A Proteomic Approach to the Investigation of Cymelarsan Resistance in *Trypanosoma brucei*

Aude Lise Foucher

Division of Infection and Immunity
Institute of Biomedical and Life Sciences

This thesis is submitted for the degree of Doctor of Philosophy
Faculty of Biomedical and Life Sciences
University of Glasgow
December 2003

Declaration

The results presented in this thesis is my own work, except when stated otherwise.

Aude Lise Foucher

Abstract

Two parasitic trypanosomatids, *T. b. gambiense* and *T. b. rhodesiense*, are responsible for around 150,000 death a year by Human African Trypanosomiasis (HAT). Prevalence values in some villages reaches up to 50%, making HAT the leading cause of mortality. There are currently only four drugs available for treating HAT, but problems with availability, administration, disease stage specificity and severe side effects have hindered their uses.

The most used drug is the arsenical, Melarsoprol, but resistance is increasingly reported in the field. Some reports have suggested that point mutations within the *TbAT1*, encoding the P2 transporter, were associated with drug resistance in both laboratory-derived strains and field isolates. However, analysis of a *ΔTbAT1* cell showed a low level of resistance and 50% of field isolates had a non-mutated P2 transporter. Consequently, it seems likely that multiple mechanisms can confer a resistance phenotype but these have been largely undescribed to date.

The advantage of a global proteomic expression profiling study of drug sensitive and drug resistant isogenic *T. brucei* lines is that, in principle, all possible mechanisms could be identified without pre-supposing a specific mechanism. The release of an almost complete *T. brucei* genome database, the advances in 2DE technology and mass spectrometry techniques, and the availability of drug sensitive and resistant isogenic *T. brucei* lines rendered this analysis feasible.

Subtractive analysis of the whole proteome of the drug sensitive and resistant isogenic line of STIB 386 identified two proteins whose loss of expression was associated with Cymelarsan resistance. Cymelarsan is a drug that is structurally and functionally closely related to Melarsoprol. Tandem mass spectrometry analysis matched the first protein onto either open reading frame (ORF) 274 or 277, present as identical tandem repeats on chromosome IX. This protein spot was shown to be an isoform of a spot located a similar molecular weight but with a more basic pI. As both ORFs have identical predicted amino acid sequences, are transcribed and no other homologues were found in the *T. brucei* genome, it is likely that the missing protein is an isoform of the adjacent protein which arose due to a post-translational modification. I have named this protein the Cymelarsan Associated Resistance (CAR) protein. The absence of the protein was not associated with Cymelarsan resistance in a pair of isogenic drug sensitive and resistant lines of 247.
The second protein spot associated with Cymelarsan resistance was identified as an arginine kinase from either ORF 342 or 343. These two ORF, present as tandem repeat on chromosome IX are not identical. However, the degree of sequencing obtained from the Tandem mass spectrometry analysis did not enable to differentiate between product of each ORF.

Finally, a diminished reduction of the Alamar Blue dye was found associated with Cymelarsan resistance in both pairs of isogenic lines analysed. An attempt was made to determine the mechanism that impaired the reduction of the Alamar blue dye in the drug resistant lines. The transport and efflux of the dye were analysed and did not account for the difference in reduction of the Alamar Blue dye. The dye was found to concentrate in vesicle-like structures thought to be acidocalcisomes.

Acknowledgements

Firstly, I would like to thanks my supervisors, Dr. Mike Turner, Dr. Jonathan Wastling and Prof. Andy Tait for their supervision, suggestions and support throughout the 3 years of my PhD.

I would also like to thank the Wellcome Trust for its financial support during the tenure of my PhD.

Thanks to all the people who have indirectly contributed to this PhD, to my friends in North lab and North/West lab who have been brilliant to work with.

I am especially grateful to Chris Wards for some useful (as well as completely useless) conversations, and for sharing the particularly difficult time of "gel pouring" with me.

I am grateful for the invaluable help and advice received from Dr Mike Barrett, Dr Sharon Kelly, Dr Richard Burchmore, Dr Gill Douce, Dr Andy Pitt, Dr Keith Mattews and Mrs Anne McIntosh.

Je tiens tout particulierement a remercier ma famille pour le soutient morale, la patience et l'amour qu'ils m'ont abondament donnee pendant toutes ces annees. Cette these n'aurais pu avoir lieu sans eux.

Table of Contents

List of Figures

Chapter I: General Introduction

Chapter II: Development of proteomic techniques for T. brucei

Chapter III: Assessment of algorithms for protein identification from peptide mass fingerprinting data

Chapter IV: A proteome map of T. brucei

Chapter V: Characterisation of isogenic drug sensitive and drug resistant lines of trypanosomes

Chapter VI: A proteomic approach to study Cymelarsan resistance in T. brucei line 386

Chapter VII Association of the CAR protein and Cymelarsan resistance in line 247 and 927

Chapter VIII. The Alamar blue phenotype

Chapter IX. General discussion

List of Tables

Chapter I: General Introduction

Chapter II: Development of proteomic techniques for T. brucei

Chapter III: Assessment of algorithms for protein identification from peptide mass fingerprinting data

Chapter IV: A proteome map of T. brucei

Chapter V: Characterisation of isogenic drug sensitive and drug resistant lines of trypanosomes

Chapter VI: A proteomic approach to study Cymelarsan resistance in T. brucei line of 386

Chapter VII: Association of the CAR protein and Cymelarsan resistance in line 247 and 927

Chapter VIII: The Alamar Blue phenotype

Chapter IX: General Discussion

List of Abbreviations

2DE	Two dimensional gel electrophoresis
APS	Ammonium persulfate
BAC	Bacterial artificial chromosome vector
BLAST	Basic Local Alignment Search Tool
Bp	basepair
CA	Carrier Ampholyte
CBSS	Carter's Balanced Salt Solution
cDNA	complementary DNA
CHAPS	3-[3-(cholamidopropyl)-dimethyl-lammonio]-1-propane sulphonate
CHCA	α cyano-4-hydroxycinamic acid
Ci	Curie
CNS	Central Nervous System
DAPI	4', 6-Diamidino-2-phenylindole
DB75	2,5 bis (4-amidinophenyl) furan
dd H_2O	Double distilled water
DEAE	Diethylaminoethyl cellulose
DEPC	Diethyl pyrocarbonate
DFMO	Diflouromethylornithine
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EATRO	East African Trypanosomiasis Research Organisation
EC_{50}	50% inhibitory concentration
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol-bis (β-aminoethyl ether) N,N,N',N'-tertaacetic acid
ELISA	Enzyme linked immuno serum assay
ES	electro spray
EST	Expressed sequence tag
g	gram
h	hour
HAT	Human African Trypanosomiasis

HCl	Hydrochloride acid
hsDNA	herring sperm DNA
IEF	Isoelectric focussing
IPG	Immobilised pH gradient
kb	kilobase
LB	Luria Bertani medium
MALDI	Matrix assisted laser desorption ionisation
Mb	Mega base
mg	milligram
MIC	minimum inhibitory concentration
min	minute
ml	millilitre
mM	millimolar
mmol	millimole
MOWSE	MOlecular Weight SEarch
MS	Mass spectrometry
NCBI	National Centre for Biotechnology Information
nm	nanometer
ORF	Open Reading Frame
PAGE	Poly-Acrylamide Gel Electrophoresis
PARP	Procyclic acidic repetitive protein
PBS	Phosphate buffer saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
pI	Isoelectric point
PIR	Protein Information Resource
PMF	Peptide Mass Fingerprinting
pmol	picomole
PMSF	Phenylmethylsulfonyl fluoride
PMT	Photomultiplier tube
PSG	Phosphate Saline Glucose
RFLP	Restriction enzyme Fragment Length Polymorphism
RNA	Ribonucleic acid
RT-PCR	Reverse Transcription polymerise chain reaction

SB	N-decyl-N,N-dimethyl-3-ammonio-1-propane sulfonate
SDM 79	Semi defined medium 79
SDS	Sodium Dodecyl Sulphate
SEM	standard error of the mean
STIB	Swiss Tropical Institute, Brunel
Taq	Thermus aquaticus
TBS	Tris buffered saline
TEMED	N,N,N',N'-tetramethyl ethylenediaminide
TFA	Trifluoroacetic acid
TIGR	The institute for genomic research
TIM	Triosephosphate isomerase
TLCK	N-α-p-Tosyl-L-lysine chloromethyl ketone
TOF	Time of flight
TPCK	N-Tosyl-L-phenylalanine chloromethyl ketone
TREU	Trypanosome Research Edinburgh University
UTR	Un-Translated Region
UV	Ultraviolet
V	Volt
VSG	Variant Surface Glycoprotein
WHO	World Health Organisation
X-gal	5-Bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside
µA	micro amperes
µl	microlitre
µM	micromolar

Chapter I
General introduction

1.1 Trypanosomes

1.1.1 Introduction

1.1.1.1 Current status of Human African Trypanosomiasis (HAT)

The latest data published by the World Health Organisation (WHO, 2003) indicate that 300,000 people suffer currently from HAT, and a further 50 million people in sub-Saharan Africa are at risk of infection. In 1996, 150,000 people died and 100,000 were handicapped by the disease and in the same year 30,000 new cases were reported. In 1998, 45,000 new cases were reported (Barrett, 1999). Sixty million people, living in 36 sub-Saharan countries (22 of which are among the poorest countries in the world) are under the daily threat of contracting HAT (Figure 1.1). Only four million of them have access to diagnosis and treatment.

There are at least 200 foci of HAT in sub-Saharan Africa and a resurgence of trypanosomiasis is occurring in the Democratic Republic of Congo (Van Nieuwenhove *et al.*, 2001), the Central African Republic (Smith *et al.*, 1998), Uganda (Olila *et al.*, 2002), Sudan (Moore and Richer, 2001) and Angola (Stanghellini and Josenando, 2001). Prevalence rates reach up to 50% in some villages, making HAT the leading cause of mortality. In two provinces of the Democratic Republic of Congo the number of deaths due to HAT have been estimated to be equal to or higher than the number of deaths due to Acquired Immune Deficiency Syndrome (AIDS) (WHO, 2003).

In addition, resistance to Melarsoprol, the drug of first choice for the treatment of HAT, has been reported in many epidemic areas of Africa. The overall relapse rate over the past 50 years was of 5 - 8% (Pepin and Milford,1994), but this rate has risen to 25% in Angola (Stanghellini and Josenando, 2001) and 30% in Uganda (Legros *et al.*, 1999) in recent years.

The increase in prevalence and distribution of HAT associated with the increase in drug resistance reported cases have initiated new research interests in the discovery of the mechanisms of drug resistance to the existing drugs and the development of new treatment for HAT.

1.1.1.2 Taxonomy/Classification

African trypanosomes are flagellated parasitic protozoa belonging to the order kinetoplastidae and the family trypanosomatidae. Kinetoplast flagellates are characterised by the possession of a unique organelle called the kinetoplast which contains deoxyribonucleic acid (DNA) and is an integral part of the mitochondrial system.

The genus *Trypanosoma* is classically subdivided into two groups of infective parasites, the stercorarian trypanosomes and the salivarian trypanosomes. In the stercorarian trypanosomes, the flagellates divide in the hind gut of the insect vector and metacyclic forms are passed out with the faeces. The most important species is *T. cruzi* which infects 12 to 24 million people in South and Central America. The disease it causes is called Chagas' disease and the most serious clinical consequences of infection are cardiac failure due to the presence of the parasite in the heart muscles or the loss of nervous control of the alimentary tract due to the presence of the parasites in the nervous system. In contrast to

the stercorarian trypanosomes, the salvarian trypanosomes typically develop in the mid gut of the vector and either in the salivary glands (*T. brucei*) or the mouthparts (*T. vivax* and *T. congolense*). One of the major species of salivarian trypanosomes is *T. brucei* which is divided into three subspecies *T. b. brucei*, *T. b. rhodesiense* and *T. b. gambiense*. These three subspecies are morphologically identical, but can be distinguished by isoenzyme analysis and restriction enzyme fragment length polymorphism (RFLP) (Gibson, 2002; Kanmogne *et al.*, 1996; Truc and Tibayrenc, 1993). *T. b. brucei* primarily infects cattle, goats, and sheep; *T. b. gambiense* is distributed in West and Central Africa and causes Western African HAT; and *T. b. rhodesiense* in Eastern and Southern Africa is the cause of Eastern African HAT (Figure 1.1) (Wang, 1995). Only two taxonomic groups of tsetse flies commonly transmit the flagellate protozoan parasite that causes HAT. The *palpalis* group is the chief carrier of *T. b. gambiense* and the *morsitans* group is the chief carrier of *T. b. rhodesiense*. There are several species within the palpalis group and within the morsitans group. Other species of salivarian trypanosomes *T. congolense* and *T. vivax* are the causative agent of Nagana, a trypanosomiasis in domesticated animals. Several species of *Glossina* transmit *T. congolense* and *T. vivax*.

Molecular phylogenetic analysis of the Trypanosoma, undertaken by sequencing the 18S RNA genes of 31 different species and subspecies, showed the existence of three distinct clades (Stevens *et al.*, 1999). The *T. brucei* clade consists of the salivarian tsetse transmitted trypanosomes from Africa. *T. evansi* and *T. equiperdum*, although non-tsetse transmitted and not restricted to Africa, also belong here by virtue of their close morphological and genetic similarity with *T. brucei*. Importantly, this clade is characterised by the expression of a dense variable surface glycoprotein coat (VSG coat) in the bloodstream forms of the parasite (section 1.1.2). The variability lies in the fact that

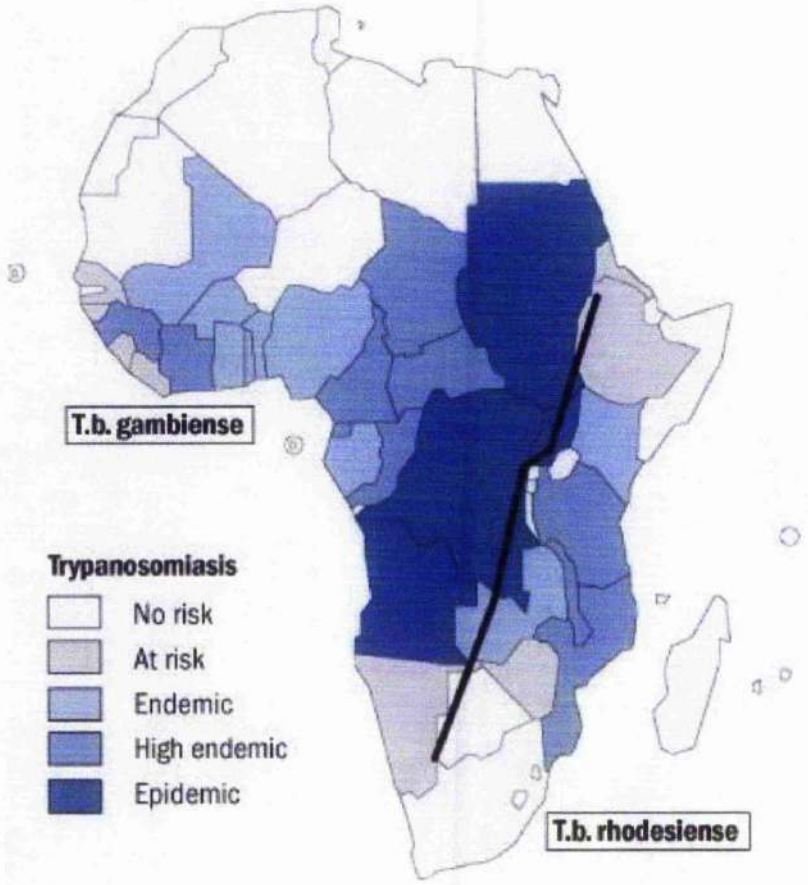


Figure 1.1: Geographic distribution of Human African Trypanosomiasis. The black line represents the approximate division for the distribution *of T. b. gambiense* and *T. b. rhodesiense*. Reproduced from medilinks.org/Features/Articles/june2002/tryps.htm

the parasite can change its VSG coat by switching between expression of multiple VSG genes, a process called antigenic variation (Barry and McCulloch, 2001). This process enables the parasite to evade the host immune response. The *T. cruzi* clade originates mainly from South American mammals, with some interesting exceptions: two species of European bat trypanosomes and one as yet unnamed species of kangaroo trypanosome from Australia. Finally, an aquatic clade comprising trypanosome species isolated from both marine and fresh water fish and amphibian has been observed (Stevens *et al*, 1999). Interestingly, the molecular phylogenetic analysis agreed with the classical classification of the Trypanosoma species.

1.1.2 Life cycle

Parasites of the *T. brucei* group are transmitted from the tsetse fly to their mammalian host where they replicate first in the bloodstream, lymphatic system and interstitial spaces, and then, during the latter stages of the infection in the central nervous system. In the mammal the bloodstream trypanosomes multiply as "long slender" forms by binary fission and eventually differentiate into a non-dividing "short stumpy" forms. The bloodstream trypanosomes have the ability to evade the host immune system by antigenic variation. A diagram of the life cycle of *T. brucei* is shown in Figure 1.2, which includes some of the intermediate forms that have been previously described (Vickerman, 1985). Trypanosomes are ingested by the tsetse fly (*Glossina* spp.) when a blood meal is taken (Vickerman, 1965). Once the short, stumpy trypomastigote reaches the mid-gut of the tsetse fly, it transforms into a long, slender procyclic form. The parasites multiply in the lumen of the midgut by binary fission. After approximately two weeks, the parasites migrate to the salivary glands through the hypopharynx, where the parasites attach to the epithelial cells of the salivary glands and then transform to the epimastigote form (Evans and Ellis, 1983).

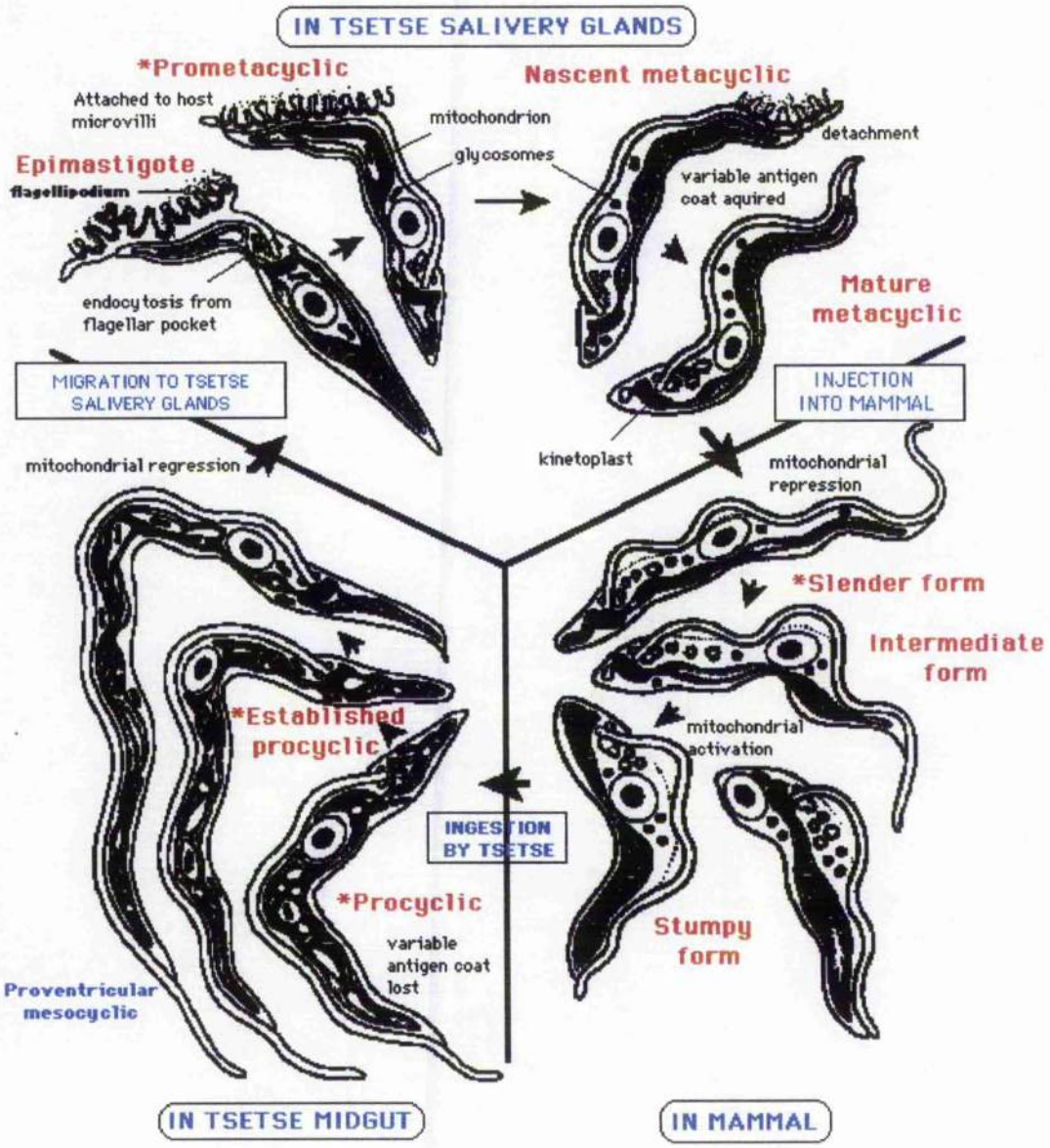


Figure 1.2: Life cycle of *Trypanosoma brucei*. Schematic representation of the developmental cycle in a mammal and in a tsetse fly vector.
Reproduced from http://www.tulane.edu/~dmsander/WWW/224/Parasitology.htm

This stage undergoes multiplication within the salivary gland, and after two to five days metacyclic (infective) forms develop from the epimastigotes (Tetley and Vickerman, 1985). With development of the metacyclic form, the parasite is now infective for the mammalian host and can be introduced into the puncture wound by the tsetse fly when it takes the next blood meal. The entire developmental cycle in the fly takes about three weeks. Once infected, the tsetse fly remains infected for life.

1.2 Control of Trypanosomiasis

1.2.1 Introduction

A concerted effort has been made to control African trypanosomiasis in both domestic livestock and humans, and there have been four main approaches taken. Firstly, avoid tsetse-infected areas, which is difficult to apply as the high risk areas are often the best grazing land. Secondly, control of the tsetse fly vector by using insecticides, fly traps and odour baits has been put to practice with varying degree of success (Schofield and Maudlin, 2001). Thirdly, the problem of animal trypanosomiasis could be overcome by increasing the use of trypanotolerant cattle that are able to resist trypanosome infection (Murray and Morrisson, 1979). Finally, the last and most widely used method for controlling both the animal and the human diseases still lies with the use of trypanocidal drugs.

1.2.2 Control of the vector

1.2.2.1 Insecticides

Following the discovery of DTT and other persistent insecticides in 1945, ground spraying became the main line of defence against tsetse flies throughout Africa for over 40 years. Selective sparying of insecticides was successfully applied to fly resting sites in riverine

fringing vegetation in Sudan in the 1950's, which led to the elimination of tsetse from 200,000 km^2 by 1978 (Snow *et al.*, 1991).

1.2.2.2 Bait technology

Bait technology involves an odour bait which attract flies to a visual target treated with insecticide. Attractants such as acetone and octenol are so effective that good control with low density of treated target (four per km^2) can be achievable (Vale *et al.*, 1988). This technique was developed further by using "live bait" such as cattle treated with insecticides. This strategy was applied with some success in Ethiopia (Leak *et al.*, 1995), Zambia (Chizyuka and Linguru, 1986) and Tanzania (Fox *et al.*, 1993).

1.2.2.3 Sterile insect technique (SIT)

Sterile insect technique involves sustained and systematic releases of sterile male insects among a wild population (Aksoy *et al.*, 2001). The insects are mass reared in large-scale insectaries where the males are sterilized by irradiation and then taken to the selected area and released by air. The sterile males fertilise wild females, which are then unable to produce progeny. Sufficient sterile males need to be released in order to achieve an over-flooding ratio sufficient to cause a decline in population size. This technique was successfully applied for the eradication of *G. austeni* from the island of Zanzibar (Shahid and Curtis, 1987).

A range of applicable techniques is available for the control of tsetse flies from simple technology such as traps, targets, treated cattle and ground spraying to more advanced technology such as aerial spraying and the release of sterile tsetse males. However, the control of tsetse flies only decrease the incidence of human and animal trypanosomiasis, and chemotherapy is still the method of choice for controlling the diseases.

1.2.3 Control of animal trypanosomiasis

Animal trypanosomiasis is an economic burden for Africa, with infection in domestic cattle causing a 20% decrease in calving, a 25% decrease in milk production and it is estimated that it causes 3 million livestock deaths per year (Seed, 2000). In addition, domestic cattle can also act as a reservoir host for *T. b. rhodesiense* (Hide, 1999). A recent HAT outbreak of *T. b. rhodesiense* in Uganda was directly associated with the import of cattle from a HAT endemic area (Fevre *et al.*, 2001). Consequently, the control of animal trypanosomiasis is important both economically and to control HAT.

1.2.3.1 Chemoprophylaxis

Animal trypanosomiasis can be prevented by the use of prophylactic drugs. The prophylactic drugs currently in use for cattle are homidium and Isometamidium. These drugs can provide protection for three months. Prophylactic drugs are particularly useful to protect animals at time of constant disease challenge. Chemoprophylaxis is an expensive, time consuming and inefficient (due to the problem of drug resistance to these drugs) solution to the problem of animal trypanosomiasis.

1.2.3.2 Chemotherapy

Treatment of animal trypanosomiasis relies on three drugs: homidium, Berenil and isometamidium chloride. Isometamedium chloride, a conjugate of ethidium and part of the Berenil molecule, is used exclusively in veterinary typanosomiasis. All three drugs are used both prophylactically and therapeutically against *T. congolense*, *T. brucei* and *T. vivax*. For all major species of trypanosomes (*T. congolense, T. vivax* and *T. evansi*) resistance to Isometamidium has been reported (Anene *et al.*, 2001). Multiple resistant population of *T. congolense* have also been described (Afewerk *et al.*, 2000). Cymelarsan, used for the treatment of trypanosomiasis in camels, has also been shown to be very effective against *T. b. brucei, T. evansi*, and *T. equiperdum* in buffalo, goats and

pigs *in vitro* (Berger and Fairlamb, 1994). Cymelarsan dissociates rapidly into Melarsen oxide, thought to be the active form of the drug (Berger and Fairlamb, 1994).

1.2.3.3 Trypanotolerant cattle

Breeds of cattle which are tolerant to trypanosomiasis occur in some part of Africa (Murray *et al.*, 1984). Although the use of trypanotolerant cattle in Africa is a way of controlling trypanosomiasis in cattle, it is not *per se* a solution to the problem, since trypanotolerant animals still suffer loss of productivity due to chronic trypanosome infection and can succumb to the disease under high challenge. In addition, these animals would still be reservoir host for the human infective trypanosomes and so will not reduce the incidence in HAT (Fevre *et al.*, 2001).

1.2.4 Control of Human African Trypanosomiasis

The process of antigenic variation has hindered efforts to develop vaccine against the parasite and so chemotherapy has remained the major means of controlling HAT (Pays, 1995). There are currently four drugs available for treating HAT: Suramin, Pentamidine, Melarsoprol, and DL-α-difluoromethylornothine (DFMO). The use of a particular drug depends on the stage of HAT being treated and relates to the ability of the drug to penetrate the blood-brain barrier. Suramin and Pentamidine treatment are restricted to the early stage of HAT while Melarsoprol and DFMO are lipophilic and thus able to cross the blood brain barrier. This characteristic makes them the drugs of choice for the treatment of the late stage of HAT when the parasite grow in the brain.

With the exception of DFMO, which was introduced for the treatment of West African HAT in 1990, all of the antitrypanosomal drugs were discovered more than 40 years ago.

1.3 Administration, uptake and mode of action of trypanocidal drugs

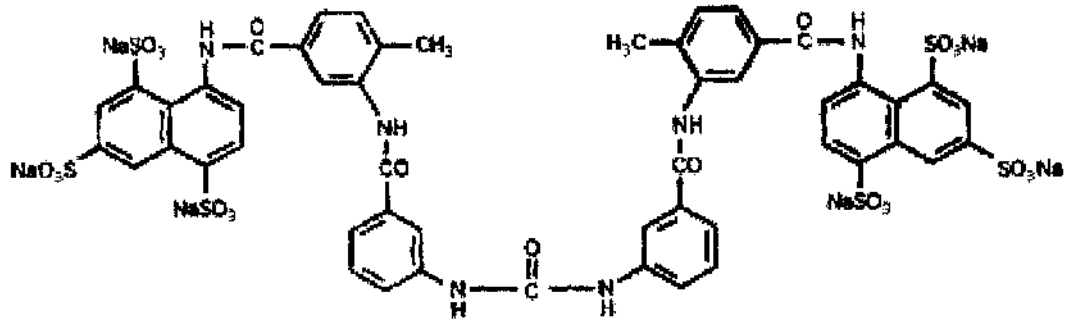
The main drugs for the treatment of the first stage of HAT are suramin and Pentamidine, which were introduced in 1922 and 1941 respectively. After the parasite has invaded the central nervous system (second stage of HAT), only two drugs are effective, namely Melarsoprol introduced in 1949 and DFMO introduced in 1990.

1.3.1 Suramin

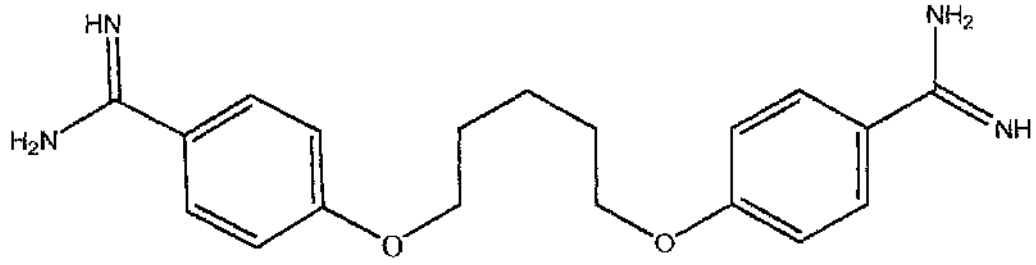
Suramin was first introduced for the treatment of early stages of HAT in 1922, and is especially effective for treating *T. b. rhodensiense* infections (Wang, 1995). It is also effective against *T. b. gambiense, T. b. brucei, T. evansi* and *T. equiperdum* (Hawking, 1963). Suramin is usually administered to patients in five intravenous injections at a dosage of 20 mg/kg body weight once every five to seven days.

The structure of Suramin is shown in Figure 1.3. It has been found to bind to albumin and low-density lipoprotein (LDL) in serum. Fairlamb & Bowman (1980) demonstrated the rapid uptake of Suramin in the presence of serum proteins in trypanosomes. The uptake was 18 fold higher than could be explained by fluid endocytosis alone, therefore suggesting a high affinity transport system. Evidence for receptor-mediated endocytosis of host LDL in both bloodstream and procyclic forms was discovered by Coppens *et al.* (1988) who proposed that Suramin was entering trypanosomes bound to LDL through LDL- receptor mediated endocytosis (Coppens *et al.*, 1988). In addition Suramin was found to inhibit LDL-binding and uptake by *T. brucei* in a concentration dependent manner, suggesting competition for the LDL-binding site on the LDL-receptor

A)



B)



C)

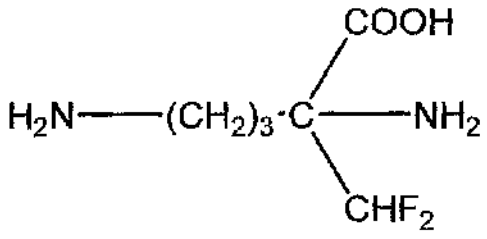


Figure 1.3: **Structure of Suramin, Pentamidine and DFMO**. Suramin (A) and Pentamidine (B) are used for treatment against early stage trypanosomiasis. DFMO (C) is in use for the treatment of late stage trypanosomiasis.

(Vansterkenburg *et al.*, 1993). However, Pal *et al.* (2002) recently demonstrated that while Suramin was transported by receptor-mediated endocytosis in *T. brucei*, the LDL-mediated endocytosis was only indirectly involved in Suramin uptake. The specific route by which Suramin enters the trypanosomes has not yet been identified.

Suramin is an inhibitor of many different dehydrogenases and kinases from mammalian, bacterial and fungal sources. In *T. brucei*, Suramin was found to inhibit dihydrofolate reductase and thymidine kinase as well as being a potent inhibitor of many of the glycolytic enzymes in *T. brucei* (hexokinase (HK), phosphoglucoisomerase (PGI), phosphofructokinase (PFK), triosephosphate isomerase (TIM), aldolase (ALDO), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), phosphoglycerate kinase (PGK), glycero-3-phosphate dehydrogenase (GPDH), and glycerol kinase (GK)) (Misset and Opperdoes, 1987). It was thus proposed that Suramin acted by blocking glycolysis in trypanosomes. This theory on the mechanism of antitrypanosomal action of Suramin is contradicted by the fact that all these glycolytic enzymes are confined within single membrane bound microbody organelles, termed the glycosome. As it is unlikely that the glycosome will take up Suramin by passive diffusion or endocytosis due to its high negative charge, the nine glycolytic enzymes would be protected from Suramin via compartmentalisation. However, the genes encoding these enzymes are located in the nucleus and the enzymes are synthesised on free ribosomes in the cytoplasm with subsequent import into the glycosome within three to five minutes without any proteolytic modification. One possible mode of action of Suramin could be that the inhibition of glycosomal proteins import leading to a gradual decrease of enzyme concentration in the glycosome and a slowing down of energy metabolism (Wang, 1995).

1.3.2 Pentamidine

The aromatic diamidine Pentamidine (1,5-bis (4'-aminidinophenory) pentane; Figure 1.3) introduced in 1937 has been used for decades in the treatment of HAT, antimony-resistant leishmaniasis (Sereno *et al.*, 2000). It is effective against the early stage of *T. b. gambiense* infection and is given at a dose of 4 mg/kg body weight by 7 to 10 intramuscular injections daily or on alternative days. Pentamidine is the only drug that has been use for large scale prophylaxis (Dukes, 1984).

African trypanosomes accumulate Pentamidine to millimolar concentrations via a high affinity transport system which is energy dependent and competitively inhibited by other diamidines (Damper and Patton, 1976). The Pentamidine transporter was identified in 1995 as the P2 adenosine/adenine transporter (Carter and Fairlamb, 1993). Two more Pentamidine transporters were recently described, a high-affinity pentamidine transporter (HAPT1) and a low affinity Pentamidine transporter (LAPT1) (de Koning, 2001).

The mechanism of Pentamidine antitrypanosomal action is unknown. Pentamidine has been shown to inhibit thymidylate synthetase, an enzyme present in various trypanosomatids, (Gutteridge, 1969) as well as S- adenosyl-L-methionine decarboxylase (Bittoni *et al.*, 1986), mitochondrial topoisomerase II (Shapiro, 1993), mitochondrial membrane potential (Vercesi and Docampo, 1992), dihydrofolate reductase (Waalkes and Maluku, 1976), thymidylate synthetase (Kaplan and Myers, 1977) and calcium transport (Benaim *et al.*, 1993). Given this range of inhibitory activities, it is difficult to define a specific selective mode of action.

1.3.3 DFMO

DFMO (Figure 1.3) was primarily developed as an anti-tumour proliferation and modulation of cell differentiation drug but the clinical trials showed poor results. DFMO was first synthesised as a potential suicide inhibitor of ornithine decarboxylase (ODC). ODC is the key enzyme in the pathway leading to biosynthesis of polyamines: putrescine, spermidine, and spermine, which are essential for the proliferation of prokaryotic as well as eukaryotic cells. DFMO depletes the polyamines putrescine and spermidine from *T. brucei* both *in vitro* and *in vivo*. Subsequently, the DFMO-treated trypanosomes are transformed to a non-dividing, short, stumpy form that is apparently incapable of changing its VSG and is eventually killed by the host immune reaction (Bitonti *et al.*, 1986b). DFMO is relatively ineffective against *T. b. rhodensiense* infections. Its inefficiency in treating *T. b. rhodesiense* infections is explained by the short half-life of ODC and high expression levels of ODC in *T. b. rhodesiense* as compared to *T. b. gambiense* (Itens *et al.*, 1995; Itens *et al.*, 1997).

The uptake of DFMO in bloodstream forms of *T. b. gambiense* and *T. b. rhodesiense* is thought to be a combination of passive and mediated diffusion, depending on the extracellular concentration (Bitonti *et al.*, 1986; de Koning, 2001). DFMO can be accumulated in trypanosomes to millimolar levels (de Koning, 2001). DFMO is a remarkably safe drug but it requires to be administered intravenously at a large dose of 400 mg/kg body weight per day in four equal doses every 6 h for 14 days (Van Nieuwenhove *et al.*, 1985).

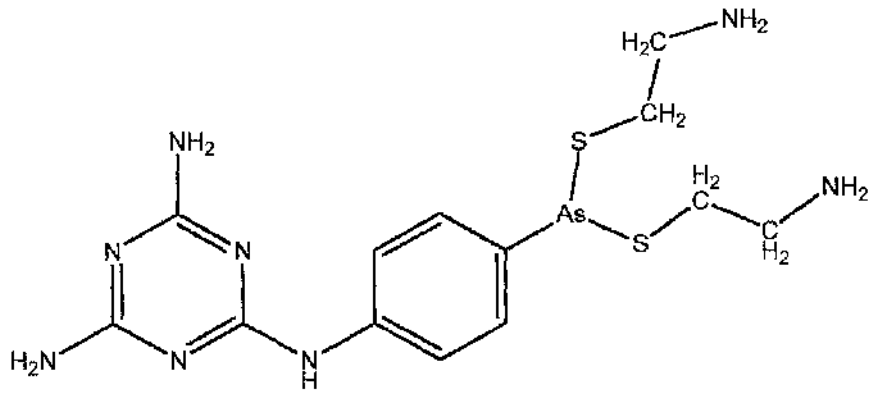
1.3.4 Arsenicals

The class of melaminophenyl arsenical drugs use for the treatment of trypanosomiasis includes Melarsoprol, Cymelarsan, Trimelarsan and Melarsen oxide (Figure 1.4). Melarsoprol is the only arsenical drug currently used for the treatment of HAT and was made available in 1949. Cymelarsan is licensed only as a veterinary drug against trypanosomiasis in camels and the other two drugs have not been licensed. Until 1990, Melarsoprol was the only drug capable of treating late stage *T. b. gambiense* and *T. b. rhodesiense* infections. It is administered intravenously at a dose of 3.6 mg/kg body weight in three to four series of four injections separated by at least one week (WHO). A major problem with Melarsoprol is its toxic side effects. The encephalopathic syndrome, which is reported to occur in 2-10% of treated cases is fatal in 50-70% of these cases, constitutes the most important complication of Melarsoprol treatment (Keiser and Burri, 2000).

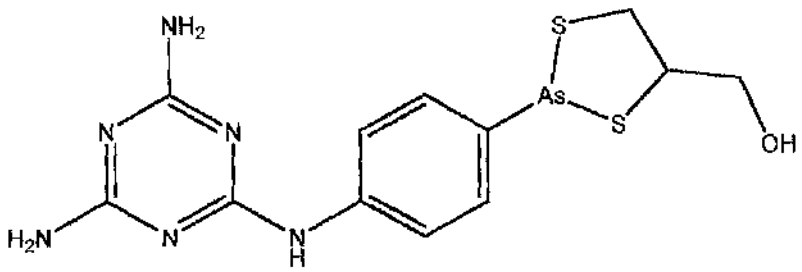
In bloodstream forms of *T. brucei*, Melarsoprol is rapidly metabolized into Melarsen oxide (Keiser and Burri, 2000; Keiser *et al.*, 2000). Melarsen oxide was shown to be the active metabolite against HAT. Water soluble melaminophenyl arsenicals such as Melarsen oxide are transported into trypanosomes by the P2 adenosine/adenine transporter (Carter and Fairlamb, 1993; de Koning and Jarvis, 1999; de Koning, 2001). It is unknown whether the lipophilic Melarsoprol is also transported by the same mechanism (de Koning, 2001). In procyclic forms however, Melarsoprol crosses the membrane by passive diffusion (Scott *et al.*, 1997). A summary of transporters involved in the uptake of trypanocydal drugs discovered so far in bloodstream trypanosomes is given in Figure 1.5.

Melarsoprol leads to the rapid lysis of trypanosomes, but its mechanism of antitrypanosomal action is not well understood. It has been proposed that trivalent

A)



B)



C)

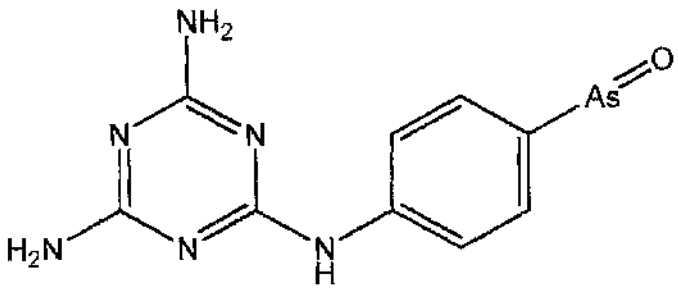


Figure 1.4: Structure of the melaminophenyl arsenical drugs. Cymelarsan (A) and Melarsoprol (B) dissociate primarily into Melarsen oxide (C).

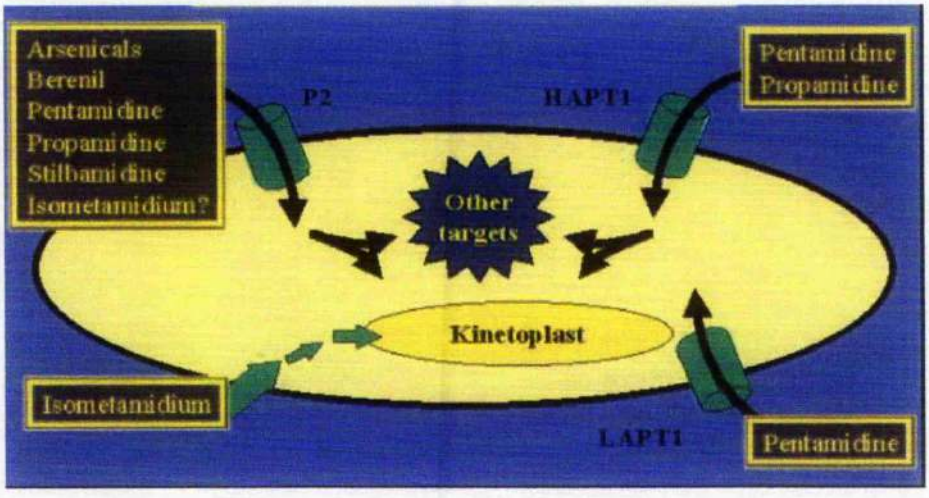


Figure 1.5: **Transport of trypanocides in bloodstream forms of *T. brucei*.** Arsenical drugs (Melarsoprol, Cymelarsan, etc.) and diamidines share a common transporter: the P2 transporter. Reproduced from de Koning (2001).

arsenicals with two replaceable ligands have a high affinity for intracellular thiols such as trypanothione and dihydropoamide plus several enzymes including trypanothione reductase (Fairlamb *et al.*, 1989), 6-phosphogluconate dehydrogenase (Hanau *et al.*, 1996) and glycerol-3 phosphate dehydrogenase (Denise *et al.*, 1999). However, these inhibitory effects have not been directly linked to cell lysis. *In vivo*, Melarsoprol is quickly metabolised into Melarsen oxide, which was shown to be an active metabolite (Keiser *et al.*, 2000). Melarsen oxide was first believed to act by blocking glycolysis in trypanosomes through inhibition of the pyruvate kinase. However, later studies indicated that the lytic effect of Melarsen oxide is not the primary result of inhibiting pyruvate kinase (Cunningham *et al.*, 1994). Thus, while a number of enzyme activities have been shown to be inhibited by the arsenicals, the nature of the primary target of these inhibitors is still unclear. Whether the basis for their selective action is due to their rapid and specific uptake by the P2 transporter followed by inhibition of a range of enzyme activities or is via some other mechanisms still need to be investigated. The rate of entry of the drug into the CSF is related to the lipid solubility of the unionised molecule. Melarsen oxide and Melarsoprol have high lipid solubility. At physiological pH of 7.4, only 0.2% of Melarsoprol is unionised, indicating that only a small amount of the drug is available to cross the lipid membrane.

1.4 Arsenical resistance in T. brucei

1.4.1 Definition of drug resistance

Treatment failures with Melarsoprol were reported as early as 1986 (WHO). However, the rates of treatment failure remained at low and constant levels of 1 to 10% (WHO, 1986; Pepin and Milord, 1994). This situation has changed in recent years with reports of much higher rates of treatment failure from northern Uganda and northern Angola. Up to 25% of

treated patients in Angola and 30% in Uganda were reportedly non-responsive to Melarsoprol therapy (Legros *et al.*, 1999, Matovu *et al.*, 2001). Treatment failure can be defined as the "absence or insufficient effect of drug action after administration of a normally effective dose". It is usually observed as the presence of trypanosomes in a sample from patients within one month of treatment completion or a complete lack of clinical response during the course of treatment. Drug resistance, however, was defined in 1963 by the WHO as the "ability of a parasite strain to multiply or survive in the presence of concentrations of a drug that normally destroy parasites of the same species or prevent their multiplication. Such resistance may be relative (yielding to increased doses of the drug tolerated by the host) or complete (withstanding maximum doses tolerated by the host)".

The difference between the two is fundamental as although it is true that drug-resistance will bring about drug failure, it is not necessarily true that drug failure is the result of drug resistance since resistance is only one of many factors affecting drug failure (Borst and Ouellette, 1995).

1.4.2 Biochemical and molecular basis of arsenical resistance in *T. brucei*

Cross-resistance between the diamidines such as Pentamidine and the melaminophenyl arsenicals has been reported in laboratory-induced drug resistant strains of *T. brucei* (Fairlamb *et al.*, 1992). Cross-resistance between two different classes of drugs generally implies an alteration in either a common uptake/efflux mechanism or a common target (or both).

Carter and Fairlamb were the first to demonstrate the transport of Melarsen oxide via an adenosine transporter, later referred to as P2 transporter, into bloodstream forms of trypanosomes (Carter and Fairlamb, 1993). This provided the first basis for the search for the molecular mechanism of arsenical resistance in *T. brucei*. Since this first report, studies of arsenical resistance in *T. brucei* have mainly centred on this transporter. The gene encoding the transporter, *TbAT1,* was cloned in 1999 and a structure for the protein was proposed from the sequence (Mäser *et al.*, 1999). Sequencing of *TbAT1* from a laboratory-derived arsenical resistant line of *T. brucei* demonstrated the presence of ten point mutations in the gene, six of which conferred an amino acid substitutions in the protein sequence (Mäser *et al.,* 1999). In addition, cloning of the *TbAT1* gene in yeast also supported the involvement of the P2 transporter in resistance to Melarsoprol. The strain of *Saccharomyces cerevisiae* used had no transporter for exogenous adenosine and was unable to use it as a purine source. These cells defective in purine biogenesis were transformed with *TbAT1* gene and grown on agar containing adenosine as sole purine source. The transformed yeasts were able to grow on adenosine as sole source of purine. In addition, expression of the *TbAT1* gene in yeast rendered the yeast sensitive to Melarsoprol. However, the expression of the mutated *TbAT1* gene in yeast did not confer sensitivity to Melarsoprol, demonstrating that the mutated gene was not able to mediate the transport of the drug in the cell (Mäser *et al.*, 1999).

Analysis of the P2 transporter by competitive uptake with adenosine revealed that the P2 transporter also mediates the transport of Berenil, Pentamidine, Propamidine and Stilbamine in addition to the arsenicals (Figure 1.5) (de Koning, 2001), explaining the cross-resistance observed between the diamidines and melaminophenyl arsenical drugs (Fairlamb *et al.*, 1992; Frommel and Balber, 1987). The P2 transporter was shown to be

selective for adenine and adenosine with little affinity for other purines, but it also mediated the transport of diverse structures such as Melarsoprol and Pentamidine (Carter *et al.*, 1995; de Koning and Jarvis, 1999). The paradox of the selectivity of the P2 transporter was resolved with the determination of its substrate recognition motif. The substrate recognition motif of the P2 transporter was determined by assessing the inhibition of adenosine transport by a series of adenosine analogs (Figure 1.6) (de Koning and Jarvis, 1999; de Koning, 2001). However, preliminary analysis of the P2 transporter in Melarsoprol resistant *T. b. rhodesiense* isolated from relapsed patients revealed that only 50% of the isolates contained the mutated transporter (Matovu *et al.*, 2001). In addition, a knockout of *TbAT1* in *T. brucei* conferred only a three fold resistance to Melarsoprol, which is well below the level of resistance obtained in laboratory-induced lines (Matovu *et al.*, 2003). In addition, two laboratory-induced Cymelarsan resistant lines (247 and 386) exhibited the resistance phenotype in procyclic stages (Scott *et al.*, 1996) for which P2 transporter activity cannot be detected (Figure 1.7) (de Koning *et al.*, 2000). Consequently the resistance to Melarsoprol exhibited by these lines cannot be accounted for by mutations in the P2 transporter.

The observation that a considerable proportion of wild type *TbAT1* were found among relapsed patients, that only three fold resistance to Melarsoprol was observed in the *TbAT1* knockout line and that procyclic forms can exhibit arsenical resistance strongly indicate that *TbAT1* is not the only gene responsible for conferring arsenical resistance in *T. brucei*. A model whereby the loss of a single transporter could mediate cross-resistance to all diamidines and melaminophenyl arsenical drugs and high level arsenical resistance appears, in view of these findings, incomplete.

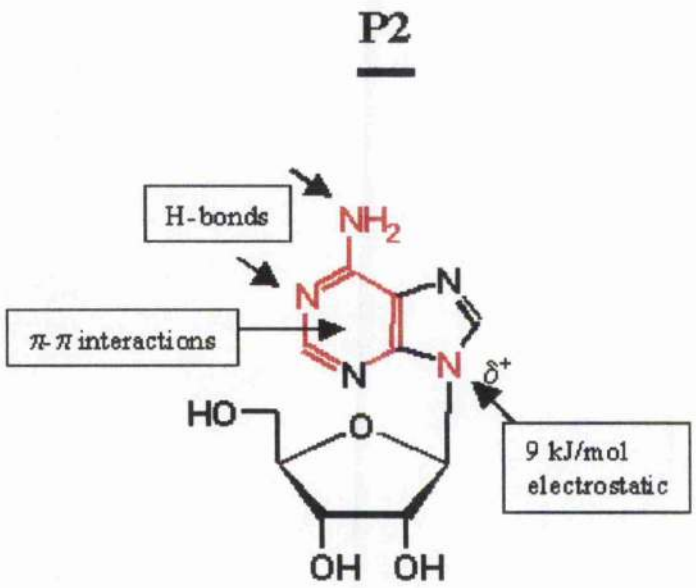


Figure 1.6: Recognition motif for the P2 transporter. Reproduced from de Koning, 2001.

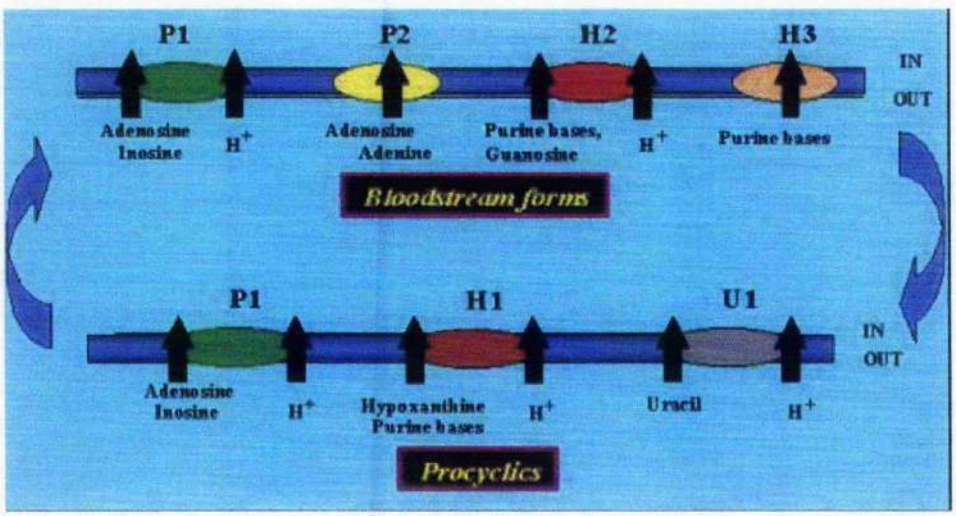


Figure 1.7: Nucleotide and nucleoside transporter in *T. brucei*. Reproduced from de Koning, 2001.

1.5 The study of mechanisms of drug resistance

From the preceding discussion it is clear that the field of drug resistance is far from fully understandood or defining the genes that determine drug resistance in trypanosomes and therefore explaining the mechanisms by which resistance occurs. In trypanosomes there are a range of techniques that could be applied to investigate the mechanisms of resistance and these are reviewed below.

1.5.1 Genetic tools

1.5.1.1 Classical genetic analysis

While *T. brucei* multiplies mainly by binary fission, a mating system exists and genetic exchange was clearly demonstrated by laboratory co-transmission studies (Sternberg and Tait, 1990). There is evidence for segregation of alleles and recombination between loci (Turner *et al.,* 1990). However, it appears that genetic exchange is non-obligatory and the frequency at which it occurs in natural population is still unknown (Sternberg *et al.*, 1989). The genetic basis of a phenotype can be determined by the analysis of the progeny from a cross of parental stocks with different phenotype (drug sensitive/ resistant stocks). To determine the region of the genome in which gene or genes associated with the particular phenotype is located, a genetic map must have been previously established. This genetic map, made using molecular markers, enables linkage analysis to be performed. The co-segregation of markers on the genetic map with the phenotype of the progeny of the cross determine the location, between two markers, of the gene or genes of interest. The efficiency of the linkage analysis depends on the resolution of the genetic map. In general linkage can be established within several hundred kilobases, which might contain several genes. Analysis using RNAi or gene knockout is then required to identify the specific gene.

Genetic maps for chromosome I and II of *T. brucei* are available (Hall *et al.*, 2003; El Sayed *et al*, 2003). A recent study has demonstrated the existence of crossing over and the size of the recombination unit was calculated which suggested that at least one crossover will occur per chromosome at meiosis (Tait *et al.*, 2002). Crosses between naturally occurring variation in drug resistance between parental stocks of 386/247, 386/927 and 247/927 were produced (Hope *et al.*, 1999). The linkage analysis of the drug resistance phenotype in the progeny from the cross between 247/927 determined the location of a single gene conferring resistance between two closely located markers on chromosome II (Dr S. Taylor, personal communication).

The main limiting factor in the study of inheritance patterns of traits is the available phenotypic variation between parasite stocks. In addition, there are difficulties in obtaining and cloning large number of hybrid progeny, which determine statistical significance of segregation phenotypes.

1.5.1.2 Reverse genetic analysis

Two powerful reverse genetic techniques are available for the analysis of gene function in *T. brucei*. Firstly the use of gene knockouts based on homologuous recombination and secondly RNA interference (RNAi) which allows the knockdown of specific transcripts. These techniques are largely used when a specific gene is being studied and can be used to define whether the gene is essential or the loss leads to a specific phenotype. RNAi is currently being used on a large scale basis to knockdown every identified gene on chromosome I. These technique are particularly useful for confirming that an identified gene has a particular function.

RNAi is the process by which gene-specific double stranded RNA (dsRNA) elicits degradation of mRNA. This phenomenon occurs post-transcriptionally, which makes it the strategy of choice to down-regulate the expression of repeated genes. The first observation of degradation of mRNA by gene-specific dsRNA in *T. brucei* was made in 1998 (Ngô *et al.*, 1998). RNAi technology has become the method of choice to down-regulate gene expression in a variety of organisms. RNAi is particularly well suited for diploid organisms such as trypanosomes as transcripts from both alleles are down regulated. Placement of integrated RNAi constructs under control of the *tet* repressor system or the T7 promoter system provide a powerful method to probe gene function in trypanosomes. RNAi technology has been applied in *T. brucei* to study the function of gamma-glutamylcysteine synthetase (Huynh *et al.*, 2003), surface protease GP63 (LaCount *et al.*, 2003), gamma tubulin (McKean *et al.*, 2003) and farnesyl phosphate synthase (Montalvetti *et al.*, 2003). The RNAi technique is simple and efficient, but the induction of mRNA degradation by RNAi failed to provide a phenotype in 50% of the cases (Ullu *et al.*, 2002). In addition, RNAi technology requires the pre-selection of genes whose function are of interest. In the case of the mechanism of Cymelarsan resistance in trypanosomes the selection of genes is not obvious as the drug targets and the mode of action of the drug are still unknown.

1.5.2 Micro-array expression analysis

One approach to identifying the mechanisms of drug resistance is by comparative analysis of gene expression between drug sensitive and resistant isolates. One of the hypothesis assumed in employing this technique is that a change in gene expression causes resistance.

Micro-array expression analysis is based on the construction of ultra-high density DNA micro-arrays on glass microscope slides followed by hybridisation with fluorescently labelled cDNA and analysis using a con-focal laser scanner. The advantages of this technique are two-fold. Firstly, the micro-arrays allow the simultaneous interrogation of thousands of cDNA clones or oligonucleotides with mRNA from various resistant lines of parasites at once. Secondly, the use of fluorescently labelled probes and con-focal laser microscopy allows partial quantitative determination of the relative expression levels of many genes in a single experiment by comparing different resistant lines (Celis *et al.*, 2000). A large amount of data are generated by this approach and the analysis of the data can be complex. Reproducibility can be quite low and so for statistical analysis micro-arrays studies requires the experiment to be repeated several times. In addition, micro-arrays study the expression of mRNA level in an organism. One mRNA molecule can give rise to anything from one to a few dozen gene-products in higher eukaryotes depending upon the degree of post-translational modification, cleavage events and the degree of differential exon splicing possible within exon-rich genes. Micro-array technology has been applied to the study of stage specific gene expression in *T. brucei* (Diehl *et al.*, 2002) and *T. cruzi* (Minning *et al.*, 2003).

1.5.3 Proteomic techniques

1.5.3.1 Definition

The term "proteome" first appeared in literature in July 1995 and refers to the "total protein complement of a genome". Wasinger *et al.* (1995) qualified this further by stating that "although 'proteome' refers to the total complement able to be encoded by a given genome, it is unlikely that: 1- the totality of this potential for protein expression will be realised at any given instant, and 2- the detection threshold of 2 dimensional gel electrophoresis (2DE) for low copy number molecules will enable all translated proteins to

be visualised. Proteome analysis must depend upon experimental design and the need to examine several physiological states both *in vitro* and *in vivo*."

In general, the proteome represents the protein pattern of an organism, a cell, an organelle, or even a body fluid determined quantitatively at a certain moment and under precisely defined conditions. Furthermore, the proteome reflects the current metabolic status of the corresponding cell (or organism) which is determined by manifold interactions of the different molecules that are currently still difficult to analyse and by a range of environmental parameters. Unlike the genome, the proteome is thus a highly dynamic system which is characteristically altered by changes in environmental conditions. Analysis of the total protein complement of a genome has only recently become possible due to improvements in the reproducibility of separation technologies in association with immobilised pH gradients, the availability of computer software for the construction of polypeptide maps and the comparison of gel separations by advanced image analysis and the sensitivity and applicability of mass spectrometry to the analysis of proteins separated by two dimensional gel electrophoresis (2DE).

1.5.3.2 Applicability of proteomics study to trypanosomes

Analysis of the complete sequences from chromosome I and II of *T. brucei* provides some information about the structure and gene content of the chromosomes. Both chromosomes have a gene-rich core, a region containing a retro-transposon hotspot, VSGs adjacent to telomeres and may carry a bloodstream expression site with upstream 50 bp repeats (El-Sayed *et al.*, 2003, Hall *et al.*, 2003). Analysis of these two chromosomes revealed long arrays of genes located on the same coding strand, each array separated by a strand switch region. This arrangement of the genes along the chromosomes is consistent with polycistronic transcription of trypanosome genes (Tschudi and Ullu, 1988), whereby long

mRNA transcripts of multiple genes are processed to individual mRNAs by trans-splicing of a 5' spliced leader and addition of a poly-A tail (Matthews *et al.*, 1994). Analysis of the strand switch regions revealed a lack of common sequence motifs that could represent specific promoters capable of controlling gene transcription during the development of *T. brucei*. Only two promoters have been identified so far in the *T. brucei* genome. One controls the transcription of the VSG genes (Zomerdijk *et al.*, 1990) the other one controls the transcription of the procyclin gene (Sherman *et al.*, 1991). With such a gene organisation, any control at the level of transcription initiation would imply that all genes belonging to the same unit are regulated in the same way. This does not appear to be the case. Instead, control of gene expression in *T. brucei* is most likely accomplished at the post-transcriptional level, by both regulating the levels of mRNA and the levels and activities of the expressed proteins (post-translational modifications). This view is supported by the study of the regulation of cytochrome c protein in *T. brucei* (Torri *et al*, 1993). The bloodstream trypanosomes synthesise apocytochrome c protein at levels comparable to that of procyclic trypanosomes, but protein turnover prevents accumulation of cytochrome c protein in the bloodstream form. A similar observation applies to cytochrome c reductase.

The importance of protein-based analysis is its ability to study post-transcriptional control as well as post-translational modifications such as phosphorylation, glycosylation, acylation and methylation. Proteins determine the biological phenotype of an organism and proteins are the primary targets for most therapeutic agents. The advantages of proteomic studies are that they give a global picture of the changes taking place, whereas previous drug resistance studies have concentrated on individual enzymes or proteins and not on total response (Barett *et al.*, 2000). Techniques such as 2DE, quantitative computer image

analysis and protein identification can be use to create a "reference map" of all detectable proteins. Such reference maps establish patterns of normal gene expression in the organism and allow the examination of some post-translational protein modifications which are functionally important for many proteins. In addition, comparative maps can be established by analyzing variants of the wild type and so identify differences associated with a mutant or a different physiological condition. It is possible to screen proteins systematically from reference maps to establish their identities (Wilkins *et al.*, 1996).

Very few proteomic studies have been performed on trypanosomes. Three proteomic studies were performed in the 1980s on *T. brucei* using carrier ampholytes for isoelectric focussing in the first dimension. These were used to detect difference in VSG expression between cloned populations (Pearson *et al.*, 1981) or to detect differences in protein expression between subspecies (Anderson *et al.*, 1985; Pearson and Jenni, 1989). The use of carrier ampholytes meant that the reproducibility of the gels was difficult. In addition, no easy/fast methods were available to identify the differentially expressed proteins. The availability of *T. brucei* genome sequences and of new technologies, for isoelectric focussing and identification of protein separated by 2DE using mass spectrometry methods, have removed most of the limitations of the proteomic studies. Recently, a protocol was developed for the extraction and separation of proteins from *T. brucei* using IPG technology (van Deursen *et al.*, 2003). It is expected that the new technologies will spark an increase in research activity concerned with the analysis of the *T. brucei* proteome.

1.6. Aim

The aim of this project is to use proteomics to understand the mechanisms of arsenical drug resistance in *T. brucei* parasites. This is a particularly important area, given that there are only four drugs available for the treatment of HAT. Further, resistance to Melarsoprol treatment, the drug of first choice for late stage trypanosomiasis, has been reported in many parts of Africa. One mechanism by which drug resistance to Melarsoprol can be generated has been described and involves the alteration of the adenosine P2 transporter. However, there is preliminary evidence suggesting that other mechanisms are very likely to be involved.

Isogenic lines of parasites that are sensitive or resistant to Cymelarsan (a close analogue of Melarsoprol) have been described previously (Scott *et al.*, 1996, 1997). Comparison of the proteins expressed by each of these lines can be made by 2DE in order to identify specific polypeptide alterations associated with the resistant phenotype. The identity of the proteins showing differences can then be determined using mass spectrometry followed by the screening of gene sequence databases.

The specific objectives of the project are:

- To find the best solubilisation method for proteins from trypanosomes to maximise sensitivity and reliability of proteomic analysis
- To construct a partial proteome map of *T. brucei*
- To identify changes in the polypeptide profile of *T. brucei* that are associated with arsenical resistance
- To characterise the polyptetides associated with Cymelarsan resistance

- To conduct downstream analyses of any proteins identified associated with drug resistance, with the aim of developing testable hypothesis of the mechanism of arsenical resistance in *T. brucei*

Chapter II

The development of proteomic techniques for *T. brucei*

2.1 Introduction

Two-dimensional gel electrophoresis (2DE) has come into widespread use since the publication, in the early seventies, of methods combining isoelectric focussing (IEF) in the first dimension with sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension (Ames and Nikaido, 1976). In recent years, the technique of 2DE was improved by the introduction of immobilised pH gradients (IPG) for the separation of proteins by IEF to replace the use of carrier ampholytes. Immobilised pH gradient gels are made by a small number of defined chemicals (acrylamido buffers) which are covalently linked to a polyacrylamide matrix. These allow true steady-state focussing, which greatly increases the reproducibility of the technique.

The protocol for 2DE using immobilised pH gradients is well developed and has recently been reviewed (Görg *et al*, 2000). However, sample preparation, i.e. the extraction and solubilisation of proteins in a solution compatible with 2DE, still requires to be determined empirically for each organism or cell fraction under study. For efficient protein solubilisation and high resolution on 2D gels it is necessary to achieve complete disruption of molecular interactions, to obtain single polypeptides, and maintain this state throughout the separation process. Many protocols have been developed to achieve this aim. Table 2.1 illustrates the evolution of sample preparation methods for proteins from protozoan parasites by 2DE. It is now generally accepted that a combination of chaotropes (i.e. urea and thiourea) and zwitterionic amphiphilic compounds (e.g. CHAPS or SB3-10)

Sample	IEF method	Solubilisation method	Detection	Reference
T. brucei	CA-IEF	0.25% NP-40, 2 mM PMSF	Iodinated lectins	Pearson *et al.*, 1981
L. tropica	CA-IEF	0.5% Triton X100, 20 mM EDTA, 2 mM PMSF, 10 mM Iodoacetamide	^{125}I labelling	Handman *et al.*, 1983
T. brucei	CA-IEF	9 M urea, 5% 2-mercaptoethanol, 4% Nonidet P-40, 2% ampholites	^{35}S labelling	Anderson *et al.*, 1985
T. brucei	CA-IEF	9 M urea, 2% 2-mercaptoethanol, 4% Nonidet P-40, 2% ampholites	^{35}S labelling	Pearson and Jenni, 1989
T. gondii	CA-IEF	9.8 M urea, 0.3% SDS, 3% NP 40, 2% ampholites	^{3}H labelling	Achbarou *et al.*, 1991
T. gondii	IPG-IEF	7 M urea, 2 M thiourea, 4% CHAPS, 1% DTT, 2% IPG buffer, 0.05% Triton X-100	Coomassie blue staining	Dlugonska *et al.*, 2001; Nischik *et al.*, 2001,
T. gondii	IPG-IEF	8 M urea, 2% CHAPS, 22 mM DTT, 0.5% IPG buffer,	Silver and Coomassie blue staining	Cohen *et al.*, 2002
L. infantum	IPG-IEF	7 M urea, 2 M thiourea, 4% CHAPS, 40 mM Tris base, 1 mM PMSF	Silver staining	El Fakhry *et al.*, 2002
T. brucei	IPG-IEF	9 M urea, 2 M thiourea, 2% CHAPS, 65 mM DTT, 0.5% IPG buffer, cocktail of protease inhibitors	Silver staining	Van Deursen *et al.*, 2003

Table 2.1: Development of solubilisation methodology for 2DE analysis of parasitic proteins. Selected examples from the literature to illustrate the evolution of solubilisation methodology and 2DE analysis. The use of IPG-IEF was a significant milestone in the analysis of parasitic proteins by 2DE.

in the extraction buffer are essential to solubilise most proteins from a wide range of organisms (Rabilloud *et al.*, 1997, Rabilloud, 1998).

The limitations of 2DE separation are well documented and is particularly evident for membrane proteins (Santoni *et al.*, 2000a; Santoni *et al.*, 2000b). The major obstacle of global 2DE separation remains the solubilisation of all the proteins from a sample with one sample preparation method. There are three reasons for this obstacle: first, many hydrophobic proteins are not solubilised in the non-detergent isoelectric focusing sample buffer (Santoni *et al.*, 2000b), second, solubilised proteins are prone to precipitation at their isoelectric point (Adessi *et al*, 1997) and third, hydrophobic proteins transfer inefficiently to the second dimensional gel (Wilkins *et al.*, 1998). The limited dynamic range of protein visualisation is also an issue because membrane proteins are typically lower in abundance compared with most soluble proteins. In addition, membrane proteins are heterogeneous due to the differential glycosylation of extracellular domains, and this results in membrane proteins focusing over a wide pH range during IEF, producing streaks on the 2D gels.

Sample preparation is the first important step of every proteome analysis. In this chapter, a comparison of three protocols for the extraction and solubilisation of proteins from *T. brucei* was made. The protocols were based on a recently published protocol for protein extraction from *T. brucei* (van Deursen *et al.*, 2003), which uses a combination of chaotropes, zwitterionic amphiphilic compounds as well as a cocktail of protease inhibitors. Protein extracts were separated on 2D gels and compared using three criteria: the resolution of the protein spots, the number of protein spots extracted and separated which could be visualised and the distribution of the protein spots at low/high molecular

weight/pI. The reproducibility of the extraction protocol and the solubilisation of membrane proteins using the selected protocol were also assessed.

2.2 Materials and Methods

2.2.1 Commonly used reagents

All equipment and reagents for IEF and SDS electrophoresis (IPGphor, Multiphor II horizontal electrophoresis unit, Hoefer Dalt multiple vertical electrophoresis unit, EPS 3500 XL power supply, Multitemp II thermostatic circular, ImageMaster 2Delite V3.01 software, Typhoon scanner 9400, Labscan, Immobiline DryStrip, IPG buffers, TEMED, CHAPS, DTT, APS, urea, glycine, Tris, ECL kit, Hybond-C nitrocellulose membrane and Protein Silver Staining Kit) were purchased from Amersham Biosciences. TPCK, TLCK, EDTA, EGTA, Leupeptin, Pepstatin A, Coomassie Brilliant blue G250 and Gentamycin were purchased from Sigma. SDM 79 culture medium and foetal calf serum were purchased from GibcoBRL. The Sypro ruby stain and 30% polyacrylamide solutions were purchased from BioRad.

2.2.2 *In vitro* culture of procyclic trypanosomes

For this study, two laboratory stocks of *T. brucei* (TREU 927 line and STIB 386 line) were used. The lines will be referred to as 927 and 386. The lines were grown in mice from a cryopreserved stabilate before being transformed into procyclic forms by transfer of parasites to Semi Defined Medium 79 (SDM79) with 3 M cis-asconitate (Brun and Schonenberger, 1981).

Procyclic trypanosomes were grown in suspension culture either in T-25 or T-75 culture flasks (Falcon, UK). The culture medium was SDM79 supplemented with 10% heat

inactivated (56°C, 30 min) foetal calf serum and 0.2 mg Gentamycin, referred to as complete SDM79 (Brun and Schonenberger, 1979). Flasks were incubated at 25°C. Trypanosomes from 5 ml of procyclic culture at $5x10^6$ trypanosomes/ml were concentrated by centrifuging at 12,000 g for 3 min. Stabilates were made by re-suspending the trypanosomes in 0.5 ml of complete SDM79 medium and 0.5 ml of freezing medium (10% glycerol in heat inactivated foetal calf serum). The trypanosomes were then transferred into freezing vials and cooled in a cotton wool lined polyester container at -70°C for 6 h prior to transfer to liquid nitrogen.

2.2.3 SDS-PAGE mini gels

SDS-PAGE mini gels were used to assess the quality and quantity of proteins extracted using the solubilisation protocols. These were also used as the second dimension for 2DE analysis with 7 cm IPG strips.

Polyacrylamide gels (12%) were made in 0.375 M Tris-HCl pH 8.8, 1% (w/v) SDS, 1% (w/v) ammonium persulfate, 5 µl TEMED, using a 30% polyacrylamide mix.

For the quantification of protein extracts, a stacking gel was added onto the 12% gel. The stacking gel consisted of a 4% polyacrylamide gel in 0.125 M Tris-HCl, pH 6.8.

The samples were prepared as follows: 10 µl of the protein extract was mixed with 10 µl of SDS loading buffer (62.5 mM Tris-HCl, pH 6.8, 10% glycerol, 2% (w/v) SDS, 0.4 ml 2-mercaptoethanol, 0.05% (w/v) bromophenol blue), boiled for 1 min and then chilled on ice prior to loading onto the SDS-PAGE gel. Separation was performed by electrophoresis at 110 V for 2 h. Gels stained with Coomassie Brilliant blue were incubated in a Coomassie staining solution (0.5 g Coomassie Brilliant blue G 250, 100 ml acetic acid, 125 ml iso-propanol and ddH_2O up to 1 L) overnight. Gels were destained by multiple washes in a solution of 10% acetic acid/ 20% methanol until background staining had disappeared. Gel images were acquired using an AmershamPharmacia Biotech Labscan. Images of the gels

were acquired at 300 dots per inch using Labscan sv3.0 software on an Umax flatbed scanner with integrated transparency adapter.

2.2.4 Sample preparation for 2DE

A published protocol (P1, van Deursen *et al.*, 2003) for the extraction and solubilisation of proteins from *T. brucei* was compared with two modified protocols (P2 and P3). Procyclic form trypanosomes ($1x10^8$ trypanosomes/ml) were concentrated by centrifuging for 8 min at 12,000 g in a centrifuge tube and the pellet was transferred into an Eppendorf tube. The pellet was washed twice in 500 µl of phosphate buffered saline with 1% glucose (PSG), pH 7.4. Trypanosomes were resuspended and lysed in 350 µl of lysis/rehydration buffer consisting of 9 M urea, 2 M thiourea, 2% CHAPS, 65 mM DTT, 0.5% IPG buffer and a trace of bromophenol blue. 35 µl of protease inhibitor cocktail 1 (1 mM EDTA, 1 mM EGTA, 1 µg/ml Pepstatin A, 10 µg/ml Leupeptin, 100 µg/ml TPCK, 50 µg/ml TLCK) was added to the trypanosomes suspension. The trypanosomes were then incubated for 1 h at room temperature with the tubes being vortexed every 10 min, and the resulting lysate was stored at -20°C (protocol P1) (van Deursen *et al*, 2003). In the first variant of this protocol, a nuclease mixture (2,000 units/ml DNAase 1,750 units/ml RNAase A, 50 mM $MgCl_2$, 0.5 M Tris pH 8.0) was added to the sample at the same time as the lysis/rehydration buffer. The tubes were incubated on ice for 10 min before the addition of the protease inhibitor cocktail 1 and then incubated for 1 h at room temperature with vortexing every 10 min (protocol P2). In the second variant, a commercially available protease inhibitor cocktail ("COMPLETE" protease inhibitor cocktail Tablets (Roche) complemented with 10 µl of a EDTA, EGTA and PMSF solution) was used (protocol P3) instead of cocktail 1. This cocktail was added immediately to the lysis/rehydration buffer. No nucleases mixture was used in this protocol. The tubes were incubated 1 h at room temperature as described for

	Protocol P1	Protocol P2	Protocol P3
Lysis/rehydration buffer	350 µl	350 µl	350 µl
Nucleases mixture		10 µl	
Proteinase inhibitor cocktail 1	35 µl	35 µl	
Proteinase inhibitor cocktail 2			35 µl

Table 2.2: Summary of protocol P1, P2 and P3. The volumes given in the table correspond to the volumes of reagents used for the extraction of proteins from $1x10^8$ trypanosomes to be loaded onto a 18 cm 2D gels.

the other protocols. A summary of the three protocols is given in Table 2.2. The lysates were centrifuged for 5 min at 12,000 g to remove insoluble material and stored at -20°C.

2.2.5 Two-dimensional gel electrophoresis

Extracts were run in the first dimension using IPG strips rehydrated with the supernatant of proteins extracted in lysis/rehydration buffer. The volume of the protein extract used for the rehydration of the strip depended on the size of the IPG strip: 125 µl was used on 7 cm IPG strips and 350 µl was used on 18 cm IPG strips. Rehydratation and IEF were carried out on an IPGphore at 20°C and using 50 µA/strip. Rehydratation was performed at 30 V for 15 h, followed by IEF at 500 V for 500 V/h, 1,000 V for 1,000 V/h, 8,000 V for 3,975 V/h by gradient and a final stage at 8,000 V/h for 4 h (7 cm gels) or 6 h (18 cm gels).

For the separation in the second dimension, the rehydrated IPG strip was first equilibrated for 15 min by rocking in a solution of 50 mM Tris-HCl pH 8.8, 6 M urea, 30% glycerol, 2% SDS and 100 mg of DTT to reduce the proteins, and then for 15 min in a solution of 50 mM Tris-HCl pH 8.8, 6M urea, 30% glycerol, 2% SDS and 250 mg of α-Iodoacetamide (Calbiochem) in order to alkalynate the cysteine residues. The IPG strip was then rinsed in electrophoresis buffer before being embedded on a 12% SDS-PAGE gel with 0.5% (w/v) agarose. Electrophoresis buffer consisted of 25 mM Tris base, 192 mM glycine and 0.1% SDS. Gels were run at 110 V for 2 h (7 cm gels) or 110 V for 24 h (18 cm gels) at 15°C.

2.2.6 2D gel staining and image acquisition

The gels were stained either in Coomassie Brilliant blue, silver or Sypro ruby (BioRad). Staining with Coomassie Brilliant blue was performed as described for SDS-PAGE gels. Silver staining was performed using the Protein Silver Staining Kit and the automatic gel stainers according to the manufacturer's instructions. Sypro ruby staining of 2D gels was

performed following the manufacturer's (BioRad) instructions. The gels stained with Coomassie Brilliant blue or silver were imaged as described previously for SDS-PAGE mini gels. Images from gels stained with the fluorescent dye, Sypro ruby, were acquired using a Typhon scanner 9400 (variable mode imager) under the following settings: emission filter set at 560 LP green/purple, PMT= 550, laser = green 532, sensitivity normal, pixel size = 100 microns (Görg *et al.*, 2000).

2.2.7 Western blotting

Following SDS-PAGE analysis, separated proteins were transferred to Hybond-C nitrocellulose membrane by electroblotting in transfer buffer (20 mM Tris, 225 mM Glycine) using a BioRad mini transblot cell, according to the manufacturer's instructions. Proteins were transferred at 110 V for 1 h 30 min. The membrane was stained with Ponceau S to allow rapid assessment of protein transfer to the nitrocellulose membrane. The nitrocellulose membrane was then placed in blocking solution consisting of 5% skimmed milk in TBS-Tween (0.02 M Tris, 0.137 M NaCl pH 7.6, 0.05% Tween 20) and incubated overnight at 4°C. The nitrocellulose membrane was washed for 10 min in TBS-Tween at room temperature three times and then incubated for 1 h 30 min in the primary antibody diluted in TBS-Tween. Two primary antibodies raised against *T. brucei* Rab2 and Rab11 proteins were used (gift from Dr M. Field, Imperial College London). The membrane was washed for 10 min in TBS-Tween three times to remove any unbound antibody, secondary antibody, horseradish peroxidase-conjugated anti-rabbit IgG (Diagnostics Scotland, 1:1,000 dilution) in TBS-Tween, was added and incubated for 1 h 30 min at room temperature. The membrane was washed for 10 min in TBS-Tween three times. The bound conjugated antibody was detected using an ECL kit and incubation for 1 min. The surplus solution was drained from the membrane which was then exposed to the autoradiographic films from 1 to 10 min as required (Kaufmann *et al.*, 2001).

Autoradiographs were developed using a film processor (X-Ograph imaging system compact X4) according to the manufacturer's instructions.

2.3 Results

2.3.1 Assessment of the three sample preparation protocols

The three sample preparation protocols developed for the extraction and solubilisation of proteins from *T. brucei* were compared using $1x10^8$ procyclic forms trypanosomes of stock 927. The concentration of protein extracted using each protocol could not be assessed with either a Bradford technique (e.g. Coomassie Plus Protein Assay from Pierce) or an assay based on the biuret reaction (e.g. BCA Protein Assay from Pierce) because of the presence of high concentrations of urea, thiourea and DTT reagents in the lysis/rehydration buffer (data not shown). Instead, 12% SDS-PAGE mini gels were run to assess the relative concentration and quality of protein extracted (Figure 2.1). Limited differences were observed between the proteins extracts visualised on the SDS-PAGE gel, but protocol P3 gave a higher number of protein bands with an especially good representation of those at high molecular weights.

In order to further assess the quality of the proteins extracted using the three protocols, the extracts were separated on 18 cm pH 3-10 non-linear (nl) IPG strips in the first dimension followed by 12% polyacrylamide gels in the second dimension. The gels were stained with silver, scanned and analysed using the ImageMaster 2DElite software (Figure 2.2A). Only 303 protein spots were detected on the 2D gel run with protein extracted using protocol P1. The protein spots were concentrated mainly between pH 4 and 7 and very few were detected at more basic pH. Detection of spots using ImageMaster 2DElite software on this

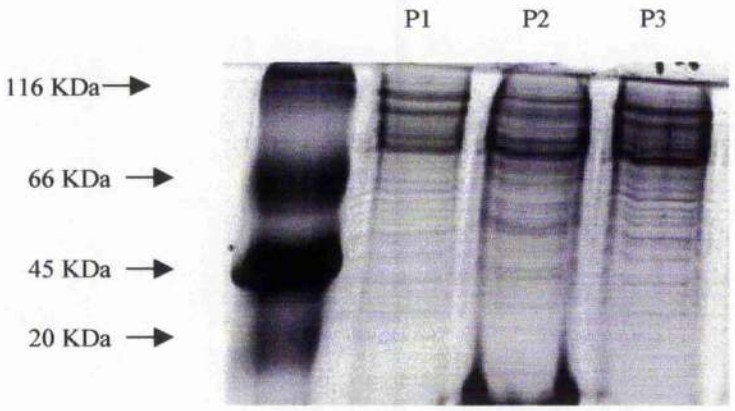


Figure 2.1: SDS-PAGE analysis of proteins extracted using protocols P1, P2 and P3. Proteins from $1x10^8$ procyclic trypanosomes were extracted. The extracts were run on a 17% SDS-PAGE gel to assess the quantity and quality of the extracts. The gel was stained with Coomassie Brilliant blue. Lane 1: broad range molecular weight marker (catalogue number C3437, Sigma).

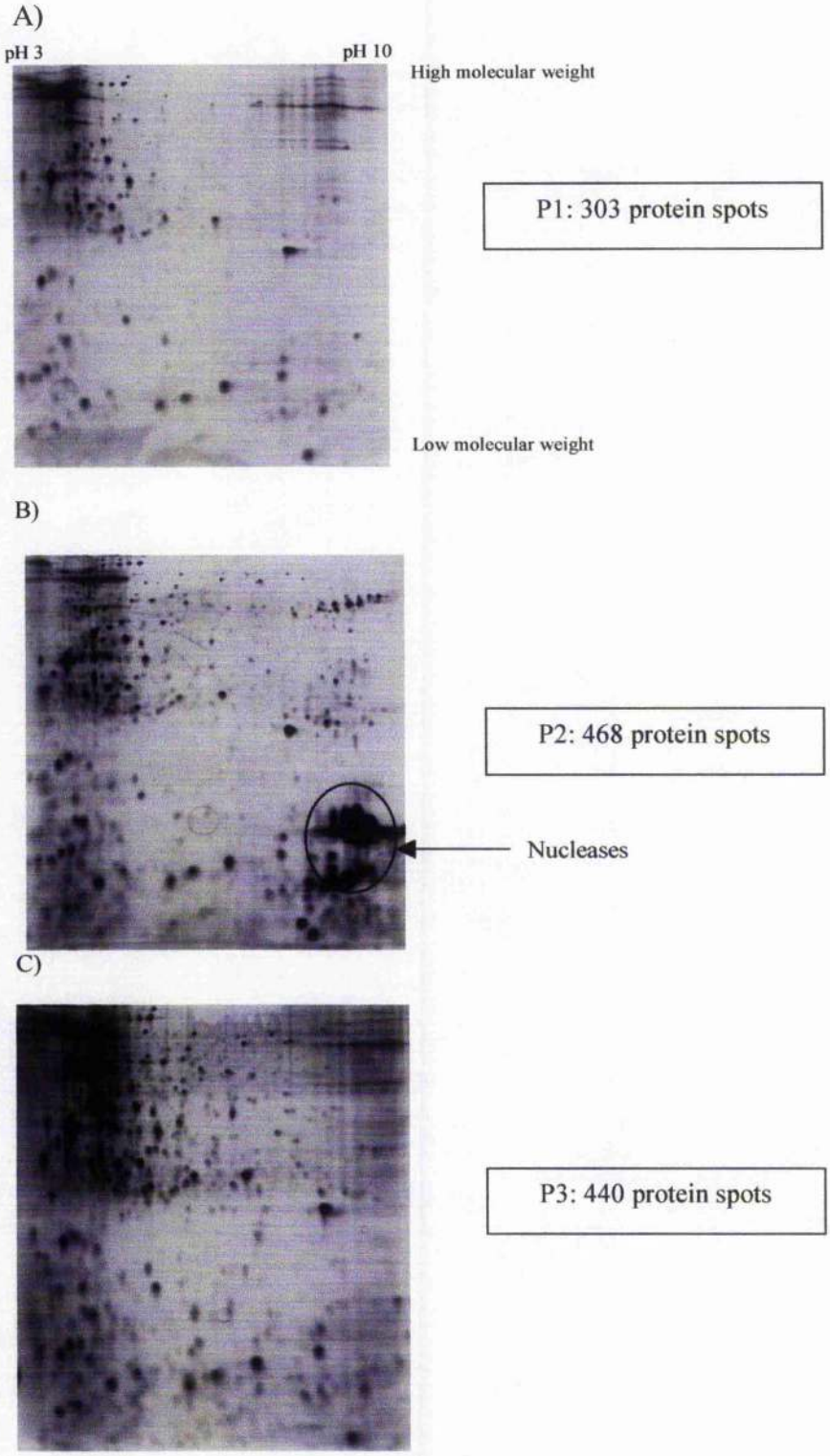


Figure 2.2: 2DE separation of proteins extracted using protocol P1, P2 and P3. Proteins were extracted from $1x10^8$ trypanosomes, separated on broad pH IEF (pH 3-10nL) 2D gels and silver stained. The gels were analysed using ImageMaster 2DElite software to determine the number of protein spots separated.

gel was difficult because of the poor resolution of high molecular weight spots. 468 and 440 protein spots were detected on silver stained 2D gels of proteins extracted using protocols P2 and P3 respectively (Figure 2.2B and C). The resolution on these gels was much higher than for protocol P1. A good representation of low and high molecular weight protein spots as well as protein spots located at acidic and basic pH was observed on these 2D gels. The 2D pattern of proteins extracted using protocol P3 showed horizontal streaking on the gel in the area of high molecular weight proteins. This prevented the complete separation of the protein spots located at this region of the gel. No streaking was observed on 2D gel run with proteins extracted using protocol P2.

Based on theses results, protocol P2, combining the lysis/rehydration buffer with nucleases and protease inhibitors, was the most efficient sample preparation for 2DE analysis of *T. brucei* proteins. For all subsequent analysis, protocol P2 was chosen since the greater number of spots were obtained and were separated with the highest resolution using this sample preparation protocol.

2.3.2 Assessment of the reproducibility of protocol P2

Three independent extractions were undertaken using protocol P2 on different days from different cultures of *T. brucei* 927. The samples were separated by 2DE (pH 4-7) on 18 cm gels. The gels were silver stained and analysed (Figure 2.3). The 2D pattern on the gels was highly reproducible. The intensity of staining of spots also seemed to vary very little between the different extractions.

Proteins from three different concentrations of trypanosomes ($1x10^7$, $1x10^8$ and $1x10^9$) were extracted using protocol P2 and separated by 2DE (pH 3-10nl) to assess the optimal

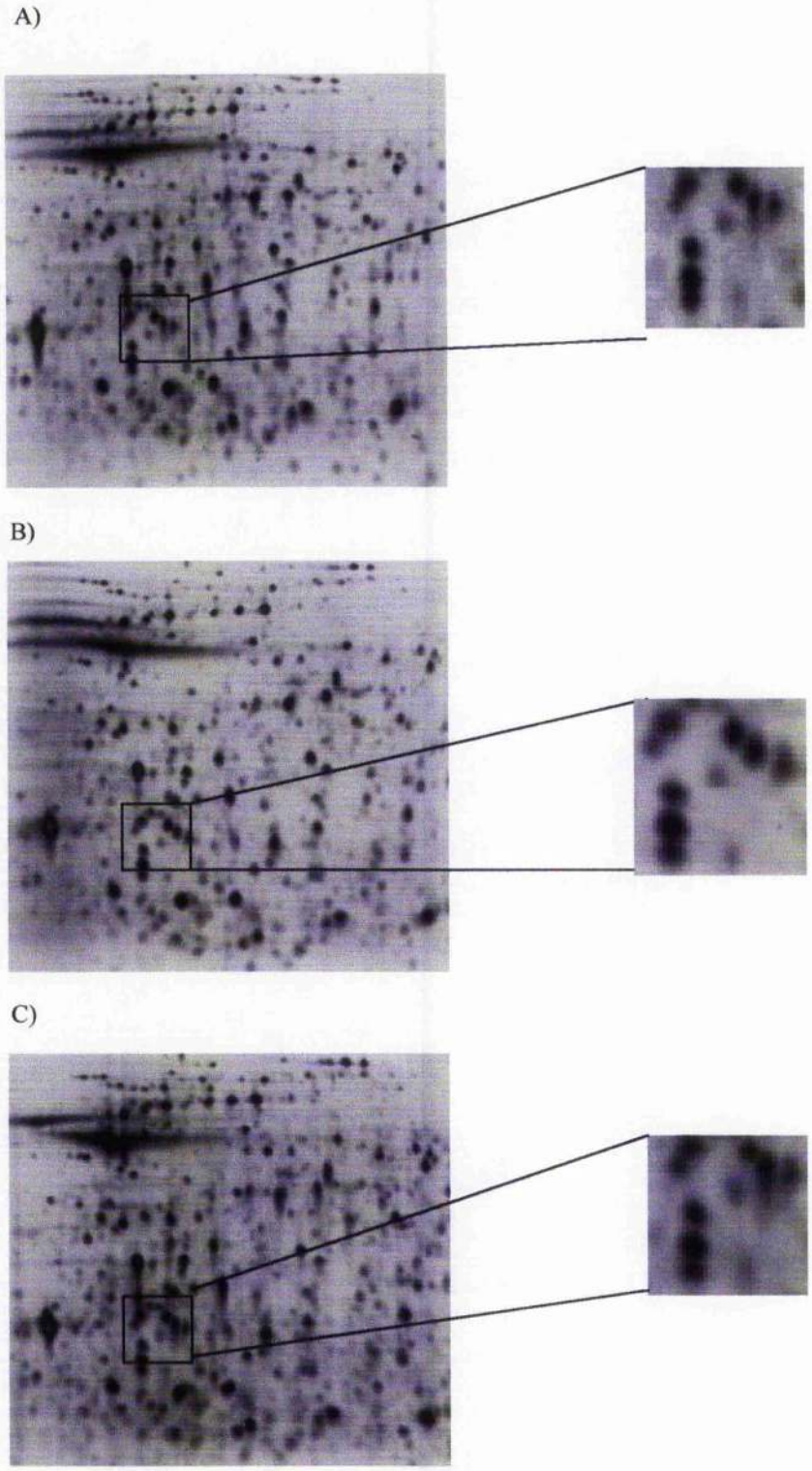


Figure 2.3: 2DE separation of three independent protein extracts using protocol P2. The reproducibility of the sample preparation using protocol P2 was assessed by running three independent protein extracts on pH 4-7 2D gels.

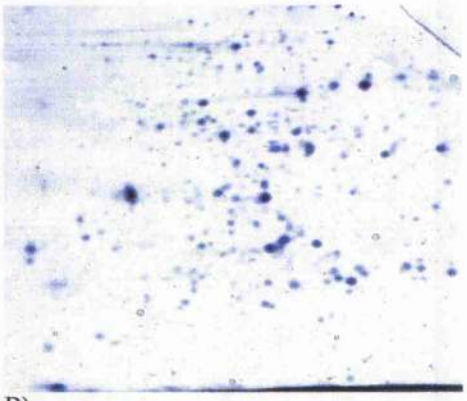
protein concentration for running over 18 cm 2D gels. The gels were silver stained and analysed with the 2DElite software (data not shown). At a concentration of $1x10^9$ trypanosomes (about 4 mg of proteins) horizontal streaking due to protein precipitation was observed, preventing the resolution of the spots. The highest number of protein spots with the best resolution were obtained from extracting proteins from $1x10^8$ trypanosomes, corresponding to about 0.7 mg of proteins, and this concentration was loaded onto gels to be silver stained for the rest of the project.

2.3.3 Comparison of staining techniques for the detection of proteins

One of the major limitations of the 2DE technique is the visualisation of the proteins once they have been separated on 2D gels. Unlike genomics, proteomics has no equivalent of amplification by PCR and thus sensitivity of the staining technique is a critical issue. Three protein stains, Coomassie Brilliant blue, silver and Sypro ruby, were compared for their sensitivity in detecting proteins separated by 2DE.

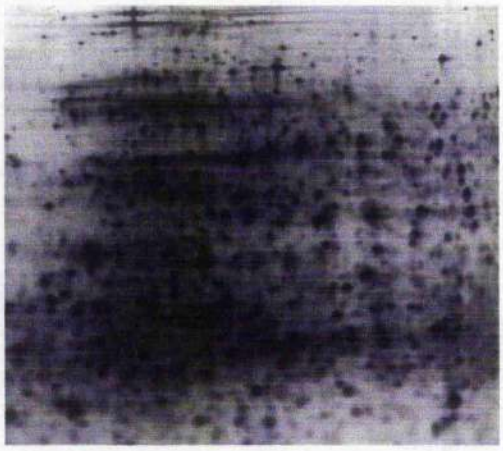
Proteins were extracted from procyclic trypanosomes using protocol P2, separated on three 18 cm 2D gels, pH 4-7 and the 2D gels were stained with either Coomassie Brilliant blue, silver or Sypro ruby (Figure 2.4). Image analysis of these gels enabled the detection of 909 protein spots on the silver stained gel, 605 protein spots on the Sypro ruby stained gel but only 203 protein spots on the Coomassie Brilliant blue stained gel. Silver staining was the most sensitive of the three stains used in this experiment. The developing of the silver stained gels, using the automatic gel stainer, gave inconsistent results (data not shown) possibly due to variation in water temperature.

A)



Coomassie: 203 protein spots

B)



Silver: 909 protein spots

C)

Sypro ruby: 605 protein spots

Figure 2.4: **A comparison of protein stains.** Solubilised proteins from procyclic trypanosomes ($1x10^8$) extracted using protocol P2 were loaded onto pH 4-7 gels. The gel images were analysed using the ImageMaster 2DElite software.

2.3.4 Assessment of the extraction of a membrane proteins using protocol P2

Membrane proteins are among the most difficult proteins to solubilise during sample preparation. This characteristic may exclude membrane proteins from the IEF step of 2D gels. Western blot analysis was used to assess the solubilisation and separation of two known *T. brucei* membrane proteins, Rab2 and Rab11, (Field *et al.*, 1999, Pal *et al.*, 2003) using protocol P2 and their separation by both one and two dimensional gel electrophoresis.

Proteins were separated on one dimensional SDS-PAGE gels and Western blotted using antibodies against the Rab2 and Rab11 membrane proteins of *T. brucei*. In addition, proteins were extracted directly from cells by boiling in SDS-PAGE loading buffer. This highly denaturing environment should extract the membrane proteins, and so this extract was used as a positive control. These extracts were run in parallel on SDS-PAGE gels. Coomassie Brilliant blue staining of the gel showed the increased number of protein extracted by boiling the trypanosome pellet in SDS loading buffer as compared with protocol P2 (Figure 2.5). More specifically, high molecular weight proteins were under represented in the protein extracts from protocol P2.

The Rab2 protein was extracted from procyclic trypanosomes by both extraction techniques as shown by Western blot analysis (Figure 2.6A). Western blot analysis with antibodies against the *T. brucei* Rab11 protein was also performed in a similar experiment (Figure 2.6B). No pre-immune serum was available as a control for either antiserum but the absence of binding with anti Rab11 allowed the results to be used as a negative control for Rab2.

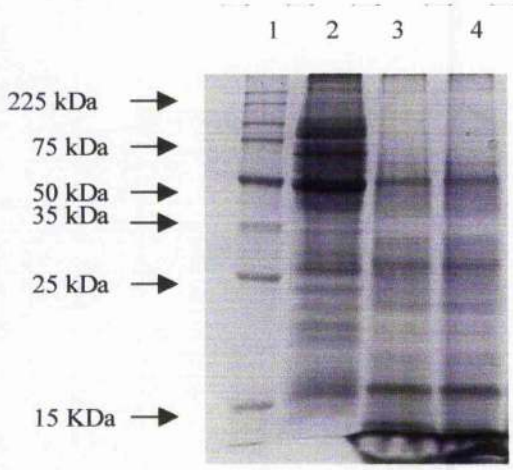


Figure 2.5: SDS-PAGE gel of protein extracted using protocol P2. Coomassie Brilliant blue stained SDS-PAGE gel of protein extracted from line 386. Proteins were extracted by dissolving $1x10^8$ trypanosomes in SDS loading buffer (lane 2) or using protocol P2 developed for 2DE analysis (lane 3 and 4).

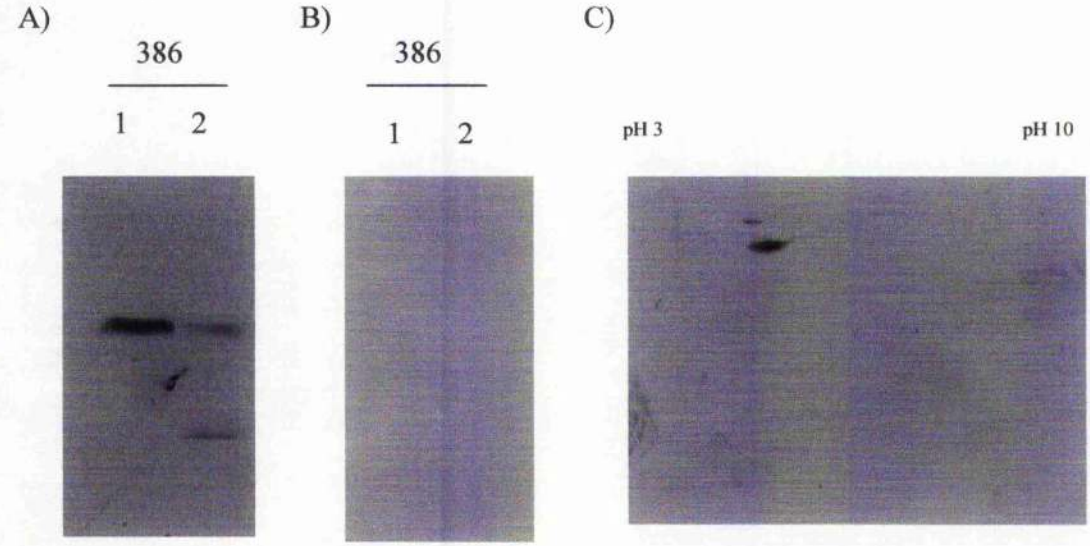


Figure 2.6: Western blot analysis of Rab2 and Rab11 extracted from *T. brucei* and separated by SDS-PAGE gel and 2DE. Proteins were extracted from procyclic forms by either boiling the trypanosome pellet in SDS loading buffer (lane 1) or using protocol P2 (lane 2) developed for 2DE sample preparation. Western blot analysis was performed using affinity-purified rabbit TbRab2 antiserum (A) and TbRab11 antiserum (B) (gift from Dr M. Field, Imperial College London). Whole cell extract from procyclic *T. brucei* 386 was separated by 2DE on a 7 cm pH 3-10nl strip. Western blot analysis was performed using affinity-purified rabbit TbRab2 antiserum (C).

To determine whether the Rab2 protein could be detected on a 2D gel, a 7 cm 2D gel (pH 3-10nl) was run with protein extracted using protocol P2 and analysed by Western blot (Figure 2.6C). The protein spots detected with the anti-Rab2 antibody were resolved as a horizontal streak, from which three protein spots located at similar molecular weight and slightly different pI could be detected within the streak (Figure 2.6C). This demonstrated that the Rab2 protein entered the IPG strip during the IEF separation step, were partially separated by IEF and transferred in the SDS-PAGE gel.

These results would suggest that a combination of chaotropes, zwitterionic amphiphilic compounds and DTT in the sample preparation buffer enables the extraction of at least some membrane proteins and their separation by 2DE.

2.4 Discussion

The "ideal" sample preparation would require extracting all the protein components, including hydrophobic membrane proteins, proteins with extreme pI a well as low copy number proteins in the presence of the most abundant proteins. Disruption of hydrogen bonds by chaotropes, prevention of hydrophobic interactions by surfactants and destruction of intramolecular and intermolecular disulfide bonds by reducing agents are essential in order to obtain a high level of protein extraction and therefore the detection of a maximum number of protein spots by 2DE and to avoid the presence of multiple proteins per spot (Rabilloud, 1998; Molloy, 2000; Lanne *et al.*, 2001). Inclusion of a cocktail of proteinase inhibitors will minimise proteolytic activity and so minimise the generation of multiple different derivatives of each protein.

To achieve this aim, protocols have been developed which include the use of a combination of urea and thiourea (chaotropes) with CHAPS (surfactant), DTT (reducing agent) and a cocktail of protease inhibitors. The addition of nucleases to the lysis/rehydration solution caused an increase in the number of protein spots and a higher resolution of these spots on 2D gels.

The sample preparation protocol needs to be as simple as possible in order to minimise sample variability. In the P2 protocol developed here, all the reagents were added to the trypanosome pellet in one step and the lysis occurred at room temperature. This gave very high reproducibility of 2D patterns from independent extractions.

Global protein expression studies using silver stained pH 4-7 gels demonstrated a variation in the number of resolved protein spots from 630 spots (Cohen *et al.*, 2002) to about 1000 spots (van Deursen *et al.*, 2003). Using the P2 protocol developed in this study 909 protein spots could be detected by silver staining of pH 4-7 gels. , which ranks the P2 solubilisation protocol in the higher range of the protocols used for global expression profiling.

The silver stain was more sensitive for protein detection on 2D gels than either Coomassie Brilliant blue or Sypro ruby. However, silver staining has several drawbacks for 2DE analysis such as a poor reproducibility, a limited dynamic range and the fact that certain proteins stain poorly, negatively or not at all (Lauber *et al.*, 2001). In addition, low detection limits often occur in mass spectrometry because the stain modifies cysteine residues and the glutaraldehyde and formaldehyde used in the procedure alkylate protein amino groups (Patton, 2000). The fluorescent stain, Sypro ruby, was slightly less sensitive

(about 66% of the protein spots detected by silver were detected by Sypro ruby), but was more reproducible and easier to perform. The detection limit of the Sypro ruby dye is in the range of 1-2 ng of protein per spot, and the linear range is over three orders of magnitude (Görg *et al.*, 2000). Consequently Sypro ruby and Coomassie Brilliant blue were the preferred detection techniques for proteomic analysis of *T. brucei*. Sypro ruby being a fluorescent stain rendered the excision of the protein spots of interest from the 2D gels difficult and requires special equipment such as the Amersham Ettan robotic workstation, which was not available at the beginning of this study. Consequently, Coomassie Brilliant Blue staining was used at first to detect highly abundant differentially expressed proteins and Sypro ruby was used in the second pass to determine differentially expressed low abundance proteins.

Two antibodies against Rab proteins were used on Western blot analysis to assess the extraction of membrane proteins using protocol P2 and the of these proteins separation by 2DE. Rab proteins are small GTPases involved in the endocytosis and exocytosis transport in trypanosomes. They control the direction and timing of vesicle fusion during intracellular trafficking between membraneous compartments. Rab proteins are not transmembrane proteins, instead they are peripheral proteins thought to be embedded in the membrane through their isoprene moieties. Rab2 is restricted to endoplasmic reticulum (ER) exit sites and ER-Golgi intermediate compartments (Field *et al.*, 1999) while Rab11 is located in the endosomal system and co-localises with internalised anti-VSG antibody and Transferrin (Pal *et al.*, 2003). The Rab2 protein was shown to be extracted by protocol P2 and to separate by 2DE. No detection of Rab11 protein could be obtained in either the sample extracted by protocol P2 or the sample obtained by boiling the trypanosome pellet in SDS buffer. Rab2 is known to be expressed at similar levels in procyclic and

bloodstream forms of *T. brucei* whereas the Rab11 protein is strongly up regulated in bloodstream trypanosomes (Field and Field, 1997; Pal *et al.*, 2003). As the samples used in this study were extracted from procyclic trypanosomes, the level of Rab11 proteins might have been too low to be detected. This experiment was performed to assess the extraction of membrane proteins by the solubilisation protocol developed. As Rab proteins are peripheral proteins, it is possible that the Rab2 protein detected in the 2DE extract was due to the detachment of the Rab protein from its isoprene moiety. Transmembrane proteins are the most difficult membrane proteins to solubilise as they contain hydrophobic regions that enable part of the protein to reside within the membrane. Extraction of transmembrane protein from *T. brucei* could be performed using anti-cystein-rich transmembrane protein (anti-CRAM) antibodies (Liu *et al.*, 2000). CRAM protein is located at the flagellar pocket and is involved in the endocytosis pathways and has been shown to be expressed in procyclic forms of trypanosomes (Liu *et al.*, 2000). As such this protein would be a good test to determine the extraction of transmembrane proteins by protocol P2.

The main aim of the work described in this chapter was to develop a simple, complete and reproducible method for sample preparation from *T. brucei*. This was achieved by comparing the proteins extracted using three different protocols followed by analysis on SDS-PAGE gels and 2D gels. Sample preparation P2 was shown to be simple, reproducible and robust, and it enabled the extraction of at least some membrane proteins, so that 2D gels of *T. brucei* proteins can now be run reproducibly. The protein spots of interest can then be analysed by mass spectrometry techniques to generate information about the protein spots. The mass spectrometry data can then be used for identification of the protein spot as described in the next chapter.

Chapter III

Assessment of algorithms for protein identification from peptide mass fingerprint data

3.1 Introduction

In the previous chapter a simple, efficient and reproducible sample preparation method was developed for 2DE analysis of protein extracted from procyclic forms of *T. brucei*. The next step, after separation of proteins by 2DE, is the identification of the proteins of interest by mass spectrometry (MS).

Various mass spectrometry techniques are available (peptide mass fingerprint, peptide sequence tags) for the identification of protein spots from 2D gels. The simplest and fastest method is peptide mass fingerprinting (PMF) which can be powerful when combined with a protein sequence database. A peptide mass fingerprint consists of the mass of each peptide generated from digestion of a protein with a specific protease as measured by mass spectrometry. A peptide mass fingerprint provides no inherent sequence information. The commonly used enzyme for digestion of proteins is trypsin which specifically cleaves the peptide bond at the C-terminus of arginine and lysine residues, except if they are followed by a proline. Peptide mass fingerprints are generated by Matrix-Assisted-Laser-Desorption-Ionsation Time-Of-Flight (MALDI-TOF) mass spectrometry. The use of peptide mass fingerprints generated from mass spectrometry for protein identification was first demonstrated in 1993 for cytochrome c from dog and pigeon, a GTP-binding protein from humans, an uteroferrin from pig, the human sex steroid binding protein and a thermostable DNA polymerase (Yates *et al.*, 1993).

Search engines have been developed to enable the rapid searches of protein and nucleotide sequence databases with PMF data. Three common steps are involved for all search engines: firstly, each protein in the database is digested *in-silico* based on the specificity of the enzyme and the masses of the resulting peptides are calculated enabling the construction of a theoretical peptide mass fingerprint; secondly, the experimentally determined peptide mass fingerprint is compared with the theoretical peptide mass fingerprint and a score is calculated which reflects the accuracy of the match between the experimental and theoretical peptide masses; thirdly, the protein sequences in the database are sorted according to their score and the protein sequence with the best score is selected (Fenyö, 2000).

All search engines developed for protein identification by PMF are based purely on a statistical approach. With the increased use of proteomic analysis, several types of search engine for protein identification by PMF have been developed and made available on the Internet (Table 3.1). All are based on the same principle: the comparison of a peptide mass fingerprint generated from a protein with the theoretical peptide mass fingerprint of all available proteins in a database (Fenyö, 2000). Each new generation of search engine has been improved by increasing the sensitivity and specificity of the protein identification. The sensitivity of a search engine can be defined as its ability to detect a match in the database, whereas its specificity is the ability to distinguish random from real matches. The main improvement made to these search engines in recent years has been the addition of algorithms for evaluating the search results. The most important aspect of the search algorithm is the need to distinguish a correct identification from a random match: i.e. specificity.

Name	Internet location
PepSea	http://pepsea.protana.com/PA_PepSeaForm.html
PeptIdent/MultiIdent	http://www.expasy.ch/tools/peptident.html
MS-Fit	http://prospector.ucsf.edu/ucsfhtml3.2/msfit.htm
ProFound	http://www.proteometrics.com/prowl-cgi/ProFound.exe
MASCOT®	http://www.matrixscience.com/cgi/searchform.pl?SEARCH=PMF

Table 3.1: List of search engine available for protein identification from peptide mass fingerprint data.

The only comparison of the various algorithms for the identification of protein using PMF data found in the literature was performed *in-silico,* using theoretical protein sequences from NCBInr cleaved using trypsin cleavage rules to obtain a random peptide monoisotopic mass pool (Tang *et al.*, 2000). This study was performed on three search engines (MS-Fit, ProFound and MASCOT®) and demonstrated a wide variation between the sensitivity and specificity of the search engines. This comparison does not appear to have been performed on real proteins separated by 2DE, digested with trypsin and analysed by MS.

In this study, 126 peptide mass fingerprints, generated by MALDI-TOF MS from *T. brucei* proteins separated by 2DE, were used to evaluate the specificity of three publicly available search engines. The MS-Fit, MASCOT® and ProFound search engines were used to search the National Centre for Biotechnology Information protein database (NCBInr). The evaluation of the sensitivity and specificity of the three search engines for the identification of proteins from *T. brucei* was performed by comparing the results obtained for each peptide mass fingerprint.

3.2 Materials and Methods

3.2.1 Commonly used reagents

Ammonium bicarbonate, TFA, CHCA and acetonitrile were purchased from Sigma. Trypsin was purchased from Promega. The ZipTip™ pipette tips were purchased from Millpore.

3.2.2 Protein extraction, 2DE and staining

Proteins were extracted from $2x10^8$ procyclic trypanosomes from line 927 according to protocol P2 developed in chapter 2. The proteins were separated by 2DE and stained using Coomassie Brilliant blue following the protocols described in chapter 2.

3.2.3 In-gel trypsin digestion and mass spectrometry

Gel pieces were excised from the gel using a scalpel and transferred into an Eppendorf tube. The gel pieces were washed for 1 h 30 min in 100 mM ammonium bicarbonate, 1 h 30 min in 50% acetonitrile/100 mM ammonium bicarbonate, then 10 min in 100% acetonitrile and then dried in a vacuum centrifuge. The gel pieces were rehydrated with 0.2 µg of trypsin in 25 mM ammonium bicarbonate for 15 min. In-gel digestion was carried out overnight at 37°C, on a horizontal shaker at 1,000 g (Hellman *et al.*, 1994).

The resulting peptides were desalted and concentrated using ZipTip™ pipette tips containing C_{18} reverse phase media. The samples were acidified by addition of 3 µl of 3% trifluoroacetic acid (TFA) prior to desalting. The ZipTip™ was equilibrated in 50% acetonitrile by aspirating 10 µl into the tip twice and then washing twice with 10 µl of 0.1% TFA. Peptides were then bound to the reverse phase and washed twice with 0.1% TFA. Peptides were eluted in 2 µl of 50% acetonitrile/ 0.1% TFA (Erdjument-Bromage *et al.*, 1998, Lamer and Jungblut, 2001).

The concentrated desalted sample (0.5 µl) was mixed with 0.5 µl of matrix solution (10 mg/ml α-cyano-4-hydroxycinamic acid (CHCA) in 50% acetonitrile/ 0.1% TFA), spotted onto a stainless steel plated MALDI sample stage, and allowed to air-dry for 30 min. Spectra were acquired using the PerSpective Biosystems Voyager-DE instrument in

reflector/delayed extraction mode. Peptides are co-crystallised with the matrix, a small weakly acidic molecule, which absorbs the energy from the laser and emits energy as heat. This results in the desorption of samples which are then accelerated into the vacuum. Since all ions are accelerated from a fixed point with the same potential, they will drift according to their mass to charge ratio. Calibration was achieved using the porcine trypsin auto-digestion peptides at 842.510 and 2211.1046 Da $[M+H]^+$. Composite spectra consisting of 200 laser shots were acquired for each protein spot.

3.2.4 Protein identification search engines

3.2.4.1 First generation algorithm: MOWSE

The first generation of algorithms developed for the use of PMF for protein identification were based on the simplest and most obvious scoring method, which is to count the number of measured peptide masses that correspond to calculated peptide masses in the theoretical peptide mass fingerprint of each protein in the database (Pappin *et al.*, 1993). This is referred to as the MOlecular Weight SEarch (MOWSE) score. The greater the number of experimental peptides matching the theoretical data, the higher the likelihood of the protein in the database being the protein excised from the gel. This technique was shown to work well for high-quality experimental data. However, it has the disadvantage that it usually gives higher scores to larger proteins as these have a larger number of proteolytic peptides and so the probability of random matching is higher. This is referred to as the protein size effect.

3.2.4.2 Second generation algorithm: the MS-Fit search engine

The MOWSE score implemented in the MS-Fit search engine (http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm) is based on the simple scoring method described above. However, the MOWSE score has been improved to compensate for the

protein size effect and also to take into account the relative abundance of the peptides in the database when calculating the score, that is, the probability of getting a random match to a larger peptide is lower and therefore a match will contribute to a higher degree to the score (Fenyö, 2000). From the literature, MS-Fit is the search engine most often employed for protein identification using PMF data, however, the scoring system employed by MS-Fit is un-reliable due to a poor scoring system and so many authors have developed their own criteria. The criteria selected for this study for the unambiguous identification of *T. brucei* protein spots using the MS-Fit search engine were as follows:

a) At least five peptides from the measured peptide mass fingerprint must match those of the theoretical peptide mass fingerprint of the protein in the database
b) The size of the peptides masses matching the theoretical peptide mass fingerprint should cover at least 15% of the protein
c) The accuracy of the matching peptides should be equal to or less than 50 ppm (differences between the measured peptide and the theoretical peptide should be lower than 50 ppm) (Mann *et al.*, 2001).

These three criteria were applied to the MS-Fit search results and used as an alternative to the MS-Fit scoring system to determine protein identification. They were considered to provide high stringency thresholds limits for unambiguous protein identification.

3.2.4.3 Third generation algorithms: the MASCOT® and the ProFound search engines

The MASCOT® search engine

The MASCOT® search engine (http://www.matrixscience.com) is based on a modified MOWSE algorithm using a probability based scoring system. The probability that the observed match between the experimental data and a protein sequence is random is

calculated for each protein sequence in the database. The proteins are classified with increasing probability of being a random match to the experimental data. In addition, the probability that a match would be significant is calculated by the search engine. This probability is based on the size of the database and the frequency of each peptide mass in the database. The MASCOT® search engine guidelines state that if the probability based score is higher than the database's probability (p) then the match is significant within 95% confidence limits. The scores are expressed as $-10*\log(P)$, where P is the probability that the observed match is a random event (Perkins *et al.*, 1999). The results are displayed as a histogram showing the distribution of the scores and the 95% confidence limit (p).

The ProFound search engine

The ProFound (http://129.85.19.192/profound_bin/WebProFound.exe) search engine also contains a probability based scoring system. The probability of the match being the best match is calculated based on Bayesian theory (Tang *et al.*, 2000). This probability takes into account available information about each individual protein sequence in the database (e.g. molecular weight, pI) and allows for the incorporation of additional experimental information (e.g. amino acid composition or sequence information). In addition, empirical information about patterns observed for the distribution of proteolytic peptides along the protein sequence is included in the algorithm (Tang *et al.*, 2000).

Database searches by the ProFound search engine can be divided into four steps:

1) assignment of mono-isotopic masses to the raw data
2) peptide mass search against the database
3) Bayesian probability calculation (P value)
4) significance testing of the results (Z score)

Bayesian probability increases with increasing number of peptide matches, increasing mass accuracy and decreasing number of digested fragments. The Bayesian probability gives a

measure of confidence that the protein selected by the algorithm from the database is the sample protein (Zang *et al.*, 1995).

The ProFound search engine also incorporates an algorithm that estimates the significance of the match (Z score) for each candidate protein. Significance testing is achieved through calculation of a score frequency function for random matches. This is derived from statistical analysis of the database being searched using random selections of peptide masses from different proteins that are then grouped as synthetic proteins and used in a peptide mass search of the database. Therefore the Z score gives an indication of how significantly different the identified protein is compared to a random mass match in the database (Ericksson *et al.*, 2000). A Z score ≥ 1.65 indicates that the candidate protein in the database is unlikely (within 95% confidence) to be a random match (Zang *et al.*, 2000). Based on the properties of this search engine, the protein with the highest P value and Z score is identified as the most likely candidate protein.

3.3 Results

Peptide mass fingerprints of *T. brucei* proteins separated by 2DE were generated by MALDI-TOF MS. These were analysed using three search engines: MS-Fit (ProteinProspector), MASCOT® (MatrixScience) and ProFound (Rockefeller). The three search engines were used to compare experimental peptide mass fingerprints against theoretical peptide mass fingerprints generated from all available protein sequences in the NCBInr database. Search parameters were selected to accommodate the three search engines and the same parameters were used throughout the searches (Table 3.2). The data entry form for each search engine were very similar. The entry form of the MASCOT® search engine is given as an example (Figure 3.1). Note that all search engines for protein

Parameter	Selected option
Database	NCBInr
Enzyme	Trypsin
Protein mass range	0 to 3000 KDa
Tolerance	50 ppm
Missed cleavage	1
Fixed modification	Carboxymethylation
Variable modification	Oxidation of methionine
Charge state	MH^+

Table 3.2: **Parameters selected for database searches using MS data**. These parameters were used with MS-Fit, MASCOT® and ProFound search engines.

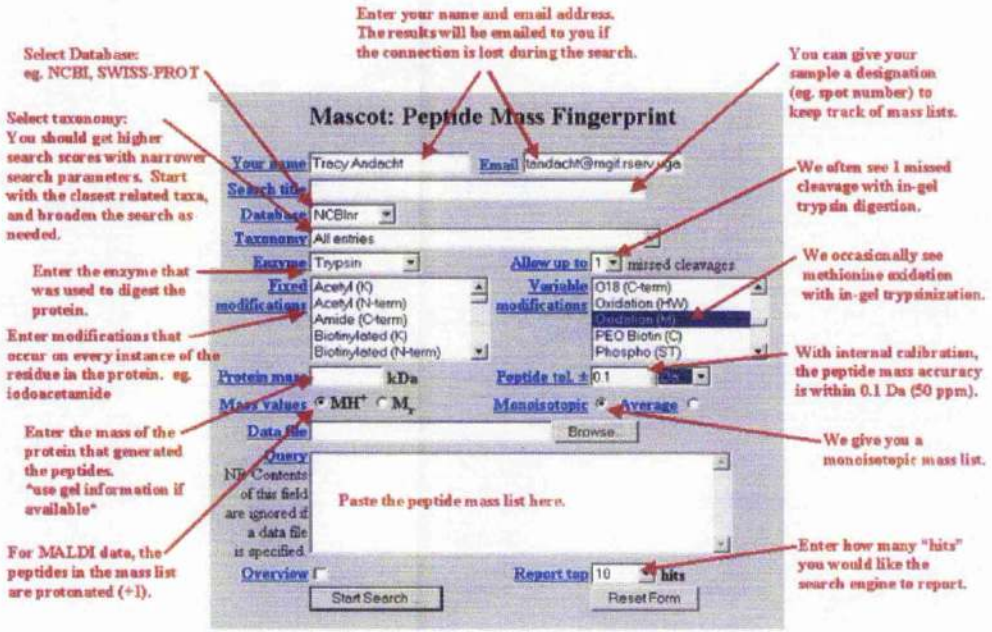


Figure 3.1: **Web page of the MASCOT® search engine.**
Reproduced from http://www.matrixscience.com/cgi/searchform.pl?SEARCH=PMF

identification using PMF data will give a "best match" candidate even if all the matching peptides are random matches. It is essential to distinguish between a significant match and a best match. This can be achieved by evaluating the specificity of the algorithms used in the search engines.

3.3.1 Searching the NCBInr database with MS-Fit, MASCOT® and ProFound search engines

126 peptides mass fingerprints were generated by MADLI-TOF MS from *T. brucei* proteins separated by 2DE. These were searched against the NCBInr database using three search engines: MS-Fit, MASCOT® and ProFound.

3.3.1.1 The MS-Fit search engine

In total, 37 protein spots were identified using the three criteria described in Materials and Methods. These proteins had a corresponding MOWSE score ranging from 10^{+3} to 10^{+10} (Table 3.3). However, 13 additional protein spots with a MOWSE score higher than 10^{+3} did not reach the threshold limits fixed by the three criteria for unambiguous identification, i.e. percentage of coverage < 15% (data not shown). This confirmed that the MOWSE score implemented in MS-Fit could not be used on its own with confidence for unambiguous protein identification by PMF. In total, 30% of the peptide mass fingerprints searched using the MS-Fit search engine (37 out of 126) reached the threshold limits set by the three criteria and were considered to be identified.

3.3.1.2 The MASCOT® search engine

The MASCOT® search engine's guidelines state that a protein score above the database score (p) would give an unambiguous identification within 95% confidence limits. This guideline was applied to the identification of the proteins from *T. brucei* and enabled the

		MS-Fit			ProFound					MASCOT			
Sample	MOWSE	Coverage	Peptide	Modification	Probability	Z score	Coverage	Peptide	Modification	Score	Coverage	Peptide	Modification
86	1.70E+07	27%	11	1	1.00E+00	2.35	27%	11	1	139	27%	11	1
87	2.96 E+9	25%	11	1	1.00E+00	2.11	24%	11	2	56	29%	13	2
89	1.75E+06	17%	9	1	9.90E-01	1.26	15%	9	1				
93	1.98 E+8	34%	10	1	1.00E+00	2.31	32%	11	1	74	33%	11	2
97	1.62 E+5	19%	9	1	9.50E-01	0.65	15%	9	1	65	17%	7	1
105	6.56E+07	16%	10	1	1.00E+00	2.09	17%	10	1	58	11%	8	1
109	1.61E+04	19%	5	none	3.30E-01	0.29	19%	5	none	46	19%	5	none
114	3.34E+08	42%	11	2	1.00E+00	2.39	42%	11	2	155	42%	12	2
115	4.78 E+3				8.50E-01	0.42	13%	6	2	48	13%	6	2
118	9.72 E+4	25%	8	none	1.00E+00	2.06	23%	8	none	68	23%	7	none
121	2.2 E+5	15%	6	1	1.00E+00	1.21	15%	6	1	42	15%	6	1
127	4.03 E+10	42%	12	2	1.00E+00	2.4	43%	12	2	153	42%	12	2
129	3.40E+06	26%	8	1	1.00E+00	2.18	27%	8	1	86	26%	8	1
131	4.19 E+6	26%	9	1	1.00E+00	1.55	23%	9	1	58	21%	7	1
133	7.21 E+6	26%	8	none	1.00E+00	1.95	22%	8	none	93	24%	8	none
136	2.18 E+4	22%	7	2	1.00E+00	1.79	17%	7	2	75	22%	7	2
140	4.97 E+3	27%	7	1	1.00E+00	1.79	23%	7	1	60	32%	6	none
145	3.52 E+4	23%	8	1	1.00E+00	1.73	21%	8	1	50	18%	6	1
148	1.84 E+5	21%	7	none	1.00E+00	1.69	19%	7	none	58	18%	6	none
155	5.7 E+4	36%	6	1	1.00E+00	1.85	37%	6	1	76	36%	6	1
158	3.90E+06	15%	9	none	1.00E+00	2.07	16%	9	none	91	15%	9	none
163	3.02 E+7	17%	9	1	1.00E+00	2.05	17%	9	1	73	17%	9	1
165					3.20E-04	0.41	10%	12	7				
168	2.85 E+6	16%	9	none	1.00E+00	1.97	16%	9	none	81	16%	9	none
172	5.17 E+3	36%	5	none	3.30E-02	0.05	27%	4	none	45	27%	4	none
173	9.16 E+8	27%	10	1	1.00E+00	2.28	27%	10	1	78	24%	9	1
175	2.83 E+3	20%	6	none	7.10E-01	0.36	16%	7	none	52	21%	6	none
176	2.64 E+3	16%	6	none	6.20E-02	0.16	15%	6	none	72	19%	6	none

179	1.41 E+4	20%	6	none		9.70E-01	0.64	20%	6	none	71	19%	6	none
180	2.92 E+3	17%	5	none		7.70E-02	0.05	16%	5	none	50	15%	5	none
181	3.59 E+4	17%	5	none		6.00E-04	0.93	20%	5	none	53	20%	5	none
187	8.38 E+5	29%	7	none		1.00E+00	1.62	29%	7	none	79	29%	7	none
193	267	16%	5		2									
194	9.99 E+3	18%	5	none		3.10E-03	0.46	18%	5	none	44	18%	5	none
195	4.02 E+4	19%	5	none		5.10E-01	0.23	22%	5	none	47	22%	5	none
201	2.3 E+4	20%	5	none		1.80E-03	0.06	17%	4	none	45	20%	5	none
203	3.67 E+9	33%	12		1	1.00E+00	2.37	31%	12	1	113	30%	11	1
204	8.31 E+7	25%	8		1	9.90E-01	0.84	21%	7	1	65	25%	8	1
209	2.41 E+8	40%	10	none		1.00E+00	2.37	41%	10	none	131	40%	10	none

Table 3.3 Results from the MS-Fit, MASCOT® and Profound search engines. The samples highlighted in yellow correspond to the spots identified by the ProFound search engine using its probability based scoring system. The samples highlighted in blue correspond to the spots identified within 95% confidence limits using the MASCOT® search engine.

identification of 16 protein spots, corresponding to 12.7% of the total peptide mass fingerprints analysed (Table 3.3).

3.3.1.3 The ProFound search engine

The ProFound search engine is implemented with a probability scoring system (P) and a significance testing system (Z). The threshold values given by the ProFound search engine for identification are a P value of 1 and a Z score above 1.65. These threshold values corresponded to an identification within the 95% confidence limits. These threshold values were used for the identification of *T. brucei* protein spots separated by 2DE and enabled the identification of 19 spots (Table 3.3). This represented 15% of the total protein spots analysed.

3.3.2 Comparison of the results obtained from the three search engines

An analysis of the specificity of algorithms implemented in the search engines can be made by submitting the same search to multiple search engines and comparing the results. This analysis was undertaken with the three search engines using peptide mass fingerprints generated from *T. brucei* protein spots separated by 2DE.

As the MS-Fit search engine had no reliable scoring system implemented in the search engine, the three criteria employed as threshold for unambiguous protein identification for the MS-Fit search engine were applied to the MASCOT© and the ProFound search engines. The best score obtained for the peptide mass fingerprints reaching the thresholds limits set by the three criteria were compared for each search engine. Using the three criteria, 29.4% of the peptide mass fingerprints analysed (37 out of 126) were identified using the MS-Fit search engine, but only 27.7% (35 out of 126) were identified using the MASCOT® search

engine and 21.4% (35 out of 126) were identified using the ProFound search engine (Table 3.4). A comparison of the three criteria used for identification showed that the differences in the results between the search engines were due to the percentage of coverage of the protein in the database (Table 3.3). The percentage was calculated directly by the search engines. For the same protein, the percentage of coverage varied from 1 to 5% between search engines. As a threshold of 15% coverage was needed for unambiguous identification, any difference in coverage calculated by the search engines meant that a protein accepted as unambiguously identified by one search engine was not identified with another. This demonstrated how critical the threshold limits set by the user became when there is no probability based scoring system used in the search engine. All search engines agreed on the name and NCBI accession number (gi number) of the best score for each peptide mass fingerprint searched, demonstrating that the three search engines had similar sensitivity.

When the probability based scoring system of the MASCOT® and the ProFound search engines was used to determine unambiguous identification only 15.8% (20 protein spots) and 12.7% (16 protein spots) were unambiguously identified respectively. The probability based search engines had a higher specificity than the three criteria used with the MS-Fit search engine.

	MS-Fit	ProFound	MASCOT®
Number of protein identified based on the scoring system of the search engine	N/A	20	16
Number of protein identified using the criteria of mass accuracy, number of matching peptides and coverage of the protein	37	35	35

Table 3.4: Summary of comparison of the results from the MS-Fit, MASCOT® and ProFound search engines. Results were obtained searching the NCBInr database with the 126 peptide mass fingerprints. The number of proteins identified by the three search engines using the scoring system and the three criteria described in the text are shown. N/A non-applicable.

3.4 Discussion

The sensitivity and specificity of the search engine together with human judgment are the three key factors in the identification of proteins by mass spectrometry. The sensitivity of a search engine relates to its ability to detect a match in a given database while the specificity relates to its ability to determine whether the match is significant (random match versus real match). Three search engines were compared for their sensitivity and specificity in protein identification by PMF. The MS-Fit search engine was one of the first developed to search databases with peptide mass fingerprints. A high score with the MS-Fit search engine (10^{-3}) did not always correspond to an unambiguous identification with the MASCOT® and the ProFound search engines. The low specificity of the MS-Fit scoring system means that the user needs to rely on judgment to accept or reject the best score as reflecting an unambiguous protein identification.

The MS-Fit search engine has been the favoured search engine for protein identification by PMF in many proteomic studies. The criteria used as threshold limits for protein spot identification employed varied remarkably, from the simplest criteria of five peptides matching to the protein in the database (McAtee *et al.*, 2001), through a minimum of 30% coverage (Jungblut *et al.*, 1999) up to a combination of criteria such as four peptides matching with an accuracy of at least 40 ppm (Smolka *et al.*, 2003). Basing the identification on the percentage of protein covered by the measured peptides biases the identification process towards small proteins, whereas an identification process based solely on five peptides matching a protein biases the identification towards large proteins.

The selection of criteria for unambiguous identification of proteins can be difficult. The accuracy obtained in the measurement of peptide masses strongly influences the specificity

of the search. Investigation of the role of accurate mass measurement in protein identification strategies demonstrated that random matches could occur for proteins matching as many as 15/23 peptides while using a range of 1,000 to 2,000 ppm (Clauser *et al.*, 1999). When dropping the mass accuracies in the range of 100 to 500 ppm, the level of random matches appeared to fall to 7/23 or 8/23 peptides respectively. Once the accuracy level reaches 50 ppm, currently possible with MALDI-TOF in reflector mode using internal calibration, the level of random matches dropped to 6/23 peptides (Clauser *et al.*, 1999). Recent calculations for proteins expressed by the genomes of *Escherichia coli* and *Saccharomyces cerevisia* indicate that at 0.1 ppm mass accuracy, 96% of the proteins will generate tryptic peptides with one or more unique masses (Figure 3.2). The remaining 4% correspond to gene products that are largely duplicates of other proteins (Bruce *et al.*, 1999). However, this level of mass accuracy is not achievable with MALDI-TOF MS. Consequently, a combination of criteria including the accuracy of the peptide match (50 ppm), the number of tryptic peptides matching (5) and in addition a limit of the coverage obtained from the proteins (15%) were used as threshold limits with the MS-Fit search engine for unambiguous identification of proteins (Mann *et al.*, 2001). However, even these stringent criteria were shown to be too relaxed compared to the probability-based scoring systems of the ProFound and the MASCOT® search engines. This demonstrated the importance of the implementation of a probability-based scoring system, removing the "human" factor for protein identification. The advantage of a probability based scoring is that a simple rule can be used to judge whether a result is significant or not. The addition of a scoring system determining the probability of the best match being a significant match is essential. It should be kept in mind that the significance testing can never definitely prove whether a result is true or false. The significance level is just the calculated risk of obtaining a false result.

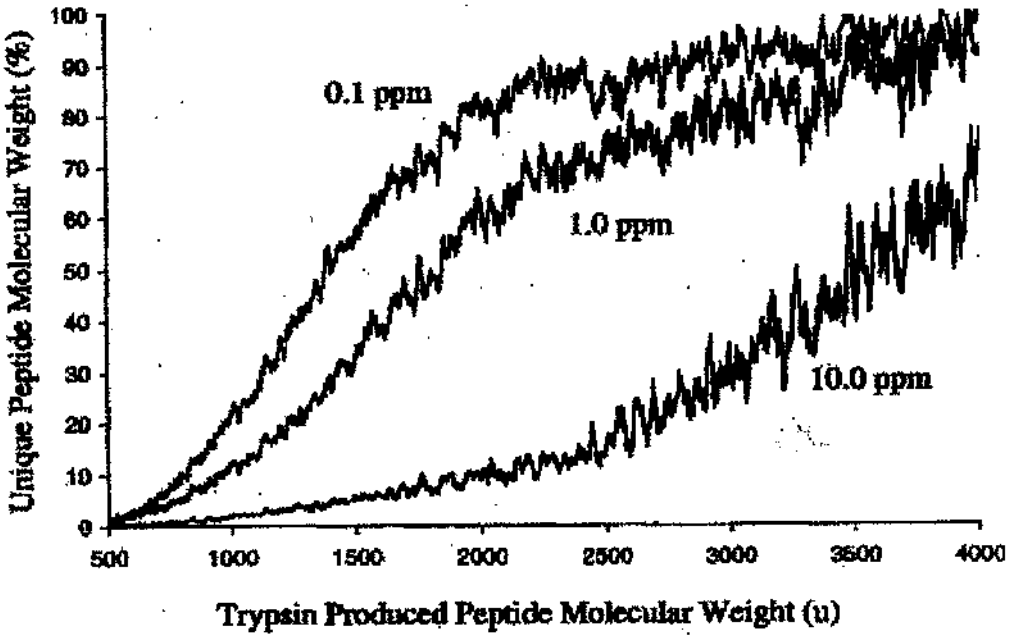


Figure 3.2: Importance of mass accuracy in protein identification. Graphical representation of all possible unique peptide molecular weights, after trypsin digestion, of all yeast proteins in the NCBI at a mass accuracy of 0.1, 1.0 and 10 ppm (reproduced from Goodlett *et al.*, 2000).

All search engines are based on the assumption that the protein being analysed is present in the database. However, this assumption is not always true. Consequently, it should be kept in mind that peptide mass fingerprint can only provide the statistically most probable identification, taking into account the state of the database at the time of the search. Even with the most sensitive and selective search algorithm, if the protein sequence is not present in the database then the algorithm's best score will at best be a closely related protein or a random match.

A comparison of the algorithm used in MS-Fit, MASCOT® and ProFound search engines, performed *in silico*, on four proteins sequences taken from NCBI database demonstrated that the algorithm employed by the ProFound was more specific than that used by the MASCOT® which in turn was more specific than the MS-Fit search engine (Tang *et al.*, 2000). The specificity of the ProFound search engine was also shown in this paper to be more consistent over a wider range of protein molecular weight (50 to 200 kDa) than the MASCOT® search engine which was only reliable up to 100 kDa. Trypsin digestion of protein spots is expected to give rise to peptide fragments of between 800 and 2,500 amino acids, depending on the distribution of the recognition sequence for trypsin in proteins (Aebersold and Goodlett, 2001). Consequently the limitation of the specificity of the MASCOT® search engine for large masses should not affect its efficiency for protein identification by PMF.

The MASCOT® search engine was shown to be both sensitive and specific for the identification of proteins from *T. brucei*. In addition the search engine is simple to use and the output of the results easy to interpret. The MASCOT® search engine can also be used for searching databases with Tandem mass spectrometry data. As MALDI-TOF and

Tandem mass spectrometry will both be used in the proteomic study of the mechanism of Cymelarsan resistance in *T. brucei* the MASCOT® search engine was selected for the rest of the proteomic analysis. Other limitations, independent of the search engines (i.e. database, post-translational modifications) in the identification of protein spots by mass spectrometry will be discussed in the next chapter.

Chapter IV
A proteome map of *T. brucei*

4.1 Introduction

The first step in proteomic analysis is the extraction and separation of proteins from trypanosomes by 2DE (chapter 2). Individually separated proteins are excised from the gel and digested with a sequence specific protease (e.g. trypsin). The masses of the proteolytic peptides are measured by mass spectrometry. A search engine is then used to search a database with the data obtained by mass spectrometry (refer to chapter 3). Many types of search engines are available. A comparison of algorithms implemented in three search engines (MS-Fit, ProFound and MASCOT®) demonstrated the sensitivity and selectivity of the MASCOT® search engine (chapter 3). The MASCOT® search engine was purchased from MatrixScience and adapted to search the NCBInr database and a local *T. brucei* database.

Proteomic expression profiling of protozoan parasites was rare, until recently, mainly due to the lack of genome sequences for parasites in general. The limited availability of genome sequence meant that proteins of interest separated by 2DE might not be represented in the database and consequently the protein spots could not be identified using mass spectrometry data. However, the situation is rapidly changing with the publication of the *Plasmodium falciparum* genome sequence (Gardner *et al.*, 2002) and full genome sequencing projects are underway for *T. brucei*, *T. vivax*, *T. congolense*, *Leishmania major*, *Leishmania infantum*, *Toxoplasma gondii*, *Babesia bovis*, *Theileria parva*,

Theileria annulata and a further six plasmodium species (http://www.sanger.ac.uk/Projects/Protozoa and http://wwwtigr.org/tdb/parasites/).

The haploid genome size of *T. brucei* is about $3.5x10^7$ base pairs with 11 homologous chromosome pairs and additional intermediate and mini-chromosomes of uncertain ploidy (El-Sayed *et al.*, 2000). The 11 megabase chromosomes are thought to contain most, perhaps all, the housekeeping genes and so were targeted by the Sanger and the TIGR centres for the sequencing of the *T. brucei* genome, using the 927 stock. The TIGR centre was allocated the sequencing of chromosomes II to VIII while the Sanger centre was in charge of the sequencing of chromosomes I, IX, X and XI. The methods employed by each centre for the sequencing of the chromosome are different. At the Sanger centre, chromosomes were separated by pulsed field gel electrophoresis, excised from the gel and sheared into 2-4 kb fragments which were inserted into a plasmid vector. The clones were then randomly end-sequenced, producing a 8-10x coverage of the chromosome. This technique is called whole chromosome shotgun sequencing. The TIGR centre produced a large number of BAC clones (bacterial artificial chromosome vector pBACe3.6, average insert size of 145 kb) and end-sequenced them. Chromosome specific BACs were selected using cDNA markers specific to the target chromosome, subcloned into plasmid and then sequenced to completion. Then the database of end-sequences was then searched to retrieve BAC clones that overlap by only 5 to 10 kb, and these are then sequence to completion. To date, the sequences of two of the 11 chromosomes have been completed and annotated (data are available on the Internet at http://www.tigr.org/tdb/mdb/tbdb/ and http://www.sanger.ac.uk/Projects/T_brucei/) (Table 4.1).

Sequencing centre	Chromosome	Size (Mb)	Status
Sanger	I	1.15/1.2	Complete
TIGR	II	1.25/1.3	Complete
TIGR	III	1.62/1.75	In progress
TIGR	IV	1.62/1.75	In progress
TIGR	V	1.68/1.75	In progress
TIGR	VI	1.7/2.0	In progress
TIGR	VII	2.1/2.1	In progress
TIGR	VIII	2.1/2.2	In progress
Sanger	IX	3.3/3.5	In annotation
Sanger	X	4.4/4.4	Finishing
Sanger	XI	5.2/5.2	In shotgun

Table 4.1: Genome sequencing project for *T. brucei*. Status of the genome sequencing project for *T. brucei* TREU 927 line, dated 26th June 2003 (reproduced from http//:www.sanger.ac.uk/Projects/T_brucei/progress.shtml). The local *T. brucei* database was made of translated sequences from the available nucleotide sequences from the 11 chromosomes. The size of the chromosomes was estimated by pulsed field gel electrophoresis.

In addition, expressed sequence tags (ESTs) made from the sequencing of randomly selected complementary DNA (cDNA) are also available from EST discovery projects. There are over 5,000 ESTs from *T. brucei* deposited in Genbank, each with a unique accession number, which represent 3,113 unique sequences (Djikeng *et al.*,1998; El Sayed *et al.*, 1995). This sequence information, provided both by the sequencing centres and the EST projects, is essential for the identification of proteins extracted from *T. brucei* and analysed by proteomic techniques.

It is estimated that the trypanosome genome contains approximately 8,000 genes based on sequence from chromosome I and II (Hall *et al.*, 2003; El-Sayed *et al.*, 2003), and while numerous post-translational modifications are known to exist in this organism, the genes in general have no introns, with the exception of the PAP gene (Mair et al., 2000). In addition, RNA editing in trypanosomes has been shown to be extensive for mitochondrial mRNAs which are encoded maxicircles (Madison-Antenucci *et al.*, 2002) but almost non-existent, except for the addition of the spliced leader sequence at the 5'end by trans-splicing, for mRNAs encoded from the 11 chromosomes. Thus unlike in higher eukaryotes, the proteome of trypanosome should not contain a drastically higher number of proteins than genes.

While the current data support the view that intron and alternative splicing do not occur in *T. brucei* and the main factor generating multiple protein product from single genes is likely to be post-translational modification, this has yet to be experimentally tested. In order to test the limitations of the incomplete genome sequence with regard to full proteomic analysis of *T. brucei*, a small scale proteome map was constructed. In addition,

the question of the level of post-translational modification of proteins and the relationship between polypeptide profile and gene number were also addressed.

4.2 Materials and Methods

4.2.1 Protein extraction, 2DE and staining

Whole soluble proteins were extracted from $2x10^8$ procyclic forms of the reference line 927 using protocol P2 as described in chapter 2. Protein extracts were separated by 2DE poly-acrylamide gels using broad pH ranges (pH 4-7 and 6-11) and stained with Coomassie Brilliant blue as described in chapter 2.

4.2.2 In-gel trypsin digestion and mass spectrometry

Protein spots were excised from the gel, digested with trypsin and PMF were acquired using the MALDI-TOF MS as described in chapter 3.

4.2.3 Database searches using the MASCOT® search engine

The "all taxa" category of the NCBInr database and a local *T. brucei* database were searched using the MASCOT® search engine. The parameters selected for the searches are described in chapter 3 (Table 3.2).

4.2.3.1 The NCBInr database

This database contains the translated protein sequences from the entire collection of annotated DNA sequences kept at GenBank, and also protein sequences in the Protein Data Bank (PDB), SWISS-PROT and Protein Information Resource (PIR) databases, covering most of the publicly available data.

4.2.3.2 The local *T. brucei* databases

The local *T. brucei* database was constructed by downloading all available sequence information from ftp sites of the Sanger and the TIGR centres. Two databases were made: a nucleotide sequence database (Pawel_nt) and a peptide sequence database (Pawel_aa). Pawel_nt database consisted of a conglomerate of three nucleotide databases:

1- A genomic database from the TIGR centre containing 60 very long contigs from chromosomes II-IX. These were arbitrarily cut into 222 fragments of manageable sizes for the MASCOT® search engine, comprising fragments of up to 50 kb in length with 600 bp overlap between fragments.

2- Sanger nucleotide genomic data from chromosomes I, IX, X and XI. Chromosome I consisted of one contig which was cut into 22 fragments, chromosome IX comprised 257 contigs and were cut into 317 fragments, chromosome X contained 427 contigs which were cut into 485 fragments and chromosome XI consisted of 896 contigs.

3- *T. brucei* GSS/EST clusters from the Sanger centre containing 20,493 sequences.

Pawel_aa corresponds to the predicted putative genes from genomic data described in 1 and 2. This corresponds to 6,920 sequences from chromosome II-IX from TIGR, 509 sequences from chromosome I, 2,348 sequences from chromosome IX, 10,590 sequences from chromosome X and 2,568 sequences from chromosome XI from the Sanger centre. In total, Pawel_aa contained 22,935 sequences as updated on the 21/06/2003. These databases were downloaded and adapted for searches with the MASCOT® search engine by Dr Pawel Herzyk (University of Glasgow).

4.3 Results

4.3.1 Efficiency of peptide mass fingerprinting

In order to produce a proteomic map of the procyclic form of the reference line 927 whole soluble cell extracts were separated on pH 4-7 and pH 6-11 2D gels (Figure 4.1). A total of 203 protein spots were detected on pH 4-7 gel and 74 protein spots on pH 6-11 gel. From these 2D gels, 126 protein spots were excised (91 protein spots from pH 4-7 gel and 35 protein spots from pH 6-11 gel), in-gel digested with trypsin and peptide mass fingerprints were acquired using MALDI-TOF. The peptide mass fingerprints were searched against the NCBInr and the local *T. brucei* database using the MASCOT® search engine, enabling the identification of 51 protein spots. The list of peptide masses obtained for each protein spot and used to search the databases is given in Appendix I.

In total, 40.5% of the peptide mass fingerprints obtained using MALDI-TOF (51 out of 126 protein spots) were matched to a protein in either the NCBInr database or the local *T. brucei* databases (Table 4.2). The criteria for identification with the MASCOT® search engine were that the protein score had to be higher than the database score (p). From the 51 protein spots identified, 17 peptide mass fingerprints matched onto a *T. brucei* entry in the NCBInr database. In addition, 24 peptide mass fingerprints matched only to proteins in the local *T. brucei* database. Most of the proteins in the local *T. brucei* database are not annotated so a BLASTp search was performed against the NCBInr database to identify homologous sequences. From the 24 proteins identified in the local database, seven sequences had no homologues in the NCBInr database; these are referred to as hypothetical proteins in Table 4.2. Detailed results from the MASCOT® search engine for each protein spot analysed by PMF are given in Appendix II.

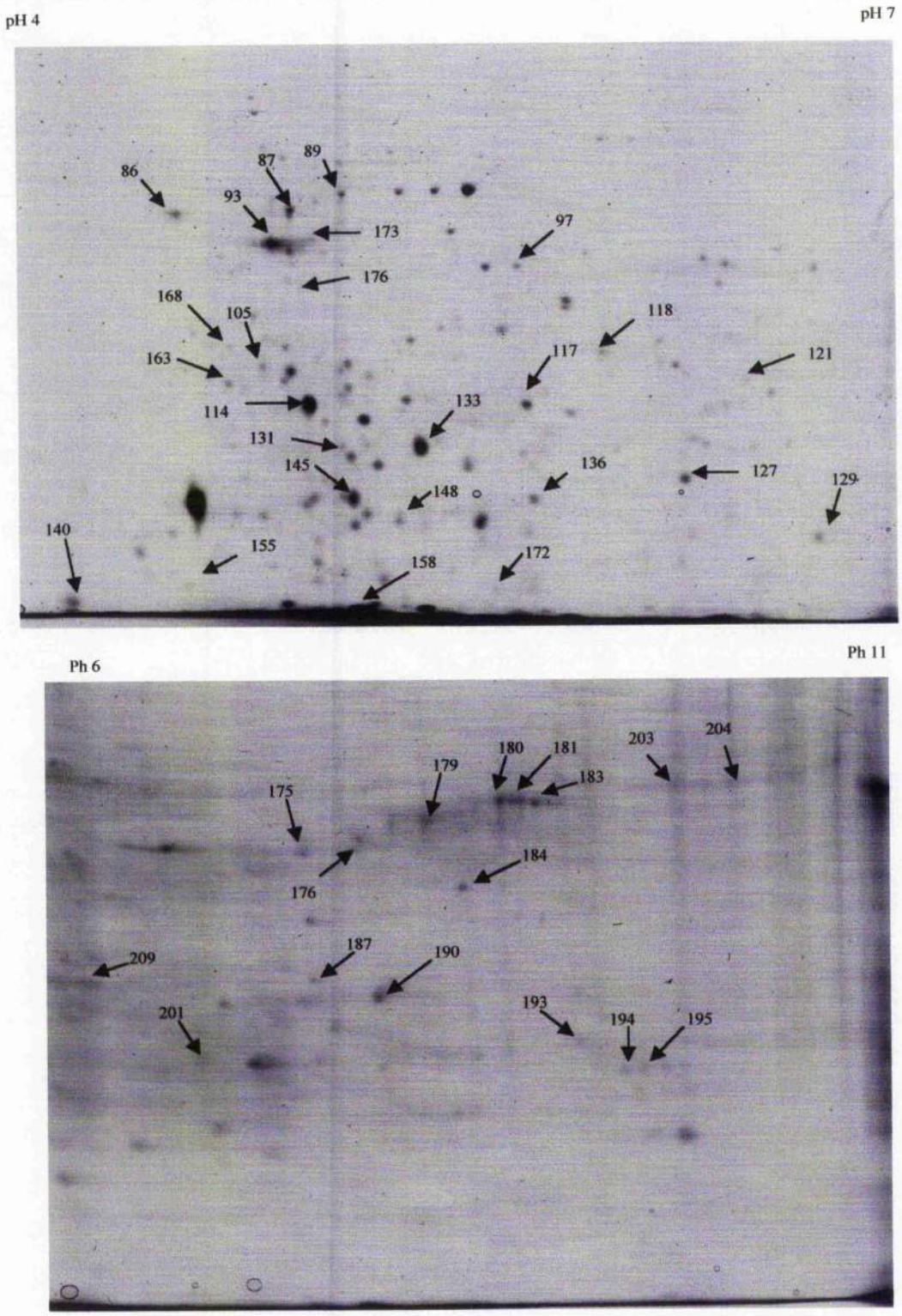


Figure 4.1: **Proteome map of *T. brucei*.** Coomassie Brilliant blue stained gel of whole cell extract from *T. brucei* line 927 separated by 2DE on pH 4-7 and pH 6-11. The proteins from which peptide mass fingerprints were generated and an identification was obtained by searching the NCBInr database and the local *T. brucei* database are indicated by an arrow.

Sample	Protein	ppm	Score	P	Coverage	Peptide
86	beta tubulin	25	139	72	27%	11
89	Blast HSP 60	25	69	56	17%	9
90	Blast: HSP 70	17	112	56	21%	11
92	Blast: HSP 70	25	102	56	16%	10
93	Tubulin alpha chain	34	74	72	33%	11
97	Paraflagellar rod protein	27	65	72	17%	7
101	Blast: dihydrolipoamide S acetyltransferase	20	76	55	19%	7
104	Hypothetical	30	56	55	34%	5
105	Blast: HSP70	21	79	56	14%	9
112	Putative phospatidylinositol 4 phosphate 5 kinase	20	63	50	19%	6
114	beta tubulin	16	155	72	42%	12
115	Elongation factor 2	30	90	56	17%	10
117	Blast: Elongation factor 2	20	120	56	15%	12
118	Blast: alpha tubulin	15	73	56	23%	7
123	Blast: Elongation factor 2	30	66	55	10%	9
127	guanine nucleotide binding	17	153	72	42%	12
129	pyridoxine/pyridoxal kinase	14	86	72	26%	8
131	Blast: cytochrome C oxidase subunit IV	20	67	56	21%	
132	HSP83	18	66	56	24%	7
133	Tubulin alpha chain	22	93	72	24%	8
136	Enolase	22	75	72	22%	7
140	beta tubulin	11	59	56	18%	6
145	Alpha tubulin	25	65	56	20%	7
153	Blast: elongation 1 beta	22	76	56	28%	6
154	Blast: elongation 1 beta	21	64	56	31%	5
155	Proteasome subunit	24	76	72	36%	6
158	Heat shock 70KD protein 4	26	91	72	15%	9
163	Heat shock 70KD protein 4	11	73	72	17%	9
168	BIP/GRP78	22	81	72	16%	9
170	Hypothetical	30	71	55	36%	6
172	Adenylate kinase	24	105	56	39%	8
173	ATPase beta subunit	19	78	72	24%	9
174	Putative DnaJ protein	30	111	50	35%	10
175	Blast: translation elongation factor 1 alpha	19	58	56	11%	5
176	Elongation factor 1 alpha	11	72	72	19%	6
179	Blast: translation elongation factor 1 alpha	22	67	56	19%	6
180	Hypothetical	25	67	56	22%	6
181	Blast: translation elongation factor 1 alpha	24	57	56	17%	5
182	Hypothetical	20	55	55	19%	5
183	Hypothetical	30	60	55	20%	6
184	Hypothetical	30	123	50	22%	11
185	Putative sucinyl CoA synthetase	25	56	50	25%	5
187	Malate dehydrogenase	25	79	72	29%	7
189	Hypothetical	20	60	55	14%	7
190	Blast: Malate dehydrogenase	50	100	55	29%	7
191	Blast: Malate dehydrogenase	30	97	55	34%	7
195	Glyceraldehyde 3 phosphate dehydrogenase	14	58	56	21%	6
203	Tubulin alpha chain	24	113	72	30%	11
206	Blast: pyruvate dehydrogenase	50	76	55	48%	5
209	guanine nucleotide binding	20	131	72	40%	10

Table 4.2: List of protein spot identified by the MASCOT® search engine. Mass fingerprints were obtained by MALDI-TOF MS. Peptide mass fingerprint were searched against NCBInr and the local *T. brucei* database using the MASCOT® search engine. The table indicates the accuracy of the match (ppm), the score of the protein, the database score (p), the coverage of the protein by the identified peptides, and the number of peptides from the fingerprint that match on the protein. The search was last carried out on the 25/06/2003. Non-annotated ORF from the local *T. brucei* database were identified by blast against the NCBI database to homologues in other species.

The majority of proteins separated by 2DE and visualised by Coomassie staining were not identified. From all the protein spots processed, 10.13% were identified as enzymes, followed by 7.14% identified as structural proteins, 7.14% identified as elongation factors and 9.52% classified as others and comtaining the hypothetical proteins (Figure 4.2).

4.3.2 Relationship of polypeptide profile complexity to gene number in *T. brucei*

An isoform is defined in the Dictionary of Cell Biology as a protein having the same function and similar (or identical) sequence but the product of a different gene and usually tissue specific (Lackie and Dow, 1995). For the purpose of this thesis, we will limit the definition of an isoform to a protein product of one or different genes with identical DNA sequences. Consequently, any isoform observed on a 2DE will be the results of a post-translational modification by either the covalent addition of a small adduct or by proteolytic processing.

The identification of protein spots by MALDI-TOF analysis showed a large proportion of α (5 spots) and β (3 spots) tubulins on the proteome map. So far, the genome sequencing project of *T. brucei* has identified 13 α tubulin genes and 14 β tubulin genes on one homologue and 9 α tubulin genes and 10 β tubulin genes on the other homologue of chromosome 1 (Hall *et al.*, 2003). Alignment of all the sequences of α tubulin genes and β tubulin genes found in the GeneDB database has shown that all α tubulin genes are identical and all β tubulin genes are identical to each other (data not shown). Consequently the tubulins spots identified from the proteome map are isoforms. The α and β tubulins spots identified are located at very different places on the proteome map, indicating that these isoforms are possibly the results of modification by proteolytic processing.

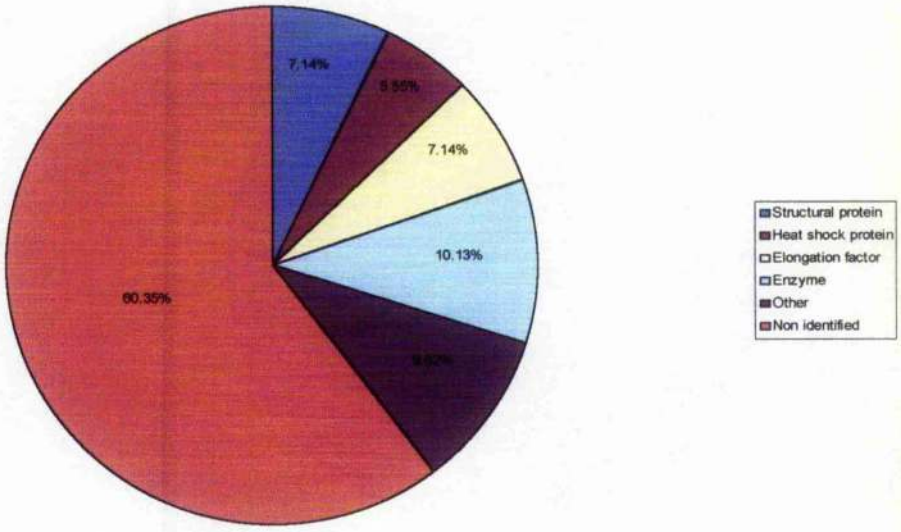


Figure 4.2: **Pie chart representation of classes of protein identified.** The class other contains the "hypothetical" proteins (4.76%) from the local *T. brucei* database for which no homologues where found in the NCBI database.

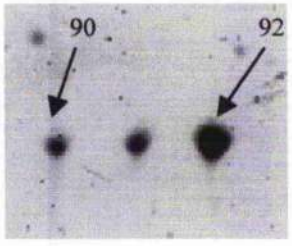


Figure 4.3: **2DE pattern of post-translationally modified proteins**. Zoom view of a part of the 2D gel (pH 4-7) showing the characteristic pattern of isoforms with similar molecular weight but slight variations in pI. Samples 90 and 92 were both identified as heat shock protein 70, transcript product of the same ORF (score: 112 and 102 respectively).

The HSP70 protein spots (spot 90 and 92) identified on the proteome map are also isoforms as they are products of the same gene. The HSP70 gene family is large but also very diverse which enabled the differentiation of the gene products by MALDI-TOF analysis, as illustrated in Figure 4.3. Protein spots 90 and 92 are located at similar apparent molecular weight but with different pI (Figure 4.4), indicating that the isoform are probably the results of post-translational modification by covalent addition of a small adduct. This suggests that the HSP70 protein is regulated at the post-transcriptional level in *T. brucei*.

In some cases however the MALDI-TOF analysis does not enable a determination of whether the protein spots identified are isoforms, or products of genes with different sequences. This was observed for spot 175, 176 and 179, all identified as elongation factor 1 α (score = 58, 72 and 67 respectively), located at similar apparent molecular weight but with differences in pI. At first glance the results from the MASCOT® search engine suggested that all three spots were transcriptional products of the same ORF (ORF 775), which had obtained the highest score. However, a closer examination of the results suggested the presence of two elongation 1 α genes (ORF 775 and ORF 774). An alignment of the transcriptional product from these two ORFs showed a difference in length between the two ORFs and 100% homology in the aligned sequences (Appendix III). The shorter ORF (ORF 775) thus obtained the best score with the MASCOT® search engine (as the percentage of coverage obtained was higher), however the mass spectrometry data obtained cannot differentiate between the two genes. Consequently, spot 175, 176 and 179 could be either post-translationally modified elongation factor 1 α or the transcripts of different genes.

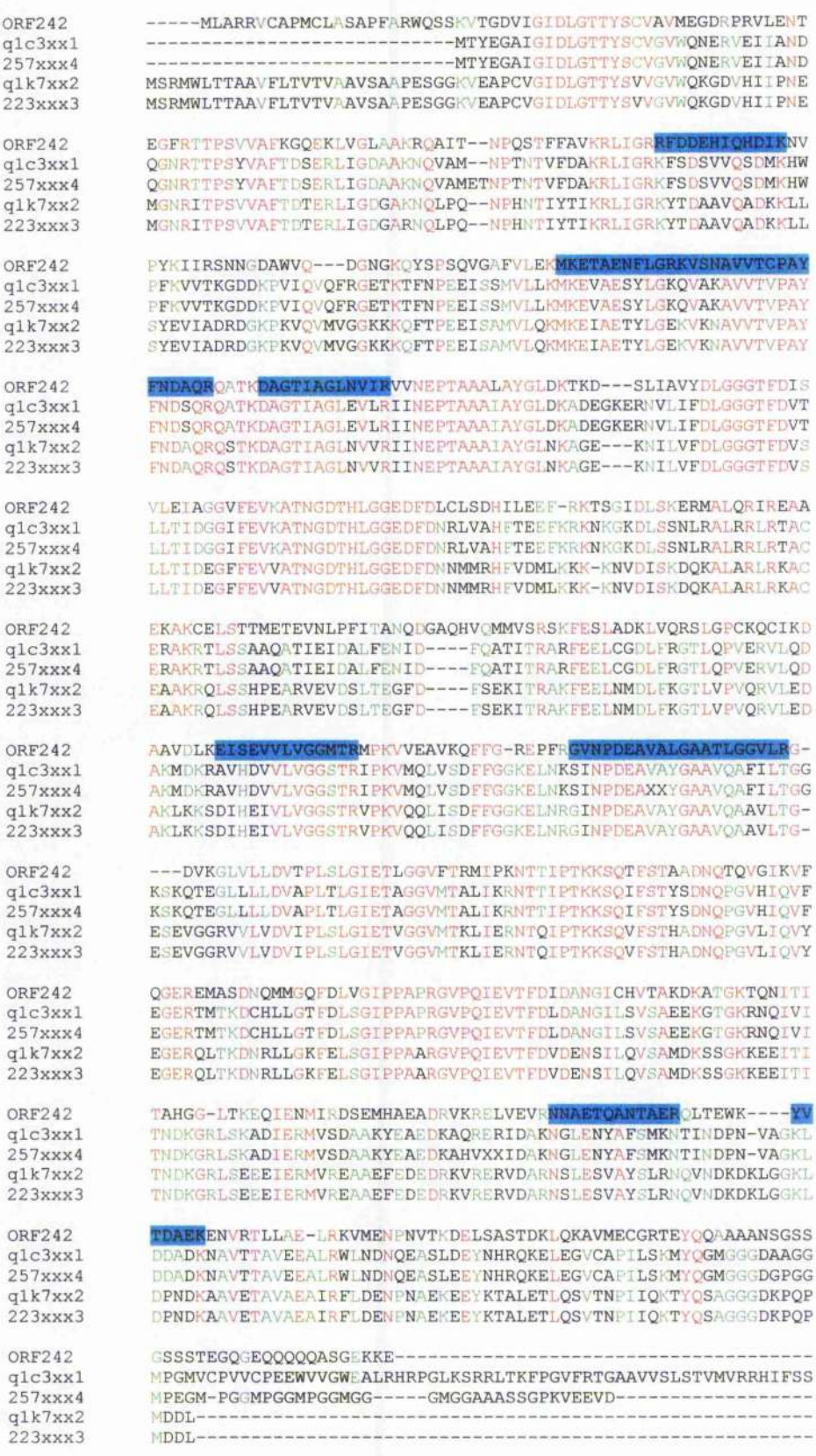


Figure 4.4: Alignment of ORFs encoding HSP70. Peptides identified by MALDI-TOF analysis of spot 90 are highlighted in blue.

4.4 Discussion

In this chapter, protein spots were digested with trypsin and 51 were identified to act as landmarks for further proteomic studies. The likelihood of obtaining identifications by peptide mass fingerprinting given the incomplete state of the *T. brucei* genome-sequencing project was determined.

There was a good correlation between spot intensity and spectrum quality, although some intense spots gave few peaks by MALDI-TOF. The proteome map revealed that 40.5% of proteins visualised by Coomassie staining can be identified using the simple peptide mass fingerprinting technique with the current state of the *T. brucei* genome database. A protein was accepted as identified only if the score was higher than the database score (p) and the accuracy under 50 ppm (chapter 3). From the 126 protein spots assessed, 10.3% were enzymes (component of the metabolic pathways), 7% were structural proteins involved in the cyto-architecture of the cell (tubulins, paraflagellar rod proteins, etc...) and 7% were elongation factors. Structural proteins are known to be highly expressed in cells, which is in accordance with the results obtained from the preliminary proteome map of *T. brucei* line 927. A significantly high proportion of identified protein spots (5.55%) were match to non-annotated, hypothetical ORFs of *T. brucei*. The identification of these genes from their protein product proves that these are real genes and that they are expressed in *T. brucei*. These ORFs were found to have no homologues in the NCBI database and so can be considered to be specific to *T. brucei*. Because of their specificity to the *T. brucei* parasite these proteins should be further analyse in view of their potential as drug targets. The sample used in this study were excised from Coomassie Brilliant blue stained gel and so there has been a unavoidable bias toward the analysis of the most highly expressed proteins.

With the available *T. brucei* database, 59% of the peptide mass fingerprints generated were not matched to a protein in the two databases searched. Closer examination of the peptide mass spectra obtained for these protein spots revealed some poor quality spectra (27 spectra) attributable to low protein concentration and so these would be unlikely to provide identification. In addition, five protein spots were lost during the in-gel trypsin digestion. Removal of the poor quality spectra and lost spots brings the percentage of non-identified proteins down to 37.6%. Inspection of the results from the MASCOT® search engine reveals that a few protein spots with good spectra were not unambiguously identified (score < database score). Closer examination of these mass spectra with the mass spectra generated for proteins located at similar molecular mass but with a slightly more acidic or basic pI revealed high similarities in the peptide masses. Examination of the results obtained with the MASCOT® search engine for these peptide mass fingerprints showed that the first scores were just under the database score necessary for unambiguous identification. In addition, the protein name and accession number associated with the highest score for these spots was the same as that for protein spots with similar molecular mass but slightly different pI. This suggested that these could be accounted for by post-translational modification. Most post-translational modifications change the masses of the modified peptides (Table 4.3). Modified peptides would not fit the theoretical values calculated from the database entries by the MASCOT® search engine and therefore these would be eliminated from the identification.

38 good quality peptide mass fingerprints had no match in either the local *T. brucei* database or the NCBI database suggesting that the sequence of the ORFs encoding these protein spots were not available in the database, and that no close homologue of the proteins

PTM type	Δ Mass (Da)	Function and notes
Phosphorylation	+ 80	Reversible, activation/inactivation of enzyme activity, modulation of molecular interactions, signaling
Acetylation	+ 42	Protein stability, protection of N terminus. Regulation of protein-DNA interactions (histones)
Methylation	+ 14	Regulation of gene expression
Glycosylation:		
N-linked	>800	Excreted proteins, cell-cell recognition/signalling
O-linked	203, >800	Reversible, regulatory functions
Hydroxyproline	+ 16	Protein stability and protein-ligand interactions
Sulfation	+ 80	Modulator of protein-protein and receptor-ligand interactions
Pyroglutamic acid	-17	Protein stability, blocked N terminus
Ubiquitination	>1,000	Destruction signal.

Table 4.3: Post-translational modifications and mass shift. Reproduced from http://www.abrf.org/index.cfm/dm.home

occur in the sequenced genome of other organisms. These spots are likely to be identified when the whole annotated genome sequence of *T. brucei* becomes available. The 59% of unidentified protein spots can be explained by a combination of some poor quality spectra (21.4%), some heavy post-translational modifications (3%) and an incomplete *T. brucei* genome sequence (34.6%).

A possible explanation for the acquisition of poor mass spectra lies in the limitations of the MALDI-TOF MS. Examination of peptide mass spectra generated by MALDI-TOF on 100 proteins from *Mycobacteria* species showed that peak intensities of arginine containing peptides were significantly higher than those of lysine peptides (Krause *et al.*, 1999). The intensity of some of the lysine peptides can be so low that these peptides cannot be separated from background noise in the peptide mass fingerprint. This characteristic of peptide mass spectrum generated by MALDI-TOF limits the sequence coverage obtainable by PMF analysis, and consequently the score given to a protein spot.

PMF has been used extensively in proteomic studies to identify proteins separated by means of 2DE. Peptide mass analysis by MALDI-TOF was shown to combine simplicity with high throughput which made it an ideal first step to large scale proteome mapping. When complete genomic sequences are available, it has been demonstrated that up to 80% of the proteins separated by 2DE and analysed by MALDI-TOF can be positively identified (Shevchenko *et al.*, 1996). However, PMF efficiency for organisms with complete genome sequenced varies enormously from as little as 33% for *Homo sapiens* to 80% for *Saccharomyces cerevisae* (Table 4.4). These large differences in PMF efficiency for organisms where the complete genome has been sequenced can be explained by the size and complexity of their genomes. The human genome (3 Gb) has been completely

Organism	**Genome sequencing project**	**Genome size**	**Predicted number of genes**	**Efficiency of PMF**	**Reference**
Saccharomyces cerevisiae	Complete	12 Mb	6,280	80%	Wildgruber *et al.*, 2002
Homo sapiens	Complete	3 Gb	924 for chromosome 21 and 22	32.6%	Ha *et al.*, 2002
Helicobacter pylori	Complete	1.66 Mb	443	50.7%	Chao *et al.*, 2001
T. brucei	Not complete	35 Mb	8,000	40.5%	

Table 4.4: Proteome mapping projects for *Saccharomyces cerevisiae*, *Homo sapiens*, *Helicobacter pylori* and *T. brucei*. The four proteomic studies were carried out by 2DE of proteins and MALDI-TOF analysis of tryptic digested proteins selected from the two dimensional gels.

sequenced but its use for proteomic studies based on protein identification by PMF is limited. This is due to the gene organisation and the mechanisms of post-transcriptional RNA processing. The average human gene contains 10 to 15 exons separated by introns. Exons only account for about 2% of the total genome and the remaining 98% constitutes introns (24%) and intergenic regions (74%). These features make it difficult to identify genes in the human genome by ORF scanning. An additional factor is the extensive occurrence of complex RNA splicing events which is a key feature of human gene transcription (Modrek and Lee, 2002; Zhang *et al.*, 2002). Analysis of chromosomes 19 and 22 showed that there are, on average, 3.2 distinct mRNA transcripts per gene, with 70% of these involving alternative splicing within the coding region (Lander *et al.*, 2001). This means that the peptides predicted directly from the ORF will not correspond to those actually translated from the alternative spliced transcripts making the identification impossible. Pseudogenes are another major problem for ORF finding in the human genome. For example, there are 80 ribosomal protein genes in the human genome but >2,000 associated pseudogenes (Zhang *et al.*, 2002). Consequently, the assignment of ORF's is particularly difficult and far from complete for the human genome. In order to have an efficient identification of peptide mass fingerprints, it is necessary to have a complete database of the total ORFs and mRNA transcripts. The need for such a "perfect" database is because of the pure statistical basis of protein identification from peptide mass data. The proteome in different tissues and cell types would then represent a sub-set of this database. However, the importance of PMF should not be under estimated as, where identification is obtained, it provides much needed confirmation that ORFs are expressed.

In comparison, the yeast genome is far simpler than the human genome. *S. cerevisiae* was the first eukaryote genome to be fully sequenced (12 Mb). The work was completed in

1996 and the detection of ORFs in *S. cerevisiae* started at the same time. The lower complexity of the yeast genome is mirrored by the efficiency of protein identification by PMF.

The *T. brucei* genome organization has been determined from the sequencing and annotation of chromosome I and II. This analysis showed a lack of promoters, no introns, but large probable polycistronic units. Except for the addition of the spliced leader on every transcript, splicing is a rare event in *T. brucei*. The simple organization of the *T. brucei* genome means that the annotation of the genome is relatively straightforward and a complete database of the total ORFs and mRNA transcripts should be obtainable directly from the genome. Consequently, proteins can be identified by direct comparison of PMF with theoretical PMF from transcripts predicted from the ORFs, and this is reflected in the relatively high rate of protein identification reported here.

Identification of protein spots on the proteome map showed that post-translationally modified proteins were separated by 2DE from the primary translated products (e.g. HSP70 protein spots). This is because most post-translational modification leads to changes in pI. This change in pI gives rise to a specific protein spot pattern on 2D gels. In this study 51 proteins spots were identified which corresponded to 30 different proteins. Consequently, 40% of the protein spots identified were likely to be isoforms. This might seem a high percentage, but comparison with proteome maps from various organisms available at the Expasy's web site (http://ca.expasy.org/ch2d/2d-index.html) showed this to be an average isoform value. Post-translational modification of a protein can determine its activity state, localisation, turnover, and interactions with other proteins. For example, kinase cascades are turned on and off by the reversible addition and removal of phosphate

groups. As such the determination of the post-translational modification state of a protein is extremely important. Proteomic studies using 2DE and mass spectrometry enable the separation and assessment of the level of modification of a protein. The particular modification and the site of modification can further be analysed by a variety of techniques which have been recently reviewed (Mann and Jensen, 2003; Mann *et al.*, 2002; Arbersold and Goodlett, 2001).

However it is difficult, sometimes impossible with PMF data to differentiate between post-translational modifications and transcripts from different but closely related genes in a gene family. For example, the number of α and β tubulins gene copies on chromosome I has been estimated (from mapping data) to be 13-14 in one homologue and at least 9-10 in the other, with high homology between them (Hall *et al.*, 2003). Remarkably, almost 20% of the annotated genes on chromosome I are members of closely related families (Table 4.5).

The production of a preliminary proteome map of *T. brucei* demonstrates the feasibility of proteomic studies on this parasite. The lack of a complete and accurately annotated genome and protein sequence databases accounted for 34.6% of the non-identified protein spots. In 3% of the cases the data collected by MALDI-TOF mass spectrometry was not sufficient for reliable identification of a protein, especially in cases were the 2DE pattern suggested post-translational modification of the protein. In cases of ambiguous identification by PMF, Tandem mass spectrometry can be used in analysis to generate peptide sequence information about a particular protein spot of interest.

Before embarking on a proteomic study of Cymelarsan resistance in *T. brucei*, a biological analysis of the isogenic lines to be used needs to be undertaken to determine potential variables which could affect the protein expression in each line. The characterization of the two pairs of isogenic lines is described in the next chapter.

Product description	Distribution	% homology (protein)	Number of copies
RHS	Clustered	20-99%	5
LRRP1	Clustered	93-99%	3
Unknown function	Clustered	75-95%	3
Phosphate permease	Tandem	97%	2
PGK	Tandem	65-89%	3
Unknown function	Dispersed	99%	2
ESAG2	Dispersed	68-82%	3
Calpain-like	Clustered	6-80%	7
β tubulin	Tandem	>99%	>9
α tubulin	Tandem	>99%	>9
Histone H3	Tandem	>99%	7
Unknown function	Tandem	100%	7
Pteridine transporter	Tandem	>99%	3

Table 4.5: Gene families on chromosome I of *T. brucei*. Partial summary of the characteristic of gene families found on chromosome I. ND, not determined. Reproduced from Hall *et al.*, 2003.

Chapter V

Characterisation of isogenic drug sensitive and drug resistant lines of trypanosomes

5.1 Introduction

In previous chapters, protocols for the extraction/solubilisation of proteins from *T. brucei* and their separation by 2DE were developed, a sensitive and specific search engine was selected for identification of proteins using mass spectrometry data and a preliminary analysis of the proteome of line 927 demonstrated that 40% of proteins separated by 2DE can be identified using the simple technique of PMF. These results have essentially established the main methodologies for undertaking proteomic analysis of *T. brucei* as well as defining the limitations. As discussed in the introduction, one application of proteomics is the analysis and identification of alterations associated with particular phenotypes and the main aim of the project was to apply this technology to analysing the alterations in drug resistant trypanosome lines.

Cymelarsan (Melarsamine hydrochloride) is a close analogue of Melarsoprol used in veterinary trypanosomiasis. Cymelarsan is a water-soluble trivalent agent that is very effective against *T. b. brucei, T. evansi* and *T. equiperdum* infections in camels, buffalo, goats and pigs (Payne *et al.*, 1994; Nyang'ao *et al.*, 1995) but is very unstable in solution and forms several dissociation products. The primary dissociation product, Melarsen oxide, is thought to be the active form of Cymelarsan (Berger and Fairlamb, 1994) as well as the primary metabolite of

Melarsoprol *in vivo* (Keiser *et al.*, 2000). Consequently, Cymelarsan generated resistance can be used as a model for Melarsoprol resistance in trypanosomes.

Cymelarsan resistance was generated *in vivo* in the STIB 386 and STIB 247 lines of trypanosomes by gradually increasing sub-curative doses of drug (Scott *et al.*, 1996). Comparison of EC_{50} values obtained from *in vitro* growth inhibition assays demonstrated that procyclic forms of the laboratory-generated Cymelarsan resistant lines were over 500 fold more resistant to Cymelarsan than the sensitive parental lines (Table 5.1). Preliminary characterisation of the two pairs of isogenic lines (386/386Mr, 247/247Mr) showed that the drug resistance phenotype was stable after passaging in mice in the absence of drug pressure as well as being stable after transmission through tsetse flies (Scott *et al.*, 1996). Cross-resistance between Cymelarsan and Melarsoprol was observed in bloodstream forms *in vivo* but not in procyclic forms in *in vitro* inhibition assays for both pairs of isogenic lines (Table 5.1). This apparent lack of cross-resistance was observed to be an artifact of the *in vitro* assay (Scott *et al.*, 1997). In addition, *in vitro* studies demonstrated that bloodstream forms of the Cymelarsan resistant 247 line was partially cross-resistant to two diamidine drugs, Pentamidine and Berenil (data not shown, Scott *et al.*, 1997).

The two pairs of isogenic lines (386/386Mr and 247/247Mr) have been previously analysed to determine whether they possess, transcribe and express the P2 transporter gene (*TbAT1*). Southern blot analysis demonstrated that the sensitive 386 and 247 lines both had the genes encoding P1 and P2 transporters. The Cymelarsan resistant 386 line possessed a *TbAT1* gene with the same restriction digestion pattern as the sensitive line, whereas the cluster of genes

	in vitro (EC_{50})		*in vivo* (MIC in mg/Kg)	
Trypanosome line	**Cymelarsan**	**Melarsoprol**	**Cymelarsan**	**Melarsoprol**
386	0.025	0.98	2.0	0.64
386Mr	14	1.3	>40	>40
247	0.06	7.0	0.3	0.32
247Mr	31	16	>40	>40

Table 5.1: Results of *in vitro* and *in vivo* cross-resistance assays for the pairs of isogenic lines. Resistance of sensitive (386 and 247) and Cymelarsan induced drug resistant (386Mr and 247Mr) lines were tested by growth inhibition assay. Average EC_{50} values (μM) of Cymelarsan and Melarsoprol calculated from growth inhibition assays for the wild type and the selected lines. The minimal inhibitory concentration (MIC) values from *in vivo* assays corresponded to the dose required to cure mice infected with the unselected or selected lines. Data from Scott *et al.*, 1996.

coding for P1 transporter was rearranged (Dr R.Burchmore, personal communication). Northern analysis showed that bloodstream and procyclic forms of the Cymelarsan resistant 386 line do not transcribe the *TbAT1* gene, whereas the sensitive 386 line expressed it in both life cycle stages. Analysis of the Cymelarsan resistant 247 line showed that the *TbAT1* gene was completely lost (Dr R. Burchmore, personal communication). These unpublished data show that the loss of expression of the P2 transporter is associated with the Cymelarsan resistance phenotype. The expression of the Cymelarsan resistance in procyclic forms of trypanosomes, for which no P2 activity can be detected (de Koning *et al.*, 2000), strongly suggested that *TbAT1* is not the only gene responsible for conferring arsenical resistance in these lines. In addition, the level of Cymelarsan resistance conferred by knockout of *TbAT1* is much lower than the levels detected in these two pairs of isogenic lines (Matovu *et al.*, 2003), further supporting the theory of a multi-factorial mechanism of drug resistance.

In this chapter the genotypes and phenotypes of the isogenic Cymelarsan sensitive and resistant lines were assessed to characterize these lines more completely. An assessment of some of the variable, which can affect the proteome of the isogenic lines was undertaken. In addition, obvious candidates for the mechanism of Cymelarsan resistance in *T. brucei*, such as export of the drug by a P-glycoprotein and analysis of cross-resistance in regard to P2 transporter alteration were assessed.

5.2 Materials and Methods

5.2.1 Commonly used reagents

The trifluoperazine (TFP), prochlorperazine (PCP), verapamil, EDTA, SDS, Tris, Gentamycin, Berenil and Pentamidine were purchased from SIGMA. Proteinase K, agarose, FCS and SDM79 were purchased from Gibco BRL. Reagents purchased from other companies are indicated in the text.

5.2.2 *In vitro* culture of procyclic trypanosomes

Procyclic forms of the pairs of isogenic lines were grown *in vitro* in complete SDM79 as described in chapter 2.

5.2.3 DNA extraction and genotyping of the isogenic lines

The pellet from 1 ml of trypanosome procyclic culture was concentrated in an Eppendorf tube by centrifuging for 3 min at 12,000 g, and washed once using 500 µl of PSG buffer at pH 7.8. The trypanosomes were lysed in 100 µl of lysis buffer (50 mM Tris pH 8.0, 10 mM EDTA pH 8, 0.5% SDS) with addition of 3.3 µl of 20 mg/ml proteinase K and incubated at 56°C for 4 h. The tubes were then incubated for 5 min at 95°C to inactivate the proteinase K and stored at -20°C. Genotyping was undertaken by PCR amplification of three highly polymorphic minisatellite markers, CRAM, MS42 and 292, using the primers listed in Table 5.2 (MacLeod *et al.*, 1999). PCR was undertaken in 10 µl reaction volumes containing 5 pmol/ml of each primer, 1 µl of PCR Master mix (ABgene), 1 unit of Taq polymerase (ABgene), and 200-300 ng of DNA template. The PCR conditions were as follow: 28 cycles of denaturation for 50 sec at 95°C, annealing for 50 sec at 64°C and elongation for 3 min at 68°C, on a Robocycler

Primer	Sequence
CRAM G	5'- CTGCTGATGCCGTACATGATGATTTC
CRAM H	5'- AACTCCCTCCCGATCGATCACAAC
MS42 F	5'- TTGTGCGGTCGTTAACGCGCGTTCAA
MS42 W	5'- GGTGATTCATCGGCTCCCTTACCA
292 G	5'- ACACCCCCTCTCCACTTCAGATAC
292 H	5'- GCTGAACCTGTGGGCCCCTCAATTG

Table 5.2: Sequence of primers for genotyping. Sequences of the primers used to PCR amplify the three minisatellites CRAM, MS42 and 292 for genotyping the isogenic lines.

Gradient 96 (Stratagene). Agarose gels for electrophoresis of DNA were prepared as described in Sambrook *et al.* (1989). Powdered agarose (1% w/v, Seakem) was mixed with 0.5x electrophoresis buffer (0.045 M Tris-borate, 0.001 M EDTA) and ethidium bromide (1%). The DNA samples (10 µl) were mixed with the gel-loading buffer (0.25% (w/v) bromophenol blue, 30% (v/v) glycerol in water) and loaded into the wells. The samples were run in the gels at 100 V for 2 h 30 min. The gels were then examined under ultraviolet light and photographed with a Gel Imager.

5.2.4 *In vitro* growth curves of the isogenic lines

Procyclic forms of each isogenic line were grown by inoculating complete SDM79 with $5x10^5$ procyclic trypanosomes, and incubating the cultures at 25°C. Growth curves were obtained by estimating trypanosome concentrations every 24 h over a 7 day period using an Improved Neubauer haemocytometer. The experiments were carried out in triplicate. The mean trypanosome concentration was plotted against time in hours and from these growth curves, the rate of growth and population doubling time (PDT) were calculated. The rate of growth (b) over the first three days was calculated by linear regression analysis ($y = bx + c$). The mean PDTs were calculated from the growth rates using the following equation: $PDT = \ln 2/b$.

5.2.5 *In vitro* trypanocidal drug sensitivity testing

5.2.5.1 Growth inhibition assays

Four trypanocidal drugs were each used in *in vitro* drug resistance assays on procyclic forms of the pairs of isogenic lines. Cymelarsan (Rhône-Mérieux) and Berenil were dissolved in filter sterile complete SDM79 medium and Pentamidine was dissolved in 20% DMSO solution in complete SDM79 medium. Melarsoprol was available as a 3.6% solution in propylene

glycol. The drugs were serially diluted in complete SDM79 medium, to obtain concentrations ranging from 0.001 μM to 1,000 μM. In microtitre plates, 100 μl of drug was mixed with 100 μl of procyclic culture at $2x10^6$ trypanosomes/ml and incubated at 25°C. The controls consisted of 100 μl of complete SDM79 mixed with 100 μl of procyclic culture at $2x10^6$ trypanosomes/ml. Growth was assessed by counting the trypanosomes in an Improved Neubauer haemocytometer after 24 h incubation. The growth inhibition experiments were performed in triplicate, and repeated at least once. The percentage inhibition is expressed relative to the cell densities in culture growing in the absence of drug. The growth was plotted against increasing drug concentration, and the values between 10% and 90% growth were approximated to a straight line by linear regression (R^2 values varying from 0.92 to 0.97). The drug concentration which inhibited growth by 50% (EC_{50} value) was calculated from the regression equation.

5.2.5.2 Growth inhibition assays with Cymelarsan and P-glycoprotein inhibitors

Three general P-glycoprotein inhibitors, prochlorperazine (PCP), trifluoperazine (TFP) and verapamil were used in combination with Cymelarsan to assess the potential involvement of drug efflux as a mechanism of Cymelarsan resistance in procyclic forms of the pairs of isogenic lines. The P-glycoprotein inhibitors were serially diluted in complete SDM79 medium. The assay was performed in microtiter plates as shown in Table 5.3. The plates were incubated for 24 h at 25°C and concentration of trypanosome was determined using an Improved Neubauer haemocytometer. The experiments were performed in triplicate and repeated at least once.

5.2.5.3 The Alamar Blue assay

	Controls			Drug combination	
Cymelarsan: 1 μM	100 μl			100 μl	100 μl
Inhibitor of P-glycoprotein: 1 μM		100 μl		100 μl	
Inhibitor of P-glycoprotein: 10 μM			100 μl		100 μl
Trypanosomes (procyclic culture)	100 μl	100 μl	100 μl	100 μl	100 μl
SDM79	100 μl	100 μl	100 μl		

Table 5.3: Growth inhibition assays with Cymelarsan and P-glycoprotein inhibitors. The effect of P-glycoprotein inhibitors on growth inhibition assays of Cymelarsan were tested on procyclic forms of the pairs of isogenic lines.

Cymelarsan was serially diluted in complete SDM79 medium, with concentrations ranging from 0.001 μM to 1,000 μM. In microtiter plates, 100 μl of Cymelarsan solution was mixed with 100 μl of procyclic culture at $2x10^6$ trypanosomes/ml and 20 μl of the Alamar Blue dye. The control was made by replacing the Cymelarsan solution with complete SDM79. The plates were incubated for 24 h at 25°C. Measurements were taken with a LS 55 Luminescence Spectrometer (Perkin Elmer instruments) at 530 nm excitation and 590 nm emission wavelengths. Three initial concentrations of procyclic cultures ($2x10^6$, $4x10^6$ and $6x10^6$ trypanosome/ml) of the Cymelarsan sensitive and resistant lines were tested. Presence and viability of the trypanosomes in each well was confirmed by microscopic examination. Each experiment was performed in duplicate and replicated at least once.

5.3 Results

5.3.1 Genotypes of isogenic lines

All procyclic lines were genotyped using three minisatellite markers (CRAM, MS42, 292) as described previously by MacLeod *et al.* (1999). These minisatellite markers were selected as they are highly polymorphic and can differentiate between strains of *T. brucei* with a high probability. The genotypes of 386 and 386Mr lines (Figure 5.1A) and of 247 and 247Mr lines (Figure 5.1B) were each identical for the three minisatellite markers used in this study. These data demonstrated that the sensitive and resistant pairs of lines are indeed isogenic as originally described (Scott *et al.*, 1995) and no inadvertent substitution of one line for another with a different genotype has occurred. This finding contrasts with the genotyping results of RU15 line, a laboratory derived Melarsoprol resistant line made from 427 line, which had a

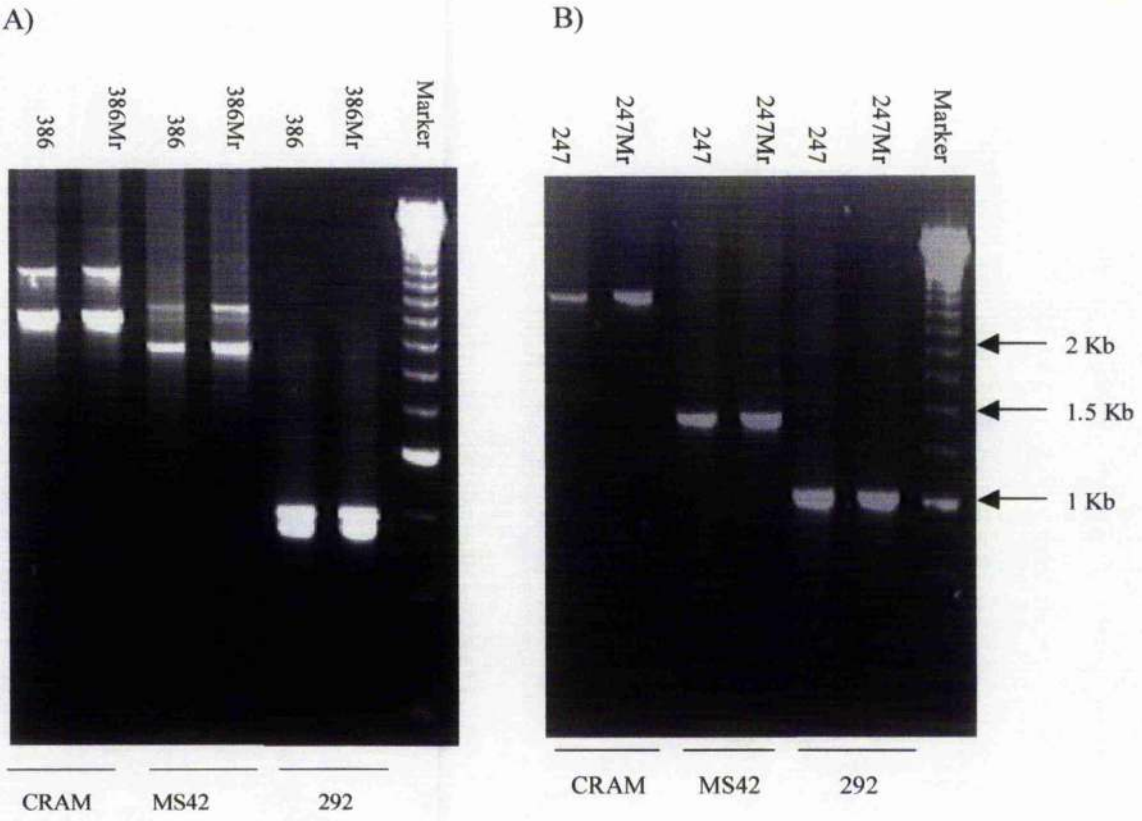


Figure 5.1: **Genotypes of the two pairs of isogenic lines**. The sizes of the PCR products correspond to the expected sizes for line 386 (2.8 Kb and 2 Kb for CRAM, 2.3 Kb and 1.8 Kb for MS42 and 1 Kb and 0.95 Kb for 292) and 247 (2.8 Kb for CRAM, 1.5 Kb for MS42 and 1.3 Kb for 292) (Dr A. MacLeod, personal communication).

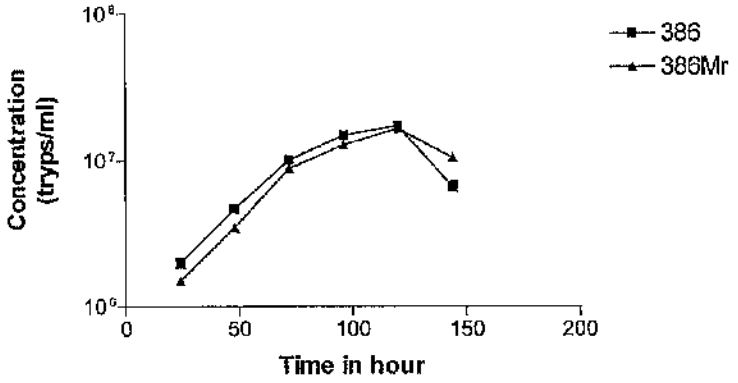
different genotype from the parent line using these three minisatellite markers (Dr A. MacLeod, personal communication).

5.3.2 Growth of isogenic lines

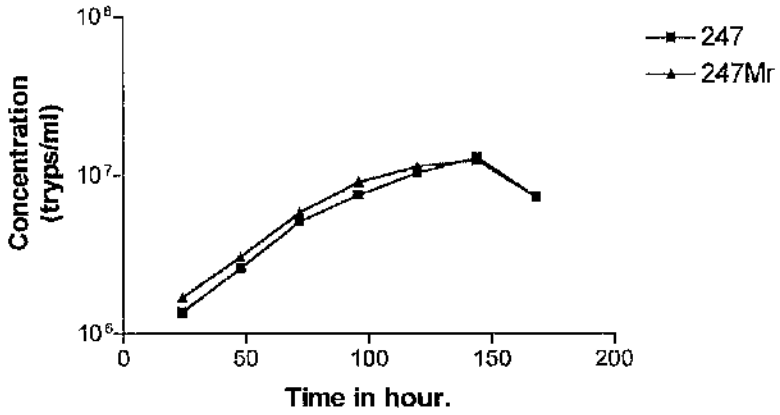
The rates of growth of procyclic forms and the population doubling times were calculated to investigate whether selection for Cymelarsan resistance had caused a change in growth rate. The assessment of *in vitro* growth of the procyclic forms of 386 and 386Mr lines showed no noticeable difference in growth between the isogenic lines (Figure 5.2A). Both procyclic cultures reached a maximum concentration of $1.7x10^7$ trypanosomes/ml in the stationary phase. Growth rates and mean PDTs were calculated for the isogenic lines from these growth curves to assess further variability between the lines (Figure 5.2C). The growth rates were very similar (0.023 and 0.025 respectively). The *in vitro* growth curves of procyclic cultures of 247 and 247Mr isogenic lines showed no detectable differences in growth between them (Figure 5.2B). The maximum trypanosome concentration reached by these iogenic lines was $1.3x10^7$ trypanosomes/ml.

Taken together these results suggest that there are no detectable differences in growth rates between Cymelarsan sensitive and resistant lines of 386 and 247. Thus, any proteomic profile differences found between the lines are unlikely to be artefacts of differences in growth that might have arisen as a pleiotropic phenotype when the lines were originally selecting for drug resistance.

A)



B)



C)

	Growth rate	Population doubling time
386	0.023	30.26 h
386Mr	0.025	27.29 h
247	0.021	32.39 h
247Mr	0.020	33.97 h

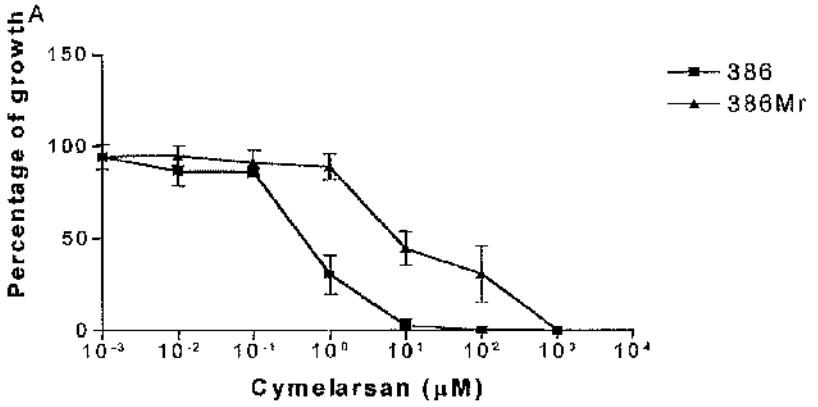
Figure 5.2: Growth curves of the two pairs of isogenic lines. Procyclic forms of isogenic lines of 386 (A) and 247 (B) were grown in SDM79 *in vitro* and the average growth rate and population doubling time (PDT) were calculated from the growth curves (C). The experiments were performed in triplicate.

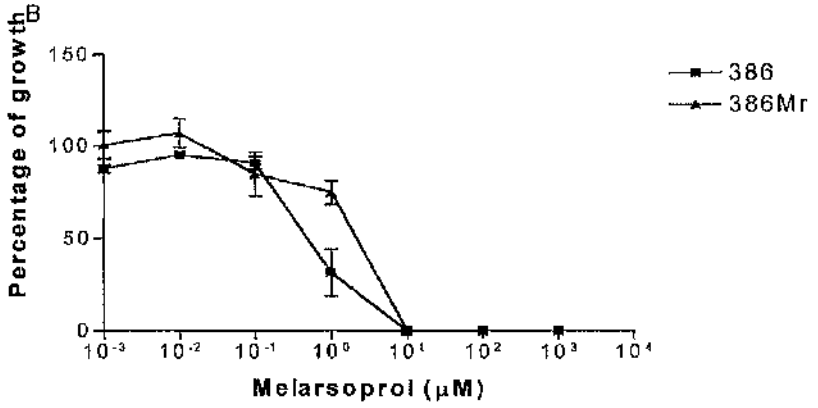
5.3.3 Sensitivity and resistance to arsenicals and diamidines

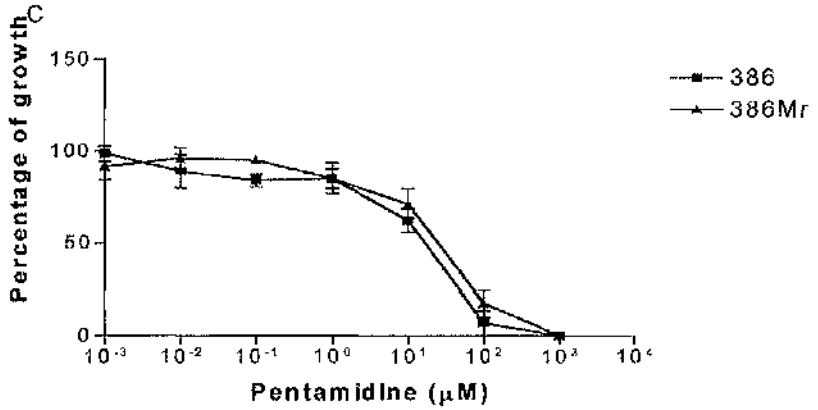
In vitro growth inhibition assays of Cymelarsan sensitive and resistant lines of 386 and 247 were used to assess sensitivity/resistance of the pairs of isogenic lines to trypanocidal drugs. Two arsenical drugs (Cymelarsan and Melarsoprol) and two diamidine drugs (Pentamidine and Berenil) were tested.

The procyclic forms of 386 and 386Mr lines exhibited a dose dependent inhibition of growth by Cymelarsan. The 386Mr line was not affected to the same extent by the presence of Cymelarsan in the medium compared to the 386 sensitive parental line (Figure 5.3A). At a concentration of 1 μM Cymelarsan, the 386Mr line showed only 25% inhibition in growth whereas the 386 line showed 75% inhibition of growth. The 386 line was completely inhibited by 10 μM Cymelarsan whereas the 386Mr line was completely inhibited only by the highest concentration of Cymelarsan (1,000 μM). The minimum dose inhibiting 50% of growth (EC_{50}) was estimated to be 0.57 μM for 386 and 14.19 μM for 386Mr, showing a 24.9 fold increase in EC_{50} (Table 5.4).

The growth inhibition assays of 386 and 386Mr lines in the presence of Melarsoprol showed no detectable differences for most concentrations of Melarsoprol (Figure 5.3B). The only detectable difference was seen at a Melarsoprol concentration of 1 μM, at which the 386Mr line showed a 25% reduction in growth whereas the 386 line showed a 70% reduction in growth. Both isogenic lines of 386 exhibited a 100% inhibition of growth at a concentration of 10 μM of Melarsoprol. Comparison of the EC_{50} values of the 386 line and the 386Mr line for Melarsoprol showed less than two fold increase in resistance (Table 5.4).


A
Percentage of growth
150
100
50
0
10^{-3} 10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 10^{3} 10^{4}
Cymelarsan (μM)
386
386Mr


B
Percentage of growth
150
100
50
0
10^{-3} 10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 10^{3} 10^{4}
Melarsoprol (μM)
386
386Mr


C
Percentage of growth
150
100
50
0
10^{-3} 10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 10^{3} 10^{4}
Pentamidine (μM)
386
386Mr

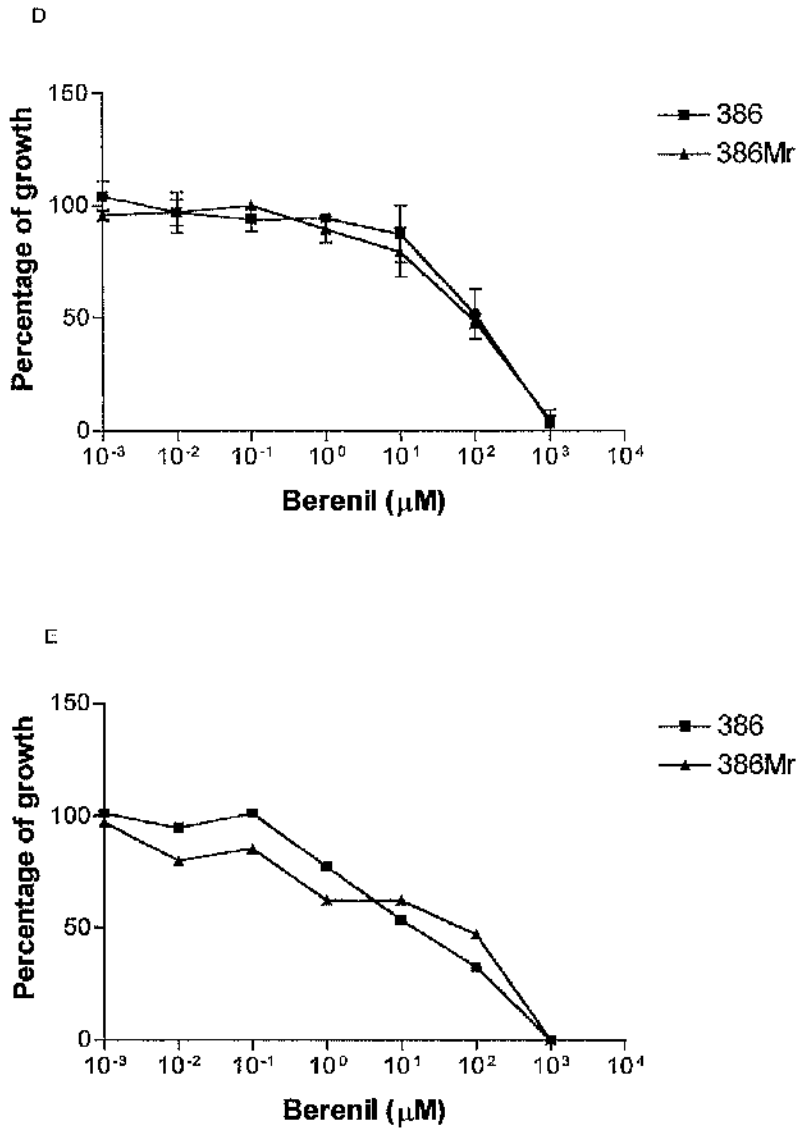


Figure 5.3: ***In vitro*** **growth inhibition assays in arsenical and diamidine drugs for the isogenic 386 lines.** Procyclic forms of 386 and 386Mr lines were grown in SDM79 in the presence of the trypanocidal drugs Cymelarsan (A), Melarsoprol (B), Pentamidine (C), Berenil for 24 h (D) and Berenil for 48 h (E). The data present the average value from duplicate experiments. The lowest concentration corresponds to the control in the absence of drug. The error bars represent the mean +/- SEM, n = 6.

	Cymelarsan	Melarsoprol	Pentamidine	Berenil
386	0.57	0.62	10.85	86.4
386Mr	14.19	1.19	17.01	69.24
247	0.17	0.76	140.83	36.93
247Mr	213.79	1.91	82.24	81.22

Table 5.4: Average EC_{50} values (μM) from *in vitro* growth inhibition assays for the two pairs of isogenic lines. The assays were performed on procyclic forms of the pairs of isogenic lines for two arsenical drugs (Cymelarsan and Melarsoprol) and two diamidine drugs (Pentamidine and Berenil).

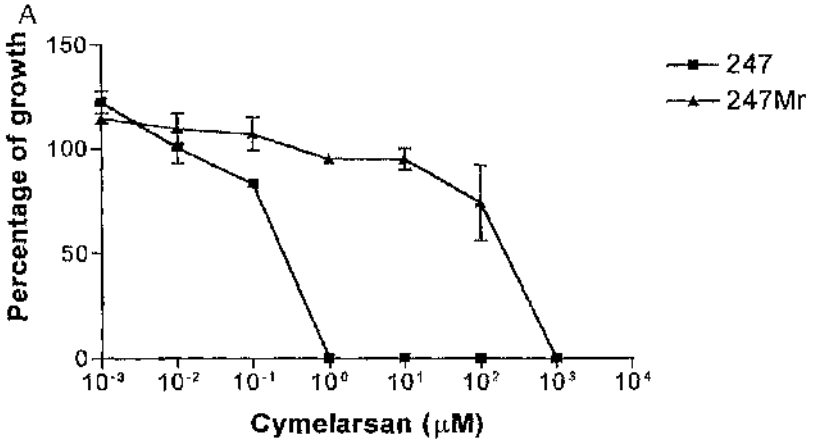
The phenomenon of cross-resistance between the diamidine, Berenil, and the arsenicals is commonly reported as well as cross-resistance with Pentamidine (Fairlamb *et al.*, 1992; Frommel and Balber, 1987; Pospichal *et al*, 1994). Cross-resistance between the arsenicals and the diamidinies (Berenil and Pentamidine) was assessed for procyclic forms of the pairs of isogenic lines by an *in vitro* growth inhibition assay.

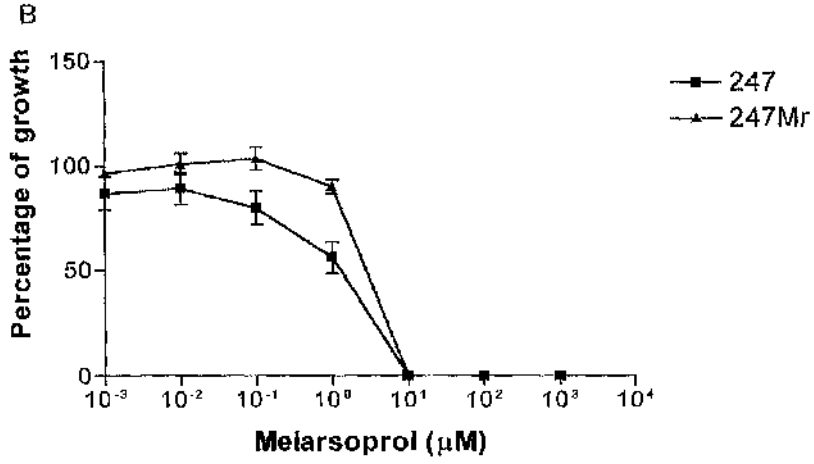
Resistance of the procyclic forms of the isogenic lines of 386 to Pentamidine was tested using the *in vitro* growth inhibition assay (Figure 5.3C). The 386 and 386Mr lines showed very similar growth inhibition curves in the presence of Pentamidine. 100% growth inhibition was detected at a concentration of 1,000 μM Pentamidine for both lines. Comparison of the EC_{50} values showed a slight increase (1.56 fold) in resistance to Pentamidine (Table 5.4). Two growth inhibition assays for 386 and 386Mr lines were conducted for the diamidine drug Berenil (Figure 5.3). Berenil is a trypanostatic drug and as such might require a longer period of exposure to inhibit the growth of the trypanosomes. Consequently, concentrations of trypanosomes were calculated after 24 h (Figure 5.3D) and 48 h (Figure 5.3E) of incubation in the presence of Berenil. For both growth inhibition assays the resistance pattern of 386 and 386Mr lines were very similar. Berenil had no noticeable effect on the growth of the isogenic lines up to a concentration of 10 μM. Both lines exhibited 100% growth inhibition when incubated in 1,000 μM Berenil for either 24 h or 48 h. The EC_{50} values for the isogenic lines were compared showing that the 386 and 386Mr lines had similar sensitivity to Berenil (Table 5.4).

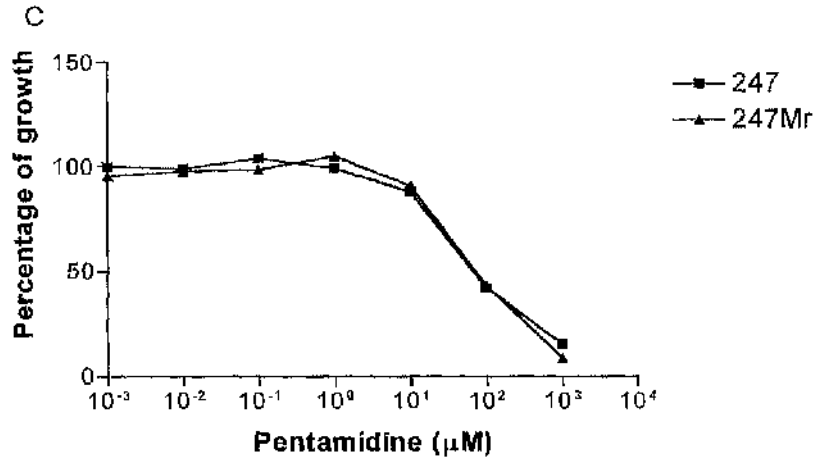
Growth inhibition assays were conducted *in vitro* on procyclic forms of Cymelarsan sensitive and resistant 247 lines to assess their sensitivity to two arsenical-based drugs (Cymelarsan and Melarsoprol) and two diamidine drugs (Pentamidine and Berenil).

The isogenic lines showed a dose dependent growth inhibition by Cymelarsan and 100% inhibition of growth of the 247 sensitive line was observed at a concentration of 1 μM Cymelarsan (Figure 5.4A). This concentration of Cymelarsan had no inhibitory effect on the growth of the resistant line. 100% inhibition of growth of the 247Mr line was observed at a Cymelarsan concentration of 1,000 μM. Comparison of the EC_{50} values for the isogenic lines showed a 1,247 fold increase in resistance to Cymelarsan by the resistant line (Table 5.4). Growth inhibition assays of the isogenic lines with Melarsoprol showed a slight difference between the lines at concentrations of 0.1 μM and 1 μM of Melarsoprol (Figure 5.4B). At a concentration of 10 μM of Melarsoprol both 247 and 247Mr lines exhibited 100% inhibition of growth. Comparison of the EC_{50} values obtained for Melarsoprol with the isogenic lines showed a 2.5 fold increase in resistance to Melarsoprol in the 247Mr line (Table 5.4).

Procyclic forms of the 247 and 247Mr lines exhibited similar growth inhibition patterns in increasing concentration of Pentamidine (Figure 5.4C). No effect on the growth of the isogenic lines was noticeable up to a concentration of 10 μM Pentamidine. At the highest concentration of Pentamidine tested (1,000 μM), 93% of growth was inhibited. The 247 showed less than a two fold increase in resistance to Pentamidine when compared with the 247Mr line (Table 5.4). The growth inhibition assays of 247 and 247Mr lines in the diamidine drug Berenil showed no detectable difference in growth (Figure 5.4D). At a concentration of


A
Percentage of growth
150
100
50
0
10^{-3} 10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 10^{3} 10^{4}
Cymelarsan (μM)
247
247Mr


B
Percentage of growth
150
100
50
0
10^{-3} 10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 10^{3} 10^{4}
Melarsoprol (μM)
247
247Mr


C
Percentage of growth
150
100
50
0
10^{-3} 10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 10^{3} 10^{4}
Pentamidine (μM)
247
247Mr

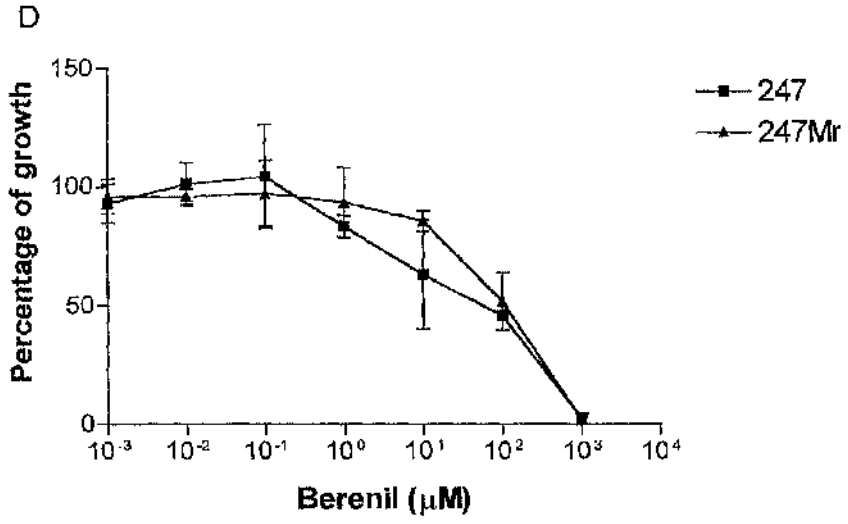


Figure 5.4: *In vitro* growth inhibition assay in arsenical and diamidine drugs for the isogenic 247 lines. Procyclic forms of sensitive and Cymelarsan resistant 247 lines were grown in SDM79 in the presence of the trypanocidal drugs Cymelarsan (A), Melarsoprol (B), Pentamidine (C), Berenil (D). The lowest concentration corresponds to the control in the absence of drug. The error bars represent the mean +/- SEM, n = 6.

1,000 μM Berenil both trypanosomes lines exhibited a 100% inhibition of growth. The EC_{50} values were calculated for both isogenic lines and a comparison of these showed a 2.2 fold increased in sensitivity to Berenil by the 247Mr line (Table 5.4).

These results confirmed that the procyclic forms of the Cymelarsan resistant lines were resistant to Cymelarsan as compared to the sensitive parents, as originally described (Scott *et al.*, 1996). Taken together these results suggest that in *in vitro* growth inhibition assays the procyclic forms of the pairs of isogenics lines had no cross-resistance between Melarsoprol and Cymelarsan, and no cross-resistance to the diamidine drugs in either of the pairs of isogenic lines.

5.3.4 The effect of P-glycoprotein inhibitors on Cymelarsan resistance

Recently, two genes encoding P-glycoproteins have been identified in the *T. brucei* genome (Shahi *et al.*, 2002). Over-expression of one of these genes, the *TbMRPA* gene, was linked to a 10 fold increase in Melarsoprol resistance (Shahi *et al.*, 2002). This report raised the question as to whether the resistance of the drug selected lines in this study could arise from a change in expression of either of these P-glycoprotein genes. To test this hypothesis, an investigation of growth inhibition by Cymelarsan in the presence of three candidate inhibitors of P-glycoprotein was carried out.

Three general P-glycoprotein inhibitors (verapamil, PCP and TFP) were used in combination with Cymelarsan in *in vitro* assays to assess Cymelarsan efflux as a potential mechanism for drug resistance in procyclic forms of the two pairs of isogenic lines. Verapamil at concentrations of 1 μM or 10 μM had no growth inhibitory effect on any lines (Figure 5.5A

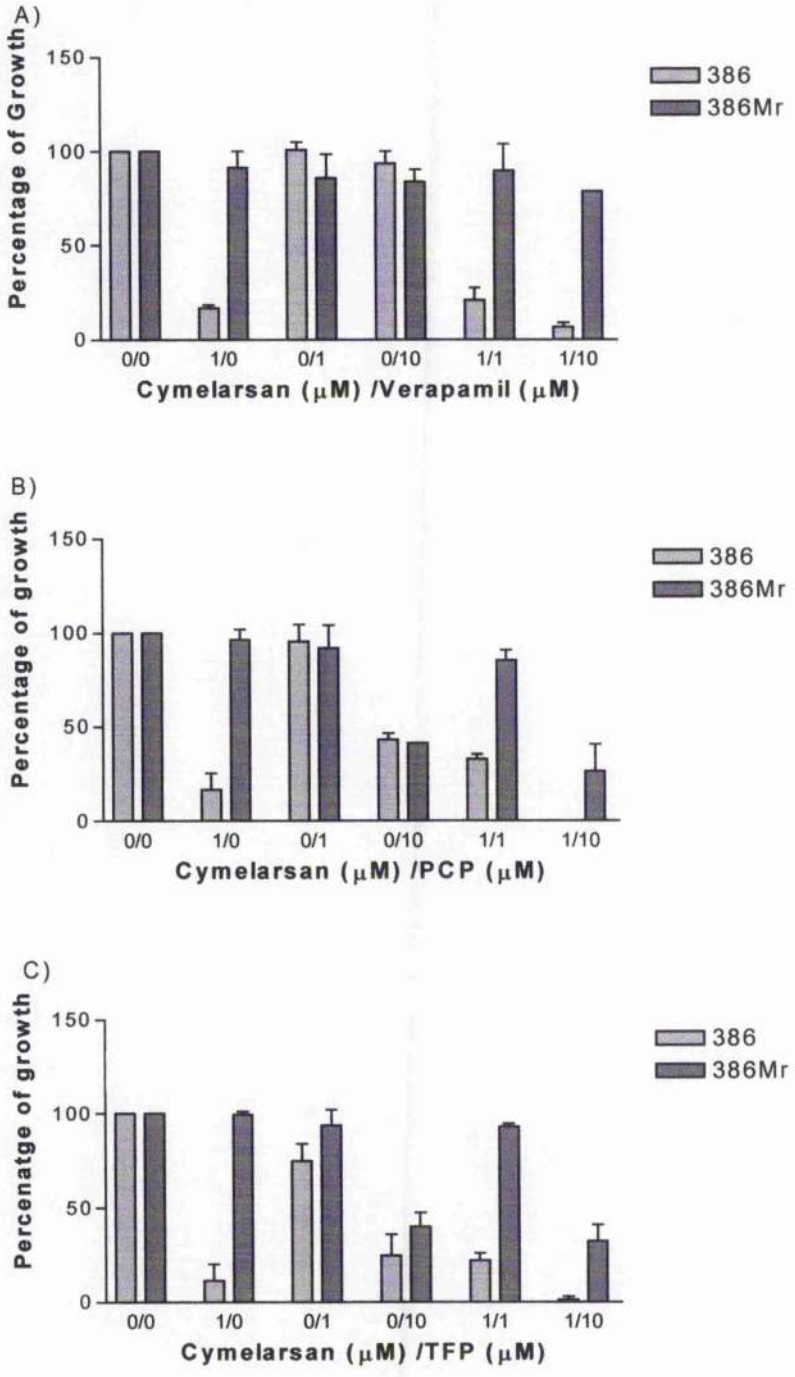


Figure 5.5: ***In vitro*** **growth inhibition assays of isogenic 386 lines in the presence of Cymelarsan and P-glycoprotein inhibitors.** The assays were performed on procyclic forms. Three general P-glycoprotein inhibitors (verapamil (A), PCP (B) and TFP(C)) were tested. The error bars represent the mean +/- SEM, n = 6.

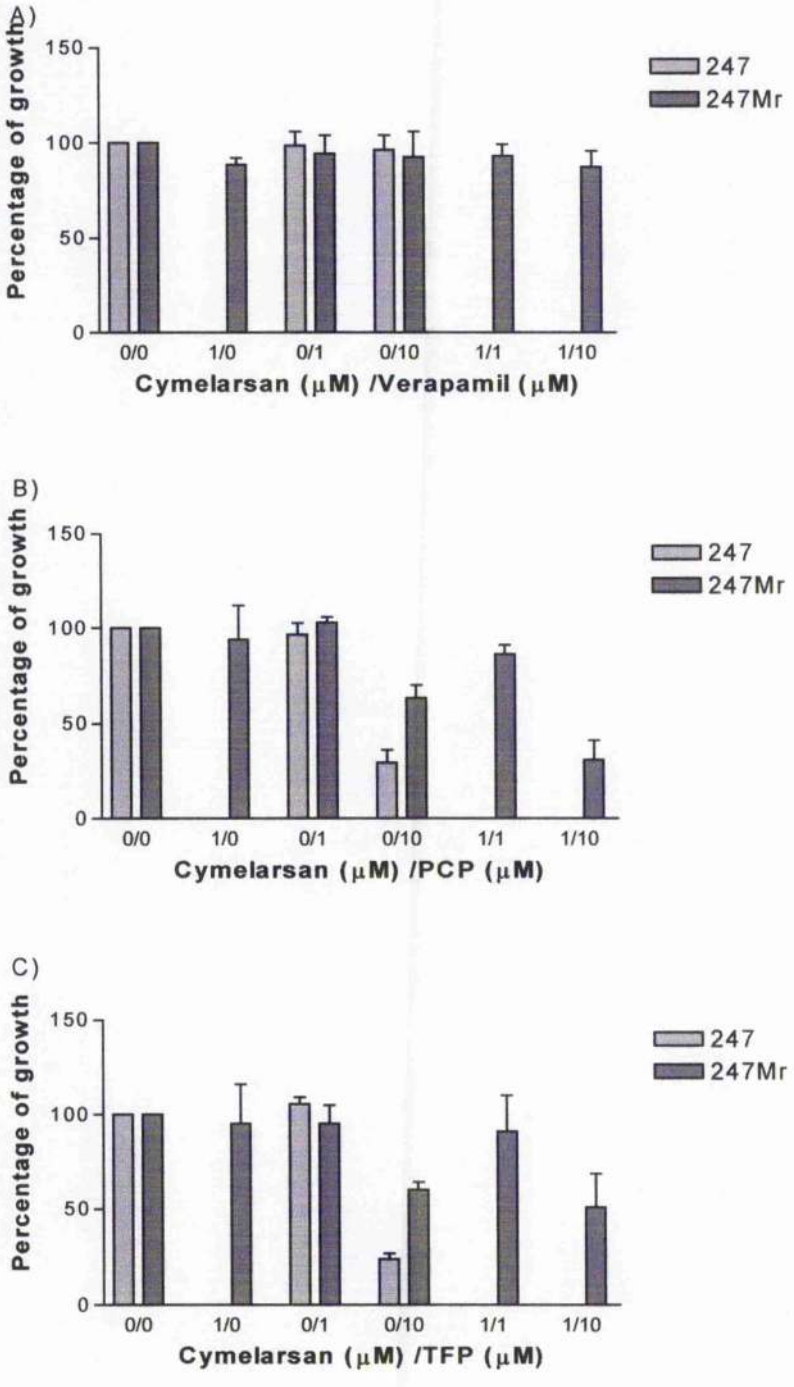


Figure 5.6: ***In vitro*** **growth inhibition assay of isogenic 247 lines in the presence of Cymelarsan and P-glycoprotein inhibitors**. The assays were performed on procyclic forms. Three general P-glycoprotein inhibitors (verapamil (A), PCP (B) and TFP(C)) were tested. The error bars represent the mean +/- SEM, n = 6.

and 5.6A). The addition of verapamil did not prevent the inhibition of growth produced by the presence of the arsenical in the medium (Figure 5.5A and 5.6A). These results were observed for all four lines tested. Both PCP and TFP were shown to be toxic to procyclic forms of all lines at a concentration of 10 μM, but had no effect on growth at 1 μM. Comparison of the growth of trypanosomes exposed to 1 μM of TCP or PCP in the presence of either 0 μM or 1 μM of Cymelarsan demonstrated that these P-glycoprotein inhibitors had no noticeable effect on the resistance to Cymelarsan (Figures 5.5B, 5.5C, 5.6B and 5.6C).

These results suggest that efflux of Cymelarsan by an over-expressed or altered P-glycoprotein in the Cymelarsan resistant 386 and 247 lines was not involved in the mechanism of Cymelarsan resistance in procyclic forms.

5.3.5 Drug sensitivity testing using the Alamar Blue assay

The use of the Alamar Blue dye is based on the ability of cellular enzymes to reduce the non-fluorescent blue dye to a pink fluorescent compound, allowing detection of cell growth by spectrofluorometry and spectrometry. The Alamar Blue assay was recently developed as an *in vitro* assay to test for drug sensitivity of procyclic trypanosomes in culture (Räz *et al.*, 1997). This assay was shown to be simple, reproducible and economical.

The Alamar Blue assay was used to assess the sensitivity or resistance of the procyclic forms of the pairs of isogenic lines in *in vitro* culture as an alternative to the growth inhibition assay. The assay was first performed on procyclic cultures in the absence of drug and demonstrated that the level of reduced Alamar Blue dye produced by Cymelarsan resistant lines was lower than the level produced by sensitive lines (Figure 5.7). The assay was then performed in

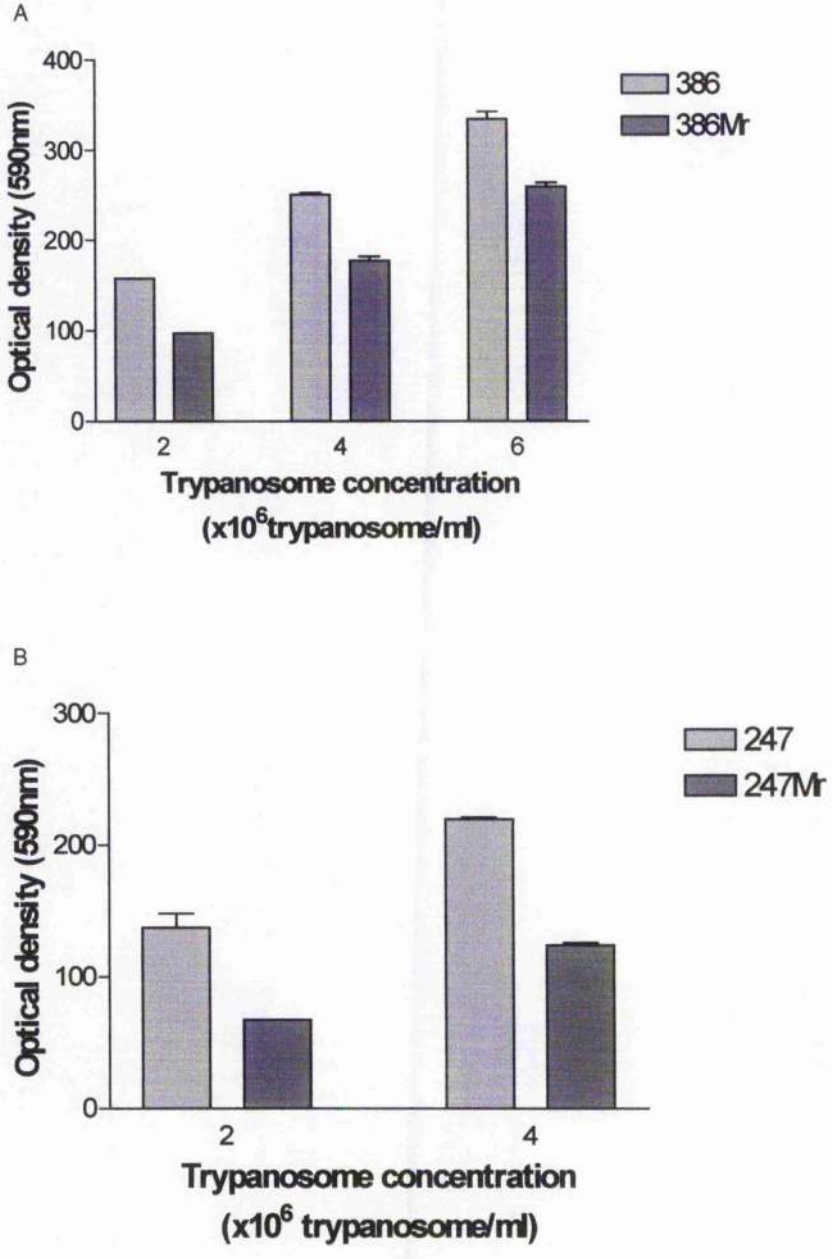


Figure 5.7: **Alamar Blue assays on the pairs of isogenic lines in the absence of drug.** *In vitro* reduction of the Alamar blue dye by procyclic forms of isogenic lines of 386 (A) and 247 (B) in SDM79. The error bars represent the mean +/- SEM, n = 4.

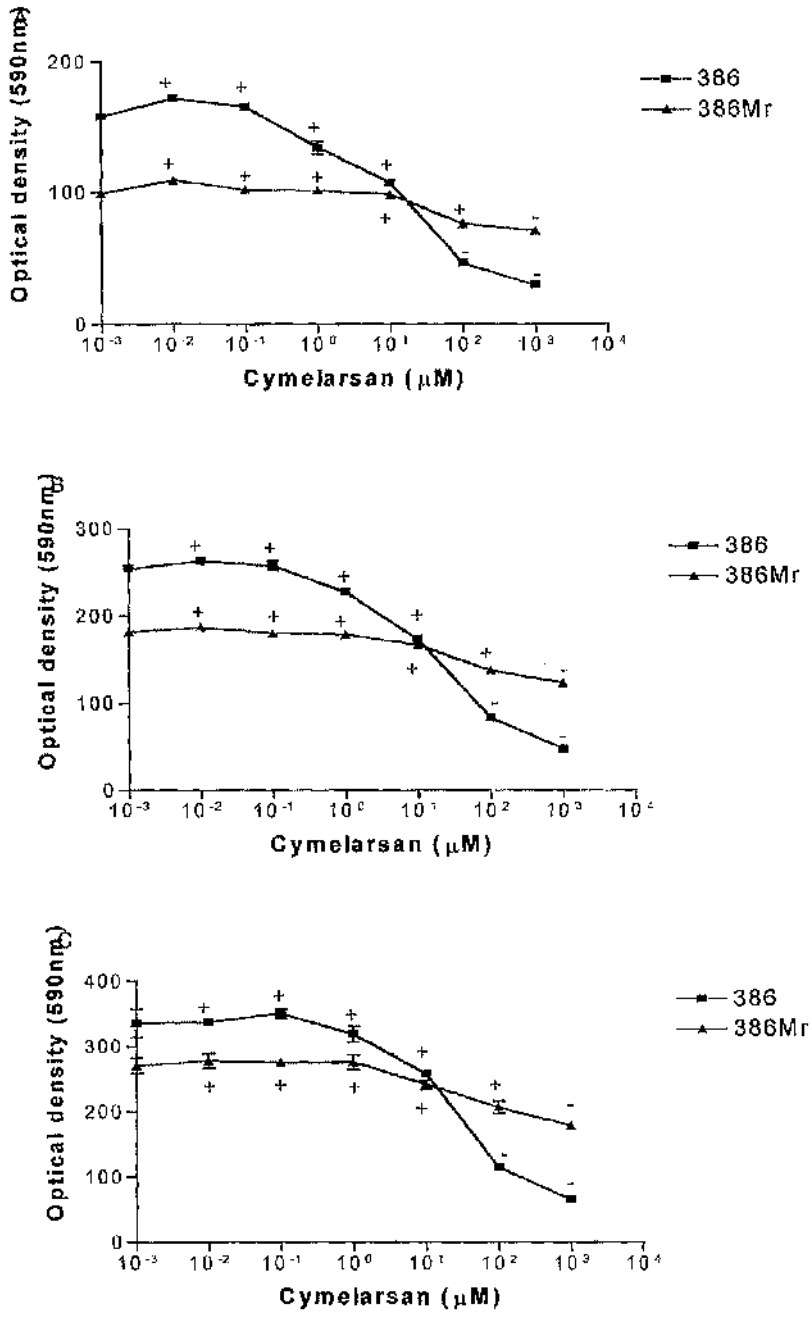


Figure 5.8: Alamar Blue assay on the isogenic 386 lines in the presence of Cymelarsan. Three concentrations of procyclic cultures were tested: $2x10^6$ (A), $4x10^6$ (B), $6x10^6$ (C) trypanosomes/ml. The status of the procyclic culture is indicated as alive (+) or dead (-) as assessed by microscopy. The lowest concentration corresponds to the control in the absence of drug. The error bars represent the mean +/- SEM, n = 4.

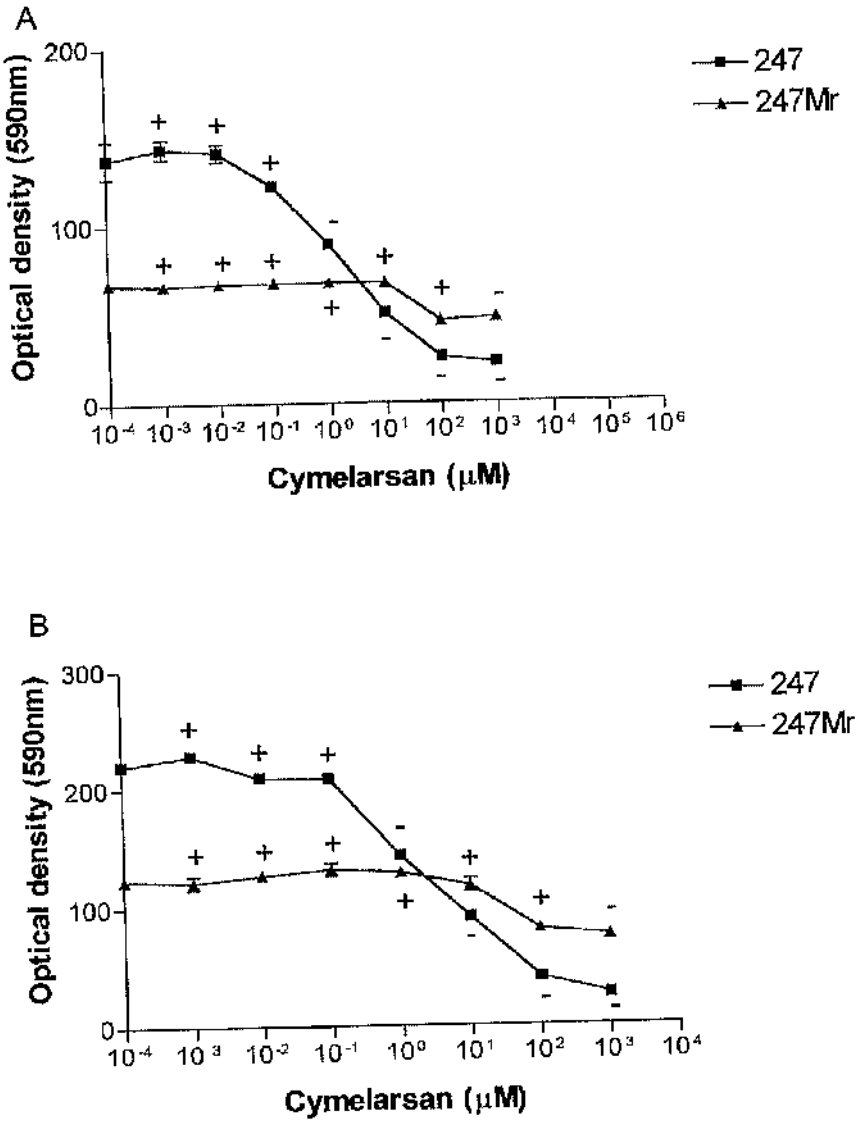


Figure 5.9: Alamar Blue assay on the isogenic 247 lines in the presence of Cymelarsan. Three concentrations of procyclic cultures were tested: $2x10^6$ (A), $4x10^6$ (B) trypanosomes/ml. The status of the procyclic culture is indicated as alive (+) or dead (-) as assessed by microscopy. The lowest concentration corresponds to the control in the absence of drug. The error bars represent the mean +/- SEM, n = 4.

increasing concentration of Cymelarsan. The 386 and 247 lines both exhibited a dose-dependent killing of the trypanosomes (Figures 5.8 and 5.9). However, the 386Mr and 247Mr lines showed very little dose-dependent killing of the trypanosomes. Microscopic examination of the culture confirmed the presence and viability of the Cymelarsan resistant lines in the culture medium for concentration ranging from 10^{-3} to 10^{2} µM of Cymelarsan.

These results suggest that the mechanism of reduction of the Alamar Blue dye has been modified during the selection for Cymelarsan resistance in *T. brucei* lines of 386 and 247. This unexpected result is investigated further in chapter 8.

5.4 Discussion

The main mechanism of arsenical resistance in trypanosomes has long been thought to be due to an alteration in uptake of the drugs mediated by the P2 transporter. However, no activity of the P2 transporter has been found in procyclic stages of trypanosomes (de Koning *et al.*, 2000). A recent study on field isolates from Uganda has demonstrated the presence of un-mutated *TbAT1* gene in 50% of the relapsed patients (Matovu *et al.*, 2001). In addition, the knockout of the *TbAT1* gene only conferred a three fold resistance to Melarsoprol (Matovu *et al.*, 2003), which is below the reported resistance observed in *in vivo* experiment in the two pairs of isogenic lines used in this study (MIC > 40 mg/Kg; Table 5.1). Consequently, the mechanism of resistance to the arsenicals drugs appears to be multi-factorial and necessitates to be investigated.

The procyclic forms of the laboratory-generated Cymelarsan resistant lines differ from the sensitive lines in their ability to survive in high concentration of Cymelarsan. Comparison of EC_{50} values calculated from growth inhibition assays evaluated the resistance as 25 fold in the 386 lines and 1,247 fold in the 247 lines. These values are different from the resistance levels described previously for procyclic forms of these lines which were a 500 fold increase in resistance for both pairs of isogenic lines (Scott *et al.*, 1996). The difference in resistance level could be explained by the use of a different assay to evaluate the EC_{50} values.

The use of growth inhibition assays with Melarsoprol, Pentamidine and Berenil demonstrated the absence of cross-resistance between Cymelarsan, Melarsoprol and the diamidines drugs. All four drugs are known to enter the trypanosomes via a nucleoside transporter, the P2 transporter, in bloodstream forms (de Koning, 2001). However, the P2 tranporter was shown to be a minor entry route for Pentamidine which can also be mediated by HAPT1 and LAT1 (de Koning, 2001), potentially explaining the lack of cross-resistance observed between Cymelarsan and Pentamidine in this study.

The lack of cross-resistance with Melarsoprol was expected in the assay of procyclics as Melarsoprol itself is not thought to enter the trypanosome via the P2 transporter despite the fact that this molecule possesses the recognition site for the P2 transporter (de Koning, 2001). Instead, Melarsoprol is thought to enter the trypanosomes via passive diffusion. In order to enter via the P2 transporter, Melarsoprol first requires to be metabolised into Melarsen oxide and this only happens *in vivo* (Keiser *et al.*, 2000). This explains why cross-resistance was

observed in the bloodstream forms of the two pairs of isogenic lines between Cymelarsan and Melarsoprol and not in procyclic forms (Scott *et al.*, 1996).

Two genes encoding P-glycoprotein were recently discovered in the genome sequencing project of *T. brucei* (Shahi *et al.*, 2002) with good homologies to both *Leishmania* and *T. cruzi* P-glycoprotein genes (Figure 5.10). The over-expression of one of the genes, *TbMRPA*, in line 927 was reported to confer a 10 fold increase in Melarsoprol resistance while over-expression of the second gene (*TbMRPE*) was associated with resistance to Suramin. From these data one hypothesis was that over-expression of the *TbMRPA* gene could account for some or all of the resistance observed in the lines studied here. Inhibition of P-glycoproteins using specific inhibitors demonstrated that an over-expression or alteration of this gene was not associated with Cymelarsan resistance in *T. brucei*.

The Alamar Blue assay has been recently shown to enable the determination of sensitivity of *T. b. gambiense* and *T. b. rhodesiense* to Melarsoprol, DFMO, Pentamidine and Suramin (Räz *et al.*, 1997). The assay is based on the reduction of the Alamar Blue dye, by intracellular enzymes, which correlates with cell growth and metabolic activity in the presence/absence of drug. The use of the *in vitro* Alamar Blue assay on each pair of isogenic lines in the absence of drug has demonstrated the reduced fluorescence obtained from the dye by the Cymelarsan resistant lines. In addition, the expected dose-dependent response to Cymelarsan, corresponding to a loss of viability, was only observed for the sensitive lines. No change in the fluorescence level due to the reduction of Alamar Blue was observed for the resistant lines with increasing level of Cymelarsan and loss of viability. The alteration in the fluorescence

level produced by the Cymelarsan resistant line could be due to a defect in the activity of the intracellular enzymes or an alteration in the transport/export of the Alamar Blue dye. The analysis of these two hypothesc has been undertaken and the results are presented in chapter 8.

The Cymelarsan resistance observed in the two pairs of isogenic lines was not associated with either a lack of expression of the *TbAT1* gene or over-expression or alteration of a P-glycoprotein. Consequently, the resistance mechanism in the procyclic forms of these lines does not follow the conventional pattern described so far in the literature. Other genes/proteins must be involved in the Cymelarsan resistance which have so far not been identified.

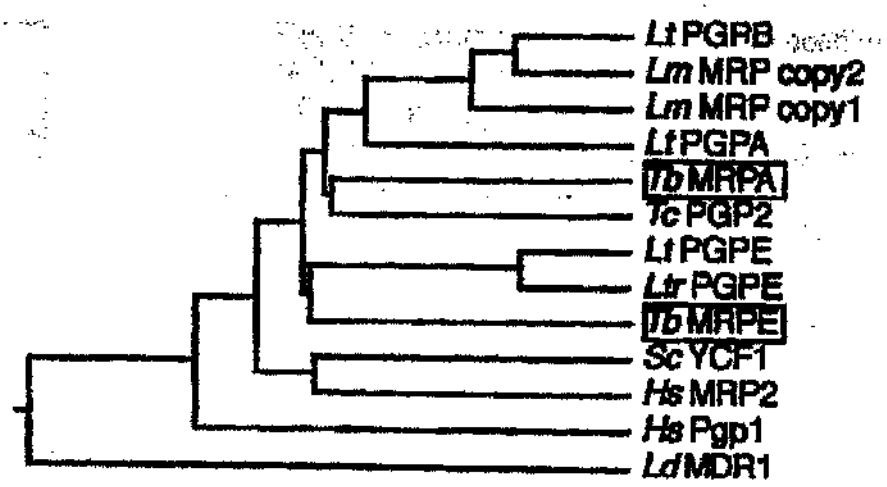


Figure 5.10: Evolutionary tree showing selected multidrug resistant proteins (MRP) and P-glycoproteins (Pgp). Species are *Homo sapiens* (Hs), *Leishmania tarenolae* (Lt), *Leishmania tropica* (Ltr), *Leishmania donovani* (Ld), *Leishmania major* (Lm), *Saccharomyces cerevisiae* (Sc), *Trypanosoma brucei* (Tb) and *Trypanosoma cruzi* (Tc). The two genes encoding MRPs from *T. brucei* are boxed. (Reproduced from Shahi *et al*, 2002)

Chapter VI

A proteomic approach to study Cymelarsan resistance in *T. brucei* line 386

6.1 Introduction

Analysis of the isogenic sensitive and resistant lines of 386 demonstrated that both procyclic and bloodstream forms exhibited the Cymelarsan resistance phenotype (Scott *et al.*, 1996). The cross-resistance analysis in procyclic forms of the 386 lines ruled out the involvement of the P2 transporter in the mechanism of arsenical resistance (chapter 5). In addition, the inhibition of P-glycoproteins had no effect on the resistance phenotype of the 386Mr line, arguing against the hypothesis that increased efflux of the drug by an over-expressed P-glycoprotein could explain the Cymelarsan resistance. With the increase number of reported cases of Melarsoprol refractory parasites in the field (Stanghellini and Josenando, 2001; Legros *et al.*, 1999; Brun *et al.*, 2001; Matovu *et al.*, 2001) it is imperative to determine the genes/proteins involved in the mechanism of arsenical resistance. The well-characterised laboratory derived isogenic lines of 386 were used as a model to determine the mechanisms of drug resistance.

There are a number of approaches to identify the genes that determine resistance to arsenical drugs in *T. brucei*. These include genetic analysis/positional cloning (Tait *et al.*, 2002), biochemical analysis (Fairlamb, 2002), cloning and functional analysis of candidate genes (Mäser *et al.*, 1999) and phenotypic analysis of isogenic parasites lines which are resistant or sensitive to the drug. The particular approach taken depends to a large extent on the hypotheses as to the mechanism of drug resistance. In the case of laboratory-selected lines, it is likely that multiple mechanisms occur to confer the final resistance

phenotype. In principle, a range of mechanisms could be put forward that include point mutations or deletions in genes involved in drug transport, in the drug target or involved in drug degradation, as well as alterations in the level of expression of any of these genes. The advantages of using either a proteomic or microarray approach in this context are that, in principle, all the possible mechanisms could be identified, although microarray analysis will only identify alterations in gene expression at the transcriptional level.

A proteomic approach, assessing the level of protein expression in a pair of isogenic lines was preferred to a micro-array approach. Measured mRNA levels are often implicitly or explicitly extrapolated to indicate levels of activity of the corresponding protein in cell. Even a bacterium with a small genome such as *Mycoplasma genitalium* had 24% more proteins expressed than the number predicted on a simple one gene, one protein system (Wasinger *et al.*, 2000). Similarly, a correlation coefficient of 0.48 was obtained between the mRNA and protein abundance determined by 2DE and gel imaging of human tissues (Anderson and Seilhamer, 1997). This correlation coefficient was confirmed using techniques such as immunoaffinity-HPLC and Northern analysis (Anderson *et al.*, 1998). These proteomic analyses illustrated the strength of 2DE analysis over micro-array technology. The low correlation coefficient obtained is probably due to the fact that the rate of degradation of individual mRNAs and protein differ. Proteins are very often the target of drugs and so measuring proteins directly gives a more accurate picture of the biology of a cell under various conditions. Proteome profiles can be scanned and quantitated to search for protein differences such as changes in levels of pre-existing proteins, induction of new products and alterations in amino acid sequence. 2DE is a relatively simple and reproducible technique. Proteome analysis provides data that are of

critical importance for the description of the state of a biological system and that are not readily apparent from the sequence and the level of expression of mRNA transcripts.

While the number of global proteomic expression profiling studies with pathogens such as bacteria and viruses are numerous, the application of these methodologies to protozoan parasites has lagged behind. This can be attributed to a number of causes, one of these is that protozoan parasites tend to be difficult to culture and manipulate *in vitro*, making it difficult to obtain enough pure material for proteomic analysis. However, procyclic forms of *T. brucei* can be cultured relatively easily *in vitro* and a reproducible solubilisation protocol for 2DE analysis of proteins from trypanosomes has been developed in this study (chapter 2). This enables a proteomic expression profiling study to be carried out and a comparative analysis made between Cymelarsan sensitive and resistant isogenic lines.

Differentially expressed protein spots can be analysed by mass spectrometry methods enabling protein identification. In the case of *T. brucei*, approximately 40% of protein spots are predicted to be identifiable using the simple technique of PMF with the current level of annotation of the *T. brucei* genome (chapter 4). For protein spots which cannot unambiguously be identified by PMF, Tandem MS can be used to generate tryptic peptide sequence information that would either confirm tentative MALDI-TOF identification or more readily identify the gene if it is present in the database. In order to analyse the basis of Cymelarsan resistance and identify pathways involved in the response to drug treatment, the strategy illustrated in Figure 6.1 was used.

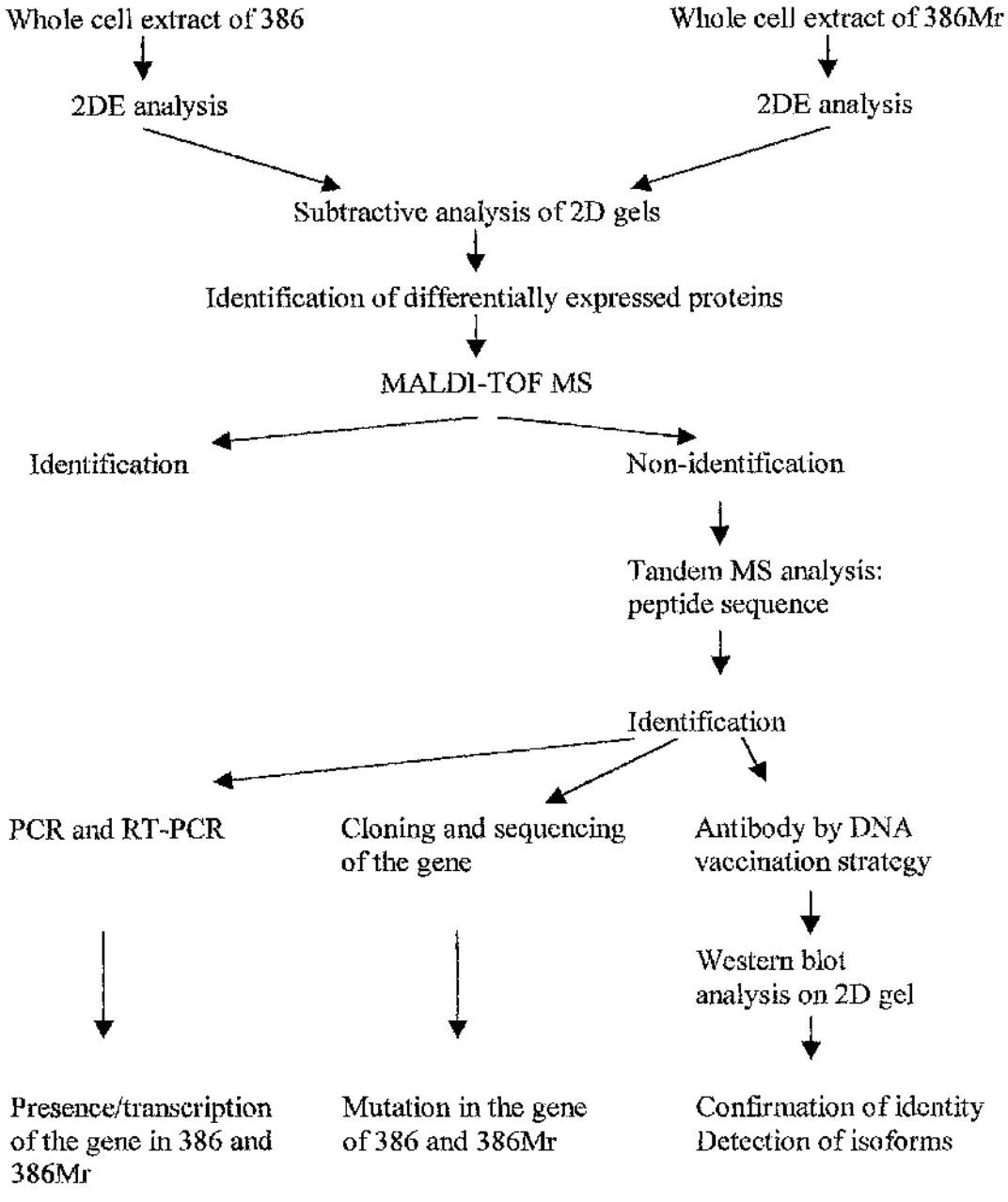


Figure 6.1: Schematic representation of strategies employed to determine protein associated with arsenical resistance in isogenic 386 lines. Two mass spectrometry approaches were used to increase the likelihood of identification of the differentially expressed protein. Mass spectrometry data were searched against the NCBInr and the local *T. brucei* database using the MASCOT® search engine. The corresponding genes were then further analysed to assess their presence/absence, transcription and mutations in the lines.

In this chapter, proteomic techniques were used to compare whole soluble protein expression patterns between Cymelarsan sensitive and resistant 386 lines and so detect variation in protein expression associated with arsenical resistance. In addition, these protein expression patterns were also compared with those of the Cymelarsan resistant 386 line growing in 1 μM Cymelarsan to determine whether some protein expressions were induced in the presence of Cymelarsan. Differentially expressed protein spots were identified using mass spectrometry techniques and the corresponding genes identified. These genes were further analysed to define the alterations associated with arsenical resistance.

6.2 Materials and Methods

6.2.1 Commonly used reagents

The QIAGEN plasmid minikit, QIAquick gel extraction kit, EndoFree plasmid Maxi kit and Ni-NTA agarose were all purchased from Qiagen. The competent cells (Top 10, DH5α and BL21AI *E. coli* cells), Rnase Out, Sperscript RT, L-arabinose and the TOPO vector were purchased from Invitrogen. The enzymes, EcoRI, HindIII, BamHI and BsaJI were purchased from New England BioLab. The antibiotics (Ampicillin and Kanamycin), X-gal, Chloroform, Isopropanol, DEPC, formaldehyde and formamide were purchased from SIGMA. Trizol, pDONR211 plasmid and pDEST17 vector were purchased from GIBCO BRL®. Custom PCR MasterMix and the Taq polymerase were purchased from ABgene. Microspin TM S 200HR, Hybond-N nylon membrane, X-ray films, pre-cast 12% polyacrylamide 24 cm gels and the DALT II vertical tank were purchased from Amersham Biosciences. Oligonucleotides were synthesised by MWG Biotech.

6.2.2 *In vitro* culture of procyclic trypanosomes

The *in vitro* procyclic cultures of Cymelarsan sensitive and resistant lines of 386 in SDM79 were performed as described in chapter 2. In addition, the Cymelarsan resistant 386 line was grown in complete SDM79 medium supplemented with 1 µM Cymelarsan solution diluted in complete SDM79 medium.

6.2.3 Protein extraction, 2DE and mass spectrometry techniques

6.2.3.1 Protein extraction and 2DE

Proteins extraction from the procyclic cultures was performed as described in chapter 2 (protocol P2) except that 470 µl of solubilisation buffer was used instead of 350 µl. 24 cm strips were used in this study to separate the proteins in the IEF. Rehydration of the strip was performed using 470 µl of protein extract and separation by IEF was performed as described in chapter 2, except that the last separation step at 8,000 V was increased to 8 h. The second dimension of the 2DE was run with 24 cm, 12% polyacrylamide pre-cast gels in a DALT II vertical tank according to the manufacturer's instructions.

6.2.3.2 MALDI-TOF and Tandem MS

Trypsin digestion of protein spots of interest and peptide mass fingerprinting by MALDI-TOF MS were carried out as described in chapter 3.

For Tandem MS, the protein spots were digested the same way as for peptide mass fingerprinting. The samples were purified and concentrated using ZipTips as previously described (chapter 3), except that the TFA solution was replaced by formic acid (10%). 1 µl of the concentrated sample was loaded into a capillary needle (Protana). The capillary

needle was linked to the Tandem MS and voltage was applied to enable the spray of the sample (nano-electrospray).

The Tandem mass spectrometer used in this study (Q-STAR) contains two analysers, separated by a collision cell. Peptide mass fingerprints were acquired in the first analyser (TOF-MS scan), the high intensity peptides were then specifically selected, fragmented in the collision cell and the mass of the fragment ions were measured in the second analyser, producing peptide sequence information.

The Tandem mass spectra were acquired in the following way. The TOF-MS scan was run for 10 s and detected doubly, triply and quadruply charged ions whose mass to charge ratio (m/z) was between 400 and 1,000. From this, mass spectra were acquired using a data-dependent acquisition mode in which each full scan mass spectrum generated by the TOF-MS was followed by collision-induced dissociation of the eight most intense parent ions whose signal had reached a threshold of 20 counts/s. Acquisition was performed with the dynamic exclusion of m/z ratios of already fragmented ions. The fragmentation was performed in positive polarity using nitrogen as the collision gas. Mass spectra were acquired for 15 s in the range of 50 to 2,000 m/z. The resulting peptide MS/MS spectra were interpreted using the MASCOT® search engine to search the NCBInr and a local *T. brucei* database to identify the genes encoding the particular protein. Partial carboxymethylation of cysteine residues and oxidation of methionine residues were considered in the search.

6.2.4 Molecular biology methods

6.2.4.1 Sequencing

Once the sequence of the gene encoding the protein of interest was identified, primers were designed to allow amplification of the complete coding sequence. Each ORF was cloned using primers 274F/274R and 277F/277R (Figure 6.2, Table 6.1). PCR was performed, on genomic DNA from 386 and 386Mr lines, in 40 µl reaction volume containing, 4 µl of Custom MasterMix buffer, 5 µM of each primer, 200-300 ng of DNA template and 4 units of Taq polymerase. Amplification was performed on a robocycler (Stratagene) by denaturing at 95°C for 50 s, annealing at 63°C for 50 s and extension at 68°C for 2 min for 30 cycles followed by incubation at 68°C for 15 min. The PCR product was run on a 1% agarose gel and extracted from the gel using the QIAquick™ gel extraction kit, according to the manufacturer's instructions. Cloning was performed using the TOPO vector and competent *E. coli* TOP 10 cells. Insertion of the PCR product into the TOPO vector was performed by mixing 1 ng of TOPO vector with the PCR product (1 µl) in water (final volume of 5 µl) and incubating at room temperature for 5 min. 1 µl of this cloning reaction was added to a vial of competent *E. coli* TOP 10 cells and incubated on ice for 30 min. Transformation of the competent cells was performed using the heat shock method (42°C for 30 s followed by 4°C for 2 min). Luria Bertani (LB) medium (250 µl) was then added to the vials which were incubated at room temperature for a further 30 min. Selection of recombinant *E. coli* cells was undertaken on LB agar plates containing Ampicillin (0.4mg/ml) and X-gal (5 µg/ml, 5-bromo-4-chloro-3-indolyl-β-D-galactoside) at 37°C overnight. Single colonies were picked and grown in 10 ml of LB medium containing Ampicillin (0.4 mg/ml) at 37°C overnight. *E. coli* cells from 3 ml of culture were concentrated in an Eppendorf by centrifuging at 12,000 g for 3 min. The TOPO vector

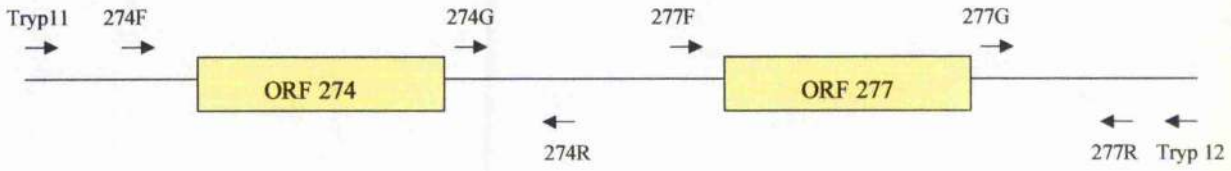


Figure 6.2: Location of primers use for the analysis of ORF 274 and 277. Schematic representation of the two ORFs encoding the CAR protein, with the location of the primers designed for RFLP analysis (Tryp11/Tryp12), cloning (274F/274R and 277F/277R) and Northern analysis (274G/274R and 277G/277R).

Name of primer	Sequence	Annealing T°C/Time	Elongation T°C/Time
Tryp9/11	5'-GAGGGAGGGGGAATTTCCTC	59°C/50s	68°C/ 2min30
Tryp9/12	5'-GAGCCATGACGAATCTTTTG	59°C/50s	68°C/ 2min30
274-F	5'-CAGCGAATTTATCTACATGC	63°C/50s	68°C/ 2min
274-R	5'-TGGCCACGGAAACACTCCTA	63°C/50s	68°C/ 2min
274-G	5'-GTGTAAGTGAAACCCCCAAT	63°C/50s	68°C/ 2min
277-F	5'-GTGCTACTACATTCGCCTCA	63°C/50s	68°C/ 2min
277-R	5'-CATCACAATGGATGGAGGAT	63°C/50s	68°C/ 2min
277-G	5'-CGATTACCGAGGTGAAAAGT	63°C/50s	68°C/ 2min
NAC Ab-F	5'-GTAACCCGGAAGCTTACGATGGGCGCTAAC	60°C/50s	68°C/ 2min30
NAC Ab-R	5'-GATCAGTTGGGATCCTAGGTCAGTTCC	60°C/50s	68°C/ 2min30
AttB1 F	5'-GGGACAAGTTTGTACAAAAAAGCAGGCTCGATG AGCGCTAACGATGTTCAA	68°C/50s	68°C/ 2min30
AttB2 R	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATCAG TTGGGATCCATCGTCAG	68°C/50s	68°C/ 2min30
TIM-A	5'-GCGTAGTGGCCTCCACCTTTG	63°C/50s	68°C/ 2min
TIM-H	5'-TGCCGTTGAGTGGGTGAAGATAGC	63°C/50s	68°C/ 2min
BJS 10	5'-GGGCAAACAACAGAT		
BJS 11	5'-CGCAAATGGGCGGTAGGCGTG		

Table 6.1: List of primers for the analysis of ORF 274 and 277. Sequence of primers used for the analysis of ORF 274 and ORF 277, and the production of anti-CAR antibodies, with the amplification conditions are given. Denaturation for all PCR was performed at 95°C for 50 s.

containing the insert was extracted from the *E. coli* clone using the QIAGEN plasmid minikit according to the manufacturer's instructions. To assess the presence of the insert in the TOPO vector, 3 µl of plasmid was digested with EcoRI (250 U) in 15 µl of ddH_2O containing 2 µl of EcoRI buffer at 37°C for 2 h. Digestion products were analysed by agarose gel electrophoresis on a 1% agarose gel. Plasmids containing the correct size insert were precipitated using 3 M Sodium acetate ($1/10^{th}$ volume) and 100% ethanol (2 volumes) by incubating at -20°C overnight (Sambrook *et al.*, 1989). The selected clones and the primers were then sent for sequencing to the MBSU of the University of Glasgow.

6.2.4.2 Analysis of transcription

6.2.4.2.1 RNA extraction

Extraction of RNA from procyclic culture was performed under sterile conditions in a culture hood. 27 ml of procyclic cultures at $7x10^6$ trypanosomes/ml were centrifuged for 8 min at 12,000 g to collect $1x10^8$ trypanosomes. The pellet was washed once in PSG buffer and then re-suspended in 1 ml of Trizol. Following 5 min incubation at room temperature, 0.2 ml of chloroform was added to the solution, the tubes were shaken vigorously and then incubated for a further 2 min at room temperature. The phases were separated by centrifuging at 4°C for 15 min at 12,000 g and the aqueous phase was transferred into a clean Eppendorf to which 0.5 ml of isopropanol was added, mixed well and incubated at room temperature for 10 min. The RNA was pelleted by centrifuging the tubes at 4°C for 10 min at 12,000 g, and then washed in 75% ethanol prepared in diethyl pyrocarbonate (DEPC) treated water. The pellet was air-dried, re-suspended in 100 µl of DEPC treated water and treated with DNA-free™ kit (Ambion) according to the manufacturer's instructions.

6.2.4.2.2 RT-PCR

Total complementary DNA (cDNA) was prepared from RNA dissolved in DEPC treated water using an oligo $(dT)_{15}$ primer for first strand synthesis. The total cDNA from the RNA was produced by incubating the PCR tubes containing the RNA on a robocycler (Stratagene) at 70°C for 10 min and then at 4°C for 1 min. To each tube, 4 µl of 5x first strand buffer, 0.01 M of DTT, 10 µM of dNTP mix and 1 µl of Rnase out was added. After the tubes were incubated at 42°C for 5 min, 1 µl of Superscript RT was added. The solution was covered with 6 µl of mineral oil to prevent evaporation of the solution. Tubes were incubated at 46°C for 50 min, followed by 70°C for 15 min and then stored at -20°C. This total cDNA was then used as a DNA template in PCR reaction using primers 274F/274R and 277F/277R to assess the transcription of ORF 274 and 277. The amplification was performed as described in section 6.2.4.1 except that the last elongation step (68°C for 15 min) was removed.

6.2.4.2.3 Northern blot analysis

The RNA extracted from trypanosomes was run on a 6% formaldehyde gel. Gels were made by mixing 2% melted agarose with 12% formaldehyde in 20x Northern gel buffer (0.036M Na_2HPO_4, 0.004M NaH_2PO_4). 3 µl of the RNA samples were mixed with 3.5 µl of formaldehyde, 10 µl of formamide, 1 µl of 20xNorthern gel buffer and double distilled water to a final volume of 20 µl. The samples were incubated at 55°C for 15min and chilled on ice. 2 µl of RNAse-free loading buffer (50% glycerol, 1 mM EDTA pH8, 0.25% Bromophenol blue, 0.25% Xylene cyanol FF) was added to the samples, and the samples were loaded onto the gel. The gel was run at 60 V for 6 h and then incubated in 20x SSC (0.3 M Tri-sodium citrate, 3 M sodium chloride, pH 7.0) at room temperature for 30 min. The RNA was transferred onto a Hybond-N nylon membrane at room temperature overnight (Sambrook *et al.*, 1989). Cross-linking of the RNA onto the membrane was performed in a Spectrolinker XL-100 UV crosslinker (Spectronics Corporation). The

membrane was incubated in the pre-hybridisation solution (7.5 ml formamide, 3.75 ml 20x SSC, 3 ml 50x Dendhart, 0.1 ml NaH_2PO_4 pH 6.5, 0.3 ml HS DNA, 0.15 ml 10% SDS and 0.15 ml 0.5 M EDTA) at 42°C for a minimum of 4 h. The probes were generated by PCR amplification of genomic DNA from the Cymelarsan sensitive 386 line, using primers 274G/274R, 277G/277R and TIM-A/TIM-F (Figure 6.2, Table 6.1). The PCR products were run on a 1% agarose gel and the fragment corresponding to the correct size product were recovered from the gel using the QIAquick™ gel extraction kit according to the manufacturer's instructions. The PCR products were labelled with α ^{32}P (NEN Life Science) using the Stratagene Prime-IT II random primer labelling kit according to the manufacturer's instructions. The labelled probe was purified using a Microspin TM S 200HR column according to the manufacturer's instructions. The probe was denatured for 5 min at 95°C, before being added to the hybridisation solution. Hybridisation was allowed to take place overnight at 42°C. The probe was then removed and the membrane washed in 0.1% SDS and 0.1xSSC at 57°C for 3 cycles of 30 min. The membrane was exposed to a X-ray film for 7 days at -80°C and the autoradiographs were developed using a Compact X4 film processor (X-graph Imaging Systems) according to the manufacturer's instructions.

6.2.4.3 Production of antibody

Antibodies to the expressed product of ORF 274 were produced in mice using a DNA vaccination strategy (Figure 6.3). Mice were first injected with the pcDNA3.1(+) plasmid containing ORF 274 and then with a recombinant protein produced from ORF 274 to boost the antibody response.

6.2.4.3.1 Cloning in pcDNA3.1(+)

The pcDNA3.1 plasmid was donated by Dr Gill Douce (University of Glasgow). The forward primer was designed from the sequence of ORF 274 and modified to contain the Kozak translation initiation sequence, an ATG codon and a HindIII restriction site. The reverse primer was designed from the sequence of ORF 274 and modified to contain a stop codon and a BamHI restriction site, to facilitate cloning in the right open reading frame and forward orientation (Figure 6.4). PCR was performed on genomic DNA from line 927 using these primers in a 40 µl reaction as previously described. The PCR products and pcDNA 3.1 (+) plasmid were digested using HindIII and BamHI restriction enzymes by incubating at 56°C for 1 h. Transformation of Top 10 *E. coli* competent cells with the plasmid and PCR product was carried out as previously described. Single colonies of the transformed *E. coli* cells were picked, grown up in LB medium and the plasmid extracted as previously described. The plasmid from these clones was sequenced by the MBSU of the University of Glasgow, using the BJS 10 and BJS 11 primers (Table 6.1), to determine whether the insert was in frame. One clone, with the insert in frame and in the right orientation, was selected and grown in 100 ml of LB medium containing Ampicillin (0.4 mg/ml). Plasmid DNA was extracted from a 100 ml culture using the EndoFree® plasmid Maxi kit, according to the manufacturer instructions. The purified plasmid was given to Dr Gill Douce (University of Glasgow) for coating onto gold particles and injection into mice by gene-gun technology.

6.2.4.3.2 Expression of recombinant protein in *E. coli* (Gateway system)

The Gateway Cloning Technology is a universal cloning system that takes advantage of the site-specific recombination properties of bacteriophage lambda (Landy, 1989) to provide a rapid and highly effective way to move a gene of interest into multiple vector systems. The Gateway Cloning Technology was used for cloning PCR products flanked by an *att*B

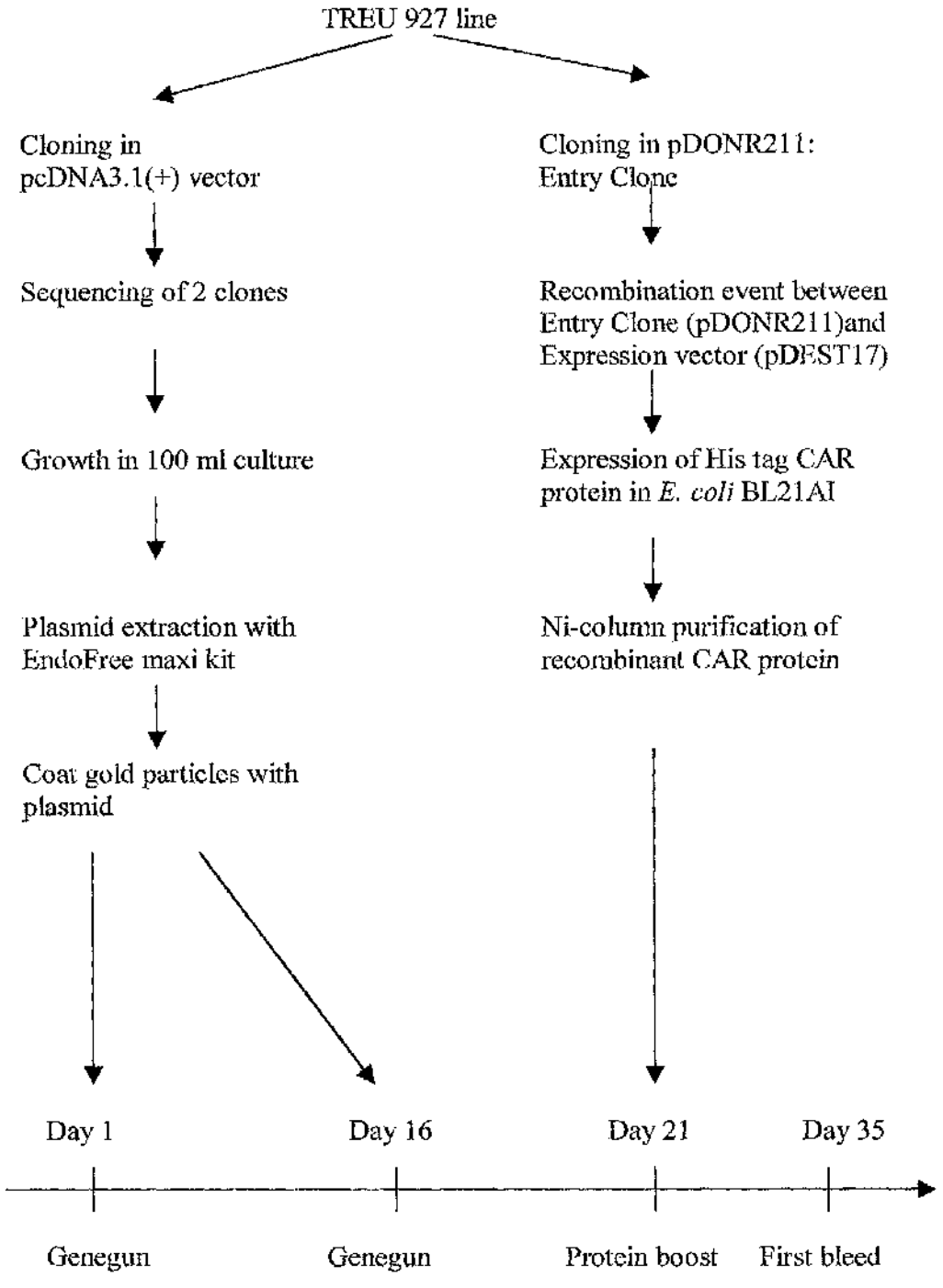


Figure 6.3: **The genegun technology**. Schematic representation of the strategy used to produce antibody against Tryp9.0.000322-274 (CAR gene) from the reference line 927. The mice was injected on day 1 and day 16 with gold particles coated with the plasmid pcDNA3.1(+) containing the Kozak initiation sequence and ORF 274 in frame with the T7 promoter and on day 28 with a purified N terminal His tagged CAR protein.

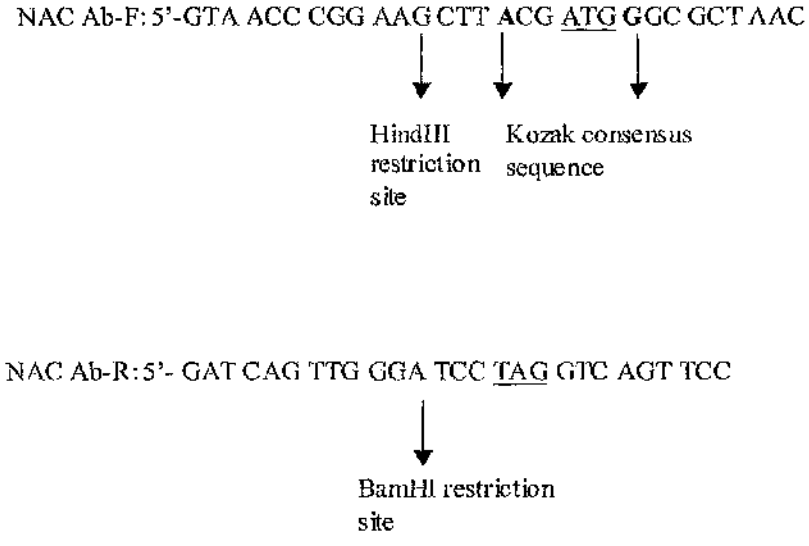


Figure 6.4: Primers for cloning in plasmid pcDNA3.1(+). Sequences of the forward and reverse primers designed to clone ORF 274 in plasmid pcDNA3.1(+) for the production of antibodies in mice. The forward primer contains the start codon (underlined), a HindIII restriction site (blue) and the Kozak translation initiation sequence (bold). The reverse primer contains the stop codon (underlined) and a BamHI restriction site (blue).

recombination sites to generate Entry Clones. Primers were designed to contain the *attB* recombination sites in frame with the start codon (forward primer) and the stop codon (reverse primer). Amplification using these primers was performed on 200-300 ng of DNA from line 927 in a 40 µl reaction, using the conditions described in Table 6.1. The PCR products were run out on a 1% agarose gel and the band corresponding to the predicted size of ORF 274 was extracted from the gel using the QIAquick™ gel extraction kit and purified using 30% PEG 8000/ 30 mM $MgCl_2$. The Entry Clone was generated using the purified PCR product and the pDONR211 plasmid with insertion into the plasmid being undertaken according to the manufacturer instructions. This plasmid was used to transform DH5α competent cells and recombinant cells were selected on LB agar plates containing Kanamycin (50 µg/ml). Single colonies were picked and grown in LB medium containing Kanamycin (50 µg/ml), and the plasmid was extracted from 3 ml of culture using the QIAGEN plasmid minikit.

The Entry Vector was generated by a recombination event between the pDONR211 plasmid, containing ORF 274 from trypanosome line 927, and the pDEST 17 vector containing a T7 promoter and 6 histidines at the N terminus. The recombinant pDEST17 vector was introduced into BL21AI competent cells. Recombinant cells were selected on LB agar plates containing Ampicillin (100 µg/ml) by growth at 37°C overnight. Single colonies were then grown in LB medium at 37°C. The recombinant protein was expressed by growing the transformed BL21AI cells on LB medium containing Ampicillin (100 µg/ml) at 37°C for 3 h on a shaker. Expression of the recombinant protein was induced by addition of L-arabinose in the culture medium and incubation at 37°C for 4 h. The cells were pelleted (4°C for 15 min at 12,000 g) and stored at -20°C. Recombinant protein, from the induced BL21AI cells, was extracted by sonicating the cells in Binding buffer (10 mM

Tris-HCl pII 7.8, 5 mM Imidazole, 0.5 M NaCl, 10% glycerol) on ice 20 times for 30 s. The recombinant protein was purified from the extract using affinity chromatography on Ni-NTA agarose according to the manufacturer's instruction. The purity of the eluted protein was assessed by running the fractions on a 12% SDS-PAGE gel. The gel was stained with Coomassie Brilliant blue.

Antiserum to the recombinant protein was obtained from female mice immunised by a genegun injection of plasmid pcDNA 3.1 (+) containing ORF 274 (day 1 and 16) followed by intraperitoneal innoculation of a booster of 0.2 mg/ml recombinant protein with 0.1 ml of complete Freund's adjuvant. All injections and bleeding of the mice were performed by Dr Gill Douce. A first bleed was obtained after 35 days, and the production of antibody tested by Western dot blot analysis against the recombinant CAR protein (data not shown). The final bleed was made on day 45.

6.2.4.3 Western blot analysis

Western blot analysis of proteins separated onto SDS-PAGE gels were carried out as described in chapter 2.

6.3 Results

6.3.1 Comparative proteomic analysis of Cymelarsan sensitive and resistant *T. brucei* using broad pH ranges

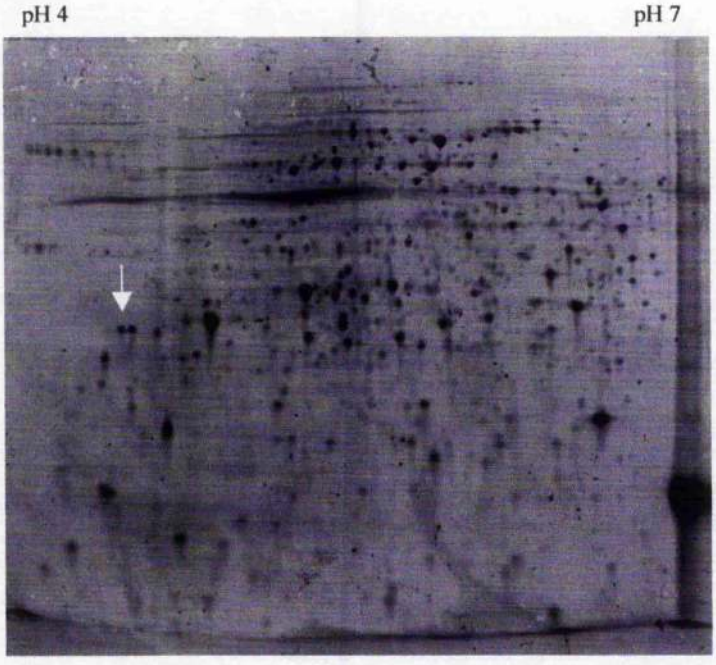
In order to compare the protein expression profiles of Cymelarsan sensitive and resistant lines of 386, soluble proteins extracted from procyclic cultures were separated by 2DE. Proteins extracted from $1x10^8$ procyclic form parasites were separated in the first

dimension according to their isoelectric point, on IPG strips (pH 4-7), and then by molecular weight on a 12% polyacrylamide gel. After staining the gels with Sypro ruby, they were compared by subtractive analysis, using the ImageMaster 2DElite software. The comparison showed that, out of 485 proteins separated on the 2D gels, there was only one protein spot (spot 2) differentially expressed between the lines (Figure 6.5). The protein spot present in the soluble protein extract of 386, at about pI = 4.5, was absent from the extract of the 386Mr line. This protein spot was located adjacent to a protein spot, present in both lines, with approximately the same molecular weight but a more acidic pI (spot 1).

In order to ensure that the difference observed between 386 and 386Mr lines was not due to minor differences in sample preparation, further protein extractions were made from procyclic form parasites. These were separated by 2DE, as before, and stained with either Sypro ruby or Coomassie Brilliant blue. Nine independent soluble trypanosome protein extracts were run on 2D gels, all of which consistently revealed the absence of the protein spot around pH 4.5 in the 386Mr line (Figure 6.6A and 6.6B). This confirmed that the differentially expressed protein spot was not due to sample preparation. In addition, the missing protein spot could not be detected in the 386Mr protein extracts using either the insensitive Coomassie Brilliant blue staining or the sensitive Sypro ruby staining. This suggests that the observed difference was not an artifact due to variation in sample loading.

2D gels were run using soluble protein extracts from procyclic forms of 386Mr line growing in 1 μM of Cymelarsan in SDM79 (Figure 6.6C). These gels also revealed the absence of the protein spot around pH 4.5 in the 386Mr line growing in the presence of the

A)



B)

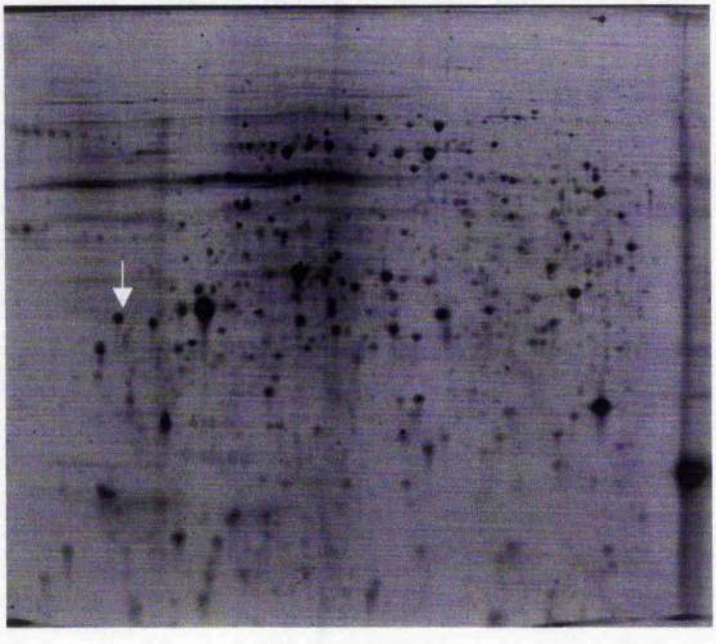
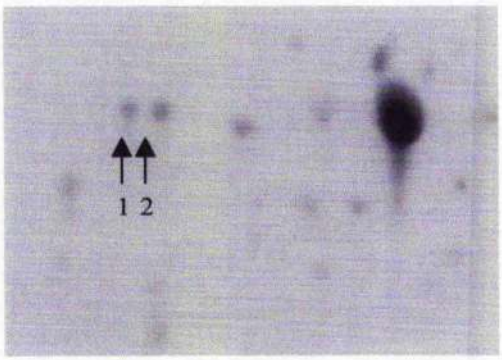
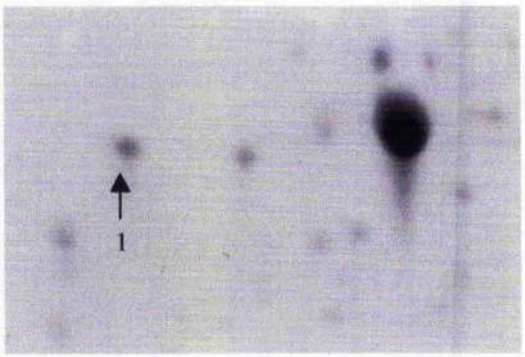


Figure 6.5: **2D gels of procyclic forms of isogenic 386 lines, pH 4-7**. Broad pH range isoelectric focusing (pH 4-7) 2D gels for 386 (A) and 386Mr (B) lines. Soluble proteins extracted from $1x10^8$ procyclic trypanosomes were loaded onto the gel. The gels were stained with Sypro ruby. Image analysis using the 2DElite software revealed over 485 distinct protein spots. One protein spot indicated by an arrow was differentially expressed between the lines.

A)



B)



C)

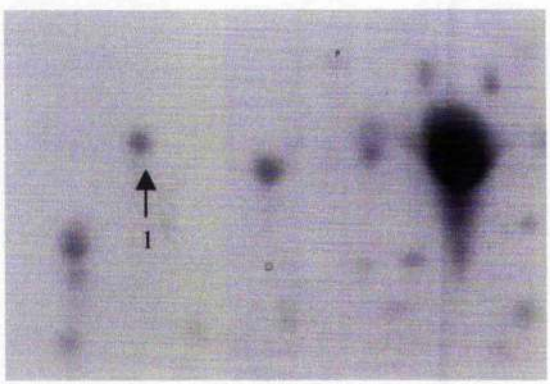


Figure 6.6: **Close-up of the differentially expressed protein spot**. Close-up of a section of the gels between pH 4.4 and 4.6 from 2D gels on broad pH range IEF (pH 4-7) for 386 (A), 386Mr (B) and 386Mr growing in the presence of 1 μM Cymelarsan (C). Soluble proteins, from $1x10^8$ parasites grown in SDM79, were separated on pH 4-7 2D gels and stained with Coomassie Brilliant blue. An arrow indicates the differentially expressed protein spot (spot 2).

drug. No other difference in the 2D protein pattern was observed in over 485 protein spots separated by 2DE and visualised by Sypro ruby staining.

6.3.2 Comparative proteomic analysis of Cymelarsan sensitive and resistant *T. brucei* using narrow pH ranges

With complex samples, such as the whole cell extract employed in this study, 2DE analysis on wide-range pH gradient reveals only a small percentage of the whole proteome because of insufficient spatial resolution. The use of narrow range focusing gels can provide increased resolution of proteins with similar electrophoretic mobilities. Zoom-in gels in the pH range between 4 and 7 (IPG 4-5, 4.5-5.5, and 5.5-6.7) were used in the first dimension separation to increase the resolving power of 2DE on whole cell extracts from procyclic forms of the two isogenic lines.

6.3.2.1 Analysis of Cymelarsan resistance on pH 4-5

To determine whether other proteins were differentially expressed between the pair of isogenic lines, whole soluble protein extract from procyclic forms were separated on a narrow pH range gel of pH 4 to 5. The gels were stained with Coomassie Brilliant blue, compared and one differentially expressed protein spot detected between the lines. One protein spot (at approximately pI = 4.5) on the 2D pattern of proteins from the 386Mr line was missing when compared with the 2D gel separation from the 386 line. On the gel from the 386 line this protein spot was located adjacent to a second protein spot, present in both lines, which had approximately the same molecular weight but a more acidic pI (data not shown).

6.3.2.2 Analysis of Cymelarsan resistance on pH 4.5-5.5

Proteins extracted from $1x10^8$ procyclic stage trypanosomes were separated on narrow isoelectric focusing, pH 4.5 to 5.5 strips, prior to separation according to their molecular weight on 12% polyacrylamide gels. Analysis of the Coomassie Brilliant blue stained gels using the ImageMaster 2DElite software revealed almost 200 spots on each gel (157 spots on the 386 gel, 198 spots on the 386Mr) (Figure 6.7A and 6.7B). The protein patterns were very similar for the two lines. The variation in the number of protein spots detected was due to very faint spots which were not detected in the extract from 386. As this was only observed for the faint spots it is likely that these differences were due to a difference in quantity of protein loaded rather than real variation between the samples associated with Cymelarsan resistance. This was confirmed by the fact that the differences in protein detected on these gels were not reproducible, based on the analysis of three independent extractions.

6.3.2.3 Analysis of Cymelarsan resistance on pH 5.5-6.7

Soluble proteins from $1x10^8$ of procyclic forms of the 386 and 386Mr lines were extracted and separated on 2D gels as described above but using a narrow pH range of pH 5.5-6.7 for IEF. Analysis of the Coomassie Brilliant blue stained gels using the ImageMaster 2DElite software revealed the presence of 124 protein spots on the gel from the 386 line and 104 protein spots on the gel from the 386Mr line. Most of the protein spots missing on the gel from sensitive parasites corresponded to very faint spots detected in the 386Mr line. These were likely to be missing from the resistant protein separation due to the difficulties in loading exactly the same amount of protein from both lines. Subtractive analysis of the 2D gels revealed that a highly abundant protein spot present in the extract of the 386 line was missing from the extract of the 386Mr line (Figure 6.8A and 6.8B). This protein spot (spot

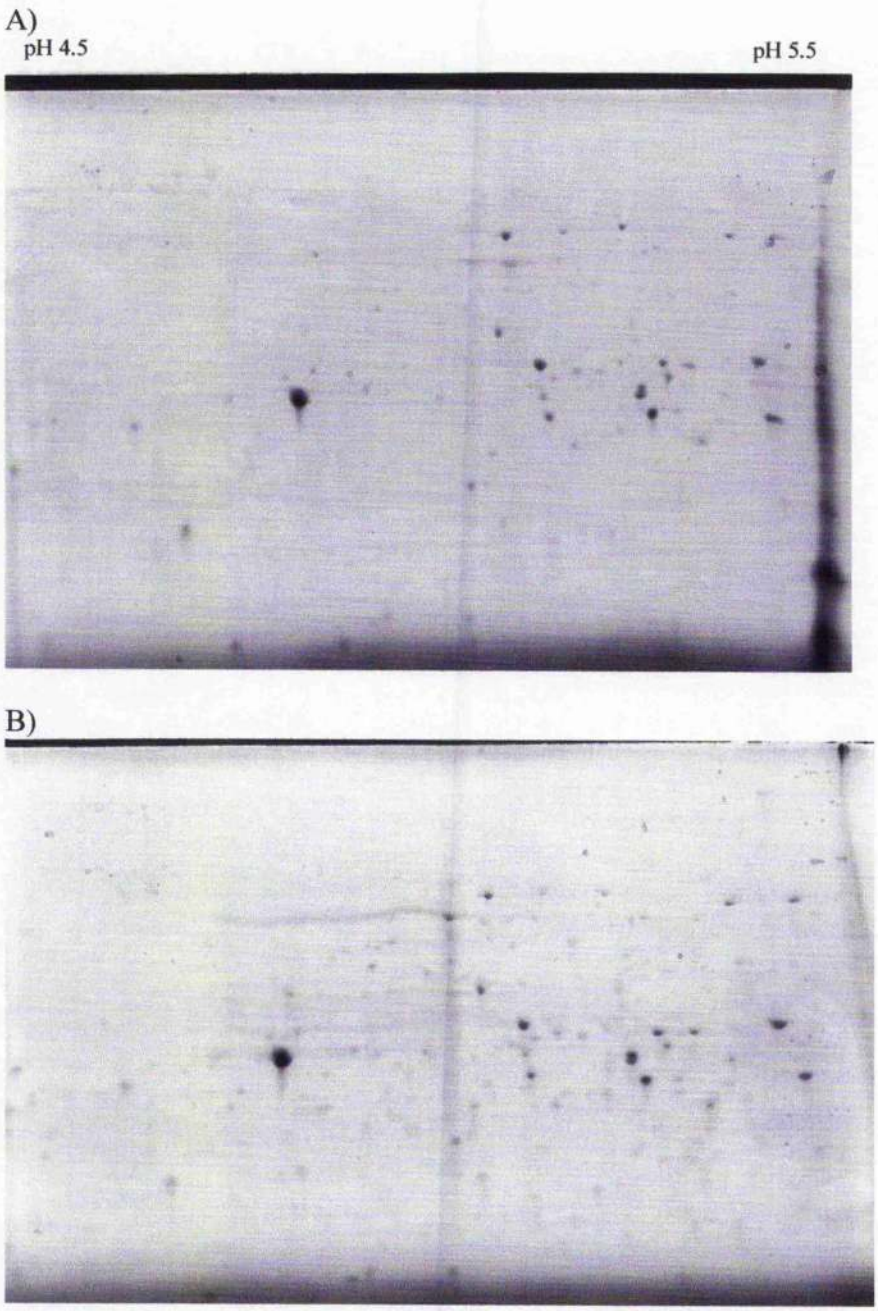


Figure 6.7: 2D gels of procyclic forms of isogenic 386 lines, pH 4.5-5.5. Narrow pH isoelectric focusing (pH 4.5 to 5.5) 2D gels of 386 (A) and 386Mr (B) lines. Soluble proteins from $1x10^8$ procyclic parasites grown in SDM79 were separated by 2DE. Subtractive analysis of Coomassie Brilliant blue stained 2D gels revealed no difference in protein expression pattern between the two isogenic lines.

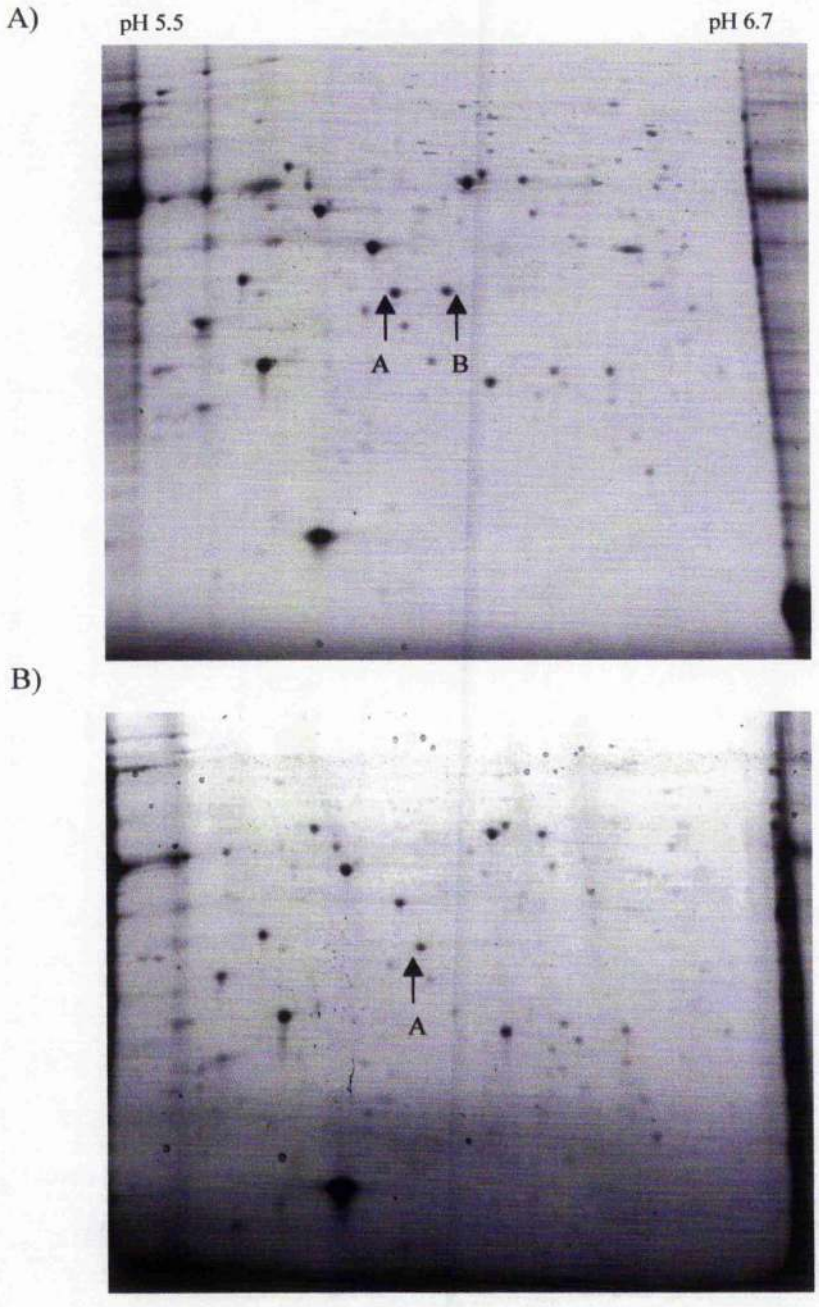


Figure 6.8: 2D gels of procyclic forms of isogenic 386 lines, pH 5.5-6.7. Narrow pH isoelectric focusing (pH 5.5 to 6.7) 2D gels of 386 (A) and 386Mr (B) lines. Soluble proteins from $1x10^8$ procyclic parasites were separated onto the 2D gel prior to Coomassie Brilliant blue staining. Subtractive analysis of the 2D gels revealed one protein spot differentially expressed (spot B) between the isogenic lines of 386.

B) had a pI of around 5.8 and in the sensitive line was adjacent to another protein spot (spot A), present in both lines, with a similar molecular weight but a more acidic pI. Due to technical difficulties with the 2DE system, this experiment was not repeated.

Taken together, the results from the 2D analysis demonstrated that two protein spots were differentially expressed between the procyclic stages of the pair of isogenic lines of 386 from over 1,000 protein spots analysed both on broad and narrow pH ranges. The two differentially expressed protein spots were visible on Coomassic Brilliant blue stained gels, suggesting that they were highly abundant in the cell. The absence of protein spot 2 (pI 4.5), in the protein extract of the 386Mr line, was confirmed in nine independent protein extractions separated on pH 4-7 gels and also on the narrow pH 4-5 range gel. The absence of spot B (pI = 5.8) in the protein extract of the 386Mr line was only shown in one experiment. Both spot 2 and spot B are adjacent to spots with similar molecular weight but a more acidic pI, suggesting that spot B could be an isoform of spot A and spot 2 could be an isoform of spot 1.

6.3.3 Identification of differentially expressed proteins

The mass spectrometry techniques, MALDI-TOF and Tandem MS, were used to determine the identity of the differentially expressed protein spots observed by 2DE between the procyclic forms of Cymelarsan sensitive and resistant 386 lines.

The differentially expressed protein spots detected on pH 4-7, pH4-5 and pH 5.5-6.7 gels, as well as nearby protein spots, were excised from the gels, digested with trypsin and analysed by mass spectrometry. Mass spectra were acquired initially by MALDI-TOF MS to provide a peptide mass fingerprint. In addition, Tandem MS was also carried out to confirm the identification obtained by PMF and to identify spots for which no

identification was obtained by PMF analysis. One of the mass spectra obtained from Tandem MS analysis of spot 1 from the 386 line and corresponding to peptide sequence FPGNTTFLIFGEAQLGDTTMQTQEAAAR is given as an example in Figure 6.9. PMF and peptide sequence tag were performed on five protein spots from the 386 line and four proteins from the 386Mr line, excised from pH 4-7 gels (Figure 6.10A and 6.10B). Analysis of the peptide mass fingerprints was carried out using the MASCOT® search engine on the NCBInr and the local *T. brucei* databases. Protein spot 4 and 5 from both 386 and 386Mr lines were identified as β tubulin (gi number: 9366734) (Figure 6.10C). No identity could be assigned to either protein spot 1 or 2 by searching either the NCBInr or the local *T. brucei* databases with their peptide mass fingerprints (Figure 6.10C). Analysis by Tandem MS confirmed the identity of spot 4 and 5 for both 386 and 386Mr lines as β tubulins, with scores of 204 and 469 for the 386 line and 142 and 437 for the 386Mr line (Figure 6.10C). Tandem MS analysis on spot 3 enabled its identification as another β tubulin (gi number: 9366734) with a score of 74 and 143 for the 386 and 386Mr line respectively (see Appendix IV). Using the sequence information generated by Tandem MS no identification for protein spots 1 and 2 could be made searching the NCBInr database. The MASCOT® search engine was then used to search the local *T. brucei* database. The two protein spots (spot 1 and 2) from the 386 line and the corresponding protein spot (spot 1) from the 386Mr line, all matched a putative open reading frame, accession number Tryp9.0.000322-274 (GeneDB), located on Chromosome IX with scores of 123, 178 and 217 respectively (Table 6.2). The differentially expressed protein spot, and the adjacent spot, detected on pH4-5 gels were analysed by Tandem MS and found to both match onto Tryp9.0.000322-274 (data not shown).

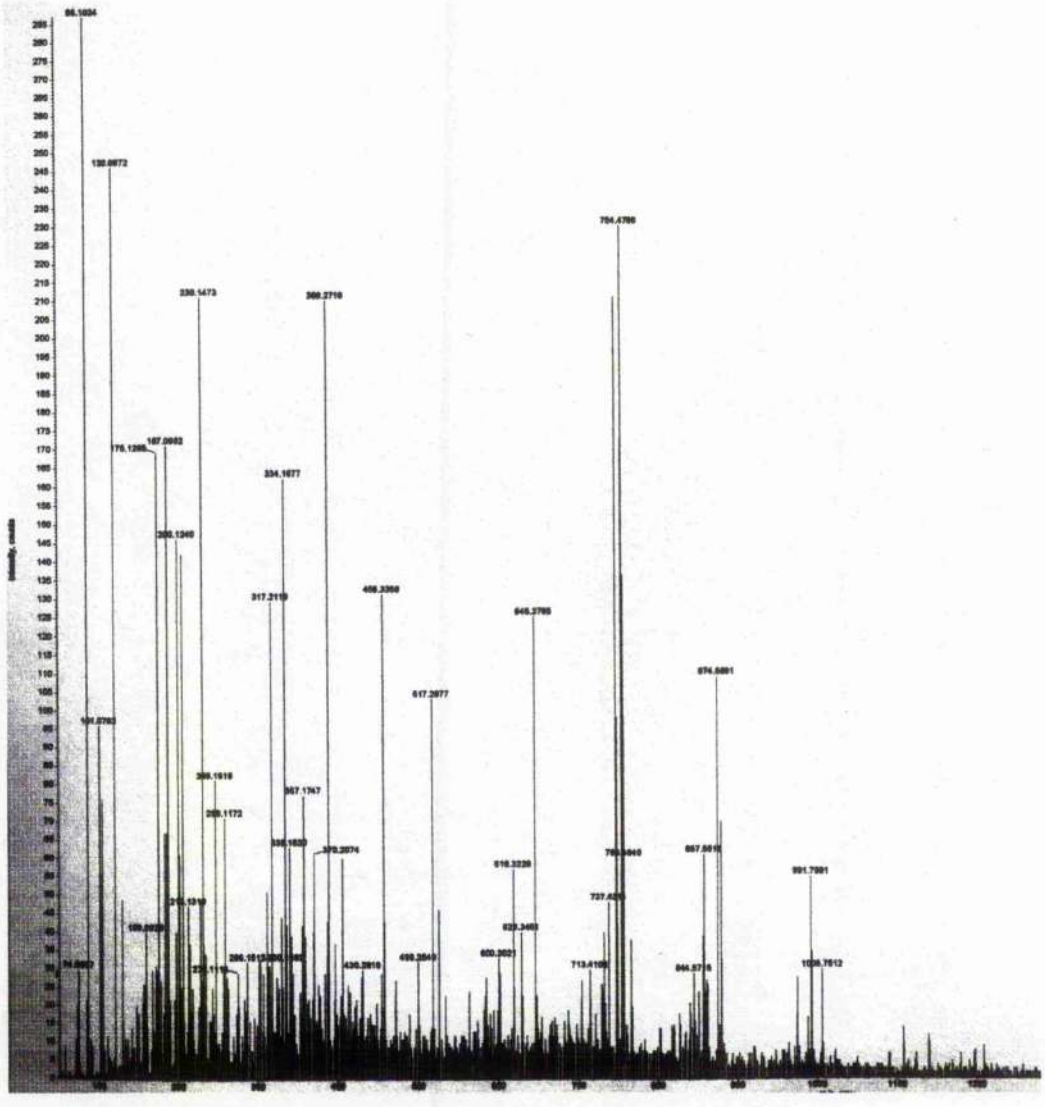
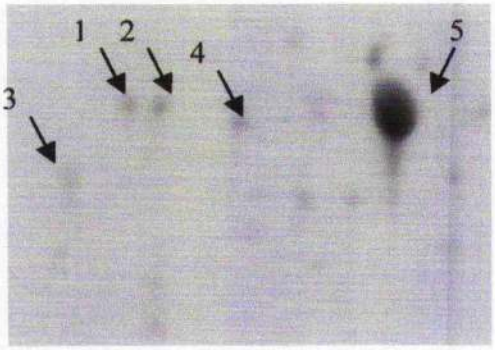
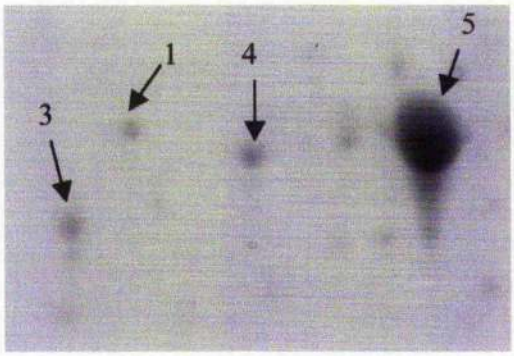


Figure 6.9: **Example of a Tandem mass spectrum**. Tandem mass spectrum of the tryptic, micropurified peptides derived from in gel digestion of protein spot 1 from 386Mr 2D gel. The parent peptide (1528.4 Da) was selected out of the mixture in the first analyser and was then dissociated by collision with nitrogen. The resulting fragments were separated in the second analyser producing the Tandem mass spectrum. The spectrum gives the mass to charge ratio (m/z) of the tryptic peptide against their intensity. Ions generally have a charge state corresponding to the number of positively charged amino acid plus the charge formally localized at the N terminus of the peptide. Thus tryptic peptides are often doubly charged, or triply if they contain a histidyl residue.

A)



B)



C)

Sample		MALDI-TOF				Tandem MS	
	gi number	Name	Score	Peptide	Coverage	ID	Score
386/1	Tryp9.0.000322-274*					Hypothetical protein	123
386/2	Tryp9.0.000322-274*					Hypothetical protein	178
386/3	9366734					β tubulin	74
386/4	9366734	β tubulin	68	5	16%	β tubulin	204
386/5	9366734	β tubulin	162	11	38%	β tubulin	469
386Mr/1	Tryp9.0.000322-274*					Hypothetical protein	217
386Mr/2	Absent						
386Mr/3	9366734					β tubulin	143
386Mr/4	9366734	β tubulin	73	5	16%	β tubulin	142
386Mr/5	9366734	β tubulin	101	9	18%	β tubulin	437

Figure 6.10: **Mass spectrometry analysis of spot 2.** Location of the differentially expressed protein spot (spot 2) on 2D gels from 386 (A) and 386Mr (B) lines. The protein spots were digested and the resulting tryptic peptides were analysed by MALDI-TOF MS and Tandem MS. The summary of results (C) showed the identity of the proteins as identified using the MASCOT® search engine to search the NCBInr database (Taxonomy: all species) and a local *T. brucei* database (*GeneDB accession number).

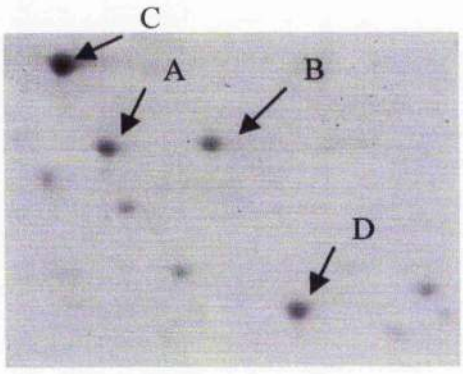
Sample	Score	Peptide score	Sequence of matching peptides	Name
386 spot 1	123	55	AALSFVVNQPEVYR	Tryp9.0.000322-274
		54	QTPDDVEIVEGEFDDK	
		13	MGLKPEPNISK	
386Mr spot 1	222	14	MGLKPEPNISK	Tryp9.0.000322-274
		61	AVSGALVEEVGTAGEETSELAGETPAAAAPSESK	
		67	QTPDDVEIVEGEFDDK	
		55	AALSFVVNQPEVYR	
		5	EIAVVMAQGK	
386 spot 2	178	60	AALSFVVNQPEVYR	Tryp9.0.000322-274
		45	MGLKPEPNISK	
		35	QTPDDVEIVEGEFDDKEIAVVMAQGK	
		37	NNKGDIVNAIMELTMDPN	

Table 6.2: Detailed results from the MASCOT[36] search engine for spot 1 and 2. The table gives the sequence of peptides from the local *T. brucei* database corresponding to the mass spectra obtained from the protein spots. The score indicates the probability that the protein in the database corresponds to the protein from the 2D gel. A score over 23 corresponds to the probability of the protein from the database to be a random match of less than 5% ($p<0.05$).

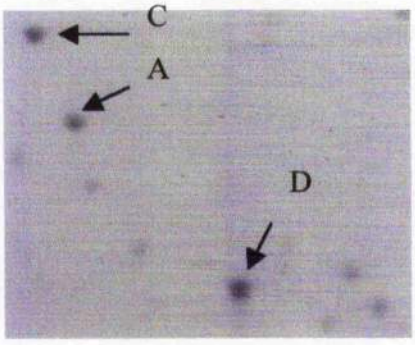
Mass spectrometry analysis was performed on the differentially expressed protein spot and surrounding protein spots from the narrow (pH 5.5-6.7) range strips (Figure 6.11). Four protein spots from the 386 line and three protein spots from the 386Mr line (Figure 6.11A and 6.11B) were excised from the 2D gels and digested with trypsin. Tandem MS analysis was performed on the tryptic peptides. The mass spectra obtained were searched against the NCBInr and the local *T. brucei* databases using the MASCOT® search engine. The protein spots surrounding the differentially expressed protein, were identified with high confidence as an enolase (spot C) and a pyridoxine kinase (spot D) (Figure 6.11C, Appendix 4). The differentially expressed protein spot (spot B) matched onto an arginine kinase (score = 510). Spot A from the 386Mr, adjacent to spot B on the 386 2D gel, was also identified as an arginine kinase (score = 487). Three gene copies of arginine kinase are present in the *T. brucei* genome (ORF 342, 343 and 345) with high homology to each other. One peptide, sequenced from both spot A and B by Tandem MS, showed that spot A and B are not isoforms as they matched onto different ORFs (Table 6.3). Spot A matched onto ORF 345 while spot B match onto either ORF 342 or 343. These results suggest that unlike the CAR protein, spot A and B are not transcripts of the same ORF that have been modified post-translationally. No identification was made for spot A of the 386 line due to the loss of the sample during Tandem MS analysis.

The analysis of the protein spots separated by 2DE using mass spectrometry enabled the identification of the differentially expressed protein spots. Spot 2 from gel pH 4-7 and from gel pH 4-5 were matched onto a non-annotated ORF, which was then designated as the Cymelarsan Resistance Associated protein (CAR protein) and spot B on gel pH 5.5- 6.7 was identified as an arginine kinase. Spot 1, adjacent to the CAR protein on 2D gels of the 386 line and present in both lines, was matched onto the same ORF as the CAR protein,

A)



B)



C)

Sample	gi number	ID	Score
386/A			
386/B	**ORF 342* / ORF 343**	**Arginine kinase**	**510**
386/C	8132069	Enolase	333
386/D	24559793	Pyridoxine kinase	218
386Mr/A	**ORF 345***	**Arginine kinase**	**487**
386Mr/B	**Absent**		
386Mr/C	8132069	Enolase	573
386Mr/D	24559793	Pyridoxine kinase	276

Figure 6.11: **Mass spectrometry analysis of spot B**. Close-up view of narrow pH range isoelectric focusing (pH 5.5-6.7) 2D gels of 386 (A) and 386Mr (B) lines showing the differentially expressed protein spot. Proteins were digested using trypsin and analysed by Tandem MS. A summary of the results obtained from searching the NCBInr and local *T. brucei* database (*GeneDB accession number) using the MASCOT® search engine is given (C). A score over 53 in this case corresponds to the probability of the protein from the database to be a random match of less than 5% ($p<0.05$).

Sample	Score	Peptide score	Sequence of matching peptides	Name
386Mr/A	487	40	YLTSDVFK	**ORF 345**
		42	LGLSEYEAVK	
		60	ACEYWPTGR	
		59	EMQDGILELIK	
		31	SLAGYPFNPCLTK	
		43	TFLVWVNEEDHLR	
		75	DFGDLNTLVDVDPEGK	
		55	ETQQQLIDDHFLFK	
		49	LGFLTFCPSNLGTTIR	
		30	GTAGEHSDSPDGIYDISNK	
		6	QPPKDFGDLNTLVDVDPEGK	
386/B	510	50	HLTSDVFK	**ORF 342/ ORF 343**
		58	LGLSEYEAVK	
		60	ACEYWPTGR	
		62	EMQDGILELIK	
		30	SLAGYPFNPCLTK	
		65	TFLVWVNEEDHLR	
		65	DFGDLNTLVDVDPEGK	
		63	ETQQQLIDDHFLFK	
		57	LGFLTFCPSNLGTTIR	

Table 6.3: Detailed results from the MASCOT® search engine for spot A and B. The table gives the sequence of peptides from the local *T. brucei* database corresponding to the mass spectra obtained from the protein spots. The score indicates the probability that the protein in the database corresponds to the protein from the 2D gel. A score over 53 in this case corresponds to the probability of the protein from the database to be a random match of less than 5% ($p<0.05$). One identified peptide, highlighted in blue, enabled to differentiate ORF 345 from ORF 342 and ORF 343.

suggesting that spot 1 was an isoform of the CAR protein. Spot A and B were both identified as arginine kinase. However, the Tandem MS analysis demonstrated that spot A matched onto ORF 345 while spot B matched onto either ORF 342 or ORF 343. These results would suggest that the transcript of either ORF 342 or ORF 343 (or both) is present in the 386 line but not in the 386Mr line at pH 5.8 on the 2D gel.

6.3.4 Molecular analysis of the CAR locus

6.3.4.1 Presence of the CAR gene in the isogenic 386 lines

The CAR protein was found to be differentially expressed between 386 and 386Mr lines. It was located at around pH 4.5 on 2D gels and was matched onto a putative ORF (Tryp9.0.000322-274) from the reference line 927. BLAST searches in GeneDB using the sequence from this protein revealed that there was another copy of the gene situated upstream from Tryp9.0.000322-274, called Tryp9.0.000322-277. These ORFs will thereafter be referred to as ORF 274 and ORF 277. Both copies have identical nucleotide (564 bp) sequences in the reference line 927. No signal peptides or transmembrane domains were found in the predicted protein sequence. Searches with the programs InterPro and Pfam (available at http://ca.expasy.org/tools/) showed the presence of a Nascent polypeptide Associated Complex domain (NAC domain) at the N terminus (amino acid residues 29-88) and an UBiquitin-Associated domain (UBA domain) at the C terminus (amino acid residues 146-186) (Figure 6.12). The DNA and corresponding amino acid sequence of ORF 274 from the reference line 927 are given in Figure 6.13.

In order to confirm the presence of the two ORFs in 927 and investigate whether both homologues occur in the 386 and 386Mr lines, primers were designed to the 5'UTR of

A)

E score	Organism	gi number	Name
2e-04	*T. cruzi*	21435928	NAC alpha mRNA
0.004	*Mus musculus*	20345404	mRNA

B)

E score	Organism	gi number	Name
1e-52	*T. cruzi*	21435929	NAC alpha
5e-40	*Leishmania major*	12311802	possible nascent polypeptide associated complex subunit, copy 2
1e-32	*Leishmania major*	12311803	possible nascent polypeptide associated complex subunit, copy 1
3e-32	*Leishmania infantum*	8671200	nascent polypeptide associated complex homologue, alpha chain

C) **Systematic Name: TRYP9.0.000322-274**

Product: POSSIBLE NASCENT POLYPEPTIDE ASSOCIATED COMPLEX SUBUNIT, COPY 1. (1 Other), predicted by automatic BLAST annotation

Location Chromosome: 9 Contig Location: 18909..19472 Length: 564 bp

Predicted Peptide Properties

Mass: 20.1 kDa Amino acids: 188 Isoelectric point: pH 4.2 Charge: -11.5

Signal Peptide: Not found Transmembrane Domains: 0 found

Domain Information

Database	Acc. No./ID	Description
InterPro	IPR002715	Nascent polypeptide-associated complex NAC
Pfam	PF01849	NAC domain
InterPro	IPR000449	Ubiquitin-associated domain
	PF00627	UBA/TS-N domain

Figure 6.12: Summary of BLAST searches in NCBI and GeneDB with ORF 274. Summary of BLASTn searches (A) and BLASTp searches (B) in NCBInr database using ORF 274 from the reference line 927. Information provided by GeneDB about ORF 274 (C).

ATG AGC GCT AAC GAT GTT CAA AAC CCT CCA GTG GAA GAC GAC GAG
M S A N D V Q N P P V E D D E

GTC CCA ACC CTC GAA GCT GCT GAC ACT CAA CAG GCT TCC AAG CCA
V P T L E A A D T Q Q A S K P

TCC AAG CGC TAC GCA AAG GCT ATG GCC AAG ATG GGT TTG AAA CCT
S K R Y A K A M A K M G L K P

GAG CCA AAC ATC TCT AAG GTA CAC ATT CGC AAA CGT GCG GCG CTC
E P N I S K V H I R K R A A L

NAC Domain

TCT TTT GTG GTG AAC CAA CCC GAG GTG TAC CGC TTT CCA GGA ACC
S F V V N Q P E V Y R F P G T

AAC ACA TTC CTA ATC TTT GGT GAA GCT CAG CTT GGC GAC ACA ACG
N T F L I F G E A Q L G D T T

ATG CAA ACG CAG GAA GCC GCA GCT CGC GCG GTG TCT GGC GCC CTT
M Q T Q E A A A R A V S G A L

GTG GAG GAG GTG GGA ACT GCA GGA GAG GAA ACC TCT GAG CTT
V E E V G T A G E E T S E L

GCT GGT GAA ACT CCT GCA GCA GCT GCC CCA TCT GAG AGC AAA CAG
A G E T P A A A A P S E S K Q

ACT CCC GAC GAT GTC GAG ATT GTG GAA GGC GAA TTC GAC GAT
T P D D V E I V E G E F D D

AAG GAG ATT GCT GTG GTT ATG GCA CAA GGC AAA ACG GAC CGC ATT
K E I A V V M A Q G K T D R I

UBA domain

GGA GCA ATC CGC GCC CTT CGG AAC AAC AAG GGA GAT ATC GTC AAT
G A I R A L R N N K G D I V N

GCT ATC ATG GAA CTG ACG ATG GAT CCC AAC
A I M E L T M D P N

Figure 6.13: Schematic representation of the DNA and amino acid sequences of ORF 274 from *T. brucei* line 927. Two domains were predicted by PedictProtein (http://cubic.bioc.columbia.edu/predictprotein/) in the ORF. A nascent polypeptide associated domain (NAC domain, underlined) and an ubiquitin-associated domain (UBA domain, underlined). Peptides, from trypsin digestion of spot 1 from 386Mr line, sequenced by Tandem MS matching onto the theoretical peptides from the translated ORF 274 are highlighted in blue.

ORF 274 and 3'UTR of ORF 277. PCR amplification of 927 genomic DNA was predicted to give a 2 Kb product that should contain two EcoRI sites and one BamHI site. PCR amplification of 927 genomic DNA confirmed the size prediction and digestion with EcoRI and BamHI confirmed the presence of the restriction sites. The same experiment carried out on genomic DNA from 386 and 386Mr lines confirmed the presence of the two ORFs in these lines (Figure 6.15). This experiment demonstrated the potential presence of two ORFs, identical to each other in the genome reference line 927, potentially each encoding the CAR protein in 386 and 386Mr lines.

A mutation in the ORF of one of the lines could account for the disappearance of a CAR protein isoform on 2D gels of the 386Mr lines. This possibility was assessed by cloning and sequencing the genes. An alignment of the ORFs and their 3' and 5'UTR regions was performed to identify regions of low homology which could be used to design ORF specific primers. As there is no robust bioinformatic method of identifying the precise boundaries of 3' and 5'UTRs, 287 bp upstream and 233 bp downstream of ORF 274 was compared with 171 bp upsteam and 260 bp downstream of ORF 277. The alignment of the ORFs confirmed the 100% homology in the translated region of the genes. The 5'UTR was shown to be highly conserved over 116 bp upstream of the start codon whereas only 19 bp aligned in the 3'UTR regions. These regions were followed by highly divergent sequences, which were used to design the ORF-specific primers (Figure 6.16). Analysis of the sequence of the expected PCR product for ORF 274 and ORF 277 using these primers enabled the detection of a BsaJI restriction site present at different positions in the 3'UTR of the ORFs, 27 bp downstream the codon stop of ORF 274 and 137 bp downstream of the codon stop of ORF 277 (Figure 6.16). The digestion of the PCR products, amplified using the ORF specific primers, with the restriction enzyme BsaJI, generating fragments of the

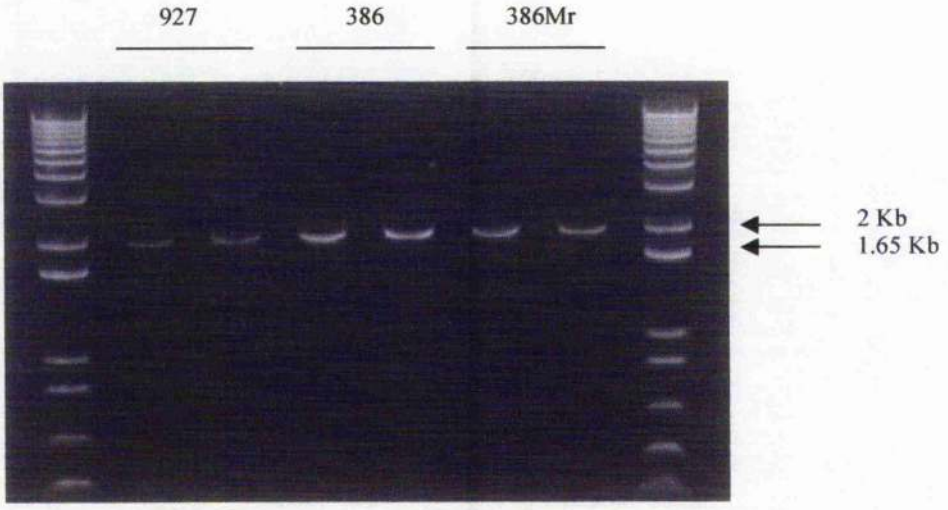


Figure 6.14: PCR amplification of the tandem repeat of ORF 274 and 277. PCR performed on line 927, 386 and 386Mr using primers Tryp9/11 and Tryp9/12, amplified a DNA fragment containing both open reading frames. The 2 Kb PCR product (expected size from reference line 927 database) was recovered for all three lines.

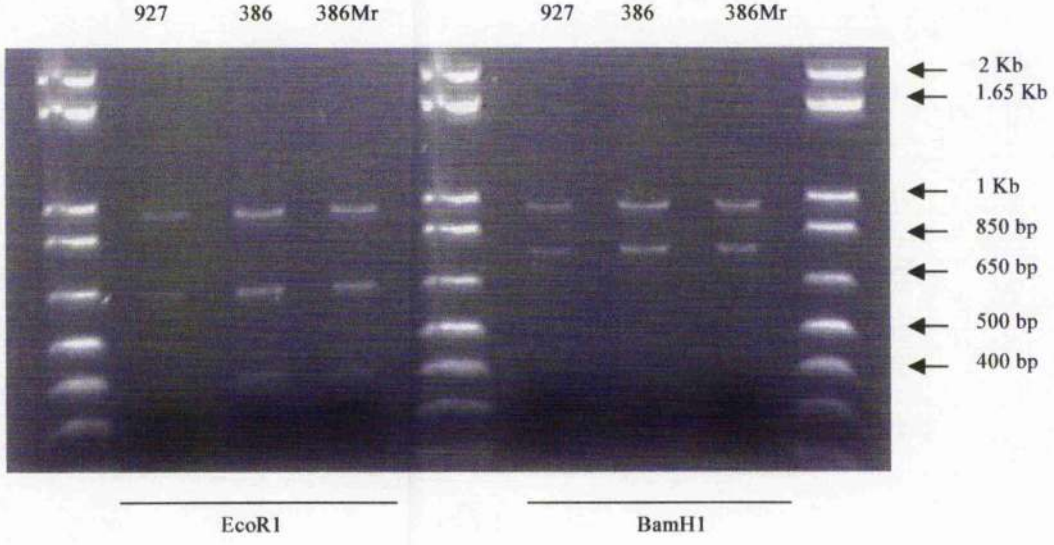


Figure 6.15: RFLP analysis on the tandem repeat of ORF 274 and 277. Digestion of the 2 Kb PCR product using the restriction enzymes EcoRI and BamHI. The restriction fragments corresponded to the predicted sizes of EcoRI digestion (968 bp, 641 bp and 391 bp) and BamHI digestion (968b bp, 764 bp and 269 bp).

```
ORF274   1 GAGGGAGGGGGAATTTCCTCTGCTTATATATTTACATTTGAATCAGCGAA      50

                                                  ..||......||
ORF277   1                                        CAAAAATTATAA      12

        51 TTTATCTACATGCATACACATCTTCCTTTACATTATTTATCCACTGCGTG      100
           ......|......|......||..|...|.|.||..||.|.|.|||||||
        13 AAATGTTTGTGCTACTACATTCGCCTCATTCCTTCGTTTTTCTCTGCGTG       62

       101 TCTGCTTTTAGATTTCTCCCTCCCTGTTTTATGCGCATTTCAACTTTTCT      150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
        63 TCTGCTTTTAGATTTCTCCCTCCCTGTTTTATGCGCATTTCAACTTTTCT      112

       151 TTTCCCTTCTTACATTTCTTTCGCTACATCCAATCTGAACAGTAACCCGG      200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       113 TTTCCCTTCTTACATTTCTTTCGCTACATCCAATCTGAACAGTAACCCGG      162

       201 ACCATCACGATGAGCGCTAACGATGTTCAAAACCCTCCAGTGGAAGACGA      250
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       163 ACCATCACGATGAGCGCTAACGATGTTCAAAACCCTCCAGTGGAAGACGA      212

       251 CGAGGTCCCAACCCTCGAAGCTGCTGACACTCAACAGGCTTCCAAGCCAT      300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       213 CGAGGTCCCAACCCTCGAAGCTGCTGACACTCAACAGGCTTCCAAGCCAT      262

       301 CCAAGCGCTACGCAAAGGCTATGGCCAAGATGGGTTTGAAACCTGAGCCA      350
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       263 CCAAGCGCTACGCAAAGGCTATGGCCAAGATGGGTTTGAAACCTGAGCCA      312

       351 AACATCTCTAAGGTACACATTCGCAAACGTGCGGCGCTCTCTTTTGTGGT      400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       313 AACATCTCTAAGGTACACATTCGCAAACGTGCGGCGCTCTCTTTTGTGGT      362

       401 GAACCAACCCGAGGTGTACCGCTTTCCAGGAACCAACACATTCCTAATCT      450
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       363 GAACCAACCCGAGGTGTACCGCTTTCCAGGAACCAACACATTCCTAATCT      412

       451 TTGGTGAAGCTCAGCTTGGCGACACAACGATGCAAACGCAGGAAGCCGCA      500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       413 TTGGTGAAGCTCAGCTTGGCGACACAACGATGCAAACGCAGGAAGCCGCA      462

       501 GCTCGCGCGGTGTCTGGCGCCCTTGTGGAGGAGGTGGGAACTGCAGGAGA      550
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       463 GCTCGCGCGGTGTCTGGCGCCCTTGTGGAGGAGGTGGGAACTGCAGGAGA      512

       551 GGAAACCTCTGAGCTTGCTGGTGAAACTCCTGCAGCAGCTGCCCCATCTG      600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       513 GGAAACCTCTGAGCTTGCTGGTGAAACTCCTGCAGCAGCTGCCCCATCTG      562

       601 AGAGCAAACAGACTCCCGACGATGTCGAGATTGTGGAAGGCGAATTCGAC      650
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       563 AGAGCAAACAGACTCCCGACGATGTCGAGATTGTGGAAGGCGAATTCGAC      612

       651 GATAAGGAGATTGCTGTGGTTATGGCACAAGGCAAAACGGACCGCATTGG      700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       613 GATAAGGAGATTGCTGTGGTTATGGCACAAGGCAAAACGGACCGCATTGG      662

       701 AGCAATCCGCGCCCTTCGGAACAACAAGGGAGATATCGTCAATGCTATCA      750
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       663 AGCAATCCGCGCCCTTCGGAACAACAAGGGAGATATCGTCAATGCTATCA      712

       751 TGGAACTGACGATGGATCCCAACTGATCGAGAGTGCTTTTCGTGAGCAAC      800
           ||||||||||||||||||||||||||||||||||||||||||||||||.|.
       713 TGGAACTGACGATGGATCCCAACTGATCGAGAGTGCTTTTCGTGAGCGAT      762

       801 AGTGTAAGTGAAACCCCCAATAAGGACATAAGGGGGATGGAGGGAAATGA      850
           .....|.||||||.......|....|..|.......|...|.........
       763 TACCGAGGTGAAAAGTTTTTTCTTAATTTCCTTTTTAAAAATAATGGCCT      812
```

```
851 AAACTGTAAGAAAAAAATATGAGATTTGAAGTGGACTCTCAGAAGTGATT    900
    .||...|||.....|...|.|..||.|..|.....|.............
813 GAAAGTTAAACGTGAGTCACGTAATGTTGACCCACCGTCTCAGTCGTGCA    862

901 TGTGGGGTTCTGCCTAGGAGTGTTTCCGTGGCCACCTGCTTGTAAGTTAC    950
    ....|..|.||.|.|.........|........|.||..||........|
863 CAGTGAATCCTCCATCCATTGTGATGTTAACATAACTTTTTCCTCACATC    912

951 GAAGGAAAGAATGTGCGAGAACAATATACAGTGATAACAGAAAAGAGTAA   1000
    ..........|..........|..|.|....|...|........|.|.
913 TTTCCTTTTTTTCCTTTGTCTGTACTTTCTTATACTTCTTGTTTCTGCAG    962

1001 GTGATA                                  1006
     ..|.|.
963 TGGTTTGATCGGCCAAAAGATTCGTCATGGCTC        995
```

Figure 6.16: **Alignment of ORF 274 with ORF 277 from line 927.** The start and stop codon for each ORF are highlighted in blue. The divergent sequences in the 3' and 5'UTR regions were used to design ORF specific primers (underlined sequences). The BsaJ1 restriction sites are highlighted in red.

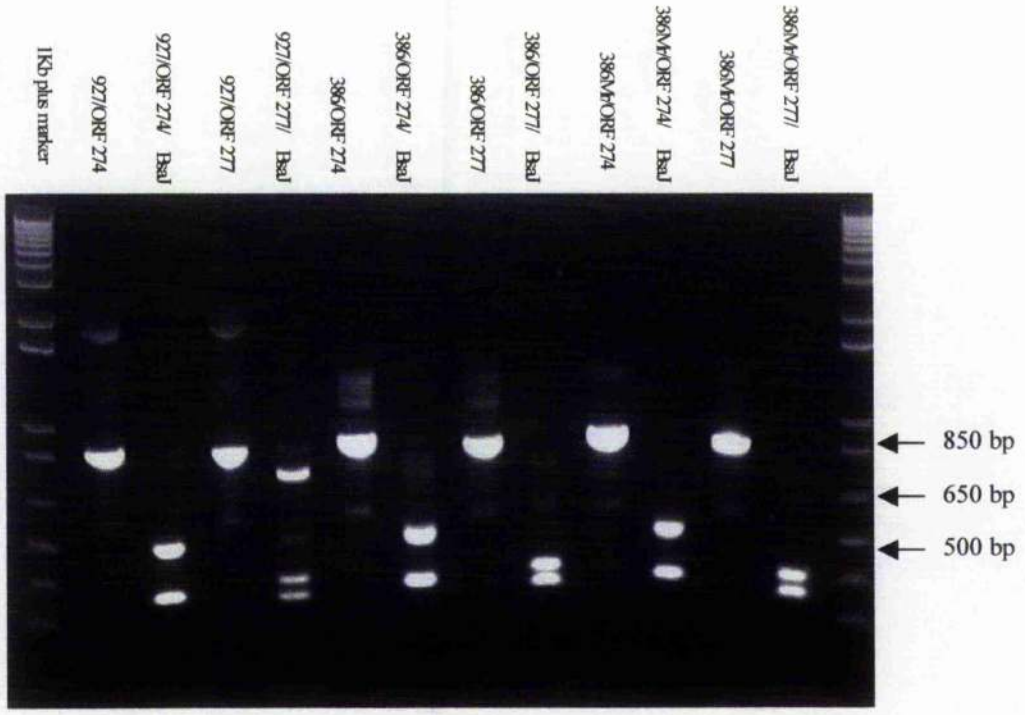


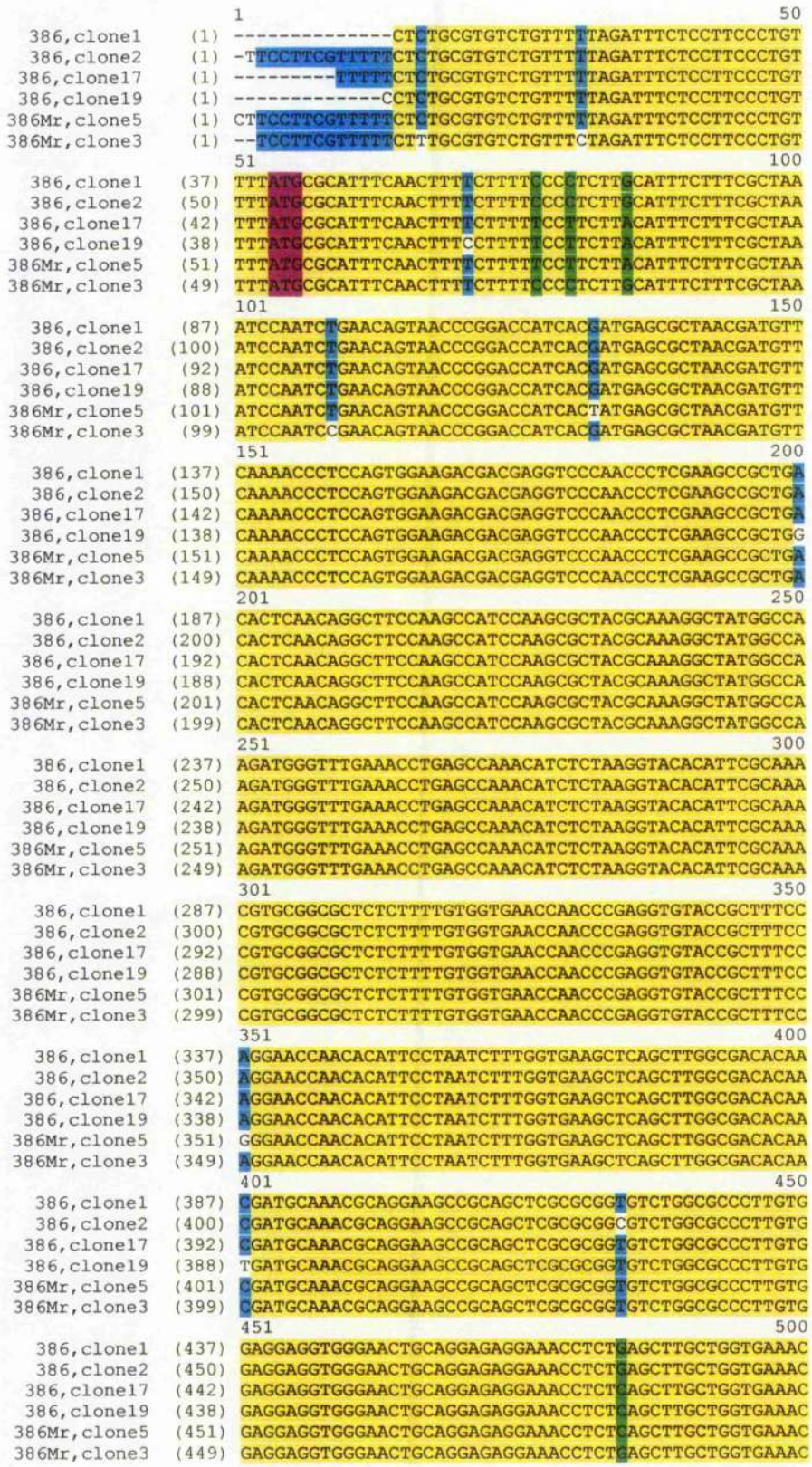
Figure 6.17: **Specificity of the primers designed for ORF 274 and 277**. PCR using the ORF specific primers on genomic DNA from 927, 386 and 386Mr amplified a fragment of the expected size (ORF 274: 891 bp and ORF 277: 868 bp). The PCR product from each ORF was digested using the BsaJI restriction enzyme. The digestion products corresponded to the expected sizes (ORF 274: 370 bp and 500 bp and ORF 277: 123 bp, 350 bp and 395 bp).

expected sizes for each ORF (Figure 6.17). Both ORFs from 386 and 386Mr lines were cloned, using the ORF specific primers, and sequenced to determine whether any nucleotide difference could be detected. Differences between sequences of different clones of ORF 274 were observed which appear consistent with allelic variation (nucleotides highlighted in green, Figure 6.18). Other differences however appear randomly in the sequences and were probably due to PCR errors by the Taq polymerase or sequencing errors. Taken together the sequencing of the ORFs would suggest that both ORF 274 and 277 from the 386 line are identical to their homologues in 386Mr line (Figure 6.18 and 6.19) and that ORF 274 is identical to ORF 277. No mutations were found in the translated gene products of ORF 274 and 277 that could account for the absence of the CAR protein spot (spot 2) in the 386Mr extract detected by 2DE.

6.3.4.2 Transcription of ORF 274 and 277 in the isogenic lines

The sequencing of the two ORFs from both the 386 and 386Mr lines demonstrated that both ORFs were identical. These results suggest that the lack of expression at the protein level of one isoform of the CAR protein is not due to an alteration in coding region of either of the genes. An alternative explanation for the differences at the protein level could be that there are differences in the level of transcripts from the two ORFs between the 386 and 386Mr lines. To analyse this possibility, Northern blot and RT-PCR analyses were performed.

Primers were designed in the 3'UTR regions of each ORF to produce 150 bp long probes which should specifically hybridise to the transcripts from one or the other ORF (Figure 6.1). In the Northern analysis, the membrane was first probed with a PCR product amplified from the coding region for TIM as a positive control. The TIM probe showed the


1 50
386,clone1 (1) ----------CTCTGCGTGTCTGTTTTTAGATTTCTCCTTCCCTGT
386,clone2 (1) -TTCCTTCGTTTTTCTCTGCGTGTCTGTTTTTAGATTTCTCCTTCCCTGT
386,clone17 (1) ---------TTTTTCTCTGCGTGTCTGTTTTTAGATTTCTCCTTCCCTGT
386,clone19 (1) ----------CCTCTGCGTGTCTGTTTTTAGATTTCTCCTTCCCTGT
386Mr,clone5 (1) CTTCCTTCGTTTTTCTCTGCGTGTCTGTTTTTAGATTTCTCCTTCCCTGT
386Mr,clone3 (1) --TCCTTCGTTTTTCTTTGCGTGTCTGTTTCTAGATTTCTCCTTCCCTGT
51 100
386,clone1 (37) TTTATGCGCATTTCAACTTTTCTTTTCCCCTCTTGCATTTCTTTCGCTAA
386,clone2 (50) TTTATGCGCATTTCAACTTTTCTTTTCCCCTCTTGCATTTCTTTCGCTAA
386,clone17 (42) TTTATGCGCATTTCAACTTTTCTTTTTCCTTCTTACATTTCTTTCGCTAA
386,clone19 (38) TTTATGCGCATTTCAACTTTCCTTTTTCCTTCTTACATTTCTTTCGCTAA
386Mr,clone5 (51) TTTATGCGCATTTCAACTTTTCTTTTTCCTTCTTACATTTCTTTCGCTAA
386Mr,clone3 (49) TTTATGCGCATTTCAACTTTTCTTTTCCCCTCTTGCATTTCTTTCGCTAA
101 150
386,clone1 (87) ATCCAATCTGAACAGTAACCCGGACCATCACGATGAGCGCTAACGATGTT
386,clone2 (100) ATCCAATCTGAACAGTAACCCGGACCATCACGATGAGCGCTAACGATGTT
386,clone17 (92) ATCCAATCTGAACAGTAACCCGGACCATCACGATGAGCGCTAACGATGTT
386,clone19 (88) ATCCAATCTGAACAGTAACCCGGACCATCACGATGAGCGCTAACGATGTT
386Mr,clone5 (101) ATCCAATCTGAACAGTAACCCGGACCATCACTATGAGCGCTAACGATGTT
386Mr,clone3 (99) ATCCAATCCGAACAGTAACCCGGACCATCACGATGAGCGCTAACGATGTT
151 200
386,clone1 (137) CAAAACCCTCCAGTGGAAGACGACGAGGTCCCAACCCTCGAAGCCGCTGA
386,clone2 (150) CAAAACCCTCCAGTGGAAGACGACGAGGTCCCAACCCTCGAAGCCGCTGA
386,clone17 (142) CAAAACCCTCCAGTGGAAGACGACGAGGTCCCAACCCTCGAAGCCGCTGA
386,clone19 (138) CAAAACCCTCCAGTGGAAGACGACGAGGTCCCAACCCTCGAAGCCGCTGG
386Mr,clone5 (151) CAAAACCCTCCAGTGGAAGACGACGAGGTCCCAACCCTCGAAGCCGCTGA
386Mr,clone3 (149) CAAAACCCTCCAGTGGAAGACGACGAGGTCCCAACCCTCGAAGCCGCTGA
201 250
386,clone1 (187) CACTCAACAGGCTTCCAAGCCATCCAAGCGCTACGCAAAGGCTATGGCCA
386,clone2 (200) CACTCAACAGGCTTCCAAGCCATCCAAGCGCTACGCAAAGGCTATGGCCA
386,clone17 (192) CACTCAACAGGCTTCCAAGCCATCCAAGCGCTACGCAAAGGCTATGGCCA
386,clone19 (188) CACTCAACAGGCTTCCAAGCCATCCAAGCGCTACGCAAAGGCTATGGCCA
386Mr,clone5 (201) CACTCAACAGGCTTCCAAGCCATCCAAGCGCTACGCAAAGGCTATGGCCA
386Mr,clone3 (199) CACTCAACAGGCTTCCAAGCCATCCAAGCGCTACGCAAAGGCTATGGCCA
251 300
386,clone1 (237) AGATGGGTTTGAAACCTGAGCCAAACATCTCTAAGGTACACATTCGCAAA
386,clone2 (250) AGATGGGTTTGAAACCTGAGCCAAACATCTCTAAGGTACACATTCGCAAA
386,clone17 (242) AGATGGGTTTGAAACCTGAGCCAAACATCTCTAAGGTACACATTCGCAAA
386,clone19 (238) AGATGGGTTTGAAACCTGAGCCAAACATCTCTAAGGTACACATTCGCAAA
386Mr,clone5 (251) AGATGGGTTTGAAACCTGAGCCAAACATCTCTAAGGTACACATTCGCAAA
386Mr,clone3 (249) AGATGGGTTTGAAACCTGAGCCAAACATCTCTAAGGTACACATTCGCAAA
301 350
386,clone1 (287) CGTGCGGCGCTCTCTTTTGTGGTGAACCAACCCGAGGTGTACCGCTTTCC
386,clone2 (300) CGTGCGGCGCTCTCTTTTGTGGTGAACCAACCCGAGGTGTACCGCTTTCC
386,clone17 (292) CGTGCGGCGCTCTCTTTTGTGGTGAACCAACCCGAGGTGTACCGCTTTCC
386,clone19 (288) CGTGCGGCGCTCTCTTTTGTGGTGAACCAACCCGAGGTGTACCGCTTTCC
386Mr,clone5 (301) CGTGCGGCGCTCTCTTTTGTGGTGAACCAACCCGAGGTGTACCGCTTTCC
386Mr,clone3 (299) CGTGCGGCGCTCTCTTTTGTGGTGAACCAACCCGAGGTGTACCGCTTTCC
351 400
386,clone1 (337) AGGAACCAACACATTCCTAATCTTTGGTGAAGCTCAGCTTGGCGACACAA
386,clone2 (350) AGGAACCAACACATTCCTAATCTTTGGTGAAGCTCAGCTTGGCGACACAA
386,clone17 (342) AGGAACCAACACATTCCTAATCTTTGGTGAAGCTCAGCTTGGCGACACAA
386,clone19 (338) AGGAACCAACACATTCCTAATCTTTGGTGAAGCTCAGCTTGGCGACACAA
386Mr,clone5 (351) GGGAACCAACACATTCCTAATCTTTGGTGAAGCTCAGCTTGGCGACACAA
386Mr,clone3 (349) AGGAACCAACACATTCCTAATCTTTGGTGAAGCTCAGCTTGGCGACACAA
401 450
386,clone1 (387) CGATGCAAACGCAGGAAGCCGCAGCTCGCGCGGTGTCTGGCGCCCTTGTG
386,clone2 (400) CGATGCAAACGCAGGAAGCCGCAGCTCGCGCGGCGTCTGGCGCCCTTGTG
386,clone17 (392) CGATGCAAACGCAGGAAGCCGCAGCTCGCGCGGTGTCTGGCGCCCTTGTG
386,clone19 (388) TGATGCAAACGCAGGAAGCCGCAGCTCGCGCGGTGTCTGGCGCCCTTGTG
386Mr,clone5 (401) CGATGCAAACGCAGGAAGCCGCAGCTCGCGCGGTGTCTGGCGCCCTTGTG
386Mr,clone3 (399) CGATGCAAACGCAGGAAGCCGCAGCTCGCGCGGTGTCTGGCGCCCTTGTG
451 500
386,clone1 (437) GAGGAGGTGGGAACTGCAGGAGAGGAAACCTCTGAGCTTGCTGGTGAAAC
386,clone2 (450) GAGGAGGTGGGAACTGCAGGAGAGGAAACCTCTGAGCTTGCTGGTGAAAC
386,clone17 (442) GAGGAGGTGGGAACTGCAGGAGAGGAAACCTCTGAGCTTGCTGGTGAAAC
386,clone19 (438) GAGGAGGTGGGAACTGCAGGAGAGGAAACCTCTGAGCTTGCTGGTGAAAC
386Mr,clone5 (451) GAGGAGGTGGGAACTGCAGGAGAGGAAACCTCTGAGCTTGCTGGTGAAAC
386Mr,clone3 (449) GAGGAGGTGGGAACTGCAGGAGAGGAAACCTCTGAGCTTGCTGGTGAAAC

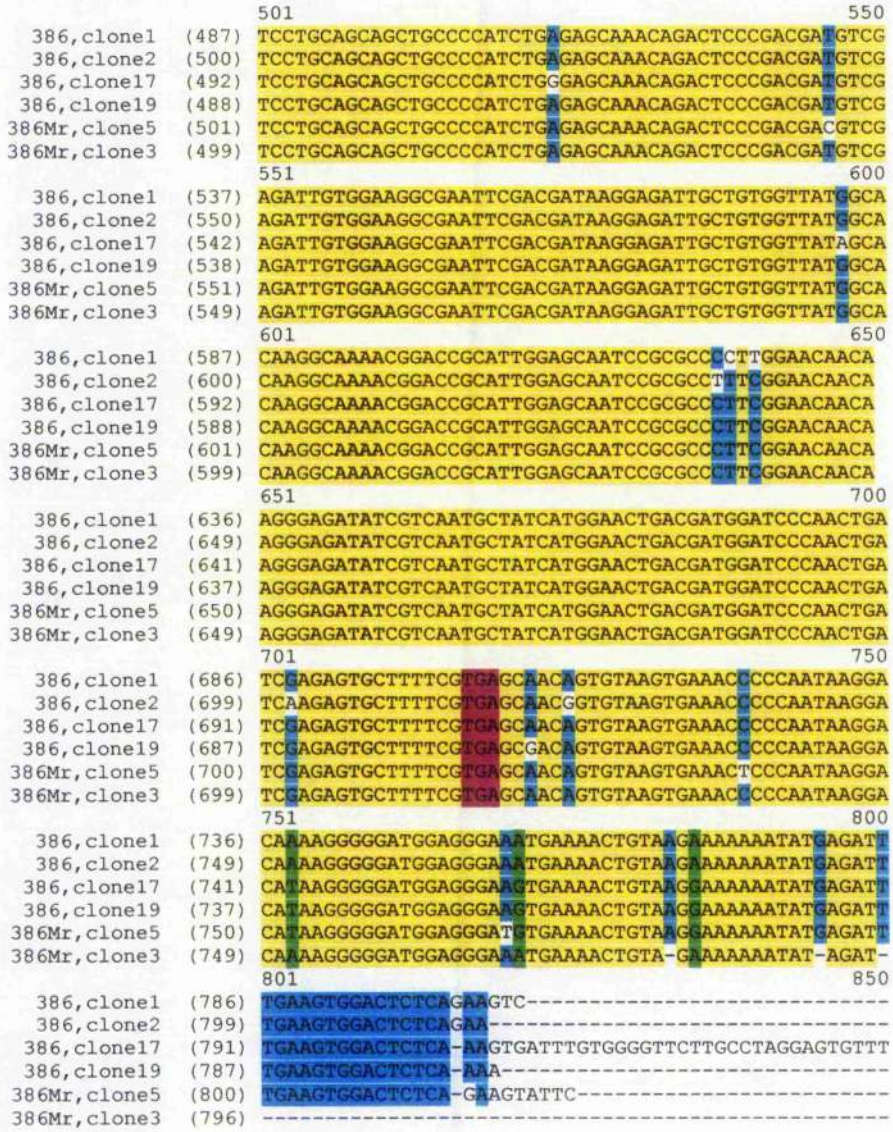
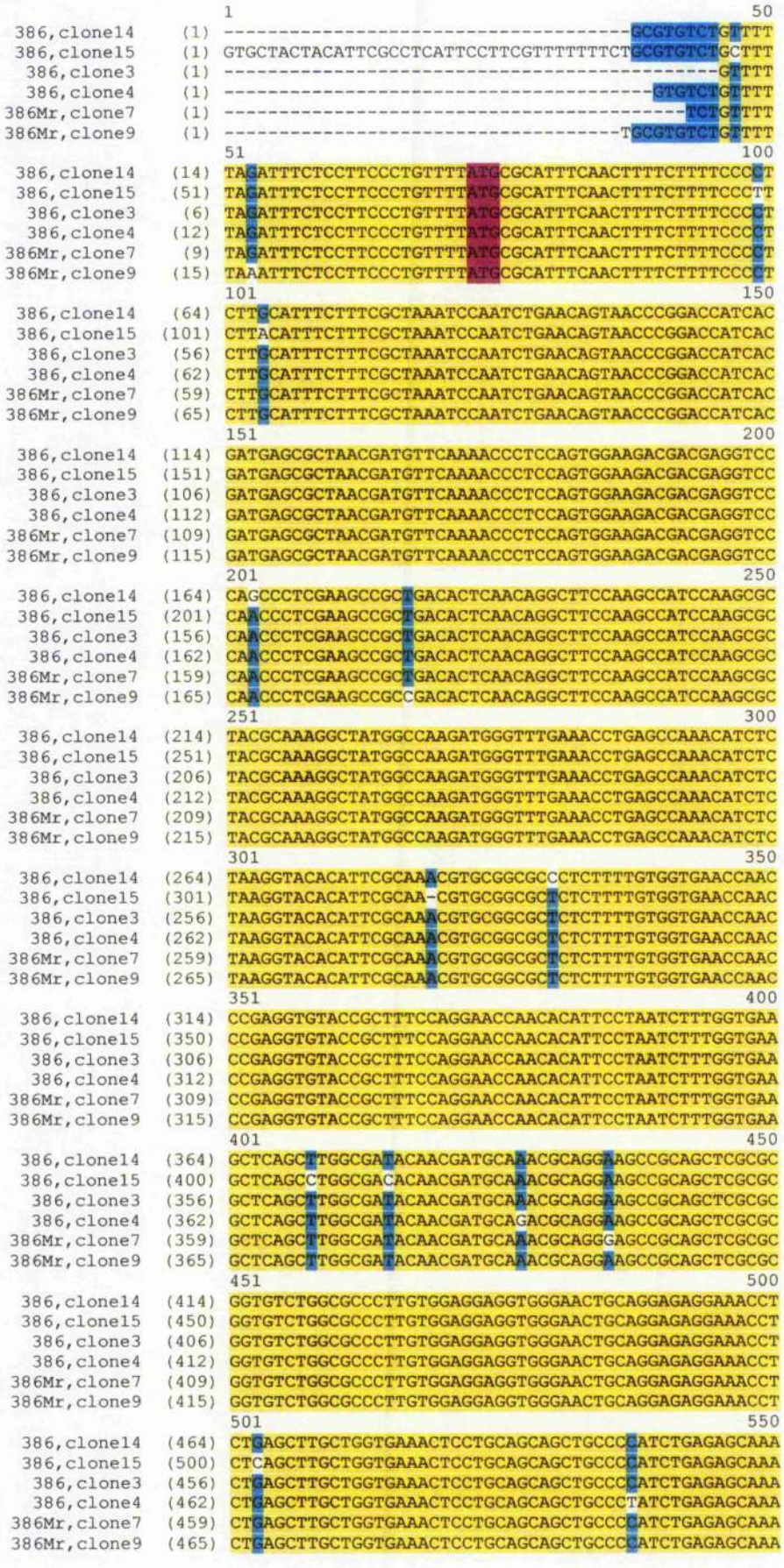


Figure 6.18: Alignment of ORF 274 sequences from isogenic 386 lines. Start codon (ATG) and stop codon (TGA) are indicated in pink. The sequences from line 386 (clone 1, 2, 17 and 19) are aligned against sequences from line 386Mr (clone 3 and 5). Allelic differences are highlighted in green.


1 50
386,clone14 (1) ----------GCGTGTCTGTTTT
386,clone15 (1) GTGCTACTACATTCGCCTCATTCCTTCGTTTTTTTCTGCGTGTCTGCTTT
386,clone3 (1) ----------GTTTT
386,clone4 (1) ----------GTGTCTGTTTT
386Mr,clone7 (1) ----------TCTGTTTT
386Mr,clone9 (1) ----------TGCGTGTCTGTTTT
51 100
386,clone14 (14) TAGATTTCTCCTTCCCTGTTTTATGCGCATTTCAACTTTTCTTTTCCCCT
386,clone15 (51) TAGATTTCTCCTTCCCTGTTTTATGCGCATTTCAACTTTTCTTTTCCCTT
386,clone3 (6) TAGATTTCTCCTTCCCTGTTTTATGCGCATTTCAACTTTTCTTTTCCCCT
386,clone4 (12) TAGATTTCTCCTTCCCTGTTTTATGCGCATTTCAACTTTTCTTTTCCCCT
386Mr,clone7 (9) TAGATTTCTCCTTCCCTGTTTTATGCGCATTTCAACTTTTCTTTTCCCCT
386Mr,clone9 (15) TAAATTTCTCCTTCCCTGTTTTATGCGCATTTCAACTTTTCTTTTCCCCT
101 150
386,clone14 (64) CTTGCATTTCTTTCGCTAAATCCAATCTGAACAGTAACCCGGACCATCAC
386,clone15 (101) CTTACATTTCTTTCGCTAAATCCAATCTGAACAGTAACCCGGACCATCAC
386,clone3 (56) CTTGCATTTCTTTCGCTAAATCCAATCTGAACAGTAACCCGGACCATCAC
386,clone4 (62) CTTGCATTTCTTTCGCTAAATCCAATCTGAACAGTAACCCGGACCATCAC
386Mr,clone7 (59) CTTGCATTTCTTTCGCTAAATCCAATCTGAACAGTAACCCGGACCATCAC
386Mr,clone9 (65) CTTGCATTTCTTTCGCTAAATCCAATCTGAACAGTAACCCGGACCATCAC
151 200
386,clone14 (114) GATGAGCGCTAACGATGTTCAAAACCCTCCAGTGGAAGACGACGAGGTCC
386,clone15 (151) GATGAGCGCTAACGATGTTCAAAACCCTCCAGTGGAAGACGACGAGGTCC
386,clone3 (106) GATGAGCGCTAACGATGTTCAAAACCCTCCAGTGGAAGACGACGAGGTCC
386,clone4 (112) GATGAGCGCTAACGATGTTCAAAACCCTCCAGTGGAAGACGACGAGGTCC
386Mr,clone7 (109) GATGAGCGCTAACGATGTTCAAAACCCTCCAGTGGAAGACGACGAGGTCC
386Mr,clone9 (115) GATGAGCGCTAACGATGTTCAAAACCCTCCAGTGGAAGACGACGAGGTCC
201 250
386,clone14 (164) CAGCCCTCGAAGCCGCTGACACTCAACAGGCTTCCAAGCCATCCAAGCGC
386,clone15 (201) CAACCCTCGAAGCCGCTGACACTCAACAGGCTTCCAAGCCATCCAAGCGC
386,clone3 (156) CAACCCTCGAAGCCGCTGACACTCAACAGGCTTCCAAGCCATCCAAGCGC
386,clone4 (162) CAACCCTCGAAGCCGCTGACACTCAACAGGCTTCCAAGCCATCCAAGCGC
386Mr,clone7 (159) CAACCCTCGAAGCCGCTGACACTCAACAGGCTTCCAAGCCATCCAAGCGC
386Mr,clone9 (165) CAACCCTCGAAGCCGCCGACACTCAACAGGCTTCCAAGCCATCCAAGCGC
251 300
386,clone14 (214) TACGCAAAGGCTATGGCCAAGATGGGTTTGAAACCTGAGCCAAACATCTC
386,clone15 (251) TACGCAAAGGCTATGGCCAAGATGGGTTTGAAACCTGAGCCAAACATCTC
386,clone3 (206) TACGCAAAGGCTATGGCCAAGATGGGTTTGAAACCTGAGCCAAACATCTC
386,clone4 (212) TACGCAAAGGCTATGGCCAAGATGGGTTTGAAACCTGAGCCAAACATCTC
386Mr,clone7 (209) TACGCAAAGGCTATGGCCAAGATGGGTTTGAAACCTGAGCCAAACATCTC
386Mr,clone9 (215) TACGCAAAGGCTATGGCCAAGATGGGTTTGAAACCTGAGCCAAACATCTC
301 350
386,clone14 (264) TAAGGTACACATTCGCAAACGTGCGGCGCCCTCTTTTGTGGTGAACCAAC
386,clone15 (301) TAAGGTACACATTCGCAA-CGTGCGGCGCTCTCTTTTGTGGTGAACCAAC
386,clone3 (256) TAAGGTACACATTCGCAAACGTGCGGCGCTCTCTTTTGTGGTGAACCAAC
386,clone4 (262) TAAGGTACACATTCGCAAACGTGCGGCGCTCTCTTTTGTGGTGAACCAAC
386Mr,clone7 (259) TAAGGTACACATTCGCAAACGTGCGGCGCTCTCTTTTGTGGTGAACCAAC
386Mr,clone9 (265) TAAGGTACACATTCGCAAACGTGCGGCGCTCTCTTTTGTGGTGAACCAAC
351 400
386,clone14 (314) CCGAGGTGTACCGCTTTCCAGGAACCAACACATTCCTAATCTTTGGTGAA
386,clone15 (350) CCGAGGTGTACCGCTTTCCAGGAACCAACACATTCCTAATCTTTGGTGAA
386,clone3 (306) CCGAGGTGTACCGCTTTCCAGGAACCAACACATTCCTAATCTTTGGTGAA
386,clone4 (312) CCGAGGTGTACCGCTTTCCAGGAACCAACACATTCCTAATCTTTGGTGAA
386Mr,clone7 (309) CCGAGGTGTACCGCTTTCCAGGAACCAACACATTCCTAATCTTTGGTGAA
386Mr,clone9 (315) CCGAGGTGTACCGCTTTCCAGGAACCAACACATTCCTAATCTTTGGTGAA
401 450
386,clone14 (364) GCTCAGCTTGGCGATACAACGATGCAAACGCAGGAAGCCGCAGCTCGCGC
386,clone15 (400) GCTCAGCCTGGCGACACAACGATGCAAACGCAGGAAGCCGCAGCTCGCGC
386,clone3 (356) GCTCAGCTTGGCGATACAACGATGCAAACGCAGGAAGCCGCAGCTCGCGC
386,clone4 (362) GCTCAGCTTGGCGATACAACGATGCAGACGCAGGAAGCCGCAGCTCGCGC
386Mr,clone7 (359) GCTCAGCTTGGCGATACAACGATGCAAACGCAGGGAGCCGCAGCTCGCGC
386Mr,clone9 (365) GCTCAGCTTGGCGATACAACGATGCAAACGCAGGAAGCCGCAGCTCGCGC
451 500
386,clone14 (414) GGTGTCTGGCGCCCTTGTGGAGGAGGTGGGAACTGCAGGAGAGGAAACCT
386,clone15 (450) GGTGTCTGGCGCCCTTGTGGAGGAGGTGGGAACTGCAGGAGAGGAAACCT
386,clone3 (406) GGTGTCTGGCGCCCTTGTGGAGGAGGTGGGAACTGCAGGAGAGGAAACCT
386,clone4 (412) GGTGTCTGGCGCCCTTGTGGAGGAGGTGGGAACTGCAGGAGAGGAAACCT
386Mr,clone7 (409) GGTGTCTGGCGCCCTTGTGGAGGAGGTGGGAACTGCAGGAGAGGAAACCT
386Mr,clone9 (415) GGTGTCTGGCGCCCTTGTGGAGGAGGTGGGAACTGCAGGAGAGGAAACCT
501 550
386,clone14 (464) CTGAGCTTGCTGGTGAAACTCCTGCAGCAGCTGCCCCATCTGAGAGCAAA
386,clone15 (500) CTCAGCTTGCTGGTGAAACTCCTGCAGCAGCTGCCCCATCTGAGAGCAAA
386,clone3 (456) CTGAGCTTGCTGGTGAAACTCCTGCAGCAGCTGCCCCATCTGAGAGCAAA
386,clone4 (462) CTGAGCTTGCTGGTGAAACTCCTGCAGCAGCTGCCCTATCTGAGAGCAAA
386Mr,clone7 (459) CTGAGCTTGCTGGTGAAACTCCTGCAGCAGCTGCCCCATCTGAGAGCAAA
386Mr,clone9 (465) CTGAGCTTGCTGGTGAAACTCCTGCAGCAGCTGCCCCATCTGAGAGCAAA

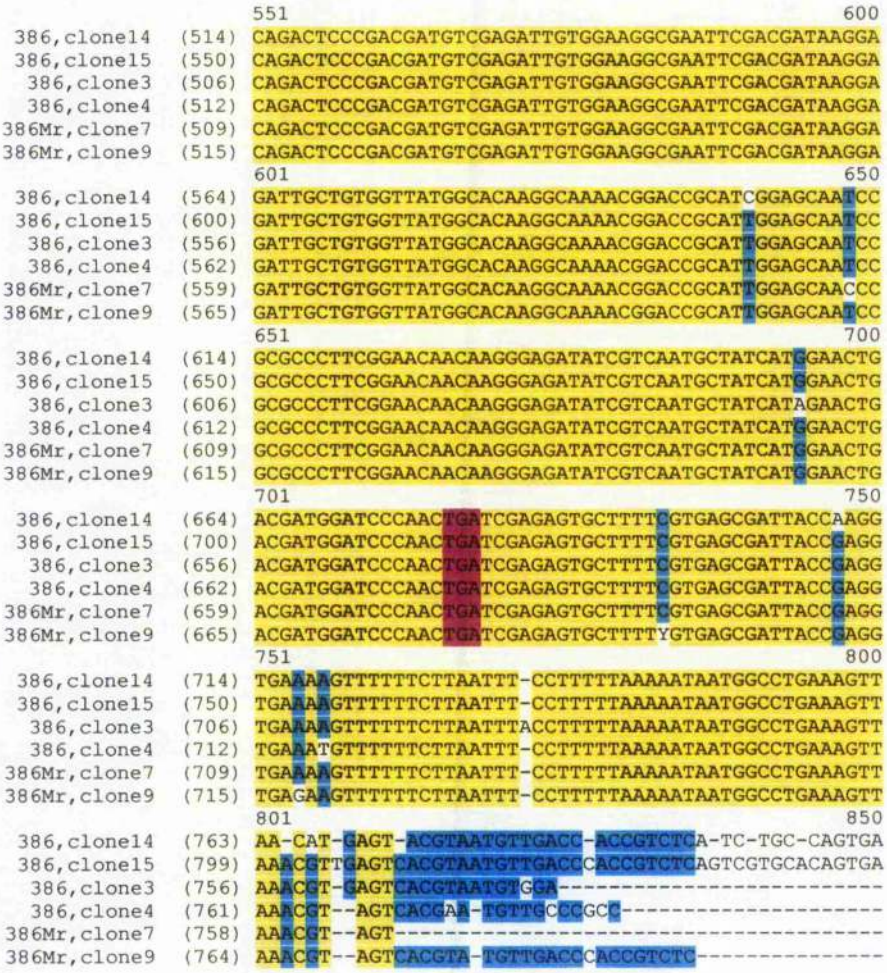


Figure 6.19: **Alignment of ORF 277 sequences from isogenic 386 lines**. Start codon (ATG) and stop codon (TGA) are indicated in pink. The sequences from line 386 (clone 3, 4, 14 and 15) are aligned against sequences from line 386Mr (clone 7 and 9) .

presence of intact RNA molecules on the membrane (Figure 6.20A). PCR products amplified from genomic DNA of line 386, for the coding region of ORF 274 were then used to probe the membrane. This probe gave a faint signal after 8 days exposure (Figure 6.20B). The membrane was then washed and hybridised in succession with probes designed in the 3'UTR of ORF 274 and the 3'UTR of ORF 277. No signal was obtained using these two probes after 8 days exposure. The lack of detection of a signal with the 3'UTR/ORF specific probes could be due to a relatively low level of transcripts from these genes making Northern blot analysis too insensitive. On this basis, RT-PCR analysis was conducted. Total cDNA was synthesized from RNA extracted from 386 and 386Mr lines as described in Materials and Methods. No product was recovered from the PCR amplification performed on the mRNA extract from 386 and 386M showing the absence of contamination in the mRNA extract. PCR amplification from these cDNA preparations was undertaken using primers for the TIM gene (positive control) and the ORF specific primers (274F/274R and 277F/277R). A PCR product corresponding to the predicted size for the ORF was identified for each ORF of 386 and 386Mr lines (Figure 6.21). Thus both ORFs are transcribed in both 386 and 386Mr lines. These results suggest that the absence of the CAR protein observed on 2D gels of the 386Mr line was not due to a mutation or a defect in transcription of one of the two ORFs.

6.3.4.3 Translation of CAR protein and detection of its isoforms

In order to confirm the identity of the CAR protein spot differentially expressed between the isogenic lines, antibodies were raised in mice against the gene products of ORF 274. These were used to identify the protein by Western blotting. The DNA vaccination technique was used in combination with a prime boost strategy to produce antibodies against ORF 274 in mice, as described in Materials and Methods.

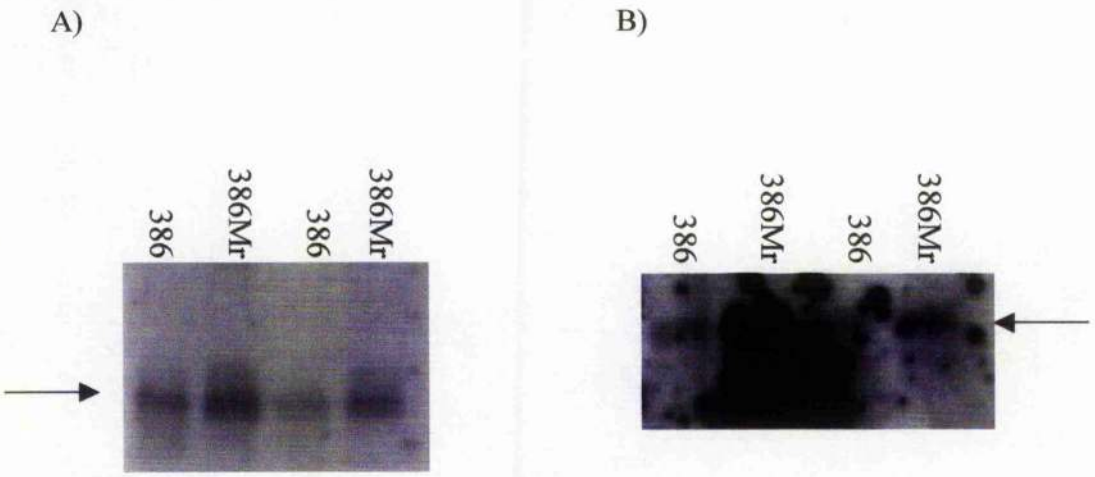


Figure 6.20: **Northern blot analysis for transcription of ORF 274 and 277.** RNA was extracted from procyclic forms of 386 and 386Mr lines. The blot was probed using the TIM product as a loading control (A) and subsequently probed in succession with a PCR product from the coding region of ORF 274 (B), a PCR product from the 3'UTR of ORF 274 and a PCR product from the 3'UTR of ORF 277. No product could be detected using the probes designed in the 3'UTR of the genes.

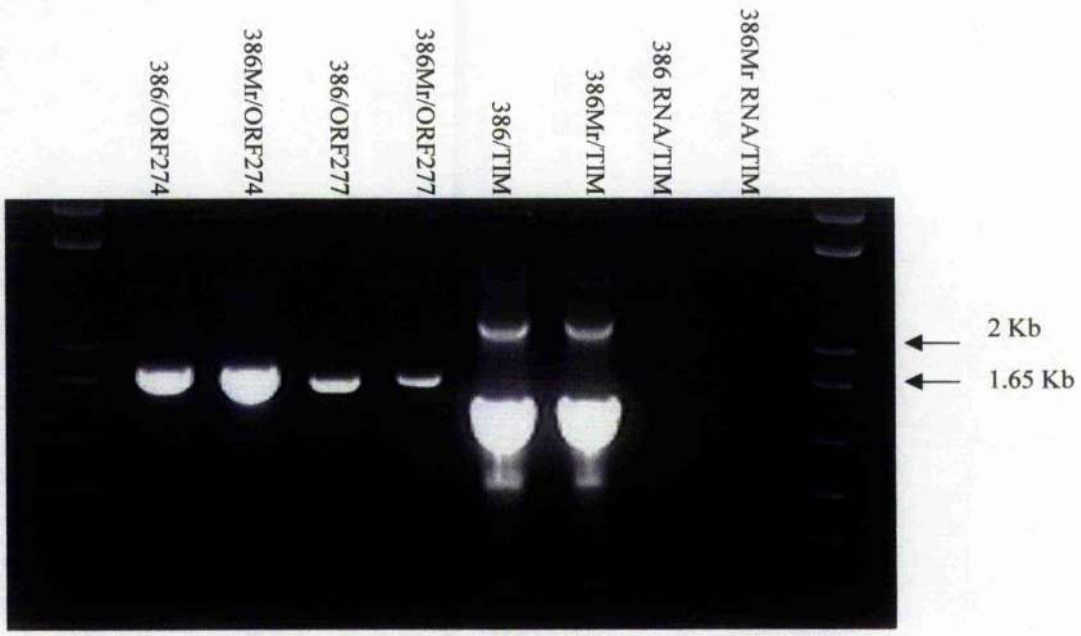
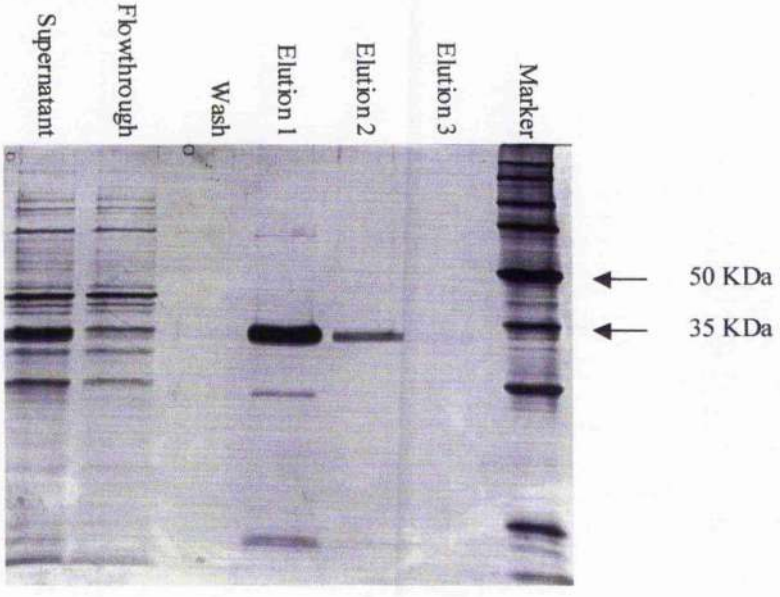


Figure 6.21: **RT-PCR analysis of ORF 274 and 277.** Total cDNA was produced from mRNA extracted from 386 and 386Mr lines. PCR reaction on total cDNA, using the ORF specific primers (274F/274R and 277F/277R), demonstrated the transcription of both ORFs in each line. The TIM primers were used in the PCR reaction as a loading control. No product could be recovered from performing the PCR reaction on the mRNA extract.

The recombinant gene product of ORF 274, to be used for boosting the antibody production, was purified from the bacterial extract using a Ni-NTA agarose column. The elution fractions from the column were run on a 12% SDS-PAGE gel and the gel stained with Coomassie Brilliant blue to assess the production and purity of the recombinant protein (Figure 6.22A). A main band was detected in elution fractions 1 and 2 with a molecular weight of about 34 kDa. This is higher than the predicted gene product size of ORF 274 (20.1 kDa). In order to confirm that this band was the gene product of ORF 274, the band was excised from the gel, trypsin digested and analysed by Tandem MS. The resulting tryptic peptides were analysed by Tandem MS and the mass spectra generated were used to search the local *T. brucei* database with the MASCOT® search engine. The Tandem mass spectra matched onto ORF 274 (score of 213) (Figure 6.22B).

Western blot analysis of whole protein extract from 386 and 386Mr lines separated on SDS-PAGE gel demonstrated the presence of the gene product of ORF 274 in both lines (Figure 6.23A) and the specificity of the antibody produced by the DNA vaccination technique (Figure 6.23B). Detection of isoforms of the gene product of ORF 274 was performed by running whole protein extract on 7 cm 2D gels and transferring the separated proteins onto a membrane. The Western blot analysis revealed the presence of three isoforms of the gene product of ORF 274 at location of 34 kDa and pH 4.5 in the 386 line and only two isoforms at this location in the 386Mr line (Figure 6.24). This experiment confirmed that the CAR protein and spot 1 were the gene products of ORF 274, as previously determined by mass spectrometry analysis. The analysis of isoforms by Western blot on 2D gel revealed the presence of another isoform of the CAR protein located adjacent to the two previously identified isoforms in the 386 line. In addition, two more protein spots were detected in both 386 and 386Mr lines using these antibodies, but these

A)



B)

Sample name	Score	Peptide sequence	Peptide score
Recombinant gene product of ORF 274	213	AALSFVVNQPEVYR	55
		GDIVNAIMELTDPN	9
		QTPDDVEIVEGEFDDK	66
		FPGTNTFLIFGEAQLGDTTM	36
		AVSGALVEEVGTAGEETSEL	48

Figure 6.22: SDS-PAGE gel of purified recombinant gene product of ORF 274. The protein was purified from the *E. coli* cell extract using a nickel NTA agarose column. The fractions from the column were run on a SDS-PAGE gel (A). The band containing the purified protein was excised and digested with trypsin. Summary of the results obtained from the MASCOT® search engine using the Tandem mass spectra generated from tryptic peptides of the purified His-tagged gene product of ORF 274 expressed in *E. coli*. The MASCOT® search engine was used to search the local *T. brucei* database (B). A score over 23 indicates that the probability of the protein from the database to be a random match is less than 5% ($p<0.05$).

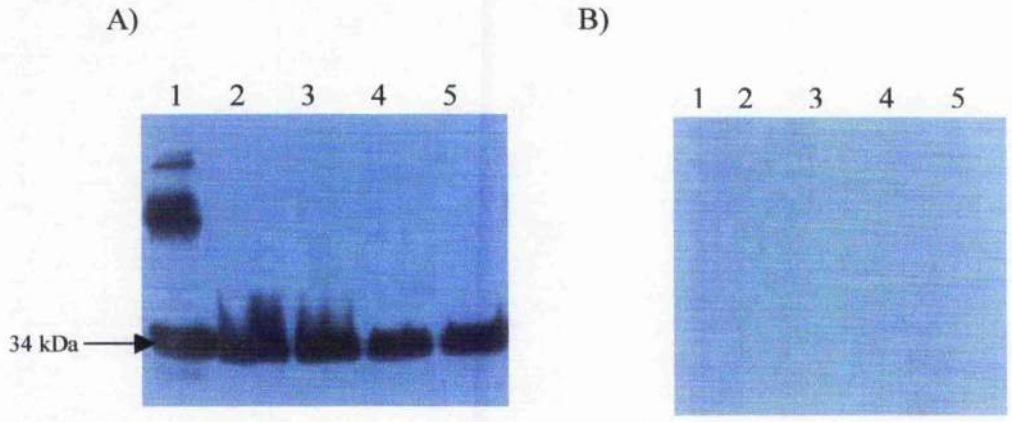


Figure 6.23: Western blot analysis of expression of the gene product of ORF 274 in isogenic 386 lines. Proteins from 386 (lane 2 and 3) and 386Mr (lane 4 and 5) lines were run on SDS-PAGE gels and transferred on a membrane. The gene product of ORF 274 was detected using antiserum from mice (1/167). The antibodies were detected with HRP conjugated anti mouse IgG (1/1,000) and the ECL detection kit. The positive control (lane 1) consisted of the recombinant gene product of ORF 274. Serum from non-immunised mice was used as a negative control (B).

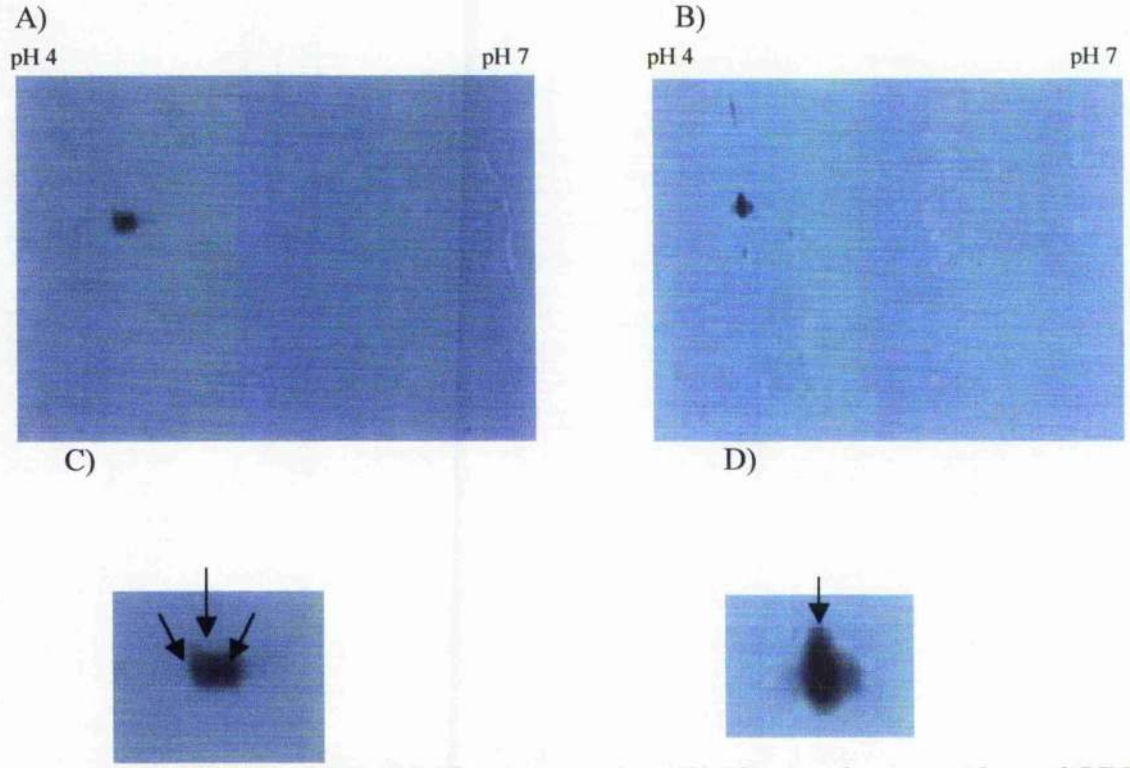


Figure 6.24: Western blot analysis of expression of isoforms of gene product of ORF 274 in isogenic 386 lines. Whole protein extract from 386 (A) and 386Mr (B) lines were separated by 2DE on pH 4-7 gels. The proteins were electro-blotted on a membrane and presence of isoforms was assessed by Western blotting using serum from immunised mice (1/10,000). Close up view of the isoforms of the gene product of ORF 274 detected in the 386 (C) and the 386Mr (D) lines.

were much less intense and were not located at 34 kDa, pH 4.5 (data not shown). These could be degradation products of spot 1 and 2.

6.3.5 Features of predicted amino acid sequence of arginine kinase

The differentially expressed protein spot located at pH 5.8 (spot B) had been identified as arginine kinase from either ORF 342 or 343. The difference in molecular weight and pI of the predicted transcript product from each ORF are shown in Table 6.4. The amino acid sequence of ORF 342 was analysed using the PredictProtein software available at the Expasy web site. The amino acid sequence contained all the characteristic consensus blocks of the guanidino kinase family, and a putative actin binding domain (CPSNLGT) (Figure 6.25). The genes are present as tandem repeat on chromosome IX of the reference line 927.

6.4 Discussion

Whole cell preparations of *T. brucei* proteins could be separated by 2DE into over 1,000 distinct protein spots on various pH ranges gels. Comparison of the 2DE pattern of proteins extracted from 386 and 386Mr confirmed the prediction that the vast majority of proteins were identical. The analysis focussed on identifying consistent differences in protein patterns between the isogenic lines in order to identify an alteration associated with Cymelarsan resistance.

The data presented provide evidence for a correlation between the absences of protein spot 2 (34 kDa, pI 4.5) and spot B (40 kDa, pI 5.8) on the 2D gels and the Cymelarsan resistance phenotype in *T. brucei* line 386. Protein spot 2 was identified as the gene product of a non-annotated ORF, and was named Cymelarsan Associated Resistance

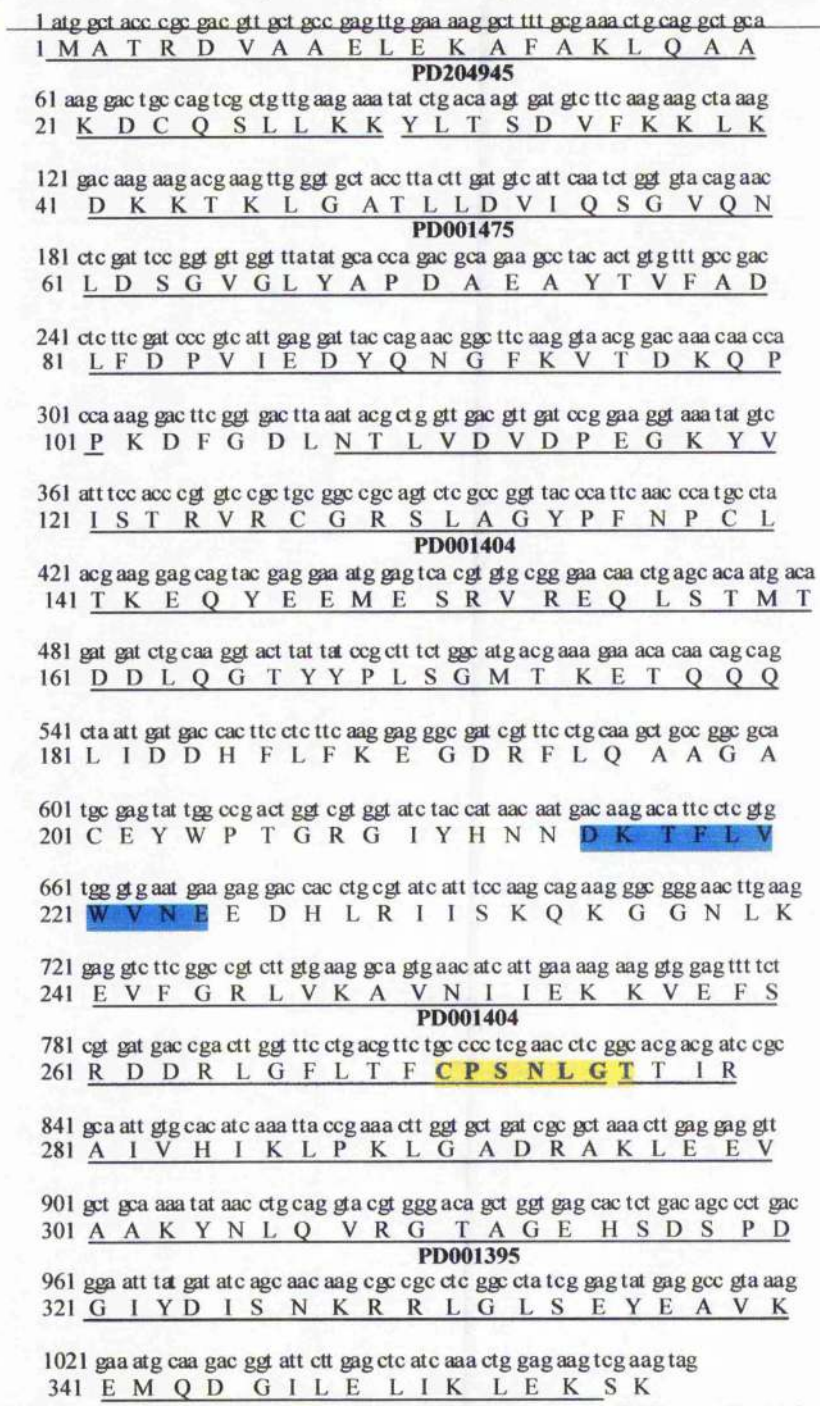


Figure 6.25: Schematic representation of the DNA and amino acid sequences of arginine kinase (ORF 345) from *T. brucei*. A signature sequence pattern of ATP:guanidine kinase (yellow) containing a reactive cysteine residue at position 271 was discovered and a putative actin binding domain (blue). Characteristic consensus sequences according to ProDom database are underlined.

Name	Location	Size	Molecular weight	pI
Chr9contig160.tmp0342	Chromosome IX	1215 bp	44.7 kDa	4.9
Chr9contig160.tmp0343	Chromosome IX	1113 bp	41.4 kDa	5.9
Chr9contig160.tmp0345	Chromosome IX	1071 bp	40.1 kDa	6.4

Table 6.4: Information about the three genes encoding arginine kinase in *T. brucei*. The information was reproduced from GeneDB.

(CAR) protein and protein spot B was identified as an arginine kinase. The correlation between the absence of the CAR protein (spot 2) and Cymelarsan resistance was demonstrated on both broad and narrow pH range gels stained with the sensitive Sypro ruby dye and insensitive Coomassie brilliant blue stain.

Tandem MS analysis identified the CAR protein as matching ORF 274 on chromosome IX of *T. brucei*. The predicted pI calculated from the sequence of ORF 274 corresponded to the location of the CAR protein on the 2D gel. However, the predicted molecular weight (20.1 kDa) was lower than the observed molecular weight (34 kDa) of the CAR protein. This was also observed on SDS-PAGE gels for the recombinant gene product of ORF 274 produced for the antibody production. A Small Ubiquitin-related MOdifier (SUMO) protein attachment site was identified in the predicted sequence of the protein. Modification of the CAR protein by a SUMO protein would add 11 kDa to the molecular weight of the protein. The SUMO proteins are covalently conjugated onto lysine residues within target sequences. Most sumoylated proteins are nuclear, and three main functional roles of SUMO have been proposed: 1) protein targeting sumoylation has been shown to be important for nuclear import, 2) enhancement of protein stability, potential competition with (and inhibition of) ubiquitination and 3) transcriptional control (i.e. negative regulation of transcription from the androgen receptor) (Hay, 2001). The molecular weight of the recombinant gene product of ORF 274 produced in *E. coli*, for the generation of antibodies in mice was measured by mass spectrometry as 22 kDa, corresponding to 2 kDa higher than the expected molecular weight. The high molecular weight observed on SDS-PAGE and 2D gels is therefore not due to the attachment of a SUMO protein. An alternative explanation for the discrepancy in molecular weight is that the CAR protein is made of 16% of acidic residues (aspartate and glutamic acid). The high proportion of

negatively charged amino acids could account for the discrepancy between expected molecular weight and the apparent molecular weight of the CAR protein on 2D and SDS-PAGE gel. The location of arginine kinase on 2D gel corresponded to the predicted molecular weight and pI determined from the gene sequence of ORF 343 in GeneDB.

A BLAST analysis of GeneDB identified the presence of another copy of ORF 274 (Figure 6.26) and two more copies of the arginine kinase gene (Figure 6.27). The genes, encoding the CAR protein and spot 1, were present as tandem repeats on chromosome IX of reference line 927. As the sequences of the two ORFs are identical one would predict that the duplication had occurred recently. Duplication is important for genes whose functions are essential for the survival of the species (Gu *et al.*, 2003).

Two protein spots, which differed in isoelectric point on 2D gels of 386, were matched onto either ORF 274 or 277. As both ORFs have identical predicted amino acid sequences and no other homologues were found in the *T. brucei* genome, it is likely that the isoform is due to a post-translational modification. In principle, comparison of the mass spectrum of the modified protein with the unmodified protein should allow the identification of the modified peptide. However, the low sequence coverage obtained in Tandem MS for these protein spots meant that the identification of the modified peptide was not possible. In order to analyse reliably the differences, a higher quantity of protein would be required. This could now be achieved by immunoprecipitation using the antiserum to the expressed product of ORF 274.

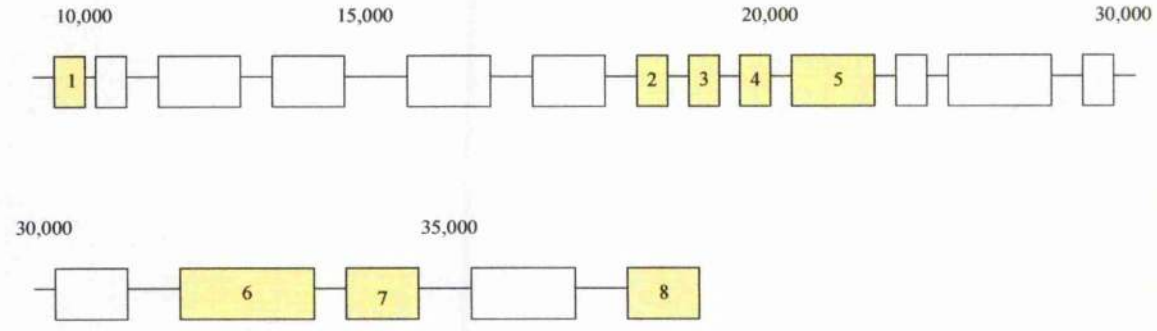


Figure 6.26: Representation of ORF 274 and 277 on chromosome IX of *T. brucei*. Complementary string of Chromosome IX (10 Kb to 35 Kb) with boxes indicating putative ORFs with homology to proteins in NCBI (yellow) or without homology to proteins in NCBI (white). Ubiquitin conjugating E2 enzyme (1), 60S ribosomal protein L10 (2), ORF 274 (3), ORF 277 (4), chaperone (5), cell cycle check point protein (6), metaloprotease like protein (7) and tyrosine phosphatase (8).

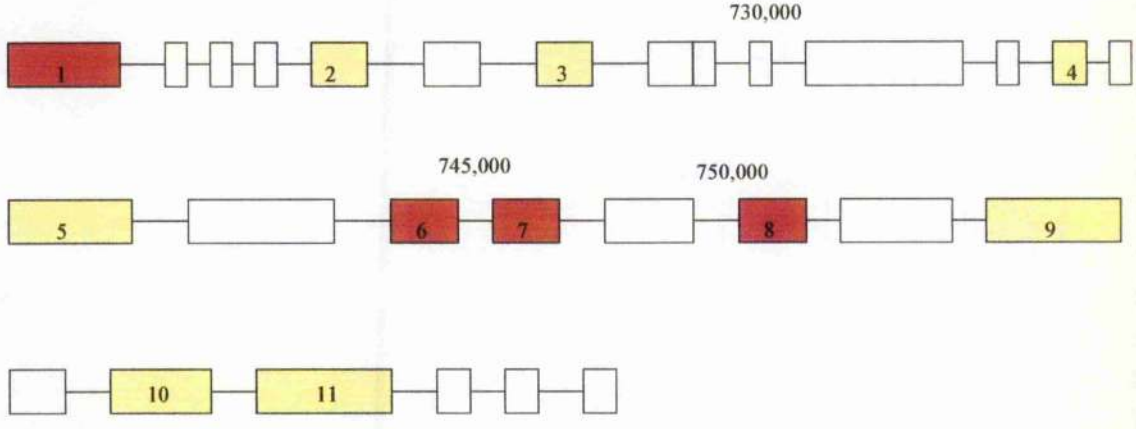


Figure 6.27: Representation of arginine kinase genes on chromosome IX of *T. brucei*. String of Chromosome IX (725 Kb to 760 Kb) with boxes indicating putative ORFs without homology to proteins in NCBI (white), putative ORFs with homology to protein in NCBI (yellow) and experimentally characterised ORFs (red). This region of the chromosome contains genes encoding glutamate dehydrogenase (1), NAD/FAD dependent dehydrogenase (2), 40s ribosomal protein S3 (3), calmodulin (4), phosphatase protein (4), conserved hypothetical protein (5), arginine kinase ORF 342 (6), arginine kinase ORF 343 (7), arginine kinase ORF 345 (8), ABC transporter (9), myo-inositol-1 phosphatase (10) and a transmembrane protein (11).

Sequence pattern searches on the predicted amino acid sequences of ORF 274 and arginine kinase revealed a number of possible modification sites, including potential phosphorylation and glycosylation sites (Figure 6.28). NetPhos predicted five threonines and two serines as putative phosphorylation sites on the gene product of ORF 274, and 14 putative phosphorylation sites on arginine kinase (ORF 342). The protein spots from the 2D gel were identified by mass spectrometry. Because post-translational modification causes a peptide mass shift, any peptide found to match onto the protein can be considered free from modification. This reduced the potential modification sites present on the gene product of ORF 274 and arginine kinase to 5 and 10 phosphorylation sites respectively (Figure 6.28). Protein phosphorylation can be involved in the activation/inactivation of enzyme activity, modulation of molecular interactions and signalling. For example, activation of a kinase by protein phosphorylation is a hallmark of intracellular signalling cascades (Hunter, 1995). Potential post-translational regulation of the gene product of ORF 274 and 277 could allow large amounts of the protein and enzyme to be present in the cell in an inactive form.

InterPro and Pfam searches revealed the presence of a nascent polypeptide associated complex (α NAC domain) at the N terminal and an ubiquitin associated domain (UBA domain) at the C terminal of the gene product of ORF 274. In humans, NAC is a heterodimeric complex of two subunits: α NAC of 33 kDa and β NAC of 21 kDa. Similarly, α (Egd2p) and β (Egd1p) subunits have been found in yeast (Shi *et al.*, 1995). The mammalian NAC complex acts in the regulation of co-translational transport of proteins into the endoplasmic reticulum, and was also shown to be involved in transcriptional regulation. In yeast however, the NAC complex has a different function as it binds ribosomes, protecting newly synthesized polypeptide chains from degradation and

A)

MSANDVQNPPVEDDEVPTLEAADTQQASKPSKRYAKAMAKMGLKPEPNIS

KVHIRKRAALSFVVNQPEVYRFPGTNTFLIFGEAQLGDTTMQTQEAAARA

VSGALVEEVGTAGEETSELAGETPAAAAPSESKQTPDDVEIVEGEFDDKE

IAVVMAQGKTDRIGAIRALRNNKGDIVNAIMELTMDPN

B)

MATRDVAAELEKAFAKLQAAKDCQSLLKKYLTSDVFKKLKKKTKLGATLLDVIQ

SGVQNLDSGVGLYAPDAEAYTVFADLFDPVIEDYQNGFKVTDKQPPKDFGDLNTL

VDVDPEGKYVISTRVRCGRSLAGYPFNPCLTKEQYEEMESRVREQLSTMTDDLQG

TYYPLSGMTKETQQQLIDDHFLFKEGDRFLQAAGACEYWPTGRGIYHNNDKTFLV

WVNEEDHLRIISKQKGGNLKEVFGRLVKAVNIIEKKVEFSRDDRLGFLTFCPSNLG

TTIRAIVHIKLPKLGADRAKLFEVAAKYNLQVRGTAGEHSDSPDGIYDISNKRRLG

LSEYEAVKEMQDGILELIKLEKSK

Figure 6.28: Predicted post-translational modification sites on gene product of ORF 274 (A) and arginine kinase (B). Potential threonine and serine phosphorylation sites (yellow), as predicted by NetPhos (http://www.cbs.dtu.dk/services/NetPhos), N-glycosylation site (blue) as predicted by NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc), and SUMO protein attachment site (red) as predicted by Abgent (http://www.abgent.com/cgi-bin/tools.pl) are shown on the amino acid sequence. No myristoylation (http://mendel.imp.univie.ac.at/myristate/SUPLpredictor.htm) or sulfination sites (http://us.expasy.org/cgi-bin/sulfinator.pl) were found. The sequences underlined correspond to the matching peptide fragmented by Tandem MS from spot 2 and spot B.

preventing any inappropriate translocation into the reticulum (Wang *et al.*, 1995). The role of such domains in *T. brucei* remains to be elucidated.

The UBA domain was first identified in 1996 in proteins that function in ubiquitin mediated proteolysis (Hofmann and Bucher, 1996). Since then this domain has been found in proteins that lack any obvious link to proteolysis. The structure and function of the UBA domain has been mainly studied in humans and yeast. The UBA domain is composed of 35 amino acid residues which form a sequence motif containing three alpha helices (Figure 6.29). Structure analysis of various UBA containing proteins suggest that conserved residues contribute to the common structure of the UBA domain, while divergent residues on the surface of the helices would permit the selective interaction with different protein targets. Target proteins include ubiquitin, multi-ubiquitin chains, ubiquitinated proteins and other effectors, predicting a role for this motif in protein-protein interaction and subcellular targeting (Wilkinson *et al.*, 2001; Mueller and Feigon, 2002; Hofmann and Bucher, 1996). The association of the NAC and UBA domains in a protein is not uncommon, as shown in yeast with Egdp2 (Shi *et al.*, 1995; Wang *et al.*, 1995). The association of the two domains in one protein could indicate that the function of this protein is in the degradation or refolding of newly synthesised proteins with folding defects.

Arginine kinase belongs to a family of conserved proteins with phosphotransferase activity, the guanidine kinases. Arginine kinase catalyse the synthesis of N-phosphorylated guanidine compounds by the reversible transfer of a phosphoryl group from ATP to an enzyme-specific guanidine acceptor. Phosphoarginine, the reaction product, plays a crucial role as an energy reserve because the high-energy phosphate can be transferred when a

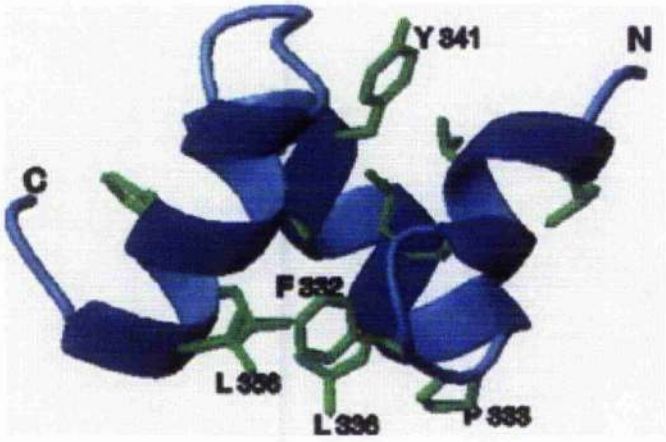


Figure 6.29: **UBA domain structure**. Ribbon representation of the structure of a human DNA repair protein UBA domain as determined by NMR spectroscopy. The representation sows the characteristic compact three-helix bundle encoded by the conserved residues of all UBA domains (reproduced from Dieckmann *et al.*, 1998).

renewal of ATP is needed. ATP consumption and high energy phosphate regeneration are required to support processes such as stage differentiation, cell division and motility. Phosphoarginine is presumed to support rapid bursts of cellular activity in the short term until metabolic pathways such as glycogenolysis, glycolysis and oxidative phosphorylation are switched on (Hird and McLean, 1986). Arginine kinase has been identified in *T. brucei* (Pereira *et al.*, 2002) and *T. cruzi* (Pereira *et al.*, 2000; Pereira *et al.*, 1999). These gene sequences contain all the characteristic consensus blocks of the guanidino kinase family and a putative actin-binding domain (Pereira *et al.*, 2000).

The results presented in this chapter suggest that the absence of the CAR protein, an isoform of spot 1, is associated with Cymelarsan resistance in *T. brucei*. The question arises as to whether the change in CAR expression is a general feature of Cymelarsan resistance in *T. brucei* or is a particular attribute of the 386Mr line only. A proteomic analysis of a second pair of isogenic lines of Cymelarsan resistant and sensitive *T. brucei* lines has been undertaken and the results are presented in the next chapter.

Chapter VII

Association of the CAR protein and Cymelarsan resistance in line 247 and 927

7.1 Introduction

The proteomic study performed on the pair of isogenic lines of 386 demonstrated a positive correlation between the absence of the CAR protein (located at pH 4.5, 34 kDa) with resistance to Cymelarsan in *T. brucei* (chapter 6). The CAR protein and its isoforms are encoded by two ORF's present as a tandem repeat on chromosome IX. Analysis of these ORFs in the 386 and 386Mr lines demonstrated the absence of sequence differences between the ORFs that could account for the presence of the CAR protein isoform of spot 1, in the sensitive 386 line. In addition, both ORFs were transcribed in both isogenic 386 lines. Consequently, it was suggested that the presence of three protein spots encoded by ORF 274 and 277 (one of which being the CAR protein spot) in the sensitive line was due to a post-translational modification, and that the protein spots encoded by ORF 274 and 277 in the resistant line is not modified.

The question arises as to whether the change in CAR protein expression is a general feature of Cymelarsan resistance or is a particular attribute of the 386Mr line only. To address this question a 2DE analysis was undertaken with of a second pair of isogenic lines of Cymelarsan sensitive and resistant lines of 247. The Cymelarsan resistant line of 247 was generated from the sensitive 247 line *in-vivo* by increasing sub-curative doses of Cymelarsan (Scott *et al.*, 1996). This line was produced following the same protocol as the 368 isogenic lines.

Resistance to Cymelarsan in the procyclic form of the laboratory induced resistant line of 247 was shown to be 1,257 fold greater than in the parental line (chapter 5).

Proteins from Cymelarsan sensitive and resistant 247 lines were separated by 2DE and analysed by Tandem MS. Subtractive analysis of the region of the gel at pH 4.5 and 34 kDa was undertaken. Western blot analysis using antibodies raised against the transcriptional product of ORF 274 was performed on 2D gels to analyse the CAR protein spot and its isoforms in 247 and 247Mr lines. In addition, a 2DE analysis of the genome reference line 927 was undertaken. In each case, the analysis was focussed on expression of transcriptional products of ORF 274 and 277 rather than undertaking a general proteome scan.

7.2 Materials and Methods

7.2.1 *In vitro* culture of procyclic trypanosomes

Procyclic forms of the three lines were grown *in-vitro* in complete SDM79 as described in chapter 2.

7.2.2 Protein extraction, 2DE and mass spectrometry

Whole soluble protein extract was obtained from $1x10^8$ procyclic forms of trypanosomes using protocol P2 as described in chapter 2. Fresh protein extracts were made for each gel as proteins from the 247 lines seemed to degrade when stored at -20°C (data not shown).
The whole soluble protein extract was separated by 2DE on pH 4-7 gels as described in chapter 6. The gels were stained with either Sypro orange or colloidal Coomassie blue stain.

For Sypro orange staining, the gels were fixed overnight in a solution of 7.5% acetic acid and 10% methanol. The gels were incubated in 0.005% SDS for 1 h and then transferred in the Sypro orange dye diluted (1/5,000) in 7.5% acetic acid in which they were left overnight. Gels were rinsed briefly in a solution of 7.5% acetic acid and 30% methanol, and images were acquired immediately with the Typhoon scanner as previously described (chapter 2) using the green laser.

For colloidal Coomassie blue staining, the gel was fixed for 3 h in a solution of acetic acid (10%) and ethanol (40%). The gel was washed twice in water for 10 min before being transferred into the colloidal Coomassie blue stain (1 part methanol/1 part colloidal stock (50 g ammonium sulphate, 6 ml ortho-phosphoric acid, 500 ml ddH_2O, 10 ml of a 5% Coomassie blue G 250 solution)). The gel was left to stain for three days. The gel images were acquired as described in chapter 2.

Protein spots of interest were excised from the gel and digested with trypsin using the in-gel protein digestion protocol described in chapter 3. Tandem MS analysis of the proteins was performed as described in chapter 4.

7.2.3 SDS-PAGE gels and Western blot analysis

SDS PAGE gels and Western blot analysis were performed as described in chapter 6.

7.3 Results

7.3.1 Proteomic analysis of the isogenic 247 lines

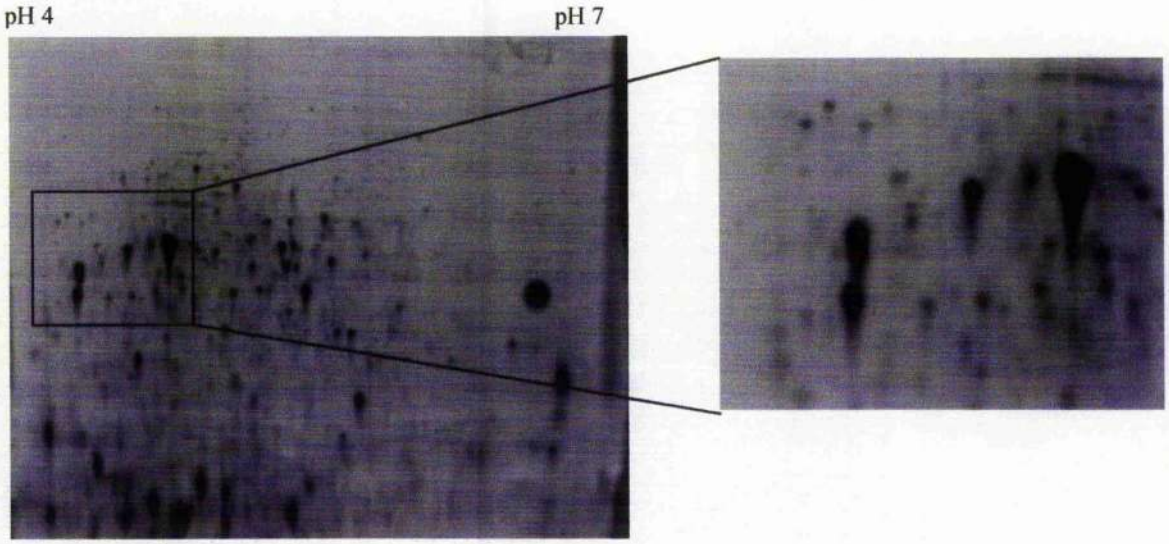
Proteins from the isogenic lines of 247 were extracted and separated by 2DE on pH 4-7 gels to detect the CAR protein spot and its isoforms in 247 and 247Mr lines. Tandem mass

spectrometry was performed to identify the protein spots located at and around the expected location of the CAR protein on the 2D gels. The gels were stained with Sypro orange. Analysis of the gels using the ImageMaster 2DElite showed 363 proteins spots on the gel from the 247 line and 388 protein spots on the gel from the 247Mr line. The general 2D pattern of proteins from the 247 and 247Mr lines was very similar, as expected from isogenic lines (Figure 7.1). In addition, comparison of the gels from the two pairs of isogenic lines (386 and 247) revealed a high similarity in 2D pattern on pH 4-7 gels. Comparison of the 2D pattern in the area of the gel at approximately 34 kDa and pH 4.5 between 247 and 247Mr lines showed no differentially expressed protein spots. Protein spots located in this area were excised from the gels, digested with trypsin and analysed by Tandem MS. Protein spots 1, 3 and 4 from the sensitive line 2D gel and 1, 2, 3, and 4 from the resistant line 2D gel were identified with high confidence as β tubulins (Figure 7.2). β tubulins had previously been found to surround the CAR protein spots on 2D gels from the isogenic lines of 386 (chapter 6). The identification of these spots demonstrated that the location on the gels from the 247 lines corresponded to the location of the CAR protein spots on gels from the 386 lines. No protein spots were detected at the expected location of the CAR protein spots on the gels from 247 and 247Mr (Figure 7.1).

7.3.2 Presence of ORF 274 and 277 in isogenic 247 lines

Since the CAR protein spot and its isoforms could not be visualised on 2D gels of 247 and 247Mr lines at the expected location, the question arises as to whether the ORFs encoding the CAR protein exist in the 247 genome. To address this question, a PCR was performed on genomic DNA of lines 247 and 247Mr using ORF specific primers (274F/274R, 277F/277R). PCR products of the expected size were observed for both ORF 274 and 277 in 247 and 247Mr lines (Figure 7.3). In addition, a smaller PCR product was also amplified using these

A)



B)

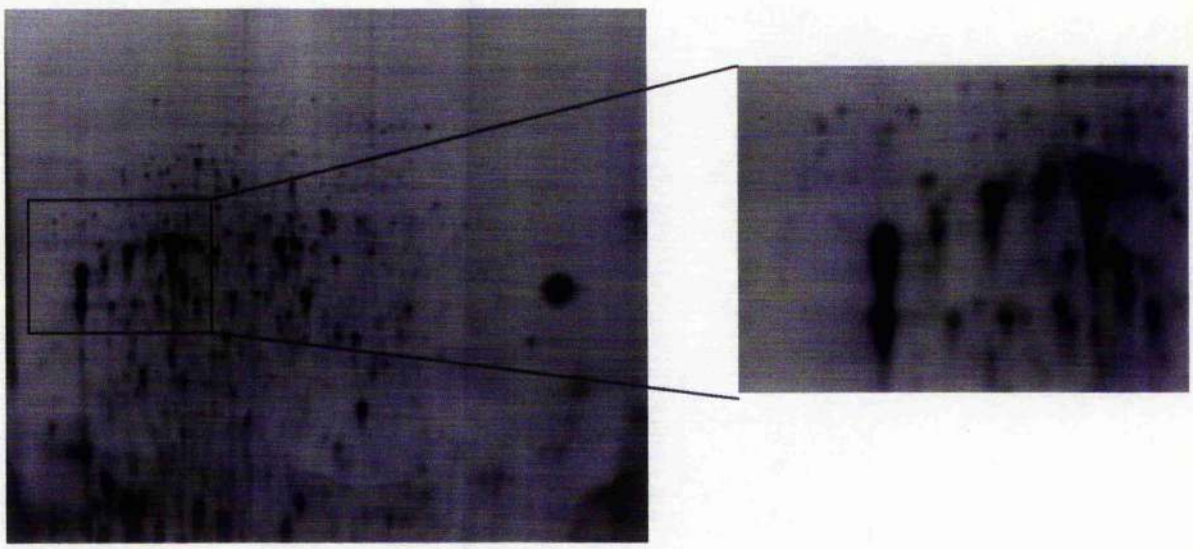
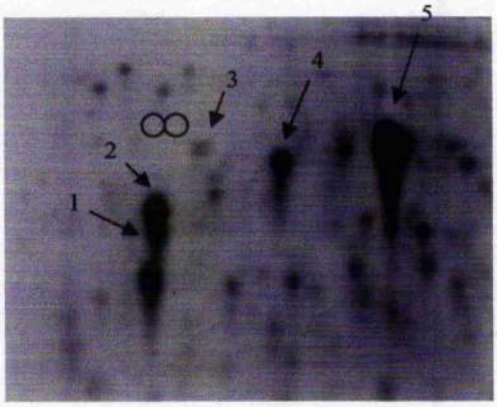
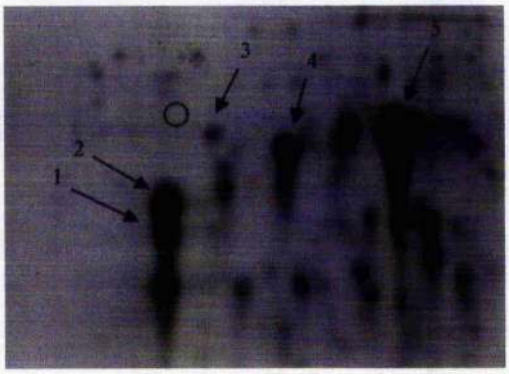


Figure 7.1: 2D gels of procyclic forms of isogenic 247 lines, pH 4-7. Broad pH range isoelectric focussing (pH4-7) 2D gels of procyclic forms of 247 (A) and 247Mr (B) lines. The gels were stained with Sypro orange.

A)



B)



C)

	Name	Score	Peptide identified	gi/number
247				
1	Tubulin beta chain	223	4	9366734
2	No ID			
3	Tubulin beta chain	133	3	9366734
4	Tubulin beta chain	116	2	9366734
5	Chaperonin HSP60	404	7	18277736
247Mr				
1	Tubulin beta chain	270	6	9366734
2	Tubulin beta chain	143	3	9366734
3	Tubulin beta chain	123	2	9366734
4	Tubulin beta chain	99	2	9366734
5	Chaperonin HSP60	65	2	18277736

Figure 7.2: Tandem mass spectrometry analysis of protein spots located in the expected region of the CAR protein. Location of the protein spots from 247 (A) and 247Mr (B) lines. The table (C) gives a summary of the results obtained searching the NCBInr database using the MASCOT® search engine with the mass spectra acquired. The empty circles indicate the expected position of the CAR protein spot and its isoform.

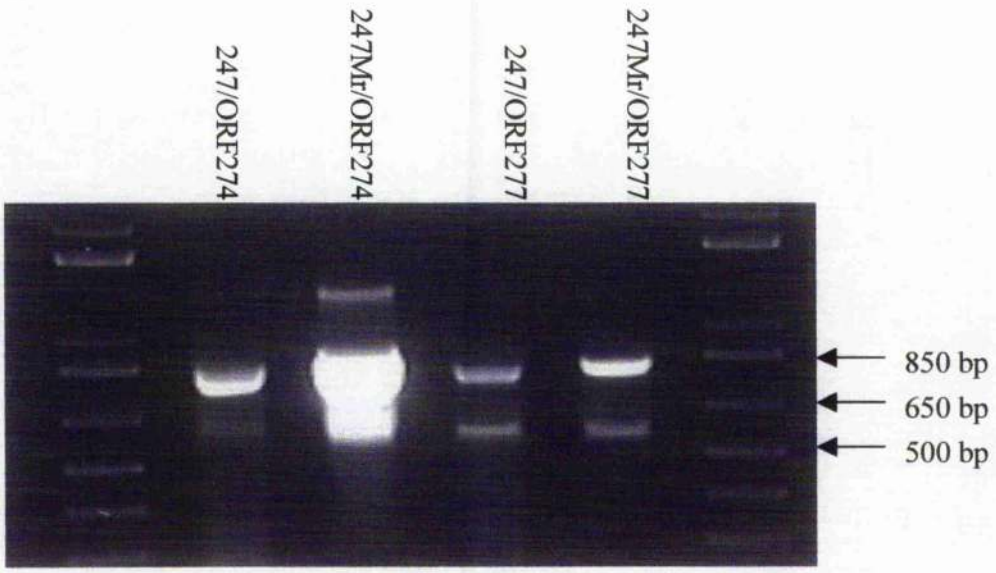


Figure 7.3: **PCR amplification of ORF 274 and 277 from isogenic 247 lines.** The PCR was performed using the ORF specific primers. The main product corresponded to the expected size for each ORF (ORF 274 = 891 bp, ORF 277 = 868 bp).

primers. This smaller product was not recovered from PCR amplification of genomic DNA from the 927, 386 and 386Mr lines.

These results would suggest that the pair of isogenic 247 lines both contain ORF 274 and 277 in their genome.

7.3.3 Presence of expressed product of ORF 274 and 277 in isogenic 247 lines

The CAR protein spot and its isoform could not be detected on 2D gels of 247 and 247Mr lines at its expected location. However, ORF 274 and 277, encoding the CAR protein, were shown to be present in each line. In view of these results a Western blot analysis of whole cell extract separated by SDS-PAGE gel was performed. This analysis revealed the presence of ORF 274 protein in both lines (Figure 7.4). The protein detected had a similar molecular weight to the recombinant His tagged transcriptional product of ORF 274 (34 kDa). Higher molecular weight bands were also detected in the lane containing purified recombinant His tagged protein. These probably correspond to dimers and trimers of the transcriptional product of ORF 274, however no molecular weight marker was run on the gel so the size cannot be determined.

To assess the number of CAR protein isoforms and their pIs in 247 and 247 Mr lines, 2D gels of these lines were run and Western blotted. Only one transcriptional product of ORF 274 could be detected at 34 kDa, pH 4.5 for both 247 and 247Mr lines (Figure 7.5). Other products were detected at lower molecular weight and basic pH in both 247 and 247Mr lines. These fainter spots might correspond to degradation products or could indicate a cross-specificity

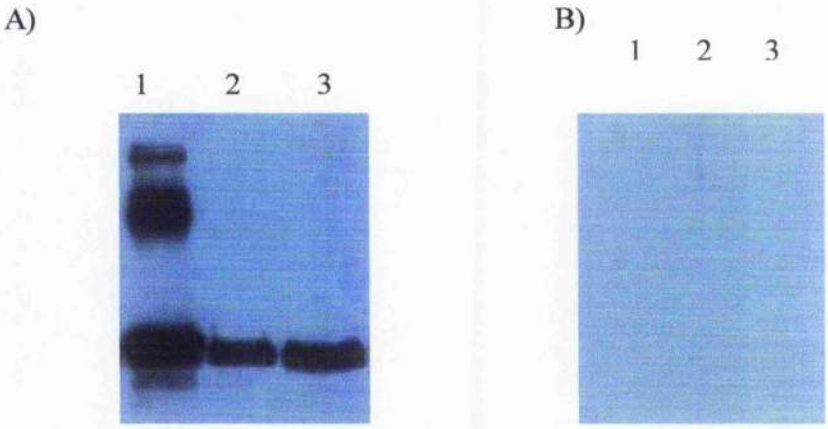


Figure 7.4: Western blot analysis of expression of the gene product of ORF 274 in the isogenic 247 lines. Protein extracts from 247 (lane 2) and 247Mr (lane 3) were separated by SDS-PAGE gels and electro-transferred onto a nitrocellulose membrane. The membrane was blotted with mouse anti-CAR antibody and detected with HRP-conjugated anti-mouse IgG (1/1000) and the ECL detection kit. The positive control was the His-tagged recombinant CAR protein (lane 1).
Serum from non-immunised mice was used as negative control (B).

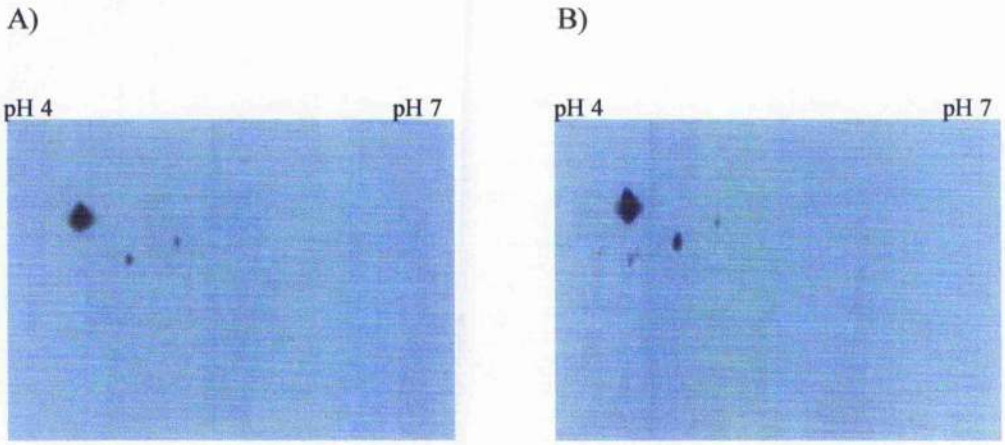


Figure 7.5: Western blot analysis of expression of isoforms of the gene product of ORF 274 in isogenic 247 lines. Cymelarsan sensitive (A) and resistant (B) 247 lines.

of the antiserum to another protein not detected on one dimensional SDS-PAGE gels.

7.3.4 Proteomic analysis of line 927

Protein extracted from procyclic forms of line 927 were separated by 2DE on pH 4-7 gel and stained with colloidal Coomassie blue. I am grateful to Anne Fladas for running and staining the gel. Examination of the 2D gel revealed a similar general 2D pattern of protein spots at location of 34 kDa and pH 4.5 as lines of 386 and 247. A protein spot located on the 2D gel at the expected position of a CAR protein isoform was detected. This spot (spot 6) was excised from the gel, digested with trypsin and analysed by Tandem MS. The MASCOT[40] search engine was used to search the NCBInr and the local *T. brucei* database. The protein spot matched onto ORF 274 encoding the CAR protein with a score of 95, indicating an correct identification within 95% confidence limits (Figure 7.6).

These results would suggest that only one isoform of the CAR protein was present in line 927 at a location of 34 kDa and pH 4.5.

7.4 Discussion

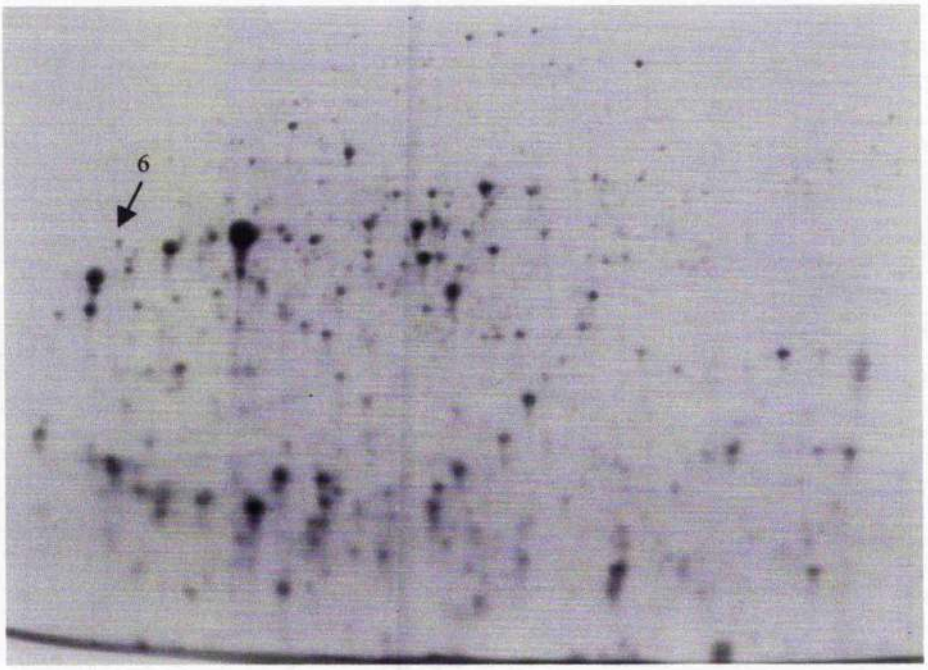
The presence of the CAR protein spot and its isoform in *T. brucei* isogenic Cymelarsan sensitive and resistant 247 lines and the genome reference 927 line was assessed by 2DE analysis and Western blot analysis as the absence of the CAR protein was associated with resistance to Cymelarsan in the 386 lines.

No protein spot could be detected by Sypro orange staining at the expected location on the CAR protein spot and its isoforms in 247 and 247Mr. However, a single protein product of ORF 274 was detected by Western blot analysis of 2D gels at the expected location for both Cymelarsan sensitive and resistant 247 lines. This suggests that the quantity of protein of ORF 274/277 expressed in the 247 lines is lower than the sensitivity of the Sypro orange stain, or alternatively that it doesn't stain with Sypro orange. In addition, these results suggest that in the isogenic 247 lines the transcriptional product of ORF 274 and 277 was perhaps expressed at a lower level than in 386 lines where the CAR protein is readily detectable by Coomassie Brilliant blue.

Only one transcriptional product of ORF 274 and 277 was detected at pH 4.5, 34 kDa in both 247 and 247Mr lines, suggesting that the CAR protein is not associated with Cymelarsan resistance in this pair of *T. brucei* isogenic lines.

A possible clue to the mechanism of Cymelarsan resistance in *T. brucei* is the diminished metabolism of the Alamar Blue dye in the 386Mr and 247Mr (chapter 5). This new phenotype associated with cymelarsan resistance in the two pairs of isogenic lines will be further investigated in the next chapter.

A)



B)

Line	Name	Score	Peptide identified	Contig name
927	CAR protein	95	3	ORF 274

C)

Contig name	Score	p score	Peptide sequence
Tryp9.0.00322-274	95	56	AALSFVVNQPEVYR
			QTPDDVEIVEGEFDDK
			GDIVNAIMELTMDPN

Figure 7.6: Mass spectrometry analysis of protein spot located at the expected region of the CAR protein in line 927. Analysis of colloidal Coomassie blue stained 2D gels of protein extracted from 927 revealed the presence of a spot at the region of the CAR protein (A). Results from the MASCOT® search engine searching the local *T. brucei* database are presented in table B and C.

Chapter VIII

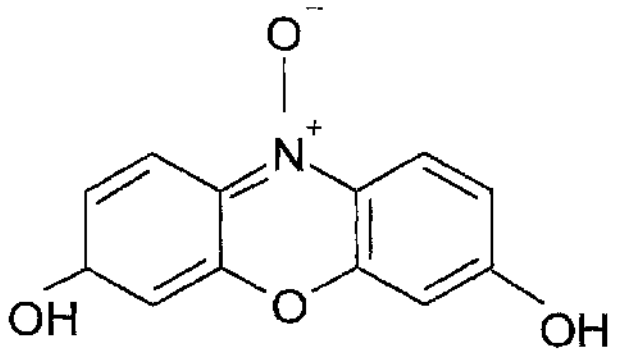
The Alamar Blue phenotype

8.1 Introduction

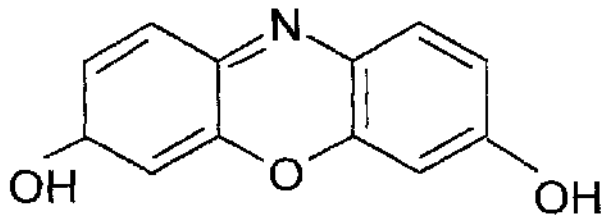
The use of the Alamar Blue dye in *in vitro* assays for testing sensitivity of cells to drugs is based on the ability of intracellular enzymes to reduce the non-fluorescent blue dye to a pink fluorescent compound (Figure 8.1). The reduction of the dye correlates with growth and metabolic activity of the cells in the presence/absence of drug, and can be measured by spectrofluorometry and spectrometry. The Alamar Blue assay was recently developed as an *in-vitro* assay to test drug sensitivity/resistance of procyclic trypanosomes in culture (Räz *et al.*, 1997). The assay was shown to be simple, fast, reproducible and economical.

The assay performed in the absence of drug on procyclic stages of each pair of isogenic lines, demonstrated a marked difference in the level of reduction of the Alamar Blue dye between the Cymelarsan sensitive and resistant lines as shown in chapter 5. In the presence of increasing concentrations of Cymelarsan, the sensitive lines (247 and 386) showed a dose dependent response with a decrease in fluorescence from Alamar Blue as the trypanosomes viability is lost. In contrast, the two resistant lines (386Mr and 247Mr) showed no change in the level of Alamar Blue fluorescence with increasing levels of Cymelarsan (0.001 μM to 1,000 μM).

A)



B)



C)

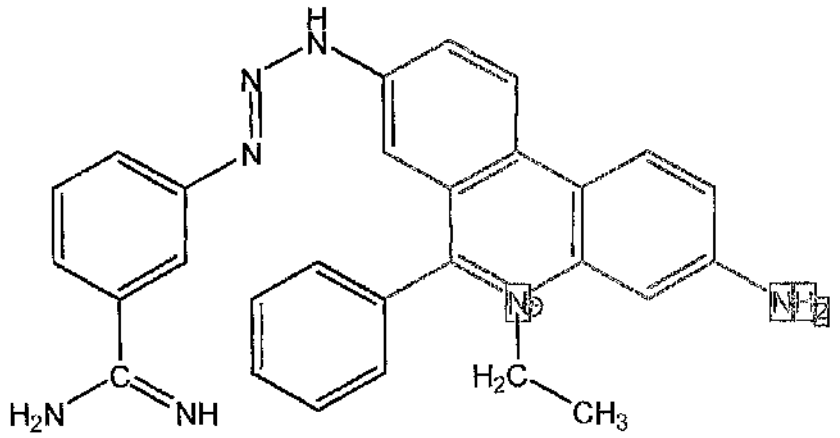


Figure 8.1: **Structure of the Alamar Blue dye and the Isometamidium.** The reduced form of the Alamar Blue dye (B) is fluorescent and can be detected by spectrofluorometry or spectrometry. The structural similarities between the oxidized Alamar Blue structure (A) and Isometamidium (C) are highlighted in blue. The structure of Isometamidium correspond to the active form (red isoform) of the drug found in Samorin®.

The fluorescence from the Alamar Blue dye produced by the Cymelarsan resistant lines was almost constant whether the trypanosomes were alive or dead. However, the fluorescence from the Alamar Blue dye was higher in the presence of Cymelarsan resistant lines than without trypanosomes in the medium, demonstrating some reduction of the Alamar Blue dye by the trypanosomes (chapter 5). The decrease ability to reduce Alamar Blue shown by the Cymelarsan resistant lines demonstrated an unexpected association of the Alamar Blue phenotype with Cymelarsan resistance.

This new phenotype associated with Cymelarsan resistance might potentially be linked to the mechanism of Cymelarsan resistance in *T. brucei*. In this chapter, an attempt is made to identify the mechanism that impaired the reduction of the Alamar Blue dye in the Cymelarsan resistant lines. The Alamar Blue dye can only be reduced by intracellular enzymes (O'Brien *et al.*, 2000), therefore two possible mechanisms of decrease in reduction of Alamar Blue were tested: the presence and activity of reductive enzymes in the lines and the transport/export of the Alamar Blue dye into/from the trypanosomes.

To test for the presence and activity of reductive enzymes, the following question was addressed:

1- is there any difference in cell lysate reductive activity of Alamar Blue between the sensitive and resistant lines?

To test whether and how the Alamar Blue dye was taken up by trypanosomes, the following questions were addressed:

1- What are the effect of P-glycoprotein inhibitors on the reduction of the Alamar Blue dye?

2- What is the activity of the transporter for Isometamidium, which possess a common structure with the Alamar Blue dye?

3- What are the actions of chloroquine and other lysosomotropic agents on the reduction of the Alamar Blue dye?

4- What is the localisation of the reduced Alamar Blue due inside trypanosomes?

8.2 Materials and Methods

8.2.1 Commonly used reagents

Ammonium chloride, Bafilomycin A1, Concanamycin A, chloroquine, verapamil, PCP, TFP were purchased from SIGMA. LS 55 Luminescence Spectrometer was from Perkin Elmer instruments. The fluorescent microscope and camera were purchased from Zeiss. Samorin® was purchased from May and Baker (UK). Alamar Blue was purchased from Trek Diagnostic System, Inc. and Cymelarsan was a kind gift from Rhône Merieux. Chloroquine was a gift from Dr Lisa Ranford-Cartwright (University of Glasgow).

8.2.2 *In vitro* culture of procyclic trypanosomes and cell lysate preparation

Procyclic stages of each pair of isogenic lines were grown *in vitro* in complete SDM79 as described in chapter 2.

Cell lysates were prepared in PBS in the presence of protease inhibitor cocktail and nucleases as described in chapter 2. The protein concentration from the cell lysate was determined using a Coomassie Plus Protein Assay (Pierce) according to the manufacturer's instructions, using bovine serum albumin as the protein standard.

8.2.3 *In vitro* drug sensitivity testing

8.2.3.1 The Alamar Blue assay

The Alamar Blue assay was performed on procyclic cultures of the two pairs of isogenic lines (386/386Mr and 247/247Mr) as described in chapter 5.

Doubling serial dilutions of test compounds (P-glycoprotein inhibitors, chloroquine, Bafilomycin A1, Concanamycin A and ammonium chloride) in complete SDM79 were set out in a 96 well plate in duplicate, in a volume of 100 µl (Table 8.1). 100 µl of trypanosome culture ($2x10^6$ trypanosome/ml) were added to each well, with 20 µl of the Alamar Blue dye. Measurements with a LS 55 Luminescence Spectrometer were taken, after 24 h incubation at 25°C, as described in chapter 5. Each experiment was performed in duplicate and replicated at least once. The presence and viability of the trypanosomes in each well was confirmed by microscopy.

The Alamar Blue assay was also performed directly on cell lysates from each pair of the isogenic lines. In the first experiment, the cell lysate (100 µl) and Alamar Blue dye (10 µl) were mixed to assess the level of reduction of the dye by the intracellular enzymes from Cymelarsan sensitive and resistant lines of 386 and 247. The measurements were taken after incubation for 24 h at 25°C. In the second experiment, the assay was performed in the presence of doubling serial dilutions of chloroquine in PBS to assess the action of chloroquine on the intracellular enzymes. 100 µl of chloroquine (0.005 to 50 µM) was mixed with cell lysate (100 µl) and Alamar Blue dye (20 µl), incubated for 24 h at 25 °C and the reduction of

		Final concentration	
Inhibitor	**Compound**	**Minimum**	**Maximum**
Inhibitor of P-glycoprotein	Verapamil	0.01 mM	1 mM
	PCP	0.01 mM	1 mM
	TFP	0.01 mM	1 mM
Inhibitor of proton pump	Chloroquine	5 μM	50 μM
	Bafilomycin A1	0.001 μM	32 μM
	Concanamycin A	0.01 μM	29 μM
	Ammonium chloride	0.01 mM	1 M

Table 8.1: Concentrations of test compounds used in the Alamar Blue assay.

the Alamar Blue dye was recorded using a LS 55 Luminescence Spectrometer as described in chapter 5. The experiment was performed in duplicate and repeated at least once.

8.2.3.2 Growth inhibition assay

To assess for the relationship between the decrease in fluorescence of the Alamar Blue dye and resistance to the trypanocidal drug Isometamidium, an *in vitro* growth inhibition assay was performed as previously described in chapter 5. The drug was serially diluted in complete SDM79 with final concentrations ranging from 0.001 μM to 1,000 μM. The percentage inhibition is expressed relative to the cell densities in cultures growing in the absence of Isometamidium. The experiment was carried out in triplicate and was repeated at least three times on each pair of the isogenic lines.

8.2.4 Fluorescence microscopy

The localisation of the reduced Alamar Blue dye and Isometamidium in procyclic stages of each pair of the isogenic trypanosomes were investigated by fluorescence microscopy using the intrinsic fluorescence of these compounds.

The localisation of the Alamar Blue dye in procyclic trypanosomes was observed by adding 20 μl of the Alamar Blue dye to 200 μl of procyclic cultures. After incubating for 1 min at room temperature, 5 μl of live trypanosomes were placed on a slide and covered with a coverslip. Cells were immediately examined on a Zeiss UV microscope (Axiovert 200M) fitted with a DAPI filter, and images taken with a camera (AxioCam) using Axiovision software, version 3.0 (Zeiss). Trypanosomes were also observed under phase contrast to assess the morphological integrity of the cells.

To assess for the effect of chloroquine on the localisation of the Alamar Blue dye inside the trypanosomes, chloroquine was mixed with the Alamar Blue dye in SDM79 at a concentration of 50 μM, and added to procyclic trypanosomes of each pair of isogenic lines. The mixture was immediately placed on a slide, covered with a coverslip and viewed using a DAPI filter and phase contrast microscopy as previously described.

Procyclic stages of the isogenic lines (10 μl at $6x10^6$ trypanosome/ml) were mixed with 10 μM Isometamidium (10 μl) in SDM79. The presence of the Isometamidium drug inside each pair of isogenic lines was assessed by fluorescence microscopy as previously described. The viability and morphological integrity of the cells were assessed by phase contrast microscopy.

8.3 Results

8.3.1 Enzymatic reduction of the Alamar Blue dye

To test whether the difference in the level of reduced Alamar Blue dye produced by Cymelarsan sensitive and resistant lines was due to a defect in the ability of the parasite's intracellular enzyme to metabolise the Alamar Blue dye, Alamar Blue assays were performed on cell lysate from each pair of isogenic lines. The fluorescence, recorded at 590 nm, showed no detectable difference in the level of reduction of the Alamar Blue dye between cell lysates from Cymelarsan sensitive and resistant lines for both pairs of isogenic lines, as compared to the difference in fluorescence observed for live trypanosomes (Table 8.2).

	Line	Average protein concentration (mg/ml)	Average optical density (at 530 nm)	Ratio: optical density/concentration
	Control	0	25.42	-
Cell lysate	386	2.98	276.06	92.63
	386Mr	2.58	217.08	84.14
	247	3.40	395.58	116.34
	247Mr	3.40	326.35	95.98
Procyclic trypanosome	386	-	251.16	-
	386Mr	-	177.08	-
	247	-	219.53	-
	247Mr	-	123.97	-

Table 8.2: *In vitro* Alamar Blue assays on cell lysates from the two pairs of isogenic lines. The cell lysates were prepared in PBS and the control consisted of PBS only. The optical density obtained for procyclic trypanosomes ($4x10^6$ trypanosome/ml) was added in the table for comparison of Alamar blue reduction from live trypanosome and cell lysates.

These results would suggest that both Cymelarsan sensitive and resistant lines possess the enzymes able to reduce the Alamar Blue dye, and that the enzymes are active in both Cymelarsan sensitive and resistant lines.

8.3.2 Assessment of the efflux of the Alamar Blue dye by P-glycoprotein

Efflux of drugs through a membrane protein, P-glycoprotein, has been reported as a mechanism of drug resistance in *T. cruzi* (Neal *et al.*, 1989) and *T. congolense* (Sutherland and Holmes, 1993). An increased efflux of the Alamar Blue dye could explain the decreased fluorescence produced by the Cymelarsan resistant lines. To assess this hypothesis the Alamar Blue assay was performed in the presence of three general inhibitors of P-glycoproteins (verapamil, PCP and TFP) on each pair of isogenic lines. These three compounds have been reported to revert drug resistance by inhibition of P-glycoproteins at concentrations of 5 μM for TFP and PCP and 15 μM verapamil in Leishmania (Basselin *et al.*, 2002) and 8 μM for verapamil in *T. cruzi* (Neal *et al.*, 1989). For both pairs of isogenic lines, verapamil, PCP and TFP were lethal to trypanosomes at concentrations of 1 mM, 0.1 mM and 0.1 mM respectively as assessed by microscopic examination of the cultures. At non-lethal concentrations, the presence of the P-glycoprotein inhibitors had no effect on the level of reduced Alamar Blue dye produced by the 386Mr and 247Mr after 24 h exposure (Figure 8.2 and 8.3). Similarly, the level of reduced Alamar Blue dye produced by the 386 and 247 lines was not altered by the presence of P-glycoprotein inhibitors at sub-lethal doses. These results suggest that the low level of fluorescence obtained from the reduction of the Alamar Blue dye observed in Cymelarsan resistant lines is not associated with an increased efflux of the dye by an over-expressed P-glycoprotein.

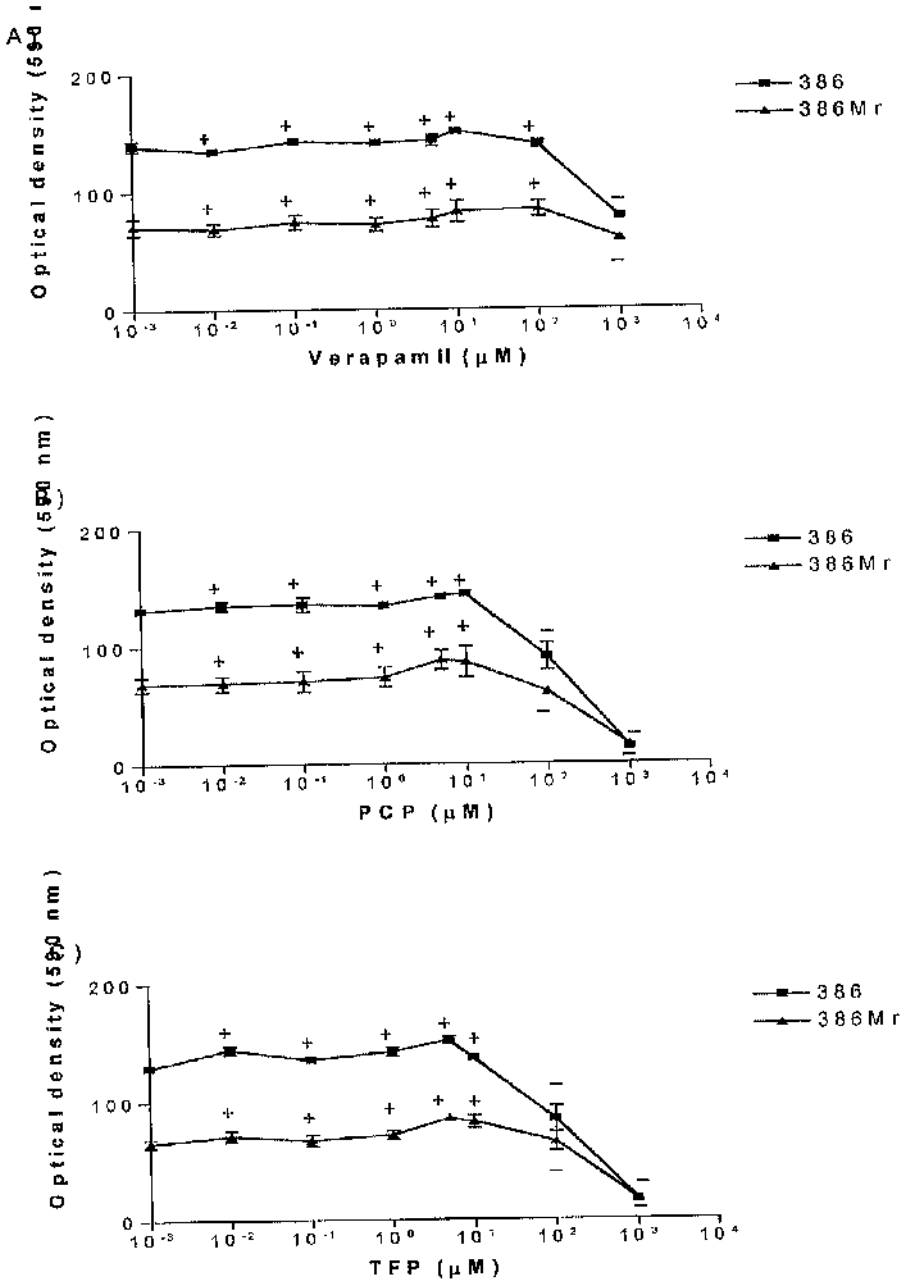


Figure 8.2: Alamar Blue assays in the presence of P-glycoprotein inhibitors for the isogenic lines of 386. Procyclic forms were incubated in the presence of the Alamar Blue dye and three common P-glycoprotein inhibitors, verapamil (A), prochlorperazine (PCP, B) and trifluoperazine (TFP, C) in SDM 79. The status of the procyclic culture is indicated for each concentration of inhibitors as alive (+) or dead (-) as assessed by microscopy. The lowest concentration corresponds to the control in the absence of drug. The error bars represent the mean +/- SEM, n = 4.

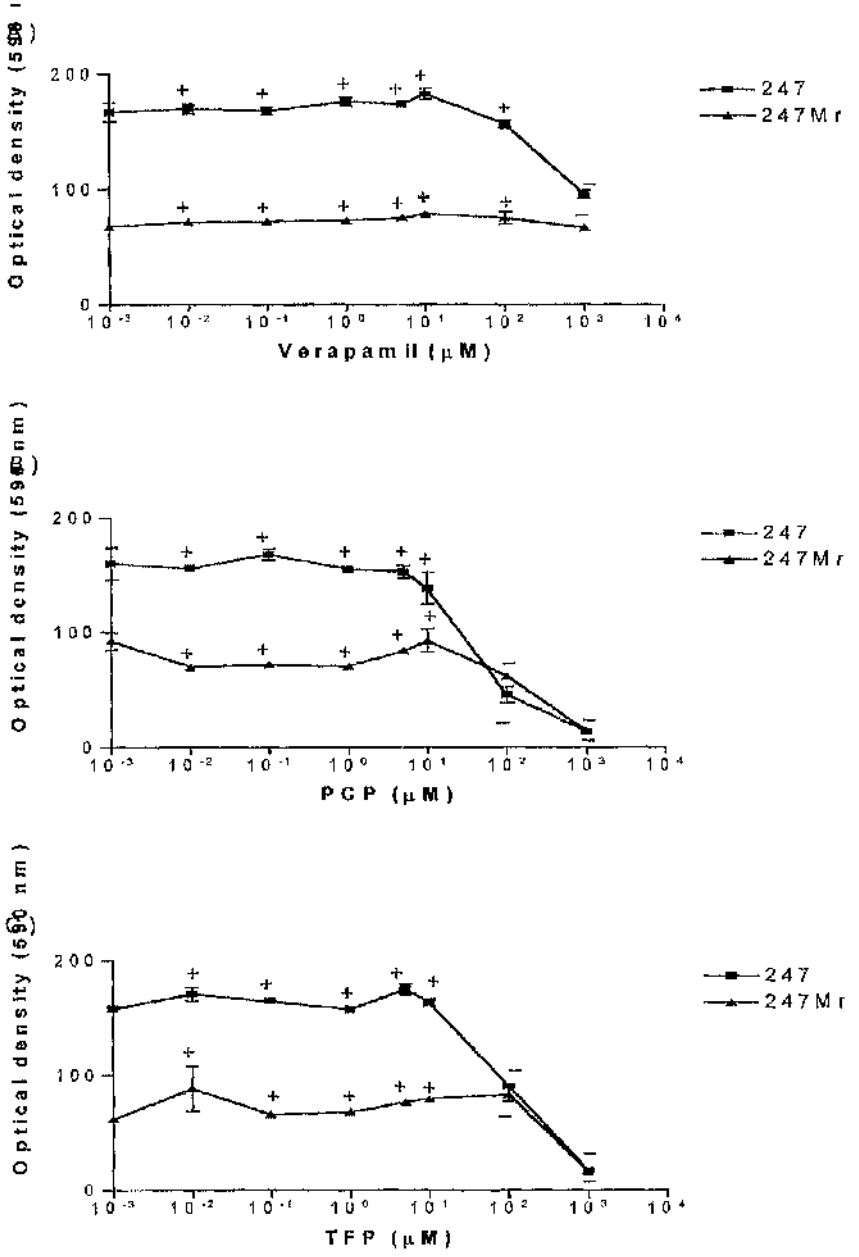


Figure 8.3: Alamar Blue assays in the presence of P-glycoprotein inhibitors for the isogenic lines of 247. Procyclic forms were incubated in the presence of the Alamar Blue and three common P-glycoprotein inhibitors, verapamil (A), prochlorperazine (PCP, B) and trifluoperazine (TFP, C) in SDM79. The status of the procyclic culture is indicated for each concentration of inhibitors as alive (+) or dead (-) as assessed by microscopy. The lowest concentration corresponds to the control in the absence of drug. The error bars represent the mean +/- SEM, n = 4.

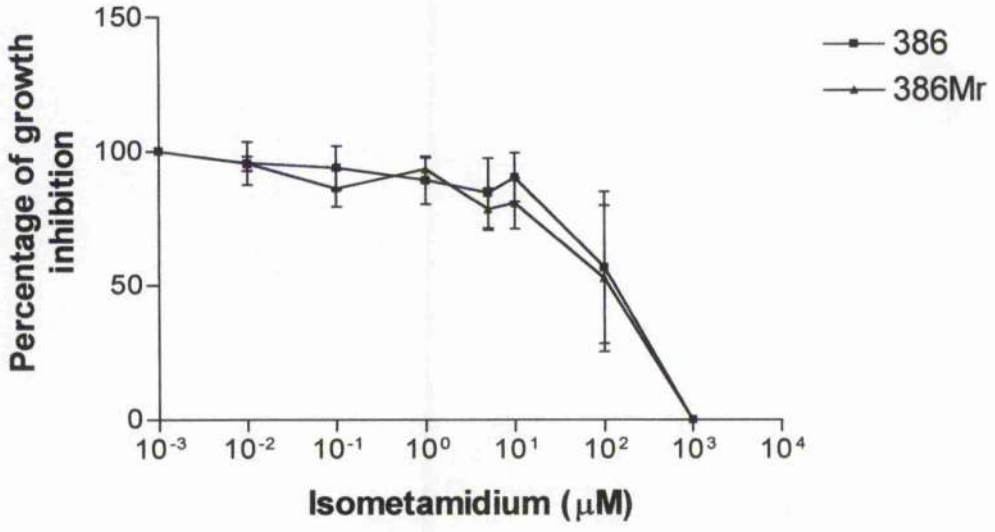
8.3.3 Mode of transport of the Alamar Blue dye

8.3.3.1 A possible transporter for the Alamar Blue dye

Isometamidium is a trypanocidal drug structurally similar to the Alamar Blue dye (Figure 8.1). It possesses the same three phenyl ring structure, a positively charged quaternary ammonium in the same position on the ring, and a NH_2 group (corresponding to a OH group in the Alamar Blue dye) able to form a hydrogen bond (Figure 8.1). These structural similarities could imply an uptake of the Alamar Blue dye mediated via the same transporter that is used by Isometamidium (de Koning, personal communication). To test whether there was a defect in the Isometamidium transporter in the two Cymelarsan resistant lines that could be linked to the decrease level of reduced Alamar Blue dye produced, a growth inhibition assay was performed on isogenic lines of 386 and 247 in the presence of Isometamidium. Isometamidium is a naturally fluorescent drug which stains the nucleus and the kinetoplast of trypanosomes. This characteristic was also used to assess the uptake of Isometamidium into the trypanosomes.

In vitro growth inhibition assays of procyclic stages of each pair of isogenic lines were performed in the presence of serially diluted Isometamidium in SDM79. Both Cymelarsan sensitive and resistant lines exhibited a dose-dependent response to the drug. No detectable differences in dose-dependent inhibition were observed between sensitive and resistant lines for both pairs of isogenic lines (Figure 8.4A and 8.5A). Fluorescence microscopy of the procyclic forms of the isogenic lines incubated in 10 μM of Isometamidium using the DAPI filter showed the staining of the kinetoplast by the drug (Figure 8.4C and 8.5C). The stained

A)



B)

C)

Line 386

Line 386Mr

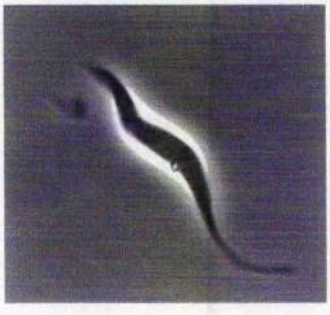
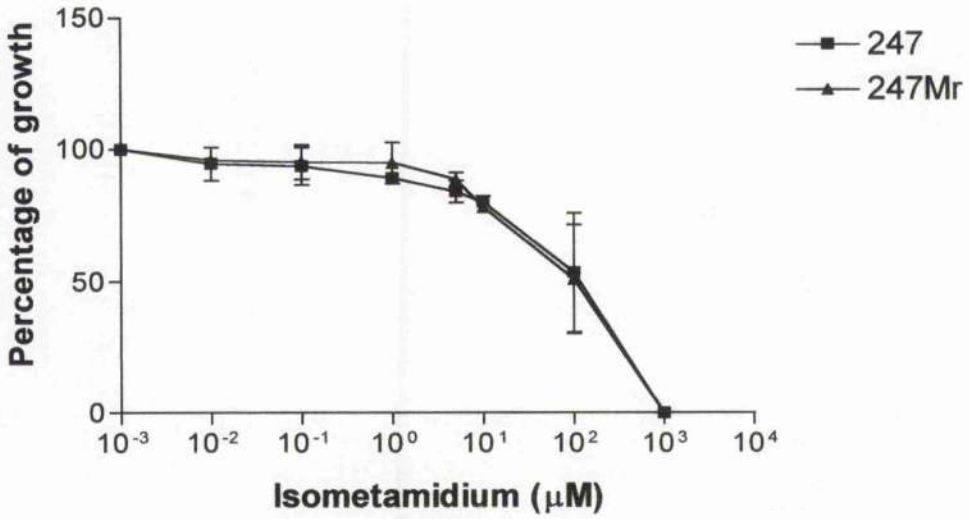


Figure 8.4: **Isometamidium in the isogenic 386 lines**. *In vitro* growth inhibition assay of procylic forms of the isogenic 386 lines (A) for Isometamidium in SDM79. Phase contrast (B) and DAPI (C) images of isogenic 386 lines showing the staining of the kinetoplast by the drug. The lowest concentration on the graph corresponds to the control in the absence of drug. The error bars represent the mean +/- SEM, n = 9.

A)



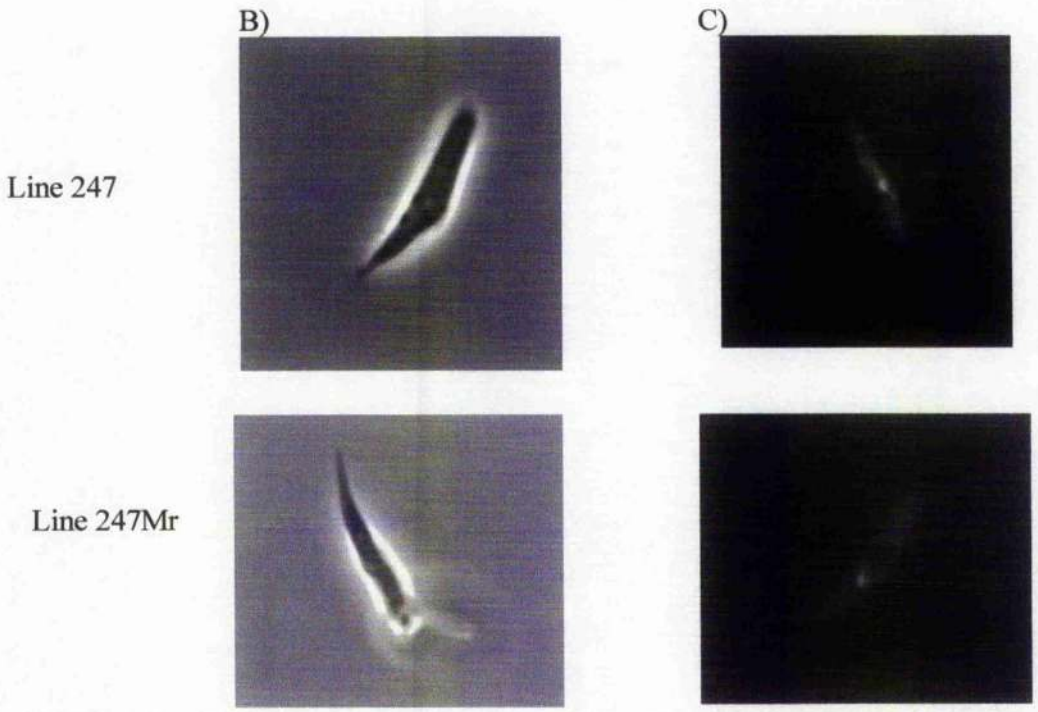


Figure 8.5: **Isometamidium in the isogenic 247 lines**. *In vitro* growth inhibition assay of procyclic forms of isogenic 247 lines in the presence of Isometamidium in SDM79. Phase contrast (B) and under DAPI filter (C) images of the lines showing the staining of the kinetoplast by Isometamidium. The lowest concentration on the graph corresponds to the control in the absence of drug. The error bars represent the mean +/- SEM, n = 9.

kinetoplast was detectable in both Cymelarsan sensitive and resistant lines, demonstrating the occurrence of Isometamidium uptake. The pairs of isogenic lines were also observed under phase contrast (Figure 8.4B and 8.5B) to assess the morphological integrity of the trypanosomes. These results suggest that a defect in the transporter mediating the uptake of Isometamidium is not associated to neither the Alamar Blue phenotype nor Cymelarsan resistance.

8.3.3.2 Localisation of the reduced Alamar Blue dye in trypanosomes

When reduction of Alamar Blue takes place the dye transforms from a charged, non-fluorescent, blue compound to a non-charged, fluorescent, pink compound. To assess the localisation and potential concentration of the reduced Alamar Blue dye inside trypanosomes, procyclic forms of each pair of isogenic lines were incubated in the presence of the Alamar Blue dye and examined by fluorescence microscopy using a DAPI filter. Fluorescence microscopy showed the presence and accumulation of the reduced Alamar Blue dye in vesicular-like organelles (Figure 8.6). These organelles were distributed throughout the cell in both sensitive and resistant lines of 386 and 247. The fluorescent organelles can be observed almost immediately after mixing the dye with the trypanosomes. The morphological integrity of the trypanosomes was intact as assessed by phase contrast microscopy (Figure 8.6).

8.3.3.2 Endocytosis

The weak base chloroquine is a lysosomotropic agent known to reduce the acidity of endosomes and lysosomes (Krogstad *et al.*, 1985). Chloroquine has been used, in many organisms, to assess the uptake and degradation of material through the endocytic pathway and is thought to act by blocking the formation of the secondary lysosome. Studies on the

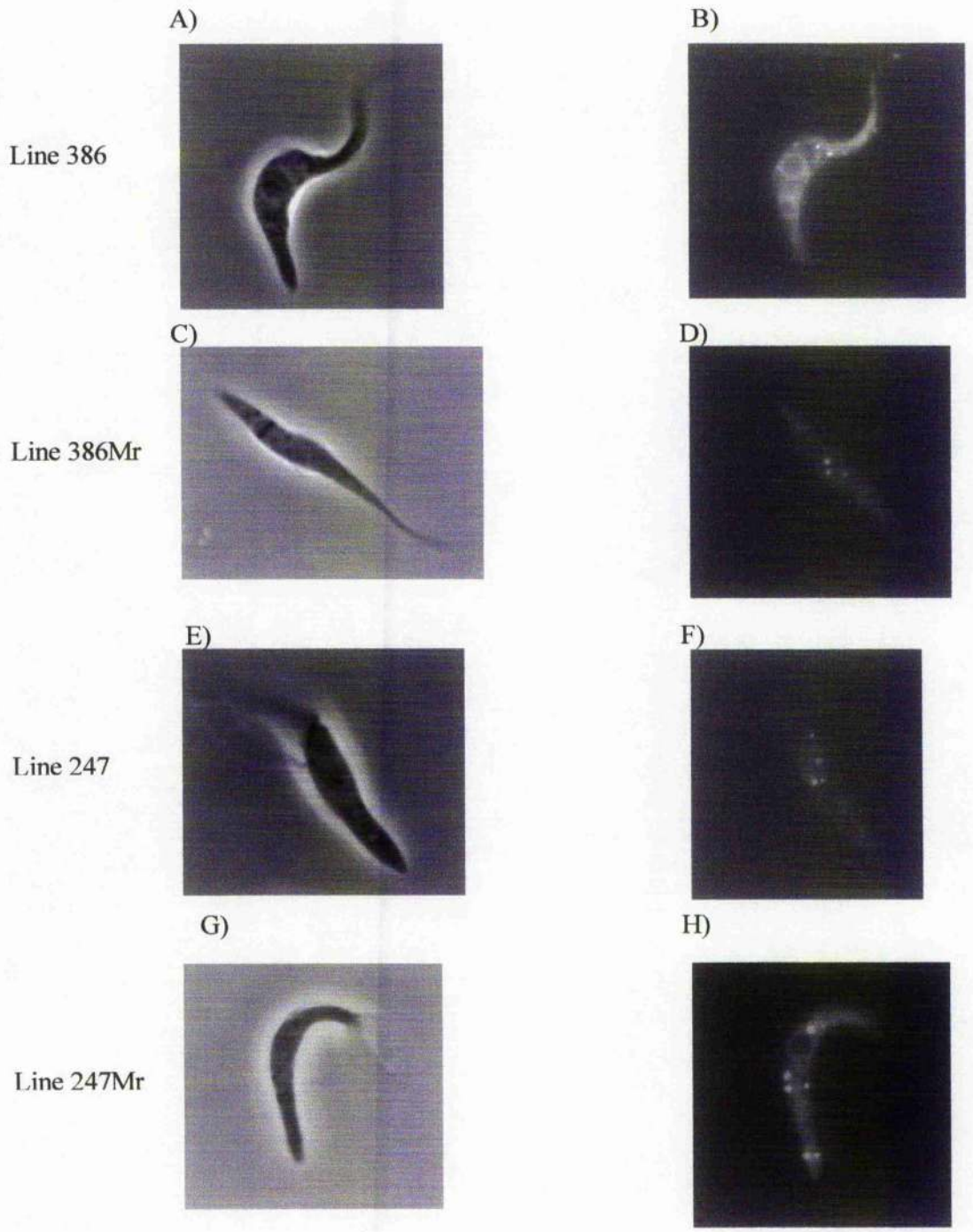


Figure 8.6: **Localisation of the Alamar Blue dye in the pairs of isogenic lines**. Phase contrast and fluorescence microscopy of trypanosomes after incubation in Alamar Blue showed the presence and concentration of the reduced Alamar Blue dye in vesicle like pockets inside the trypanosomes.

internalisation of HDL and anti-CRAM IgG by endocytosis, for example, revealed that exposure of procyclic trypanosomes to 50 μM of chloroquine completely abolished the lysosomal degradation of these molecules (Lee *et al.*, 1999; Liu *et al.*, 2000).

Chloroquine was used to determine whether the Alamar Blue dye was transported into trypanosomes via the endocytic pathway. The Alamar Blue assay was performed on procyclic stages of each pair of isogenic lines in concentrations of chloroquine up to 50 μM. A dose dependent decrease in the fluorescence from the Alamar Blue dye was observed for the sensitive lines and at a concentration of 50 μM of chloroquine, the fluorescence had dropped to a level similar to that produced by the Cymelarsan resistant lines (Figure 8.7A and 8.8A). The presence of chloroquine in the medium had no detectable effect on the fluorescence from the Alamar Blue dye produced by the Cymelarsan resistant lines.

Fluorescence microscopy performed on the procyclic forms of each pair of isogenic lines incubated in the presence of chloroquine and Alamar Blue showed the reduced dye in vesicle-like organelles in both sensitive and resistant lines (Figure 8.7C/E and 8.8C/E). These structures were distributed throughout the cell, as previously seen in the absence of chloroquine (Figure 8.6). The organelles appeared larger and more numerous in the trypanosomes incubated in the presence of chloroquine. The trypanosomes were intact as assessed by phase contrast microscopy (Figure 8.7B/D and 8.8B/D).

Chloroquine is believed to act on the endocytic pathway at the level of secondary lysosome formation by reducing the acidity of endosomes and lysosomes through the inhibition of the

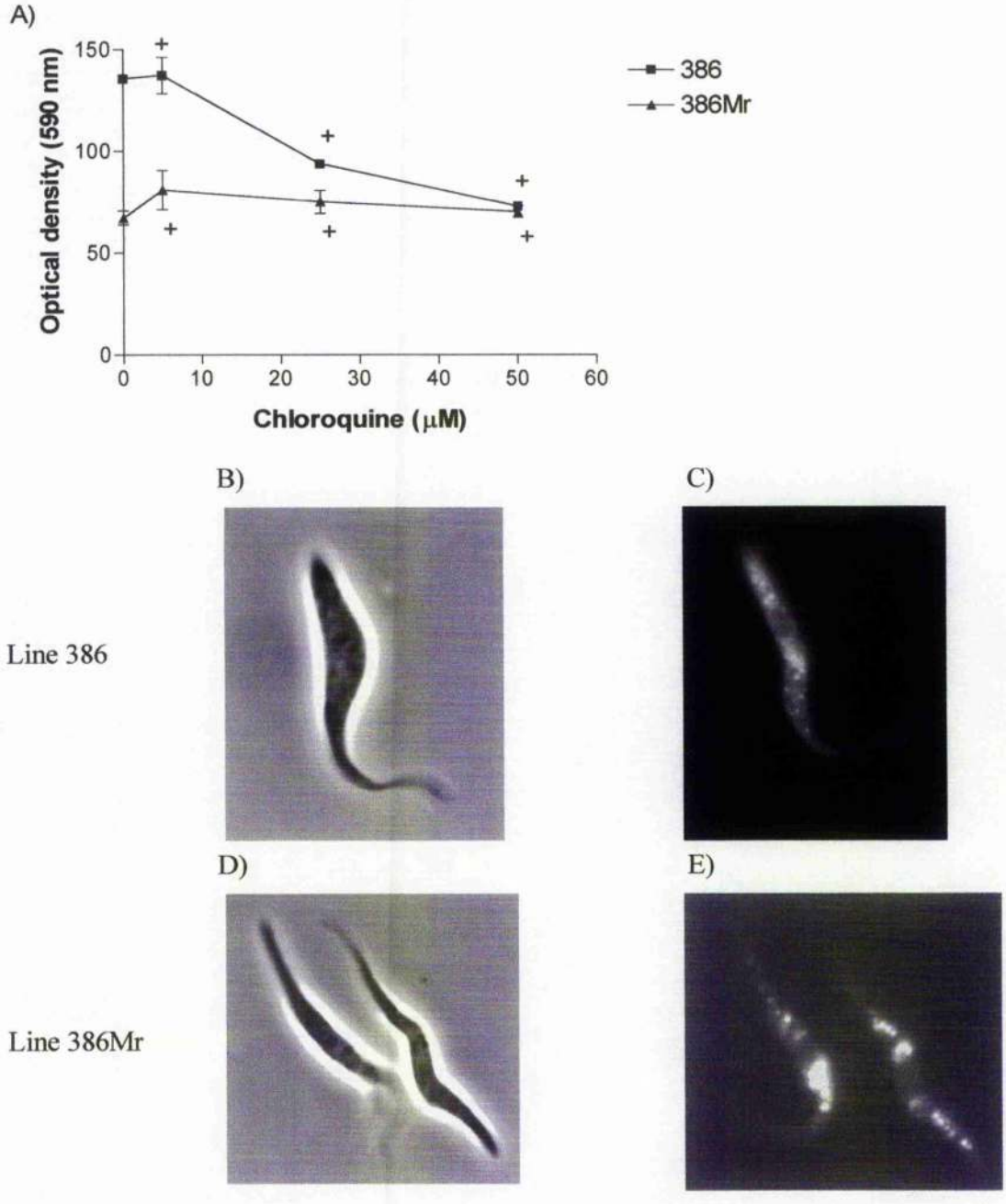


Figure 8.7: Alamar Blue assays of the isogenic 386 lines incubated in the presence of chloroquine (A). The status of the procyclic culture is indicated, for each concentration of drug, as alive (+) or dead (-) as assessed by microscopy. Phase contrast (B and D) and fluorescent (C and E) images of the procyclic lines in the presence of Alamar Blue and 50 μM of chloroquine. The error bars represent the mean +/- SEM, n = 4.

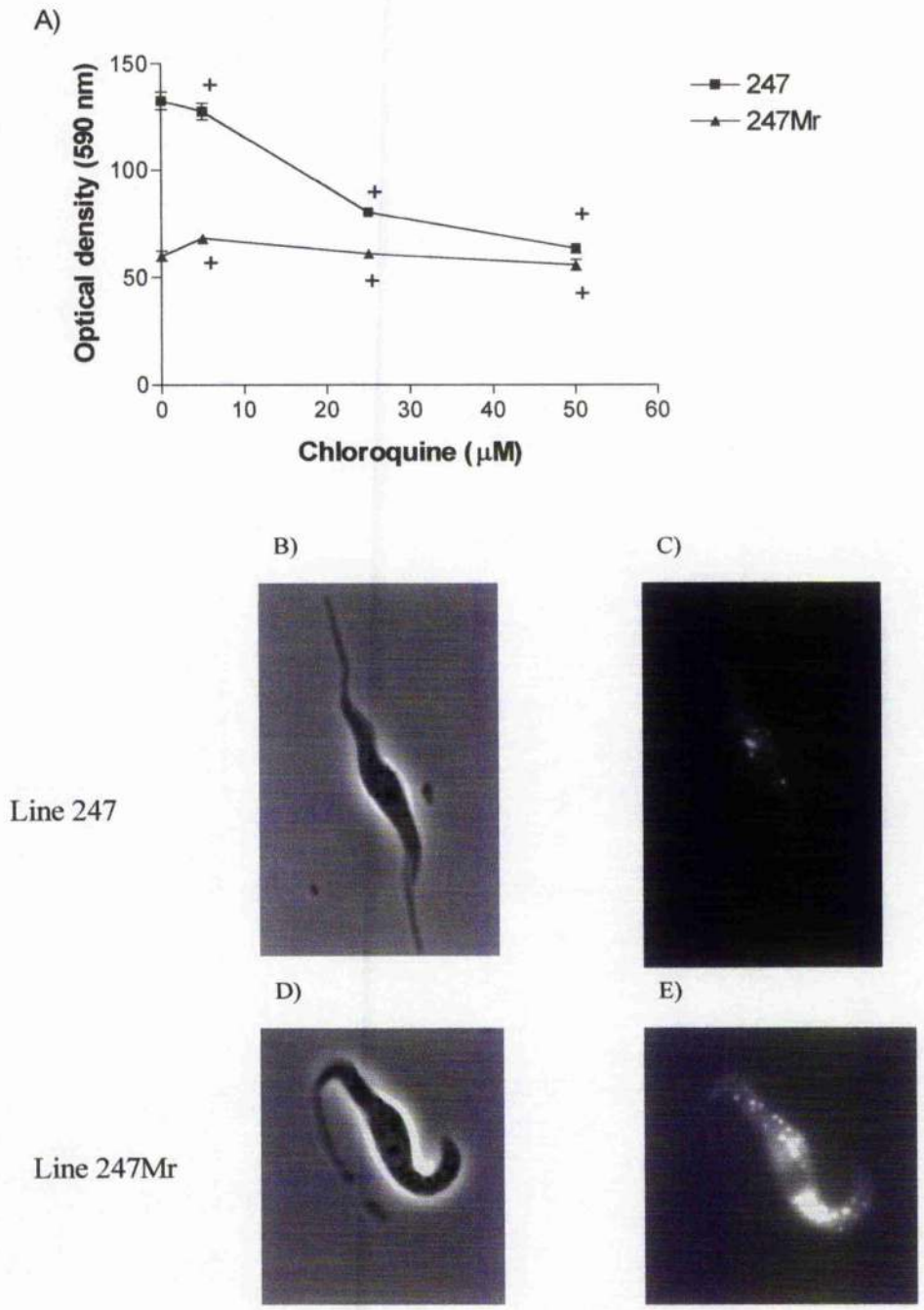


Figure 8.8: **Alamar Blue assays of the isogenic 247 lines incubated in the presence of chloroquine** (A). The status of the procyclic culture is indicated, for each concentration of drug, as alive (+) or dead (-) as assessed by microscopy. Phase contrast (B and D) and fluorescent (C and E) images of the procyclic lines in the presence of Alamar Blue and 50 μM of chloroquine. The error bars represent the mean +/- SEM, n = 4.

proton pumps present on these organelles. This in turn is thought to inhibit the formation of the secondary lysosomes. To assess whether the difference in the level of reduced Alamar Blue dye produced by Cymelarsan sensitive and resistant lines was due to a reduced acidification of endosomes and lysosomes, three known inhibitors of proton pumps were tested for their ability to decrease the level of reduced dye produced by the sensitive lines. Bafilomycin A1 and Concanamycin A are macrolide antibiotics know to specifically inhibit two types of proton pumps: V-ATPase at nM concentration and P-ATPases at mM concentration (Drose and Altendorf, 1997). The third inhibitor, ammonium chloride, is a weak base that has been shown to increase *Plasmodium falciparum* vesicles pH at concentration of 50 mM, in a similar way as chloroquine (Krogstad *et al.*, 1985).

Bafilomycin A1 (up to 3.2 μM), Concanamycin A (up to 29 μM) and ammonium chloride (up to 1 M) were used in the Alamar Blue assay on procyclic forms of each pair of isogenic lines to determine whether the reduced acidification of endosomes and lysosomes in trypanosomes was linked to the level of reduction of the Alamar Blue dye. Microscopic examination of the cultures after 24 h demonstrated that concentrations of 3.2 μM of Bafilomycin A1 and 1 M ammonium chloride were lethal to procyclic trypanosomes (Figure 8.9A/C and 8.10A/C). Sub-lethal concentrations of Bafilomycin A1 and ammonium chloride had no detectable effect on the level of reduced Alamar Blue dye produced by Cymelarsan sensitive and resistant lines. Similarly, Concanamycin A up to 29 μM had no effect on the level of reduced Alamar Blue dye produced by sensitive and resistant lines (Figure 8.9B and 8.10B).

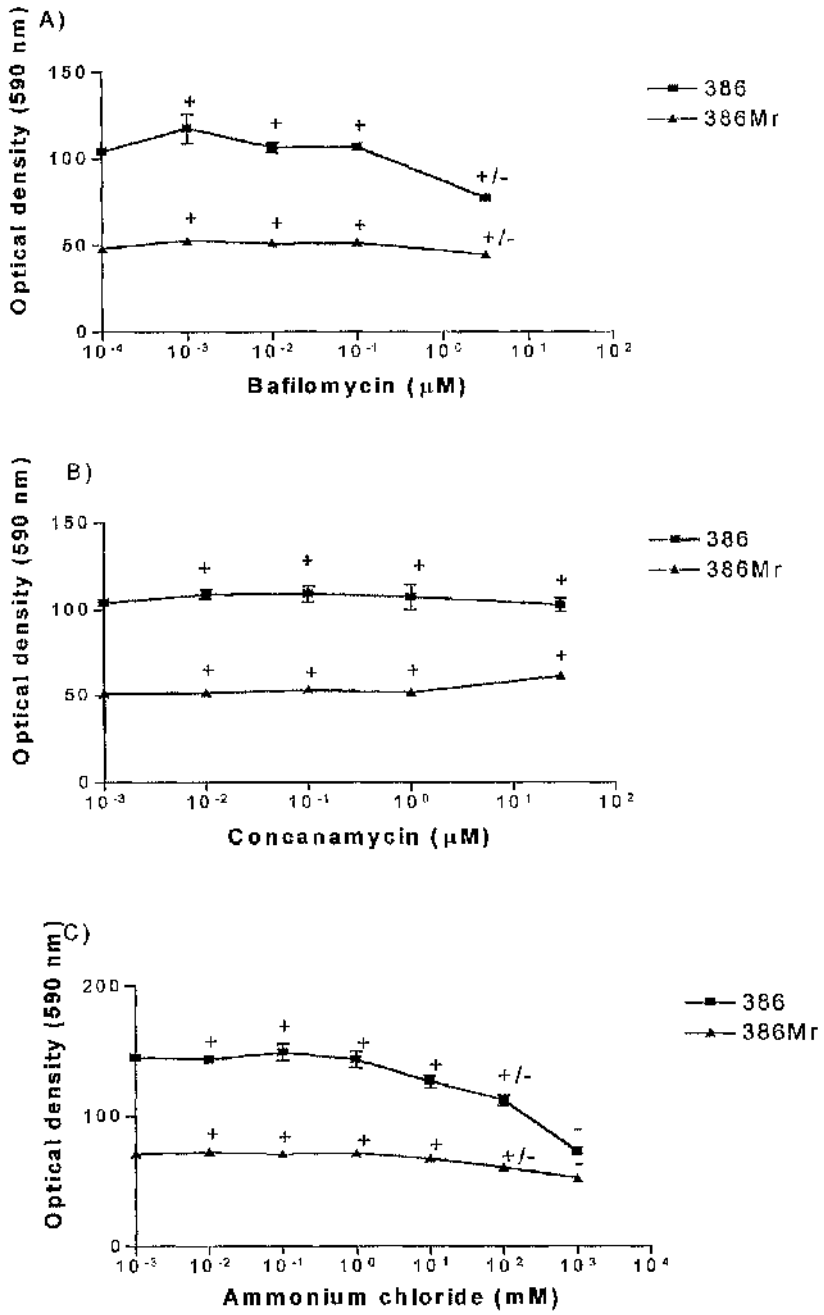


Figure 8.9: Alamar Blue assays of the isogenic 386 lines in the presence of lysosomotropic agents. Three lysosomotropic agents were tested (Bafilomycin A1 (A), Concanamycin A (B) and ammonium chloride (C)) *in vitro* in SDM79. The status of the culture is indicated as alive (+) or dead (-) as assessed by microscopy for each concentration of inhibitors. The lowest concentration corresponds to the control in the absence of drug. The error bars represent the mean +/- SEM, n = 4.

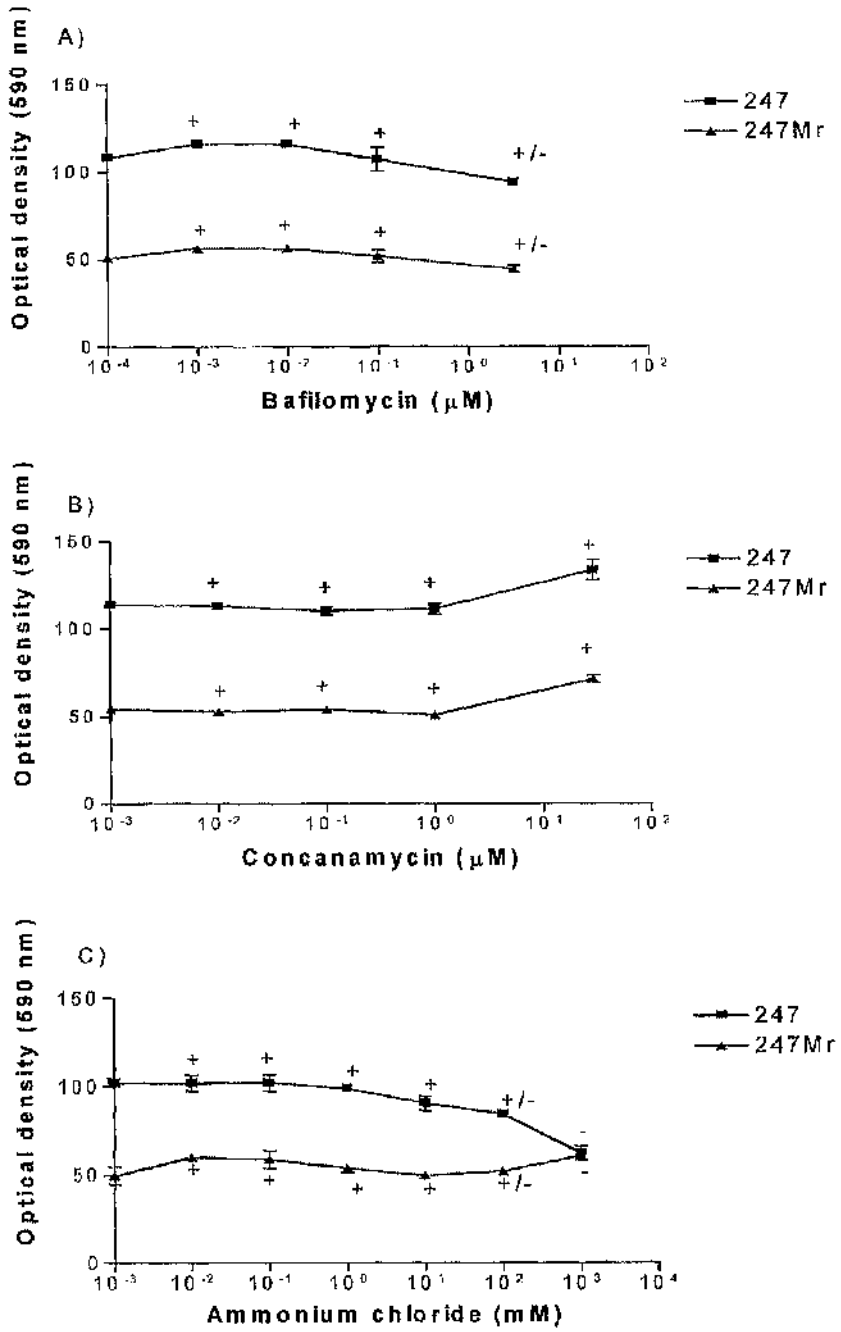


Figure 8.10: Alamar Blue assays of the isogenic 247 lines in the presence of lysosomotropic agents. Three lysosomotropic agents were tested (Bafilomycin A1 (A), Concanamycin A (B) and ammonium chloride (C)) *in vitro* in SDM79. The status of the procyclic culture is indicated as alive (+) or dead (-) as assessed by microscopy for each concentration of inhibitors. The lowest concentration corresponds to the control in the absence of drug. The error bars represent the mean +/- SEM, n = 4.

Taken together the results suggest that the reduction of the Alamar Blue dye is sensitive to the presence of chloroquine in the medium. However, the action of chloroquine on the production of the reduced Alamar Blue dye is not associated to its ability to decrease the acidity level of endosomes and lysosomes. The question arises as to the mode of action of chloroquine on the reduction of the Alamar Blue dye. One possibility is that chloroquine might inhibit the intracellular enzymes responsible for the reduction of the Alamar Blue dye. To test this hypothesis, an *in vitro* Alamar Blue assay was performed on cell lysates from sensitive and resistant lines of 386 and 247 in the presence of chloroquine (Figure 8.11). At a concentration of 50 μM of chloroquine the fluorescence has dropped slightly in both sensitive and resistant lines. However, the drop in fluorescence observed on cell lysates was lower that that observed for live trypanosomes. These results would suggest that the intracellular enzymes responsible for the reduction of the Alamar Blue dye are not inhibited by chloroquine.

Suramin has been reported to bind to serum proteins because it is negatively charged and enter the trypanosome via a LDL receptor mediated endocytosis (Vansterkenburg *et al.*, 1993). The Alamar Blue dye is also negatively charged and could potentially use an equivalent route of entry into trypanosomes. An *in vitro* Alamar Blue assay on procyclic trypanosomes incubated in the presence and the absence of serum however showed no detectable differences in level of fluorescence (Figure 8.12). These results suggest that the entry of the Alamar Blue dye into trypanosomes does not require binding to serum components.

8.3.3.4 The effect of Cymelarsan on the localisation of the reduced Alamar Blue dye in trypanosomes

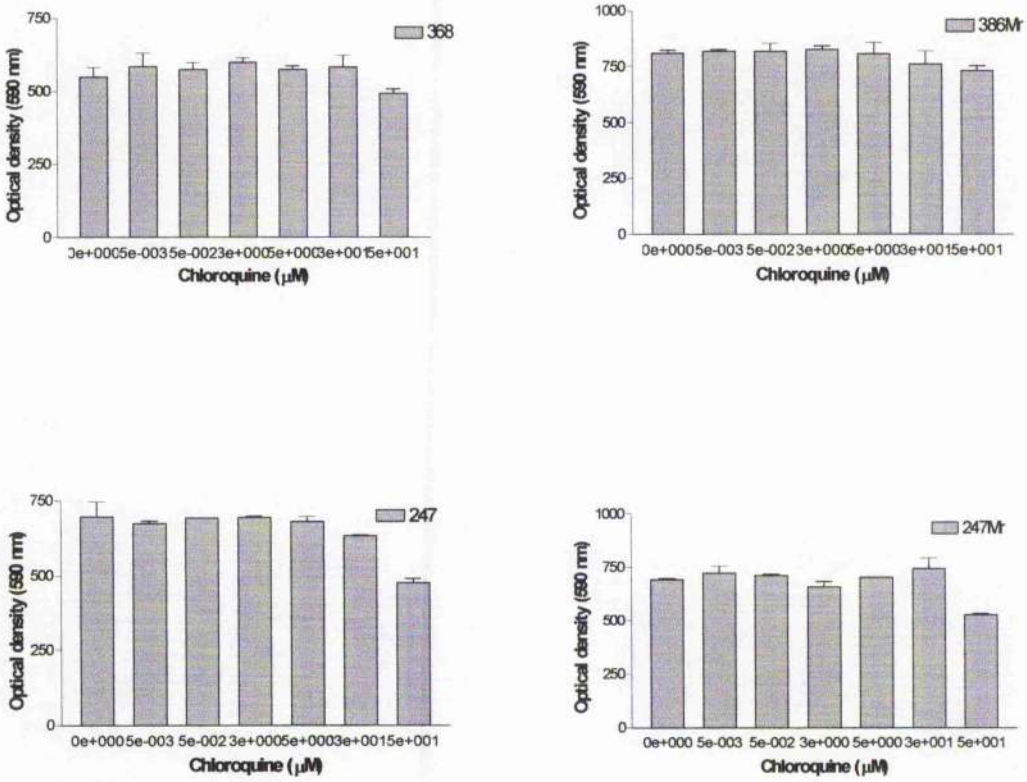


Figure 8.11: Alamar Blue assays on cell lysates in the presence of chloroquine. Chloroquine was added to cell lysates from each pair of isogenic lines to assess its effect on the intracellular enzymes responsible for the reduction of the Alamar Blue dye. The error bars represent the mean +/- SEM, n = 2.

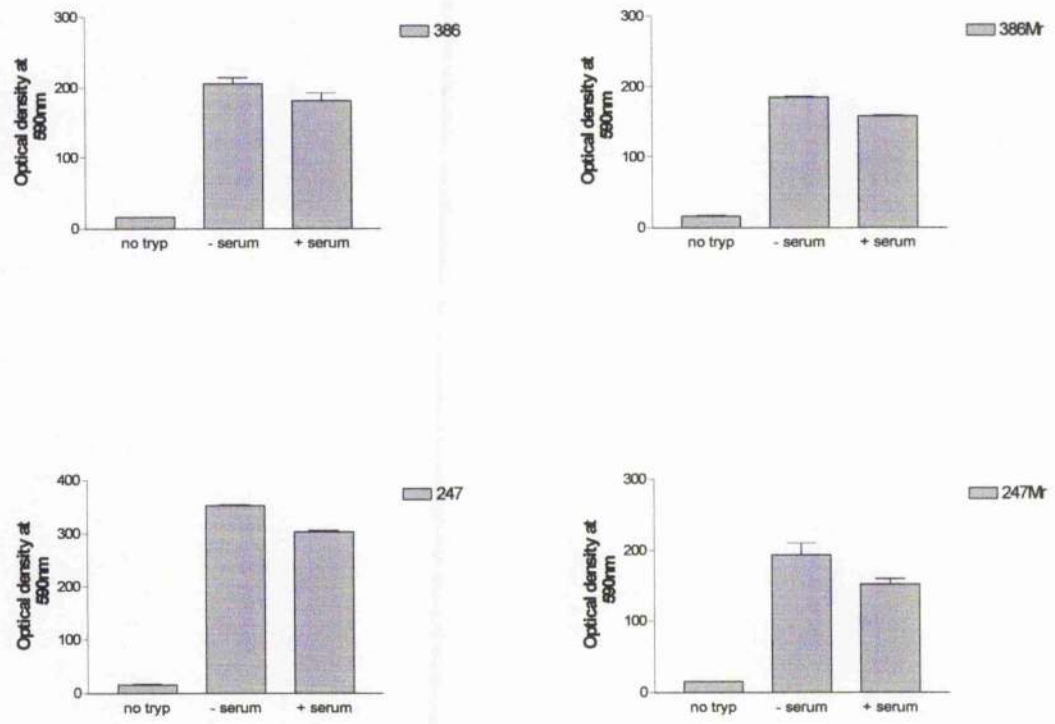


Figure 8.12: **Uptake of the Alamar Blue dye in the absence of serum**. Procyclic trypanosomes were incubated in Alamar Blue in the presence and absence of serum to test whether uptake occurred via binding to serum components. The error bars represent the mean +/- SEM, n = 2.

To assess whether Cymelarsan had an effect of the reduction and localisation of the Alamar Blue dye in the trypanosome cells, each pair of isogenic lines was incubated in the presence of Alamar Blue and observed by fluorescence microscopy using a DAPI filter prior to the addition of Cymelarsan. After addition of Cymelarsan the fluorescence from the vesicle-like structure disappears rapidly while the trypanosomes were still morphologically intact (Table 8.3). This phenomenon was observed for both sensitive and resistant lines, and suggests that the Cymelarsan might affect the membrane of the vesicle-like structures in a way that enables the fluorescent Alamar Blue dye to leak out.

8.4 Discussion

The Alamar Blue assay was developed by Räz *et al.* (1997) to assess the sensitivity of African trypanosomes to various drugs *in vitro*. The trypanocidal activity of drugs such as Melarsoprol, DFMO, Pentamidine and Suramin against *T. b. gambiense* (STIB 930) and *T. b. rhodesiense* (STIB 900) was shown to be readily assayed (Räz *et al.*, 1997). In the results presented here, assaying Cymelarsan sensitive and resistant lines showed an association of Cymelarsan resistance with a decreased production of reduced Alamar Blue dye. Two mechanisms by which the Cymelarsan resistant lines could produce less reduced Alamar Blue were assessed: a) the ability of cellular enzymes to reduce the Alamar Blue dye and b) the mode of transport/efflux of the Alamar Blue dye.

Line	Drug		Fluorescence microscopy	Phase contrast microscopy
386	Alamar Blue	_	Vesicles like structure	Intact
	Alamar Blue	Cymelarsan	None	Intact
386Mr	Alamar Blue	_	Vesicles like structure	Intact
	Alamar Blue	Cymelarsan	None	Intact
247	Alamar Blue	_	Vesicles like structure	Intact
	Alamar Blue	Cymelarsan	None	Intact
247Mr	Alamar Blue	_	Vesicles like structure	Intact
	Alamar Blue	Cymelarsan	None	Intact

Table 8.3: Fluorescence microscopy of each pair isogenic lines in the presence of Alamar Blue and Cymelarsan. The effect of Cymelarsan of the distribution of the reduced Alamar Blue dye inside the trypanosomes was assessed by fluorescence microscopy. The experiment was performed only once for each pair of isogenic lines.

The Alamar Blue assay is based on the ability of intra-cellular enzymes to reduce the non-fluorescent blue dye to a pink fluorescent compound. Cell lysate from sensitive and Cymelarsan resistant lines reduced the Alamar Blue dye *in vitro*. Many enzymes have been reported to be able to reduce the Alamar Blue dye *in vitro*. The list includes mitochondrial enzymes such as flavin mononucleotide dehydrogenase, flavin adenine dinucleotide dehydrogenase, nicotinamide adenine dinucleotide dehydrogenase, nicotinamide adenine phosphate dehydrogenase, dihydrolipoamine dehydrogenase, NAD(P)H:quinone oxidoreductase and cytochromes (Gonzalez and Tarloff, 2001). It would seem unlikely that all these enzymes could be altered in the Cymelarsan resistant line, and this is confirmed by the experiment carried out on the cell lysates.

The Alamar Blue assay is based on the fact that the dye can only be reduced by intra-cellular enzymes and therefore requires the uptake of the dye in order for it to be reduced to the fluorescent product, and then excreted into the medium (O'Brien *et al.*, 2000). The route of entry of Alamar Blue into the trypanosome is unknown. Uptake by the parasite can be the result of passive diffusion, receptor-mediated uptake, transporter-mediated uptake or endocytosis. The chemical structure of the oxidised Alamar Blue dye (Figure 8.1) would indicate that the molecule is unable to diffuse inside the cell due to its charged state so it is likely that a mode of transport is required. There are two possible modes of transport by which a compound can enter a trypanosome: through a transporter or via the endocytic pathway.

Comparison of the structure of the oxidised Alamar Blue dye with various trypanocial drugs revealed a structural similarity with Isometamidium. This structural similarity suggested that

the Isometamidium transporter was a possible transporter for the dye. Transport of Isometamidium occurred in both sensitive and Cymelarsan resistant lines and no increase in Isometamidium resistance was observed in the Cymelarsan resistant lines suggesting that the transport of Isometamidium was similar between the lines. This suggested that the Alamar Blue phenotype was not linked to Isometamidium uptake.

The identification of vesicle-like structures in which the reduced Alamar Blue dye accumulates inside the trypanosomes and the partial inhibition of the reduction of the dye in sensitive lines by 50 μM of chloroquine, suggested uptake via the endocytic pathway. Lysosomotropic agents are compounds which reduce the acidity of organelles in the endocytic/lysosomal pathways. Chloroquine is a lysosomotropic agent thought to act on the endocytic pathway (Figure 8.13) by increasing the pH of lysosomes preventing the formation of secondary lysosomes (Malik *et al.*, 2001; Blanchard *et al.*, 2000). The use of three other lysosomotropic agents (Bafilomycin A1, Concanamycin A and ammonium chloride) had no effect on the production of reduced Alamar Blue dye by Cymelarsan sensitive and resistant lines. This would suggest that acidic pH in lysosomes is not essential for the production of reduced Alamar Blue dye.

Studies on the internalisation by endocytosis of HDL and anti-CRAM IgG revealed that exposure of procyclic forms of trypanosomes to 50 μM of chloroquine completely abolished the lysosomal degradation of these molecules (Lee *et al.*, 1999; Liu *et al.*, 2000). The trypanocidal drug such as suramin can enter via the endocytic pathway bound to LDL (Coppens *et al.*, 1988; Vansterkenburgh *et al.*, 1993). The oxidised Alamar Blue dye is

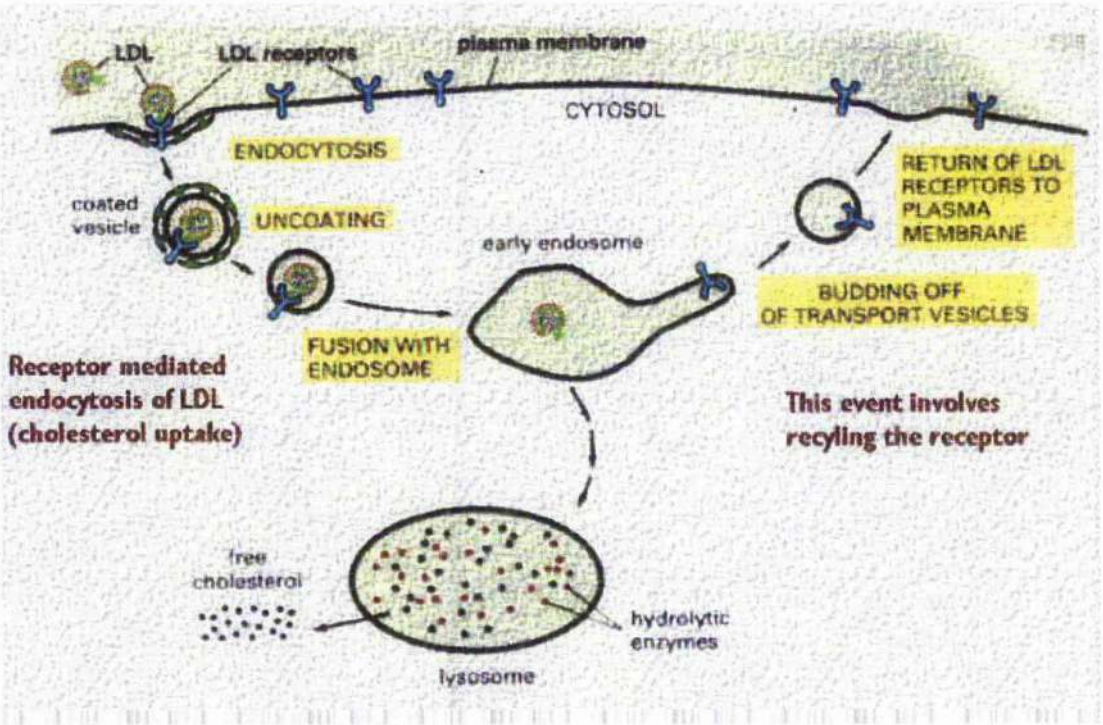


Figure 8.13: **Model of an endocytic uptake of LDL**. (Reproduced from http://cellbio.utmb.edu/cellbio/recend2.htm)

charged and could bind to HDL or LDL in the medium before internalisation via endocytosis. A preliminary experiment showed that Alamar Blue dye was still able to enter the trypanosome and be reduced in the absence of serum in the medium. This would suggest that the Alamar Blue dye enters the trypanosome via a different route. The elimination of the endocytic pathway as the route of entry of the Alamar Blue dye in the cell is further supported by the fact that procyclic trypanosomes show low levels of endocytosis (Langreth and Balber, 1975) and so it is difficult to account for the rapid appearance of the fluorescent vesicular-like organelles through this pathway. In addition, the vesicular-like structures accumulating the reduced Alamar Blue dye were located throughout the trypanosome cell, while endosomes and lysosomes are only observed around the flagellar pocket region (Langreth and Balber, 1975). Consequently the action of chloroquine on the reduction of the Alamar Blue dye appears to be unrelated to its ability to decrease the acidity of lysosomes. In addition, 50 μM of chloroquine had no inhibitory action of the cell lysate and so the mode of action of chloroquine on the production of the reduced Alamar Blue dye in the trypanosome has not been determined so far.

The vesicular-like structures accumulating the reduced Alamar Blue dye appeared to correspond to dense granules that are observed under phase contrast microscopy (Figure 8.14). The location of these vesicular-like structures and their appearance are characteristic of acidocalcisomes (Docampo and Moreno, 2001). In addition acidocalcisomes have been shown to swell in the presence of chloroquine (Coppens *et al.*, 1993) which would explain the larger fluorescent vesicles observed when trypanosomes are incubated in the presence of the Alamar Blue dye and chloroquine. Acidocalcisomes are acidic organelles described as cytoplasmic electron dense granules containing calcium, magnesium, sodium, zinc, iron and pyrophosphate

(Figure 8.15; Miranda *et al.*, 2000; Scott and Docampo, 1998; Rodrigues *et al.*, 1999). The low content of S found in these organelles would suggest that they have a low protein content. The function of these organelles is unknown although Docampo and Moreno recently (1999) have suggested four possible functions based on the information available: a) a role in the process of calcium storage, b) a role as an energy storage organelle containing large amount of pyrophosphate, c) a possible role in the regulation of cytoplasmic pH through the H^+ATPases present in the membrane and finally d) a role in the control of the process of osmoregulation. Analysis of the association between the lowered level of Alamar Blue reduction and Cymelarsan resistance in procyclic form of *T. brucei* has lead to the identification of vesicle-like structures that may be acidocalcisomes. The rapid disappearance of the fluorescence from the vesicles was observed after the addition of Cymelarsan to the medium. However the integrity of the dense granules was not influenced by the addition of Cymelarsan. The diverse functions suggested for acidocalcisomes clearly support the view that these are important organelles for the trypanosome. One hypothesis based on the results presented in this chapter, is that Cymelarsan could act on the membrane of the acidocalcisomes, rendering them leaky. This would explain the disappearance of the reduced Alamar Blue dye from the vesicles when trypanosomes are incubated in Cymelarsan, as well as suggesting a release of the acidocalcisomes' contents into the cytoplasm. The release of the contents of the acidocalcisome is likely to lead to a decrease cytosolic pH with a consequent inhibition of metabolic pathways. To test this hypothesis it would first be necessary to prove that the vesicular-like structures accumulating the Alamar Blue dye are acidocalcisomes and this could be undertaken by using acidocalcisome specific antibodies (Lemercier *et al.*, 2002) and Alamar Blue, in a co-localisation study.

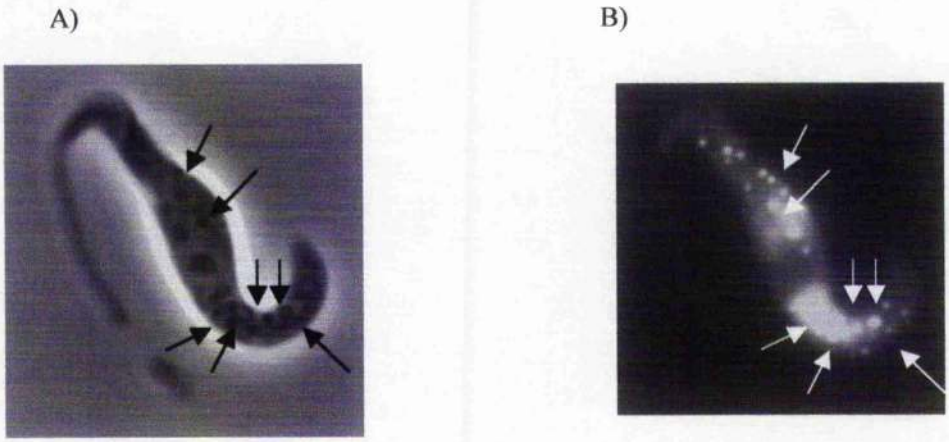


Figure 8.14: Localisation of the Alamar Blue dye in procyclic trypanosomes. Apparent co-localisation of the dense granules under phase contrast (A) and the vesicular-like organelles (B) accumulating the reduced Alamar Blue dye.

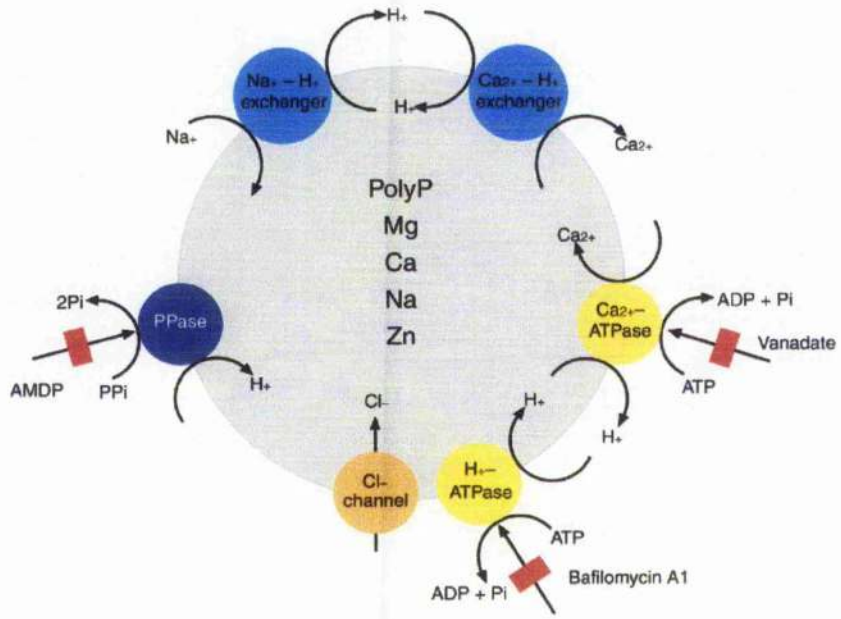


Figure 8.15: Model of an acidocalcisome in trypanosomes. Calcium ions uptake occurs in exchange for H^+ by a reaction catalysed by a vacuolar Ca^{2+}-ATPase which is inhibited by vanadate. A H^+ gradient is established by a Bafilomycin A1-sensitive vacuolar H^+-ATPase and an aminophenylethylenephosphonate (AMDP)-sensitive vacuolar H^+-Ppase. Chloride (Cl^-) transport, through a chloride channel is associated with the function of the vacuolar H^+-ATPase. Ca^{2+} release occurs in exchange for H^+ and is favored by sodium-proton exchange. The acidocalcisome is rich in pyrophosphate, magnesium, calcium, sodium and zinc. (Docampo and Morena, 2001).

To properly assess the function and involvement of the acidocalcisome in Cymelarsan resistance in *T. brucei* it would be necessary to study the membranes of acidocalcisomes. This would require the isolation of a sub-cellular fraction containing well-preserved acidocalcisomes and no other cell membranes. A method has recently been developed for the preparation of acidocalcisomes (Scott and Docampo, 2000). Measurements of the pH of a solution containing acidocalcisomes in the presence and absence of Cymelarsan would determine the effect of Cymelarsan on the membranes of acidocalcisomes.

Chapter IX
General discussion

The aim of this project was to investigate mechanisms of arsenical resistance in the protozoa parasite *T. brucei*. This aim was pursued by subtractive analysis of the whole protome from drug sensitive and resistant isogenic lines. It was anticipated that differentially expressed proteins between the isogenic lines would be associated with Cymelarsan resistance and could determine the mechanisms of arsenical resistance in *T. brucei*.

Most drug resistance studies performed so far on *T. brucei* have relied on biochemical, genetic or functional analysis of a candidate gene. The drawback of these approaches is the need for the pre-selection of a gene to analyse. The proteomic approach employed in this study enabled the global proteomic expression profiling of a drug sensitive *T. brucei* line and its laboratory-derived isogenic drug resistant line. The advantage of this approach is that all the possible mechanisms of arsenical resistance can potentially be analysed at once.

A recently published protocol for the solubilisation of proteins from *T. brucei* (van Deursen *et al.*, 2003) was developed for the study of Cymelarsan resistance in procyclic forms of the parasite. The protocol was simple, reproducible and extracted membrane proteins. The limitations of protein identification due to the limited *T. brucei* genome database were assessed with the production of a preliminary proteome map. Forty-two percent of the protein

spots separated by 2DE and analysed by PMF were identified. Only 30% of the protein spot not identified was due to the incomplete *T. brucei* genome database.

Two pairs of isogenic lines were used in this study. Preliminary characterisation demonstrated that the resistance phenotype was stable after passage in mice and cyclic transmission through tsetse flies (Scott *et al.*, 1996). In addition, both bloodstream and procyclic forms of the parasites were shown to express the resistance phenotype and cross-resistance in bloodstream forms between Melarsoprol and Melarsen oxide had been shown. Procyclic forms of the parasites were further characterised in this study. No cross-resistance was detectable to either Melarsoprol, Pentamidine or Berenil for both pairs of isogenic lines. This was consistent with the fact that no P2 transporter activity can be detected in procyclic forms (de Koning *et al.*, 2000) which rules out the involvement of the P2 transporter as the main mechanism of Cymelarsan resistance in procyclic forms of these lines. Another known mechanism of drug resistance in trypanosomes and other parasites involves the excretion of drugs via an over-expressed P-glycoprotein (Shahi *et al.*, 2002). This mechanism was assessed on both pairs of isogenic lines by growth inhibition assay in presence of general P-glycoprotein inhibitors. The presence of the inhibitors had no effect on the resistance to Cymelarsan. The P-glycoprotein inhibitor, verapamil, has been shown to reverse drug resistance in *T. cruzi* (Neal *et al.*, 1989) and *T. congolense* (Sutherland and Holmes, 1993), and consequently was expected to inhibit the P-glycoprotein from *T. brucei*. The other two P-glycoprotein inhibitors, TFP and PCP, have never been shown to be effective against P-glycoprotein from trypanosomes. The two obvious mechanism of drug resistance, the alteration of transport by the P2 transporter and the increased efflux by a P-glycoprotein, were tested and found to be not associated with arsenical

resistance in procyclic forms of the 386 and 247 lines. To identify protein associated with arsenical resistance a subtractive analysis of whole cell extract from sensitive and resistant lines was performed.

The subtractive analysis of the global expression profile of 386 and 386Mr lines enabled the detection of two protein spots whose lack of expression was associated with Cymelarsan resistance in the pair of isogenic lines of 386.

The differential expression of the CAR protein (spot 2) was reproducible over nine independent protein extractions separated on pH 4-7 gels and on one protein extraction separated on pH 4-5 gel. The CAR protein is adjacent on the 2D gel to a protein spot (spot 1) with similar molecular weight but more basic pI. Both the CAR protein and spot 1 from the proteins extracted from 386 and 386Mr were identified as products of either ORF 274 or 277. The presence of an extra form of the CAR protein in the 386 line could be explained by a mutation in either one of the ORFs, an absence of transcription of one of the two ORFs in the 386Mr line, or a post-translational modification of spot 1 which occurs only in line 386.

Analysis of the sequences obtained for the two genes from both 386 and 386Mr lines showed that ORF 274 and 277 from the 386 line were identical to their homologues in the 386Mr line. In addition, the predicted sequence of the transcript of ORF 274 was identical to the predicted sequence of the transcript of ORF 277. PCR on total cDNA from both lines demonstrated that both ORFs were transcribed in each line. These results suggested that the CAR protein was a post-translationally modified spot 1. The simplest way of regulating the activity of a protein is

by changing the number of protein molecules in the cell. This is generally achieved by regulating gene transcription, mRNA translation, or the rate of mRNA and protein turnover. A more rapid increase in protein activity can be achieved by post-translational modification. For example, activation of a kinase by protein phosphorylation is a hallmark of intracellular signalling cascades (Hunter, 1995). Post-translational regulation of the CAR protein could allow large amounts of the protein to be present in the cell in an "inactive" form.

The identity of spot 2 was confirmed by Western blot analysis using an antibody raised by DNA vaccination strategy against the expressed product of ORF 274. The number of isoforms present in each line was detected by Western blot analysis of 2D gels. In total, three isoforms were detected at pH 4.2, 34 kDa in the extract from the 386 line and only two isoforms at this location in the extract of 386Mr. The association of the absence of the CAR protein with Cymelarsan resistance was confirmed in independent extractions, and the identity of the protein was confirmed by mass spectrometry and Western blot analysis.

From these results the question arise as to whether the CAR protein is associated with Cymelarsan resistance in other *T. brucei* lines. Analysis of the pair of isogenic 247 lines showed that the CRA protein was not associated with Cymelarsan resistance in these lines. In addition, the analysis showed that only one isoform of the CAR protein could be detected in both 247 and 247Mr lines. Currently the antibody raised against the transcript of ORF 274 is being used for immuno-localisation of the CAR protein in *T. brucei* to extend our knowledge of its potential function.

Analysis of the amino acid sequence of the CAR protein showed the presence of a NAC domain and a UBA domain. The function of this protein in *T. brucei* is unknown. From the presence of the NAC and UBA domains, it seems highly likely that the protein is part of a multi-protein complex. Many cellular functions are performed not by individual proteins but by large multi-protein complexes. The analysis of complexes formed with the CAR protein could help on understanding of the function of the protein with respect to Cymelarsan resistance in *T. brucei*. This experiment could be performed by immuno-precipitation using the antibodies raised against the expressed product of ORF 274. Biochemical purification of complexes followed by mass spectrometry has enabled the identification of multi-protein complexes in other contexts (Neubauer *et al.*, 1997).

The second protein spot associated with cymelarsan resistance in the pair of isogenic lines of 386 was identified by mass spectrometry as an arginine kinase. Using the mass spectrometry analysis it was not possible to determine whether the protein spot was the transcript of ORF 243 or 242. However, the absence of a protein product from either one of the two ORFs (or both) at pH 5.8 on 2D gels was associated with Cymelarsan resistance.

The comparative proteomic analysis of Cymelarsan sensitive and resistant lines of 386 by 2DE enabled the direct comparison of expression of 1,000 proteins spots separated on 2D gels. It is estimated that the trypanosome genome contains approximately 8,000 genes based on sequence from chromosome I and II (Hall *et al.*, 2003; El-Sayed *et al.*, 2003) and while all the genes will not be expressed at the procyclic stages of culture, it is expected that many proteins were not separated or detected in this study. Consequently, the proteomic analysis performed

on the sensitive and resistant lines was biased towards highly abundant and soluble proteins. The limitation of spatial resolution, which causes spots to overlap on the 2D gel, was first reported in 1982 for albumin, for which what was considered a pure protein spot was further separated into additional components by using IPG strips of 0.01 pH/cm (Bjellqvist *et al.*, 1982). The use of narrow pH range had been shown to increase the separation and resolution of protein spots (Cohen *et al.*, 2002). The problem of spatial resolution on 2D gels was addressed in this study by using narrow pH range strips. The use of narrow pH IPG strips has enabled the detection of another differentially expressed protein spot between sensitive and resistant line of 386. However, even using narrow pH range strips, only the most abundant proteins were detected. In order to increase the proteomic analysis of sensitive and resistant 386 lines, the sample would have had to be pre-fractionated.

Only a few comparative proteomic analysis of disease/resistant cells/organisms with control cells/organisms have been published to date. Comparative proteomic analysis in these studies revealed only a small number of proteins differentially expressed between the two cells/organisms. There seems to be a general trend towards the identification of a higher number of differentially expressed proteins when using a quantitative proteomic approach. For example, the comparative analysis of p-53 deficient mice (a tumour suppressor gene) with control mice showed only 7 protein spots, out of 886 analysed by 2DE, were specific to the p53 deficient mice (Anaki *et al.*, 2000). Whereas, the comparative and quantitative proteomic study of a cancer cell line (T47D) and its antiestrogen-resistant derivative showed 38 protein spots differentially expressed (up- or down-regulated by a factor of more than 2) out of 2,500 spots analysed by 2DE (Huber *et al.*, 2003). Similarly, a quantitative and comparative

proteomic study of mouse-virulent and avirulent *Toxoplasma gondii* showed 35 protein spots, out of 300 spots analysed by 2DE, were up- or down-regulated between the two lines (Nischik *et al.*, 2001).

The resistance to Cymelarsan in this line was induced by prolonged exposure to sublethal doses of Cymelarsan, a treatment which is likely to be highly mutagenic and may induce changes in addition to drug resistance. Consequently, the results presented here will need to be confirmed using other methods. It will be difficult to confirm the association between the absence the CAR protein (isoform of spot 1) with arsenical resistance as most technologies available will remove all forms of the protein (knockout, RNAi) or examine transcript levels (micro-arrays) which are probably not involved.

The complete proteome analysis is limited by the technology. It is possible that other differentially expressed proteins between the sensitive and Cymelarsan resistant lines were not identified. Fluorescent two dimensional difference gel electrophoresis (DIGE) allows samples labeled with different dyes to be co-separated on the same gel, thereby eliminating the gel-to-gel variation (Beaumont *et al.*, 2001). This technique will be very useful in quantitative and comparative proteomic studies.

Synteny in coding regions and conservation of strand-switch regions are observed between several trypanosomatid parasites, suggesting an evolutionary pressure to conserve the close association of certain genes within clusters. This organisation suggests that clusters encode proteins involved in related cellular processes, requiring co-regulation or similar expression

level. ORF 274 and 277 are present in a cluster containing mainly putative ORFs with homology to enzymes and chaperonine. Of interest is the presence of the Ubiquitin conjugating E2 enzyme, containing a UBA domain and involved in the degradation of proteins by the ubiquitination pathway (Madura, 2002; Hoffman and Bucher, 1996). The CAR protein also contained a UBA domain thought to enable it to bind to ubiquitin and other proteins associated with the ubiquitination pathway (Wilkinson *et al.*, 2001). The presence of the E2 enzyme gene and ORF 274 and 277 in this cluster on chromosome IX, all containing UBA domain, would suggest the involvement of the expressed products of ORF 274 and 277 in the ubiquitination pathway. The mechanism of resistance to arsenical drugs could thus be linked perhaps to a higher turnover of proteins in the cell. Arsenical drugs are known to have a high affinity for intracellular thiols such as trypanothione and dihydropoamide plus several enzymes including trypanothione reductase (Fairlamb *et al.*, 1989), 6-phosphogluconate dehydrogenase (Hanau *et al.*, 1996) and glycerol-3 phosphate dehydrogenase (Denise *et al.*, 1999), all of which are essential for trypanosome survival. A higher turnover of these proteins, which might be the cellular targets of arsenicals, would enable the parasite to survive in the presence of the drug.

These proteomic analyses of sensitive and resistant line of 386 have demonstrated that the resistance phenotype was associated with the absence of an isoform of the CAR protein. Consequently, the mechanism of arsenical drug resistance in these lines might be due to the absence of a post-translationmal modification pathway.

The biochemical analysis of the two pairs of isogenic lines showed the association of the Cymelarsan resistance phenotype with a decrease fluorescence produced from the metabolism of the Alamar Blue dye. Further analysis of the metabolism of the Alamar blue dye demonstrated that the activity of the reductive enzymes were similar between sensitive and resistant lines. The reduced Alamar Blue dye was shown to concentrate into vesicle-like structure thought to be acidocalcisomes. The addition of Cymelarsan to the lines incubated on the presence of Alamar Blue demonstrated the disappearance of the fluorescence from the vesicle-like structure, suggesting that the drug might act on the membrane of the acidocalcisomes. Acidocalcisomes are acidic compartments containing pyrophosphate, magnesium, calcium, sodium and zinc (Docampo and Morena, 2001). An alteration of their membranes could release the high proton content into the cytoplasm, which would affect many metabolic pathways. A protocol has recently been developed for the extraction of pure acidocalcisomes (Scott and Docampo, 2000). These could be used to analyse the effect of Cymelarsan on the membrane of acidocalcisome, by measuring the pH of a solution prior and after the addition of Cymelarsan.

Proteomic methods are "hypothesis generating". Whilst the significance of the CAR protein and arginine kinase in Cymelarsan resistance in *T. brucei* line 386 remains to be established, this study has demonstrated the potential of proteomics to provide unexpected insights into molecular mechanisms by virtue of its ability to study protein expression in its entirety without pre-supposing a specific mechanism. It appears that the differential expression of the CAR protein represents a change in the level of post-translational modification of an existing gene

product which could not have been identified by the application of purely nucleic-acid based technologies.

Reference List

Achbarou, A., Mercereau-Puijalon, O., Autheman, J. M., Fortier, B., Camus, D., and Dubremetz, J. F. (1991). Characterization of microneme proteins of *Toxoplasma gondii*. *Mol.Biochem.Parasitol.* **47**, 223-233.

Adessi, C., Miege, C., Albrieux, C. and Rabilloud, T. (1997). Two dimensional gel electrophoresis of membrane proteins: a current challenge for immobilised pH gradients. *Electrophoresis*, **18**, 127-135.

Aebersold, R. and Goodlett, D. R. (2001). Mass spectrometry in proteomics. *Chem.Rev.* **101**, 269-295.

Afewerk, Y., Clausen, P.-H., Abede, G., Tilahun, G. and Mehlitz, D. (2000). Multiple resistant *Trypanosoma congolense* populations in village cattle of Metekel district, north-west Ethiopia. *Acta Tropica*, **76**, 231-238.

Aksoy, S., Maudlin, I., Dale, C., Robinson, A. S. and O'Niell, S. L. (2001). Prospects for control of African trypanosomiasis by tsetse vector manipulation. *Trends Parasitol.* **17**, 29-35.

Alonso, G. D., Pereira, C. A., Remedi, M. S., Paveto, M. C., Cochella, L., Ivaldi, M. S., Gerez de Burgos, N. M., Torres, H. N., and Flawia, M. M. (2001). Arginine kinase of the flagellate protozoa *Trypanosoma cruzi*. Regulation of its expression and catalytic activity. *FEBS Lett.* **498**, 22-25.

Ames, G. F. and Nikaido, K. (1976). Two dimensional gel electrophoresis of membrane proteins. *Biochem.*, **15**, 616-623.

Anderson, L. and Seilhamer, J. (1997). A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533-537.

Anderson, N. L., Parish, N. M., Richardson, J. P., and Pearson, T. W. (1985). Comparison of African trypanosomes of different antigenic phenotypes, subspecies and life cycle stages by two-dimensional gel electrophoresis. *Mol.Biochem.Parasitol.* **16**, 299-314.

Ancne, B. M., Onah, D., N., and Nawa, Y. (2001). Drug resistance in pathogenic African trypanosomes: what hopes for the future? *Veterinary Parasitol.* **96**, 83-100.

Barrett, J., Jefferies, J. R., and Brophy, P. M. (2000). Parasite proteomics. *Parasitol.Today* **16**, 400-403.

Barrett, M. P. (1999). The fall and rise of sleeping sickness. *Lancet* **353**, 1113-1114.

Basselin, M., Denise, H., Coombs, G. H. and Barrett, M. P. (2002). Resistance to pentamidine in *Leishmania mexicana* involves exclusion of the drug from the mitochondrion. *Antimicriob.Agents Chemother.*, **46**, 3731-3738.

Benaim, G., Lopez-Estano, C., Docampo, R., and Moreno, S. N. J. (1993). A calmodulin-stimulated Ca^{2+}-pump in plasma membrane vesicles from *Trypanosoma brucei*. Selective inhibition by pentamidine. *Biochem.J.*, **296**, 759-763.

Berger, B. J., Carter, N. S., and Fairlamb, A. H. (1995). Characterisation of pentamidine-resistant *Trypanosoma brucei brucei*. *Mol.Biochem.Parasitol.* **69**, 289-298.

Berger, B. J. and Fairlamb, A. H. (1994). Properties of melarsamine hydrochloride (Cymelarsan) in aqueous solution. *Antimicrob.Agents Chemother.* **38**, 1298-1302.

Bitonti, A.J., Bacchi, C.J., McCann, P.P. and Sjoerdsma, A. (1985). Catalytic irreversible inhibition of *Trypanosoma brucei brucei* ornithine decarboxylase by substrate and product analogs and their effects on murine trypanosomiasis. *Biochem.Pharmacol.* **34**, 1773-1777.

Bitonti, A.J., Bacchi, C.J., McCann, P.P and Sjoerdsma, A. (1986). Uptake of a-difluoromethylornithine by *Trypanosoma brucei brucei*. *Biochem.Pharmacol.* **35**, 351-154.

Bjellqvista, B., Eka, C., Righettib, P. G., Gianazzab, E., Gorg, A., Westermeirec R. and Postelc, W. (1982). Isoelectricfocussing in immobilised pH gradients: principle, methodology and some applications. *J.Bioch.Biophysic.Methods* **6**, 317-339.

Blanchard, F., Duplomb, L., Wang, Y., Robledo, O., Kinzie, E., Pitard, V., Godard, A., Jacques, Y., and Baumann, H. (2000). Stimulation of leukemia inhibitory factor receptor degradation by extracellular signal-regulated kinase. *J.Biol.Chem.* **275**, 28793-28801.

Borst, P. and Ouellette, M. (1995). New mechanisms of drug resistance in parasitic protozoa. *Annu.Rev.Microbiol.* **49**, 427-460.

Bruce, J. E., Anderson, G. A., Wen, J., Harkewicz, R., and Smith, R. D. (1999). High-mass-measurement accuracy and 100% sequence coverage of enzymatically digested bovine serum albumin from an ESI-FTICR mass spectrum. *Anal.Chem.* **71**, 2595-2599.

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**, 289-292.

Brun, R. and Schonenberger, M. (1981). Stimulating effect of citrate and cis-Aconitate on the transformation of *Trypanosoma brucei* bloodstream forms to procyclic forms *in vitro*. *Z.Parasitenkd.* **66**, 17-24.

Brun, R., Schumacher, R., Schmid, C., Kunz, C. and Burri, C. (2001). The phenomenon of treatment failures in Human African trypanosomiasis. *Trop.Med.Int.Health*, **6**, 906-914.

Carter, N. S. and Fairlamb, A. H. (1993). Arsenical-resistant trypanosomes lack an unusual adenosine transporter. *Nature* **361**, 173-176.

Carter, N. S., Berger, B. J. and Fairlamb, A. H., (1995). Uptake of diamidine drugs by the P2 nucleoside transporter in melarsen-sensitive and -resistant *Trypanosoma brucei brucei*. *J.Biol.Chem.* **270**, 28153-28157.

Celis, J. E., Kruhoffer, M., Gromova, I., Frederiksen, C., Ostergaard, M., Thykjaer, T., Gromov, P., Yu, J., Palsdottir, H., Magnusson, N., and Orntoft, T. F. (2000). Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett.* **480**, 2-16.

Chao, D. and Remi, G. (2001). *Helicobacter pylori*: an update and a practical approach. *J.La State Med.Soc.* **153**, 540-546.

Chizyuka, H. G. B. and Liguru, S. M. K. (1986). Dipping to control vectors of cattle parasites. *Parasitol.Today* **2**, 123

Clauser, K. R., Baker, P., and Burlingame, A. L. (1999). Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal.Chem.* **71**, 2871-2882.

Cohen, A. M., Rumpel, K., Coombs, G. H., and Wastling, J. M. (2002). Characterisation of global protein expression by two-dimensional electrophoresis and mass spectrometry: proteomics of *Toxoplasma gondii*. *Int.J.Parasitol.* **32**, 39-51.

Coppens, I., Baudhuin, P., Opperdoes, F. R., and Courtoy, P. J. (1988). Receptors for the host low density lipoproteins on the hemoflagellate *Trypanosoma brucei*: purification and involvement in the growth of the parasite. *Proc.Natl.Acad.Sci.U.S.A* **85**, 6753-6757.

Cunningham, M. L., Zvelebil, M. J., and Fairlamb, A. H. (1994). Mechanism of inhibition of trypanothione reductase and glutathione reductase by trivalent organic arsenicals. *Eur.J.Biochem.* **221**, 285-295.

Damper, D. and Patton, C. L. (1976). Pentamidine transport in *Trypanosoma brucei* - kinetics and specificity. *Biochem.Pharmacol.* **25**, 271-276.

De Koning, H. P. (2001). Uptake of pentamidine in *Trypanosoma brucei brucei* is mediated by three distinct transporters: implications for cross-resistance with arsenicals. *Mol.Pharmacol.* **59**, 586-592.

De Koning, H. P. and Jarvis, S. M. (1999). Adenosine transporters in bloodstream forms of *Trypanosoma brucei brucei*: substrate recognition motifs and affinity for trypanocidal drugs. *Mol.Pharmacol.* **56**, 1162-1170.

De Koning, H. P., Watson, C. J., Sutcliffe, L., and Jarvis, S. M. (2000). Differential regulation of nucleoside and nucleobase transporters in *Crithidia fasciculata* and *Trypanosoma brucei brucei*. *Mol.Biochem.Parasitol.* **106**, 93-107.

Denise, H. and Barrett, M. P. (2001). Uptake and mode of action of drugs used against sleeping sickness. *Biochem.Pharmacol.* **61**, 1-5.

Denise, H., Giroud, C., Barrett, M. P., and Baltz, T. (1999). Affinity chromatography using trypanocidal arsenical drugs identifies a specific interaction between glycerol-3-phosphate dehydrogenase from *Trypanosoma brucei* and Cymelarsan. *Eur.J.Biochem.* **259**, 339-346.

Dieckmann, T., Withers-Ward, E. S., Jarosinski, M. A., Liu, C. F., Chen, I. S., and Feigon, J. (1998). Structure of a human DNA repair protein UBA domain that interacts with HIV-1 Vpr. *Nat.Struct.Biol.* **5**, 1042-1047.

Diehl, S., Diehl, F., El Sayed, N. M., Clayton, C. and Hoheisel, J. D. (2002). Analysis of stage-specific expression in the bloodstream and the procyclic form of *Trypanosoma brucei* using a genomic DNA-microarray. *Mol.Biochem.Parasitol.* **123**, 115-123.

Djikeng, A., Agufa, C., Donelson, J. E. and Majiwa, P. A. O. (1998). Generation of expressed sequence tags as physical landmarks in the genome of *Trypanosoma brucei*. *Gene*, **221**, 93-106.

Dlugonska, H., Dytnerska, K., Reichmann, G., Stachelhaus, S., and Fischer, H. G. (2001). Towards the *Toxoplasma gondii* proteome: position of 13 parasite excretory antigens on a standardized map of two-dimensionally separated tachyzoite proteins. *Parasitol.Res.* **87**, 634-637.

Docampo, R. and Moreno, S. N. (1999). Acidocalcisome: A novel Ca2+ storage compartment in trypanosomatids and apicomplexan parasites. *Parasitol.Today* **15**, 443-448.

Docampo, R. and Moreno, S. N. (2001). The acidocalcisome. *Mol.Biochem.Parasitol.* **114**, 151-159.

Drose, S. and Altendorf, K. (1997). Bafilomycins and concanamycins as inhibitors of V-ATPases and P-ATPases. *J.Exp.Biol.* **200 (Pt 1)**, 1-8.

Drummelsmith, J., Brochu, V., Girard, I., Messier, N., and Ouellette, M. (2003). Proteome mapping of the protozoan parasite leishmania and application to the study of drug targets and resistance mechanisms. *Mol.Cell Proteomics.* **2**, 146-155.

Dukes, P. (1984). Arsenic and old taxa: subspeciation and drug sensitivity in *Trypanosoma brucei*. *Trans.R.Soc.Trop.Med.Hyg.* **78**, 711-725.

El Fakhry, Y., Ouellette, M., and Papadopoulou, B. (2002). A proteomic approach to identify developmentally regulated proteins in *Leishmania infantum*. *Proteomics.* **2**, 1007-1017.

El Sayed, N. M., Ghedin, E., Song, J., MacLeod, A., Bringaud, F., Larkin, C., Wanless, D., Peterson, J., Hou, L., Taylor, S., Tweedie, A., Biteau, N., Khalak, H. G., Lin, X., Mason, T., Hannick, L., Caler, E., Blandin, G., Bartholomeu, D., Simpson, A. J., Kaul, S., Zhao, H., Pai, G., Van Aken, S., Utterback, T., Haas, B., Koo, H. L., Umayam, L., Suh, B., Gerrard, C., Leech, V., Qi, R., Zhou, S., Schwartz, D., Feldblyum, T., Salzberg, S., Tait, A., Turner, C. M., Ullu, E., White, O., Melville, S., Adams, M. D., Fraser, C. M., and Donelson, J. E. (2003). The sequence and analysis of *Trypanosoma brucei* chromosome II. *Nucleic Acids Res.* **31**, 4856-4863.

El Sayed, N. M., Hegde, P., Quackenbush, J., Melville, S. E., and Donelson, J. E. (2000). The African trypanosome genome. *Int.J.Parasitol.* **30**, 329-345.

El Sayed, N. M. A., Alarcon, C. M., Beck, J. C., Sheffield, V. C. and Donelson, J. E. (1995). cDNA expressed sequence tags of *Trypanosoma brucei rhodesiense* provide new insights into the biology of the parasite. *Mol. Biochem.Parasitol.* **73**, 75-90.

Erdjument-Bromage, H., Lui, M., Lacomis, L., Grewal, A., Annan, R. S., McNulty, D. E., Carr, S. A., and Tempst, P. (1998). Examination of micro-tip reversed-phase liquid chromatographic extraction of peptide pools for mass spectrometric analysis. *J.Chromatogr.A* **826**, 167-181.

Eriksson, J., Chait, B. T., and Fenyo, D. (2000). A statistical basis for testing the significance of mass spectrometric protein identification results. *Anal.Chem.* **72**, 999-1005.

Evans, D. A. and Ellis, D. S. (1983). Recent observations on the behaviour of certain trypanosomes within their insect hosts. *Adv.Parasitol.* **22**, 1-42.

Fairlamb, A. H. (2002). Metabolic pathway analysis in trypanosomes and malaria parasites. *Philos.Trans.R.Soc.Lond B Biol.Sci.* **357**, 101-107.

Fairlamb, A. H. and Bowman, I. B. (1980). Uptake of the trypanocidal drug suramin by bloodstream forms of *Trypanosoma brucei* and its effect on respiration and growth rate *in vivo*. *Mol.Biochem.Parasitol.* **1**, 315-333.

Fairlamb, A. H., Henderson, G. B., and Cerami, A. (1989). Trypanothione is the primary target for arsenical drugs against African trypanosomes. *Proc.Natl.Acad.Sci.U.S.A* **86**, 2607-2611.

Fairlamb, A. H., Carter, N. S., Cunningham, M. and Smith, K. (1992). Characterisation of melarsen resistant *Trypanosoma brucei brucei* with respect to cross-resistance to other drugs and trypanothione metabolism. *Mol.Biochem.Parasitol.* **53**, 213-222.

Fenyo, D. (2000). Identifying the proteome: software tools. *Curr.Opin.Biotechnol.* **11**, 391-395.

Fevre, E. M., Coleman, P. O., Oditt, M., Magona, J. W., Welburn, S. C. and Woolhouse, M. E. (2001). The origins of a new *Trypanosoma brucei rhodesiense* sleeping sickness outbreak in eastern Uganda. *Lancet*, **358**, 625-628.

Field, H., Ali, B. R., Sherwin, T., Gull, K., Croft, S. L., and Field, M. C. (1999). TbRab2p, a marker for the endoplasmic reticulum of *Trypanosoma brucei*, localises to the ERGIC in mammalian cells. *J.Cell Sci.* **112 (Pt 2)**, 147-156.

Field, H. and Field, M. C. (1997). Tandem duplication of rab genes followed by sequence divergence and acquisition of distinct functions in *Trypanosoma brucei*. *J.Biol.Chem.* **272**, 10498-10505.

Fox, R. G. R., Mmbando, S. O., Fox, M. S. and Wilson, A. (1993). Effect on herd health and productivity of controlling tsetse and trypanosomiasis by applying deltamethrin to cattle. *Trop.Anim.Health Prod.* **25**, 203-214.

Frommel, T. O. and Balber, A. E. (1987). Flow cytofluorimetric analysis of drug accumulation by multidrug-resistant *Trypanosoma brucei brucei* and *T. b. rhodesiense*. *Mol.Biochem.Parasitol.* **26**, 183-192.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Manguall, C., Venter, J. C., Carucci, D. J., Hoffman S. L., Newbold, C., Davis, R. W., Fraser, C. M. and Barrell, B. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, **419**, 498-511.

Gibson, W. (2002). Epidemiology and diagnosis of African trypanosomiasis using DNA probes. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **96**, Suppl 1, S141-S143.

Gonzalez, R. J. and Tarloff, J. B. (2001). Evaluation of hepatic subcellular fractions for Alamar blue and MTT reductase activity. *Toxicol.In Vitro* **15**, 257-259.

Goodlett, D. R., Bruce, J. E., Anderson, G. A., Rist, B., Pasa-Tolic, L., Fiehn, O., Smith, R. D., and Aebersold, R. (2000). Protein identification with a single accurate mass of a cysteine-containing peptide and constrained database searching. *Anal.Chem.* **72**, 1112-1118.

Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., and Weiss, W. (2000). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21**, 1037-1053.

Gu, Z., Steinmetz, L. M., Gu, X., Scharfe, C., Davis, R. W., and Li, W. H. (2003). Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63-66.

Gutteridge, W. E., Jaffe, J. J., and McCormack, J. J., Jr. (1969). Presence and properties of dihydrofolate reductases in Trypanosomatid flagellate. *Trans.R.Soc.Trop.Med.Hyg.* **63**, 119.

Ha, G. H., Lee, S. U., Kang, D. G., Ha, N. Y., Kim, S. H., Kim, J., Bae, J. M., Kim, J. W., and Lee, C. W. (2002). Protcome analysis of human stomach tissue: separation of soluble proteins by two-dimensional polyacrylamide gel electrophoresis and identification by mass spectrometry. *Electrophoresis* **23**, 2513-2524.

Hall, N., Berriman, M., Lennard, N. J., Harris, B. R., Hertz-Fowler, C., Bart-Delabesse, E. N., Gerrard, C. S., Atkin, R. J., Barron, A. J., Bowman, S., Bray-Allen, S. P., Bringaud, F., Clark, L. N., Corton, C. H., Cronin, A., Davies, R., Doggett, J., Fraser, A., Gruter, E., Hall, S., Harper, A. D., Kay, M. P., Leech, V., Mayes, R., Price, C., Quail, M. A., Rabbinowitsch, E., Reitter, C., Rutherford, K., Sasse, J., Sharp, S., Shownkeen, R., MacLeod, A., Taylor, S.,

Tweedie, A., Turner, C. M., Tait, A., Gull, K., Barrell, B., and Melville, S. E. (2003). The DNA sequence of chromosome I of an African trypanosome: gene content, chromosome organisation, recombination and polymorphism. *Nucleic Acids Res.* **31**, 4864-4873.

Hanau, S., Rippa, M., Bertelli, M., Dallocchio, F., and Barrett, M. P. (1996). 6-Phosphogluconate dehydrogenase from *Trypanosoma brucei*. Kinetic analysis and inhibition by trypanocidal drugs. *Eur.J.Biochem.* **240**, 592-599.

Handman, E., Hocking, R. E., Mitchell, G. F., and Spithill, T. W. (1983). Isolation and characterization of infective and non-infective clones of *Leishmania tropica*. *Mol.Biochem.Parasitol.* **7**, 111-126.

Hay, R. T. (2001). Protein modification by SUMO. *Trends Biochem.Sci.* **26**, 332-333.

Hawking, F. (1963). Action of drugs upon *Trypanosoma congolense*, *T. vivax* and *T. rhodesiense* in tsetse flies and in culture.*Ann.Trop.Med.Parasitol.* **57**,255-261.

Hellman, U., Wernstedt, C., Gonez, J., and Heldin, C. H. (1995). Improvement of an "In-Gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Anal.Biochem.* **224**, 451-455.

Hide, G. (1999). History of sleeping sickness in East Africa. *Clinical Microbiology Review*, **12**, 112-125.

Hird, J. F., Cianciosi, S. C. and McLean, R. M. (1986). Investigations on the originand metabolism of the carbon skeleton of ornithine, arginine and proline in selected animals. *Comp.Biochem.Physiol.B.* **83**, 179-184.

Hofmann, K. and Bucher, P. (1996). The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biochem.Sci.* **21**, 172-173.

Hope, M., MacLeod, A., Leech, V., Melville, S., Sasse, J., Tait, A. and Turner, C. M. (1999). Analysis of ploidity (in megabase chromosome) in *Trypanosoma brucei* after genetic exchange. *Mol.Biochem.Parasitol.* **104**, 1-9.

Hunter, T. (1995). Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signalling. *Cell*, **80**, 225-236.

Huynh, T. T., Huynh, V. T., Harmon, M. A., and Phillips, M. A. (2003). Gene knockdown of gamma -glutamylcysteine synthetase by RNAi in the parasitic protozoa *Trypanosoma brucei* demonstrates that it is an essential enzyme. *J.Biol.Chem.*

Itens, M., Matovu, E., Brun, R. and Kaminsky, R. (1995). Innate lack of susceptibility of Ugandan *Trypanosoma brucei rhodesiense* to DL-a-difluoromethylornithine (DFMO). *Trop.Med.Parasitol.* **46**, 190-194.

Itens, M., Mett, H., Evans, A., Enyaru, J.C.K., Brun, R. and Kaminsky, R (1997). Alternations in ornithine decarboxylase characteristics account for tolerance for *Trypanosoma brucei rhodesiense* to DL-a-difluoromethylornithine. *Antimicrob.Agents Chemother.* **41**, 1922-1925.

Jungblut, P. R., Schaible, U. E., Mollenkopf, H. J., Zimny-Arndt, U., Raupach, B., Mattow, J., Halada, P., Lamer, S., Hagens, K., and Kaufmann, S. H. (1999). Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. *Mol.Microbiol.* **33**, 1103-1117.

Kanmogne, G. D., Stevens, J. R., Asonganyi, T. and Gibson, W. C. (1996). Characterisation of *Trypanosoma brucei gambiense* isolates using restriction fragment length polymorphism in 5 variant surface glycoprotein genes. *Acta Tropica*, **61**, 239-254.

Kaplan, H. G. and Myers, C. E. (1977). Complex inhibition of thymidylate synthase by aromatic diamidines: evidence for both rapid, freely reversible and slowly progressive, nonequilibrium inhibition. *J.Pharmacol.Exp.Ther.* **201**, 554-563.

Kaufmann, H., Bailey, J. E., and Fussenegger, M. (2001). Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. *Electrophoresis* **22**, 194-199.

Keiser, J. and Burri, C. (2000). Physico-chemical properties of the trypanocidal drug melarsoprol. *Acta Trop.* **74**, 101-104.

Keiser, J., Ericsson, O., and Burri, C. (2000). Investigations of the metabolites of the trypanocidal drug melarsoprol. *Clin.Pharmacol.Ther.* **67**, 478-488.

Krause, E., Wenschuh, H., and Jungblut, P. R. (1999). The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. *Anal.Chem.* **71**, 4160-4165.

Krogstad, D. J., Schlesinger, P. H. and Gluzman, I. Y. (1985). Antimalarials increase vesicles pH in *Plasmodium falciparum*. *J.Cell.Biol*, **101**, 2302-3309.

Lackie, J. M. and Dow, J. A. T. (1995). Dictionary of Cell Biology. London, Academic press.

LaCount, D. J., Gruszynski, A. E., Grandgenett, P. M., Bangs, J. D., and Donelson, J. E. (2003). Expression and function of the *Trypanosoma brucei* major surface protease (GP63) genes. *J.Biol.Chem.* **278**, 24658-24664.

Lamer, S. and Jungblut, P. R. (2001). Matrix-assisted laser desorption-ionization mass spectrometry peptide mass fingerprinting for proteome analysis: identification efficiency after on-blot or in-gel digestion with and without desalting procedures. *J.Chromatogr.B Biomed.Sci.Appl.* **752**, 311-322.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A.,

Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la, B. M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., Szustakowki, J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., and Choi, S. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

Landy, A. (1989). Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu.Rev.Biochem.* **58**, 913-949.

Langreth, S. G. and Balber, A. E. (1975). Protein uptake and digestion in bloodstream and culture forms of *Trypanosoma brucei*. *J.Protozool.* **22**, 40-53.

Lanne, B., Potthast, F., Hoglund, A., Brockenhuus, v. L., Nystrom, A. C., Nilsson, F., and Dahllof, B. (2001). Thiourea enhances mapping of the proteome from murine white adipose tissue. *Proteomics.* **1**, 819-828.

Lauber, W. M., Carroll, J. A., Dufield, D. R., Kiesel, J. R., Radabaugh, M. R., and Malone, J. P. (2001). Mass spectrometry compatibility of two-dimensional gel protein stains. *Electrophoresis* **22**, 906-918.

Leak, S. G. A., Woudyalew, M, Rowlands, M. and D'Ieteren, G. D. M. (1995). A trial of cypermethrin "pour-on" to control *Glossina pallidipes*, *G. fusciles*, and *G. morsitans submorsitans* (Dipteria: Glossinidae) in south-west Ethiopia. *Bull.Ent.Res.* **85**, 241-251.

Lee, M. G., Yen, F. T., Zhang, Y., and Bihain, B. E. (1999). Acquisition of lipoproteins in the procyclic form of *Trypanosoma brucei*. *Mol.Biochem.Parasitol.* **100**, 153-162.

Legros, D., Fournier, C., Gastellu, E. M., Maiso, F., and Szumilin, E. (1999). [Therapeutic failure of melarsoprol among patients treated for late stage *T. b. gambiense* human African trypanosomiasis in Uganda]. *Bull.Soc.Pathol.Exot.* **92**, 171-172.

Lemercier, G., Dutoya, S., Luo, S., Ruiz, F. A., Rodrigues, C. O., Baltz, T., Docampo, R., and Bakalara, N. (2002). A vacuolar-type H+-pyrophosphatase governs maintenance of functional acidocalcisomes and growth of the insect and mammalian forms of *Trypanosoma brucei*. *J.Biol.Chem.* **277**, 37369-37376.

Liu, J., Qiao, X., Du, D., and Lee, M. G. (2000). Receptor-mediated endocytosis in the procyclic form of *Trypanosoma brucei*. *J.Biol.Chem.* **275**, 12032-12040.

MacLeod, A., Turner, C. M., and Tait, A. (1999). A high level of mixed *Trypanosoma brucei* infections in tsetse flies detected by three hypervariable minisatellites. *Mol.Biochem.Parasitol.* **102**, 237-248.

Madison-Antenucci, S., Grams, J. and Hajduck, S. L. (2002). Editing machines: the complexities of trypanosomes RNA editing. *Cell*, **108**, 435-438.

Madura, K. (2002). The ubiquitin-associated (UBA) domain: on the path from prudence to prurience. *Cell Cycle*, **1**, 235-244.

Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, H. O., Bishop, J. R., Falcone, F. H., Gavrilescu, C., Montgomery, J. L., Santori, M. I., Stern, L. S., Wang, Z., Ullu, E. and Tschudi, C. (2000). A new twist in trypanosome RNA metabolism: cis-plicing of pre-mRNA. *RNA*, **6**, 163-169.

Malik, Z. A., Iyer, S. S., and Kusner, D. J. (2001). *Mycobacterium tuberculosis* phagosomes exhibit altered calmodulin-dependent signal transduction: contribution to inhibition of phagosome-lysosome fusion and intracellular survival in human macrophages. *J.Immunol.* **166**, 3392-3401.

Mann, M., Hendrickson, R. C., and Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. *Annu.Rev.Biochem.* **70**, 437-473.

Mann M., Ong S.-E., Gronborg, M., Jensen, O. N. and Pandey, A. (2002). Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotech.* **20**, 261-268.

Mann M. and Jensen, O.N (2003). Proteomic analysis of post-translational modifications. *Nature Biotech.*, **21**, 255-261.

Maser, P., Sutterlin, C., Kralli, A., and Kaminsky, R. (1999). A nucleoside transporter from *Trypanosoma brucei* involved in drug resistance. *Science* **285**, 242-244.

Matovu, E., Enyaru, J. C., Legros, D., Schmid, C., Seebeck, T., and Kaminsky, R. (2001). Melarsoprol refractory *T. b. gambiense* from Omugo, north-western Uganda. *Trop.Med.Int.Health* **6**, 407-411.

Matovu, E., Geiser, F., Schneider, V., Maser, P., Enyaru, J. C., Kaminsky, R., Gallati, S., and Seebeck, T. (2001). Genetic variants of the *TbAT1* adenosine transporter from African trypanosomes in relapse infections following melarsoprol therapy. *Mol.Biochem.Parasitol.* **117**, 73-81.

Matovu, E., Stewart, M.L., Geiser, F., Brun, R., Maser, P., Wallace, L. J. M., Burchmore, R. J., Enyaru, J. C. K., Barrett, M. P., Kaminsky, R., Seebeck, T. and de Koning, H. P. (2003). Mechanisms of arsenical and diamidine uptake and resistance in *Trypanosoma brucei*. *Eukaryotic Cell.* **2**, 1003-1008.

Matthews, K. R., Tschudi, C., and Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev.* **8**, 491-501.

McAtee, C. P., Hoffman, P. S., and Berg, D. E. (2001). Identification of differentially regulated proteins in metronidozole resistant *Helicobacter pylori* by proteome techniques. *Proteomics.* **1**, 516-521.

McKean, P. G., Baines, A., Vaughan, S., and Gull, K. (2003). gamma-Tubulin Functions in the Nucleation of a Discrete Subset of Microtubules in the Eukaryotic Flagellum. *Curr.Biol.* **13**, 598-602.

Melville, S.E., Leech, V., Gerrard, C.S., Tait, A., and Blackwell, J.M. (1998). The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. *Mol.Biochem.Parasitol.* **94**, 155-173

Melville, S. E., Leech, V., Navarro, M., and Cross, G. A. (2000). The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* stock 427. *Mol.Biochem.Parasitol.* **111**, 261-273.

Minning, T. A., Bua, J., Garcia, G. A., McGraw, R. A. and Tarleton, R. L. (2003). Microarray profiling of gene expression during trypanomastigote to amastigote transition in *Trypanosoma cruzi*. *Mol.Biochem.Parasitol.* **131**, 55-64.

Miranda, K., Benchimol, M., Docampo, R., and de Souza, W. (2000). The fine structure of acidocalcisomes in *Trypanosoma cruzi*. *Parasitol.Res.* **86**, 373-384.

Misset, O. and Opperdoes, F. R. (1987). The phosphoglycerate kinases from *Trypanosoma brucei*. A comparison of the glycosomal and the cytosolic isoenzymes and their sensitivity towards suramin. *Eur.J.Biochem.* **162**, 493-500.

Misset, O., Van Beeumen, J., Lambeir, A. M., Van der, M. R., and Opperdoes, F. R. (1987). Glyceraldehyde-phosphate dehydrogenase from *Trypanosoma brucei*. Comparison of the glycosomal and cytosolic isoenzymes. *Eur.J.Biochem.* **162**, 501-507.

Modrek, B. and Lee, C. (2002). A genomic view of alternative splicing. *Nat.Genet.* **30**, 13-19.

Modrek, B., Resch, A., Grasso, C., and Lee, C. (2001). Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.* **29**, 2850-2859.

Molloy, M. P. (2000). Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients. *Anal.Biochem.* **280**, 1-10.

Montalvetti, A., Fernandez, A., Sanders, J. M., Ghosh, S., Van Brussel, E., Oldfield, E., and Docampo, R. (2003). Farnesyl pyrophosphate synthase is an essential enzyme in *Trypanosoma brucei*. *In vitro* RNA interference and *in vivo* inhibition studies. *J.Biol.Chem.* **278**, 17075-17083.

Moore, A. and Richer, M. (2001). Re-emergence of epidemic sleeping sickness in southern Sudan. *Trop.Med.Int.Health* **6**, 342-347.

Mueller, T. D. and Feigon, J. (2002). Solution structures of UBA domains reveal a conserved hydrophobic surface for protein-protein interactions. *J.Mol.Biol.* **319**, 1243-1255.

Murray, M. and Morrison, W. I. (1979). Non-specific induction of increased resistance in mice to *Trypanosoma congolense* and *Trypanosoma brucei* by immunostimulants. *Parasitology* **79**, 349-366.

Murray, M., Trail, J. C. M., Davis, C. E. and Black, S. J. (1984). Genetic resistance to Africa trypanosomiasis. *J.Inf.Dis.* **149**, 311-319.

Neal, R. A., van Bueren, J., McCoy, N. G. and Iwobi, M. (1989). Reversal of drug resistance in *Trypanosoma cruzi* and *Leishmania donovani* by verapamil. *Trans.R.Soc.Trop.Med.Hyg.*, **83**, 197-198.

Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. (1998). Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat.Genet.* **20**, 46-50.

Ngo, H., Tschudi, C., Gull, K., and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc.Natl.Acad.Sci.U.S.A* **95**, 14687-14692.

Nischik, N., Schade, B., Dytnerska, K., Dlugonska, H., Reichmann, G., and Fischer, H. G. (2001). Attenuation of mouse-virulent *Toxoplasma gondii* parasites is associated with a

decrease in interleukin-12-inducing tachyzoite activity and reduced expression of actin, catalase and excretory proteins. *Microbes.Infect.* **3**, 689-699.

Nyang'ao, J. M., Olaho-Mukani, W., Maribei, J. M., and Omuse, J. K. (1995). Evaluation of the efficacy of melarsenoxyde cysteamine (Cymelarsan) in treatment of camels experimentally infected with *Trypanosoma evansi* using antigen trapping enzyme-linked immunosorbent assay. *J.Vet.Pharmacol.Ther.* **18**, 468-470.

O'Brien, J., Wilson, I., Orton, T., and Pognan, F. (2000). Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *Eur.J.Biochem.* **267**, 5421-5426.

Olila, D., McDermot, J. J., Esiler, M. C., Mitema, E. S., Patzelt, R. J. and Peregrine, A. S. (2002). Drug sensitivity of trypanosome populations from cattle in a peri-urban dairy production system in Uganda. *Acta Tropica*, **84**, 19-30.

Pal, A., Hall, B. S., and Field, M. C. (2002). Evidence for a non-LDL-mediated entry route for the trypanocidal drug suramin in *Trypanosoma brucei*. *Mol.Biochem.Parasitol.* **122**, 217-221.

Pal, A., Hall, B. S., Jeffries, T. R., and Field, M. C. (2003). Rab5 and Rab11 mediate transferrin and anti-variant surface glycoprotein antibody recycling in *Trypanosoma brucei*. *Biochem.J.* **374**, 443-451.

Pappin, D. J. C., Hojrup, P. and Bleasby, A. J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Current Biol.* **3**, 327-332.

Patton, W. F. (2000). A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis* **21**, 1123-1144.

Payne, R. C., Sukanto, I. P., Partoutomo, S., and Jones, T. W. (1994). Efficacy of Cymelarsan treatment of suramin resistant *Trypanosoma evansi* in cattle. *Trop.Anim Health Prod.* **26**, 92-94.

Pays, E. (1995). [Antigenic variation and the problem of vaccines against African trypanosomes]. *Bull.Mem.Acad.R.Med.Belg.* **150**, 123-131.

Pearson, T. W. and Jenni, L. (1989). Detection of hybrid phenotypes in African trypanosomes by high resolution two-dimensional gel electrophoresis. *Parasitol.Res.* **76**, 63-67.

Pearson, T. W., Kar, S. K., McGuire, T. C., and Lundin, L. B. (1981). Trypanosome variable surface antigens: studies using two-dimensional gel electrophoresis and monoclonal antibodies. *J.Immunol.* **126**, 823-828.

Pepin, J. and Milord, F. (1994). The treatment of human African trypanosomiasis. *Adv.Parasitol.* **33**, 1-47.

Pereira, C. A., Alonso, G. D., Paveto, M. C., Flawia, M. M., and Torres, H. N. (1999). L-arginine uptake and L-phosphoarginine synthesis in *Trypanosoma cruzi*. *J.Eukaryot.Microbiol.* **46**, 566-570.

Pereira, C. A., Alonso, G. D., Paveto, M. C., Iribarren, A., Cabanas, M. L., Torres, H. N., and Flawia, M. M. (2000). *Trypanosoma cruzi* arginine kinase characterization and cloning. A novel energetic pathway in protozoan parasites. *J.Biol.Chem.* **275**, 1495-1501.

Pereira, C. A., Alonso, G. D., Torres, H. N., and Flawia, M. M. (2002). Arginine kinase: a common feature for management of energy reserves in African and American flagellated trypanosomatids. *J.Eukaryot.Microbiol.* **49**, 82-85.

Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567.

Pospichal, H., Brun, R., Kaminsky, R. and Jenni, L. (1994). Induction of resistance to melarsenoxide cysteamine (Mel Cy) in *Trypanosoma brucei brucei*. *Acta Trop.* **58**, 187-197.

Rabilloud, T. (1998). Use of thiourea to increase the solubility of membrane proteins in two-dimensional electrophoresis. *Electrophoresis* **19**, 758-760.

Rabilloud, T., Adessi, C., Giraudel, A., and Lunardi, J. (1997). Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **18**, 307-316.

Raz, B., Iten, M., Grether-Buhler, Y., Kaminsky, R., and Brun, R. (1997). The Alamar Blue assay to determine drug sensitivity of African trypanosomes (*T. b. rhodesiense* and *T. b. gambiense*) in vitro. *Acta Trop.* **68**, 139-147.

Rodrigues, C. O., Scott, D. A., and Docampo, R. (1999). Characterization of a vacuolar pyrophosphatase in *Trypanosoma brucei* and its localization to acidocalcisomes. *Mol.Cell Biol.* **19**, 7712-7723.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning, A laboratory manual. Cold Spring Harbour Laboratory Press.

Santoni, V., Kieffer, S., Desclaux, D., Masson, F. and Rabilloud, T. (2000). Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. *Electrophoresis*, **21**, 3329-3344.

Santoni, V., Malloy, M. and Rabilloud, T. (2000). Membrane proteins and proteomics: un amour impossible? *Electrophoresis*, **21**, 1054-1070.

Schofield, C. J. and Maudlin, I. (2001). Trypanosomiasis control. *Int.J.Parasitol.* **31**, 614-619.

Scott, A. G., Tait, A., and Turner, C. M. (1996). Characterisation of cloned lines of *Trypanosoma brucei* expressing stable resistance to MelCy and suramin. *Acta Trop.* **60**, 251-262.

Scott, A. G., Tait, A., and Turner, C. M. (1997). *Trypanosoma brucei*: lack of cross-resistance to melarsoprol *in vitro* by cymelarsan-resistant parasites. *Exp.Parasitol.* **86**, 181-190.

Scott, D. A. and Docampo, R. (1998). Two types of H+-ATPase are involved in the acidification of internal compartments in *Trypanosoma cruzi*. *Biochem.J.* **331 (Pt 2)**, 583-589.

Scott, D. A. and Docampo, R. (2000). Characterization of isolated acidocalcisomes of *Trypanosoma cruzi*. *J.Biol.Chem.* **275**, 24215-24221.

Seed, J. R. (2000). Current status of African trypanosomiasis. *ASM News*, **66**, 395-402.

Sereno, D., Holzmuller, P. and Lemesre, J. L. (2000). Efficacy of second line drugs on antimonyl-resistant amastigotes of *Leishmania infantum*. *Acta Trop.* **74**, 25-31.

Shahi, S. K., Krauth-Siegel, R. L., and Clayton, C. E. (2002). Overexpression of the putative thiol conjugate transporter TbMRPA causes melarsoprol resistance in *Trypanosoma brucei*. *Mol.Microbiol.* **43**, 1129-1138.

Shahid, M. A. and Curtis, C. F. (1987). Radiation sterilization and cytoplasmic incompatibility in a "tropicalised" strain of *Culex pipiens* complex (Dipteria: Culicidae). *J.Med.Entomol.* **24**, 273-274.

Shapiro, T. A. (1993). Inhibition of topoisomerases in African trypanosomes. *Acta trop.* **54**, 251-260.

Sherman, D. R., Jang, L., Hug, M. and Clayton, C. (1991). Anatomy of the parp gene promoter of *Trypanosoma brucei*. *EMBO J.* **10**, 3379-3386.

Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., and Mann, M. (1996). Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc.Natl.Acad.Sci.U.S.A* **93**, 14440-14445.

Shi, X., Parthun, M. R., and Jaehning, J. A. (1995). The yeast EGD2 gene encodes a homologue of the alpha NAC subunit of the human nascent-polypeptide-associated complex. *Gene* **165**, 199-202.

Stanghellini, A. and Josenando, T. (2001). The situation of sleeping sickness in Angola: a calamity. *Trop.Med.Int.Health* **6**, 330-334.

Smith, D. H., Pepin, J. and Stich, A. H. (1998). Human African trypanosomiasis: an emerging public health crisis. *Br.Med.Bull.* **54**, 341-355.

Smolka, M. B., Martins, D., Winck, F. V., Santoro, C. E., Castellari, R. R., Ferrari, F., Brum, I. J., Galembeck, E., Filho, H. D. C., Machado, M. A., Marangoni, S. and Novello, J. C. (2003). Proteome analysis of the plant pathogen *Xylella fastidiosa* reveals major cellular and extracellular proteins and a peculiar codon bias distribution. *Proteomics* **3**, 224-237.

Snow W. F., Declercq, J. and van Nieuwenhove, S. (1991). Watering sites in *Glossina fuscipes* habitat as the major foci for the transmission of Gambiense sleeping sickness in an endemic area of southern Sudan. *Ann.Soc.Belg.Med.Trop.* **71**, 27-38.

Sternberg, J. and Tait, A. (1990). Genetic exchange in African trypanosomes. *Trends Genet.* **6**, 317-322.

Sternberg, J., Turner, C. M., Wells, J. M., Ranford-Cartwright, L. C., Le Page, R. W., and Tait, A. (1989). Gene exchange in African trypanosomes: frequency and allelic segregation. *Mol.Biochem.Parasitol.* **34**, 269-279.

Stevens, J. R., Noyes, H. A., Dover, G. A. and Gibson, W. C. (1999). The ancient and divergent origins of the human pathogenic trypanosomes, *Trypaosoma brucei* and *T. cruzi*. *Parasitol.* **118**, 107-116.

Sutherland, I. A. and Holmes, P. H. (1993). Alterations in drug transport in resistant *Trypanosoma congolense*. *Acta Tropica*, **54**, 271-278.

Tait, A., Masiga, D., Ouma, J., MacLeod, A., Sasse, J., Melville, S., Lindegard, G., McIntosh, A., and Turner, M. (2002). Genetic analysis of phenotype in *Trypanosoma brucei*: a classical approach to potentially complex traits. *Philos.Trans.R.Soc.Lond B Biol.Sci.* **357**, 89-99.

Tang, C., Zhang, W., Chait, B., T. and Fenyo, D. (2000). A method to evaluate the quality of database search results. http://129.85.19.192/profound/zscore.pdf

Tang, C., Zhang, W., Fenyo, D. and Chait, B., T. (2000). Assessing the performance of different protein identification algorithms. http://129.85.19.192/profound/result_evaluation.pdf

Tetley, L. and Vickerman, K. (1985). Differentiation in *Trypanosoma brucei*: host-parasite cell junctions and their persistence during acquisition of the variable antigen coat. *J.Cell Sci.* **74**, 1-19.

Torri, A. F., Bertrand, K. I., and Hajduk, S. L. (1993). Protein stability regulates the expression of cytochrome c during the developmental cycle of *Trypanosoma brucei*. *Mol.Biochem.Parasitol.* **57**, 305-315.

Truc, P. and Tibayrenc, M. (1993). Population genetics of *Trypanosoma brucei* in central Africa: taxonomic and epidemiological significance. *Parasitol.* **106**, 137-149.

Tschudi, C. and Ullu, E. (1988). Polygene transcripts are precursors to calmodulin mRNAs in trypanosomes. *EMBO J.* **7**, 455-463.

Turner, C. M., Sternberg, J., Buchanan, N., Smith E, Hide, G. and Tait, A. (1990). Evidence that the mechanism of gene exchange in *Trypanosoma brucei* involves meiosis and syngamy. *Parasitol.* **110**, 377-386.

Ullu, E., Djikeng, A., Shi, H., and Tschudi, C. (2002). RNA interference: advances and questions. *Philos.Trans.R.Soc.Lond B Biol.Sci.* **357**, 65-70.

Vale, G. A., Lovemore, D. F., Flint, S. and Cockbill, G. F. (1988). Odour-baited targets to control tsetse flies, *Glossina* spp. (Dipteria: Glossinidae), in Zimbabwe. *Bull.Ent.Res.* **78**, 31-49.

Van Deursen, F. J., Thornton, D. J., and Matthews, K. R. (2003). A reproducible protocol for analysis of the proteome of *Trypanosoma brucei* by 2-dimensional gel electrophoresis. *Mol.Biochem.Parasitol.* **128**, 107-110.

Vanhamme, L. and Pays, E. (1995). Control of gene expression in trypanosomes. *Microbiol.Rev.* **59**, 223-240.

Van Nieuwenhove, S., Schechter, P. J., Declercq, J., Bone, G., Burke, J., and Sjoerdsma, A. (1985). Treatment of gambiense sleeping sickness in the Sudan with oral DFMO (DL-alpha-difluoromethylornithine), an inhibitor of ornithine decarboxylase; first field trial. *Trans.R.Soc.Trop.Med.Hyg.* **79**, 692-698.

Van Nieuwenhove, S., Betu-Ku-Mesu, V. K, Diabakana, P. M., Declercq, J. and Bilenge C. M. M. (2001). Sleeping sickness resurgence in the DRC: the past decade. *Trop.Med.Int.Health* **6**, 335-341.

Vansterkenburg, E. L., Coppens, I., Wilting, J., Bos, O. J., Fischer, M. J., Janssen, L. H., and Opperdoes, F. R. (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**, 237-250.

Vercesi, A. E. and Docampo, R. (1992). Ca^{2+} transport by digitonin-permeabilised *Leishmania donovani*. Effect of Ca^{2+}, pentamidine and WR-6026 on mitochondrial membrane potential *in situ*. *Biochem.J.*, **284**, 463-467.

Vickerman, K. (1965). Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. *Nature* **208**, 762-766.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. *Br.Med.Bull.* **41**, 105-114.

Waalkes, T. P. and Muluku, D. R. (1976). Pharmacologic aspects of pentamidine. *Natl.Cancer Inst.Monogr.*, **43**, 171-177.

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

Wang, S., Sakai, H., and Wiedmann, M. (1995). NAC covers ribosome-associated nascent chains thereby forming a protective environment for regions of nascent chains just emerging from the peptidyl transferase center. *J.Cell Biol.* **130**, 519-528.

Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., and Humphery-Smith, I. (1995). Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* **16**, 1090-1094.

Wasinger, V. C., Pollack, J. D., and Humphery-Smith, I. (2000). The proteome of *Mycoplasma genitalium*. Chaps-soluble component. *Eur.J.Biochem.* **267**, 1571-1582.

Wildgruber, R., Harder, A., Obermaier, C., Boguth, G., Weiss, W., Fey, S. J., Larsen, P. M., and Gorg, A. (2000). Towards higher resolution: two-dimensional electrophoresis of *Saccharomyces cerevisiae* proteins using overlapping narrow immobilized pH gradients. *Electrophoresis* **21**, 2610-2616.

Wildgruber, R., Reil, G., Drews, O., Parlar, H., and Gorg, A. (2002). Web-based two-dimensional database of *Saccharomyces cerevisiae* proteins using immobilized pH gradients from pH 6 to pH 12 and matrix-assisted laser desorption/ionization-time of flight mass spectrometry. *Proteomics.* **2**, 727-732.

Wilkins, M. R., Sanchez, J. C., Williams, K. L., and Hochstrasser, D. F. (1996). Current challenges and future applications for protein maps and post- translational vector maps in proteome projects. *Electrophoresis* **17**, 830-838.

Wilkins, M. R., Gasteiger, E., Sanchez, J.-C., Bairoch, A., and Hochstrasser, D. F. (1998). Two dimensional gel electrophoresis for proteome projects: the effects of protein hydrophobicity and copy number. *Electrophoresis*, **19**, 1501-1505.

Wilkinson, C. R., Seeger, M., Hartmann-Petersen, R., Stone, M., Wallace, M., Semple, C., and Gordon, C. (2001). Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. *Nat.Cell Biol.* **3**, 939-943.

Yarlett, N., Goldberg, B., Nathan, H. C., Garofalo, J. and Bacchi, C. (1991). Differential sensitivity of *Trypanosoma brucei rhodesiense* isolates to *in vitro* lysis by arsenicals. *Exp.Parasitol.*, **72**, 205-215.

Yates, J. R., III, Speicher, S., Griffin, P. R., and Hunkapiller, T. (1993). Peptide mass maps: a highly informative approach to protein identification. *Anal.Biochem.* **214**, 397-408.

Zhang, M. Q. (2002). Computational prediction of eukaryotic protein-coding genes. *Nat.Rev.Genet.* **3**, 698-709.

Zhang, W. and Chait, B. T. (2000). ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal.Chem.* **72**, 2482-2489.

Zhang, W. and Chait, B. T. (1995). Protein identification by database searcing: a Bayesian algorithm. http://prowl.rockefeller.edu/preprint/zhang/zhang-asms-1995.html.

Zhang, X., Lee, J., and Chasin, L. A. (2003). The effect of nonsense codons on splicing: a genomic analysis. *RNA.* **9**, 637-639.

Zhang, Z., Harrison, P., and Gerstein, M. (2002). Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome. *Genome Res.* **12**, 1466-1482.

Zomerdijk, J. C., Ouellette, M., ten Asbroek, A. L., Kieft, R., Bommer, A. M., Clayton, C. and Borst, P. (1990). The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. *EMBO J.* **9**, 2791-2801.

Spot 86
1041.5875
1076.5659
1125.6867
1128.6177
1146.6239
1227.6660
1253.7993
1341.6841
1344.5850
1631.8984
1654.8755
1679.8131
1847.9068
2032.9688
2048.9784
2285.1632
2814.3631

Spot 87
877.0680
885.3124
1066.0831
1093.2466
1173.7304
1338.5169
1554.0112
1663.9053
1735.8503
1803.9832
1811.8783
1918.7002
1994.0883
2196.7214
2225.0826
2244.0891
2285.0872
2436.1253
2656.8334
2699.2678
2808.3740
2851.5175

Spot 88
Lost

Spot 89
833.4450
861.1168
877.1293
960.7006
1036.5423
1295.7636
1301.1047
1318.7155
1360.6261
1376.6555
1472.1307
1578.9490
1581.7595
1595.8142
1611.7872
1639.4897
1703.7910
1743.9769
1849.9297
1912.9711
1956.4421
2129.8322
2230.3628
2384.9706
2476.5145
2625.3458
2835.1060

Spot 90
877.0901
894.3222
926.4662
1032.5117
1036.5177
1160.4566
1199.6811
1349.7585
1766.7501
1389.7719
1497.7624
1552.7584
1626.5366
1656.3718
1795.9000
1880.0513
1911.9206
2039.9477
2093.0882
2370.4568
2410.1697
2565.8991
2807.4102
2811.2409

Spot 91
Lost

Spot 92
882.4037
1018.3845
1036.4998
1199.6607
1311.5839
1318.5517
1323.6628
1345.4626
1349.7263
1371.6906
1389.7139
1497.7460
1552.6887
1631.7732
1638.8130
1796.7599
1879.9793
1911.8802
2039.9416
2386.1391
2564.0575

Spot 93
919.3633
1064.6084
1132.5812
1283.5214
1379.7275
1396.7413
1443.9016
1446.8440
1525.7592
1664.9552
1701.9691
1705.8723
1718.9267
1823.9835
2041.1774
2286.1180
2330.1063
2409.3171

Spot 94
877.1253
906.3609
918.3600
1045.5354
1129.5829
1135.5155
1206.2620
1330.6491
1396.7964
1412.8203
1435.7726
1439.8397
1457.8619
1521.8319
1587.7747
1621.8125
1637.6653
1828.8262
1848.9549
1884.0443
2065.1811
2239.1926
2284.1807
2314.0618
2531.6561
2809.4913

Spot 95
877.1371
1086.5457
1116.2265
1186.5016
1266.5060
1280.7093
1323.5916
1406.6952

1442.7812
1459.6922
1728.8925
1904.5915
2272.7236
2482.7892
2633.6177

Spot 96
917.3916
1062.6126
1225.5987
1282.6599
1479.0598
1532.8411
1535.7718
1678.8998
1800.9081
1878.9892
2135.9794
2285.2611
2525.4272
2705.2419
2807.3478

Spot 97
855.3284
1034.5321
1106.6553
1225.5907
1477.6612
1532.8299
1577.8306
1678.9630
1801.9843
1837.9612
1878.9369
2225.2191
2284.1930
2450.8923
2753.2922
2807.4508

Spot 98
919.4670
922.3160
929.3589
1062.2995
1079.5379
1117.5449
1318.7044
1461.7263
1495.7535
1619.6981
1688.8119
1692.7660
1907.8966
1923.8576
2136.0931
2141.3029
2170.1087
2272.1001
2321.0899
2500.7796
2750.5167
2807.3809

Spot 99
847.3699
870.4827
933.1896
968.3747
1025.5341
1109.4820
1118.5268
1165.6051
1179.6084
1277.7482
1301.6205
1344.6803
1365.6585
1475.7784
1707.7859
1699.8623
1876.1934
1994.0067
2124.8628
2367.2955
2380.8649
2605.3118
2808.3029

Spot 100
Lost

Spot 101
889.3893
980.4792
1045.5524
1095.5840
1092.3815
1319.6205
1372.6673
1389.6477
1503.7185
1813.4393
2017.0678
2034.9755
2049.0272
2145.1152
2240.1113
2284.1449
2504.0363
2525.2451
2802.2987
2807.3325

Spot 102
832.3342
952.4180
1054.5569
1101.5322
1187.5859
1267.5730
1385.6827
1460.8203
1476.7578
1648.8068
1680.9377
1732.9356
1834.9972
1883.0695
1900.0125
2092.3689
2239.1775
2276.2027
2285.1742
2301.9876
2566.1265
2807.3962
2811.3023

Spot 103
862.1674
870.4652
1045.6407
1167.8628
1267.6417
1338.5812
1348.6739
1498.7213
1560.7025
1631.9045
1680.0077
1685.2506
1855.9083
2037.4043
2225.1759
2239.1792
2437.1300
2745.8076

Spot 104
861.0598
883.3925
966.4601
1023.5692
1133.5407
1152.5434
1159.5725
1230.6569
1396.6230
1437.7582
1546.8612
1567.8284
1580.7241
1691.8831
1750.8879
1851.8779
2017.3615
2239.1816
2254.0653
2285.1959
2360.1049

2469.2518
2523.4968
2786.6636
2798.2762
2808.4347

Spot 105
899.5029
1273.6732
1288.5165
1309.7281
1327.6197
1343.6013
1415.6080
1812.9290
1930.9603
1966.0155
1968.9667
2003.8935
2007.8026
2239.1890
2284.1932
2300.3406
2643.1712
2807.3110

Spot 106
864.3755
993.5073
1049.5325
1159.4945
1301.5933
1351.6716
1536.8172
1664.7776
1692.8880
1784.3474
1804.8209
1820.9351
2053.8253
2225.0360
2341.0363
2494.1333
2654.9383
2774.1791
2811.1430

Spot 107
967.4616
1069.5149
1121.5094
1283.7063
1305.6738
1310.6566
1439.8175
1479.7922
1502.6401
1537.8319
1567.7890
1578.9384
1639.9688
1710.8699
1763.6771
1880.8951
1907.8761
1927.7592
2057.1848
2239.1563
2284.1777
2421.7330
2564.2060
2577.1685
2808.2972

Spot 108
861.1257
925.4844
1158.8608
1171.3839
1367.6268
1469.8824
1475.7754
1513.8720
1525.8780
1671.1380
1902.8501
1988.1075
1994.0571
2144.2111
2239.1576
2284.1687
2387.1699
2587.7650
2658.6843
2781.0817
2807.4176

Spot 109
862.2648
1008.4855
1016.3843
1045.5423
1050.5785
1233.6998
1297.4043
1402.6578
1501.7603
1534.7585
1570.4992
1730.2498
1905.9161
1946.8762
2068.9083
2239.1524
2288.0019
2516.1008
2688.4009
2807.422o

Spot 110
873.4682
982.4687
1126.5988
1176.6501
1213.7142
1324.5756
1463.7822
1478.5131
1641.0474
1746.9238
1768.8164
1915.8478
1960.0396
2147.0608
2165.1643
2341.9955
2517.2355
2680.6737
2807.3559

Spot 111
847.4571
973.5366
993.5155
1066.4964
1109.5320
1118.4859
1165.5898
1179.5879
1234.6532
1285.5678
1320.5819
1383.6348
1427.7370
1475.7156
1614.7838
1699.8099
1707.7543
1716.8236
1965.9640
1993.9622
2003.9037
2078.6143
2368.2266
2384.8308
2538.8548
2638.2676
2767.2000

Spot 112
958.3737
979.4847
1107.5699
1160.5257
1176.6084
1209.4820
1279.5862
1381.6417
1393.6155
1471.7325
1570.7777
1599.8002
1796.6974
1848.7974

1933.9533
2091.0103
2239.1051
2292.9855
2456.9571
2661.1839

Spot 113
948.4158
989.4363
1050.5227
1118.4926
1141.5947
1188.5144
1205.5971
1293.5450
1370.8315
1390.6757
1500.7479
1513.7777
1587.7858
1654.8367
1821.9918
2070.0327
2152.0608
2225.1149
2284.1259
2419.1894
2576.9565
2680.4658
2807.3700
2811.3442
2847.0237

Spot 114
1059.4419
1076.5408
1146.5995
1253.7840
1274.7260
1341.6608
1344.6088
1481.7469
1631.8342
1654.8324
1670.7877
1679.7777
1847.8794
2165.9632
2812.4061

Spot 115
1078.5956
1085.7269
1191.4636
1235.6542
1238.5612
1251.6058
1289.7211
1432.7804
1626.1480
1765.9137
1821.0027
1881.7495
2063.0507
2065.9962
2240.1451
2494.2283
2509.2094
2616.4129
2620.3487
2807.3942

Spot 116
881.9716
1019.4710
1136.5740
1173.6314
1179.6582
1224.6333
1413.6297
1430.6175
1512.7819
1558.7109
1592.7386
1821.9967
1932.9505
1972.9948
2086.1297
2101.1311
2207.1540
2284.1496
2363.2628
2525.8648
2571.2865
2576.2370
2587.2859
2807.3630

Spot 117
856.5818
906.5262
969.5696
1033.5699
1045.5607
1122.5478
1138.4455
1180.5706
1222.5389
1238.4518
1308.6466
1390.6724
1437.7221
1453.6457
1518.8039
1595.7636
1696.8676
1699.8117
1712.9466
1905.2081
2285.0933
2572.1932
2807.3466

Spot 118
877.0588
1379.7077
1396.7199
1443.8600
1446.7187
1525.7383
1664.8989
1668.6160
1701.9190
1718.8604
1774.9700
1880.9375
2040.2090
2231.2839
2298.1278
2409.2686
2490.2435
2678.2153
2807.3596

Spot 119
847.3174
922.3970
942.3566
1106.5453
1127.5123
1144.5927
1333.7580
1344.6708
1446.7786
1551.5389
1708.8050
1763.8375
1994.0457
2214.0365
2239.1606
2284.1662
2447.4458
2606.7610
2777.2861
2807.3696
2828.4366

Spot 120
865.4275
1186.6693
1189.5970
1237.6073
1321.6927
1347.6031
1593.8344
1867.0761
1887.8289
2214.0506
2386.1796
2614.6444
2807.4373

Spot 121
861.1352
877.0936
952.4722
1197.6654
1204.6610
1389.7536
1461.6118
1560.7492
1569.8191
1609.8137
1781.3263
1913.0522
1929.0326
1947.9026
2106.9586
2261.1575
2283.1747
2453.5530
2807.3575
2811.4570

Spot 122
861.1673
871.3455
883.3741
1045.6199
1139.5343
1151.3794
1179.6440
1339.5894
1475.7750
1657.8508
1717.8343
1733.7992
1797.0015
1894.0731
2018.0988
2103.1309
2225.0905
2239.0546
2300.2065
2537.7298
2807.3018

Spot 123
856.5761
861.2135
898.3428
906.5296
1033.5563
1067.4750
1122.5343
1175.3052
1180.5357
1217.6007
1289.2640
1437.7104
1453.7294
1518.8114
1595.8031
1696.8635
1699.7986
1859.4209
2052.0036
2219.1928
2239.1544
2284.1691
2372.1127
2635.6047
2807.2933

Spot 124
861.1410
1013.4428
1139.5268
1318.5365
1467.6800
1568.7606
1657.8460
1761.8544
1771.2133
1796.9447
1963.9462
2225.1043
2231.2802
2239.1092
2284.1301
2408.0548
2609.6581
2807.3357
2810.2709

Spot 125
877.2238
891.4060
958.2684
1036.4276
1045.5870
1066.3905
1072.5980
1204.4897
1244.5944
1381.3509
1466.1916
1568.6575
1759.7904
1862.8471
1940.9096
2096.0172
2193.9511
2221.1638
2239.1567
2283.1137
2390.2584
2551.6376
2564.2342
2808.3409
2847.5803

Spot 126
Lost

Spot 127
917.4935
1159.4926
1175.4710
1255.6556
1287.5967
1319.6755
1343.6785
1407.6154
1535.7482
1552.6736
1846.8517
2101.0945
2333.1348
2349.1609
2808.3968

Spot 128
861.1076
950.2158
1045.5295
1058.4986
1082.3173
1090.5218
1245.5295
1397.6066
1440.8159
1448.7338
1621.7765
1800.8555
1837.9882
2020.7978
2033.0676
2194.6551
2225.0890
2239.1846
2284.1766
2300.1545
2426.4038
2577.0531
2679.2436
2753.7985
2807.3546

Spot 129
1057.4804
1098.5420
1247.6381
1250.6517
1403.7180
1444.8489
1460.7825
1463.7306
1655.1156
1699.7839
1723.9969
1727.8398
1849.9665
1940.9139
2225.0849

2239.1178
2281.1286
2606.9437
2807.3105

Spot 130
Lost

Spot 131
861.1316
879.4777
1045.5412
1207.5277
1451.7324
1503.6594
1519.6882
1522.6448
1680.0890
1732.8156
1874.2323
1994.9916
2037.1930
2239.2806
2344.1509
2352.1611
2402.2183
2807.3885

Spot 132
948.4411
1051.5199
1141.5286
1188.5711
1263.7519
1390.7709
1568.7624
1590.6996
1742.0205
1794.6673
1940.9357
1994.0041
2069.0399
2225.0847
2239.1278
2284.1521
2298.2209
2405.0310
2421.0738
2781.3848
2807.3396

Spot 133
1023.4775
1132.5837
1379.7005
1396.7271
1418.7662
1443.8922
1527.8829
1664.9368
1701.9308
1705.7886
1718.8686
1774.9109
2040.1563
2409.2415
2412.1875
2807.3197

Spot 134
960.4900
1114.9328
1164.1173
1432.8713
1448.8354
1464.6540
1508.4027
1511.1402
1712.5244
1746.5257
1749.2106
1762.0927
1813.5155
1905.9187
2056.6879
2237.4109
2390.5337
2394.0046
2421.0621
2748.9314

Spot 135
861.1616
900.3767
926.4347
1017.5099
1102.4414
1236.4465
1309.7118
1366.2521
1396.7326
1428.8557
1438.7633
1560.0718
1664.9020
1733.7057
1779.9996
1833.2568
1906.1998
1989.1087
2086.1324
2115.2932
2239.1357
2273.0095
2285.2007
2462.5628
2537.1929
2743.1644
2807.3686
2810.3096

Spot 136
932.3689
981.4681
1079.5765
1170.4774
1194.5181
1411.8242
1513.7407
1531.7873
1661.8068
1742.6344
1885.9725
1904.2386
1933.8738
2284.1495
2615.8879
2807.4478

Spot 137

Spot 138
980.3368
1036.4792
1196.2670
1341.6126
1424.1857
1578.9422
1581.8449
1595.8308
1611.7494
1743.9378
1746.8808
1849.9554
1852.9136
2239.1787
2314.0298
2564.1139
2776.7991
2807.3870
2810.3320

Spot 139
855.1166
1060.1680
1076.5378
1146.6081
1274.6709
1341.6612
1344.6186
1481.7338
1631.8536
1857.6858
2094.2651
2283.1401
2444.8631
2641.4423
2810.2689

Spot 140
861.2108
1076.5594
1341.6604

1363.6905
1631.8457
1647.8426
1848.8265
1853.8914
2059.0404
2165.9425
2631.1486
2807.3398
2812.3755
2846.7399

Spot 141
1076.5217
1125.6948
1142.5490
1285.6226
1341.6459
1365.6868
1632.8019
1692.7707
1812.0270
1941.0139
2048.9640
2225.1681
2239.1989
2283.2089
2408.2380
2456.7421
2628.6289
2747.2544
2807.4021
2813.4723

Spot 142
846.3486
990.3833
1071.4103
1125.6847
1146.5956
1252.6644
1300.7034
1341.6414
1390.7153
1462.6754
1502.7248
1590.7051
1692.8753
1753.8345
1863.8240
2123.0527
2239.1983
2283.1643
2298.2946
2466.1778
2635.0768
2793.1471
2807.3441
2814.2527

Spot 143
891.3429
912.3681
1045.5573
1197.6858
1213.5466
1319.7536
1335.7216
1379.7046
1403.7984
1664.8976
1701.9297
1855.0671
1861.0105
1880.0713
2004.3557
2237.1258
2253.1693
2370.4672
2409.2174
2582.0691
2743.4224
2807.4198

Spot 144
919.5152
994.4893
1071.6147
1285.7103
1288.6663
1461.7252
1500.7528
1528.7705
1665.7607
1692.8328
1718.9010
1801.9382
1880.0894
1962.0639
2051.1080
2283.1067
2440.2316
2711.0332
2807.4847

Spot 145
861.1525
884.3008
946.3356
1023.4688
1128.3856
1153.4765
1312.5725
1379.6827
1396.7209
1401.7297
1418.7605
1443.9155
1664.9801
1701.9694
1718.9110
1769.5306
2040.2092
2223.1167
2409.3377
2412.2808
2506.2511
2808.4618

Spot 146
984.2869
1073.6825
1175.6757
1293.7227
1476.8338
1664.8801
1707.8842
1815.0059
1854.9931
1898.9908
2009.9013
2161.1090
2284.1845
2616.2833
2807.4270
2810.3820

Spot 147
906.4849
969.5411
1033.5722
1045.6121
1122.5153
1294.5416
1308.6152
1378.7101
1437.7051
1494.6357
1513.7869
1595.7726
1696.8642
1699.7509
1705.7691
1941.9345
2143.7423
2225.0986
2239.1410
2446.2292
2807.4041

Spot 148
861.1257
1023.4638
1045.5694
1379.6547
1396.7228
1443.8557
1446.8129
1527.8778
1664.8967
1701.9100
1705.7891
1718.8489
2040.1410

2105.8373
2283.1249
2409.2251
2808.2976

Spot 149
982.4251
1179.3804
1195.3334
1236.4207
1366.3251
1476.3486
1514.3476
1517.2906
1542.3003
1662.3174
1717.2273
1886.1837
2102.9115
2163.0563
2448.1502
2806.7548

Spot 150
856.4474
870.4832
928.4056
1080.2071
1186.7109
1189.5650
1193.2955
1321.7137
1352.1342
1400.1590
1575.0317
1653.8434
1750.1359
1991.2697
2231.3239
2285.2216
2436.4243
2755.9879
2807.4310
2811.5170

Spot 151
Lost

Spot 152
861.1808
928.3869
1087.5161
1137.5468
1307.4023
1528.7697
1630.5567
1652.7834
1897.9356
2081.9232
2239.1377
2461.9208
2790.7939
2807.3446

Spot 153
943.5509
967.5647
1040.5031
1128.4821
1137.5644
1284.7806
1435.3810
1517.8116
1652.8183
1656.7067
1824.0213
2008.7211
2120.0661
2239.1441
2496.5388
2644.7124
2808.3369

Spot 154
877.1681
896.2444
1006.2610
1137.5348
1247.5397
1253.5261
1517.7880
1652.8181
1816.1227
2120.1137
2125.4390
2283.1409
2291.3737
2357.0479
2666.3353
2816.2769
2807.4475

Spot 155
861.1088
920.3020
1051.3261
1063.5374
1395.6688
1428.8389
1433.7651
1600.8696
1815.0703
1875.9874
1959.9425
1984.1350
2041.0453
2187.2541
2225.1319
2283.1873
2504.0766
2744.8547
2807.3515

Spot 156
861.0641
864.3722
944.4657
1066.1439
1132.5114
1222.6199
1276.9249
1450.5339
1481.8271
1593.7967
1641.7659
1361.9969
1676.9752
1794.5787
1815.7518
1903.6235
2002.9624
2077.1325
2088.0333
2213.9589
2234.1903
2477.9175
2694.7743
2764.9364
2808.4211

Spot 157
Lost

Spot 158
987.5347
1076.5886
1146.6206
1214.7252
1220.6396
1228.6730
1234.5610
1331.6613
1376.7608
1401.7849
1404.6839
1473.7221
1566.8304
1569.8200
1659.8675
1718.7769

Spot 159
1087.5619
1220.8449
1228.8629
1377.0485
1402.0821
1474.0777
1502.1012
1567.2406
1666.2453
1720.2388
1724.1858

1958.7484
2209.9143
2513.5871

Spot 160
861.1651
855.3069
917.3690
1032.4851
1048.5256
1195.9341
1237.4179
1294.5169
1318.6269
1323.6775
1349.7545
1455.8482
1582.6262
1732.6804
1795.8457
1968.9162
2131.0317
2225.0520
2300.0491
2325.9369
2451.9880
2469.0349
2511.7928
2605.5245
2689.1190
2807.3535
2810.1755

Spot 161
928.4466
1045.5720
1189.7928
1282.7323
1438.7404
1448.7130
1569.7683
1607.7485
1784.6644
1918.0201
1925.7849
1940.9236
2083.0119
2284.1248
2467.3386
2471.31898
2648.5279
2808.3076

Spot 162
928.2775
969.4153
1045.4804
1214.6345
1228.5605
1258.6250
1376.7477
1401.7357
1434.7531
1474.6388
1566.7781
1623.6993
1659.8392
1719.7358
1788.7648
1994.0241
2070.9991
2214.0185
2239.0973
2248.2478
2302.1743
2500.3618
2565.4619
2715.6535
2731.3098
2808.3436

Spot 163
987.5107
1273.6728
1285.5982
1309.7109
1327.6605
1343.6546
1512.7059
1657.7782
1744.8645
1932.0154
1940.9913
1965.9824
2003.8648
2239.0969
2284.1610
2299.1065
2431.3152
2531.3152
2638.3205
2725.3924
2807.3557

Spot 164
862.2108
1045.5933
1058.1426
1205.6140
1263.4304
1447.6977
1578.9261
1595.7324
1628.7596
1743.8988
1788.8132
1924.1545
1940.9125
1986.9773
2141.4652
2222.0369
2225.1419
2230.2430
2239.1705
2283.1791
2413.2847
2565.2295
2721.2806
2781.4695

Spot 165
861.2674
917.2857
1004.3324
1045.5698
1155.7029
1179.7276
1309.7506
1341.7031
1456.6890
1484.9154
1649.7894
1654.9252
1667.6735
1673.8618
1940.9842
1994.0919
2201.3449
2215.1872
2225.1656
2239.2140
2283.2268
2420.2747
2645.1759
2679.2944
2809.2977

Spot 166
870.3774
1034.4184
1045.6165
1271.2121
1329.6560
1368.6197
1422.6901
1493.8384
1695.9561
1859.9450
1869.9930
1919.0067
1940.9383
1974.9614
2159.9849
2225.1694
2239.1430
2254.0925
2283.1681
2298.2367
2490.1904
2642.2016
2789.2210
2808.3683

Spot 167
1045.5494
1319.2762
1485.1247
1637.8964
1934.6268
1940.9554
2083.1222
2231.2625
2225.1336
2239.1305
2283.1817
2297.1806
2409.1187
2566.2547
2791.1744

Spot 168
869.4562
1024.4606
1157.6210
1200.6836
1210.4452
1219.4696
1338.6513
1482.8005
1485.7298
1610.8865
1701.8397
1725.7896
1755.9971
1940.8517
1977.0794
2221.0173
2225.1446
2285.1607
2409.0195
2575.3658
2808.3888
2811.2999

Spot 169
963.3794
1046.5427
1130.4076
1207.6892
1307.6561
1413.8257
1460.5394
1699.8615
1862.5934
2014.9014
2135.2227
2225.1456
2239.1079
2297.2251
2410.2319
2565.3420
2676.3178
2747.2662
2807.3892
2810.3429

Spot 170
861.0943
904.2795
1046.5792
1220.6044
1308.5981
1330.6313
1510.8639
1581.8508
1627.8237
1791.7625
1941.9020
2092.1303
2233.1185
2239.1688
2283.2251
2287.1609
2435.5260
2496.2401
2561.2726
2786.8451
2807.3960
2810.3376
2829.5032

Spot 171
847.4679
891.4196
1045.5857
1139.4664
1314.6067
1352.6624
1405.6973
1566.6961
1569.7827
1607.6603
1736.1377
1911.8921
1918.0249
1966.5910
2225.1047
2370.2745
2552.7611
2808.4043

Spot 172
917.3239
958.3712
1076.5688
1216.4188
1278.7524
1333.7423
1399.6887
1430.8202
1433.7002
1441.9123
1525.8324
1680.9407
1893.9433
1940.9548
2231.2506
2234.0788
2239.1543
2271.9936
2283.2113
2497.1430
2663.3974
2808.3365

Spot 173
1023.4149
1045.5241
1266.5816
1341.7812
1360.7265
1376.6947
1396.7198
1443.8737
1486.7513
1492.7890
1502.7044
1664.9446
1701.9202
1705.7591
1718.8696
1893.0326
1912.9364
2004.0134
2044.0893
2269.2325
2272.0844
2346.0499
2409.3130
2564.2705
2792.3606
2807.3662

Spot 174
966.4594
1042.5482
1024.4286
1243.6553
1301.6904
1446.8491
1598.7229
1608.7769
1774.8973
1790.9038
1867.8670
1985.9934
2023.0342
2055.9979
2150.1202
2163.2452
2216.0096
2225.1272
2231.0634
2259.0883
2306.2258
2467.0283
2484.2577

2634.8580
2807.3704

Spot 175
894.4858
966.4958
972.6127
982.5667
1025.6480
1129.6329
1298.7578
1539.7521
1720.8437
2285.2139
2545.3800
2809.3399

Spot 176
894.4606
966.4831
982.5474
1025.6193
1036.5441
1298.7455
1539.7173
1720.8515
2178.1715
2283.2188
2544.3626
2810.3419

Spot 177
966.5254
972.6016
982.5849
1115.3833
1129.6982
1298.7179
1357.5760
2539.7988
1720.7191
1874.9519
2065.6086
2231.3099
2505.2991
2524.1369
2726.4553
2808.3017

Spot 178
830.4200
862.3587
966.4927
982.5596
1036.4279
1060.5092
1067.5942
1179.6890
1219.2546
1277.7902
1337.8582
1428.1008
1475.7977
1550.9146
1572.1023
1716.9942
1791.8388
1843.7586
1928.9635
1994.2019
2117.5192
2298.0028
2381.1517
2518.9724
2811.7879

Spot 179
928.5414
966.4429
972.5548
982.5187
1025.6041
1162.6483
1539.7533
1720.8155
1767.8524
2179.1841
2351.1615
2544.4187
2807.4168

Spot 180
966.4929
972.6002
982.5534
1025.6184
1137.4762
1162.6840
1208.5084
1539.7805
1544.8227
1554.7577
1713.9199
1720.7919
2017.0120
2075.1094
2225.1240
2523.1898
2609.3126
2807.4631

Spot 181
966.4681
972.5708
982.5236
1025.5837
1162.6623
1463.0845
1539.7412
1554.6722
1720.7661
1890.9703
2075.0815
2226.0524
2523.1327
2544.3905
2805.2785

Spot 182
972.5883
979.5633
982.5618
1025.6086
1049.5533
1162.6845
1298.6724
1539.8049
2554.7358
1720.8024
2016.9861
2075.0796
2225.1325
2526.0788
2609.2838
2807.3616

Spot 183
966.4676
982.5520
1025.6477
1055.5261
1137.4798
1162.6756
1264.4368
1539.6875
1543.7357
1554.7013
1712.8195
1964.1435
1990.0363
2075.0218
2194.8873
2239.0668
2523.0386
2526.0409
2609.1300
2692.5805
2807.2467
2850.0959

Spot 184
899.5133
1013.4571
1045.4968
1055.4857
1194.6443
1200.6349
1255.5607
1271.4995
1346.6983
1356.7764
1473.7955
1654.8006

1672.1437
1915.7933
2152.1690
2220.0742
2279.1427
2455.8902
2617.3524
2792.0494
2807.3411

Spot 185
861.1705
903.4708
1051.4191
1102.6102
1259.7849
1446.8457
1472.8940
1581.9349
1676.8174
1743.6820
1852.9862
1904.0147
1937.0236
2186.3928
2225.1103
2298.1694
2493.1635
2720.0191
2807.3268

Spot 186
861.1798
878.2884
1023.5406
1038.5089
1110.6044
1233.7840
1246.5776
1262.5386
1321.6990
1446.8716
1540.8669
1563.7415
1593.7956
1723.8434
1839.0482
2060.0242
2226.1367
2299.1150
2449.6824
2710.5130
2807.3641

Spot 187
832.4033
863.3263
1045.5447
1110.5869
1233.7295
1249.6542
1307.6237
1363.7248
1550.8486
1553.7811
1777.0180
1916.9489
1997.7137
2225.1281
2690.1109
2807.3387

Spot 188
861.2050
1045.5593
1186.5424
1194.6563
1230.6491
1291.7345
1346.7463
1441.7056
1444.8077
1523.0678
1733.9455
1757.7573
1753.9790
1916.9579
2152.1739
2239.1402
2353.4547
2539.2838
2751.3884
2807.4825

Spot 189
905.3227
1005.5325
1020.5426
1105.5293
1178.6238
1175.5734
1192.5865
1230.6626
1253.6587
1278.6823
1529.1238
1603.9603
1701.7966
1705.6898
1930.0133
2293.1278
2310.2543
2384.1412
2606.1410
2637.2016
2804.3034
2807.3053

Spot 190
894.4588
939.3601
1120.4970
1233.7425
1276.7640
1425.7665
1508.8628
1511.8111
1777.0299
1831.1101
1916.9915
2129.5186
2369.1920
2574.0609
2805.2325

Spot 191
894.4405
1110.5897
1233.7529
1370.6440
1425.7655
1508.8665
1550.8741
1567.6979
1794.8670
1917.0043
1980.0244
2149.1510
2191.2345
2271.1473
2286.1310
2299.1906
2545.7653
2804.3695
2807.3365

Spot 192
Lost

Spot 193
853.4973
1045.5401
1124.6059
1162.6655
1226.5487
1348.6051
1406.8042
1557.6414
1790.4561
2004.1979
2225.1177
2239.1498
2298.1279
2609.3155
2613.5243
2808.4030

Spot 194
856.4009
878.3158
926.3230
972.3679
1207.6499
1246.5998

1269.7016
1446.8552
1591.7876
1663.8530
1680.6991
1754.9376
1758.6904
1804.0600
1811.8679
1844.0758
2040.5075
2221.9921
2225.0960
2245.9935
2278.1020
2500.7198
2722.6229
2808.2557

Spot 195
870.4567
997.4866
1042.3342
1045.5721
1169.5689
1259.8414
1269.6533
1399.6603
1446.9112
1531.6062
1619.7500
1663.7943
1696.7241
1803.9956
1806.9517
1811.8378
1941.8808
2176.2938
2225.1434
2239.1853
2244.0969
2443.9870
2750.1027
2807.3713

Spot 196
861.0863
877.0680
885.3124
1066.0831
1093.2466
1173.7304
1338.5169
1554.0112
1663.9053
1735.8503
1803.9832
1811.8783
1918.7002
1994.0883
2196.7214
2225.0826
2244.0891
2285.0872
2436.1253
2699.2678
2808.3740

Spot 197
927.5663
970.5565
1241.6779
1253.7685
1287.6032
1291.5767
1369.8194
1445.9854
1632.9965
1804.0538
1948.7635
2134.9202
2225.1085
2239.0940
2284.2629
2302.9220
2562.2416
2741.4020
2781.8805
2807.3967

Spot 198
1040.4305
1056.4665
1086.671
1103.7226
1242.6947
1359.4990
1369.6247
1372.5761
1530.8948
1558.6759
1590.6454
1601.7995
1719.9297
1937.9733
1988.0130
2066.0588
2215.0252
2225.1372
2239.1488
2331.1978
2408.1804
2564.1877
2568.7931
2759.4775
2808.3236

Spot 199
882.2428
1045.5290
1166.6198
1210.6419
1226.6109
1429.6742
1615.7809
1845.8855
1941.9437
2109.7469
2225.1327
2274.3018
2298.1940
2444.8437
2619.6721
2720.3379
2808.3647
2828.6379

Spot 200
861.2053
945.3925
1027.5136
1045.5785
1199.1442
1362.6622
1518.0520
1569.6701
1779.9599
1939.8822
2161.8400
2222.1230
2225.1202
2239.1363
2408.1587
2564.1893
2615.2317
2758.9829
2807.4622
2810.3103
2835.2999

Spot 201
1044.4653
1110.5873
1179.6468
1233.7507
1307.7220
1363.7899
1425.7312
1475.8010
1550.9188
1632.7453
1716.7454
1791.7815
1853.8983
1917.0300
1994.1408
2082.1920
2225.1809
2285.1639
2297.2300
2526.9576
2706.1150

2807.4780

Spot 202
894.4160
1051.4705
1132.5554
1233.7263
1337.7724
1425.7655
1508.8732
1511.7708
1703.9629
1916.9850
1940.9445
1982.7316
2138.1403
2225.1283
2293.1239
2301.2114
2553.0651
2807.3533
2812.2318

Spot 203
861.2115
978.4902
1132.5794
1267.5650
1283.5289
1396.7336
1443.8919
1525.7838
1532.7676
1664.9489
1701.9494
1704.8669
1718.9440
1774.9667
1823.9415
2225.1590
2330.0824
2409.3522
2807.4161

Spot 204
862.3224
939.5475
1024.5304
1132.6323
1267.6280
1169.5593
1360.7737
1370.7276
1379.6754
1396.7743
1443.9426
1623.1074
1664.9892
1701.9681
1704.8061
1718.9115
1774.9577
1823.9797
1969.2034
2225.1327
2285.1173
2330.1213
2345.9526
2409.3289
2536.4692
2807.3910
2810.3021

Spot 205
862.3286
949.5259
1042.5230
1081.5156
1146.4965
1230.6103
1235.7054
1267.6143
1292.6481
1396.7253
1444.8002
1525.7767
1606.8201
1609.7400
1701.8867
1758.9109
1776.9084
1870.9692
1907.9974
1919.9841
2191.3114
2225.1001
2233.0094
2296.1977
2330.0821
2575.2553
2593.1028
2807.3886
2812.5041

Spot 206
890.4355
1025.3962
1046.5664
1139.5879
1143.5497
1265.6673
1303.7178
1489.7448
1505.6857
1658.8361
1786.8504
1814.1244
1959.9085
1962.8597
2067.9877
2086.9544
2239.1681
2286.1202
2320.2045
2442.0625
2487.2606
2655.2356
2761.7099
2808.3719

Spot 207
878.1894
1036.4701
1080.4913
1193.4128
1201.4977
1475.7555
1656.7849
1659.6982
1702.8845
1918.1567
2074.5150
2239.1916
2264.0949
2409.1633
2610.8828
2807.5000
2810.3572

Spot 208
870.4075
929.5374
1003.4887
1134.5213
1217.6203
1233.5456
1396.6443
1493.7411
1655.8334
1903.8863
2022.9413
2204.0216
2367.8415
2564.1345
2807.5273

Spot 209
917.4770
979.4859
1146.4707
1255.6376
1287.5762
1319.6381
1343.6599
1407.6188
1498.7705
1535.8295
1552.6948
1693.8908
1846.8313
1866.9252
2032.0156
2101.1197
2283.1589

2333.1156
2807.3656

Spot 210
861.1406
950.4834
984.4715
1013.4189
1053.5769
1219.7891
1222.3347
1231.6360
1417.6273
1613.8268
1630.8076
1649.5562
1724.8840
1913.8248
1940.9498
2123.7980
2284.1168
2289.4715
2469.1120
2679.4007
2807.3951
2843.5926

Spot 211
862.0753
926.3137
1054.4480
1080.4193
1154.3019
1228.4678
1342.4154
1614.3255
1631.3039
1679.2333
1730.5781
1801.9735
1986.7920
2225.1211
2239.0957
2242.0477
2299.0950
2409.1462
2603.8658
2757.2073
2807.8159
2851.5814

Values of peptide masses ($M+H^+$) entered for searching the NCBInr and *T. brucei* databases with the MASCOT® search engine. The underline masses correspond to the peptides matching onto a protein in the database.

Sample	Protein	Accession	Score	P	ppm	Coverage	Peptide	MW	pI	Modification
86	beta tubulin	135500	139	72	25	27%	11	50413	4.72	1
87	HSP60	18277736	56	72	36	29%	13	59777	5.31	2
88										
89	Blast: HSP 60	Tryp10.0.0001723-6	69	56	17	21%	9	59751	5.31	1
90	Blast: HSP 70	92.m00242	112	56	17	21%	11	72000	5.77	0
91										
92	Blast: HSP 70	92.m00242	102	56	25	16%	10	72000	5.77	
93	Tubulin alpha chain	135440	74	72	34	33%	11	50383	4.93	2
94										
95	Hypothetical protein	68.m00229	51	55	51	13%	5	48380	5.39	0
96	Paraflagellar rod protein	129883	54	72	25	12%	6	69953	5.74	1
97	Paraflagellar rod protein	129883	65	72	27	17%	7	69953	5.74	1
98										
99										
100										
101	Blast: dihydrolipoamide S acetyltransferase	Tryp10.0.000241-7	77	56	13	20%	7	45437	6.06	0
102	Blast: chaperonin (Mice)	CONTIG 11912			50	21.2	8			1
103										
104	Hypothetical	Tryp10.0.000709-2	56	55	23	34%	5	23349	6.74	0
105	Blast : HSP 70	Tryp11.0.001503-2	79	56	21	14%	9	71649	5.3	1
106	Blast: Paraflagellar rod protein	Tryp10.0.000423-2	53	55	25	15%	7	52090	5.87	0
107										
108										
109	Actin B	113244	46	72	28	19%	5	42140	5.3	none
110										
111										
112	Putative phospatidylinositol 4 phosphate 5 kinase	43.m00211	63	50	20	19%	6	41003	5.29	none
113										
114	beta tubulin	9366734	155	72	16	42%	12	39170	5.3	2
115	Blast: elongation factor 2	Tryp10.0.000576-168	90	56	30	17%	12	82923	5.52	3

116										
117	Blast : Elobgation factor 2	Tryp10.0.000567-168	120	56	20	15%	12	82923	5.52	0
118	Tubulin alpha chain	Tryp10.0.000280-2	73	56	15	23%	7	50383	4.93	0
119										
120										
121	Dihydrolipoamide dehydrogenase	416906	42	72	9	15%	6	50815	5.94	1
122										
123	Blast: Elongation factor 2	Tryp10.0.000021-2	66	55	16	10%	9	95300	5.83	0
124	Blast : arginine kinase (T.brucei)	CONTIG 11896			25	14.50%	5			0
125										
126										
127	guanine nucleotide binding	6226620	153	72	17	42%	12	35181	5.98	2
128										
129	pyridoxine/pyridoxal kinase	2459793	86	72	14	26%	8	33596	6.21	1
130										
131	Blast: Cytochrome C oxidase subunit IV	Tb927.1.4100	67	56	20	21%	7	40680	5.53	1
132	Blast: HSP 83	Tryp10.0.000967-3	66	56	18	24%	7	36058	5.09	2
133	Tubulin alpha chain	135440	93	72	22	24%	8	50383	4.93	0
134										
135										
136	Enolase	8132069	75	72	22	22%	7	47133	5.93	2
137										
138	Mitochondrial chaperonin 60	TRYP_x-280c07.q1c_1	70	55	28	43%	5	17183	4.75	1
139	beta tubulin	9366734	51	72	12	13%	5	39170	5.3	0
140	beta tubulin	Trp10.0.000280-4	59	56	11	18%	6	50413	4.72	1
141										
142										
143										
144	Blast: Spermidine synthase	Tryp9.0.000477.355	54	55	50	20%	5	33274	5.03	0
145	Tubulin alpha chain	Tryp10.0.000280-2	65	56	25	20%	7	50383	4.93	1
146										

147	Blast: Elongation factor 2	tryp10.0.000021.2	62	55	20	9%	8	95300	5.83	0
148	Tubulin alpha chain	135440	58	72	13	18%	6	50383	4.93	0
149										
150										
151										
152	Blast: elongation factor 1	CONTIG 11472			50	22.70%	6			0
153	Blast: Elongation factor 1 beta	91.m00148	76	56	22	28%	6	28614	5.52	0
154	Blast: Elongation factor 1 beta	91.m00148	64	56	21	31%	5	28614	5.52	0
155	Proteasome subunit	12229950	76	72	24	36%	6	27274	4.9	1
156										
157										
158	Heat shock 70KD protein 4	123633	91	72	26	15%	9	71676	5.25	0
159										
160										
161										
162	Hypothetical	Tryp10.0.003348.8	49	55	50	33%	5	21086	5.3	1
163	Heat shock 70KD protein 4	123633	73	72	11	17%	9	71676	5.25	1
164										
165										
166										
167										
168	BiP/GRP78	295361	81	72	22	16%	9	71533	5.34	0
169										
170	Hypothetical	Tryp11.0.000893-4	71	56	17	36%	6	26642	5.53	2
171	Phosphoenolpyruvate carboxikinase	130765	43	72	23	14%	5	52937	8.26	0
172	Blast: Adenylate kinase	22.m00078	162	56	24	39%	8	29492	5.63	1
		75.m00344	56	56	24	25%	5	38373	9.07	0
173	ATPase beta subunit	11908130	78	72	19	24%	9	55837	5.29	1
174	Putative DnaJ protein	24m.00012	111	50	30	35%	10	45356	7.17	2
175	Blast: translation elongation factor 1 alpha	Tryp10.0.000008-1	58	56	19	11%	5	49474	9.01	0
176	Elongation factor 1 alpha	1078705	72	72	11	19%	6	44149	8.32	0

177											
178											
179	Blast: translation elongation factor 1 alpha	Tryp10.0.000008-1	67	56	22	19%	5	49474	9.01	0
180	Hypothetical	Tryp10.0.000534-4	67	56	20	22%	6	45884	9.1	0
181	Blast: translation elongation factor 1 alpha	Tryp10.0.000008-1	57	56	24	17%	5	49474	9.01	0
182	Hypothetical	Tryp10.0.000040.65	55	55	20	19%	5	45884	9.1	0
183	Hypothetical	Tryp10.0.000040.65	60	55	30	20%	6	45884	9.1	0
184	Hypothetical	72.m00219	123	50	30	22%	11	43396	9.08	1
185	Putative sucinyl CoA synthetase	64.m00253	56	50	25	25%	5	31844	8.78	0
186	Hypothetical	Tryp9.000334.59	50	55	50	18%	5	34711	7.88	2
187	Malate dehydrogenase	2605782	79	72	25	29%	7	33548	8.11	0
188										
189	Hypothetical	Tryp10.0.000031.71	60	55	20	14%	7	71892	9.46	0
190	Blast: Malate dehydrogenase	Tryp10.0.0003708.2	100	55	50	29%	7	33181	8.6	0
191	Blast: Malate dehydrogenase	Tryp10.0.0003708.2	97	55	30	34%	7	33181	8.6	0
192										
193										
194	Glyceraldehyde 3P dehydrogenase	120678	44	72	30	18%	5	39241	9.4	0
195	Glycosomal glyderaldehyde 3P	43.m00292	58	56	14	21%	6	44211	9.31	0
196										
197										
198										
199										
200										
201	Malate dehydrogenase	2605782	45	72	32	20%	5	33548	8.11	0
202	Hypothetical	Tryp10.0.003708.2	53	55	20	19%	5	33181	8.6	0
203	Tubulin alpha chain	135440	113	72	24	30%	11	50383	4.93	1
204	Tubulin alpha chain	135440	65	72	31	25%	8	50383	4.93	1
205										
206	Hypothetical	Tryp9.0.004195.12	76	55	50	48%	5	13785	6.73	0
207	possible chord containing protein	9366769	41	72	31	15%	4	35093	6.46	0

208										
209	guanine nucleotide binding	6226620	131	72	20	40%	10	35181	5.98	0
210										

Results from the MASCOT searches in the NCBI and the local *T. brucei* database for all the samples analysed. The results highlighted in blue correspond to protein spots identified in the NCBI database while the results highlighted in yellow correspond to the protein spots identified in the local *T. brucei* database. The results non-highlighted correspond to protein spot matched in a database but the score was lower than the database score, and so the protein identification falls outside the 95% confidence limits. The molecular weight and pI given in the table correspond to the expected values from the protein sequences in the database.

```
775xxx0  ----------------------------------------------------------------------------------------------------
774xxx1  MGKEKVHMNLVVVGHVDAGKSTATGHLIYKCGGIDKRTIEKFEKEAADIGKASFKYAWVL

775xxx0  ------------------------------------------------------------MITGTSQADAAILIIASAQ
774xxx1  DKLKAERERGITIDIALWKFESPKSVFTIIDAPGHRDFIKNMITGTSQADAAILIIASAQ

775xxx0  GEFEAGISKDGQTREHALLAFTLGVKQMVVCCNKMDDKTVNYGQERYDEIVKEVSAYIKK
774xxx1  GEFEAGISKDGQTREHALLAFTLGVKQMVVCCNKMDDKTVNYGQERYDEIVKEVSAYIKK

775xxx0  VGYNVEKVRFVPISGWQGDNMIEKSEKMPWYKGPTLLEALDMLEPPVRPSDKPLRLPLQD
774xxx1  VGYNVEKVRFVPISGWQGDNMIEKSEKMPWYKGPTLLEALDMLEPPVRPSDKPLRLPLQD

775xxx0  VYKIGGIGTVPVGRVETGVMKPGDVVTFAPANVTTEVKSIEMHHEQLAEATPGDNVGFNV
774xxx1  VYKIGGIGTVPVGRVETGVMKPGDVVTFAPANVTTEVKSIEMHHEQLAEATPGDNVGFNV

775xxx0  KNVSVKDIRRGNVCGNTKNDPPKEAADFTAQVIILNHPGQIGNGYAPVLDCHTSHIACKF
774xxx1  KNVSVKDIRRGNVCGNTKNDPPKEAADFTAQVIILNHPGQIGNGYAPVLDCHTSHIACKF

775xxx0  AEIESKIDRRSGKELEKAPKSIKSGDAAIVRMVPQKPMCVEVFNDYAPLGRFAVRDMRQT
774xxx1  AEIESKIDRRSGKELEKAPKSIKSGDAAIVRMVPQKPMCVEVFNDYAPLGRFAVRDMRQT

775xxx0  VAVGIIKAVTKKDGSGGKVTKAAVKASKK
774xxx1  VAVGIIKAVTKKDGSGGKVTKAAVKASKK
```

Figure 1: Protein isoform versus gene family. Alignment of the protein sequences from ORF 775 and 774 of *T. brucei* encoding elongation 1α demonstrate high homology between different genes of the same family. The highlighted sequences correspond to the peptides identified by PMF.

Sample	Score	Peptide	Score	Sequence	gi/number
386/3	74	4	28	INVYFDEATGGR	9366734
			19	SVLIDLEPGTMDSV	
			18	GHYTEGAELIDSVL	
			10	FWEVISDEHGVDPT	
386/4	204	4	42	FPGQLNSDLR	9366734
			55	INVYFDEATGGR	
			47	SVLIDLEPGTMDSV	
			61	GHYTEGAELIDSVL	
386/5	469	10	30	EQYPDR	9366734
			31	LREQYPDR	
			37	LAVNLVPFPR	
			45	FPGQLNSDLR	
			46	KLAVNLVPFPR	
			51	INVYFDEATGGR	
			70	SVLIDLEPGTMDSV	
			58	EIVCVQAGQCGNQI	
			58	GHYTEGAELIDSVL	
			45	AGPYGQIFRPDNFI	
386Mr/3	143	5	31	EQYPDR	9366734
			54	INVYFDEATGGR	
			21	SVLIDLEPGTMDSV	
			27	GHYTEGAELIDSVL	
			11	AGPYGQIFRPDNFI	
386Mr/4	142	5	30	EQYPDR	9366734
			50	INVYFDEATGGR	
			27	SVLIDLEPGTMDSV	
			21	GHYTEGAELIDSVL	
			15	AGPYGQIFRPDNFI	
386Mr/5	437	11	29	EQYPDR	9366734
			17	LREQYPDR	
			31	LAVNLVPFPR	
			45	FPGQLNSDLR	
			68	INVYFDEATGGR	
			51	SVLIDLEPGTMDSV	
			65	EIVCVQAGQCGNQI	
			55	GHYTEGAELIDSVL	
			44	AGPYGQIFRPDNFI	
			27	FWEVISDEHGVDPT	
			6	EAESCDCLQGFQIC	

Table 1: Detailed results from the MASCOT® search engine for spot 3, 4 and 5 for 386 and 386Mr lines.

Sample	Score	Peptide	Score	Sequence	gi/number
247/1	223	4	51	LAVNLPFPR	9366734
			58	FPGQLNSDLR	
			49	INVYFDEATGGR	
			65	SVLIDLEPGTMDSV	
274/2					
274/3	133	3	52	INVYFDEATGGR	9366734
			58	SVLIDLEPGTMDSV	
			24	EIVCVQAGQCGNQI	
274/4	116	2	57	INVYFDEATGGR	9366734
			58	SVLIDLEPGTMDSV	
274/5	404	7	29	LSGGVAVIR	18277736
			19	ITDALCSTR	
			79	VGGASEVEVNEK	
			69	YVNMFEAGIIDPAR	
			42	DDTVLLNGGGDVAMMK	
			81	ALDGLLQDQSLTADQR	
			86	VLENTDAAVGYDAQLDR	
247Mr/1	270	6	14	LREQYPDR	9366734
			51	LAVNLVPFPR	
			49	FPGQLNSDLR	
			43	INVYFDEATGGR	
			63	SVLIDLEPGTMDSV	
			45	EIVCVQAGQCGNQI	
247Mr/2	143	3	45	FPGQLNSDLR	9366734
			47	INVYFDEATGGR	
			51	SVLIDLEPGTMDSV	
247Mr/3	123	2	57	INVYFDEATGGR	9366734
			65	SVLIDLEPGTMDSV	
247Mr/4	99	2	53	SVLIDLEPGTMDSV	9366734
			46	GIIYTEGAELIDSVL	
247Mr/5	65	2	21	LSGGVAVIR	18277736
			44	VGGASEVEVNEK	

Table 2: Detailed results from the MASCOT® search engine for spot 1, 2, 3, 4 and 5 for 247 and 247Mr lines.

Name	Old Accession number	New Accession number
CAR protein	Tryp9.0.000322-274	Tb09.211.0120
CAR protein	Tryp9.0.000322-277	Tb09.211.0130
Arginine kinase	ORF 342	Tb09.160.4560
Arginine kinase	ORF 343	Tb09.160.4570,AK1
Arginine kinase	ORF 345	Tb09.160.1590,AK1

Table 1: Update of accession numbers given by GeneDB (http://www.genedb.org/) for the CAR protein and arginine kinase, as searched on the 11/02/2004.

What we see depends mainly on what we look for.

John Lubbock