately 750 bp if the cassette had integrated correctly at the 3' flank. It is evident that six clones were positive for integration at the 5' and 3' recombination sites (Table 4-2).

Table 4-2. Summary of PCR analysis for potential TKT single allele gene knockouts

PCR Product Olignucleotides	ITS 1 MB95 MB96	*HYG* MB39 MB40	episome Sp6 MB212	5' integration MB91 MB212	3' integration MB94 MB213
Clone A1	●	●	-	●	●
Clone A2	●	●	●	-	-
Clone A3	●	●	-	●	●
Clone A4	●	●	-	●	●
Clone A5	-	nd	nd	nd	nd
Clone A6	●	●	-	●	-
Clone B1	●	●	-	●	●
Clone B2	●	●	●	-	-
Clone B3	●	●	●	-	-
Clone B4	●	●	-	●	●
Clone B5	●	●	●	-	-
Clone B6	●	●	-	-	●
Clone B7	●	●	-	●	●

Legend; ●, positive; -, negative; nd, not done.

In summary, the PCR analysis indicated that of the twelve clones analysed, all were positive for the hyg gene, four were episomal, one had apparently integrated at the 5' flank but not at the 3' flank, whereas another had integrated at the 3' flank but not at the 5' flank, and six were identified as being single allele knockout cell lines. Clone B1 was selected for a second round transfection to delete the second allele. It is imperative that the single allele knockout clones are analysed thoroughly prior to transfection with the second knockout cassette to ensure that the correct clone is selected.

4.4.4 Second round transfection

The single allele knockout cell line (clone B1) was transfected with the tkt knockout construct conferring resistance to nourseothricin in an attempt to disrupt the second transketolase allele.

A total of seven potential double knockout clones (5 from population A, and 2 from population B) were recovered from the second round transfection (summarised in Table 4-3), although one of these did not survive long enough to allow genomic DNA to be extracted. These were screened for the presence of the ITS1 repeat, the *SAT* resistance gene, and most importantly, the transketolase gene.

Table 4-3. Analysis of the first round transfection.

	Population A	**Population B**
Plate 1 (6-fold)	100.0	100.0
Plate 2 (72-fold)	58.3	44.8
Plate 3 (864-fold)	5.2	2.1

The percentage of positive wells in the three dilution plates for each population of potential Δlmtkt cells.

All of the clones were positive for the ITS1 repeat region, indicating the presence of *Leishmania* genomic DNA (Figure 4-8A). Amplification of the transketolase gene was possible from five of the clones, however, not from clone B1.A1 (Figure 4-8B). These results suggest that deletion of the transketolase gene has been successful in clone B1.A1, although had been

retained in the other clones isolated. Interestingly, only two of the clones (B1.A1 and B1.A2) were positive for the *SAT* antibiotic resistance marker (data not shown), indicating that the drug pressure may have been too low to successfully select for parasites expressing the *SAT* resistance gene. A summary of the PCR results are shown in Table 4-4.

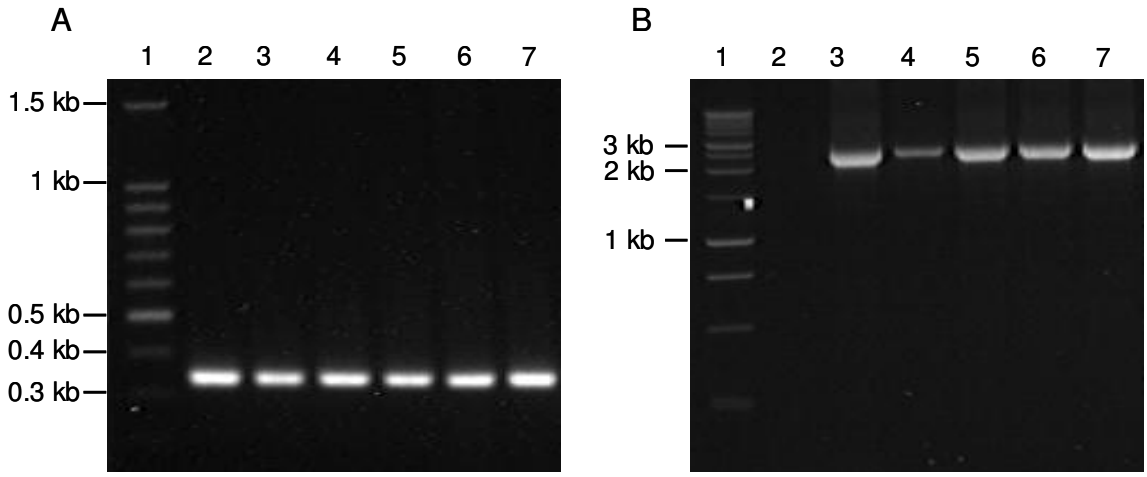


Figure 4-8. PCR analysis of *L. mexicana* double allele transketolase knockout clones.

A - Screen for ITS1 region using MB95 and MB96 to generate an expected product of 350 bp; B - Screen for transketolase using MB250 and MB251 to generate an expected product of 2100 bp. The lanes on each of the images are as follows, A; 1, 100 bp ladder; 2, clone B1.A1; 2, clone B1.A1; 3, clone B1.A2; 4, clone B1.A3; 5, clone B1.A4; 6, clone B1.A5; 7, clone B1.B1. B; 1, 1 kbp ladder; 2, clone B1.A1; 2, clone B1.A1; 3, clone B1.A2; 4, clone B1.A3; 5, clone B1.A4; 6, clone B1.A5; 7, clone B1.B1. Both DNA ladders were supplied by Promega.

Table 4-4. Summary of PCR analysis for potential TKT double allele gene knockouts

PCR product	**ITS 1**	***SAT***	***TKT***
Oligo (s)	MB95	MB210	MB250
Oligo (as)	MB96	MB211	MB251
Clone B1.A1	●	●	-
Clone B1.A2	●	●	●
Clone B1.A3	●	-	●
Clone B1.A4	nd	nd	nd
Clone B1.A5	●	-	●
Clone B1.B1	●	-	●
Clone B1.B2	●	-	●

Legend; ●, positive; -, negative; nd, not done.

Whilst PCR is a suitable approach to screen potential double knockout clones, it is not the most reliable. Therefore, the deletion of the transketolase gene was confirmed by Southern Blotting.

4.4.5 Southern Blot Analysis

In order to confirm that tkt had been deleted from the genome, the technique of Southern blotting was used. The *L. major* tkt DNA sequence was subjected to a GeneDB BLAST search against the *L. mexicana* contigs. The sequence of *L. mexicana* contig_0000588 was viewed and annotated using the vector NTI software (Invitrogen, Paisley). Digestion of genomic DNA with NruI would generate appropriate sized fragments of 2.2 kb and 3.2 kb when probed with a 1.5 kb section of the tkt ORF. A second restriction enzyme, BamHI, was chosen as this had been previously tested (Nicola Veitch, PhD thesis), however the exact location of a downstream BamHI restriction site remains unknown as it is located on another contig. A schematic diagram of the *TKT* locus is shown in Figure 4-9.

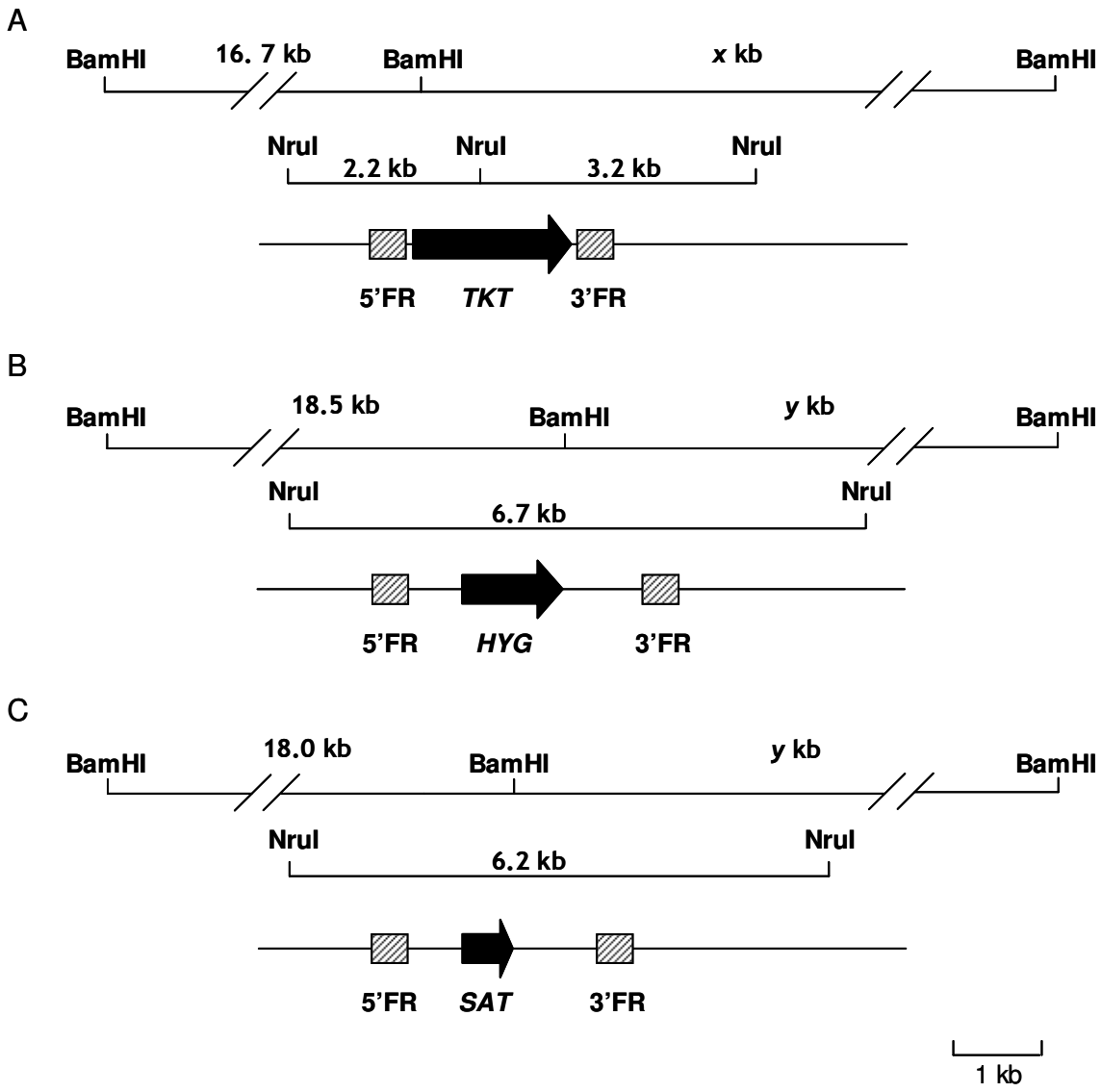


Figure 4-9. Schematic representation of the *L. mexicana* Transketolase locus.

A - endogenous transketolase locus. B - transketolase locus upon successful integration of the *HYG* knockout cassette. C - transketolase locus upon successful integration of the *SAT* knockout cassette. The ORF's are indicated by a solid black arrow, the TKT flanking regions are indicated by a striped box, and the restriction sites are indicated by a vertical line. The exact size of the downstream BamHI fragement (denoted 'x') is currently unknown due to incompletion of the *L. mexicana* genome project. The size of the downstream BamHI fragment (denoted 'y') for either knockout cassette integration is 0.5 kb smaller than 'x'. Scale bar represents 1 kb.

Initially, the six potential double allele knockout clones were analysed by Southern blot, which indicated that five of the clones had retained the tkt gene (data not shown). This result was consistent with the PCR screen which indicated that transketolase failed to amplify in clone B1A1; however, it was present in the five other clones (Table 4-4). Therefore, genomic DNA was isolated from

wild type *lmTKT*+/+, single allele knockout (Clone B1) and the *lmtkt*-/- (Clone B1A1) *L. mexicana* promastigotes, and digested with either NruI or BamHI.

The blot was probed with a 1.5 kb section of the full length ORF obtained from pMB-G52 using a BamHI / SacI double digest (Figure 4-10A). Wild type gDNA digested with BamHI was predicted to give a single band greater than 10 kb (Nicola Veitch, PhD Thesis). This is evident with a single large band greater than 10 kb (lane 2), a signal half as intense in the single allele knockout (lane 3), and no band in the double allele knockout (lane 4). Similarly, the expected bands of 2.2 kb and 3.2 kb produced by an NruI digest are evident for the wild type gDNA (lane 5), with the signal half as intense in the single allele knockout (lane 6), and no bands in the double allele knockout (lane 7), indicating that the transketolase gene had been deleted. The membrane was stripped and re-probed with the 5' transketolase UTR which is approximately 0.4 kb (Figure 4-10B). It was expected that the probe would anneal to a fragment of 16.7 kb for the WT gDNA (Figure 4-9A), a fragment of 18.5 kb for DNA containing the *HYG* cassette Figure 4-9B), and a fragment of 18.0 kb for DNA containing the *SAT* cassette Figure 4-9C). For the NruI digest it was expected that the probe would anneal to a 2.2 kb fragment for WT gDNA (Figure 4-9A), a 6.7 kb fragment for gDNA containing the *HYG* cassette (Figure 4-9B), and 6.2 kb fragment for gDNA containing the *SAT* cassette (Figure 4-9C). These expectations are consistent with the results obtained (Figure 4-10B).

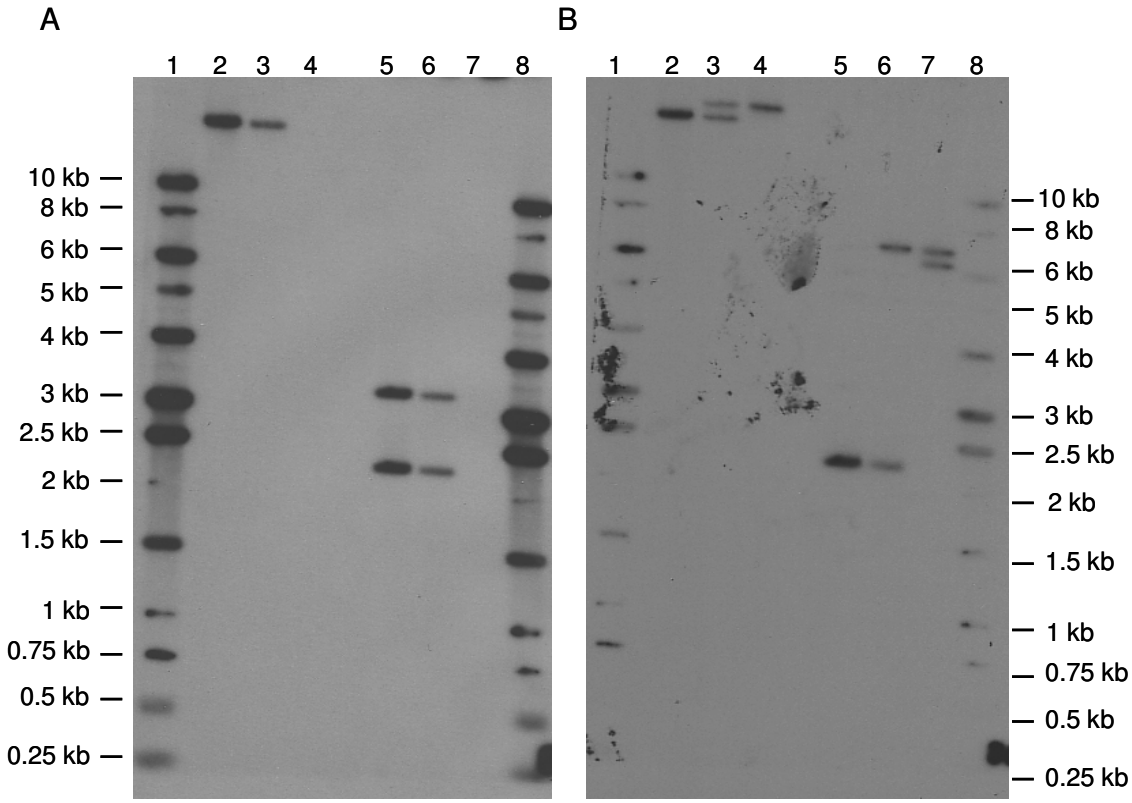


Figure 4-10. Southern blot demonstrating knockout of the transketolase gene for L. mexicana.

A: - membrane probed with a 1.5 kb section of the *TKT* ORF. B: - membrane striped and re-probed with a 0.4 kb section of the 5' flank. Each sample contains 10 µg gDNA and are loaded as follows - 1, 1 kb DNA ladder (Promega); 2, BamHI digest of *lm*TKT+/+ gDNA; 3, BamHI digest of *lm*TKT+/- gDNA; 4, BamHI digest of clone B1.A1 gDNA; 5, NruI digest of *lm*TKT+/+ gDNA; 6, NruI digest of *lm*TKT+/- gDNA; 7, NruI digest of clone B1.A1 gDNA; 8, 1 kb DNA ladder (Promega).

The results indicate that transketolase is clearly absent from the double allele knockout, confirming targeted gene replacement. Additionally, the intensity of the bands in the single allele knockout are half as intense as in the wild type, indicating a 50% reduction in the DNA coding for tkt.

4.5 Ribosomal expression of transketolase

For stable integration into the ribosomal locus, a derivative of pSSU-int (Misslitz, A *et al*. 2000), pGL631 (Prof. JC Mottram, University of Glasgow) was used. The tkt PCR fragment was ligated into the pGEM-T Easy vector, and sub-cloned into pGL631 using NotI and XhoI. The cassette can integrate at six locations on each allele at the ribosomal locus, with possible variation in gene expression, depending on integration site (Hilley JD, personal communication).

Unfortunately the coverage for this region of the *L. mexicana* genome was poor, possibly due to the complexity in assembling repeat regions, therefore making it difficult to determine the location(s) of the tkt gene(s). Consequently, it was decided to focus on designing an over-expression construct, using that to over-express tkt in the wild type, and generate tkt re-expression in the null mutants.

4.6 Over-expression of TKT in L. mexicana

To over-express TKT the pNUS-HnN vector was used (Tetaud, E *et al*. 2002). This vector acts as an episome within the *Leishmania* cell. An intergenic region between PGKA and PGKB of *Crithidia fasciculata* was used to direct maturation of the target gene and *NEO* drug resistance marker. Upstream of the target gene is the 5' untranslated region of PGKB, which directs trans-splicing, and the 3' untranslated region from *C. fasciculata* GSPS was used for polyadenylation of the *NEO* drug resistance marker. No protomer is required to drive transcription.

Unfortunately the XhoI and NotI restriction sites flanking the *TKT* could not be used to sub-clone transketolase from pMB-G51 into the pNUS-HnN vector. Consequently, an alternative cloning strategy was required. This involved removing the N-terminal HIS Tag, thrombin cleavage site and MCS by an EcoRI restriction digestion and dephosphorylation of the vector backbone to prevent re-ligation. The *TKT* gene from pMB-G51 was digested with EcoRI, and ligated with the dephosphorylated over-expression construct. A total of 27 bacterial clones were screened by PCR for the presence of the tkt gene. As the tkt gene could integrate in either orientation, five clones positive for tkt were selected at random and screened for orientation by restriction digest. Restriction digests of plasmid DNA with either BamHI or SacI would give rise to different sized products depending on the orientation of the gene (Figure 4-11A and Figure 4-11B). If the gene is orientated correctly then a BamHI restriction digest will generate two products approximately 3.8 kB and 4.5 kB in size, whereas incorrect orientation will generate two products approximately 6.2 and 2.1 kB in size. Similarly, a SacI restriction digest will yield products approximately 7.9 kB and 0.4 kB if the tkt gene is correctly orientated, or products approximately 6.5 kB and 1.8 kB if incorrectly orientated. The restriction digest profiles for both BamHI and SacI indicate that in clones 7 and 10 the transketolase ORF was correctly orientated (Figure 4-11C).

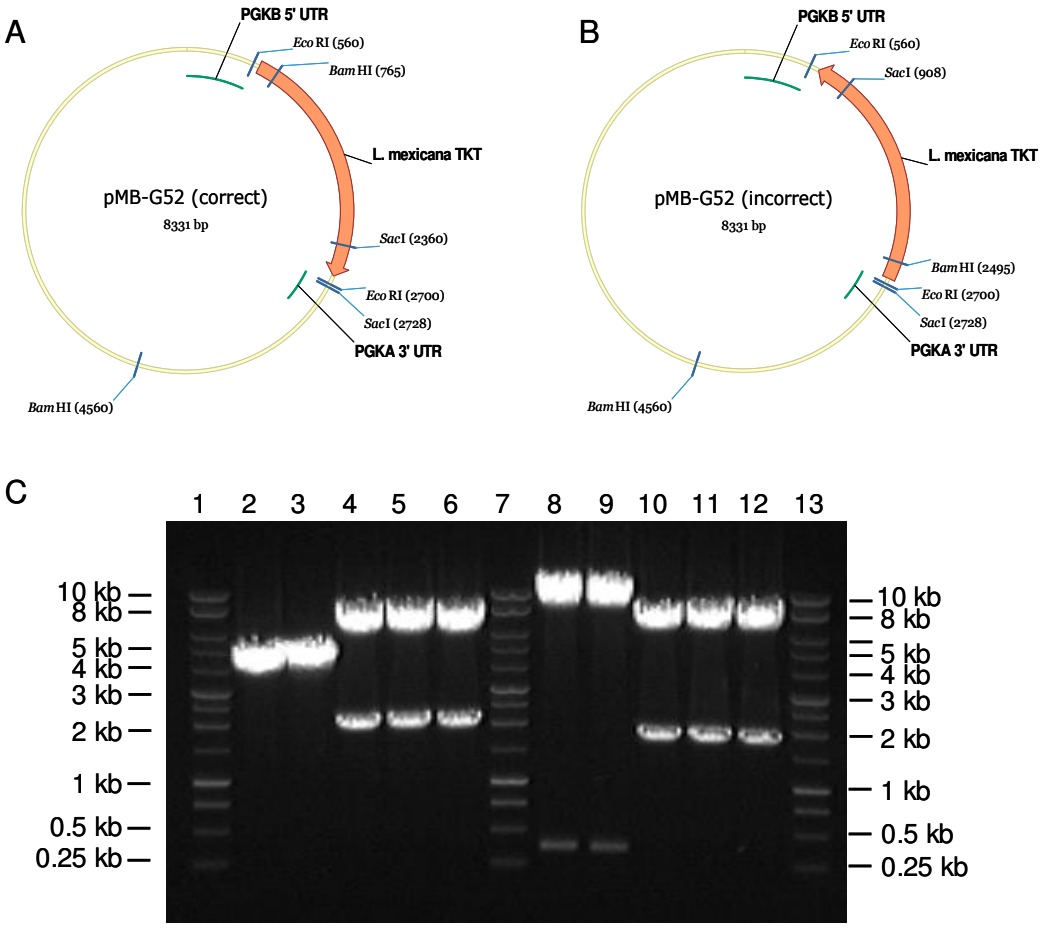


Figure 4-11. Determining the correct orientation of the transketolase ORF.

A – Plasmid map with the TKT ORF in the correct orientation. BamHI digest should yield products 3795 bp and 4536 bp. SacI digest should yield products 7963 bp and 368 bp. B – Plasmid map with the *TKT* ORF in the incorrect orientation. BamHI digest should give products 6266 bp and 2065 bp. SacI digest should give products 6511 bp and 1820 bp. C – Restriction Digest of plasmid DNA to screen for correct orientation of transketolase. The samples are as follows; 1, 1 Kb ladder; 2, clone 7; 3, clone 10; 4, clone 11; 5, clone 17; 6, clone 27; 7, 1 Kb ladder; 8, clone 7; 9, clone 10; 10, clone 11; 11, clone 17; 12, clone 27; 13, 1 Kb ladder. The DNA in lanes 2-6 have been digested with BamHI, wherease the DNA in lanes 8-12 have been digested with SacI. The 1 Kb ladder was supplied by Promega.

To generate a construct which could be used as a negative control, the tkt ORF was digested with NotI and XhoI, and the vector backbone ligated with the 100 bp MCS from the pGL631 vector, to give plasmid pMB-G53. This construct was used as a negative control for transfections involving the tkt over-expression construct. Wild type and transketolase knockout promastigotes successfully transfected with the ectopic constructs were analysed by PCR. Promastigotes were screened for the presence of a 2 kB fragment from within the tkt gene to within the Neomycin resistance marker (specific to the tkt over-expression

construct) (Figure 4-12A), and for the presence of the Neomycin phosphotransferase gene (present in all constructs) (Figure 4-12B). These results confirm the transfection of the correct plasmids into the correct cell lines.

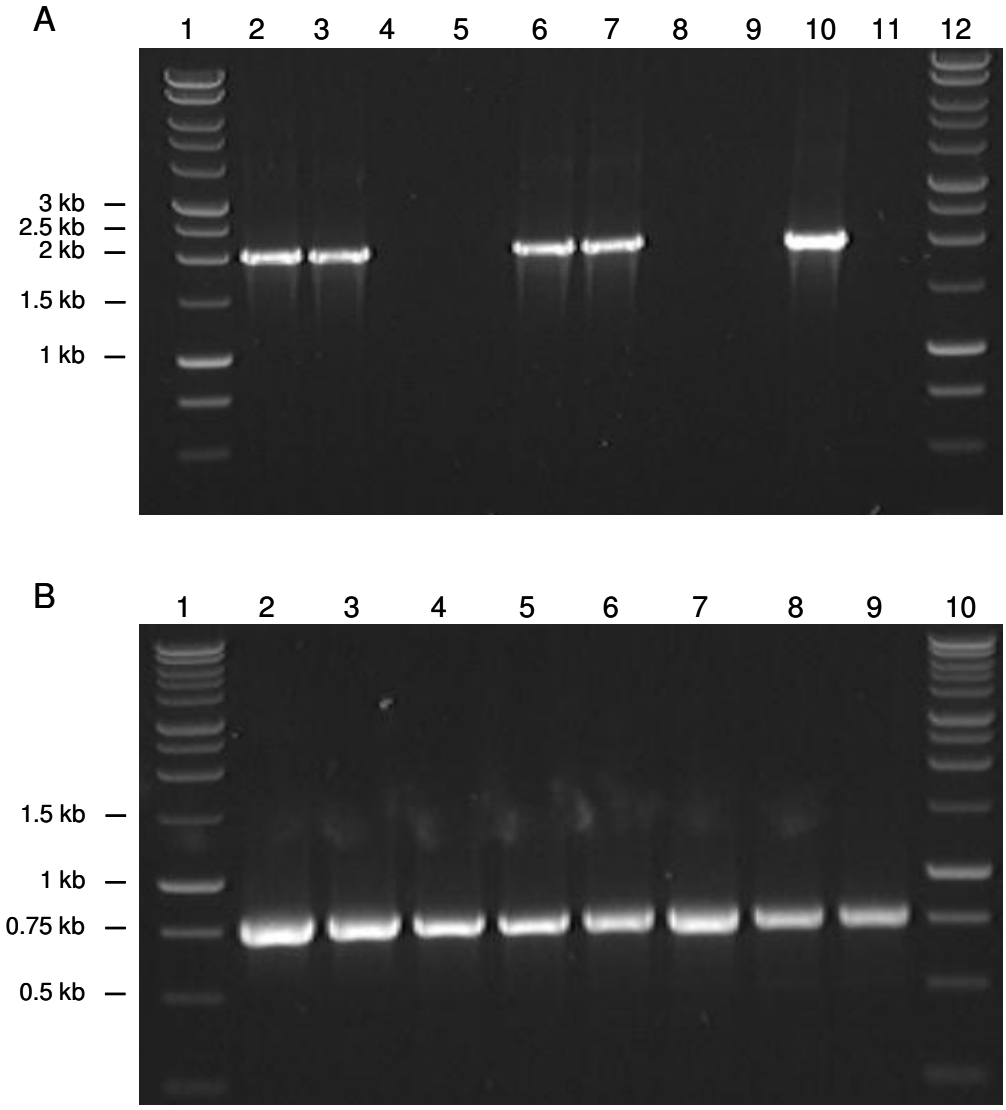


Figure 4-12. PCR analysis of *L. mexicana* TKT over-expression cell lines.

A - Screen for TKT-NEO fragment using MB334 and MB334 to generate an expected product of 2 kb; B - Screen for neomycin using MB37 and MB38 to generate an expected product of 750 bp. The lanes on each of the images are as follows, A; 1, 1 kb ladder; 2, wt pMB-G52A; 3, wt pMB-G52B; 4, wt pMB-G53A; 5, wt pMB-G53B; 6, tkt -/- pMB-G52A; 7, tkt -/- pMB-G52B; 8, tkt -/- pMB-G53A; 9, tkt -/- pMB-G53B; 10, pMB-G52; 11, pMB-G53; 12, 1 kb ladder. B; 1, 1 kb ladder; 2, wt pMB-G52A; 3, wt pMB-G52B; 4, wt pMB-G53A; 5, wt pMB-G53B; 6, tkt -/- pMB-G52A; 7, tkt -/- pMB-G52B; 8, tkt -/- pMB-G53A; 9, tkt -/- pMB-G53B; 10, 1 kb ladder. DNA ladders were supplied by Promega.

4.7 Transketolase sequence

To generate a construct for the re-expression of transketolase in the L. mexicana tkt knockouts, the ORF was amplified from gDNA, and ligated into the pGEM-T easy vector (Promega, Southampton, UK). The DNA sequences of two independently derived tkt clones were aligned with the *L. mexicana* tkt sequence (GenBank accession ID AJ427448) (Supplementary Figure 8-2). Two single base substitutions were identified at nucleotide positions 234 and 875. The first was identified as a silent mutation at amino acid position 78, whereas the second mutation induced an amino acid change at amino acid position 275 (A275T) (Supplementary Figure 8-1). Whilst a change from alanine to threonine may have consequences in the folding or activity of the protein, particularly post-translation modifications exploiting the threonine hydroxy group, sequence alignment with tkt from other *Leishmania* species all revealed a threonine residue at position 275 (Figure 4-13). This would suggest that the *L. mexicana* tkt sequence (GenBank accession ID AJ427448) is incorrect at amino acid position 275.

```
                          241                                  *  280
    L. braziliensis TKT MIVQTTTIGYGSSKQGTEKVHGAPLGDEDIASVKTKFGRD
        L. infantum TKT MIVQTTTIGFGSSKQGTEKVHGAPLGEEDIANVKTKFGRD
           L. major TKT MIVQTTTIGFGSSKQGTEKVHGAPLGEEDIANVKTKFGRD
        L. mexicana TKT MIVQTTTIGFGSSKQGTEKVHGAPLGEEDIANIKAKFGRD
L. mexicana TKT Clone 1 MIVQTTTIGFGSSKQGTEKVHGAPLGEEDIANIKTKFGRD
```

Figure 4-13. Alignment of transketolase residues 241 to 280 from several *Leishmania spp*.

Gene sequences for *L. infantum* (LinJ.24.21.50), *L. braziliensis* (LbrM.24.2140) and *L. major* (LmjF.24.2060) were obtained from TriTrypDB, whereas the gene sequence from *L. mexicana* was obtained from GenBank (accession ID AJ427448). Both *L. mexicana* strains were MNYC/BZ/62/M379, and from the same laboratory. The amino acid difference at residue 275 is indicated with an asterisk.

4.8 Phenotypic Analysis

4.8.1 Growth curve

The growth of the *lmTKT*+/-, *lmtkt*-/-, and *lmtkt*-/- TKT ribosomal re-expression cell lines were compared to that of the WT *L. mexicana* promastigotes (Figure 4-14). Both the single allele knockout and the *lmtkt*-/- cell lines have a slightly longer initial lag phase than the WT cell line; however there is little difference

in the mean generation time (WT, 13.5 hours; *lmTKT+/-* cell single allele knockout, 14 hours; *lmtkt-/-*, 15 hours). Additionally, the stationary phase concentration of the WT cell line is approximately 2 x 10^7 cells per ml, compared with approximately 1 x 10^7 cells per ml for the *lmtkt-/-* cell line. Ribosomal re-expression of TKT had similar growth rates and maximum densities when compared to the wild type cells. These results were displayed on a linear scale to highlight a potential difference in maximum growth density. Therefore, TKT is not essential for growth or survival under the standard laboratory conditions used in these experiments. The growth phenotype described here for *L. mexicana* is similar to the growth phenotype of *T.brucei lmtkt-/-* procyclic trypanosomes which was unaltered when compared to WT procyclic trypanosomes (Stoffel, SA *et al*. 2011).

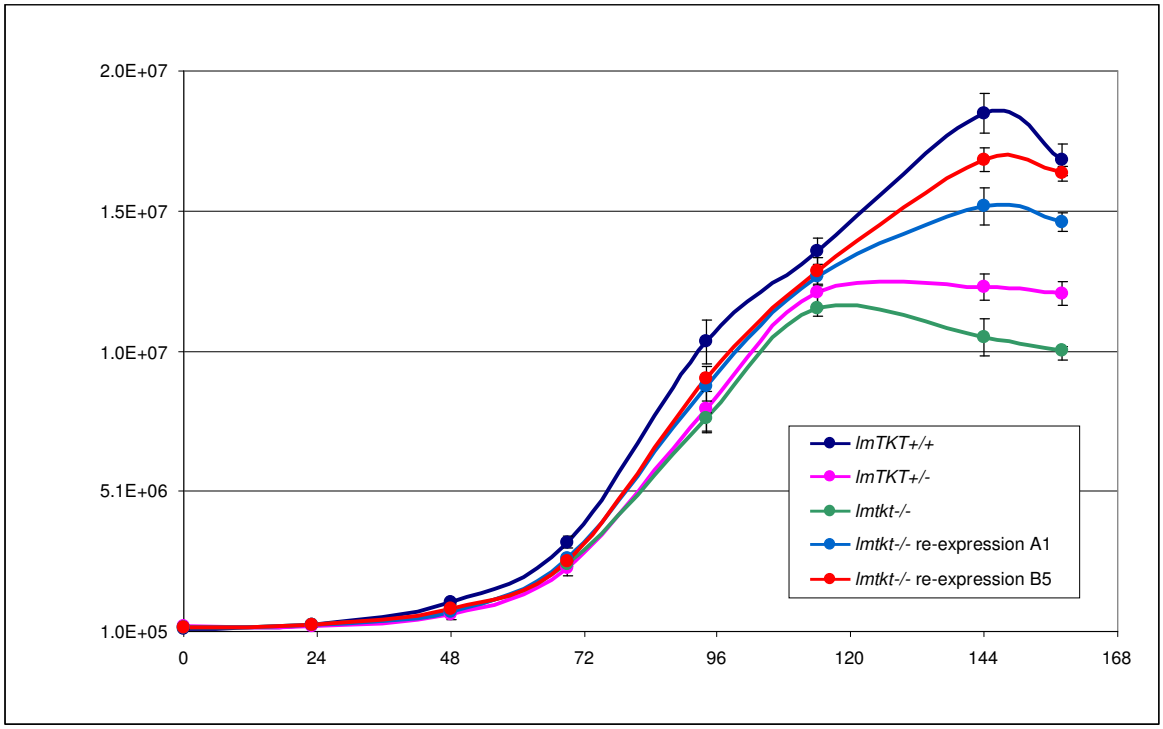


Figure 4-14. Growth of genetically manipulated *L. mexicana* promastigotes.

Wild type (WT), *TKT* single allele knockout (*lmTKT+/-*), *tkt*-null (*lmtkt-/-*), and *tkt*-null:*TKT*-complemented (*lmtkt-/- TKT* ribosomal re-expression) *L. mexicana* promastigotes cell lines were counted over a 7 day period. No antibiotics were applied to the *lmTKT+/-* or *lmtkt-/-*cell lines, however the *TKT* re-expression cell lines were grown in the presence of 10 µg/ml puromycin. Three cultures of each cell line were counted to generate a mean value +/- SEM (n = 3).

4.8.2 Over-expression of Transketolase

The growth of the wild type *L. mexicana* promastigote cell lines transfected with the transketolase over-expression construct were compared to those which were transfected with the empty vector (Figure 4-15). The populations containing the empty vector had a slower lag phase than those with the transketolase over-expression construct; however there is no obvious effect on growth with each population possessing a similar mean generation time

(

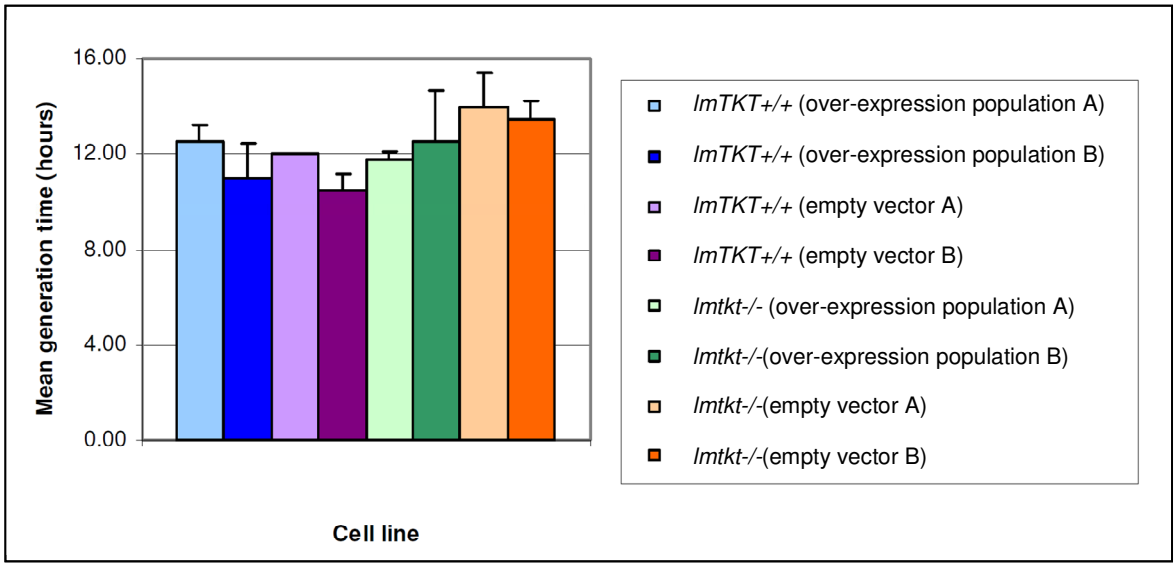


Figure 4-17). The maximum density of the stationary phase promastigotes was approximately the same for each population irrespective of whether the cells were transfected with the empty vector or the tkt over-expression construct.

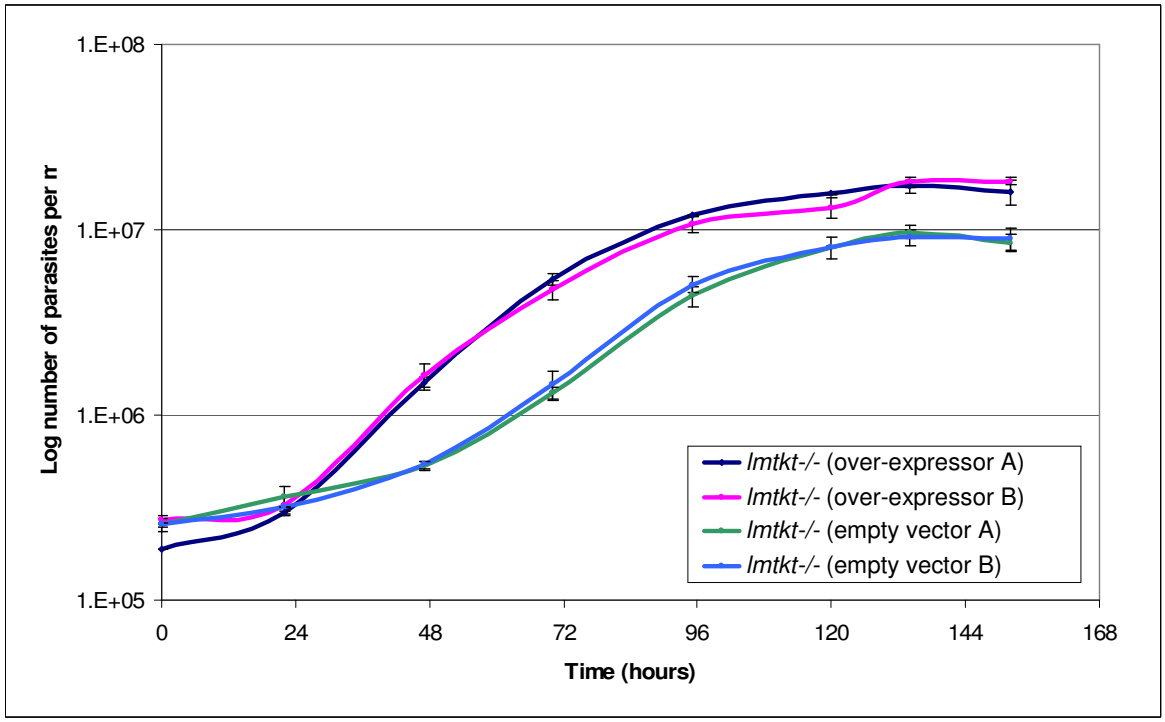


Figure 4-15. Effect on the growth of *L. mexicana* wild type cells transfected with episomal constructs.

Two distinct populations over-expressing the TKT enzyme were compared with two distinct populations containing the empty vector over a 7 day period. All cell lines were grown in the presence of 50 µg/ml G418. Two cultures of each cell line were counted to generate a mean value +/- SEM.

The growth of *lmtkt-/-* mutant populations transfected with the transketolase over-expression construct were compared to those which were transfected with the empty vector (Figure 4-16). The populations which contained the empty vector had a significantly increased lag time, in addition to a stationary phase density of around 9 x 10^6 cells per ml. The populations with the transketolase over-expression construct had a stationary phase density of approximately 1.5 x 10^7 cells per ml. Additionally, the *lmtkt-/-* cell populations containing the empty vector were not viable for a period of more than three weeks, despite the stable nature of *lmtkt-/-* null cell line with no episomal construct.

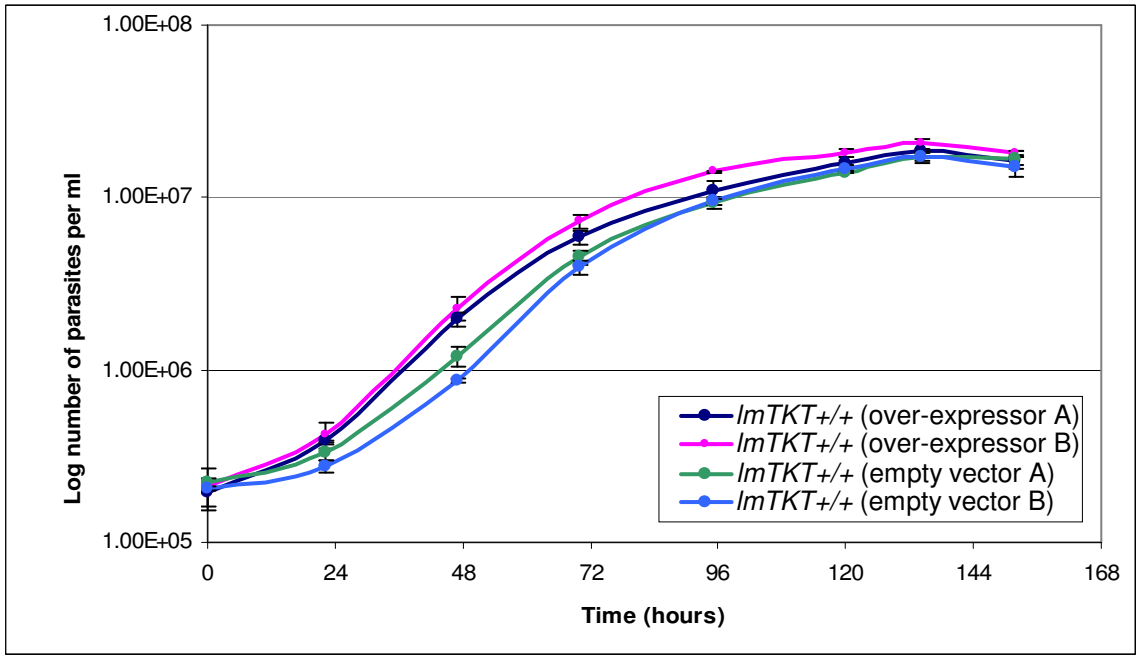


Figure 4-16. Effect on the growth of *L. mexicana* TKT null (*lmtkt-/-*) cells transfected with episomal constructs.

Two distinct populations over-expressing the TKT enzyme were compared with two distinct populations containing the empty vector over a 7 day period. All cell lines were grown in the presence of 50 µg/ml G418. Two cultures of each cell line were counted to generate a mean value +/- SEM.

The mean generation time of the *lmtkt-/-* cells over-expressing TKT were slightly lower than those which contained the empty vector, and in addition were comparable to the mean generation times of the wild type cells with the over-expression construct or empty vector (

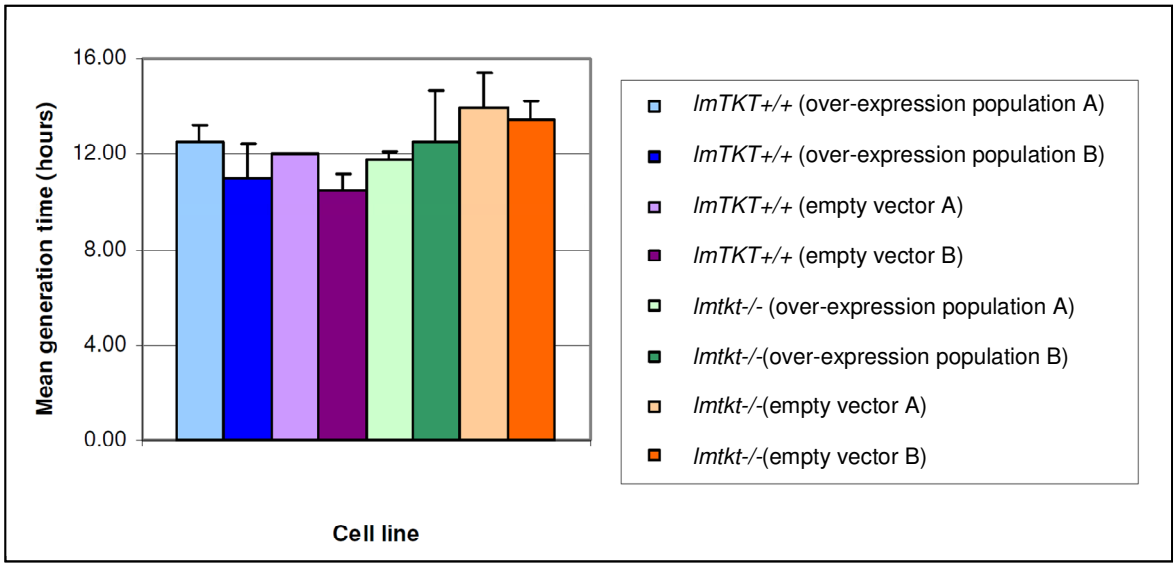


Figure 4-17).

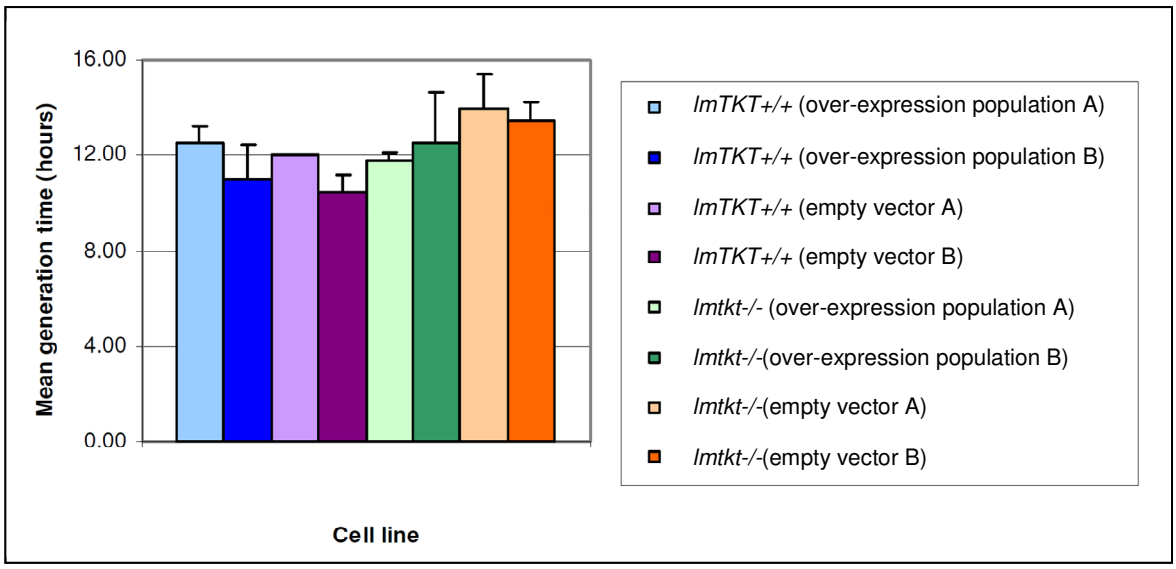


Figure 4-17. Mean generation time of *L. mexicana* promastigotes transfected with episomal constructs.

The over-expression of tkt has no effect on the growth of wild type *L. mexicana* promastigotes. Additionally, there is no difference in growth between wild type and *lmtkt-/-* cells when TKT is over-expressed.

4.8.3 Medium colour change

Initial observations indicated differences in the colour of the medium between the wild type and tkt knockout cell line. A similar phenotype was observed in the glucose transporter knockout (*lmgt-/-*) cell line (Burchmore, RJ *et al*. 2003). To further investigate this, wild type, *lmTKT+/-*, *lmtkt-/-*, and *lmgt-/- L. mexicana* promastigotes were grown for a period of two weeks. After two weeks of growth, the cells will have reached stationary phase (for growth curves refer to Figure 4-14; (Burchmore, RJ *et al*. 2003), with the number of viable cells declining steadily.

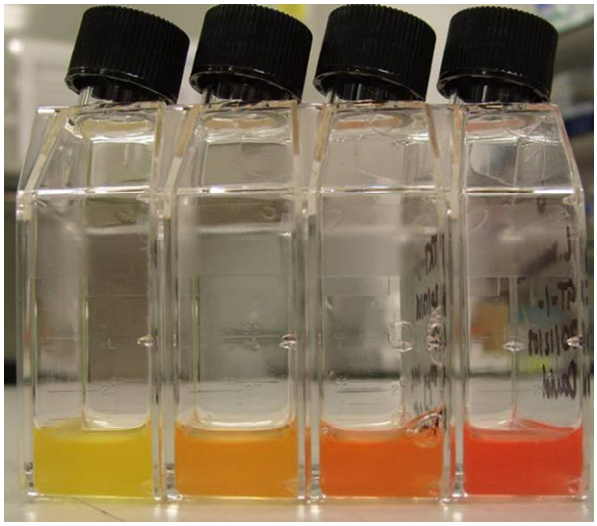


Figure 4-18. *L. mexicana* promastigote HOMEM acidification.

Promastigotes in the stationary phase of growth exhibit various degrees of medium acidification. The acidification of the *lmtkt-/-* is not as pronounced as the wild type cell line, despite similar cell densities. The *lmgt-/-*, which is unable to transport glucose, is shown as a control.

The medium from the wild type cell lines was yellow, whereas the medium from the tkt knockout cell line was orange, indicating a higher pH. The medium colour from the single allele knockout cell line was between yellow and orange. The glucose transporter knockout cell line, designated *lmgt-/-*, is unable to metabolise glucose, and as a consequence, the medium does not change colour.

These observations indicate that the colour change of the medium may be linked to the expression of transketolase. Furthermore, the phenotypic similarity to the glucose transporter knockout cell line suggests a possible link between transketolase and glucose metabolism. The colour change that occurs during the

growth of *L. mexicana* promastigotes in cultures may be the result of the excretion of organic acids that are produced from the partial metabolism of glucose. The *lmgt-/-* promastigotes cannot utilize exogenous glucose, and release significantly reduced amounts of succinate and pyruvate due to a metabolic shift from glycolysis to amino acid metabolism (Dhilia Lamasudin, personal communication), which may account for the slower rate of medium acidification. The slower rate of medium colour change in the *lmtkt-/-* cell line is also consistent with reduced metabolism of glucose. These observations do not take into consideration cell density, which could have a significant influence.

4.8.4 Glucose Utilisation

To investigate a potential metabolic shift, the glucose consumption of the Δlmtkt null cell line was compared to that of the parental wild type cell line using an enzymatic assay. The Glucose Oxidase (GO) Assay was used to determine the quantity of free glucose remaining in the supernatant at various time points (0 hours, T_0; 30 hours, T_{30}; 90 hours, T_{90}) from three independent experiments (section 2.9.1).

The linear range of the glucose oxidase assay is between 20 and 80 µg of glucose. HOMEM contains 3 g of D-glucose per litre (16.66 mM); therefore, 30 µl of supernatant will contain a maximum of 80 µg glucose. A series of glucose standards were quantified resulting in the following equation: $y = 53.724x - 13.031$; where '*y*' represents the quantity of glucose in µg and 'x' represents the absorbance at 595 nm. The R-squared value was 0.9582.

Three independent cell cultures were set up for the *L. mexicana* wild type and *lmtkt-/-* cell lines, in the late log phase of growth, at approximately 1×10^7 cells per ml. The quantity of glucose in supernatant and the cell densities were determined at T_0, T_{30} and T_{90}. This would allow the glucose consumption to be normalised against cell numbers.

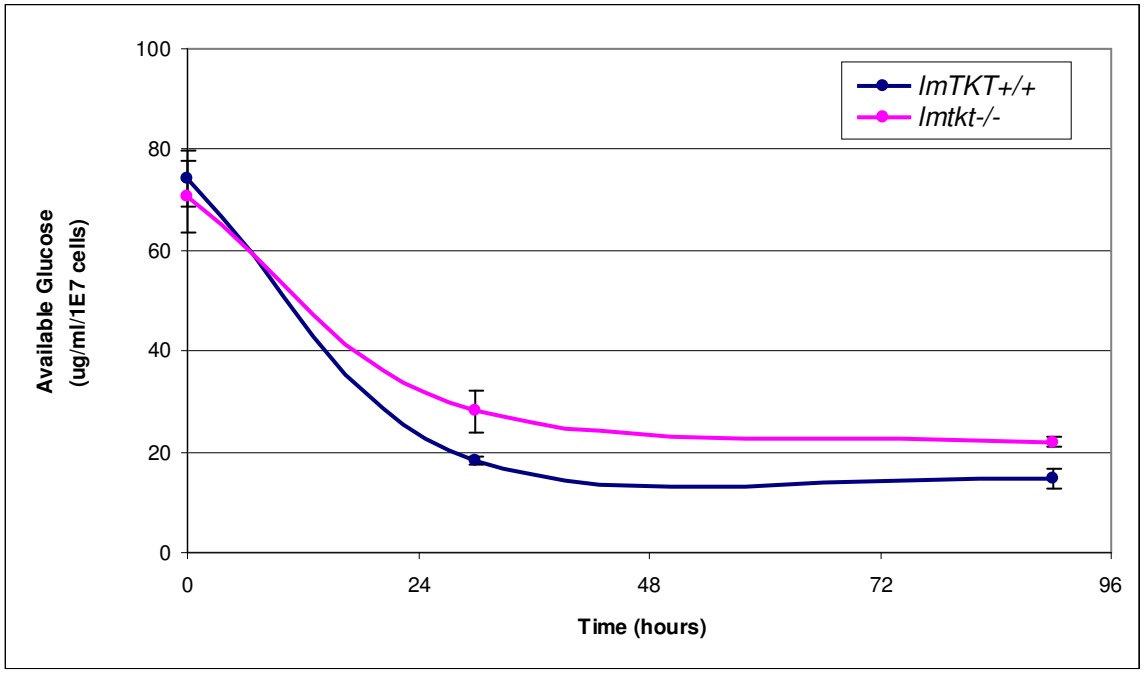


Figure 4-19. Utilisation of glucose by *L. mexicana* promastigotes.

The quantity of glucose remaining in the medium, and cell numbers were determined for three time points; T_0, T_{30}, and T_{90}. The available glucose is expressed as ug/ml normalised to 1x 10^7 cells. The assays were performed on three separate cultures for each cell line (n = 3).

The quantity of glucose available in the medium for the wild type promastigotes decreased from approximately 70 µg/ml/1x10^7 cells to 20 µg/ml/1x10^7 cells in the first 30 hours; compared to a decrease from approximately 70 µg/ml/1x10^7 cells to 30 µg/ml/1x10^7 cells for the *lmtkt-/-* promastigotes. After 90 hours, the medium from both cell lines contained approximately 20 µg of glucose/ml/1x10^7 cells. These results indicate that there is no significant difference in glucose consumption between the wild type and *lmtkt-/-* cell lines, despite the clear difference in acidification of the medium. This would suggest that whilst *lmtkt-/-* promastigotes are not excreting organic acids at the same rate as the wild type promastigotes, they do metabolise similar levels of glucose. A change in secreted end products rather than change in rate of glucose consumption is thus inferred.

The concentration of glucose determined by the GO assay at T_0 was 10.2 mM, however, the HOMEM recipe in the laboratory states a concentration of approximately 16 mM. After 90 hours, this decreases to 7.0 mM and 9.3 mM for the wild type and *lmtkt-/-* promastigotes respectively. This suggests that once

the cells reach the maximum density of approximately $2.5x10^7$ cells per ml, glucose consumption decreases significantly.

The promastigotes require NADPH to defend against oxidative stress. Whilst the oxidative branch of the PPP can function normally in the *lmtkt-/-* cell line, the absence of TKT means that R5P cannot be converted to glycolytic intermediates by the non-oxidative branch. Therefore, a significant proportion of G6P enters the PPP to regenerate NADPH, and these molecules cannot re-enter the glycolytic pathway as another phosphorylated sugar at a later stage.

4.8.5 Sensitivity of L. mexicana to oxidative stress and drugs

In procyclic form *T.brucei*, transketolase knockouts were initially reported to be more sensitive than wild type cells to oxidative stress induced by exposure to methylene blue and H_2O_2. Conversely, trypanosomes over-expressing transketolase were less sensitive to these compounds than trypanosomes expressing endogenous levels of TKT (Stoffel SA, personal communication). Furthermore, transketolase knockouts were hypersensitive to pentamidine, whereas the cells over-expressing transketolase were less sensitive to pentamidine. These results suggest that a possible mode of action for pentamidine in procyclic form trypanosomes is through the generation oxidative stress, and that transketolase is important in defending the cells against oxidative stress. Therefore, the sensitivity of the *L. mexicana* wild type cell line to methylene blue and pentamidine were compared to the transketolase knockout cell line and the transketolase knockout cell lines re-expressing transketolase.

The transketolase knockout cell line was twice as sensitive to pentamidine as the wild type cell line (wild type, 241.4 nM; *lmtkt-/-*, 118.3 nM). The tkt null cell line re-expressing transketolase from the ribosomal locus had a similar level of sensitivity to the wild type cell lines (*lmtkt-/-* ribosomal re-expression G49, 267.1 nM). However, the cells re-expressing transketolase from an ectopic plasmid were slightly less sensitive to pentamidine than the *lmtkt-/-* cell line, although were still more sensitive than the wild type cell line (*lmtkt-/-* episomal re-expression G52A, 177.8 nM; *lmtkt-/-* episomal re-expression G52B, 147.7 nM).

There is very little difference in sensitivity to methylene blue between the wild type and *lmtkt-/-* cell lines (wild type, 4.309 μM; LmΔTKT null, 3.542 μM).The *lmtkt-/-* cell lines re-expressing transketolase were more sensitive to methylene blue than the wild type or *lmtkt-/-* cell lines. As the cell lines re-expressing transketolase were not doing so from the endogenous locus, the cells were grown in the presence of G418 or puromycin to maintain a high level of gene expression. This may have had an effect of the EC_{50} values observed.

Table 4-5. Pentamidine EC_{50} values for *L. mexicana* promastigotes.

Cell Line	EC_{50} Pentamdine (nM)	95% Confidence Interval
lmWt	241.4	212.3 to 273.8 nM
lmtkt-/-	118.3	104.8 to 133.3 nM
lmtkt-/- re-expression (G49)	267.1	227.9 to 301.4 nM
lmtkt-/- re-expression (G52A)	177.8	153.7 to 202.8 nM
lmtkt-/- re-expression (G52B)	147.7	132.0 to 164.6 nM

Cell lines are as follows; wild type (*lm*WT), *tkt* null mutant (*lmtkt-/-*), ribosomal TKT re-expression (*lmtkt-/-* re-expression G49) and episomal TKT expression (*lmtkt-/-* re-expression G52A and *lmtkt-/-* re-expression G52B). (n=3)

Table 4-6. Methylene blue EC_{50} values for *L. mexicana* promastigotes.

Cell Line	EC_{50} Methylene Blue (μM)	95% Confidence Interval
lmWt	4.3	4.2 to 4.5 μM
lmtkt-/-	3.5	3.3 to 3.8 μM
lmtkt-/- re-expression (G49)	1.6	1.3 to 2.0 μM
lmtkt-/- re-expression (G52A)	1.0	0.9 to 1.2 μM
lmtkt-/- re-expression (G52B)	1.1	0.9 to 1.3 μM

Cell lines are as follows; wild type (*lm*WT), *tkt* null mutant (*lmtkt-/-*), ribosomal TKT re-expression (*lmtkt-/-* re-expression G49) and episomal TKT expression (*lmtkt-/-* re-expression G52A and *lmtkt-/-* re-expression G52B). Started at 4×10^5, (n=3)

In the absence of transketolase, *L. mexicana* promastigotes are twice as sensitive to pentamidine as the wild type promastigotes; re-introducing transketolase to the knockout cell line decreased sensitivity to pentamidine. However, there was little difference in sensitivity to methylene blue between the wild type and TKT knockout cell line.

As previously mentioned, the procyclic form *T.brucei* cells over-expressing tkt were shown to be less sensitive to pentamidine and methylene blue. Therefore, the alamar blue assay was used to compare the sensitivity of *L. mexicana* WT cell lines transfected with episomal copies of the tkt gene or the empty vector to both pentamidine and methylene blue. The two cell populations over-expressing transketolase were not significantly more resistant to pentamidine or methylene blue than the populations that contained the empty vector (Table 4-7 and Table 4-8). The cell populations that contained the empty vector were approximately 4-fold less sensitive to pentamidine (1.117 μM and 1.09 μM) than the wild type cell line (241.4 nM), and approximately 2-fold less sensitive to methylene blue (8.93 μM and 9.07 μM) than the wild type cell line (4.309 μM). These assays were performed whilst the cells were grown under drug pressure with G418 which makes a direct comparison difficult. However, this may account for the difference between the EC_{50} values obtained for the wild type cell line and those obtained for the two wild type cell populations transfected with the empty vector.

Table 4-7. Pentamidine EC_{50} values for *L. mexicana* wild type promastigotes transfected with the TKT over-expression construct.

Cell Line	EC_{50} Pentamdine (μM)	95% Confidence Interval
lm WT TKT overexpressor A	1.24	1.20 to 1.29 μM
lm WT TKT overexpressor B	1.41	1.37 to 1.46 μM
lm WT empty vector A	1.12	1.05 to 1.20 μM
lm WT empty vector B	1.09	1.06 to 1.13 μM

Started at 2 x 10^5 cells per ml (n=3)

Table 4-8. Methylene Blue EC_{50} values for *L. mexicana* wild type promastigotes transfected with the TKT over-expression construct.

Cell Line	EC_{50} Methylene Blue (μM)	95% Confidence Interval
lm WT TKT overexpressor A	11.47	8.58 to 14.18 μM
lm WT TKT overexpressor B	10.70	8.43 to 12.98 μM
lm WT empty vector A	8.93	8.12 to 9.74 μM
lm WT empty vector B	9.07	7.86 to 10.39 μM

Started at 2 x 10^5 cells per ml (n=3)

These results suggest that the over-expression of transketolase does not play a significant role in the defence against oxidative stress in *L. mexicana* promastigotes. The flux through the oxidative branch of the PPP is determined by cellular requirements for NADPH and R5P; with the non-oxidative branch functioning to convert sugar phosphates to glycolytic intermediates, which can be diverted back through the oxidative branch of the PPP if necessary. However, if the cells are grown in a glucose rich environment, it may not be necessary to recycle the sugar phosphates produced by the oxidative branch of the PPP as G6P is readily available. As the *L. mexicana* promastigotes are grown in HOMEM, a glucose rich medium, the over-expression of transketolase does not increase the flux of G6P through the oxidative branch of the PPP. Therefore, the rate NADPH regeneration is unaltered, and thus the sensitivity to oxidative stress remains the same.

4.8.6 Metabolomic Analysis

A metabolomic analysis was carried out in collaboration with the Metabolomics Facility in Toulouse. Parasite samples were sent to colleagues in Bordeaux (Bringaud F, Université Victor Segalen, Bordeaux) and prepared for analysis in Toulouse.

The *lmtkt-/-* cells do not acidify the medium to the same extent as the wild type cells (section 4.8.3); yet consume similar quantities of glucose per 1 x 10^7 cells over a 90 hour time period.

Transketolase is a key enzyme in the non-oxidative branch of the PPP that provides a link between the ribulose 5-phosphate and the glycolytic intermediates F6P and G3P. If there is significant flux through the non-oxidative branch of the PPP then the promastigotes must respond by dealing with the accumulation of Ru5P.

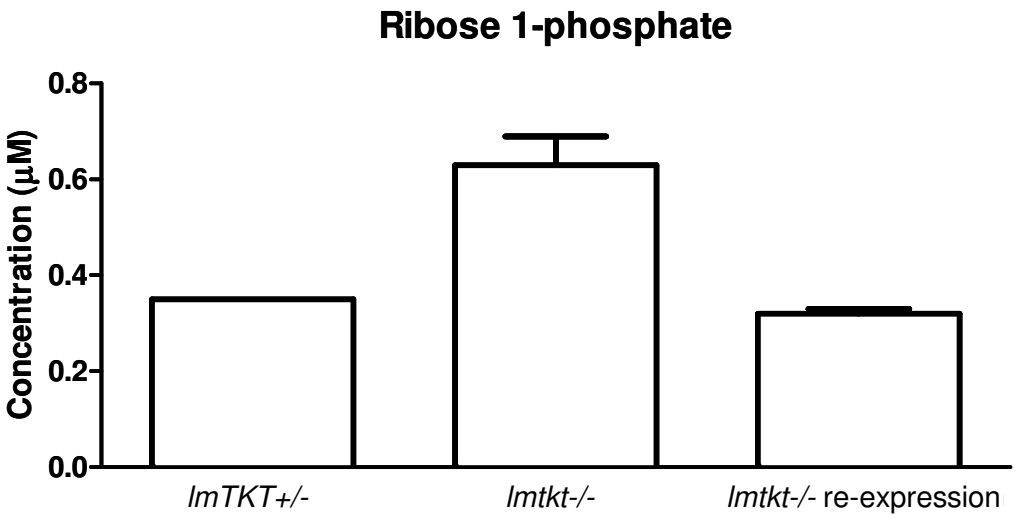


Figure 4-20. A comparison of intracellular levels of ribose 1-phosphate (R1P) in *L. mexicana*.

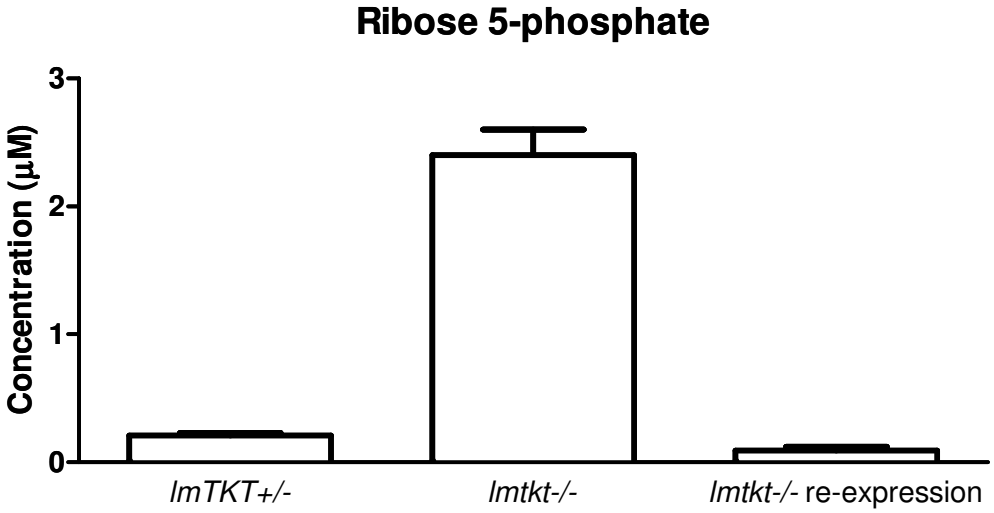


Figure 4-21. A comparison of intracellular levels of ribose 5-phosphate (R5P) in *L. mexicana*.

The metabolomic analysis indicates that there is an accumulation of ribose 1-phosphate (R1P) Figure 4-20) and ribose 5-phosphate (R5P) (Figure 4-21) in the *lmtkt-/-* cell line when compared to the wild type cells. The re-expression of TKT from the ribosomal locus in the *lmtkt-/-* cell line reverses this phenotype to concentrations similar to that of the wild type.

The first enzyme in the non-oxidative branch of the PPP is ribulose-5-phosphate isomerase (EC 5.3.1.6) which converts Ru5P to R5P. Deletion of TKT results in the accumulation of R5P, consistent with its production by the isomerase and

loss of consumption by transketolase. According to KEGG, R5P can be converted to a range of metabolites in *Leishmania* parasites (Table 4-9).

Whilst it is possible to produce Xu5P via Ru5P, the absence of TKT prevents the conversion of Xu5P to G3P. Therefore, the deletion of TKT prevents any flux through the oxidative branch of the PPP returning to the glycolytic pathway. If the concentration of R5P increases it is not surprising that the concentration of R1P also increases since R1P can be produced from R5P by phosphoglucomutase (EC 5.4.2.2), although it is unknown whether the Leishmania enzyme can catalyse this reaction.

Leishmania parasites which do not express TKT would not be expected to be able to produce S7P, an intermediate metabolite in the non-oxidative branch of the PPP, from R5P. When the wild type and *lmtkt*-/- cell lines are compared, there is a 14-fold reduction in sedoheptulose 7-phosphate in the tkt knockout cell line (Figure 4-22). However, the re-expression of TKT in the null cell line does not return the sedoheptulose 7-phosphate levels to that of the wild type, instead only a 2.5-fold increase is observed.

Table 4-9. Metabolic fates of Ribose 5-phosphate.

Metabolite	Abbreviation	KEGG CompCode	Enzyme	EC number	Accession_ID
Sedoheptulose 7-phosphate	S7P	C05382	TKT	2.2.1.1	LmjF24.2060
Ribulose 5-phosphate	Ru5P	C00199	ribose 5-phosphate isomerase	5.3.1.6	LmjF28.1970
Ribose 1-phosphate	R1P	C00620	phosphoglucomutase	5.4.2.2	LmjF21.0640
D-Ribose		C00121	ribokinase	2.7.1.-	LmjF27.0420
5-Phospho-alpha-D-ribose 1-diphosphate	PRPP	C00119	ribose-phosphate pyrophosphokinase	2.7.6.1	LmjF08.0510

The five reversible reactions listed by KEGG for ribose 5-phosphate

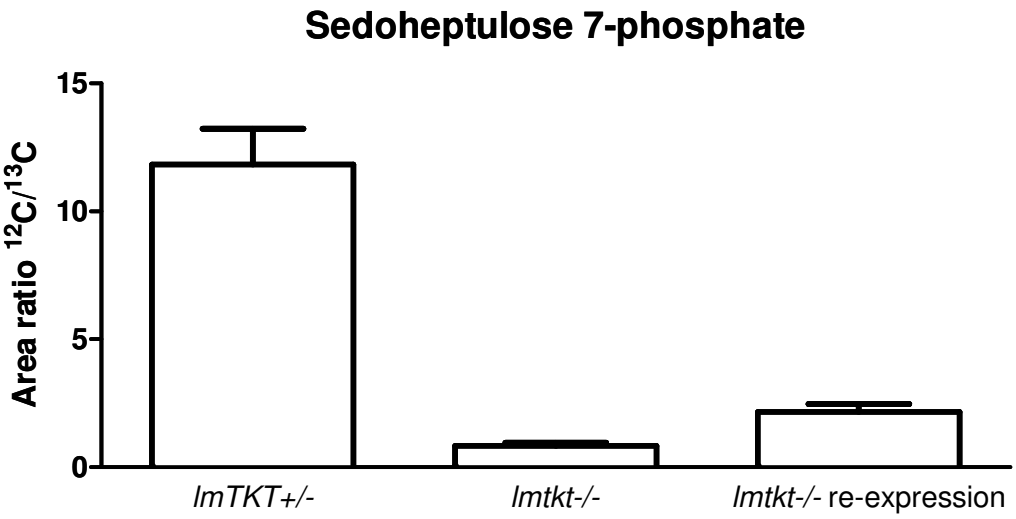


Figure 4-22. A comparison of intracellular levels of sedoheptulose 7-phosphate (S7P) in *L. mexicana*.

No standard commercially available for sedoheptulose 7-phosphate.

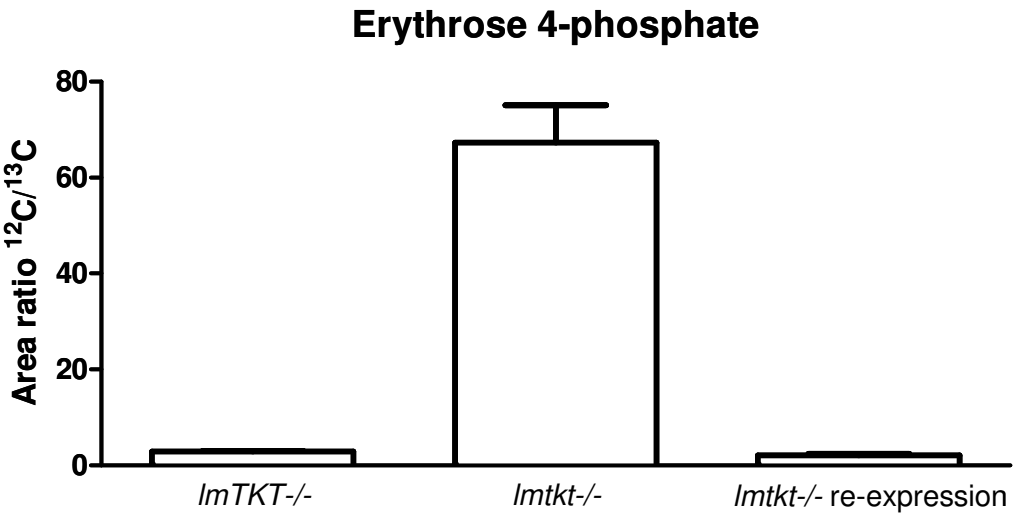


Figure 4-23. A comparison of intracellular levels of erythrose 4-phosphate (E4P) in *L. mexicana*.

No calibration curve for erythrose 4-phosphate.

In the tkt null cell line there is a 23-fold increase in E4P when compared to the wild type cell line (Figure 4-23). Re-expressing tkt in the tkt null cell line reduces E4P.

The reduced level of S7P and increased level of E4P is intriguing. S7P is usually produced by transketolase when a 2-carbon subunit is transferred from Xu5P to R5P. E4P is then produced when transaldolase transfers a 3-carbon from S7P to GA3P, and transketolase transfers a 2-carbon subunit from Xu5P to E4P, creating F6P. The increase in E4P indicates a source of this metabolite independent of TKT.

If there is an alternative method of producing S7P, then transaldolase can create F6P and E4P if GA3P is present as an acceptor. It is possible to produce S7P from both E4P and DHAP, sedoheptulose 1,7-bisphosphate acting as an intermediate metabolite (Figure 4-24A). However, as E4P would be required as a substrate, this reaction alone would not account for the elevated levels of E4P.

Another possibility is that there is an enzyme with phosphoketolase activity, in which F6P is converted to E4P and acetyl phosphate (C00227) (Figure 4-24B). No trypanosomatid enzyme with this function has currently been identified, although it would be of considerable interest to find such an enzyme. If this were the case, why would the cells produce excess E4P at the expense of consuming F6P, an important glycolytic intermediate?

Whilst there is no known biochemical or enzymatic mechanism to transfer a 1-carbon subunit from R5P or Xu5P to GA3P creating E4P, it does remain a theoretical, yet unlikely possibility.

Leishmania parasites have been shown to possess ribokinase which converts D-ribose to R5P (Ogbunude, PO *et al*. 2007). However, it is unknown whether this reaction can operate in reverse. If so, this would allow the parasites to metabolise R5P and excrete D-ribose. The *T. brucei* ribokinase enzyme is capable of dephosphorylating R5P (Kerkhoven E, personal communication). Additionally, trypanosomes grown in C^{13} labelled D-glucose produce C^{13} labelled D-ribose (Anderson JJ, personal communication), indicating that ribokinase may dephosphorylate R5P *in vivo* and excrete ribose.

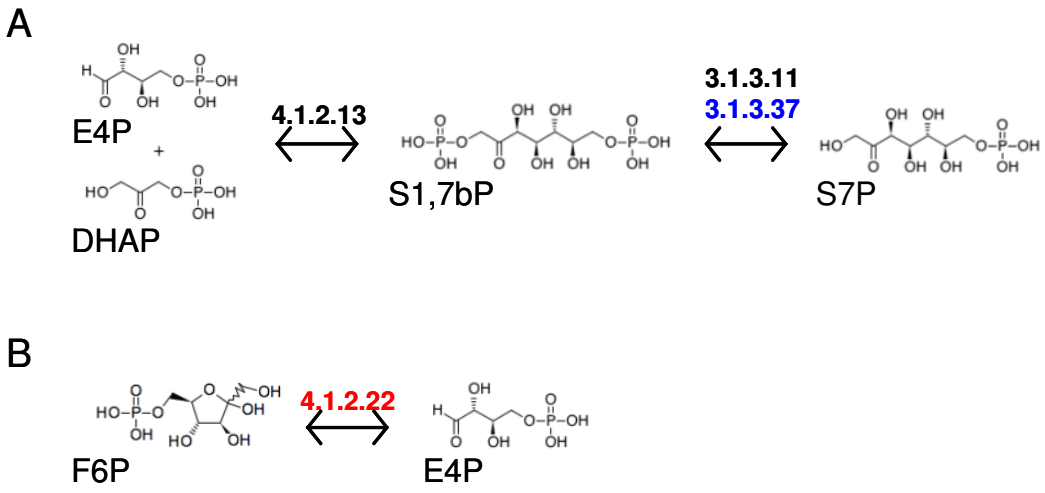


Figure 4-24. Alternative methods of E4P production.

A - Production of S7P from E4P and DHAP. B - Production of E4P from F6P. Only the main reaction pair metabolites (defined by KEGG) are shown. E4P, Erythrose 4-phosphate (C00279); DHAP, Dihydroxyacetone phosphate (C00111); S1,7bP, Sedoheptulose 1,7-bisphosphate (C00447); S7P, Sedoheptulose 7-phosphate (C05382); 4.1.2.13, aldolase; 3.1.3.11, fructose-bisphosphatase; 3.1.3.37, sedoheptulose-bisphosphatase; 4.1.2.22, fructose-6-phosphate phosphoketolase. Enzymes not known to be present in *T. brucei* or *L. major* are shown in red, those present in *T. brucei* but not *L. major* are shown in blue, and those present in both *T. brucei* and *L. major* are shown in black.

Finally, R5P can be converted to nucleic acids via PRPP. It was reported that 10.9% of D-glucose is incorportated into nucleic acids in *L. mexicana* (Maugeri, DA *et al*. 2003); therefore, D-glucose flux through the oxidative branch of the PPP must be at least 10%. Additionally, 56.7% of D-Ribose was incorporated into nucleic acids (Maugeri, DA *et al*. 2003). This would suggest that 43.3% of the D-Ribose has an alternative fate which may include flux through the non-oxidative branch of the PPP to form glycolytic intermediates. It would be unlikely for *Leishmania* to produce excess ribonucleotides if they were not required by the cell.

Interestingly, the concentration of phosphoenolpyruvate (PEP) decreases from 150 nM in the wild type to 20 nM in TKT knockout cell line. Re-expression of TKT in the null background increases the PEP concentration to 70 nM (Figure 4-25). These results suggest that less D-glucose is converted to PEP, although do not

necessarily indicate that the flux of G6P through the glycolytic pathway is reduced.

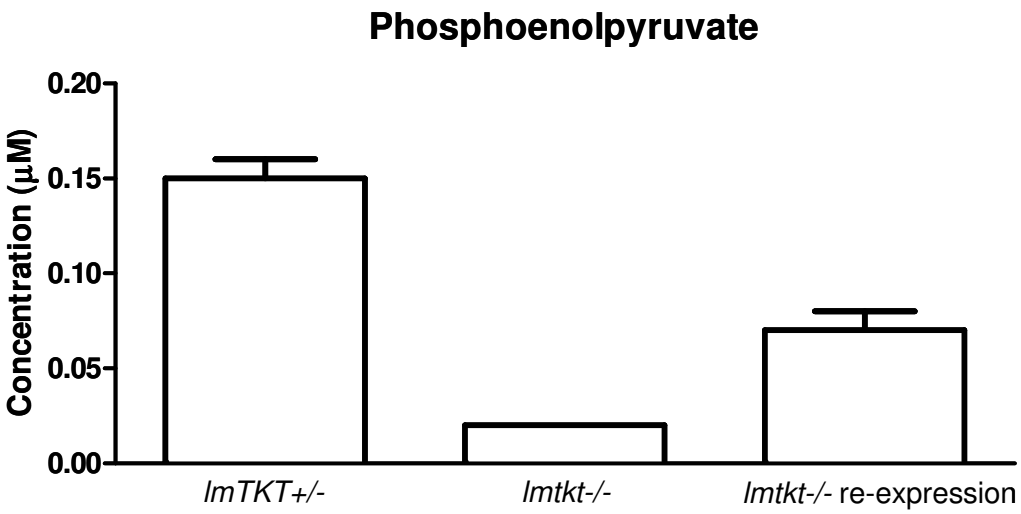


Figure 4-25. A comparison of intracellular concentrations of phosphoenolpyruvate (PEP) in *L. mexicana*.

A proposed model for glucose catabolism in the glycosome is shown in Figure 4-26. Whilst cellular requirements are responsible for the fate of G6P; environmental conditions may be responsible for the fate of R5P. If D-Glucose is unlimited then there is no need to utilise the non-oxidative branch of the PPP to divert flux through the glycolytic pathway, or back through the oxidative branch of the PPP. However, there is an ATP-ADP balance in the glycosome, with the net gain of ATP occurring in the cytosol during the conversion of PEP to pyruvate by pyruvate kinase (PK; EC 2.7.1.40). The phosphorylation of glucose in the glycosome requires 1 mole of ATP per mole of D-glucose, and thus any flux through the oxidative branch of the PPP would consume more ATP than could be replenished by the glycolytic pathway. It has been reported that *L. mexicana* promastigotes are capable of phosphorylating free ribose for the synthesis of nucleotides (Maugeri, DA *et al*. 2003). If ribokinase is able to dephosphorylate R5P, and this reaction occurred in the glycosome, then this would provide a mechanism to re-generate ATP and maintain the glycosomal ATP-ADP balance without utilising the non-oxidative branch of the PPP. If the cells are in a glucose deficient environment, for example in the midgut of a tsetse fly or sandfly, it is advantageous to convert any R5P to glycolytic intermediates to fuel the oxidative branch of the PPP and regenerate NADPH. In a glucose rich

environment, for example the mammalian bloodstream, the trypanosome is able to divert more G6P through the oxidative branch of the PPP if under increased levels of stress.

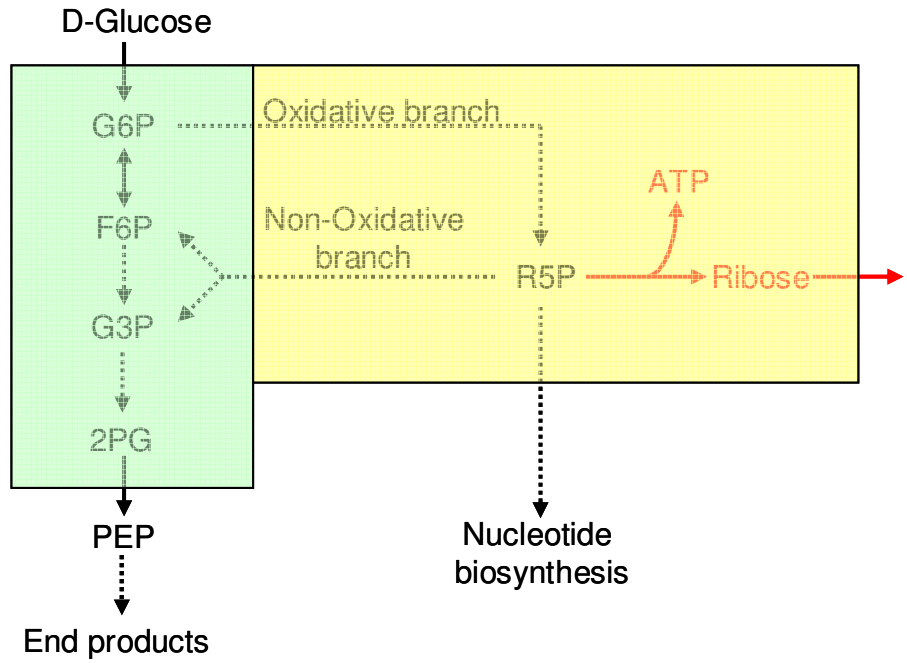


Figure 4-26. Proposed model for glucose catabolism in the glycosome.

The selected metabolites of the glycolytic pathway and the oxidative and non-oxidative branches of the PPP are shown. The solid arrows indicate a single metabolic step and dashed arrows indicate multiple metabolic steps. The green box shows metabolic reactions known to occur in the glycosome and include the first seven steps of glycolysis, whereas the yellow box indicates reactions that may occur in the glycosome based on the presence of a peroxisome targeting signal. It is unknown whether the dephosphorylation of ribose 5-phosphate to form D-ribose and ATP, shown in red, is possible *in vivo*. End products of aerobic glucose catabolism include incompletely oxidised products such as succinate, L-alanine, acetate; the complete oxidisation of glucose results in the production of CO2. Abbreviations - G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; G3P, glyceraldehyde 3-phosphate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate; R5P, ribose 5-phosphate.

Table 4-10. Impact of transketolase deletion in *L. mexicana* on the concentration of intracellular metabolites.

Metabolite	Abbreviation	KEGG ID	*lmTKT+/+*		*lmtkt-/-*		*lmtkt-/-* re-expression	
			mean (µM)	SD	mean (µM)	SD	mean (µM)	SD
6-phosphogluconate	6PG	C00345	0.26	0.04	0.26	0.08	0.03	0.01
ribose 1-phosphate	R1P	C00620	0.35	0.00	0.63	0.06	0.32	0.01
ribose 5-phosphate	R5P	C00117	0.21	0.02	2.40	0.20	0.09	0.03
erythrose 4-phosphate	E4P	C00279	2.87*	0.14*	67.30*	7.78*	2.07*	0.34*
sedoheptulose 7-phosphate	S7P	C05382	11.83†	1.40†	0.83†	0.13†	2.16†	0.31†
phosphoenolpyruvate	PEP	C00074	0.15	0.01	0.02	0.00	0.07	0.01
fructose 6-phosphate	F6P	C05345	0.20	0.01	0.14	0.02	0.05	0.01
fructose 1,6-bisphosphate	F1,6BP	C05378	0.05	0.00	0.03	0.00	0.02	0.01
2-phosphoglycerate 3-phosphoglycerate	2PG 3PG	C00631 C00197	0.56	0.04	0.97	0.06	0.28	0.05
glucose 1-phosphate	G1P	C00103	0.40	0.05	0.24	0.04	0.07	0.01
glucose 6-phosphate	G6P	C00668	1.58	0.09	1.26	0.21	0.45	0.12
glycerol 3-phosphate	Gly3P	C00093	0.39	0.07	0.16	0.01	0.14	0.02
mannose 6-phosphate	M6P	C00275	0.20	0.02	0.05	0.01	0.04	0.00
fumarate		C00122	0.30	0.06	0.20	0.02	0.16	0.02
malate		C00149	1.32	0.09	0.91	0.14	0.88	0.09
succinate		C00042	5.10	0.06	4.15	0.38	12.37	1.09
2-Oxoglutarate	2OG	C00026	0.28	0.01	0.33	0.00	0.19	0.03
cis-aconitate	CA	C00417	0.02	0.01	0.02	0.00	0.02	0.00
citrate		C00158	1.47	0.09	2.31	0.09	1.69	0.13
orotate		C00295	0.01	0.00	0.01	0.00	0.01	0.00

* no calibration curve for E4P, values expressed as area ratio $^{12}C/^{13}C$;†, no standard commercially available for sedoheptulose-7-phosphate, values expressed as area ratio $^{12}C/^{13}C$; SD, standard deviation.

4.8.7 Interaction of TKT with PEX5

In *L. mexicana*, transketolase was shown to have dual localisation between cytosolic and glycosomal compartments (Veitch, NJ *et al*. 2004) (section 4.2.3). However, it is not known how only a proportion of the enzyme localises to the glycosome. One possible explanation involves a conformational change in quaternary structure of the enzyme.

Glycosomal enzymes have a peroxisomal targeting signal (PTS), which interacts with the peroxin receptor on the glycosome surface, to facilitate localisation to the glycosome (Jardim, A *et al*. 2000). There are two types of PTS; a C-terminal PTS-1 which interacts with Pex5, and N-terminal PTS-2 which interacts with Pex7.

The PTS-1 is a carboxy terminal tripeptide motif, generally accepted as SKL, or derivatives with conservative substitutions at any of the three residues (Sommer, JM *et al*. 1992); the consensus sequence has been reported as (S/C/A),(K/R/H),(L/M) (Galland, N *et al*. 2007). For example, in *T. brucei* 427, the TKT PTS-1 motif is SHL, whereas in *L. donovani*, the TKT PTS-1 motif is SKM. The PTS-2 is an amino terminal nonapeptide motif with the following consensus sequence (R/K),(L/V/I),X_5,(H/Q),(L/A), where X denotes any amino acid (Galland, N *et al*. 2007). The majority of glycosomal enzymes are thought to enter the glycosome via Pex5; consequently there is little data on Pex7 and PTS-2 interactions.

The interaction between the human peroxisomal targeting signal receptor 1 (PXR1; GenBank accession ID U19721) (Dodt, G *et al*. 1995) of *HsPex5* and the *HsPST1* recognition motif is shown in Figure 4-27. The equivalent amino acids for the *L. donovani* peroxin 5 *(LdPEX5*; GenBank accession ID AAF67841) are shown in parenthesis.

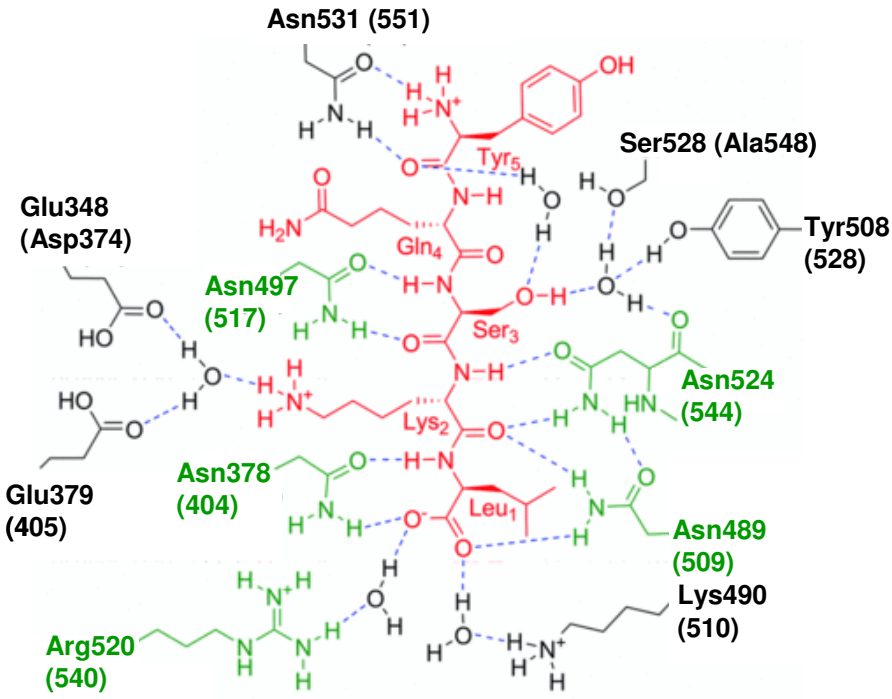


Figure 4-27. Interaction of Pex5 and PTS1.

Residues involved in the interaction of *HsPex5* and PTS1. The PTS1 targeting signal (highlighted in red) is shown as YQSKL (in *L. donovani* this corresponds to SFSKM), hydrogen bonds are highlighted in blue, residues forming invariant interactions are highlighted in green, and those forming variable interactions are highlighted in black. Amino acid residue positions are shown for the interaction of HsPex5 and PTS1; the equivalent residues for *L. donovani* Pex5 (GenBank accession ID AAF67841) are shown in parenthesis. Figure adapted from Lanyon-Hogg T *et al* 2010. Permission to reproduce this image has been granted by John Wiley and Sons.

The crystal structure of *L. mexicana* TKT revealed that the PTS-1 protrudes from a crevice at the surface of the enzyme (Veitch, NJ *et al*. 2004). A conformational change, possibly due to substrate binding, may alter the flexibility of the C-terminal tail, and enable the SKM glycosomal targeting motif to alternate between a buried and exposed (Figure 4-28). A comparative sequence analysis between the *L. donovani* and *L. mexicana* sequences revealed that both enzymes have a serine-lysine-methionine (SKM) glycosomal targeting signal at the C-terminus of the TKT subunit, and the enzymes have a 92.1 % amino acid sequence identity (Supplementary Figure 8-3).

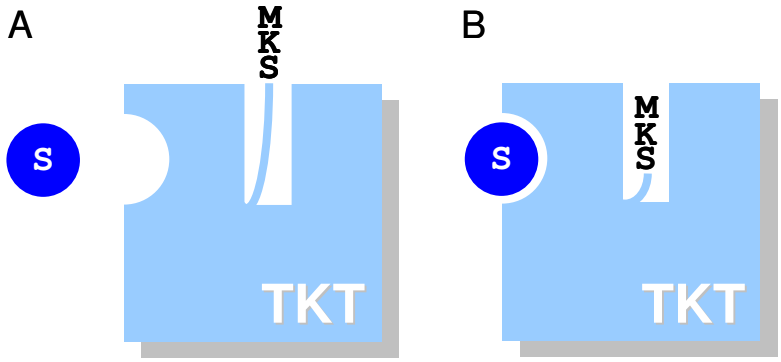


Figure 4-28. The alternating position of the *Leishmania* transketolase (TKT) PTS-1 (SKM) targeting signal.

A – Transketolase enzyme with an 'exposed' PTS-1 (SKM) motif prior to binding of the substrate 'S'. B – Transketolase enzyme with a 'buried' PTS-1 (SKM) motif after binding of the substrate 'S'.

To investigate the association between the glycosomal *LdPEX5* receptor and the *Ld*TKT enzyme, the gene was cloned and expressed in *E. coli* (Clarke A, University of Glasgow). The cells expressing the *Ld*TKT enzyme were sent to Professor Armando Jardim (McGill University, Montreal, Canada). The interaction between *Ld*PEX5 and *Ld*TKT was investigated in the presence of a range of substrates. As the concentration of transketolase increases, the absorbance measurements increase indicating an increase in interaction between transketolase and PEX5 (Figure 4-29). However, in the presence of ribose 5-phosphate there is an increase in the dissociation constant for the TKT-*Ld*PEX5 interaction, and consequently, there is a decrease in the quantity of TKT binding to the *Ld*PEX5. This does not occur in the presence of either F6P or Xu5P, unless R5P is also present. These results suggest that the in the presence of R5P there is a conformational change in TKT enzyme, and the SKM targeting signal is no longer exposed and able to interact with PEX5.

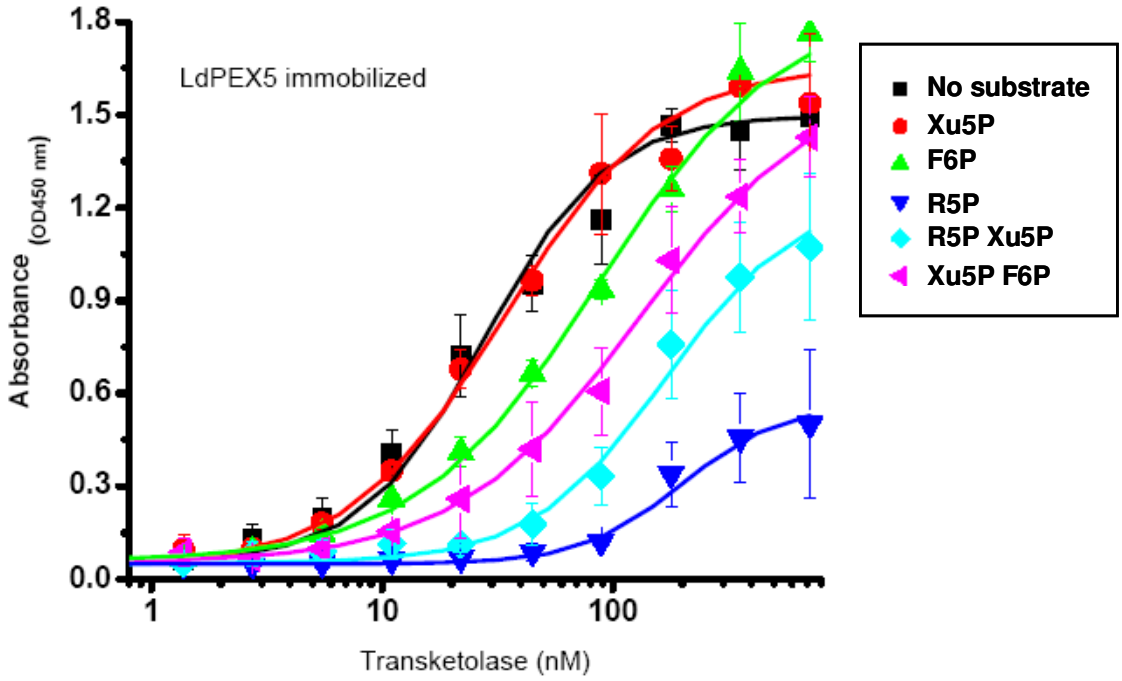


Figure 4-29. Interaction of *Ld*TKT with *Ld*PEX5.

Abbreviations; *LdPEX5*, *Leishmania donovani* peroxin 5; Xu5P, Xylulose 5-phosphate; F6P, Fructose 6-phosphate; R5P, Ribose 5-phosphate.

These preliminary results indicate a decrease in interaction between the PEX5 receptor and the *L. donovani* PTS-1 in the presence of R5P. This suggests that substrate availability; in this case R5P, has reduced the accessibility of the PTS-1 motif to the PEX5 receptor, and consequently, might influence the sub-cellular localisation of TKT. This hypothesis could be investigated by generating a mutant enzyme which has additional amino acid residues prior to the PTS-1 motif, and is therefore unable to retract the PTS-1 in response to a conformational change. It would also be useful to create a tkt enzyme with no PTS-1 motif as a control.

4.9 Discussion

To investigate whether transketolase (LmjF24.2060, EC 2.2.1.1), a key enzyme in the non-oxidative branch of the pentose phosphate pathway, was essential in *L. mexicana* promastigotes, the gene was targeted for deletion.

The successful generation of transketolase null cell line indicates that the enzyme is non-essential in *L mexicana* promastigotes, when grown in a rich growth medium. There was no significant reduction in growth rate between the tkt null cells and the wild type cells under standard laboratory conditions, although the maximum cell density in the knockout cells is slightly lower than the wild type.

The *lmtkt-/-* cells were twice as sensitive to pentamidine compared to the wild type cells, although there was almost no difference for methylene blue. It would appear that the absence of tkt has no effect on the regeneration of NAPDH (in contrast to the results observed by Stoffel SA). It may be of more interest to investigate glucose consumption in tkt knockout cells that are subject to oxidative stress. However, this would only be of interest if glucose was in limiting supply, and that the non-oxidative branch of the PPP was essential to return metabolites to the glycolytic pathway. In the case of *T. brucei* BSF cells, glucose is in high abundance, and as a result the cells can increase the uptake of glucose to meet NADPH demands, without jeopardising ATP production via glycolysis. This would explain why no tkt enzyme was detected in BSF trypanosomes (Cronin, CN *et al*. 1989).

The pentose phosphate pathway functions to produce R5P for nucleotide biosynthesis and regenerate NADPH to defend against oxidative stress (Barrett, MP 1997). There was an accumulation of R5P in the knockout cell line, indicating that a functional transketolase enzyme, and thus the non-oxidative branch of the PPP, is responsible for metabolising a significant proportion of the R5P. The elevated levels of E4P would suggest that an alternative method of E4P production, independent of transketolase, is functional in *L. mexicana*, although the exact biochemical reactions remain uknown.

Interestingly, the *lmtkt-/-* promastigotes do not acidify the medium to the same extent as the wild type cell line. Analysis of the spent medium indicated that

there was no significant difference in the quantity of glucose consumed per cell, and thus the cells undergo a metabolic shift in the absence of transketolase. It would appear that a functional non-oxidative branch of the PPP is required to recycle phosphorylated sugars, and return them to the glycolytic pathway. If transketolase is absent, then any flux through the oxidative branch of the PPP is unable to return to the glycolytic pathway. This may account for the reduced acidification of the medium, as less glucose is converted to organic acids for excretion. Instead, intracellular levels of R5P and R1P increase.

There are several logical experiments which follow up from the initial observations regarding the colour change of the medium, presence of E4P, and the theory of a hidden or exposed PTS-1 targeting signal.

1. **Incubate cells in C13 labelled D-glucose**

In the absence of transketolase there remains a question as to what happens with the Ru5P, the end product of the oxidative branch of the PPP. Is there any mechanism by which flux can return to the glycolytic pathway? The classical pentose phosphate pathway would indicate not. R5P produced could be incorporated into nucleic acids or excreted as D-Ribose, both of which could be detected if labelled D-glucose was used. It is also possible that in the presence of methylene blue, the flux through the oxidative branch of the PPP would increase to meet the cellular NADPH demand.

2. **Incubate cells in C13 labelled D-ribose.**

D-ribose is converted to R5P, and, in the absence of TKT, cannot continue through the non-oxidative branch of the PPP. If D-ribose is converted to intermediate metabolites of the non-oxidative branch of the PPP or the glycolytic pathway, it would indicate there are additional enzymes responsible for these metabolic changes.

3. **Investigate transaldolase activity in the wild type and ΔTKT null cells.**

Typically, E4P is produced when transaldolase transfers a 3-carbon subunit from S7P to GA3P. If transaldolase activity in the tkt null cells is detected, then this would imply that there is an additional source of S7P. If there is no transaldolase

activity, then it should be possible to generate a translaldolase knockout in the tkt null cell line. This would confirm a source of E4P independent of the classical non-oxidative branch of the PPP.

4. Introduce a mutant enzyme into the TKT null background

One of the most useful aspects of generating a knockout cell line is the ability to introduce a mutant copy of the gene back into the cell, and investigate the effects of that mutant enzyme. Initial studies could focus on the comparing metabolic profiles and enzyme localisation studies if changes are made to the PTS-1 targeting signal. If the PTS-1 targeting signal is capable of alternating between a buried and exposed state, then it should be possible to introduce several glycine residues prior to the PTS1. This would ensure that the targeting signal was always exposed, regardless of the substrates available. Furthermore, the PTS-1 targeting signal could be deleted to prevent glycosomal localisation. These mutant proteins could be introduced into the *lmtkt-/-* cell line; metabolic profiles and TKT localisation could be compared.

5 Pentamidine resistance in *L. mexicana*

5.1 Introduction

Aromatic diamidines (AD), such as pentamidine isethionate, have been extensively studied and shown to have potent activity against a wide range of microorganisms, including the protozoan parasites of the *Trypanosoma* and *Leishmania spp* (Bray, PG *et al*. 2003).

Pentamidine isethionate, first synthesised in the 1930's (Sands, M *et al*. 1985), has been used in the treatment of the first stage of *Trypanosoma brucei gambiense* infections. However, the drug is less useful against late stage infections; pentamidine does not cross the BBB efficiently, and consequently, does not accumulate in the CNS to sufficient levels (Bray, PG *et al*. 2003). Pentamidine is less useful in the treatment of *Trypanosoma brucei rhodeseinse* infections, as invasion of the CNS is rapid in comparison to *T.b. gambiense* infections. There is no widespread resistance to pentamidine in the field (Bray, PG *et al*. 2003), and in addition, was found to be effective against cell lines resistant to arsenicals (Van Hoof, LM 1947). The drug has also been used in the treatment of both visceral and mucocutaneous leishmaniasis since 1952, as an alternative to pentavalent antimonials which for many years were used in the first line of defence (Sands, M *et al*. 1985; Becker, I *et al*. 1999; Croft, SL *et al*. 2006). Antimony resistant *Leishmania mexicana amazonensis* (WR 227) and *L. donovani* (WR 271) amastigotes are susceptible to concentrations of pentamidine that can be achieved in human serum (Berman, JD 1982).

Whilst pentamidine is effective in the treatment of HAT, the drug does pose several problems; these include toxicity issues, high production costs, distribution difficulties, and the fact the drug can only be administered parenterally, either intramuscularly or intravenously. Pentamdine and other aromatic diamidine compounds are cationic, and thus are not effective by oral administration.

5.1.1 Cellular target of diamidines

The definitive cellular target and mode of action of the AD compounds remains unknown (Paine, MF *et al*. 2010); however, several theories have been postulated.

There have been many observations that pentamidine accumulates in the mitochondrion (Ludewig, G *et al*. 1994; Lanteri, CA *et al*. 2004), binding to the kinetoplast DNA (kDNA). Disruption of the kDNA network is evident with an increase in free minicircles, indicative of inhibition of mitochondrial topoisomerase II (Shapiro, TA 1993).

The kDNA has a higher adenine and thymine content than nuclear DNA which might explain a preferential binding to this organelle given the higher binding efficiency of diamidines for the minor groove of AT rich DNA. It has been demonstrated that mice infected with a dyskinetoplast strain of trypanosomes do not survive when treated with berenil. However, mice infected with a wild type strain are cured upon the induction of berenil treatment, thus indicating that the presence of an intact kinetoplast is an essential target for diamidine action. (Agbe, A and Yielding, KL 1995). A dyskinetoplastid strain of T. brucei was approximately 4-fold more resistant to pentamidine when compared to the isogenic parental cell line (Gould MK, personal communication) indicating that whilst the kinetoplast is a target for pentamidine, other cellular targets are likely to play a prominent role.

5.1.2 Biological target of diamidines

Crystalisation studies between AD compounds and the DNA duplex d(CGCGAATTCGCG) indicates these compounds bind with high affinity to a central AATT sequence within the minor groove of the DNA through a series of non-covalent and non-intercalative interactions (Hajduk, SL 1978; Nguyen, B *et al*. 2002). There is also a clear correlation between the molecular structure and biological activity of these compounds, with the molecular curvature of the compound critical for recognition. Ideally the amidine groups should be located at the para- positions of the phenol rings, as is the case with pentamidine and furamidine (also know as DB75, an aromatic diamidine that in its methoxy prodrug form DB289 entered clinical trials against human African trypanosomiasis (Paine MF *et al*., 2010)). Conversely, meta- substituted analogues such as DB359 have reduced anti parasitic properties as the compound does not complement the curvature of the DNA, and consequently binds more weakly to the central AATT sequence. The amidine residues are able to form H-bonds bases at the bottom of the minor groove (Nguyen, B *et al*. 2002). Further stabilisation of the complex is achieved via Wan der Waals interactions between the groove wall and

the diamidine compound, and electrostatic interactions between anionic charges on the DNA backbone and the amidine charges (Paine, MF *et al*. 2010).

5.2 Pentamidine and trypanosomatids

In recent years, a number of pentamidine resistant trypanosome (Berger, BJ *et al*. 1995; Bridges, DJ *et al*. 2007) and pentamidine resistant *Leishmania* (Basselin, M *et al*. 1997; Basselin, M *et al*. 2002; Mukherjee, A *et al*. 2006) cell lines have been generated *in vitro*. Whilst there may be significant differences between laboratory-derived drug resistant cells and field isolates, the ability to generate drug resistant lines in the laboratory and compare them to isogenic progenitor cell lines has been crucial in developing our understanding of resistance mechanisms. For example, protozoan parasites are unable to synthesise purines *de novo*, and consequently obtain these metabolites from their hosts (Bellofatto, V 2007). Therefore, research has focused on the transport mechanisms in which pentamidine, and other related diamidine compounds, enter the cells.

5.2.1 Pentamidine transport in trypanosomes

Transporters have been identified as the mechanism by which AD compounds cross biological membranes; the rate of diffusion is slow, and thus it takes several days for these compounds to kill parasites *in vitro*. The aminopurine transporter TbAT1/P2 (Tb927.5.286b) (Carter, NS and Fairlamb, AH 1993; Maser, P *et al*. 1999) recognises compounds with adenine-, adenosine-, amidine-, and melamine moieties (Paine, MF *et al*. 2010), and is the main transporter for diamidine compounds, responsible for approximately half of the pentamidine uptake in trypanosomes (de Koning, HP 2008). Loss of the P2 transporter confers increased resistance to pentamidine (Matovu, E *et al*. 2003), DB75 (Lanteri, CA *et al*. 2006), and diminazene (Barrett, MP *et al*. 1995). However, the loss in sensitivity towards these diamidine compounds is minor, indicating the existence of multiple pentamidine transporters (de Koning, HP 2001; de Koning, HP and Jarvis, SM 2001). These include the High-affinity pentamidine transporter (HAPT1), and the Low-affinity pentamidine transporter (LAPT1) (de Koning, HP and Jarvis, SM 2001).

High level resistance to pentamidine was achieved through resistance induction in a cell line lacking the TbAT1 gene. Biochemical characterisation of this cell line, designated TbAT1 -/- B48, indicated the loss of the HAPT, however the LAPT activity remaining unchanged (Bridges, DJ *et al*. 2007). The genes responsible for encoding HAPT and LAPT have not currently been identified at the molecular level.

The existence of multiple transporters involved in pentamidine uptake may offer an explanation as to why clinical resistance is not widespread; despite extensive use, it remains the drug of choice in more than 90% of cases.

5.2.2 Pentamidine transport and efflux in Leishmania

Whilst pentamidine enters trypanosomes via an aminopurine transporter, P2, and other transporters whose substrate specificity is not known, this is not the case in *Leishmania*. In *L. mexicana*, pentamidine is transported into the cells by an unknown pentamidine transporter, designated LmexPT1, with similar affinity to that of the LAPT in trypanosomes. The drug rapidly accumulates in the mitochondrion via an unknown carrier or transporter, with significant quantities binding to the kDNA within minutes (Basselin, M *et al*. 2002). However, it has also been postulated that polyamine transporters are responsible for uptake in both *L. donovani* and *L. amazonesis* as pentamidine competively inhibits polyamine uptake (Basselin, M *et al*. 1996). However, polyamines do not inhibit pentamidine uptake in *L. mexicana* ruling these carriers out (Basselin, M *et al*. 2002).

A common transport system does exist for the aromatic diamidine compounds in *L. mexicana* (Basselin, M *et al*. 2002) and *L. donovani* (Mukherjee, A *et al*. 2006), as demonstrated by the competitive inhibition of pentamidine transport using several aromatic diamdine compounds including propamidine, stilbamidine, berenil and DAPI.

Interestingly, pentamidine is a competitive inhibitor of arginine transport in *L. mexicana* promastigotes and amastigotes; however conversely, arginine does not inhibit pentamidine transport. This implies that whilst pentamidine interacts with the arginine transporter, it does not enter the parasite via this transporter (Basselin, M *et al*. 2002). Similarly, pentamidine was found to be a non-

competitive inhibitor of both putrescine and spermidine transport in various *Leishmania* species (Basselin, M *et al*. 2000), although does not enter the parasites via these transporters.

To determine the specificity of the pentamidine transporter, a total of 52 commonly occurring *Leishmania* metabolites, including polyamines, amino acids, nucleosides, and nucleobases, were tested to see if they inhibited uptake of pentamidine (Basselin, M *et al*. 2002). None of these were shown to inhibit pentamidine transport. The only inhibition observed was with two peptides rich with basic amino acids, various inorganic divalent cations, and Taurocholate (C05122). Interestingly, and in contrast to the situation in trypanosomes, pentamidine was not inhibited by nucleosides and nucleobases, despite considerable homology between the trypanosome nucleoside transporter TbAT1 and *L. donovani* nucleoside transporters LdNT1.1 and LdNT1.2.

A comparison between drug resistant and drug sensitive *L. mexicana* promastigotes and amastigotes indicated that there was no difference in the affinity for pentamidine or the rates of uptake between the two cell lines (Basselin, M *et al*. 2002). This suggests that in these particular cell lines, and in contrast to trypanosomes, there was no genetic alteration to the transport system responsible for uptake at the plasma membrane, and pentamidine enters both wild type and resistant cells via the same plasma membrane transporter.

DAPI, a fluorescent analogue of pentamidine, binds to the kDNA in wild type cells, but not pentamidine resistant cells (Basselin, M *et al*. 2002). However, if the resistant cells are treated with digitonin, a compound which permeabilises cell membranes, then DAPI is able to accumulate in the mitochondrion and bind to kDNA in a manner similar to that in wild type cells. Therefore, the resistant cells have a diminished ability to transport pentamidine and similar compounds into the mitochondrion and kinetoplast, as opposed to a reduction in the ability of these compounds to bind to the kDNA.

In pentamdine resistant *L. mexicana* promastigotes there is a decreased mitochondrial membrane potential (MMP), resulting in less pentamidine accumulating in the mitochondrion and kinetoplast (Basselin, M *et al*. 2002). Consequently, the majority of pentamidine remains in the cytosol, or as

membrane-associated pentamidine, both of which are removed from the cell by efflux pumps. This model is depicted in Figure 5-1.

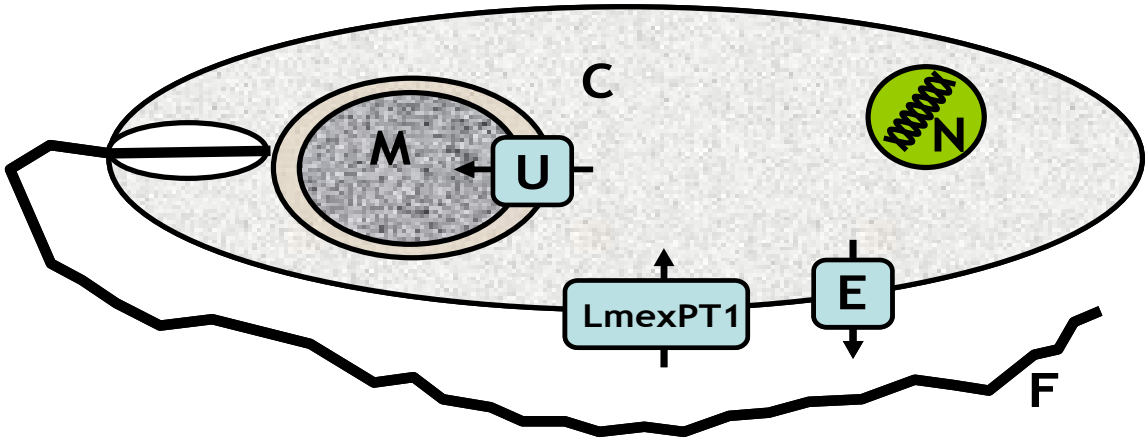


Figure 5-1. Model for pentamidine uptake in *Leishmania mexicana* promastigotes.

Pentamidine is transported to the cytoplasm by a pentamidine transporter, with similar affinity to the Trypanosome LAPT1, and is concentrated in the mitochondrion by an unknown carrier or channel. Abbreviations: M, mitochondrion; C, cytoplasm; N, nucleus; F, flagellum; LmexPT1, *Leishmania mexicana* pentamidine transporter 1; E, efflux pump; U, unknown carrier or transporter.

An energy dependent efflux pump is thought to be responsible for removing membrane associated and cytosolic pentamidine from *L. mexicana* (Basselin, M *et al.* 2002). The resistance phenotype has no influence on the rate of pentamidine efflux, with similar rates reported for both the wild type and resistant cells, implying that the efflux system cannot explain the resistance phenotype. The efflux pump can be inhibited with Trifluoperazine (TFP; D08636), PCP, and Verapamil (D02356), all of which are known inhibitors of the P-glycoprotein transporter. Consequently, similar levels of pentamidine accumulate in the cytoplasm of resistant cells as would normally accumulate in the wild type cells (Basselin, M *et al.* 2002). A similar study in *L. donovani* indicated that pentamidine resistant cells in the presence of these P-glycoprotein inhibitors accumulated pentamidine to similar levels as cells in the absence of inhibitors, thus suggesting that the *L. donovani* pentamidine efflux system is not as prominent as in *L. mexicana* (Mukherjee, A *et al.* 2006).

5.2.3 Accumulation of pentamidine in trypanosomatids

The effectiveness of AD has been attributed to selective accumulation by the pathogen, before the host cell. Pentamidine, and other aromatic diamidine

compounds, rapidly accumulate to toxic levels in trypanosomatids (millimolar), concentrating in the nucleus and mitochondrion. Whilst these compounds bind to DNA in both the nucleus and kinetoplast, the cationic nature of these drugs suggests that pentamidine binds to numerous proteins in the cell, and thus interferes with numerous cellular processes.

Pentamidine resistant trypanosomes accumulated pentamidine at a slower rate than pentamidine sensitive trypanosomes, although both were equally as sensitive to the same millimolar concentrations (Damper, D and Patton, CL 1976). The deletion of the P2 transporter (Matovu, E *et al*. 2003) and loss of the HAPT through stepwise exposure to pentamidine (Bridges, DJ *et al*. 2007) in bloodstream form *T. brucei* leads to a reduction in the uptake of pentamidine and related diamidine compounds. These data imply that the mode of resistance in the cell lines investigated was due to alterations to the transporters on the plasma membrane, and consequently leads to a reduction in drug accumulation.

In *Leishmania* the situation is more complex, as the rate of uptake remains similar for both wild type and pentamidine resistant parasites. However, pentamidine resistant *L. mexicana* (Basselin, M *et al*. 2002) and pentamidine resistant *L. donovani* (Mukherjee, A *et al*. 2006) have a reduced mitochondrial membrane potential (MMP), demonstrated using rhodamine-123 (C11190). Therefore, less pentamidine accumulates in the mitochondrion, with the majority remaining in the cytosol. Consequently, cytosolic or membrane-associated pentamidine is removed from the cell by efflux pumps (section 5.2.2).

5.2.4 Summary of pentamidine and trypanosomatids

Whilst pentamidine and related diamidine compounds target the mitochondrion, binding to the kDNA, in both *Leishmania* and trypanosomes, the mechanisms of drug resistance are distinctly different. In trypanosomes, alterations to various nucleoside and nucleobase transporters on the plasma membrane causes a reduction in drug uptake resulting in less pentamidine accumulating in the cell. In *Leishmania* there is no difference in drug uptake between wild type and pentamidine resistant strains, however a reduced mitochondrial membrane potential causes less pentamidine to accumulate in the mitochondrion, with efflux pumps removing the cytosolic pentamidine.

5.2.5 Pentamidine in Yeast

In *Saccharomyces cerevisiae*, low concentrations of pentamidine inhibit growth on non-fermentable energy sources; whereas high concentrations induce a 'petite' phenotype, typical of mutations in the mitochondrial genome (Ludewig, G *et al*. 1994). DB75, a structural analogue of pentamidine, is fluorescent at high concentrations. In both yeast and trypanosomes, DB75 localises to, and accumulates in, the mitochondrion; and as a result inhibits mitochondrial function (Lanteri, CA *et al*. 2004; Lanteri, CA *et al*. 2008). These data suggest that the mitochondrion is the primary cellular target for the diamidine compounds. Additionally, pentamidine has been reported to inhibit ATP-dependent DNA topoisomerases (Ludewig, G *et al*. 1994) and mitochondrial topoisomerase II in trypanosomes (Shapiro, TA 1993).

A pentamidine resistance gene, mitochondrial integral inner membrane protein PNT1, was identified in *S. cerevisiae* (Ludewig, G *et al*. 1994). The copy number of this gene was linked with pentamidine resistance. Cells in which the PNT1 gene had been disrupted were hypersensitive to pentamidine, whereas those expressing more copies of the gene were hyposensitive to pentamidine (Ludewig, G *et al*. 1994).

Pentamidine resistant *Leishmania mexicana* have a reduced mitochondrial membrane potential (Basselin, M *et al*. 2002), and therefore less pentamidine accumulates in the mitochondrion, with the majority remaining in the cytosol. Cytosolic or membrane-associated pentamidine is removed from the cell by efflux pumps.

5.3 Summary of L. donovani pentamidine resistance

Pentamidine resistant *Leishmania donovani* promastigotes were generated *in vitro* by gradually increasing drug pressure until cells were capable of growth in 8 μM pentamidine (Mukherjee, A *et al*. 2006). This pentamidine resistant cell line was designated R8.

In that study, the pentamidine resistant cells were cross-resistant to other diamidine compounds including propamidine, stilbamidine, berenil, DAPI, and DB99; however, there was no cross-resistance to substrates of multidrug resistant proteins, in contrast to *L. mexicana* (Basselin, M *et al*. 2002), or antimonial compounds.

Also in that study, wild type parasites treated with pentamidine and the pentamide resistant line were shown to have a reduced ability to take up Rhodamine -123 in comparison to untreated wild type cells, thus indicating a decrease in the MMP.

Furthermore, short term radio-labelled uptake assays did not reveal any difference in pentamidine uptake between the wild type and resistant lines. However a 5.4 fold difference was evident over three hours. The intracellular concentration of wild type and resistant cells were 3.8 μM and 0.77 μM respectively.

Whilst no diamidine transporter has been identified at the molecular level, there is biochemical evidence of two pentamidine transporters in *L. donovani* promastigotes. One of which is a saturable high affinity transporter present at the plasma membrane, and the second is a non-saturable low affinity transporter, which is not as prominent in the resistant line, at the mitochondrial membrane. In contrast to *L. mexicana*, there is no evidence of pentamidine efflux in *L. donovani*.

A proteomic analysis was carried out on the R8 line compared with its wild type precursor (Poonam Tewary, NCI NIH, Washington D.C.). Using Difference gel electrophresis (DiGE), several modulated proteins were identified. Interestingly, the glycolytic enzymes phosphoglucose isomerise (PGI; EC 5.3.1.9) and enolase (EC 4.2.1.11) were up-regulated 3.55-fold and 6.67-fold respectively, in the

resistant cell line, suggesting an increased demand for glucose. However, the 5.97-fold down-regulation of pyruvate kinase (PK; EC 2.7.1.40) in the resistant cell line indicates a reduction in the conversion of phosphoenolpyruvate (PEP; C00074) to pyruvate. In the proposed model for *Leishmania* central carbon metabolism, PEP is first converted to pyruvate, before re-entering the glycosome where it is converted to succinate (Rosenzweig, D *et al*. 2008). Down regulation of PK would result in a reduction in glycolytic flux as this step would form a 'bottle neck' in the pathway. However, the model for procylic form trypanosomes is different; PEP is thought to re-enter the glycosome, where it is converted to succinate (Ebikeme, C *et al*. 2010). Consequently, there would be a down regulation in acetate production, compensated by an increase in succinate production. The accumulation of pentamidine in the cytosol of resistant parasites could have an influence on pyruvate kinase levels, given the enzyme is cytosolic.

Table 5-1. Proteins down-regulated in the *L. donovani* pentamidine resistant (R8) cells.

Protein	Protein Accesion ID	Average Ratio
Pyruvate kinase	LmjF35.0030	5.97
Beta tubulin	LmjF33.0798	4.83
Aspartate aminotransferase	LmjF35.0820	3.19
Proteosome reg subunit 5	LmjF22.0620	3.02

Peptides were matched against the L. major protein database. Only proteins greater than a three-fold difference are shown.

Table 5-2. Proteins up-regulated in the *L. donovani* pentamidine resistant (R8) cells.

Protein	GeneDB Accesion ID	Average Ratio
NADP dependent alcohol dehydrogenase	LmjF23.0360	27.04
Methionine synthase	LmjF31.0010	13.58
Carboxypeptidase	LmjF33.2540	9.58
Transketolase	LmjF24.2060	8.19
Enolase	LmjF14.1160	6.67
Hypothetical protein, conserved	LmjF25.2010	6.57
Amino peptidase	LmjF11.0630	5.95
Proteosome beta 1 subunit	LmjF12.0030	5.73
ATP dependent dead box helicase	LmjF35.0370	4.44
Carboxypeptidase	LmjF33.2540	4.38
Aspartyl tRNA synthatase	LmjF30.0460	4.26
Glucose-6-phosphate isomerase	LmjF12.0530	3.55
t complex prt	LmjF23.1220	3.42
thiol dependent reductase	LmjF33.0240	3.17
proteosome reg non ATPase subunit 5	LmjF21.0760	3.06
fumarate hydratase	LmjF29.1960	3.04

Peptides were matched against the L. major protein database. Only proteins greater than a three-fold difference are shown.

The comparative proteomic analysis of the wild type and pentamidine resistant (R8) *L. donovani* cell lines indicated a significant number of modulated proteins. However, prior to further proteomic analysis, and subsequent gene knockout investigations and metabolomic analysis, it was necessary to confirm that these cell lines were isogenic to confirm that the adaptations were due to pentamidine resistance, and not to differences between species.

5.4 Species Identification

Leishmania species are morphologically very similar, and are therefore difficult to distinguish from one another by microscopic examination. It is important to distinguish medically relevant species of *Leishmania* from one another so the most appropriate treatment can be administered; especially in geographical areas where multiple *Leishmania* species co-exist. Furthermore, identification based on clinical symptoms is not always possible as some species may cause multiple clinical manifestations of the disease. A diagnostic PCR and restriction digest was developed to distinguish between medically relevant species (Schonian, G *et al*. 2003). Although this approach was intended for the identification of field isolates; it is also a suitable method for the identification of laboratory species.

Amplification of the ribosomal ITS1 located between the genes coding for the ssu and the 5.8S rRNA (Schonian, G *et al*. 2003) produces a PCR product that varies in size and nucleotide sequence depending on the *Leishmania* species (Davila, AM and Momen, H 2000). Subsequent digestion of these PCR products using the restriction endonuclease HaeIII produces a RFLP pattern which can be used to distinguish between clinically relevant species as illustrated in Figure 5-2A (Schonian, G *et al*. 2003). This approach was used to genotype the wild type and pentamidine resistant (R8) *L. donovani* cell lines

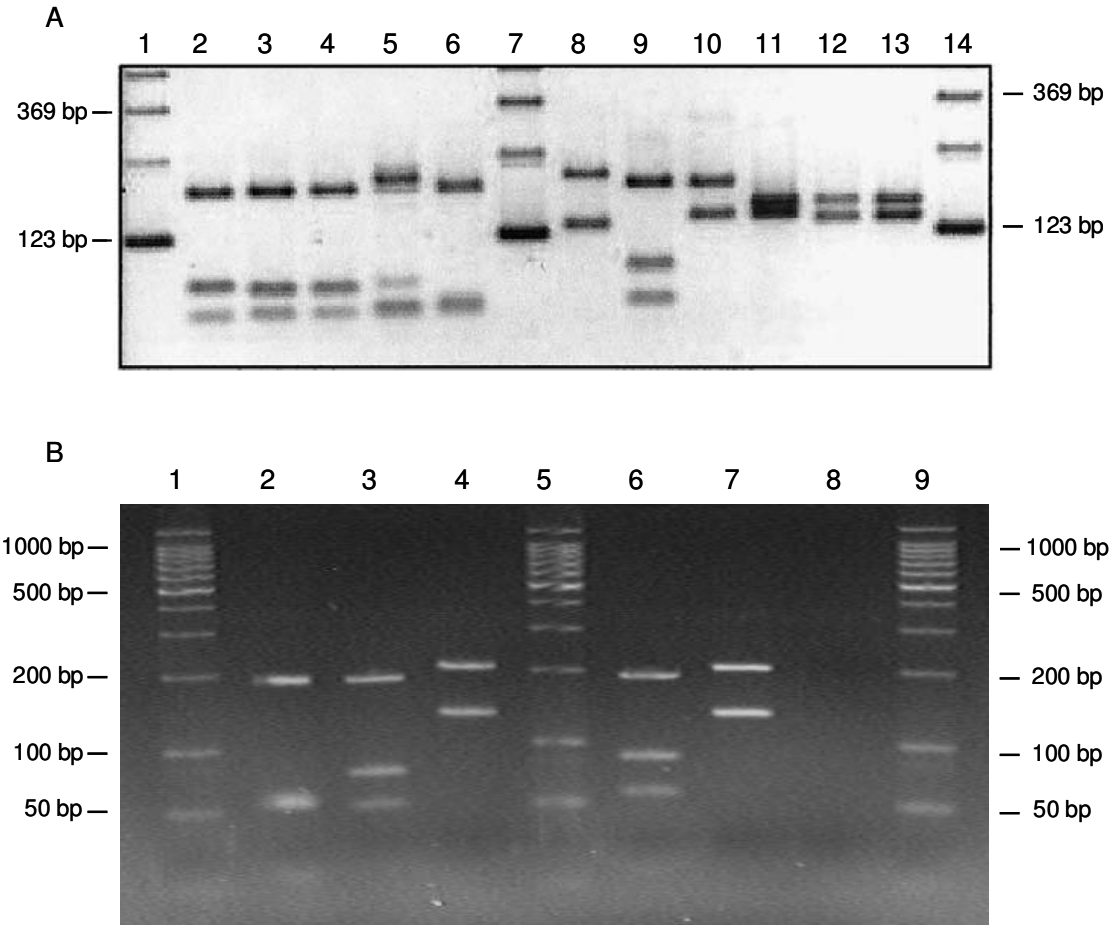


Figure 5-2. *Leishmania* Species Identification.

Digestion of amplified ITS1 regions of different *Leishmania* species with the restriction endonuclease HaeIII. A – Restriction digest profiles of medically relevant *Leishmania spp*. Reproduced from Schönian G *et al* 2003. Permission to reproduce this image has been granted by Elsevier. B – Restriction digest profiles of the *L. donovani* wild type (wt) and pentamidine resistant (R8) cell lines, and other species used in the laboratory. The lanes are as follows, A; 1, 123 bp ladder; 2, *L. donovani* MHOM/IN/80/DD8; 3, *L. infantum* MHOM/TN/80/IPT1; 4, *L. chagasi* MHOM/PA/79/WR317; 5, *L. aethiopica* MHOM/ET/94/Gere; 6, *L. tropica* MHOM/SU/74/SAF-K27; 7, 100 bp ladder; 8, *L. major* MHOM/SU/73/5ASKH; 9, *L. mexicana* MHOM/MX/85/SOLIS; 10, *L. amazonensis* MHOM/BZ/73/M2269; 11, *L. braziliensis* MHOM/BZ/75/M2903; 12, *L. guyanensis* MHOM/BR/75/M4147; 13, *L. panamensis* MHOM/CR/87/NEL3; 14, 123 bp ladder. B; 1, 100 bp ladder; 2, *L. donovani* MHOM/IN/1983/AG83 wild type; 3, *L. donovani* MHOM/IN/1983/AG83 pentamidine resistant (R8); 4, *L. infantum* JPCM5 MCAN/ES/98/LLM-877; 5, 100 bp ladder; 6, *L. mexicana* MNYC/BZ/62/M379; 7, *L. major* LV39c5 MRHO/SU/1959/Neal P; 8; negative control; 9, 100 bp ladder. The 100 bp marker was supplied by Promega.

The results of the species identification test indicate that the wild type *L. donovani* AG83 and the derived pentamidine resistant (R8) cell lines currently present in the laboratory were not isogenic as they did not produce identical HaeIII restriction digest profiles (Figure 5-2B – lanes 2 and 3). Subsequently, a

comparison with the HaeIII restriction digest profiles published by Schönian G. et al 2003 suggested that the *L. donovani* AG83 wild type cell line may be *L. tropica* (Figure 5-2A – lane 6). Furthermore, the *L. infantum* control gDNA used in the experiment (Figure 5-2B – lane 4) has the same restriction digest profile as *L. major* (Figure 5-2A - lane 8; Figure 5-2B – lane 7). These results highlight the importance of regularly testing parasite cell lines to minimise the chances of any mistakes occurring. The RFLP profiles of the *L. donovani (*Figure 5-2A – lane 2), *L. infantum* (Figure 5-2A – lane 3), and *L. mexicana* (Figure 5-2A – lane 9) are similar, but the following differences are observed. The two lower bands of the *L. mexicana* profile run slightly higher than the corresponding bands for the *L. donovani* and *L. infantum*; additionally, there difference in size between the two bands is greater in *L. mexicana*. Given the similarities between these cell lines it is essential to use the appropriate experimental controls to discriminate between these species which are routinely used in the Wellcome Trust Centre for Molecular Parasitology, Glasgow.

As the *L. donovani* wild type and pentamidine resistant cell lines kept at the University of Glasgow did not appear to be isogenic, it was necessary to investigate the genotype of the original cell lines stored at Jawaharlal Nehru University, New Dehli, India. The analysis of theses cell lines, courtesy of Prof. Rentala Madhubala (School of Life Sciences, Jawaharlal Nehru University, New Delhi, India), was inconclusive in determining that these lines were *L. donovani* and isogenic. There appeared to be two fragments in the wild type sample (Figure 5-3 – lane 1), corresponding to ~180 bp and ~ 100 bp; however, an additional fragment of ~250 bp was present in the resistant line (Figure 5-3 – lane 3). It is possible that this fragment could arise from the incomplete restriction digest of ITS1. The expected sizes of the fragments for the *L. donovani* are approximately 200 bp, 90 bp, and 60 bp. The absence of a definitive 3rd fragment in both wild type AG83 and the R8 are unlikely to be due to PCR error, as the ITS1 is a tandem repeat region, and therefore should amplify easily. Sequence analysis of the ITS1 region may provide a clear indication as to whether these particular cell lines are *L. donovani*, or another *Leishmania spp*.

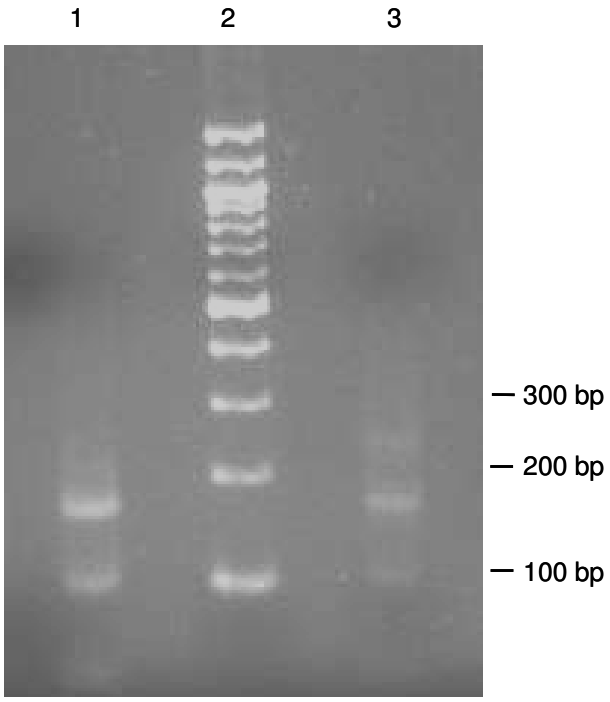


Figure 5-3. *L. donovani* species identification.

Digestion of amplified ribosomal internal transcribed spacer 1 (ITS 1) regions of the *L. donovani* AG83 (MHOM/IN/1983/AG83) wild type and R8 cell lines with the restriction endonuclease HaeIII. The lanes are as follows; 1, *L. donovani* AG83 wild type; 2, 100 bp ladder; 3, *L. donovani* R8. The supplier of the 100 bp ladder is unknown. Image supplied by Prof. Rentala Madhubala, School of Life Sciences, Jawaharlal Nehru University, New Delhi, India.

In summary, these results do not demonstrate that the lines are that of *L. donovani* and isogenic, and as a consequence it was decided to select new pentamidine resistant lines for further analysis.

5.5 Generating pentamidine resistant promastigotes

As the analysis of the ITS1 spacer region indicated that the *L. donovani* wild type and pentamidine resistant (R8) cell lines in our possession were non-isogenic, and therefore of limited value to try and understand specific changes associated with drug resistance. Consequently, it was necessary to generate an alternative pentamidine resistant cell line.

Initially, wild type *L. mexicana* promastigotes were grown in various concentrations of pentamidine (Table 5-3); and cell growth compared to that of control cells grown in the absence of pentamidine. These results indicate that 250 nM pentamidine is too toxic for wild type cells, and that 125 nM pentamidine is the maximum concentration that is tolerated by the cells. Additionally, cells at 62.5 nM and 31.25 nM pentamidine appeared to grow normally, indicating the pentamidine accumulation may not be sufficient to perturb the physiological state of the cell. In order to maximise the chance of generating a resistant cell line, the pentamidine drug pressure was applied to two independent wild type cell lines.

Table 5-3. Growth of wild type *L. mexicana* at various concentration of pentamidine

Pentamidine [nM]	Growth rate observations
500	cells died
250	cells died
125	cells grew slowly
62.5	cells grew at a similar rate to the control
31.25	cells grew at a similar rate to the control
0	cells appeared to grow normally

Pentamidine resistant *L. mexicana* were raised in the laboratory by stepwise exposure to increasing drug pressure, starting at 125 nM pentamidine, until cells capable of growth in 8 μM pentamidine had been selected. This drug concentration was chosen based on the final dose selected under drug pressure achieved for the R8 pentamidine resistant *L. donovani* cell line (Mukherjee, A *et al*. 2006).

The progress of the pentamidine resistance induction is shown in Figure 5-4. The scatter graph depicts the maximum pentamidine concentration the cells were subjected to, and survived in, over the course of 330 days. Whilst two independent cell lines were generated, only one cell line was chosen for

analysis. It is important to remember that two independently derived clonal cell lines will not necessarily have the same mechanism of resistance, and that any mechanism of resistance could be one of many.

The cells were normally grown at two different pentamidine concentrations; a concentration equal to previous passage, and one that was double the previous passage. This process ensured that if the higher concentration of pentamidine was toxic to the cells, there was a back-up cell culture at a tolerated concentration of pentamidine. A reduced growth phenotype was evident when the cells were introduced to a higher drug pressure than the previous passage. However, if the cells were maintained at a constant selection pressure, the reduced growth phenotype was less pronounced at further passages. At various points, the cells were cloned by limiting dilution in the absence of drug pressure; these are indicated by red dots (Figure 5-4). Immediately after a clonal line was generated, the cells were returned to a concentration of pentamidine to which the cells were previously exposed. The resistant strain generated, designated B5111, was capable of surviving in pentamidine concentrations 64-fold higher than those tolerated by the wild type parental strain.

The acquisition of resistance was non-linear; for example, 4 μM pentamidine was the maximum concentration that the cells were able to tolerate for around 100 days between day 190 and day 290, however, approximately 20 days later (day 310) the cells were able to survive at a concentration of 8 μM. This would suggest that a single event is responsible for a sudden increase in resistance, or that multiple events contribute to a sudden increase in resistance. Similar observations have been noted in the generation of pentamidine resistant *Trypanosoma brucei* (Bridges D 2006, PhD Thesis) and diminazine resistance in *Trypanosoma evansi* (Osman, AS *et al*. 1992).

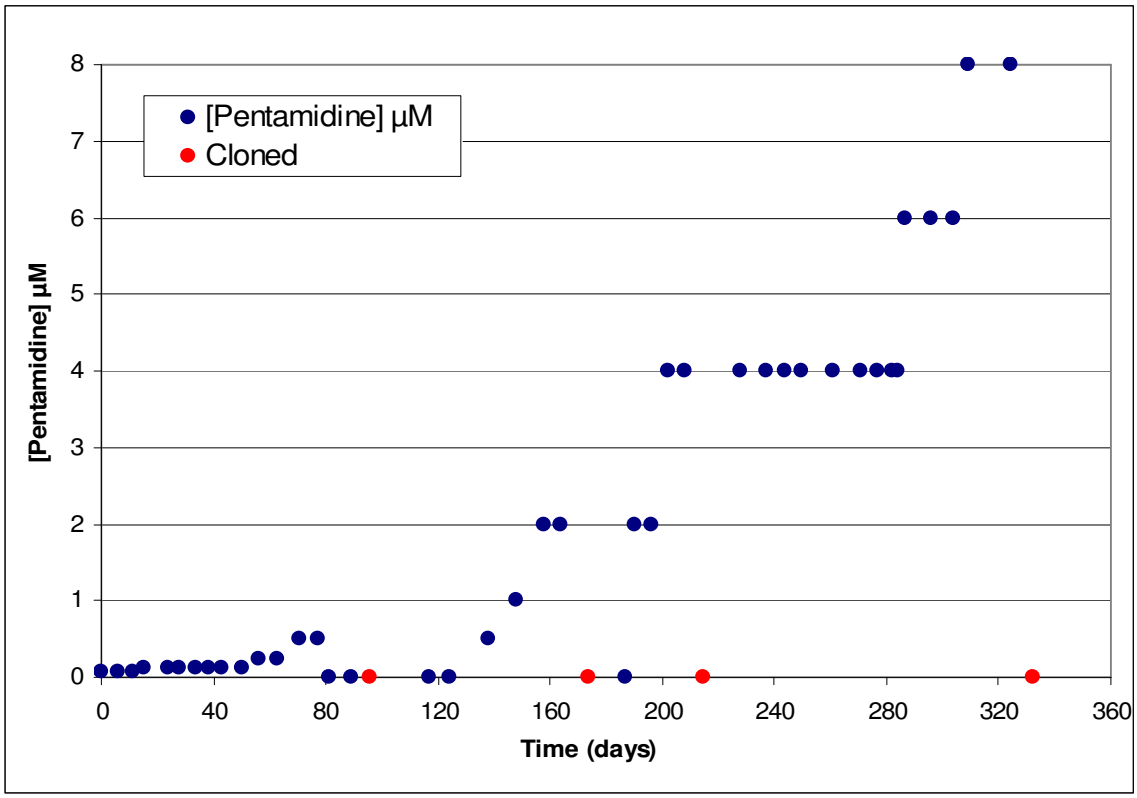


Figure 5-4. Acquisition of resistance to pentamidine over time in *L. mexicana* wild type promastigotes.

The blue dots indicate the maximum concentration of pentamidine in which the cells were growing at a rate comparable to the wild type cell line. The red dots indicate the points at which the resistant parasites were cloned (in the absence of pentamidine) and cryo-preserved.

5.6 Species Identification

The discrepancy between the authenticity of the *L. donovani* AG83 wild type and pentamidine resistant R8 cell lines reiterates the importance of genotyping cell lines prior to any comparative analysis. Therefore, the *L. mexicana* wild type and pentamidine resistant cell lines were genotyped to ensure the both cell lines were isogenic as previously described (section 5.4). The gDNA isolate from other *Leishmania* species routinely used in the lab were used as controls.

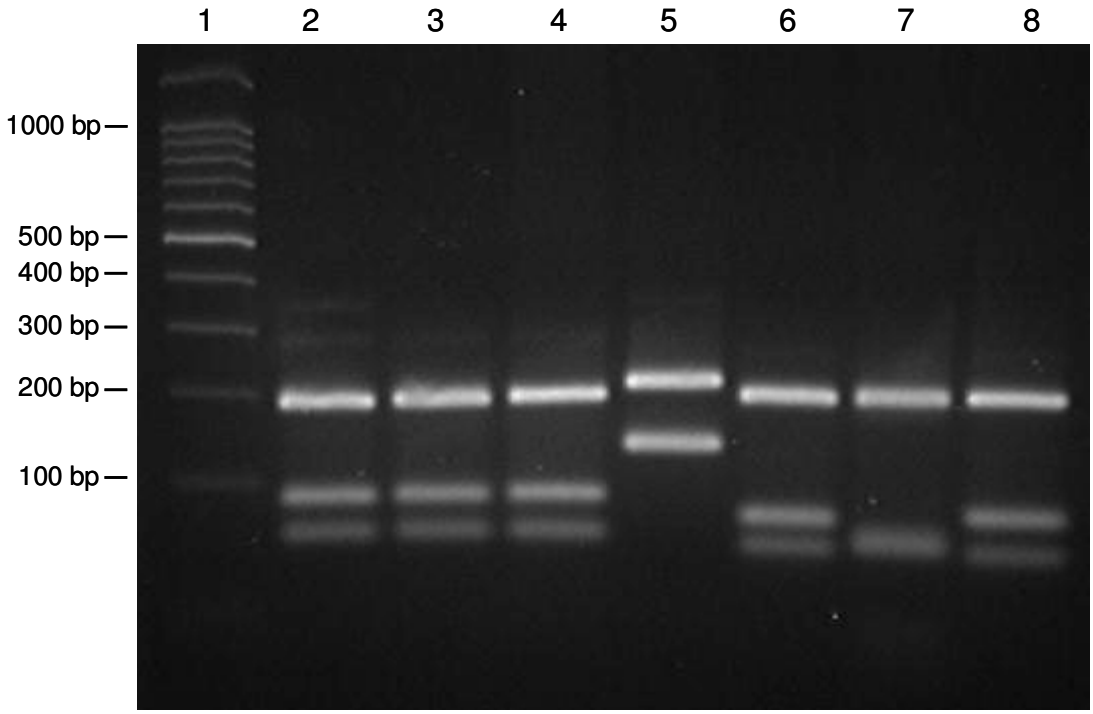


Figure 5-5. Species Identification.

Digestion of amplified ribosomal internal transcribed spacer 1 (ITS 1) regions of the different *Leishmania* species with the restriction endonuclease HaeIII. The lanes are as follows; 1, 100 bp ladder; 2, *L. mexicana* wild type; 3, *L. mexicana* pentamidine resistant; 4, *L. mexicana* positive control; 5, *L. major*; 6, *L. infantum*; 7, *L. donovani* wild type; 8, *L. donovani* pentamidine resistant (R8). The 100 bp ladder is supplied by Promega.

The RFLP profiles generated by HaeIII restriction digests for both the *L. mexicana* wild type and *L. mexicana* pentamidine resistant lines (Figure 5-5 – lanes 2 and 3 respectively) were identical to the profiles obtained from the *L. mexicana* control cell line (Figure 5-5 – lane 4; provided by Professor JC Mottram), and those outlined in Schonian G. et al. 2003. Furthermore, the RFLP profiles generated from the *L. major*, *L. infantum*, and *L. donovani* control gDNA samples were identical to those described in Schonian G. et al. 2003.

These results demonstrated that the *L. mexicana* wild type and *L. mexicana* pentamidine resistant lines described in this thesis are isogenic, and that any differences between them are not due to speciation differences.

5.7 Phenotypic Analysis

5.7.1 Growth phenotype

The growth of *L. mexicana* pentamidine resistant cells in both the presence and absence of 8 µM pentamidine was compared to that of the wild type cell line. The resistant cells grown in the absence of pentamidine had a similar growth phenotype to the wild type cell line; however, the wild type cells reached a stationary phase density of around 1.8 x10^7 cells per ml, compared to 1.6 x 10^7 cells per ml for the resistant cell line. The pentamidine resistant cell line grown in the presence of pentamidine had a slight increase in the lag time when compared to those grown in the absence of pentamidine; however, both these cell lines a maximum density of 1.6 x 10^7 cells per ml.

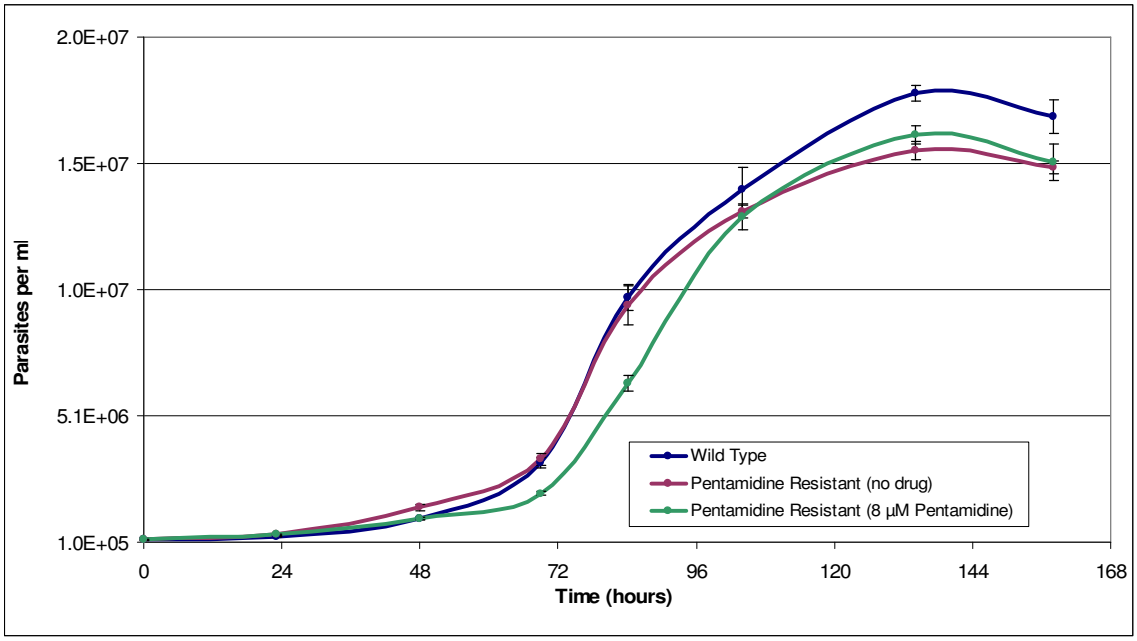


Figure 5-6. Effect on the growth of wild type and pentamidine resistant *L. mexicana* promastigotes.

The cell lines were counted over a 7 day period, with the pentamidine resistant line grown in both the absence and presence of 8 µM pentamidine. Three cultures of each cell line were counted to generate a mean value +/- SEM.

5.7.2 Alamar Blue Assay

Whilst the pentamidine resistant line generated was capable of growth in 8 µM pentamidine, compared to 0.125 µM pentamidine for the parental cell line. It was also necessary to determine the EC_{50} value, the concentration of drug which

inhibits 50 % of cellular function, of both the wild type and pentamidine resistant lines to determine the toxicity of pentamidine and other potential leishmaniacidal compounds.

The Alamar Blue assay is an effective assay to measure metabolic function and cellular activity in a variety of organisms (Raz, B *et al*. 1997). The assay involves the conversion of resazurin, a blue non-fluorescent dye, to resorufin, a pink fluorescent dye. The EC_{50} value can be easily calculated for each compound, and then compared with one another. This assay was used to determine the sensitivity to pentamidine, and a range of leishmaniacidal compounds, of both the pentamidine resistant clone B5111 and the parental cell lines. These results are displayed in Figure 5-7 and summarised in Table 5-4.

The pentamidine resistant cell line expressed cross-resistance to both diminazine (8.7-fold) and DB75 (7.6-fold) (Table 5-4). These data suggest that diminazine and DB75 share a common transport mechanism and mode of action to that of pentamidine.

Interestingly, there was only moderate cross resistance (2.0-fold) to DB766, an arylimidamine (AIA) compound, indicating that this compound may have an alternative mode of action. DB766 is a structural analogue of cationic diamidines, although differ from diamidine compounds in their physicochemical properties (they are more lipophilic at a neutral pH) and their bioavailability (Wang, MZ *et al*. 2010). Also described as 'reverse amidines', they are more active against intracellular parasites, such as *Leishmania* and *T. cruzi*, than dimaidine compounds (Wang, MZ *et al*. 2010). Previous studies have demonstrated they exhibit sub-micromollar 50 % inhibitory concentrations for drug assays involving *Leishmania* amastigotes (Stephens, CE *et al*. 2003), and sub-nanomolar concentrations in *Leishmania* promastigotes assays (Rosypal, AC *et al*. 2008). The EC_{50} value obtained for the wild type *L. mexicana* promastigotes is 1.25 nM (with a 95 % CI of 1.18 to 1.32 nM). These results are around the expected value for this class of compound.

The mechanism in which AIA compounds act is not yet known, however studies involving *T. cruzi* indicate ultrastructural alterations in the nucleus, mitochondrion, and organisation of microtubules (Silva, CF *et al*. 2007). Also

affected is the proton electrochemical potential gradient of the mitochondrial membrane. These observations are consistent with *L. mexicana* treated with pentamidine (Basselin, M *et al*. 2002). This may be one of the reasons why there is a 2-fold cross resistance for DB766 with the pentamidine resistant cell line.

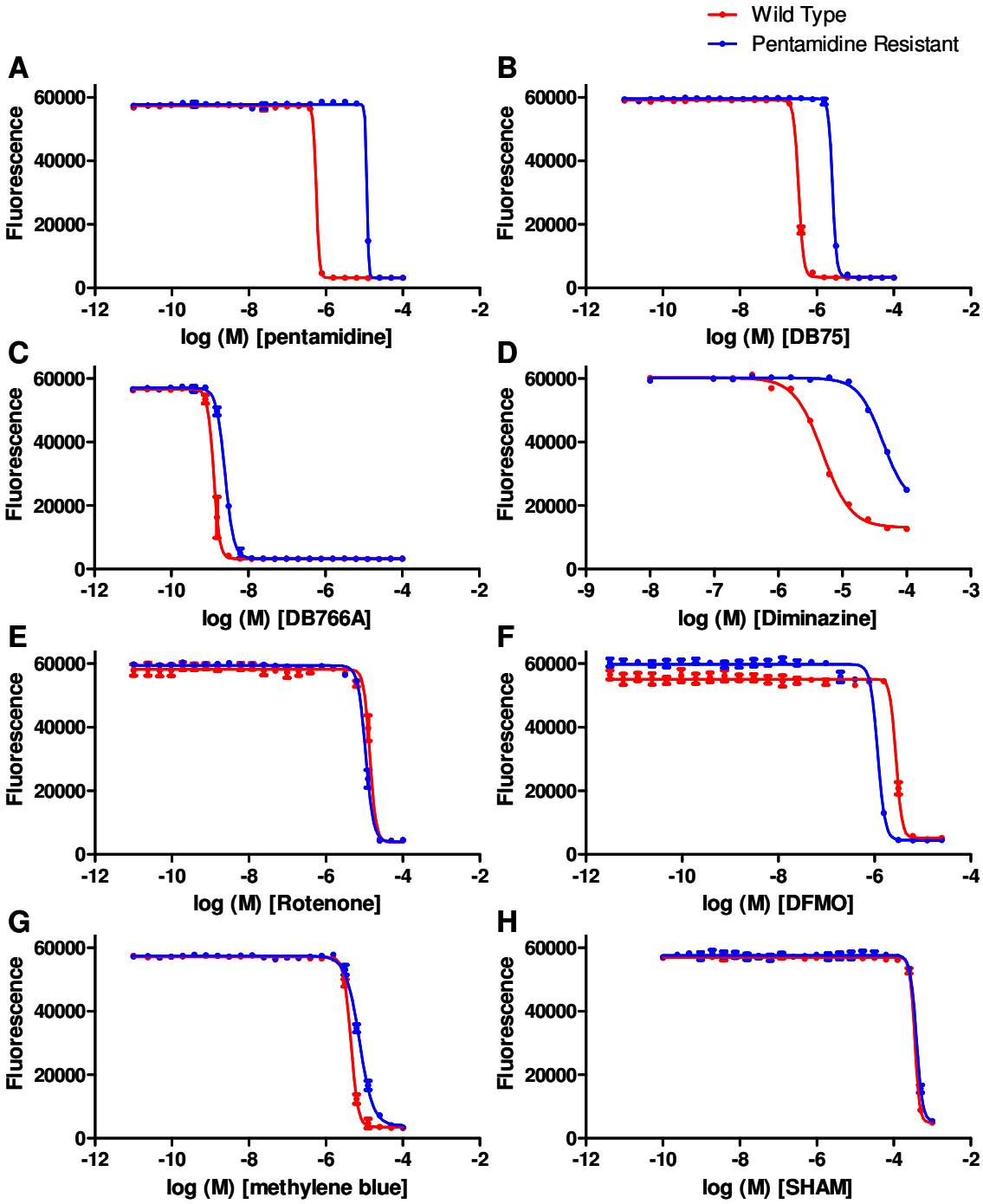


Figure 5-7. Alamar Blue Assay Graphs.

A – pentamidine. B – DB75. C – DB766A. D – Diminazine. E – Rotenone. F – DFMO. G – methylene blue. H – SHAM. The *L. mexicana* wild type are shown in red, and the L. mexicana pentamidine resistant are shown in blue. n = 3, (diminazene; n = 2); error bars = ±SEM.

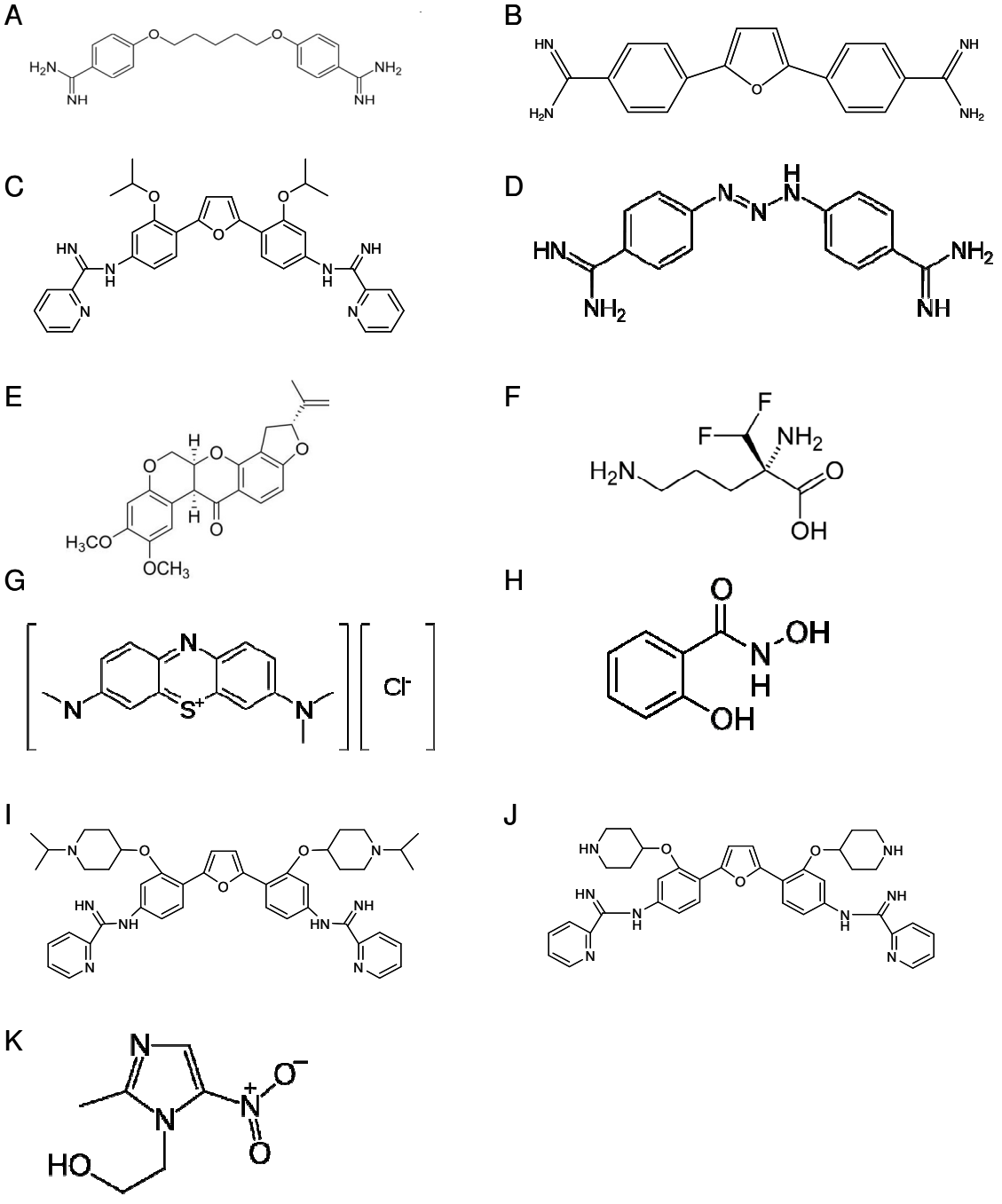


Figure 5-8. Drug structures of compounds tested

A – Pentamidine; B – DB75; C – DB766; D – Diminazine; E – Rotenone; F – DFMO; G – methylene blue; H – SHAM; I – DB1876; J – DB1880; K – Metronidazole.

Table 5-4. Summary of EC_{50} values for compound tested for their Leishmaniacidal effects.

Drug	*L. mexicana* wt		*L.mexicana* PentRes		Difference
	EC_{50}	95 % CI	EC_{50}	95 % CI	
Pentamidine	0.57 µM	0.54 to 0.61 µM	11.27 µM	8.45 to 15.05 µM	19.8
Diminazene	4.89 µM	4.4 to 5.4 µM	42.5 µM	36.8 to 49.0 µM	8.7
DB75	0.34 µM	0.33 to 0.34 µM	2.58 µM	2.52 to 2.64 µM	7.6
DB766	1.25 nM	1.18 to 1.32 nM	2.47 nM	2.43 to 2.52 nM	2.0
Methylene Blue	4.46 µM	4.36 to 4.57 µM	7.43 µM	7.13 to 7.75 µM	1.7
SHAM	351 µM	341 to 362 µM	388 µM	374 to 402 µM	1.1
Rotenone	14.0 µM	13.2 to 14.9 µM	10.9 µM	10.6 to 11.2 µM	0.8
DFMO	2.80 µM	2.61 to 3.01 µM	1.16 µM	1.11 to 1.22 µM	0.4

EC_{50}, effective concentration 50%; 95 % CI, 95 % confidence interval; n = 3, (diminazene; n= 2);

Pentamidine resistant cells are slightly more sensitive to methylene blue (1.7-fold) and eflornithine (2-fold) than the progenitor wild type cell lines. Methylene blue oxidises NADPH, thus reduces cellular levels of NADPH, generating a situation similar to oxidative stress (Maugeri, DA *et al*. 2003). NADPH can be produced by the PPP (Chapter 4). These data suggests that the flux through the PPP is increased in the pentamidine resistant cells, or that NADPH production by an alternative mechanism, for example, the conversion of malate to pyruvate by NADP-dependent malate dehydrogenase (EC 1.1.1.38), is increased. Eflornithine is an inhibitor of ornithine decarboxylase (ODC; EC 4.1.1.17) (section 1.10.3). Pentamidine resistant *L. donovani* and *L. amazonesis* were both reported to have reduced levels of putrescine; this was attributed to a decrease in putrescine uptake, increase in putrescine efflux, and lower levels of ODC (Basselin, M *et al*. 1997). If pentamidine resistant cells have reduced levels of ODC, then it could be expected that these cells will be more sensitive to eflornithine.

There was no difference in sensitivity to rotenone (0.8-fold), an inhibitor of Complex I of the electron transport chain, or SHAM (1.1-fold), an inhibitor of the *T. brucei* alternative oxidase (AO), between the wild type and pentamidine resistant cells. These results suggest that pentamidine, and subsequent resistance to pentamidine, does not affect, nor is involved in changes to Complex I or the AOX (although the presence of a bona fide alternative oxidase like protein in *Leishmania* is not certain. (Van Hellemond, JJ *et al*. 1998) reported its absence although an ortholgue of the *T. brucei* AO gene has been annotated as a putative alternative oxidase in *L. infantum* (LinJ.36.4600) and other *Leishmania* species but not yet characterized) (section 1.11.4).

Several compounds tested were not effective against *L. mexicana* wild type or the pentamidine resistant promastigotes, as observed by the failure to convert resazurin to resarufin. These included the following drugs; DB1876 starting at 1 μM, DB1880 starting at 1 μM, and metronidazole starting at 1 mM. The structure of DB766 is closely related to both DB1876 and DB1880; however DB1876 and DB1880 have piperidine rings attached via the ether linakge. The compounds all have the melamine recognition motif, necessary for transport into *T. brucei* by the P2 transporter (Carter, NS and Fairlamb, AH 1993); although this mechanism may not be used for transport by *Leishmania* spp. These additions may make the molecule too large, or too inflexible, for the transporter(s) which recognises and

transports DB766, thus reducing the effectiveness of these compounds. The mode of action of metronidazole involves the reduction of nitro group to generate short-lived cytotoxic intermediates which interact with macromolecules and DNA (Muller, M 1983). Failure for this compound to act on *L. mexicana* would indicate that the drug does not enter the cell (unlikely as uptake is by diffusion), or that metabolic pathways of low redox potential, effective at reducing nitroimidazoles, are not as prevalent in *Leishmania* as in other species.

5.8 DiGE Analysis

The proteome of the *L. mexicana* wild type and pentamidine resistant cells were compared by differential gel electrophoresis (DiGE) (Unlu, M *et al*. 1997).

A total of 20 protein spots at least 3-fold up-regulated in the pentamidine resistant line were detected by the DeCyder software, with 17 of them identified by MASCOT[25] (Table 5-5). A further four protein spots were found to be down-regulated at least 3-fold in the pentamidine resistant line, and subsequently identified.

Table 5-5. Proteins upregulated greater than 3-fold in the pentamidine resistant cell line.

Protein	Protein Accesion ID	EC number
2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase, putative	LmjF25.2010	4.1.2.-
eEF1B beta 1	LmjF34.0820	n/a
glutamine synthetase, putative	LmjF06.0370	6.3.1.2
glyceraldehyde 3-phosphate dehydrogenase, glycosomal	LmjF30.2970	1.2.1.12
hypothetical protein, conserved	LmjF24.1970	n/a
hypothetical protein, conserved	LmjF03.0820	n/a
hypothetical protein, conserved	LmjF35.2450	n/a
hypothetical protein, conserved	LmjF19.0340	n/a
mannose-1-phosphate guanyltransferase	LmjF23.0110	2.7.7.13
microtubule associated protein-like protein	LmjF19.0860	n/a
peptidase m20/m25/m40 family-like protein	LmjF31.1890	n/a
peroxidoxin	LmjF23.0040	1.11.1.-
potassium channel subunit-like protein	LmjF01.0820	n/a
protein kinase-like protein	LmjF27.1800	n/a
S-adenosylhomocysteine hydrolase	LmjF36.3910	1.16.1.8
seryl-tRNA synthetase, putative	LmjF11.0100	6.1.1.11
tryparedoxin peroxidase	LmjF15.1040	1.11.1.15

The peptide fragments isolated from *L. mexicana* were compared with the *L. major* protein database

[25] http://www.matrixscience.com/

5.8.1 The down-regulation of flagellar proteins

The paraflagellar road (PFR) is network of woven cytoskeletal filaments connected to axoneme along the length of the flagellum and has an essential role in cell motility (Bastin, P *et al*. 1998; Maga, JA and LeBowitz, JH 1999; Maga, JA *et al*. 1999). A transmission electron microscopy (TEM) image of the flagellum is shown in (Figure 5-9).

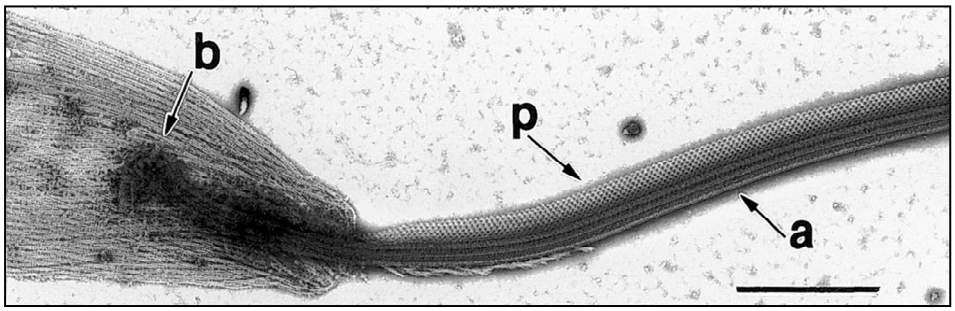


Figure 5-9. TEM image of the *L. mexicana* cytoskeletan

The flagellum, originating at the basal body and emerging from the flagellar pocket, consists of the paraflagellar rod, axoneme and attachment fibres. Abbreviations; P, paraflagellar rod; A, axoneme; B, basal body. Scale bar 1 µM. Maga, JA and LeBowitz, JH (1999).

Arising from the basal body, the flagellum is composed of the paraflagellar rod and the axoneme. In *Trypanosoma* and *Leishmania* there are two PFR proteins, PFR-1 and PFR-2, with molecular masses of 70-80 kDa and 68-72 kDa respectively. These proteins are highly conserved, with over 60 % amino acid identity (Bastin, P *et al*. 1998; Maga, JA and LeBowitz, JH 1999). Furthermore, there is at least 80 % amino acid identity between *T. brucei*, *T. cruzi*, and *L. major* for both PFR-1 and PFR-2 (Maga, JA and LeBowitz, JH 1999). Several additional minor proteins are known to localise to the flagellum, however the roles of these remain unknown (Woods, A *et al*. 1989; Woodward, R *et al*. 1994).

Several proteins associated with the flagellum were down regulated in the *L. mexicana* pentamidine resistant cell line. These included PFR-1D (LmxM.08_29.1750, LmxM.08_29.1760), PFR 2C (LmxM.16.1430), and a component of the axoneme (LmxM.20.1400). The DiGE comparison images for the PFR proteins and the component of the axoneme are displayed in Figure 5-10 and Figure 5-11 respectively.

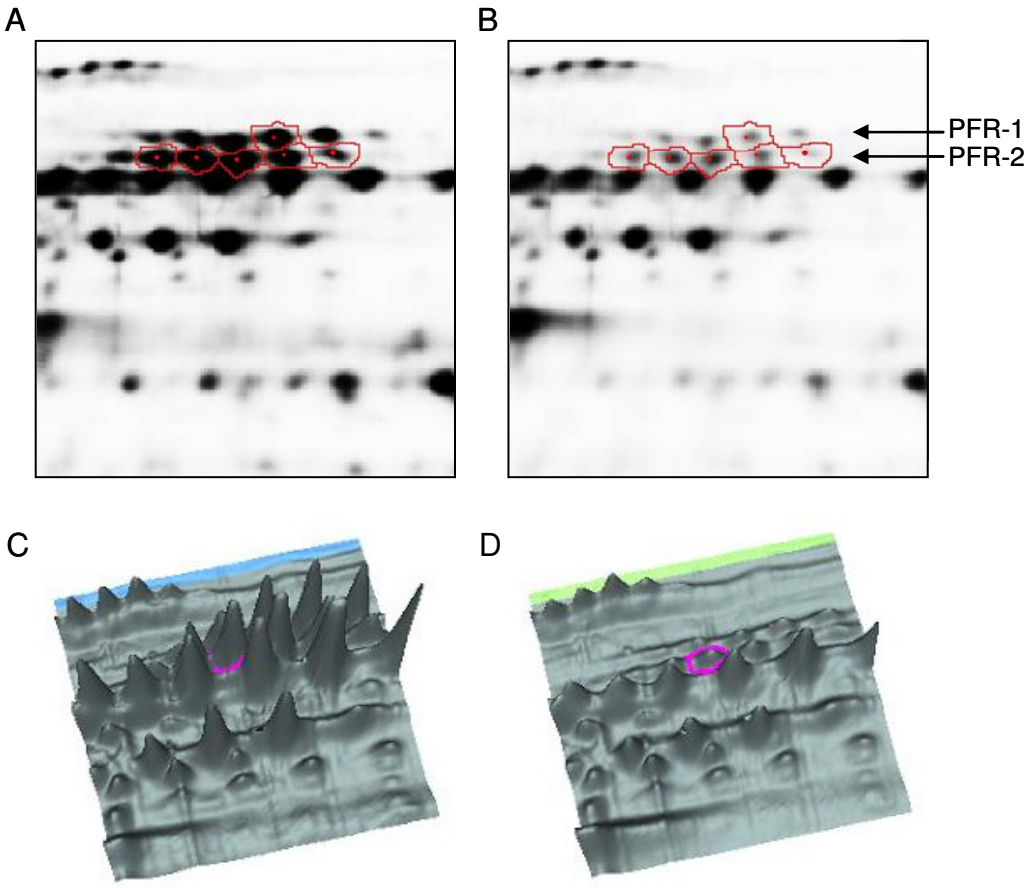


Figure 5-10. DIGE comparison images of the Paraflagellar Rod 1 (PFR-1D) and Paraflagellar Rod 2 (PFR-2C) proteins

An approximate 3-fold down-regulation in the pentamidine resistant cell line; A – 2D DIGE image of Pentamidine sensitive cell line (wild type); B – 2D DIGE image of Pentamidine resistant (clone B5111); C – 3D Intensity map Pentamidine sensitive cell line (wild type); D – 3D intensity map Pentamidine resistant (clone B5111).

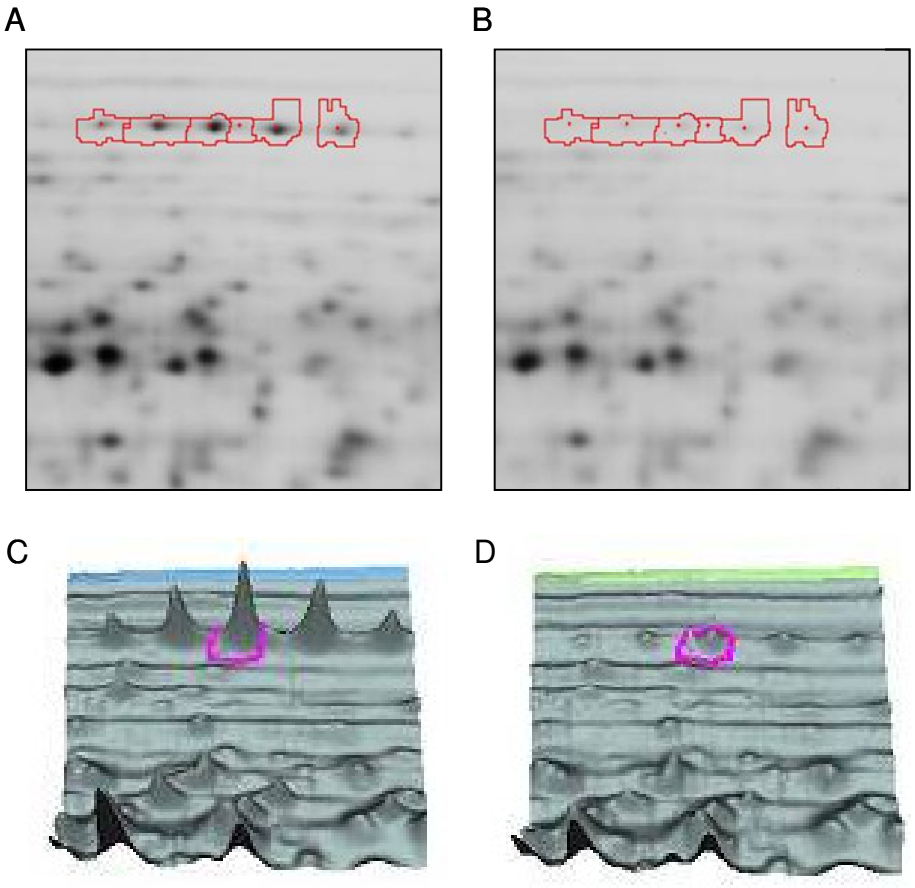


Figure 5-11. DIGE comparison images of a component of the axoneme.

An approximate 4-fold down-regulation in the pentamidine resistant cell line; A – 2D DIGE image of Pentamidine sensitive cell line (wild type); B – 2D DIGE image of Pentamidine resistant (clone B5111); C – 3D Intensity map Pentamidine sensitive cell line (wild type); D – 3D intensity map Pentamidine resistant (clone B5111).

When the pentamidine resistant promastigotes are observed using a light microscope, they appear to have reduced motility, rounder bodies, and a shorter flagellum. Morphologically, they appear more similar to amastigotes than promastigotes.

In *L. mexicana*, deletion of PFR-2C (Santrich, C *et al*. 1997), PFR-1D (Maga, JA *et al*. 1999), or both PFR-1D and PFR-2C (Maga, JA *et al*. 1999) are possible in promastigotes, indicating that these genes are not essential for parasite survival. However, in these null mutants a native PFR structure was not formed, with cells moving four to five times slower than wild type cells. Consequently, both these proteins are essential for a functional flagellum. The knockout cell lines were viable as axenic amastigotes, however preliminary results indicate that

PFR2 is required for the effective colonisation of the sandfly midgut (Maga, JA *et al*. 1999); no follow-up data is known to have been published. In *T. brucei*, loss of PFR-A, analogous to PFR-2C, is lethal (Hunger-Glaser, I and Seebeck, T 1997), whilst RNAi knockdown of the corresponding mRNA transcript resulted paralysed cells, although growth was reported to be normal (Bastin, P *et al*. 1998). The phenotype of the pentamidine resistant lines with respect to reduced motility is consistent with the reduction in paraflagellar rod proteins. However, it is not clear how a causal link between reduced paraflagellar rod proteins would cuase pentamidine resistance.

The promastigotes described in this chapter were selected in the presence of pentamidine over the course of a year, and as a consequence, any changes at the –omic or phenotypic level may be due to exposure to pentamidine, the length of time in continuous culture, or a combination of both these factors.

If pentamidine does have an effect on mitochondrial metabolism, then the ability to regenerate ATP by oxidative phosphorylation may be reduced. One possible explanation for the down-regulation of the flagellar proteins in the presence of pentamidine could be an attempt by the parasite to conserve ATP. This ATP could subsequently be used to fuel the efflux pump(s) (Bradley, G *et al*. 1988) which are responsible for the exclusion of pentamidine from the cytosol (Basselin, M *et al*. 2002), or alternatively the ATP could be used for other cellular processes.

On the other hand, these data could be explained by the length of time in continuous culture. Promastigotes grown *in vitro* may have no selective advantage in retaining a full length flagellum, therefore flagellar proteins are suitable proteins to down-regulate. This hypothesis is difficult to investigate as there may be multiple mechanisms in which *Leishmania* promastigotes may become resistant to pentamidine, or indeed any drug. A second independent culture was generated, but not analysed due to time constraints. However, even if the two cell lines did not have similar proteomic changes, it does not necessarily imply that the down-regulation of flagellar proteins is not a genuine response when the cells are challenged with pentamidine. Instead it would appear it is either a direct or indirect response, to one mechanism of pentamidine resistance in *Leishmania* promastigotes.

As previously mentioned, *L. mexicana* PFR1D and/or PFR2C knockout cell lines exist (Maga, JA *et al*. 1999). It would be interesting to compare the sensitivity to pentamidine, and other diamidine compounds, in both these knockout cell lines, with the progenitor wild type cell line.

A range of further experiments to learn more about possible links between the observed changes to the flagellum and the pentamidine resistance phenotype will be of interest in future work.

5.8.2 Unidentfied protein

There was an unidentified protein which was 12-fold upregulated in the pentamidine resistance line. This protein spot was situated near the base of the gel, suggesting a protein of low molecular weight (15-20 kDa). Additionally, the protein was slightly left of centre, indicating a pI in the region of 4.5 - 5.5. Identification of this protein, and the subsequent over-expression of the protein, would be of considerable interest in relation to pentamidine resistance.

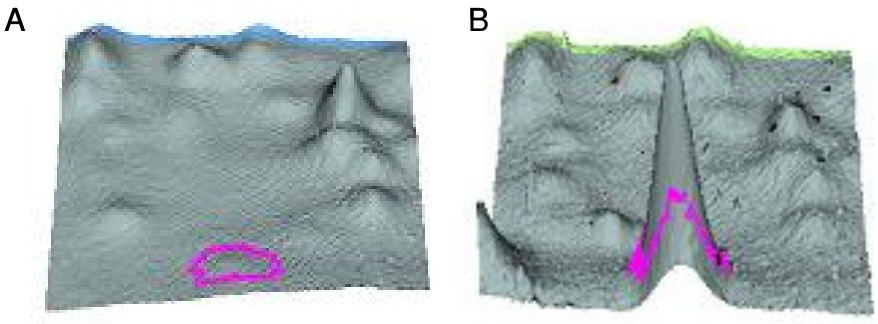


Figure 5-12. DIGE comparison of the unidentified protein illustrated by 3D intensity map

An approximate 12-fold up-regulation in the pentamidine resistant cell line; A – Pentamidine sensitive cell line (wild type); B – Pentamidine resistant (clone B5111).

5.9 Voltage Dependent Anion Channel

The voltage dependent anion channel (VDAC; Tb927.2.2510) is the main metabolite transporter in the outer mitochondrial membrane of procyclic form *T. brucei* (Pusnik, M *et al.* 2009). The VDAC could be the mechanism by which pentamidine enters the mitochondrion.

The TbVDAC locus was deleted by homologous recombination (Pusnik, M *et al.* 2009). Phenotypic analysis indicated that the VDAC -/- mutants failed to grow in the absence of glucose, whereas the control cells, designated as VDAC +/+, grew at a normal rate. These data suggests that in the absence of VDAC, SLP is essential; proline and other amino acids are unable to enter the mitochondrion and fuel OP.

Given a possible role of VDAC in pentamidine uptake, a BLAST search of the *L. mexicana* genome using the *T. brucei* VDAC ORF was performed. It revealed two hypothetical genes with 56.1 % and 51.3 % sequence identity. These were identified as LmxM.02.0460 and LmxM.02.0450, and were consequently named as VDAC-1 and VDAC-2 respectively. These two proteins have a 40.6 % sequence identity with one another (Supplementary Figure 8-8). Suitable primers were designed for the amplification of the both VDAC-1 and VDAC-2 by PCR. The PCR was performed in triplicate to enable the sequencing of three independent clones from both the wild type and pentamidine resistant cell lines.

Sequence alignments of the replicates indicated no differences (data not shown). The sequences derived from the wild type and resistant cell lines were compared with a reference sequence obtained from tritrypDB (Aslett, M *et al.* 2011)[26]. For VDAC-1, there was a single nucleotide difference at position 785 (reference strain - cytosine; lab strains - thymine) (Supplementary Figure 8-4), which results in an amino acid difference at position 262 (reference strain - alanine, GCC; lab strains - valine, GTC) (Supplementary Figure 8-6). There were no differences for VDAC-2 (Supplementary figures 8-5 and 8-7).

To investigate whether the *Leishmania* promastigotes were sensitive to pentamidine when VDAC was over-expressed, transgenic cell lines expressing VDAC-1 from an episomal locus were generated. The pNUS-HnN over-expression

[26] http://tritrypdb.org

construct (Tetaud, E *et al*. 2002) was digested with EcoRI, releasing the MCS and N-terminal His tag, and the resulting plasmid backbone was treated with SAP to prevent self re-ligation. The VDAC-1 ORF was digested from the intermediate pGEM-T easy vector, pMB-G105, using EcoRI, and sub-cloned into the over-expression construct, yielding VDAC-1 over-expression construct pMB-G111.

Wild type *L. mexicana* promastigote cells were transfected with the VDAC-1 over-expression construct or the empty vector, and grown in 50 µg/ml G-418, selecting parasites that express the neomycin acetyltransferase gene present in the over-expression construct. The cells over-expressing VDAC-1 did not survive in G-418, whereas the control cell line with no over-expression of VDAC-1 did survive. These data suggest that the over-expression of VDAC-1 in *L. mexicana* is lethal. One possible method to circumvent this problem would be to reduce the concentration of G-418 used to select successful transfected cell lines, in an attempt to reduce the expression of VDAC-1. It has been reported that increased G-418 concentrations can increase the expression of the desired protein (Tetaud, E *et al*. 2002).

No differences in VDAC are apparent in the selection of pentamidine resistance in *Leishmania*. However, a failure to successfully achieve over-expression and lack of time to produce knockout cells leaves the question open as to whether VDAC is involved in pentamidine uptake into the mitochondrion of *Leishmania*.

5.10 Pentamidine susceptibility of cells grown in different carbon sources

As diamidine compounds target the mitochondrion (Ludewig, G *et al*. 1994; Lanteri, CA *et al*. 2008), it was hypothesised that cells grown in a glucose deficient environment would be more sensitive to pentamidine and oxidative stress, than cells grown in a glucose rich environment, given the relative contribution of mitochondrial metabolism in these two conditions.

Procyclic form trypanosomes offered a model to test this hypothesis due to the availability of different growth media. These cells are normally cultured in SDM79, although can be easily adapted to grow in SDM80, a culture medium lacking glucose (Lamour, N *et al*. 2005). Initially, cells were grown in SDM80 supplemented with 5 mM D-glucose (C00031) and 5 mM L-proline (C00148). To induce a metabolic shift, cells were then grown in a glucose-rich environment (10 mM D-glucose and 0.1 mM L-proline) or proline-rich environment (10 mM L-proline and 0.1 mM D-glucose). Those grown in a glucose-rich environment will derive energy primarily via substrate level phosphorylation through glycolysis, whereas those grown in the proline-rich environment will obtain energy primarily via mitochondrial oxidative phosphorylation. As a control, cells were also maintained in SDM80 supplemented with 5 mM glucose and 5 mM proline.

Those grown in the proline-rich environment had a pentamidine EC_{50} of 5.0 nM (95 % CI of 4.2 to 5.9 nM), whereas those grown in a glucose-rich environment had an EC_{50} of 21.6 nM (95 % CI of 19.5 to 23.9 nM). This corresponds to a 4- fold difference in the EC_{50} values between these two carbon sources. Cells grown in 5 mM glucose and 5 mM proline had an EC_{50} of 24.6 nM (95 % CI of 20.0 to 31.7 mM), and as expected, were similar to those grown in the glucose rich medium. These results are summarised in Table 5-6.

Table 5-6. Pentamidine EC_{50} values for *T. brucei* 427 procylic form trypanosomes grown in different carbon sources.

	EC_{50} Pentamidine (nM)	95% Confidence Interval
5 mM Glucose + 5 mM Proline	24.6	20.0 to 31.7 nM
10 mM Glucose + 0.1 mM Proline	21.6	19.5 to 23.9 nM
10 mM Proline + 0.1 mM Glucose	5.0	4.2 to 5.9 nM

n=3

Methylene blue is an agent capable of stimulating oxidative stress. The methylene blue EC_{50} value derived from the cells grown in a proline-rich environment was 0.23 µM, although the 95 % Confidence Interval could not be calculated using GraphPad Prism software. However, exclusion of 7 data points (Supplementary Table 8-1) enabled the calculation of an EC_{50} value at 0.22 µM and 95% Confidence intervals between 0.13 µM and 0.37 µM. Those grown in the glucose-rich environment had an EC_{50} of 0.42 µM (95 % CI of 0.39 to 0.44 µM). This corresponds to a 2- fold difference in the EC_{50} values between these two carbon sources. Cells grown in 5 mM glucose and 5 mM proline had an EC_{50} of 0.61 µM (95 % CI of 0.58 to 0.65 µM). These results is summarised in Table 5-7.

Table 5-7. Methylene blue EC_{50} values for *T. brucei* 427 procylic form trypanosomes grown in different carbon sources.

	EC_{50} Methylene Blue (µM)	95% Confidence Interval
5 mM Glucose + 5 mM Proline	0.61	0.58 to 0.65 µM
10 mM Glucose + 0.1 mM Proline	0.42	0.39 to 0.44 µM
10 mM Proline + 0.1 mM Glucose	0.22	0.13 to 0.37 µM

n=3

These results indicate that procylic trypanosomes grown in a glucose-deficient environment are more sensitive to both pentamidine and methylene blue, and that low concentrations of D-glucose may affect the organisms ability to regenerate NADPH (Chapter 4). This could be related to an ability of pentamidine and methylene blue to generate oxidative stresses more readily in cells using respiration as their primary energy source.

5.11 Discussion

The initial aim of this chapter was a proteomic comparison between the *L. donovani* wild type and pentamidine resistant R8 cell lines (Mukherjee, A *et al*. 2006) , although this was not possible as the PCR diagnostic test described in this chapter indicated these cell lines were not isogenic. These data demonstrated the importance of ensuring that cell lines are what they are believed to be, prior to further comparative analyses. The diagnostic PCR test described by Schonian G. *et al* (2003) is simple, accurate and rapid. Where cell lines have been genetically manipulated, for example gene knockout, appropriate PCR experiments can confirm the gene of interest is no longer present, and has been replaced with drug resistance markers. There is no justifiable reason to not test cell lines on a regular basis, as unfortunately, mistakes do happen. The *L. mexicana* wild type and pentamidine resistant lines described in Basselin M. *et al* (2002) could not be located, and for this reason, it was necessary to generate a new pentamidine resistant *in vitro*, by step wise exposure to pentamidine. This resistant cell line was capable of growth in 8 μM pentamidine, approximately 64-fold greater than the wild type, and had an EC_{50} value of 11.3 μM, approximately 20-fold greater than the wild type. These data are consistent with previous studies that indicated it was possible to induce resistance to pentamidine at concentrations in the low micromolar range (Basselin, M *et al*. 2002; Mukherjee, A *et al*. 2006). The pentamidine resistant line described in this chapter was cross resistant to the diamidine compounds diminazene and furamidine, have a similar mode of action. Other pentamidine resistant *Leishmania* and trypanosome cell lines show cross resistance to diamidine compounds (Basselin, M *et al*. 2002; Mukherjee, A *et al*. 2006; Bridges, DJ *et al*. 2007).

Whether assessment of drug modes of action through selection of resistance *in vitro* is the best way of identifying modes of action is a topic of debate.

One of the main arguments against this approach is the length of time it can take to generate a drug resistant line. Any changes could be due to the time in culture, rather than the selective pressure of pentamidine. For example, the changes in the flagellar proteins could be due to pentamidine resistance, or alternatively, the length of time these parasites spent in continuous culture. A DiGE experiment can indicate numerous changes to the proteome, although if these changes are not genuinely associated with drug resistance, the data are

meaningless. Selecting multiple resistant lines, and determining whether they have similar changes associated with resistance can add significantly to the interpretation of such data. However, multiple mechanisms of resistance may be possible. Two independent lines were selected here, but time restraints prevented analysis of the second line.

An alternative method would be to investigate differences between field isolates which are either sensitive or resistant to a specific drug. Whilst it would be easy to confirm that the field isolates were the same species, it would be more necessary to demonstrate they were the same strain (Lewin, S *et al*. 2002; Schonian, G *et al*. 2003). Strain variations amongst a *Leishmania* species could account for subtle variations at the genomic, proteomic or metabolomic level. However, by increasing the number of drug-sensitive and drug-resistant isolated analysed, it would be possible to identify any changes occurring across multiple datasets.

Using DiGE to investigate drug resistance mechanisms in field isolates might be a suitable approach as not only has the technique become well established in recent years, the proteome is stable for a short period of time, sufficient for protein sample preparation. However, in the absence of isogenic backgrounds, there is likely to be a significant degree of protein level variability between isolated. The metabolome is even more fluid and dynamic than the proteome, and as a consequence, more susceptible to erratic fluctuations. However, a metabolomic comparison of field isolates indicated changes in the phospholipid composition of the membrane (t'Kindt, R *et al*. 2010), suggesting an increase in membrane fluidity may be related to the changes in drug uptake observed in these drug-resistant isolates. Alterations in lipid metabolism was also observed in pentamidine resistant *L. mexicana* (Basselin, M and Robert-Gero, M 1998).

Another question relates to the use of promatigotes in learning about drug resistance. The promastigote is the insect stage of the parasite, and whilst culturing promastigotes is routine and well established, it might be more appropriate to give more consideration to the effect of drugs on amastigotes, the clinically relevant life cycle stage. The process of culturing amastigotes is more complex than promastigotes, as this requires the harvesting and infection of macrophages. It has therefore been common practise to identify biochemical

modes of action in the easier to manipulate promastigote form, and then to determine whether such mechanisms translate to amastigotes and given our ultimate aim of dissecting biochemical pathways associated with resistance this study also studied the procyclic cells.

The fact that it was necessary to select new pentamidine resistant lines for analysis meant that substantial time constraints were introduced meaning that the planned metabolomic work and additional follow up on the proteomics work were not possible. However, the cell lines will now be available for such follow up.

In addition to the hypothesis-free pursuit of mechanisms of diamidine resistance I also pursued specific hypotheses that emerged based on new information derived over the time frame of the project. For example, the VDAC has been identified as the main mitochondrial metabolite transporter in *T. brucei* (Pusnik, M *et al*. 2009) and we hypothesised that this protein might be a main carrier of pentamidine into mitochondria, and given loss of mitochondrial accumulation of diamidines has been related to resistance, its loss might be associated with resistance. However, we could not identify differences at the genomic level in either of two VDAC orthologues in our sensitive or pentamidine resistant *L. mexicana* lines. Attempts to over-express one of these proteins in *L. mexicana* promastigotes was unsuccessful which is consistent with over-expression of VDAC in a range of organisms inducing apoptotic cell death (Shoshan-Barmatz, V *et al*. 2010). In *T. brucei*, procyclics which lack the VDAC protein are still capable of replicating the mitochondrion, and given that mitochondrial substrate-level phosphorylation is essential, this implies an additional mechanism for nucleotide transport into the mitochondrion (Bochud-Allemann, N and Schneider, A 2002; Pusnik, M *et al*. 2009). Increasing numbers of mitochondrial membrane protein are now being characterised in *T. brucei* and it will be of interest to determine which are involved in pentamidine uptake (Pusnik, M *et al*. 2011).

In addition, I studied a hypothesis that metabolic changes could influence sensitivity to pentamidine. For this purpose I used procyclic *T. brucei* due to the presence of different media in which either substrate level phosphorylation (glucose rich medium) or oxidative phosphorylation (proline rich medium) are the major mechanisms of ATP production. Procyclic form trypanosomes were

more sensitive to both pentamidine and the oxidiatve stress inducing agent methylene blue when grown in a glucose-deficient environment. Possible explanations for these results could involve the trypanosomes inability to rapidly regenerate NADPH to protect against the oxidative stress induced by methylene blue, or that the amino acid metabolism is affected due to the accumulation of pentamidine in the mitochondrion. However, when trypanosomes are grown in 5 mM D-glucose and 5 mM L-proline, there is a slight reduction in sensitivity to both pentamidine and methylene blue, when compared to cells grown in a glucose-rich and proline-deficient environment. A possible explanation for this result could involve the presence of sufficient glucose for NADPH regeneration, and the presence of sufficient proline to sustain growth via OP. When glucose is the only carbon source present, this must fuel both NADPH and ATP production. However, more detailed analysis of metabolism would be required to resolve this.

It would be useful to adapt *Leishmania* promastigotes to a custom made medium, either HOMEM or SDM-79/SDM80, and compare the susceptibility to both pentamidine and methylene blue when grown in glucose-deficient and glucose-rich environments. As *Leishmania* promastigotes are able to metabolise ribose, using it as a carbohydrate source (Berens, RL *et al*. 1980) it would be interesting to determine if the presence of ribose, in the absence of glucose, is sufficient for NADPH regeneration, thus reducing sensitivity to oxidative stress. These experiments would be relevant to the *L. mexicana* tkt null cell line, which, under standard growth conditions, show little difference in sensitivity to pentamidine when compared with the wild type cells (Chapter 4).

6 General Discussion

There is an urgent need to develop new compounds to combat trypanosomiasis and leishmaniasis, as many of the current drugs are toxic, expensive, and difficult to administer. Drug resistance is also a growing concern. If we understand how these drugs work (i.e. the mode of action), or how resistance is acquired, then our ability to control these diseases will improve.

In this study, a pentamidine resistant *L. mexicana* promastigote cell line was generated *in vitro*, by step wise exposure to pentamidine. These cells were cross resistant to other diamidine compounds, consistent with previous studies (Basselin, M *et al*. 2002; Mukherjee, A *et al*. 2006). Proteomic analysis using DiGE revealed differences in proteins associated with the flagellum. These changes may be genuinely associated with pentamidine resistance or, alternatively, they may be due to the length of time spent in continuous culture during resistance induction. It may be more beneficial to investigate changes to the proteome (and metabolome) over a shorter period of time to minimise indirect changes.

The successful deletion of the transketolase gene from the Leishmania genome indicates that this gene is not essential in the promastigote form of the parasite, at least under standard laboratory conditions. There was a moderate increase in sensitivity to pentamidine in the knockout cell line, although no difference in the sensitivity to methylene blue was detected, indicating that NADPH regeneration is not affected in the knockout cell line. Interestingly, a reduction in the acidification of the growth medium observed during these experiments may indicate changes in the end products that are excreted by the cell (either a reduction in organic acids, or an increase in ammonia). Metabolomic profiling revealed that an alternative mechanism of E4P production may be present in Leishmania, although further experiments are necessary. A targeted metabolomics approach, using labelled substrates, might reveal the existence of alternative enzymatic reactions. The presence of additional enzymes in carbohydrate metabolism is not unrealistic given the range of sugars that the Leishmania promastigotes are exposed to in the sandfly midgut (Opperdoes, FR and Coombs, GH 2007). If substrate availability is able to initiate a conformational change in transketolase through exposure of the PTS targeting signal, the knockout cell lines provide a null background of endogenous TKT, to allow the characterisation of cell lines expressing mutant TKT.

This thesis presents a simple metabolite database that is able to query complex metabolomic datasets, and provide a biological context in relation to KEGG (Kanehisa, M and Goto, S 2000). KEGG is an excellent resource for metabolites, biochemical reactions, and metabolic pathways, although the rigid nature of depicting classical biochemical pathways can be a hindrance when considering a global metabolomics approach. The Human Metabolome Database (HMDB) (Wishart, DS *et al*. 2009) is another excellent resource, although it is obviously specific for human metabolites. However, the metabolite information in these databases is often acquired through the annotation of enzymes, with an assumption that metabolites associated with a particular enzyme are present in the metabolome.

It is imperative that the metabolomics community resolve the metabolome of specific biological systems so as to provide reference metabolomes to aid untargeted metabolomics. There are no shortages of cell lines to analyse using metabolomic techniques, and there are no shortage of biological questions to ask. Unfortunately, there is a lack of appropriate software programs to analyse the data, and as a consequence data is produced faster than it can be analysed. However, there are encouraging signs with the introduction of IDEOM (Creek, DJ *et al*. 2012) and PeakML (Scheltema, RA *et al*. 2011) for the analysis of raw metabolomic data, although the interpretation of peaks (and related peaks) is subjective. Alternatively, MassTrix (Suhre, K and Schmitt-Kopplin, P 2008) and Pathos (Leader, DP *et al*. 2011) were designed to communicate with the KEGG database. Whilst metaSearch is a basic program, the ability to rapidly match data acquired from a high resolution mass spectrometer against a list of known masses to provide a tabular view of results is certainly useful. The conventional method to view relationships between metabolites is through classical biochemical pathways, although attitudes towards this approach are changing through the introduction of programs such as MetExplore to visualise metabolic changes (Cottret, L *et al*. 2010). Efforts should concentrate on integrating the elements of raw data analysis and network visualisation.

Adopting the approaches used in other disciplines may accelerate the rate at which metabolomics, particularly untargeted metabolomics, develops as a technique. For example, the advent of DNA sequencing by Frederick Sanger in the 1970's, eventually enabled the sequencing of entire genomes. At the end of

2011, complete genome projects existed for 3080 organisms (Kyrpides, NC 1999). The Human Genome Project (HGP)27, started in 1990, took 13 years to complete (and is constantly revised). More recently, projects such as TriTrypDB (Aslett, M *et al*. 2011), specific to Leishmania and Trypanosoma, and PlasmoDB (Aurrecoechea, C *et al*. 2009), specific to Plasmodium, have become essential resources for genomic or functional genomic studies. These databases provide reference genomes that allow straightforward comparisons with sequence data derived in the laboratory. A reference metabolome, derived from metabolomic data, would become a fundamental analysis tool for the metabolomics community. There are however various challenges that have to be addressed. These include the standardisation of metabolite sample preparation method(s), and the robust identification of metabolites by orthogonal approaches. The coverage of such a reference metabolome would be by no means complete. However, it would provide a solid foundation for the interpretation of untargeted metabolomic studies, in particularly how parasites respond to individual drugs.

[27] http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml

7 Appendices

7.1 Cell Culture medium

7.1.1 HOMEM

NaH_2PO_4	157 mg
NaCl	6.8 g
$MgSO_4.7H_2O$	200 mg
KCl	400 mg
$NaHCO_3$	300 mg
Sodium Pyruvate	110 mg
D-Glucose	3 g
L-Glutamine	292 mg
P-aminobenzoic Acid	1 mg
HEPES	5.96 g
D-Biotin	0.1 mg
Phenol red (dissolved in 5 ml 0.1 M NaOH)	15 mg
Vitamins 100 X (Sigma, M6895)	10 ml

Choline chloride	100 mg/l
D-Calcium pantothenate	100 mg/l
Folic Acid	100 mg/l
Nicotinamide	100 mg/l
Pyridoxal hydrochloride	100 mg/l
Riboflavin	10 mg/l
Thiamine hydrochloride	100 mg/l
I-Inositol	200 mg/l
KCl	200 mg/l
KH_2PO_4	200 mg/l
NaCl	8 g/l
Na_2HPO_4	1.15 g/l

MEM non-essential amino acids 100 X (Gibco, 11140035)	10ml

Glycine	750 mg/l
L-Alanine	890 mg/l
L-Asparagine	1320 mg/l
L-Aspartic acid	1330 mg/l
L-Glutamic Acid	1470 mg/l
L-Proline	1150 mg/l
L-Serine	1050 mg/l

Essential amino acids 50 X (Sigma, M5550)	30 ml

L-Arginine	6.32 g/l
L-Cystine.2HCl	1.564 g/l
L-Histidine.HCl.H_2O	2.1 g/l
L-Isoleucine	2.625 g/l
L-Leucine	2.62 g/l
L-Lysine	3.625 g/l
L-Methionine	755 mg/l
L-Phenylalanine	1.65 g/l
L-Threonine	2.38 g/l
L-Tryptophan	510 mg/l
L-Tyrosine	1.8 g/l
L-Valine	2.34 g/l

Produced by Invitrogen; For 1 litre; pH 7.3; store at 4ºC.

Foetal Calf Serum	100 ml

7.1.2 SDM-80

NaH_2PO4	157 mg
NaCl	6.8 g
$MgSO_4.7H_2O$	151 mg
KCl	400 mg
$CaCl_2.2H_2O$	250 mg
L-Arginine	100 mg
L-Methionine	70 mg
L-Phenylalanine	80 mg
L-Threonine	350 mg
L-Tyrosine	100 mg
Taurine	160 mg
L-Alanine	200 mg
L-Asparagine	13.2 mg
L-Aspartate	13.3 mg
L-Glutamate	14.7 mg
L-Glutamine	200 mg
Glycine	7.5 mg
L-Serine	60 mg
HEPES	8 g
MOPS	5 g
$NaHCO_3$	2.2 g
Pyruvate	220 mg
Mercaptoethanol (neat)	35 µl
Hypoxanthine	14 mg
Thymidine	4 mg
Vitamins 100 X (Sigma, M6895)	10 ml
Essential amino acids 50 X (Gibco BRL, 1130-036)	20 ml
Hemin (dissolved in 0.1 M NaOH)	5 mg
Phenol red (dissolved in 0.1 M NaOH)	15 mg
Dialysed Foetal Calf Serum	100 ml

Glucose-rich; 1.802 g D-glucose and 0.0115 g L-proline
Proline-rich; 1.151 g L-proline and 0.0180 g D-glucose

For 1 litre; pH 7.4; filter sterilise; store at 4ºC.

7.2 Molecular Biology Reagents

7.2.1 Cell lysis buffer A

10 mM Tris-Cl pH 8.0
100 mM EDTA
1% Sarkosyl
100 µg Proteinase K per 0.5 ml (add fresh each time)

7.2.2 1x TAE

Tris acetate (pH 8.5) 40 mM
EDTA 1 mM

7.2.3 Depurination solution

22.7 ml 11 M HCl
977.3 ml dH_20

7.2.4 Denaturation solution

20g NaOH
87.664g NaCl
Make to I litre

7.2.5 Neutralisation solution

87g NaCl
1 M Tris-HCl pH7.4
Make to 1 litre

7.2.6 20x SSC

88.23 g NaCl
175.32 g Trisodiumcitrate
Make to 1 litre
pH7-8

7.2.7 50x Denhardts

1% (w/v) Ficoll 400
1% (w/v) Polyvinylpyrrolidone
1% (w/v) Bovine Serum Albumin (Sigma-Aldrich, Poole, UK, Fraction V)
Filter and store at -20 °C

7.3 Proteomics

7.3.1 Cell Lysis Buffer B

6 M Urea
2 M Thiourea
4 % CHAPS
25 mM Tris

7.3.2 Protein Inhibitor Cocktail

Leupetin 100 µg/ml
Pefabloc 500 µg/ml
Pepstatin 5 µg/ml
Phenantrholine 1 mM
EDTA 1 mM
EGTA 1 mM

7.3.3 DiGE Rehydration Buffer

6 M Urea
2 M Thiourea
4% CHAPS
0.5% IPG buffer
65 mM DTT
Trace bromophenol blue

7.3.4 DiGE Equilibration Buffer 1

2 % SDS
50 mM Tris-HCl pH 8.8
6 M urea
30 % (v/v) glycerol
0.002% bromophenol blue
100 mg of DTT per 10 ml

7.3.5 DiGE Equilibration Buffer 2

2 % SDS
50 mM Tris-HCl pH 8.8
6 M urea
30 % (v/v) glycerol
0.002% bromophenol blue
250 mg of Idoacetamide per 10 ml

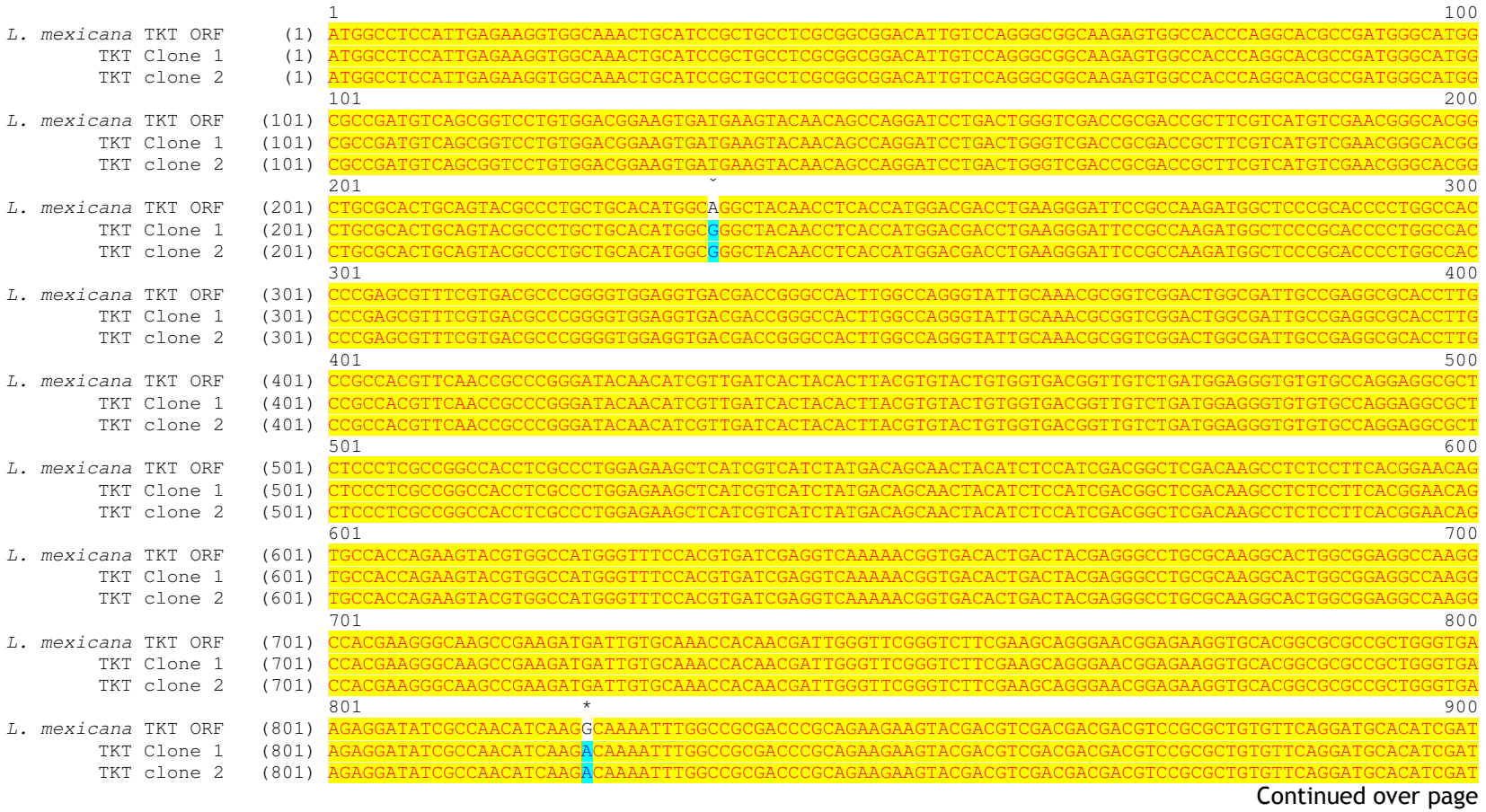
7.3.6 DiGE fixing solution

40% ethanol
10% acetic acid
0.05% SDS

8 Supplementary Data

Figure 8-1. *L. mexicana* Transketolase amino acid seqeuence alignment.

Alignment of the *L. mexicana* TKT amino acid sequence (GeneBank accession ID AJ427448) with two independently derived clones. An amino acid change from alanine to threonine, donated by an asterisk, is located at residue 275.



Continued over page

Figure 8-2. *L. mexicana* Transketolase DNA sequence alignment.

Alignment of the *L. mexicana* TKT ORF sequence (GeneBank accession ID AJ427448) with two independently derived clones. Point mutations are denoted by a caron (silent mutation) or an asterisk (mutation conferring an amino acid change), and are located at nucleotides 234 and 823 respectively.

```
                                 901                                                                          1000
L. mexicana TKT ORF    (901) AAGTGTTCCGCGGAACAGAAGGCGTGGGAGGAACTCTTGGCGAAGTACACAGCCGCGTTCCCGGCCGAGGGTGCCGCCTTTGTGGCGCAGATGAGGGGCG
        TKT Clone 1    (901) AAGTGTTCCGCGGAACAGAAGGCGTGGGAGGAACTCTTGGCGAAGTACACAGCCGCGTTCCCGGCCGAGGGTGCCGCCTTTGTGGCGCAGATGAGGGGCG
        TKT clone 2    (901) AAGTGTTCCGCGGAACAGAAGGCGTGGGAGGAACTCTTGGCGAAGTACACAGCCGCGTTCCCGGCCGAGGGTGCCGCCTTTGTGGCGCAGATGAGGGGCG
                                1001                                                                         1100
L. mexicana TKT ORF   (1001) AGCTGCCGTCTGGGTGGGAGGCGAAGCTCCCGACGAACTCCTCGGCCATCGCGACGCGCAAGGCGAGCGAGAACTGCCTGGCTGTGCTCTTCCCGGCCAT
        TKT Clone 1    (1001) AGCTGCCGTCTGGGTGGGAGGCGAAGCTCCCGACGAACTCCTCGGCCATCGCGACGCGCAAGGCGAGCGAGAACTGCCTGGCTGTGCTCTTCCCGGCCAT
        TKT clone 2    (1001) AGCTGCCGTCTGGGTGGGAGGCGAAGCTCCCGACGAACTCCTCGGCCATCGCGACGCGCAAGGCGAGCGAGAACTGCCTGGCTGTGCTCTTCCCGGCCAT
                                1101                                                                         1200
L. mexicana TKT ORF   (1101) CCCGGCTCTCATGGGCGGATCGGCTGACCTCACGCCGAGCAACCTGACGCGCCCCGCGTCGGCAAACTTGGTGGACTTCTCGTCGAGCAGCAAGGAGGGT
        TKT Clone 1    (1101) CCCGGCTCTCATGGGCGGATCGGCTGACCTCACGCCGAGCAACCTGACGCGCCCCGCGTCGGCAAACTTGGTGGACTTCTCGTCGAGCAGCAAGGAGGGT
        TKT clone 2    (1101) CCCGGCTCTCATGGGCGGATCGGCTGACCTCACGCCGAGCAACCTGACGCGCCCCGCGTCGGCAAACTTGGTGGACTTCTCGTCGAGCAGCAAGGAGGGT
                                1201                                                                         1300
L. mexicana TKT ORF   (1201) CGCTACATTCGCTTCGGTGTCCGTGAACATGCCATGTGCGCCATCCTCAACGGTCTCGACGCCCATGATGGTATCATCCCGTTCGGCGGCACCTTCCTCA
        TKT Clone 1    (1201) CGCTACATTCGCTTCGGTGTCCGTGAACATGCCATGTGCGCCATCCTCAACGGTCTCGACGCCCATGATGGTATCATCCCGTTCGGCGGCACCTTCCTCA
        TKT clone 2    (1201) CGCTACATTCGCTTCGGTGTCCGTGAACATGCCATGTGCGCCATCCTCAACGGTCTCGACGCCCATGATGGTATCATCCCGTTCGGCGGCACCTTCCTCA
                                1301                                                                         1400
L. mexicana TKT ORF   (1301) ACTTCATCGGCTACGCCCTTGGTGCAGTGCGCCTCGCCGCGATTTCTCACCACCGCGTCATCTACGTGGCGACACACGACAGCATCGGCGTTGGCGAGGA
        TKT Clone 1    (1301) ACTTCATCGGCTACGCCCTTGGTGCAGTGCGCCTCGCCGCGATTTCTCACCACCGCGTCATCTACGTGGCGACACACGACAGCATCGGCGTTGGCGAGGA
        TKT clone 2    (1301) ACTTCATCGGCTACGCCCTTGGTGCAGTGCGCCTCGCCGCGATTTCTCACCACCGCGTCATCTACGTGGCGACACACGACAGCATCGGCGTTGGCGAGGA
                                1401                                                                         1500
L. mexicana TKT ORF   (1401) CGGGCCAACCCACCAGCCTGTCGAGTTGGTGGCTGCCCTGCGTGCCATGCCAAACCTGCAGGTGATACGTCCTAGCGACCAGACAGAGACGAGCGGTGCG
        TKT Clone 1    (1401) CGGGCCAACCCACCAGCCTGTCGAGTTGGTGGCTGCCCTGCGTGCCATGCCAAACCTGCAGGTGATACGTCCTAGCGACCAGACAGAGACGAGCGGTGCG
        TKT clone 2    (1401) CGGGCCAACCCACCAGCCTGTCGAGTTGGTGGCTGCCCTGCGTGCCATGCCAAACCTGCAGGTGATACGTCCTAGCGACCAGACAGAGACGAGCGGTGCG
                                1501                                                                         1600
L. mexicana TKT ORF   (1501) TGGGCTGTTGCACTGTCTAGTATTCACACTCCAACGGTTCTGTGTCTGAGCCGCCAGAACACCGAGCCGCAGTCGGGGTCGAGCATCGAGGGTGTGAGGC
        TKT Clone 1    (1501) TGGGCTGTTGCACTGTCTAGTATTCACACTCCAACGGTTCTGTGTCTGAGCCGCCAGAACACCGAGCCGCAGTCGGGGTCGAGCATCGAGGGTGTGAGGC
        TKT clone 2    (1501) TGGGCTGTTGCACTGTCTAGTATTCACACTCCAACGGTTCTGTGTCTGAGCCGCCAGAACACCGAGCCGCAGTCGGGGTCGAGCATCGAGGGTGTGAGGC
                                1601                                                                         1700
L. mexicana TKT ORF   (1601) ACGGCGCCTACTCGGTGGTGGATGTGCCCGACCTGCAGCTCGTGATCGTGGCGAGCGGCTCGGAGGTGTCGCTGGCGGTGGATGCTGCCAAAGCGCTCTC
        TKT Clone 1    (1601) ACGGCGCCTACTCGGTGGTGGATGTGCCCGACCTGCAGCTCGTGATCGTGGCGAGCGGCTCGGAGGTGTCGCTGGCGGTGGATGCTGCCAAAGCGCTCTC
        TKT clone 2    (1601) ACGGCGCCTACTCGGTGGTGGATGTGCCCGACCTGCAGCTCGTGATCGTGGCGAGCGGCTCGGAGGTGTCGCTGGCGGTGGATGCTGCCAAAGCGCTCTC
                                1701                                                                         1800
L. mexicana TKT ORF   (1701) GGGTGAGCTGCGCGTAAGGGTCGTGTCGATGCCGTGCCAGGAGCTCTTCGACGCACAACCGGATACGTACCGCCAGGCTGTGCTCCCCGCGGGTGTGCCG
        TKT Clone 1    (1701) GGGTGAGCTGCGCGTAAGGGTCGTGTCGATGCCGTGCCAGGAGCTCTTCGACGCACAACCGGATACGTACCGCCAGGCTGTGCTCCCCGCGGGTGTGCCG
        TKT clone 2    (1701) GGGTGAGCTGCGCGTAAGGGTCGTGTCGATGCCGTGCCAGGAGCTCTTCGACGCACAACCGGATACGTACCGCCAGGCTGTGCTCCCCGCGGGTGTGCCG
                                1801                                                                         1900
L. mexicana TKT ORF   (1801) GTGGTGTCGGTGGAGGCGTACGTCAGTTTTGGCTGGGAGAAATACTCCCATGCGCACGTGGGCATGTCCGGCTTCGGTGCCTCGGCCCCGGCGGGTGTGC
        TKT Clone 1    (1801) GTGGTGTCGGTGGAGGCGTACGTCAGTTTTGGCTGGGAGAAATACTCCCATGCGCACGTGGGCATGTCCGGCTTCGGTGCCTCGGCCCCGGCGGGTGTGC
        TKT clone 2    (1801) GTGGTGTCGGTGGAGGCGTACGTCAGTTTTGGCTGGGAGAAATACTCCCATGCGCACGTGGGCATGTCCGGCTTCGGTGCCTCGGCCCCGGCGGGTGTGC
                                1901                                                                         2000
L. mexicana TKT ORF   (1901) TATACAAGAAGTTTGGAATTACCGTCGAGGAAGTGGTGAGGACGGGCCGTGAGTTGGCCAAGCGCTTCCCCGATGGCACGGCGCCGCTCAAGAACTCTTC
        TKT Clone 1    (1901) TATACAAGAAGTTTGGAATTACCGTCGAGGAAGTGGTGAGGACGGGCCGTGAGTTGGCCAAGCGCTTCCCCGATGGCACGGCGCCGCTCAAGAACTCTTC
        TKT clone 2    (1901) TATACAAGAAGTTTGGAATTACCGTCGAGGAAGTGGTGAGGACGGGCCGTGAGTTGGCCAAGCGCTTCCCCGATGGCACGGCGCCGCTCAAGAACTCTTC
                                2001       2016
L. mexicana TKT ORF   (2001) ATTCAGCAAGATGTAA
        TKT Clone 1    (2001) ATTCAGCAAGATGTAA
        TKT clone 2    (2001) ATTCAGCAAGATGTAA
```

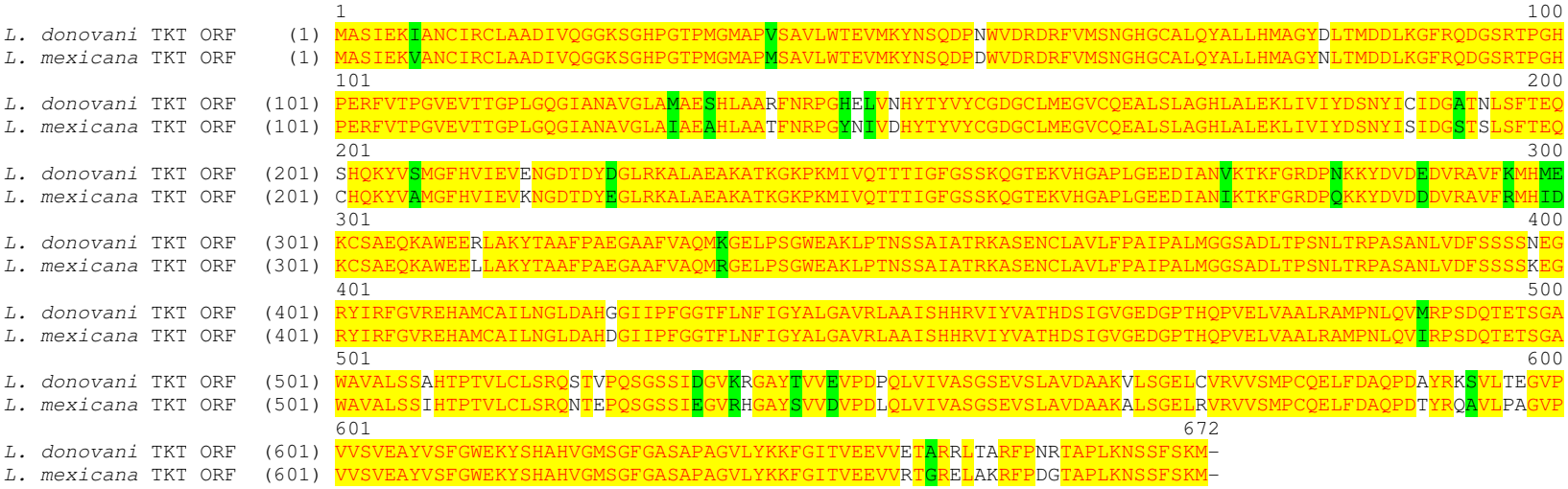


Figure 8-3. Alignment of transketolase amino acid sequences from *L. donovani* and *L. mexicana*.

Alignment of the *L. donovani* TKT (MHOM/NP/03/BPK282A1; EMBL accession ID E9BH77) and *L. mexicana* TKT (Laboratory strain) amino acid sequences. There is a 92.1 % sequence identity.

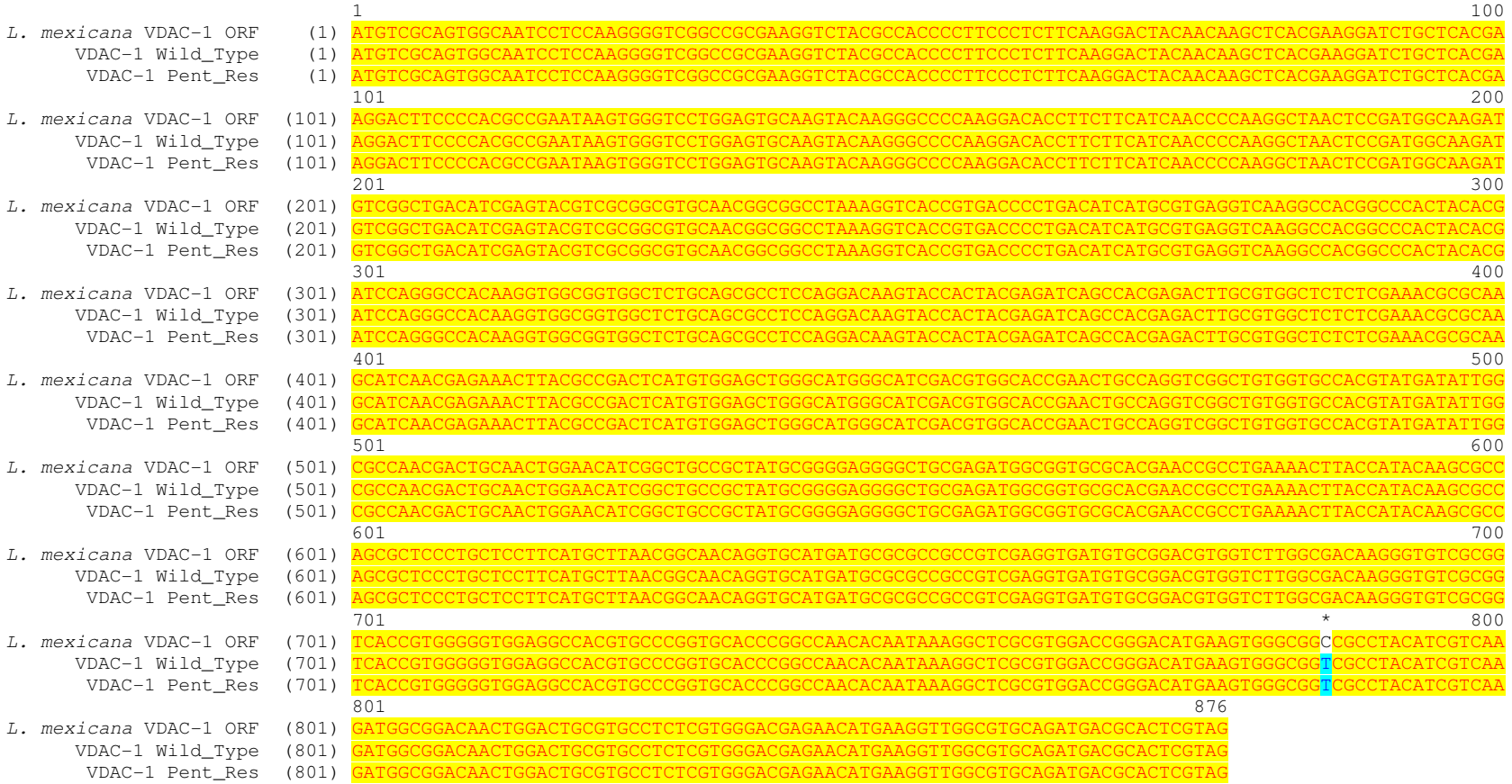


Figure 8-4. VDAC-1 DNA sequence alignment.

Alignment of the *L. mexicana* VDAC-1 ORF sequence (LmxM.02.0460) with the sequences derived from the wild type and pentamidine resistant lab strains. Point mutations are denoted by an asterisk (mutation conferring an amino acid change), and is located at nucleotide 875.

```
                                 1                                                                                   100
L. mexicana VDAC-2 ORF     (1)   ATGACAACCCTCTTCAAGGATTACAGCAAGGGCAGCAATGACTTGCTTACCAAGAACTTTTCCAGCTGCGGCGCGTGGAAGGTCGAGAGCAAGAGCAAGG
      VDAC-2 Wild_Type     (1)   ATGACAACCCTCTTCAAGGATTACAGCAAGGGCAGCAATGACTTGCTTACCAAGAACTTTTCCAGCTGCGGCGCGTGGAAGGTCGAGAGCAAGAGCAAGG
       VDAC-2 Pent_Res     (1)   ATGACAACCCTCTTCAAGGATTACAGCAAGGGCAGCAATGACTTGCTTACCAAGAACTTTTCCAGCTGCGGCGCGTGGAAGGTCGAGAGCAAGAGCAAGG
                                 101                                                                                 200
L. mexicana VDAC-2 ORF   (101)   CCCCCAAAGGCACGTATGCCCTCACGACCACCTCCAACACCCACGGTGACGTGAAGTTGGATATCGAGGGGCTGACGGGCGACGGCGCCTATTATGGCAA
      VDAC-2 Wild_Type   (101)   CCCCCAAAGGCACGTATGCCCTCACGACCACCTCCAACACCCACGGTGACGTGAAGTTGGATATCGAGGGGCTGACGGGCGACGGCGCCTATTATGGCAA
       VDAC-2 Pent_Res   (101)   CCCCCAAAGGCACGTATGCCCTCACGACCACCTCCAACACCCACGGTGACGTGAAGTTGGATATCGAGGGGCTGACGGGCGACGGCGCCTATTATGGCAA
                                 201                                                                                 300
L. mexicana VDAC-2 ORF   (201)   GCTCTGCTTCACGTCGAAGGACCTCACCGACATCCAGCTCACCGTCCGCGCCGAGGACCTCGATAACCACCGCGTGGAGGCCATCATTGGCCACAAGGGC
      VDAC-2 Wild_Type   (201)   GCTCTGCTTCACGTCGAAGGACCTCACCGACATCCAGCTCACCGTCCGCGCCGAGGACCTCGATAACCACCGCGTGGAGGCCATCATTGGCCACAAGGGC
       VDAC-2 Pent_Res   (201)   GCTCTGCTTCACGTCGAAGGACCTCACCGACATCCAGCTCACCGTCCGCGCCGAGGACCTCGATAACCACCGCGTGGAGGCCATCATTGGCCACAAGGGC
                                 301                                                                                 400
L. mexicana VDAC-2 ORF   (301)   CCGGCGGTGTGCGATATATCGGTGGAGGTCAACCACCAGACGGTAAGGCCGATAGCGGGGGGCTGCCTGAGCGTCAACGAAAAGTTTACGCAGAAGGCCG
      VDAC-2 Wild_Type   (301)   CCGGCGGTGTGCGATATATCGGTGGAGGTCAACCACCAGACGGTAAGGCCGATAGCGGGGGGCTGCCTGAGCGTCAACGAAAAGTTTACGCAGAAGGCCG
       VDAC-2 Pent_Res   (301)   CCGGCGGTGTGCGATATATCGGTGGAGGTCAACCACCAGACGGTAAGGCCGATAGCGGGGGGCTGCCTGAGCGTCAACGAAAAGTTTACGCAGAAGGCCG
                                 401                                                                                 500
L. mexicana VDAC-2 ORF   (401)   TGGAGTTGGCGCTCAGCATGGCCGCCGTTGACGGCGTTCAGGTCGGCTGCGGCACCAAGTATGACCTCAAGTCGAAGACAATCGACTGGACGGCGGGGTG
      VDAC-2 Wild_Type   (401)   TGGAGTTGGCGCTCAGCATGGCCGCCGTTGACGGCGTTCAGGTCGGCTGCGGCACCAAGTATGACCTCAAGTCGAAGACAATCGACTGGACGGCGGGGTG
       VDAC-2 Pent_Res   (401)   TGGAGTTGGCGCTCAGCATGGCCGCCGTTGACGGCGTTCAGGTCGGCTGCGGCACCAAGTATGACCTCAAGTCGAAGACAATCGACTGGACGGCGGGGTG
                                 501                                                                                 600
L. mexicana VDAC-2 ORF   (501)   CCGCATGGAGGCCAAGAACGGTCTTGTGATGACGGCGCAGACGAACCGGTCGCTCGACGTCACGGCGAGCATGATTTCCAAGGCCGCACTGCACCCAAAG
      VDAC-2 Wild_Type   (501)   CCGCATGGAGGCCAAGAACGGTCTTGTGATGACGGCGCAGACGAACCGGTCGCTCGACGTCACGGCGAGCATGATTTCCAAGGCCGCACTGCACCCAAAG
       VDAC-2 Pent_Res   (501)   CCGCATGGAGGCCAAGAACGGTCTTGTGATGACGGCGCAGACGAACCGGTCGCTCGACGTCACGGCGAGCATGATTTCCAAGGCCGCACTGCACCCAAAG
                                 601                                                                                 700
L. mexicana VDAC-2 ORF   (601)   TTCCAGCCGTGCGTGGCAGCCACCGTGACGATGAACCCGCAATCGATGACATGGGACGGCTCCGTGGCACTGGAGTGGGGTTGTCAGGTGATTCTCGGCA
      VDAC-2 Wild_Type   (601)   TTCCAGCCGTGCGTGGCAGCCACCGTGACGATGAACCCGCAATCGATGACATGGGACGGCTCCGTGGCACTGGAGTGGGGTTGTCAGGTGATTCTCGGCA
       VDAC-2 Pent_Res   (601)   TTCCAGCCGTGCGTGGCAGCCACCGTGACGATGAACCCGCAATCGATGACATGGGACGGCTCCGTGGCACTGGAGTGGGGTTGTCAGGTGATTCTCGGCA
                                 701                                                                                 800
L. mexicana VDAC-2 ORF   (701)   ACGCGGCCAAGATTCGCTTCAGCAAGAACCTCGACTGGGTCGCCTCGTACATCGCCAACCTGCGCGACGGCTGGACGGTGGTGCTGTCGATGGACAAGAC
      VDAC-2 Wild_Type   (701)   ACGCGGCCAAGATTCGCTTCAGCAAGAACCTCGACTGGGTCGCCTCGTACATCGCCAACCTGCGCGACGGCTGGACGGTGGTGCTGTCGATGGACAAGAC
       VDAC-2 Pent_Res   (701)   ACGCGGCCAAGATTCGCTTCAGCAAGAACCTCGACTGGGTCGCCTCGTACATCGCCAACCTGCGCGACGGCTGGACGGTGGTGCTGTCGATGGACAAGAC
                                 801                          834
L. mexicana VDAC-2 ORF   (801)   CATGAGGGCGGGTCTGACGCTCACTCGCAACTGA
      VDAC-2 Wild_Type   (801)   CATGAGGGCGGGTCTGACGCTCACTCGCAACTGA
       VDAC-2 Pent_Res   (801)   CATGAGGGCGGGTCTGACGCTCACTCGCAACTGA
```

Figure 8-5. VDAC-2 DNA sequence alignment.

Alignment of the *L. mexicana* VDAC-2 ORF sequence (LmxM.02.0450) with the sequences derived from the wild type and pentamidine resistant lab strains.

```
                                1                                                                                   100
VDAC-1  (LmxM.02.0460)  (1)     MSQWQSSKGSAAKVYATPSLFKDYNKLTKDLLTKDFPTPNKWVLECKYKGPKDTFFINPKANSDGKMSADIEYVAACNGGLKVTVTPDIMREVKATAHYT
VDAC-1  (Lab)           (1)     MSQWQSSKGSAAKVYATPSLFKDYNKLTKDLLTKDFPTPNKWVLECKYKGPKDTFFINPKANSDGKMSADIEYVAACNGGLKVTVTPDIMREVKATAHYT
                                101                                                                                 200
VDAC-1  (LmxM.02.0460)  (101)   IQGHKVAVALQRLQDKYHYEISHETCVALSKRASINEKLTPTHVELGMGIDVAPNCQVGCGATYDIGANDCNWNIGCRYAGRGCEMAVRTNRLKTYHTSA
VDAC-1  (Lab)           (101)   IQGHKVAVALQRLQDKYHYEISHETCVALSKRASINEKLTPTHVELGMGIDVAPNCQVGCGATYDIGANDCNWNIGCRYAGRGCEMAVRTNRLKTYHTSA
                                201                                                 *                   291
VDAC-1  (LmxM.02.0460)  (201)   SAPCSFMLNGNRCMMRAAVEVMCGRGLGDKGVAVTVGVEATCPVHPANTIKARVDRDMKWAAAYIVKMADNWTACLSWDENMKVGVQMTHS
VDAC-1  (Lab)           (201)   SAPCSFMLNGNRCMMRAAVEVMCGRGLGDKGVAVTVGVEATCPVHPANTIKARVDRDMKWAVAYIVKMADNWTACLSWDENMKVGVQMTHS
```

Figure 8-6. Alignment of VDAC-1 amino acid sequences from *L. mexicana*.

Alignment of the *L. mexicana* VDAC-1 amino acid sequence (LmxM.02.0460) with the sequence derived from the wild type lab strain. The amino acid change, located at residue 262, is denoted by an asterisk.

```
                                1                                                                                   100
VDAC-2  (LmxM.02.0450)  (1)     MTTLFKDYSKGSNDLLTKNFSSCGAWKVESKSKAPKGTYALTTTSNTHGDVKLDIEGLTGDGAYYGKLCFTSKDLTDIQLTVRAEDLDNHRVEAIIGHKG
VDAC-2  (Lab)           (1)     MTTLFKDYSKGSNDLLTKNFSSCGAWKVESKSKAPKGTYALTTTSNTHGDVKLDIEGLTGDGAYYGKLCFTSKDLTDIQLTVRAEDLDNHRVEAIIGHKG
                                101                                                                                 200
VDAC-2  (LmxM.02.0450)  (101)   PAVCDISVEVNHQTVRPIAGGCLSVNEKFTQKAVELALSMAAVDGVQVGCGTKYDLKSKTIDWTAGCRMEAKNGLVMTAQTNRSLDVTASMISKAALHPK
VDAC-2  (Lab)           (101)   PAVCDISVEVNHQTVRPIAGGCLSVNEKFTQKAVELALSMAAVDGVQVGCGTKYDLKSKTIDWTAGCRMEAKNGLVMTAQTNRSLDVTASMISKAALHPK
                                201                                                 277
VDAC-2  (LmxM.02.0450)  (201)   FQPCVAATVTMNPQSMTWDGSVALEWGCQVILGNAAKIRFSKNLDWVASYIANLRDGWTVVLSMDKTMRAGLTLTRN
VDAC-2  (Lab)           (201)   FQPCVAATVTMNPQSMTWDGSVALEWGCQVILGNAAKIRFSKNLDWVASYIANLRDGWTVVLSMDKTMRAGLTLTRN
```

Figure 8-7. Alignment of VDAC-2 amino acid sequences from *L. mexicana*.

Alignment of the *L. mexicana* VDAC-2 amino acid sequence (LmxM.02.0450) with the sequence derived from the wild type lab strain.

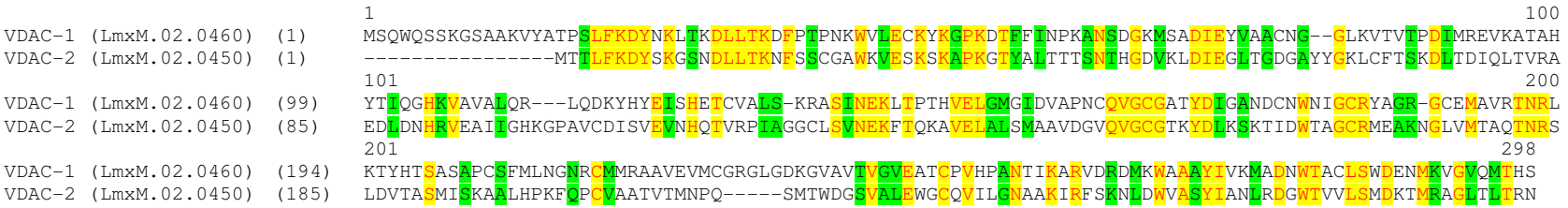


Figure 8-8. Alignment of VDAC-1 and VDAC-2 amino acid sequences from *L. mexicana*.

Amino acid sequences were obtained from TriTrypDB (Aslett, M *et al.* 2011)[28]. There is a 40.6 % sequence identity.

[28] http://tritrypdb.org

```
INSERT INTO 131_Results ( [Exact_Mass_(MIM)], Formula, CompCode )
SELECT DISTINCT Metabolites.[Exact_Mass_(MIM)], Metabolites.Formula, [109_Total KEGG database].CompCode
FROM Masses, [109_Total KEGG database] INNER JOIN Metabolites ON [109_Total KEGG database].Formula = Metabolites.Formula
WHERE (((Metabolites.[Exact_Mass_(MIM)])>(([Masses].[Mass]-([Forms]![Analysis]![PPMrange]/1000000)*[Masses].[Mass])) And
(Metabolites.[Exact_Mass_(MIM)])<([Masses].[Mass]+(([Forms]![Analysis]![PPMrange]/1000000)*[Masses].[Mass]))) AND ((Masses.Experiment)=[Forms]![Analysis]![ExpList]) AND
(([Forms]![Analysis]![TTestDW])='No')) OR (((Metabolites.[Exact_Mass_(MIM)])>(([Masses].[Mass]-([Forms]![Analysis]![PPMrange]/1000000)*[Masses].[Mass])) And
(Metabolites.[Exact_Mass_(MIM)])<([Masses].[Mass]+(([Forms]![Analysis]![PPMrange]/1000000)*[Masses].[Mass]))) AND ((Masses.Experiment)=[Forms]![Analysis]![ExpList]) AND
((Masses.PValue)<=[Forms]![Analysis]![PValueDW]));
```

Figure 8-9. SQL: 065_Analysis

```
DELETE [131_Results].Formula, [131_Results].[Exact_Mass_(MIM)], [131_Results].CompCode
FROM 131_Results
WHERE ((([131_Results].Formula) Like "*") AND (([131_Results].[Exact_Mass_(MIM)]) Like "*") AND (([131_Results].CompCode) Like "*"));
```

Figure 8-10. SQL: 071_Delete_all_records_in_131

```
DELETE [132_Unique_Formula_RESULTS].Formula, [132_Unique_Formula_RESULTS].[Exact_Mass_(MIM)]
FROM 132_Unique_Formula_RESULTS
WHERE ((([132_Unique_Formula_RESULTS].Formula) Like "*") AND (([132_Unique_Formula_RESULTS].[Exact_Mass_(MIM)]) Like "*"));
```

Figure 8-11. SQL: 070_Delete_all_records_in_132

```
INSERT INTO 132_Unique_Formula_RESULTS ( Formula, [Exact_Mass_(MIM)], KEGG_Isomers )
SELECT [131_Results].Formula, [131_Results].[Exact_Mass_(MIM)], Count([131_Results].Formula) AS CountOfFormula
FROM 131_Results
GROUP BY [131_Results].Formula, [131_Results].[Exact_Mass_(MIM)]
HAVING (((Count([131_Results].Formula))>0))
ORDER BY Count([131_Results].Formula);
```

Figure 8-12. SQL: 069_Add_Unique_Masses_To_New_Table_132

```
DELETE [139_Specific_Metabolites].[Exact_Mass_(MIM)]
FROM 139_Specific_Metabolites
WHERE ((([139_Specific_Metabolites].[Exact_Mass_(MIM)]) Like "*"));
```

Figure 8-13. SQL: 078_delete_values_from_139

```
INSERT INTO 139_Specific_Metabolites ( Metabolite, [Exact_Mass_(MIM)], Formula, CompCode )
SELECT [109_Total KEGG database].Metabolite, [109_Total KEGG database].[Exact_Mass_(MIM)], [109_Total KEGG database].Formula, [109_Total KEGG database].CompCode
FROM [109_Total KEGG database]
WHERE ((([109_Total KEGG database].Metabolite) Like "*" & [Forms]![Analysis]![metabolite] & "*"));
```

Figure 8-14. SQL: 077_Specific_Metabolite_Search

```
SELECT Masses.MASS, Masses.Experiment, Forms!Specific_Metabolite_Search_2!Metabolite_List AS [MIM of chosen metabolite]
FROM Masses
WHERE (((Masses.MASS) Between ([Forms]![Specific_Metabolite_Search_2]![Metabolite_List]-
(0.000001*[Forms]![Specific_Metabolite_Search_2]![Metabolite_List]*[Forms]![Specific_Metabolite_Search_2]![ppm_tolerance])) And
([Forms]![Specific_Metabolite_Search_2]![Metabolite_List]+(0.000001*[Forms]![Specific_Metabolite_Search_2]![Metabolite_List]*[Forms]![Specific_Metabolite_Search_2]![ppm_tolerance]))));
```

Figure 8-15. SQL: 081_search_the_masses_table_for_specific_metabolite

```
SELECT [109_Total KEGG database].Shortlist, [109_Total KEGG database].Metabolite, [109_Total KEGG database].[Exact_Mass_(MIM)], [109_Total KEGG database].Formula, [109_Total KEGG
database].CompCode, [109_Total KEGG database].KEGG_Link, [109_Total KEGG database].Image_Link, [109_Total KEGG database].InChi_ID
FROM [109_Total KEGG database]
WHERE ((([109_Total KEGG database].[Exact_Mass_(MIM)]) Between [Forms]![Multiple CompCode Search]![LOWER_MASS] And [Forms]![Multiple CompCode Search]![UPPER_MASS]))
ORDER BY [109_Total KEGG database].[Exact_Mass_(MIM)];
```

Figure 8-16. SQL: Mass Search (037)

```
SELECT [109_Total KEGG database].Shortlist, [109_Total KEGG database].Metabolite, [109_Total KEGG database].[Exact_Mass_(MIM)], [109_Total KEGG database].Formula
FROM [109_Total KEGG database]
WHERE ((([109_Total KEGG database].[Exact_Mass_(MIM)]) Between ([Forms]![Multiple CompCode Search]![Input MIM]-(0.000001*[Forms]![Multiple CompCode Search]![Input MIM]*[Forms]![Multiple CompCode
Search]![PPM_Tolerance])) And ([Forms]![Multiple CompCode Search]![Input MIM]+(0.000001*[Forms]![Multiple CompCode Search]![Input MIM]*[Forms]![Multiple CompCode Search]![PPM_Tolerance])))
ORDER BY [109_Total KEGG database].[Exact_Mass_(MIM)];
```

Figure 8-17. SQL: Mass (MIM) Search (038)

```
SELECT [109_Total KEGG database].Shortlist, [109_Total KEGG database].Metabolite, [109_Total KEGG database].[Exact_Mass_(MIM)], [109_Total KEGG database].Formula, [109_Total KEGG
database].CompCode, [109_Total KEGG database].KEGG_Link, [109_Total KEGG database].Image_Link, [109_Total KEGG database].InChi_ID
FROM [109_Total KEGG database]
WHERE ((([109_Total KEGG database].Formula)=[Forms]![multiple CompCode Search]![Input Molecular Formula]));
```

Figure 8-18. SQL: Molecular Formula Search (019)

```
SELECT [109_Total KEGG database].Shortlist, [109_Total KEGG database].Metabolite, [109_Total KEGG database].[Exact_Mass_(MIM)], [109_Total KEGG database].Formula, [109_Total KEGG
database].CompCode, [109_Total KEGG database].KEGG_Link, [109_Total KEGG database].Image_Link, [109_Total KEGG database].InChi_ID
FROM [109_Total KEGG database]
WHERE ((([109_Total KEGG database].Metabolite) Like "*" & [Forms]![Multiple CompCode Search]![Enter Metabolite Name] & "*"));
```

Figure 8-19. SQL: Metabolite Search (018) (A11)

```
SELECT [109_Total KEGG database].Shortlist, [109_Total KEGG database].Metabolite, [109_Total KEGG database].[Exact_Mass_(MIM)], [109_Total KEGG database].Formula, [109_Total KEGG
database].CompCode, [109_Total KEGG database].KEGG_Link, [109_Total KEGG database].Image_Link, [109_Total KEGG database].InChi_ID
FROM [109_Total KEGG database]
WHERE ((([109_Total KEGG database].CompCode)=[Forms]![multiple CompCode Search]![Input KEGG CompCode]));
```

Figure 8-20. SQL: KEGG CompCode Search (031) (A12)

Table 8-1. Fluorescence readings used to calculate the methylene blue EC_{50} values

log (M) [Methylene Blue]	Glucose and Proline			Glucose-rich			Proline-rich		
-5.1	6207.3	5558.7	5313.3	5024.3	4482.7	4441.7	5799.0	4674.7	5101.0
-5.4	7802.3	6817.7	6735.0	6062.0	5816.0	5119.0	7017.7	5613.3	6278.7
-5.7	10400.7	9008.3	8719.0	7844.3	7358.7	6663.3	8957.3	7296.0	8101.7
-6.0	16225.7	13924.0	13257.7	15832.7	10777.7	10600.3	9016.0	10556.7	7901.3
-6.3	44757.3	41395.7	34112.7	24283.0	20258.3	22283.0	5739.7	9208.3	5241.7
-6.6	53414.7	51828.7	53464.0	57089.0	48651.7	55694.3	20059.3	15671.0	9200.0
-6.9	56118.0	53779.0	55259.3	58183.7	57614.7	58380.0	52556.7	54403.3	54051.7
-7.2	57458.0	55677.7	56893.0	58056.0	58000.3	58172.7	52777.0	54212.0	54180.0
-7.5	57841.7	56309.7	56964.7	57969.7	57981.0	58005.7	52640.0	54019.3	54476.0
-7.8	57494.3	56866.7	57025.0	57686.7	58055.3	58071.3	*52044.3*	54217.3	53856.7
-8.1	57263.0	56465.0	56824.7	57585.0	57778.0	57974.3	*51715.3*	*53747.7*	*53775.3*
-9.0	57308.3	55539.3	54275.3	56674.3	55677.3	57681.7	*47630.7*	*53525.7*	*53580.7*

The seven values excluded from the proline-rich growth conditions are highlighted in grey.

9 References

Achterberg, V. and G. Gercken (1987). "Cytotoxicity of ester and ether lysophospholipids on Leishmania donovani promastigotes." Mol Biochem Parasitol **23**(2): 117-22.

Acosta-Serrano, A., E. Vassella, M. Liniger, C. Kunz Renggli, R. Brun, I. Roditi and P. T. Englund (2001). "The surface coat of procyclic Trypanosoma brucei: programmed expression and proteolytic cleavage of procyclin in the tsetse fly." Proc Natl Acad Sci U S A **98**(4): 1513-8.

Aebersold, R. and M. Mann (2003). "Mass spectrometry-based proteomics." Nature **422**(6928): 198-207.

Aga, E., D. M. Katschinski, G. van Zandbergen, H. Laufs, B. Hansen, K. Muller, W. Solbach and T. Laskay (2002). "Inhibition of the spontaneous apoptosis of neutrophil granulocytes by the intracellular parasite Leishmania major." J Immunol **169**(2): 898-905.

Agbe, A. and K. L. Yielding (1995). "Kinetoplasts play an important role in the drug responses of Trypanosoma brucei." J Parasitol **81**(6): 968-73.

Aksoy, S., W. C. Gibson and M. J. Lehane (2003). "Interactions between tsetse and trypanosomes with implications for the control of trypanosomiasis." Adv Parasitol **53**: 1-83.

Alban, A., S. O. David, L. Bjorkesten, C. Andersson, E. Sloge, S. Lewis and I. Currie (2003). "A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard." Proteomics **3**(1): 36-44.

Alexander, J., G. H. Coombs and J. C. Mottram (1998). "Leishmania mexicana cysteine proteinase-deficient mutants have attenuated virulence for mice and potentiate a Th1 response." J Immunol **161**(12): 6794-801.

Alpert, A. J. (1990). "Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds." J Chromatogr **499**: 177-96.

Alsford, S., S. Eckert, N. Baker, L. Glover, A. Sanchez-Flores, K. F. Leung, D. J. Turner, M. C. Field, M. Berriman, et al. (2012). "High-throughput decoding of antitrypanosomal drug efficacy and resistance." Nature **482**(7384): 232-6.

Aman, R. A., G. L. Kenyon and C. C. Wang (1985). "Cross-linking of the enzymes in the glycosome of Trypanosoma brucei." J Biol Chem **260**(11): 6966-73.

Aslett, M., C. Aurrecoechea, M. Berriman, J. Brestelli, B. P. Brunk, M. Carrington, D. P. Depledge, S. Fischer, B. Gajria, et al. (2011). "TriTrypDB: a functional genomic resource for the Trypanosomatidae." Nucleic Acids Res **38**(Database issue): D457-62.

Aurrecoechea, C., J. Brestelli, B. P. Brunk, J. Dommer, S. Fischer, B. Gajria, X. Gao, A. Gingle, G. Grant, et al. (2009). "PlasmoDB: a functional genomic database for malaria parasites." Nucleic Acids Res **37**(Database issue): D539-43.

Bacchi, C. J., J. Garofalo, D. Mockenhaupt, P. P. McCann, K. A. Diekema, A. E. Pegg, H. C. Nathan, E. A. Mullaney, L. Chunosoff, et al. (1983). "In vivo effects of alpha-DL-difluoromethylornithine on the metabolism and morphology of Trypanosoma brucei brucei." Mol Biochem Parasitol **7**(3): 209-25.

Bacchi, C. J., H. C. Nathan, T. Livingston, G. Valladares, M. Saric, P. D. Sayer, A. R. Njogu and A. B. Clarkson, Jr. (1990). "Differential susceptibility to DL-alpha-difluoromethylornithine in clinical isolates of Trypanosoma brucei rhodesiense." Antimicrob Agents Chemother **34**(6): 1183-8.

Bakker, B. M., M. C. Walsh, B. H. ter Kuile, F. I. Mensonides, P. A. Michels, F. R. Opperdoes and H. V. Westerhoff (1999). "Contribution of glucose transport to the control of the glycolytic flux in Trypanosoma brucei." Proc Natl Acad Sci U S A **96**(18): 10098-103.

Balogun, R. A. (1974). "Amino acids in the excreta of the tsetse fly, Glossina palpalis." Experientia **30**(3): 239-40.

Barrett, M. P. (1997). "The pentose phosphate pathway and parasitic protozoa." Parasitol Today **13**(1): 11-6.

Barrett, M. P. (2006). "The rise and fall of sleeping sickness." Lancet **367**(9520): 1377-8.

Barrett, M. P., D. W. Boykin, R. Brun and R. R. Tidwell (2007). "Human African trypanosomiasis: pharmacological re-engagement with a neglected disease." Br J Pharmacol **152**(8): 1155-71.

Barrett, M. P., R. J. Burchmore, A. Stich, J. O. Lazzari, A. C. Frasch, J. J. Cazzulo and S. Krishna (2003). "The trypanosomiases." Lancet **362**(9394): 1469-80.

Barrett, M. P., Z. Q. Zhang, H. Denise, C. Giroud and T. Baltz (1995). "A diamidine-resistant Trypanosoma equiperdum clone contains a P2 purine transporter with reduced substrate affinity." Mol Biochem Parasitol **73**(1-2): 223-9.

Barry, J. D., S. L. Hajduk, K. Vickerman and D. Le Ray (1979). "Detection of multiple variable antigen types in metacyclic populations of Trypanosoma brucei." Trans R Soc Trop Med Hyg **73**(2): 205-8.

Basselin, M., M. A. Badet-Denisot, F. Lawrence and M. Robert-Gero (1997). "Effects of pentamidine on polyamine level and biosynthesis in wild-type, pentamidine-treated, and pentamidine-resistant Leishmania." Exp Parasitol **85**(3): 274-82.

Basselin, M., G. H. Coombs and M. P. Barrett (2000). "Putrescine and spermidine transport in Leishmania." Mol Biochem Parasitol **109**(1): 37-46.

Basselin, M., H. Denise, G. H. Coombs and M. P. Barrett (2002). "Resistance to pentamidine in Leishmania mexicana involves exclusion of the drug from the mitochondrion." Antimicrob Agents Chemother **46**(12): 3731-8.

Basselin, M., F. Lawrence and M. Robert-Gero (1996). "Pentamidine uptake in Leishmania donovani and Leishmania amazonensis promastigotes and axenic amastigotes." Biochem J **315 (Pt 2)**: 631-4.

Basselin, M., F. Lawrence and M. Robert-Gero (1997). "Altered transport properties of pentamidine-resistant Leishmania donovani and L. amazonensis promastigotes." Parasitol Res **83**(5): 413-8.

Basselin, M. and M. Robert-Gero (1998). "Alterations in membrane fluidity, lipid metabolism, mitochondrial activity, and lipophosphoglycan expression in pentamidine-resistant Leishmania." Parasitol Res **84**(1): 78-83.

Bastin, P., T. J. Pullen, F. F. Moreira-Leite and K. Gull (2000). "Inside and outside of the trypanosome flagellum:a multifunctional organelle." Microbes Infect **2**(15): 1865-74.

Bastin, P., T. Sherwin and K. Gull (1998). "Paraflagellar rod is vital for trypanosome motility." Nature **391**(6667): 548.

Bates, P. A. (2007). "Transmission of Leishmania metacyclic promastigotes by phlebotomine sand flies." Int J Parasitol **37**(10): 1097-106.

Becker, I., P. Volkow, O. Velasco-Castrejon, N. Salaiza-Suazo, M. Berzunza-Cruz, J. S. Dominguez, A. Morales-Vargas, A. Ruiz-Remigio and R. Perez-Montfort (1999). "The efficacy of pentamidine combined with allopurinol and immunotherapy for the treatment of patients with diffuse cutaneous leishmaniasis." Parasitol Res **85**(3): 165-70.

Bellofatto, V. (2007). "Pyrimidine transport activities in trypanosomes." Trends Parasitol **23**(5): 187-9; discussion 190.

Berens, R. L., L. C. Deutsch-King and J. J. Marr (1980). "Leishmania donovani and Leishmania braziliensis: hexokinase, glucose 6-phosphate dehydrogenase, and pentose phosphate shunt activity." Exp Parasitol **49**(1): 1-8.

Berger, B. J., N. S. Carter and A. H. Fairlamb (1995). "Characterisation of pentamidine-resistant Trypanosoma brucei brucei." Mol Biochem Parasitol **69**(2): 289-98.

Berman, J. D. (1982). "In vitro susceptibility of antimony-resistant Leishmania to alternative drugs." J Infect Dis **145**(2): 279.

Berman, J. D. (1997). "Human leishmaniasis: clinical, diagnostic, and chemotherapeutic developments in the last 10 years." Clin Infect Dis **24**(4): 684-703.

Besteiro, S., M. P. Barrett, L. Riviere and F. Bringaud (2005). "Energy generation in insect stages of Trypanosoma brucei: metabolism in flux." Trends Parasitol **21**(4): 185-91.

Besteiro, S., M. Biran, N. Biteau, V. Coustou, T. Baltz, P. Canioni and F. Bringaud (2002). "Succinate secreted by Trypanosoma brucei is produced by a novel and unique glycosomal enzyme, NADH-dependent fumarate reductase." J Biol Chem **277**(41): 38001-12.

Blum, J. J. and F. R. Opperdoes (1994). "Secretion of sucrase by Leishmania donovani." J Eukaryot Microbiol **41**(3): 228-31.

Bochud-Allemann, N. and A. Schneider (2002). "Mitochondrial substrate level phosphorylation is essential for growth of procyclic Trypanosoma brucei." J Biol Chem **277**(36): 32849-54.

Boelaert, M., S. Rijal, S. Regmi, R. Singh, B. Karki, D. Jacquet, F. Chappuis, L. Campino, P. Desjeux, et al. (2004). "A comparative study of the effectiveness of diagnostic tests for visceral leishmaniasis." Am J Trop Med Hyg **70**(1): 72-7.

Boersema, P. J., S. Mohammed and A. J. Heck (2008). "Hydrophilic interaction liquid chromatography (HILIC) in proteomics." Anal Bioanal Chem **391**(1): 151-9.

Bouteille, B., O. Oukem, S. Bisser and M. Dumas (2003). "Treatment perspectives for human African trypanosomiasis." Fundam Clin Pharmacol **17**(2): 171-81.

Boyer, P. D. (1997). "The ATP synthase--a splendid molecular machine." Annu Rev Biochem **66**: 717-49.

Bradley, G., P. F. Juranka and V. Ling (1988). "Mechanism of multidrug resistance." Biochim Biophys Acta **948**(1): 87-128.

Bray, P. G., M. P. Barrett, S. A. Ward and H. P. de Koning (2003). "Pentamidine uptake and resistance in pathogenic protozoa: past, present and future." Trends Parasitol **19**(5): 232-9.

Bridges, D. J., M. K. Gould, B. Nerima, P. Maser, R. J. Burchmore and H. P. de Koning (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 **71**(4): 1098-108.

Bringaud, F., L. Riviere and V. Coustou (2006). "Energy metabolism of trypanosomatids: adaptation to available carbon sources." Mol Biochem Parasitol **149**(1): 1-9.

Brun, R. and Schonenberger (1979). "Cultivation and in vitro cloning or procyclic culture forms of Trypanosoma brucei in a semi-defined medium. Short communication." Acta Trop **36**(3): 289-92.

Brun, R., R. Schumacher, C. Schmid, C. Kunz and C. Burri (2001). "The phenomenon of treatment failures in Human African Trypanosomiasis." Trop Med Int Health **6**(11): 906-14.

Burchmore, R. J. and M. P. Barrett (2001). "Life in vacuoles--nutrient acquisition by Leishmania amastigotes." Int J Parasitol **31**(12): 1311-20.

Burchmore, R. J., D. Rodriguez-Contreras, K. McBride, P. Merkel, M. P. Barrett, G. Modi, D. Sacks and S. M. Landfear (2003). "Genetic characterization of glucose transporter function in Leishmania mexicana." Proc Natl Acad Sci U S A **100**(7): 3901-6.

Burri, C. and R. Brun (2003). "Eflornithine for the treatment of human African trypanosomiasis." Parasitol Res **90 Supp 1**: S49-52.

Burri, C., S. Nkunku, A. Merolle, T. Smith, J. Blum and R. Brun (2000). "Efficacy of new, concise schedule for melarsoprol in treatment of sleeping sickness caused by Trypanosoma brucei gambiense: a randomised trial." Lancet **355**(9213): 1419-25.

Butikofer, P., S. Ruepp, M. Boschung and I. Roditi (1997). "'GPEET' procyclin is the major surface protein of procyclic culture forms of Trypanosoma brucei brucei strain 427." Biochem J **326 (Pt 2)**: 415-23.

Capul, A. A., S. Hickerson, T. Barron, S. J. Turco and S. M. Beverley (2007). "Comparisons of mutants lacking the Golgi UDP-galactose or GDP-mannose transporters establish that phosphoglycans are important for promastigote but not amastigote virulence in Leishmania major." Infect Immun **75**(9): 4629-37.

Carter, N. S. and A. H. Fairlamb (1993). "Arsenical-resistant trypanosomes lack an unusual adenosine transporter." Nature **361**(6408): 173-6.

Caspi, R., T. Altman, J. M. Dale, K. Dreher, C. A. Fulcher, F. Gilham, P. Kaipa, A. S. Karthikeyan, A. Kothari, et al. (2010). "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases." Nucleic Acids Res **38**(Database issue): D473-9.

Cazzulo, J. J. (1992). "Aerobic fermentation of glucose by trypanosomatids." Faseb J **6**(13): 3153-61.

Chappuis, F., L. Loutan, P. Simarro, V. Lejon and P. Buscher (2005). "Options for field diagnosis of human african trypanosomiasis." Clin Microbiol Rev **18**(1): 133-46.

Chappuis, F., S. Sundar, A. Hailu, H. Ghalib, S. Rijal, R. W. Peeling, J. Alvar and M. Boelaert (2007). "Visceral leishmaniasis: what are the needs for diagnosis, treatment and control?" Nat Rev Microbiol **5**(11): 873-82.

Chassagnole, C., N. Noisommit-Rizzi, J. W. Schmid, K. Mauch and M. Reuss (2002). "Dynamic modeling of the central carbon metabolism of Escherichia coli." Biotechnol Bioeng **79**(1): 53-73.

Chaudhuri, M., R. D. Ott and G. C. Hill (2006). "Trypanosome alternative oxidase: from molecule to function." Trends Parasitol **22**(10): 484-91.

Chou, C. H., W. C. Chang, C. M. Chiu, C. C. Huang and H. D. Huang (2009). "FMM: a web server for metabolic pathway reconstruction and comparative analysis." Nucleic Acids Res **37**(Web Server issue): W129-34.

Colasante, C., M. Ellis, T. Ruppert and F. Voncken (2006). "Comparative proteomics of glycosomes from bloodstream form and procyclic culture form Trypanosoma brucei brucei." Proteomics **6**(11): 3275-93.

Coler, R. N., Y. Goto, L. Bogatzki, V. Raman and S. G. Reed (2007). "Leish-111f, a recombinant polyprotein vaccine that protects against visceral Leishmaniasis by elicitation of CD4+ T cells." Infect Immun **75**(9): 4648-54.

Coler, R. N. and S. G. Reed (2005). "Second-generation vaccines against leishmaniasis." Trends Parasitol **21**(5): 244-9.

Connor, R. J. (1994). "The impact of nagana." Onderstepoort J Vet Res **61**(4): 379-83.

Cottret, L., D. Wildridge, F. Vinson, M. P. Barrett, H. Charles, M. F. Sagot and F. Jourdan (2010). "MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks." Nucleic Acids Res **38**(Web Server issue): W132-7.

Coustou, V., S. Besteiro, M. Biran, P. Diolez, V. Bouchaud, P. Voisin, P. A. Michels, P. Canioni, T. Baltz, et al. (2003). "ATP generation in the Trypanosoma brucei procyclic form: cytosolic substrate level is essential, but not oxidative phosphorylation." J Biol Chem **278**(49): 49625-35.

Coustou, V., S. Besteiro, L. Riviere, M. Biran, N. Biteau, J. M. Franconi, M. Boshart, T. Baltz and F. Bringaud (2005). "A mitochondrial NADH-dependent fumarate reductase involved in the production of succinate excreted by procyclic Trypanosoma brucei." J Biol Chem **280**(17): 16559-70.

Coustou, V., M. Biran, S. Besteiro, L. Riviere, T. Baltz, J. M. Franconi and F. Bringaud (2006). "Fumarate is an essential intermediary metabolite produced by the procyclic Trypanosoma brucei." J Biol Chem **281**(37): 26832-46.

Coustou, V., M. Biran, M. Breton, F. Guegan, L. Riviere, N. Plazolles, D. Nolan, M. P. Barrett, J. M. Franconi, et al. (2008). "Glucose-induced remodeling of intermediary and energy metabolism in procyclic Trypanosoma brucei." J Biol Chem **283**(24): 16342-54.

Creek, D. J., A. Jankevics, K. E. Burgess, R. Breitling and M. P. Barrett (2012). "IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data." Bioinformatics **28**(7): 1048-9.

Croft, S. L., M. P. Barrett and J. A. Urbina (2005). "Chemotherapy of trypanosomiases and leishmaniasis." Trends Parasitol **21**(11): 508-12.

Croft, S. L., R. A. Neal, W. Pendergast and J. H. Chan (1987). "The activity of alkyl phosphorylcholines and related derivatives against Leishmania donovani." Biochem Pharmacol **36**(16): 2633-6.

Croft, S. L., K. Seifert and V. Yardley (2006). "Current scenario of drug development for leishmaniasis." Indian J Med Res **123**(3): 399-410.

Croft, S. L., S. Sundar and A. H. Fairlamb (2006). "Drug resistance in leishmaniasis." Clin Microbiol Rev **19**(1): 111-26.

Croft, S. L. and V. Yardley (2002). "Chemotherapy of leishmaniasis." Curr Pharm Des **8**(4): 319-42.

Cronin, C. N., D. P. Nolan and H. P. Voorheis (1989). "The enzymes of the classical pentose phosphate pathway display differential activities in procyclic and bloodstream forms of Trypanosoma brucei." FEBS Lett **244**(1): 26-30.

Damper, D. and C. L. Patton (1976). "Pentamidine transport and sensitivity in brucei-group trypanosomes." J Protozool **23**(2): 349-56.

Davies, C. R., P. Kaye, S. L. Croft and S. Sundar (2003). "Leishmaniasis: new approaches to disease control." Bmj **326**(7385): 377-82.

Davila, A. M. and H. Momen (2000). "Internal-transcribed-spacer (ITS) sequences used to explore phylogenetic relationships within Leishmania." Ann Trop Med Parasitol **94**(6): 651-4.

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 **59**(3): 586-92.

de Koning, H. P. (2008). "Ever-increasing complexities of diamidine and arsenical crossresistance in African trypanosomes." Trends Parasitol **24**(8): 345-9.

de Koning, H. P. and S. M. Jarvis (1999). "Adenosine transporters in bloodstream forms of Trypanosoma brucei brucei: substrate recognition motifs and affinity for trypanocidal drugs." Mol Pharmacol **56**(6): 1162-70.

de Koning, H. P. and S. M. Jarvis (2001). "Uptake of pentamidine in Trypanosoma brucei brucei is mediated by the P2 adenosine transporter and at least one novel, unrelated transporter." Acta Trop **80**(3): 245-50.

de Koning, W. and K. van Dam (1992). "A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH." Anal Biochem **204**(1): 118-23.

Delespaux, V. and H. P. de Koning (2007). "Drugs and drug resistance in African trypanosomiasis." Drug Resist Updat **10**(1-2): 30-50.

Dettmer, K., P. A. Aronov and B. D. Hammock (2007). "Mass spectrometry-based metabolomics." Mass Spectrom Rev **26**(1): 51-78.

Docampo, R. and S. N. Moreno (1999). "Acidocalcisome: A novel Ca2+ storage compartment in trypanosomatids and apicomplexan parasites." Parasitol Today **15**(11): 443-8.

Dodt, G., N. Braverman, C. Wong, A. Moser, H. W. Moser, P. Watkins, D. Valle and S. J. Gould (1995). "Mutations in the PTS1 receptor gene, PXR1, define complementation group 2 of the peroxisome biogenesis disorders." Nat Genet **9**(2): 115-25.

Domon, B. and R. Aebersold (2006). "Mass spectrometry and protein analysis." Science **312**(5771): 212-7.

Dunn, W. B. (2008). "Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes." Phys Biol **5**(1): 011001.

Ebikeme, C., J. Hubert, M. Biran, G. Gouspillou, P. Morand, N. Plazolles, F. Guegan, P. Diolez, J. M. Franconi, et al. (2010). "Ablation of succinate production from glucose metabolism in the procyclic trypanosomes induces metabolic switches to the glycerol 3-phosphate/dihydroxyacetone phosphate shuttle and to proline metabolism." J Biol Chem **285**(42): 32312-24.

El Rayah, I. E., R. Kaminsky, C. Schmid and K. H. El Malik (1999). "Drug resistance in Sudanese Trypanosoma evansi." Vet Parasitol **80**(4): 281-7.

Enanga, B., R. J. Burchmore, M. L. Stewart and M. P. Barrett (2002). "Sleeping sickness and the brain." Cell Mol Life Sci **59**(5): 845-58.

Englund, P. T., S. L. Hajduk and J. C. Marini (1982). "The molecular biology of trypanosomes." Annu Rev Biochem **51**: 695-726.

Faijes, M., A. E. Mars and E. J. Smid (2007). "Comparison of quenching and extraction methodologies for metabolome analysis of Lactobacillus plantarum." Microb Cell Fact **6**: 27.

Fairlamb, A. H. (2003). "Chemotherapy of human African trypanosomiasis: current and future prospects." Trends Parasitol **19**(11): 488-94.

Fairlamb, A. H., P. Blackburn, P. Ulrich, B. T. Chait and A. Cerami (1985). "Trypanothione: a novel bis(glutathionyl)spermidine cofactor for glutathione reductase in trypanosomatids." Science **227**(4693): 1485-7.

Fairlamb, A. H., G. B. Henderson, C. J. Bacchi and A. Cerami (1987). "In vivo effects of difluoromethylornithine on trypanothione and polyamine levels in bloodstream forms of Trypanosoma brucei." Mol Biochem Parasitol **24**(2): 185-91.

Fenn, J. B., M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science **246**(4926): 64-71.

Fiehn, O., J. Kopka, R. N. Trethewey and L. Willmitzer (2000). "Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry." Anal Chem **72**(15): 3573-80.

Galland, N., F. Demeure, V. Hannaert, E. Verplaetse, D. Vertommen, P. Van der Smissen, P. J. Courtoy and P. A. Michels (2007). "Characterization of the role of the receptors PEX5 and PEX7 in the import of proteins into glycosomes of Trypanosoma brucei." Biochim Biophys Acta **1773**(4): 521-35.

Gavgani, A. S., M. H. Hodjati, H. Mohite and C. R. Davies (2002). "Effect of insecticide-impregnated dog collars on incidence of zoonotic visceral leishmaniasis in Iranian children: a matched-cluster randomised trial." Lancet **360**(9330): 374-9.

Gossage, S. M., M. E. Rogers and P. A. Bates (2003). "Two separate growth phases during the development of Leishmania in sand flies: implications for understanding the life cycle." Int J Parasitol **33**(10): 1027-34.

Greengauz-Roberts, O., H. Stoppler, S. Nomura, H. Yamaguchi, J. R. Goldenring, R. H. Podolsky, J. R. Lee and W. S. Dynan (2005). "Saturation labeling with cysteine-reactive cyanine fluorescent dyes provides increased sensitivity for protein expression profiling of laser-microdissected clinical specimens." Proteomics **5**(7): 1746-57.

Gupta, N., N. Goyal, U. K. Singha, V. Bhakuni, R. Roy and A. K. Rastogi (1999). "Characterization of intracellular metabolites of axenic amastigotes of Leishmania donovani by 1H NMR spectroscopy." Acta Trop **73**(2): 121-33.

Hajduk, S. L. (1978). "Influence of DNA complexing compounds on the kinetoplast of trypanosomatids." Progress in Molecular and Subcellular Biology **6**: 158–200.

Hall, B. S., C. Bot and S. R. Wilkinson (2011). "Nifurtimox activation by trypanosomal type I nitroreductases generates cytotoxic nitrile metabolites." J Biol Chem **286**(15): 13088-95.

Handman, E. (1997). "Leishmania vaccines: old and new." Parasitol Today **13**(6): 236-8.

Handman, E. (2001). "Leishmaniasis: current status of vaccine development." Clin Microbiol Rev **14**(2): 229-43.

Handman, E. and D. V. Bullen (2002). "Interaction of Leishmania with the host macrophage." Trends Parasitol **18**(8): 332-4.

Hart, D. T., O. Misset, S. W. Edwards and F. R. Opperdoes (1984). "A comparison of the glycosomes (microbodies) isolated from Trypanosoma brucei bloodstream form and cultured procyclic trypomastigotes." Mol Biochem Parasitol **12**(1): 25-35.

Hemstrom, P. and K. Irgum (2006). "Hydrophilic interaction chromatography." J Sep Sci **29**(12): 1784-821.

Herwaldt, B. L. (1999). "Leishmaniasis." Lancet **354**(9185): 1191-9.

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

Hu, Q., R. J. Noll, H. Li, A. Makarov, M. Hardman and R. Graham Cooks (2005). "The Orbitrap: a new mass spectrometer." J Mass Spectrom **40**(4): 430-43.

Hunger-Glaser, I. and T. Seebeck (1997). "Deletion of the genes for the paraflagellar rod protein PFR-A in Trypanosoma brucei is probably lethal." Mol Biochem Parasitol **90**(1): 347-51.

Iten, M., E. Matovu, R. Brun and R. Kaminsky (1995). "Innate lack of susceptibility of Ugandan Trypanosoma brucei rhodesiense to DL-alpha-difluoromethylornithine (DFMO)." Trop Med Parasitol **46**(3): 190-4.

Iten, M., H. Mett, A. Evans, J. C. Enyaru, R. Brun and R. Kaminsky (1997). "Alterations in ornithine decarboxylase characteristics account for tolerance of Trypanosoma brucei rhodesiense to D,L-alpha-difluoromethylornithine." Antimicrob Agents Chemother **41**(9): 1922-5.

Jandera, P. (2011). "Stationary and mobile phases in hydrophilic interaction chromatography: a review." Anal Chim Acta **692**(1-2): 1-25.

Jannin, J. and P. Cattand (2004). "Treatment and control of human African trypanosomiasis." Curr Opin Infect Dis **17**(6): 565-71.

Jardim, A., W. Liu, E. Zheleznova and B. Ullman (2000). "Peroxisomal targeting signal-1 receptor protein PEX5 from Leishmania donovani. Molecular, biochemical, and immunocytochemical characterization." J Biol Chem **275**(18): 13637-44.

Jenni, L., S. Marti, J. Schweizer, B. Betschart, R. W. Le Page, J. M. Wells, A. Tait, P. Paindavoine, E. Pays, et al. (1986). "Hybrid formation between African trypanosomes during cyclical transmission." Nature **322**(6075): 173-5.

Jensen, B. C., D. Sivam, C. T. Kifer, P. J. Myler and M. Parsons (2009). "Widespread variation in transcript abundance within and across developmental stages of Trypanosoma brucei." BMC Genomics **10**: 482.

Jha, T. K., S. Sundar, C. P. Thakur, P. Bachmann, J. Karbwang, C. Fischer, A. Voss and J. Berman (1999). "Miltefosine, an oral agent, for the treatment of Indian visceral leishmaniasis." N Engl J Med **341**(24): 1795-800.

Jonscher, K. R. and J. R. Yates, 3rd (1997). "The quadrupole ion trap mass spectrometer--a small solution to a big challenge." Anal Biochem **244**(1): 1-15.

Jourdan, F., R. Breitling, M. P. Barrett and D. Gilbert (2008). "MetaNetter: inference and visualization of high-resolution metabolomic networks." Bioinformatics **24**(1): 143-5.

Kamleh, A., M. P. Barrett, D. Wildridge, R. J. Burchmore, R. A. Scheltema and D. G. Watson (2008). "Metabolomic profiling using Orbitrap Fourier transform mass spectrometry with hydrophilic interaction chromatography: a method with wide applicability to analysis of biomolecules." Rapid Commun Mass Spectrom **22**(12): 1912-8.

Kanehisa, M. and S. Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes." Nucleic Acids Res **28**(1): 27-30.

Kedzierski, L., Y. Zhu and E. Handman (2006). "Leishmania vaccines: progress and problems." Parasitology **133 Suppl**: S87-112.

Kell, D. B. (2006). "Systems biology, metabolic modelling and metabolomics in drug discovery and development." Drug Discov Today **11**(23-24): 1085-92.

Kell, D. B., M. Brown, H. M. Davey, W. B. Dunn, I. Spasic and S. G. Oliver (2005). "Metabolic footprinting and systems biology: the medium is the message." Nat Rev Microbiol **3**(7): 557-65.

Kellina, O. I. (1963). "[Tests of the Chemotherapeutic Activity of Monomycin in Experimental Cutaneous Leishmaniasis in White Mice]." Med Parazitol (Mosk) **32**: 572-6.

Kennedy, P. G. (2006). "Diagnostic and neuropathogenesis issues in human African trypanosomiasis." Int J Parasitol **36**(5): 505-12.

Khoo, S. H., J. Bond and D. W. Denning (1994). "Administering amphotericin B--a practical approach." J Antimicrob Chemother **33**(2): 203-13.

Killick-Kendrick, R., D. H. Molyneux and R. W. Ashford (1974). "Leishmania in phlebotomid sandflies. I. Modifications of the flagellum associated with attachment to the mid-gut and oesophageal valve of the sandfly." Proc R Soc Lond B Biol Sci **187**(1089): 409-19.

Killick-Kendrick, R., D. H. Molyneux and R. W. Ashford (1974). "Ultrastructural observations on the attachment of Leishmania in the sandfly." Trans R Soc Trop Med Hyg **68**(4): 269.

Knipling, E. F. (1946). "DDT to control insects affecting man." J Econ Entomol **39**: 360-6.

Krauth-Siegel, R. L., S. K. Meiering and H. Schmidt (2003). "The parasite-specific trypanothione metabolism of trypanosoma and leishmania." Biol Chem **384**(4): 539-49.

Kuriyama, M., M. C. Wang, L. D. Papsidero, C. S. Killian, T. Shimano, L. Valenzuela, T. Nishiura, G. P. Murphy and T. M. Chu (1980). "Quantitation of prostate-specific antigen in serum by a sensitive enzyme immunoassay." Cancer Res **40**(12): 4658-62.

Kyrpides, N. C. (1999). "Genomes OnLine Database (GOLD 1.0): a monitor of complete and ongoing genome projects world-wide." Bioinformatics **15**(9): 773-4.

Lamour, N., L. Riviere, V. Coustou, G. H. Coombs, M. P. Barrett and F. Bringaud (2005). "Proline metabolism in procyclic Trypanosoma brucei is down-regulated in the presence of glucose." J Biol Chem **280**(12): 11902-10.

Lanteri, C. A., M. L. Stewart, J. M. Brock, V. P. Alibu, S. R. Meshnick, R. R. Tidwell and M. P. Barrett (2006). "Roles for the Trypanosoma brucei P2 transporter in DB75 uptake and resistance." Mol Pharmacol **70**(5): 1585-92.

Lanteri, C. A., R. R. Tidwell and S. R. Meshnick (2008). "The mitochondrion is a site of trypanocidal action of the aromatic diamidine DB75 in bloodstream forms of Trypanosoma brucei." Antimicrob Agents Chemother **52**(3): 875-82.

Lanteri, C. A., B. L. Trumpower, R. R. Tidwell and S. R. Meshnick (2004). "DB75, a novel trypanocidal agent, disrupts mitochondrial function in Saccharomyces cerevisiae." Antimicrob Agents Chemother **48**(10): 3968-74.

Le Roch, K. G., J. R. Johnson, H. Ahiboh, D. W. Chung, J. Prudhomme, D. Plouffe, K. Henson, Y. Zhou, W. Witola, et al. (2008). "A systematic approach to understand the mechanism of action of the bisthiazolium compound T4 on the human malaria parasite, Plasmodium falciparum." BMC Genomics **9**: 513.

Leader, D. P., K. Burgess, D. Creek and M. P. Barrett (2011). "Pathos: a web facility that uses metabolic maps to display experimental changes in metabolites identified by mass spectrometry." Rapid Commun Mass Spectrom **25**(22): 3422-6.

Lewin, S., G. Schonian, N. El Tai, L. Oskam, P. Bastien and W. Presber (2002). "Strain typing in Leishmania donovani by using sequence-confirmed amplified region analysis." Int J Parasitol **32**(10): 1267-76.

Li, F., S. B. Hua, C. C. Wang and K. M. Gottesdiener (1996). "Procyclic Trypanosoma brucei cell lines deficient in ornithine decarboxylase activity." Mol Biochem Parasitol **78**(1-2): 227-36.

Lilley, K. S. and D. B. Friedman (2004). "All about DIGE: quantification technology for differential-display 2D-gel proteomics." Expert Rev Proteomics **1**(4): 401-9.

Lindqvist, Y., G. Schneider, U. Ermler and M. Sundstrom (1992). "Three-dimensional structure of transketolase, a thiamine diphosphate dependent enzyme, at 2.5 A resolution." Embo J **11**(7): 2373-9.

Liu, B., Y. Liu, S. A. Motyka, E. E. Agbo and P. T. Englund (2005). "Fellowship of the rings: the replication of kinetoplast DNA." Trends Parasitol **21**(8): 363-9.

Ludewig, G., J. M. Williams, Y. Li and C. Staben (1994). "Effects of pentamidine isethionate on Saccharomyces cerevisiae." Antimicrob Agents Chemother **38**(5): 1123-8.

Lyons, T. J. and A. Basu (2012). "Biomarkers in diabetes: hemoglobin A1c, vascular and tissue markers." Transl Res **159**(4): 303-12.

Maga, J. A. and J. H. LeBowitz (1999). "Unravelling the kinetoplastid paraflagellar rod." Trends Cell Biol **9**(10): 409-13.

Maga, J. A., T. Sherwin, S. Francis, K. Gull and J. H. LeBowitz (1999). "Genetic dissection of the Leishmania paraflagellar rod, a unique flagellar cytoskeleton structure." J Cell Sci **112 (Pt 16)**: 2753-63.

Magez, S., G. Caljon, T. Tran, B. Stijlemans and M. Radwanska (2009). "Current status of vaccination against African trypanosomiasis." Parasitology **137**(14): 2017-27.

Mallick, P. and B. Kuster (2010). "Proteomics: a pragmatic perspective." Nat Biotechnol **28**(7): 695-709.

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

Maser, P., C. Sutterlin, A. Kralli and R. Kaminsky (1999). "A nucleoside transporter from Trypanosoma brucei involved in drug resistance." Science **285**(5425): 242-4.

Mashego, M. R., K. Rumbold, M. De Mey, E. Vandamme, W. Soetaert and J. J. Heijnen (2007). "Microbial metabolomics: past, present and future methodologies." Biotechnol Lett **29**(1): 1-16.

Matovu, E., F. Geiser, V. Schneider, P. Maser, J. C. Enyaru, R. Kaminsky, S. Gallati and T. Seebeck (2001). "Genetic variants of the TbAT1 adenosine transporter from African trypanosomes in relapse infections following melarsoprol therapy." Mol Biochem Parasitol **117**(1): 73-81.

Matovu, E., M. L. Stewart, F. Geiser, R. Brun, P. Maser, L. J. Wallace, R. J. Burchmore, J. C. Enyaru, M. P. Barrett, et al. (2003). "Mechanisms of arsenical and diamidine uptake and resistance in Trypanosoma brucei." Eukaryot Cell **2**(5): 1003-8.

Maugeri, D. A. and J. J. Cazzulo (2004). "The pentose phosphate pathway in Trypanosoma cruzi." FEMS Microbiol Lett **234**(1): 117-23.

Maugeri, D. A., J. J. Cazzulo, R. J. Burchmore, M. P. Barrett and P. O. Ogbunude (2003). "Pentose phosphate metabolism in Leishmania mexicana." Mol Biochem Parasitol **130**(2): 117-25.

McConville, M. J., K. A. Mullin, S. C. Ilgoutz and R. D. Teasdale (2002). "Secretory pathway of trypanosomatid parasites." Microbiol Mol Biol Rev **66**(1): 122-54; table of contents.

McCulloch, R. and D. Horn (2009). "What has DNA sequencing revealed about the VSG expression sites of African trypanosomes?" Trends Parasitol **25**(8): 359-63.

McNamara, L. E., M. J. Dalby, M. O. Riehle and R. Burchmore (2010). "Fluorescence two-dimensional difference gel electrophoresis for biomaterial applications." J R Soc Interface **7 Suppl 1**: S107-18.

Michels, P. A., V. Hannaert and F. Bringaud (2000). "Metabolic aspects of glycosomes in trypanosomatidae - new data and views." Parasitol Today **16**(11): 482-9.

Misset, O., O. J. Bos and F. R. Opperdoes (1986). "Glycolytic enzymes of Trypanosoma brucei. Simultaneous purification, intraglycosomal concentrations and physical properties." Eur J Biochem **157**(2): 441-53.

Misslitz, A., J. C. Mottram, P. Overath and T. Aebischer (2000). "Targeted integration into a rRNA locus results in uniform and high level expression of transgenes in Leishmania amastigotes." Mol Biochem Parasitol **107**(2): 251-61.

Mitschke, L., C. Parthier, K. Schroder-Tittmann, J. Coy, S. Ludtke and K. Tittmann (2010). "The crystal structure of human transketolase and new insights into its mode of action." J Biol Chem **285**(41): 31559-70.

Morris, J. C., Z. Wang, M. E. Drew and P. T. Englund (2002). "Glycolysis modulates trypanosome glycoprotein expression as revealed by an RNAi library." Embo J **21**(17): 4429-38.

Mottram, J. C., A. E. Souza, J. E. Hutchison, R. Carter, M. J. Frame and G. H. Coombs (1996). "Evidence from disruption of the lmcpb gene array of Leishmania mexicana that cysteine proteinases are virulence factors." Proc Natl Acad Sci U S A **93**(12): 6008-13.

Mukherjee, A., P. K. Padmanabhan, M. H. Sahani, M. P. Barrett and R. Madhubala (2006). "Roles for mitochondria in pentamidine susceptibility and resistance in Leishmania donovani." Mol Biochem Parasitol **145**(1): 1-10.

Muller, M. (1983). "Mode of action of metronidazole on anaerobic bacteria and protozoa." Surgery **93**(1 Pt 2): 165-71.

Naidong, W. (2003). "Bioanalytical liquid chromatography tandem mass spectrometry methods on underivatized silica columns with aqueous/organic mobile phases." J Chromatogr B Analyt Technol Biomed Life Sci **796**(2): 209-24.

Nguyen, B., M. P. Lee, D. Hamelberg, A. Joubert, C. Bailly, R. Brun, S. Neidle and W. D. Wilson (2002). "Strong binding in the DNA minor groove by an aromatic diamidine with a shape that does not match the curvature of the groove." J Am Chem Soc **124**(46): 13680-1.

Nilsson, U., L. Meshalkina, Y. Lindqvist and G. Schneider (1997). "Examination of substrate binding in thiamin diphosphate-dependent transketolase by protein crystallography and site-directed mutagenesis." J Biol Chem **272**(3): 1864-9.

O'Farrell, P. H. (1975). "High resolution two-dimensional electrophoresis of proteins." J Biol Chem **250**(10): 4007-21.

Ogbunude, P. O., N. Lamour and M. P. Barrett (2007). "Molecular cloning, expression and characterization of ribokinase of Leishmania major." Acta Biochim Biophys Sin (Shanghai) **39**(6): 462-6.

Olsen, J. V., L. M. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning, et al. (2005). "Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap." Mol Cell Proteomics **4**(12): 2010-21.

Opperdoes, F. R. and G. H. Coombs (2007). "Metabolism of Leishmania: proven and predicted." Trends Parasitol **23**(4): 149-58.

Opperdoes, F. R. and J. P. Szikora (2006). "In silico prediction of the glycosomal enzymes of Leishmania major and trypanosomes." Mol Biochem Parasitol **147**(2): 193-206.

Osman, A. S., F. W. Jennings and P. H. Holmes (1992). "The rapid development of drug-resistance by Trypanosoma evansi in immunosuppressed mice." Acta Trop **50**(3): 249-57.

Paine, M. F., M. Z. Wang, C. N. Generaux, D. W. Boykin, W. D. Wilson, H. P. De Koning, C. A. Olson, G. Pohlig, C. Burri, et al. (2010). "Diamidines for human African trypanosomiasis." Curr Opin Investig Drugs **11**(8): 876-83.

Pal, A., B. S. Hall and M. C. Field (2002). "Evidence for a non-LDL-mediated entry route for the trypanocidal drug suramin in Trypanosoma brucei." Mol Biochem Parasitol **122**(2): 217-21.

Pandey, A. and M. Mann (2000). "Proteomics to study genes and genomes." Nature **405**(6788): 837-46.

Pappin, D. J., P. Hojrup and A. J. Bleasby (1993). "Rapid identification of proteins by peptide-mass fingerprinting." Curr Biol **3**(6): 327-32.

Pays, E., B. Vanhollebeke, L. Vanhamme, F. Paturiaux-Hanocq, D. P. Nolan and D. Perez-Morga (2006). "The trypanolytic factor of human serum." Nat Rev Microbiol **4**(6): 477-86.

Peacock, L., V. Ferris, R. Sharma, J. Sunter, M. Bailey, M. Carrington and W. Gibson (2011). "Identification of the meiotic life cycle stage of Trypanosoma brucei in the tsetse fly." Proc Natl Acad Sci U S A **108**(9): 3671-6.

Pepin, J. and F. Milord (1994). "The treatment of human African trypanosomiasis." Adv Parasitol **33**: 1-47.

Perkins, D. N., D. J. Pappin, D. M. Creasy and J. S. Cottrell (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." Electrophoresis **20**(18): 3551-67.

Phillips, M. A. and C. C. Wang (1987). "A Trypanosoma brucei mutant resistant to alpha-difluoromethylornithine." Mol Biochem Parasitol **22**(1): 9-17.

Pollakis, G., R. W. Grady, H. A. Dieck and A. B. Clarkson, Jr. (1995). "Competition between inhibitors of the trypanosome alternative oxidase (TAO) and reduced coenzyme Q9." Biochem Pharmacol **50**(8): 1207-10.

Poulin, R., L. Lu, B. Ackermann, P. Bey and A. E. Pegg (1992). "Mechanism of the irreversible inactivation of mouse ornithine decarboxylase by alpha-difluoromethylornithine. Characterization of sequences at the inhibitor and coenzyme binding sites." J Biol Chem **267**(1): 150-8.

Priotto, G., C. Fogg, M. Balasegaram, O. Erphas, A. Louga, F. Checchi, S. Ghabri and P. Piola (2006). "Three drug combinations for late-stage Trypanosoma brucei gambiense sleeping sickness: a randomized clinical trial in Uganda." PLoS Clin Trials **1**(8): e39.

Priotto, G., S. Kasparian, W. Mutombo, D. Ngouama, S. Ghorashian, U. Arnold, S. Ghabri, E. Baudin, V. Buard, et al. (2009). "Nifurtimox-eflornithine combination therapy for second-stage African Trypanosoma brucei gambiense trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial." Lancet **374**(9683): 56-64.

Pusnik, M., F. Charriere, P. Maser, R. F. Waller, M. J. Dagley, T. Lithgow and A. Schneider (2009). "The single mitochondrial porin of Trypanosoma brucei is the main metabolite transporter in the outer mitochondrial membrane." Mol Biol Evol **26**(3): 671-80.

Pusnik, M., O. Schmidt, A. J. Perry, S. Oeljeklaus, M. Niemann, B. Warscheid, T. Lithgow, C. Meisinger and A. Schneider (2011). "Mitochondrial preprotein translocase of trypanosomatids has a bacterial origin." Curr Biol **21**(20): 1738-43.

Ramos, H., E. Valdivieso, M. Gamargo, F. Dagger and B. E. Cohen (1996). "Amphotericin B kills unicellular leishmanias by forming aqueous pores permeable to small cations and anions." J Membr Biol **152**(1): 65-75.

Raz, B., M. Iten, Y. Grether-Buhler, R. Kaminsky and R. Brun (1997). "The Alamar Blue assay to determine drug sensitivity of African trypanosomes (T.b. rhodesiense and T.b. gambiense) in vitro." Acta Trop **68**(2): 139-47.

Reithinger, R. and J. C. Dujardin (2007). "Molecular diagnosis of leishmaniasis: current status and future applications." J Clin Microbiol **45**(1): 21-5.

Ribeiro-Gomes, F. L. and D. Sacks (2010). "The influence of early neutrophil-Leishmania interactions on the host immune response to infection." Front Cell Infect Microbiol **2**: 59.

Ritmeijer, K., C. Davies, R. van Zorge, S. J. Wang, J. Schorscher, S. I. Dongu'du and R. N. Davidson (2007). "Evaluation of a mass distribution programme for fine-mesh impregnated bednets against visceral leishmaniasis in eastern Sudan." Trop Med Int Health **12**(3): 404-14.

Rittig, M. G. and C. Bogdan (2000). "Leishmania-host-cell interaction: complexities and alternative views." Parasitol Today **16**(7): 292-7.

Roditi, I., A. Furger, S. Ruepp, N. Schurch and P. Butikofer (1998). "Unravelling the procyclin coat of Trypanosoma brucei." Mol Biochem Parasitol **91**(1): 117-30.

Roditi, I., H. Schwarz, T. W. Pearson, R. P. Beecroft, M. K. Liu, J. P. Richardson, H. J. Buhring, J. Pleiss, R. Bulow, et al. (1989). "Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of Trypanosoma brucei." J Cell Biol **108**(2): 737-46.

Rodrigues, C. O., D. A. Scott and R. Docampo (1999). "Characterization of a vacuolar pyrophosphatase in Trypanosoma brucei and its localization to acidocalcisomes." Mol Cell Biol **19**(11): 7712-23.

Rogers, M. E., M. L. Chance and P. A. Bates (2002). "The role of promastigote secretory gel in the origin and transmission of the infective stage of Leishmania mexicana by the sandfly Lutzomyia longipalpis." Parasitology **124**(Pt 5): 495-507.

Rogers, M. E., T. Ilg, A. V. Nikolaev, M. A. Ferguson and P. A. Bates (2004). "Transmission of cutaneous leishmaniasis by sand flies is enhanced by regurgitation of fPPG." Nature **430**(6998): 463-7.

Rosenzweig, D., D. Smith, F. Opperdoes, S. Stern, R. W. Olafson and D. Zilberstein (2008). "Retooling Leishmania metabolism: from sand fly gut to human macrophage." Faseb J **22**(2): 590-602.

Rosypal, A. C., K. A. Werbovetz, M. Salem, C. E. Stephens, A. Kumar, D. W. Boykin, J. E. Hall and R. R. Tidwell (2008). "Inhibition by Dications of in vitro growth of Leishmania major and Leishmania tropica: causative agents of old world cutaneous leishmaniasis." J Parasitol **94**(3): 743-9.

Sacks, D. and A. Sher (2002). "Evasion of innate immunity by parasitic protozoa." Nat Immunol **3**(11): 1041-7.

Sands, M., M. A. Kron and R. B. Brown (1985). "Pentamidine: a review." Rev Infect Dis **7**(5): 625-34.

Santrich, C., L. Moore, T. Sherwin, P. Bastin, C. Brokaw, K. Gull and J. H. LeBowitz (1997). "A motility function for the paraflagellar rod of Leishmania parasites revealed by PFR-2 gene knockouts." Mol Biochem Parasitol **90**(1): 95-109.

Saunders, E. C., D. E. S. DP, T. Naderer, M. F. Sernee, J. E. Ralton, M. A. Doyle, J. I. Macrae, J. L. Chambers, J. Heng, et al. (2010). "Central carbon metabolism of Leishmania parasites." Parasitology **137**(9): 1303-13.

Scheltema, R. A., A. Jankevics, R. C. Jansen, M. A. Swertz and R. Breitling (2011). "PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis." Anal Chem **83**(7): 2786-93.

Scheltema, R. A., A. Kamleh, D. Wildridge, C. Ebikeme, D. G. Watson, M. P. Barrett, R. C. Jansen and R. Breitling (2008). "Increasing the mass accuracy of high-resolution LC-MS data using background ions: a case study on the LTQ-Orbitrap." Proteomics **8**(22): 4647-56.

Schonian, G., A. Nasereddin, N. Dinse, C. Schweynoch, H. D. Schallig, W. Presber and C. L. Jaffe (2003). "PCR diagnosis and characterization of Leishmania in local and imported clinical samples." Diagn Microbiol Infect Dis **47**(1): 349-58.

Shapiro, T. A. (1993). "Inhibition of topoisomerases in African trypanosomes." Acta Trop **54**(3-4): 251-60.

Shaw, J., R. Rowlinson, J. Nickson, T. Stone, A. Sweet, K. Williams and R. Tonge (2003). "Evaluation of saturation labelling two-dimensional difference gel electrophoresis fluorescent dyes." Proteomics **3**(7): 1181-95.

Shin, M. H., Y. Lee do, K. H. Liu, O. Fiehn and K. H. Kim (2010). "Evaluation of sampling and extraction methodologies for the global metabolic profiling of Saccharophagus degradans." Anal Chem **82**(15): 6660-6.

Shoshan-Barmatz, V., V. De Pinto, M. Zweckstetter, Z. Raviv, N. Keinan and N. Arbel (2010). "VDAC, a multi-functional mitochondrial protein regulating cell life and death." Mol Aspects Med **31**(3): 227-85.

Silva, C. F., M. B. Meuser, E. M. De Souza, M. N. Meirelles, C. E. Stephens, P. Som, D. W. Boykin and M. N. Soeiro (2007). "Cellular effects of reversed amidines on Trypanosoma cruzi." Antimicrob Agents Chemother **51**(11): 3803-9.

Simpson, L. and J. Shaw (1989). "RNA editing and the mitochondrial cryptogenes of kinetoplastid protozoa." Cell **57**(3): 355-66.

Singh, V. P., A. Ranjan, R. K. Topno, R. B. Verma, N. A. Siddique, V. N. Ravidas, N. Kumar, K. Pandey and P. Das (2012). "Estimation of under-reporting of visceral leishmaniasis cases in Bihar, India." Am J Trop Med Hyg **82**(1): 9-11.

Sitek, B., J. Luttges, K. Marcus, G. Kloppel, W. Schmiegel, H. E. Meyer, S. A. Hahn and K. Stuhler (2005). "Application of fluorescence difference gel electrophoresis saturation labelling for the analysis of microdissected precursor lesions of pancreatic ductal adenocarcinoma." Proteomics **5**(10): 2665-79.

Skeiky, Y. A., R. N. Coler, M. Brannon, E. Stromberg, K. Greeson, R. T. Crane, J. R. Webb, A. Campos-Neto and S. G. Reed (2002). "Protective efficacy of a tandemly linked, multi-subunit recombinant leishmanial vaccine (Leish-111f) formulated in MPL adjuvant." Vaccine **20**(27-28): 3292-303.

Smith, C. A., E. J. Want, G. O'Maille, R. Abagyan and G. Siuzdak (2006). "XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification." Anal Chem **78**(3): 779-87.

Sommer, J. M., Q. L. Cheng, G. A. Keller and C. C. Wang (1992). "In vivo import of firefly luciferase into the glycosomes of Trypanosoma brucei and mutational analysis of the C-terminal targeting signal." Mol Biol Cell **3**(7): 749-59.

Soto, J., B. A. Arana, J. Toledo, N. Rizzo, J. C. Vega, A. Diaz, M. Luz, P. Gutierrez, M. Arboleda, et al. (2004). "Miltefosine for new world cutaneous leishmaniasis." Clin Infect Dis **38**(9): 1266-72.

Steen, H. and M. Mann (2004). "The ABC's (and XYZ's) of peptide sequencing." Nat Rev Mol Cell Biol **5**(9): 699-711.

Stephens, C. E., R. Brun, M. M. Salem, K. A. Werbovetz, F. Tanious, W. D. Wilson and D. W. Boykin (2003). "The activity of diguanidino and 'reversed' diamidino 2,5-diarylfurans versus Trypanosoma cruzi and Leishmania donovani." Bioorg Med Chem Lett **13**(12): 2065-9.

Sternberg, J. M. (1998). "Immunobiology of African trypanosomiasis." Chem Immunol **70**: 186-99.

Stoffel, S. A., V. P. Alibu, J. Hubert, C. Ebikeme, J. C. Portais, F. Bringaud, M. E. Schweingruber and M. P. Barrett (2011). "Transketolase in Trypanosoma brucei." Mol Biochem Parasitol **179**(1): 1-7.

Stuart, K. (1983). "Kinetoplast DNA, mitochondrial DNA with a difference." Mol Biochem Parasitol **9**(2): 93-104.

Suhre, K. and P. Schmitt-Kopplin (2008). "MassTRIX: mass translator into pathways." Nucleic Acids Res **36**(Web Server issue): W481-4.

Sundar, S., N. Agrawal, R. Arora, D. Agarwal, M. Rai and J. Chakravarty (2009). "Short-course paromomycin treatment of visceral leishmaniasis in India: 14-day vs 21-day treatment." Clin Infect Dis **49**(6): 914-8.

Sundar, S., L. B. Gupta, M. K. Makharia, M. K. Singh, A. Voss, F. Rosenkaimer, J. Engel and H. W. Murray (1999). "Oral treatment of visceral leishmaniasis with miltefosine." Ann Trop Med Parasitol **93**(6): 589-97.

Sundar, S., T. K. Jha, C. P. Thakur, J. Engel, H. Sindermann, C. Fischer, K. Junge, A. Bryceson and J. Berman (2002). "Oral miltefosine for Indian visceral leishmaniasis." N Engl J Med **347**(22): 1739-46.

Sundar, S., T. K. Jha, C. P. Thakur, M. Mishra, V. P. Singh and R. Buffels (2003). "Single-dose liposomal amphotericin B in the treatment of visceral leishmaniasis in India: a multicenter study." Clin Infect Dis **37**(6): 800-4.

Sundar, S., D. K. More, M. K. Singh, V. P. Singh, S. Sharma, A. Makharia, P. C. Kumar and H. W. Murray (2000). "Failure of pentavalent antimony in visceral leishmaniasis in India: report from the center of the Indian epidemic." Clin Infect Dis **31**(4): 1104-7.

Sundar, S. and M. Rai (2002). "Advances in the treatment of leishmaniasis." Curr Opin Infect Dis **15**(6): 593-8.

Sundar, S., F. Rosenkaimer, M. K. Makharia, A. K. Goyal, A. K. Mandal, A. Voss, P. Hilgard and H. W. Murray (1998). "Trial of oral miltefosine for visceral leishmaniasis." Lancet **352**(9143): 1821-3.

Szoor, B., I. Ruberto, R. Burchmore and K. R. Matthews (2010). "A novel phosphatase cascade regulates differentiation in Trypanosoma brucei via a glycosomal signaling pathway." Genes Dev **24**(12): 1306-16.

t'Kindt, R., R. A. Scheltema, A. Jankevics, K. Brunker, S. Rijal, J. C. Dujardin, R. Breitling, D. G. Watson, G. H. Coombs, et al. (2010). "Metabolomics to unveil and understand phenotypic diversity between pathogen populations." PLoS Negl Trop Dis **4**(11): e904.

Tautenhahn, R., G. J. Patti, D. Rinehart and G. Siuzdak (2010). "XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data." Anal Chem.

Tetaud, E., I. Lecuix, T. Sheldrake, T. Baltz and A. H. Fairlamb (2002). "A new expression vector for Crithidia fasciculata and Leishmania." Mol Biochem Parasitol **120**(2): 195-204.

Turco, S. J., G. F. Spath and S. M. Beverley (2001). "Is lipophosphoglycan a virulence factor? A surprising diversity between Leishmania species." Trends Parasitol **17**(5): 223-6.

Tyler, K. M., P. G. Higgs, K. R. Matthews and K. Gull (2001). "Limitation of Trypanosoma brucei parasitaemia results from density-dependent parasite differentiation and parasite killing by the host immune response." Proc Biol Sci **268**(1482): 2235-43.

Unlu, M., M. E. Morgan and J. S. Minden (1997). "Difference gel electrophoresis: a single gel method for detecting changes in protein extracts." Electrophoresis **18**(11): 2071-7.

Unwin, R. D., C. A. Evans and A. D. Whetton (2006). "Relative quantification in proteomics: new approaches for biochemistry." Trends Biochem Sci **31**(8): 473-84.

Uzonna, J. E., G. F. Spath, S. M. Beverley and P. Scott (2004). "Vaccination with phosphoglycan-deficient Leishmania major protects highly susceptible mice from virulent challenge without inducing a strong Th1 response." J Immunol **172**(6): 3793-7.

Van der Ploeg, L. H., K. Gottesdiener and M. G. Lee (1992). "Antigenic variation in African trypanosomes." Trends Genet **8**(12): 452-7.

Van Hellemond, J. J., B. Simons, F. F. Millenaar and A. G. Tielens (1998). "A gene encoding the plant-like alternative oxidase is present in Phytomonas but absent in Leishmania spp." J Eukaryot Microbiol **45**(4): 426-30.

Van Hoof, L. M. (1947). "Observations on trypanosomiasis in the Belgian Congo." Trans R Soc Trop Med Hyg **40**(6): 728-61.

van Weelden, S. W., B. Fast, A. Vogt, P. van der Meer, J. Saas, J. J. van Hellemond, A. G. Tielens and M. Boshart (2003). "Procyclic Trypanosoma brucei do not use Krebs cycle activity for energy generation." J Biol Chem **278**(15): 12854-63.

van Weelden, S. W., J. J. van Hellemond, F. R. Opperdoes and A. G. Tielens (2005). "New functions for parts of the Krebs cycle in procyclic Trypanosoma brucei, a cycle not operating as a cycle." J Biol Chem **280**(13): 12451-60.

Vansterkenburg, E. L., I. Coppens, J. Wilting, O. J. Bos, M. J. Fischer, L. H. Janssen and F. R. Opperdoes (1993). "The uptake of the trypanocidal drug suramin in combination with low-density lipoproteins by Trypanosoma brucei and its possible mode of action." Acta Trop **54**(3-4): 237-50.

Vassella, E., J. V. Den Abbeele, P. Butikofer, C. K. Renggli, A. Furger, R. Brun and I. Roditi (2000). "A major surface glycoprotein of trypanosoma brucei is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia." Genes Dev **14**(5): 615-26.

Vassella, E., M. Probst, A. Schneider, E. Studer, C. K. Renggli and I. Roditi (2004). "Expression of a major surface protein of Trypanosoma brucei insect forms is controlled by the activity of mitochondrial enzymes." Mol Biol Cell **15**(9): 3986-93.

Vassella, E., B. Reuner, B. Yutzy and M. Boshart (1997). "Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway." J Cell Sci **110 (Pt 21)**: 2661-71.

Veitch, N. J., D. A. Maugeri, J. J. Cazzulo, Y. Lindqvist and M. P. Barrett (2004). "Transketolase from Leishmania mexicana has a dual subcellular localization." Biochem J **382**(Pt 2): 759-67.

Vickerman, K. (1985). "Developmental cycles and biology of pathogenic trypanosomes." Br Med Bull **41**(2): 105-14.

Vickerman, K. and A. G. Luckins (1969). "Localization of variable antigens in the surface coat of Trypanosoma brucei using ferritin conjugated antibody." Nature **224**(5224): 1125-6.

Vickerman, K., P. J. Myler and K. Stuart (1993). "African Trypanosomiasis. In 'Immunology and molecular biology of parasitic infections'." Warren, K. S. (Editor). Blackwell Scientific Publications **Chapter 10**: 170-212.

Villas-Boas, S. G., J. Hojer-Pedersen, M. Akesson, J. Smedsgaard and J. Nielsen (2005). "Global metabolite analysis of yeast: evaluation of sample preparation methods." Yeast **22**(14): 1155-69.

Vincent, I. M., D. Creek, D. G. Watson, M. A. Kamleh, D. J. Woods, P. E. Wong, R. J. Burchmore and M. P. Barrett (2010). "A molecular mechanism for eflornithine resistance in African trypanosomes." PLoS Pathog **6**(11): e1001204.

Vincent, I. M., D. J. Creek, K. Burgess, D. J. Woods, R. J. Burchmore and M. P. Barrett (2012). "Untargeted Metabolomics Reveals a Lack Of Synergy between Nifurtimox and Eflornithine against Trypanosoma brucei." PLoS Negl Trop Dis **6**(5): e1618.

Vreysen, M. J. (2001). "Principles of area-wide integrated tsetse fly control using the sterile insect technique." Med Trop (Mars) **61**(4-5): 397-411.

Vreysen, M. J., K. M. Saleh, M. Y. Ali, A. M. Abdulla, Z. R. Zhu, K. G. Juma, V. A. Dyck, A. R. Msangi, P. A. Mkonyi, et al. (2000). "Glossina austeni (Diptera: Glossinidae) eradicated on the island of Unguja, Zanzibar, using the sterile insect technique." J Econ Entomol **93**(1): 123-35.

Walther, T. C. and M. Mann (2010). "Mass spectrometry-based proteomics in cell biology." J Cell Biol **190**(4): 491-500.

Wang, M. Z., X. Zhu, A. Srivastava, Q. Liu, J. M. Sweat, T. Pandharkar, C. E. Stephens, E. Riccio, T. Parman, et al. (2010). "Novel arylimidamides for treatment of visceral leishmaniasis." Antimicrob Agents Chemother **54**(6): 2507-16.

Weininger, D. (1988). "Smiles, a Chemical Language and Information-System .1. Introduction to Methodology and Encoding Rules." Journal of Chemical Information and Computer Sciences **28**(1): 31-36.

Wilkinson, S. R., C. Bot, J. M. Kelly and B. S. Hall (2011). "Trypanocidal activity of nitroaromatic prodrugs: current treatments and future perspectives." Curr Top Med Chem **11**(16): 2072-84.

Wilkinson, S. R. and J. M. Kelly (2009). "Trypanocidal drugs: mechanisms, resistance and new targets." Expert Rev Mol Med **11**: e31.

Wishart, D. S., C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong, et al. (2009). "HMDB: a knowledgebase for the human metabolome." Nucleic Acids Res **37**(Database issue): D603-10.

Woods, A., T. Sherwin, R. Sasse, T. H. MacRae, A. J. Baines and K. Gull (1989). "Definition of individual components within the cytoskeleton of Trypanosoma brucei by a library of monoclonal antibodies." J Cell Sci **93 (Pt 3)**: 491-500.

Woodward, R., M. J. Carden and K. Gull (1994). "Molecular characterisation of a novel, repetitive protein of the paraflagellar rod in Trypanosoma brucei." Mol Biochem Parasitol **67**(1): 31-9.

Wu, C. C. and J. R. Yates, 3rd (2003). "The application of mass spectrometry to membrane proteomics." Nat Biotechnol **21**(3): 262-7.

Xiao, Y., D. E. McCloskey and M. A. Phillips (2009). "RNA interference-mediated silencing of ornithine decarboxylase and spermidine synthase genes in Trypanosoma brucei provides insight into regulation of polyamine biosynthesis." Eukaryot Cell **8**(5): 747-55.

Yardley, V. and S. L. Croft (2000). "A comparison of the activities of three amphotericin B lipid formulations against experimental visceral and cutaneous leishmaniasis." Int J Antimicrob Agents **13**(4): 243-8.

Yi, Z. B., Y. Yan, Y. Z. Liang and Z. Bao (2007). "Evaluation of the antimicrobial mode of berberine by LC/ESI-MS combined with principal component analysis." J Pharm Biomed Anal **44**(1): 301-4.