Characterisation of proline dehydrogenase and its contribution to energy metabolism in *Trypanosoma brucei*

Nadia Lamour

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
February 2004

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Declaration

The results presented in this thesis are my own work, except where there is a statement to the contrary.

Nadia Lamour

Abstract

Both *Leishmania major* and *Trypanosoma brucei* undergo a complicated developmental cycle involving insect vectors and a mammalian host, subjecting these parasites to dramatic changes in their environment.

It has been demonstrated that with some strains of *T. brucei* procyclic forms in culture, the amino acid proline is the favoured substrate for energy production. This correlates well with the normal environment of these forms in the tsetse fly, where free proline is an abundant source of energy.

Proline utilisation requires two enzymes to convert proline to glutamate. In the first step, the oxidation of proline to pyrroline-5-carboxylate by proline dehydrogenase (PRODH) is coupled to the reduction of a cofactor, flavin adenine dinucleotide (FAD). In the second step, pyrroline-5-carboxylate is hydrolysed to give glutamic semialdehyde, which is oxidised by pyrroline-5-carboxylate dehydrogenase (P5CDH) to glutamate using the cofactor nicotinamide adenine dinucleotide (NAD^+).

Proline dehydrogenase genes have been identified in *T. brucei* and *L. major*. Both genes show significant homology to other eukaryotic proline dehydrogenase genes. Well conserved motifs, considered essential for the activity of the enzyme, are found in the predicted *T. brucei* and *L. major* protein sequences. As in other eukaryotes, the genes for the two enzymes (PRODH and P5CDH) involved in proline degradation are found in two different genomic locations, and encode two distinct polypeptides. In prokaryotes a single gene encodes both activities. A mitochondrial targeting motif present on the trypanosomatid enzymes, suggests that the mitochondrion is the subcellular localisation of this enzyme.

This study went on to analyse the role of proline dehydrogenase in the energy metabolism of *T. brucei* procyclic forms. Growth studies using defined media revealed that

procyclic forms *in vitro* can use either proline or glucose as an energy source. Interestingly, proline dehydrogenase activity is repressed in cells exposed to abundant glucose and, under these conditions, proline transport is also lower than in trypanosomes grown in low glucose concentrations.

RNA interference was used to ablate proline dehydrogenase activity. These RNA interference experiments confirmed that the gene does encode the trypanosomal PRODH. Cells lacking this enzyme grew well in the presence of glucose but were unable to use proline as an energy source. Moreover, proline transport in this line also differed from that in wild type parasites. Proline could not be replaced by any other amino acid as an energy source in trypanosomes deprived of glucose. Even glutamate, which is an intermediate in the proline catabolic pathway, did not support growth.

Acknowledgements

Firstly I would like to thank my supervisors Dr M. Barrett and Prof G. Coombs for the help, support and advice given throughout the three years of my project. Thank you to both of you who have been an inspiration to me as a scientist.

Thank you Mike for all the chats we had about the philosophy of life, and for permitting the transformation of the caterpillar I was, into a butterfly!

Thanks to Dr F. Bringaud for his advice for culturing my parasites.

Thanks to everyone in the North lab, all my friends and flatmates (past and present for their patience and their nice meals). Thanks to all of you for the fun nights out, lunches and chats we had together and for being there when I needed you.

I would especially like to thank Hubert and Aoibhinn, Richard, Dan, Sebastien, Styv and Kevin who focused my ideas into correct english and Rod for reading this manuscript.

Hamish, just thank you for being there to share with me this "fantastic" step of writing up!

Et le plus important pour moi, je voudrais remercier du plus profond de mon cœur mes parents, pour avoir compris que ma vie pendant ces trois dernières années était ma thèse. Merci d'avoir toujours eu confiance en moi et merci de m'avoir fait savoir tous les jours que quoi que je fasse, vous étiez fiers de moi.

Table of contents

Chapter 3. Identification of genes encoding proline dehydrogenase in *Leishmania major* and *Trypanosoma brucei* and bioinformatic analysis of the "predicted "proteins

Chapter 4. Regulation of expression of *T. brucei* and *L. major* proline dehydrogenase

List of figures

Chapter 1. Introduction

Chapter 2. Materials and Methods

Chapter 3. Identification of genes encoding proline dehydrogenase in *Leishmania major* and *Trypanosoma brucei* and bioinformatic analysis of the "predicted "proteins

Chapter 4. Regulation of expression of *T. brucei* and *L. major* proline dehydrogenase

Chapter 5. Cloning and expression of *T. brucei* and *L. major* proline dehydrogenases

Chapter 6 The physiological role of proline dehydrogenase in *T. brucei* determined through RNA interference

List of tables

Chapter 2

Chapter 3

Chapter 5

List of abbreviations

ADP	adenosine diphosphate
ATP	adenosine triphosphate
ASCT	acetate:succinate CoA transferase Succinyl CoA synthetase
BSF	bloodstream form
bp	base pair
CBSS	Carter's Balanced Salt Solution
cDNA	complementary DNA
°C	degree celcius
CNS	central nervous system
DCPIP	dichlorophenolindophenol
DMFO	diflofuromethylornithine
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
2-DOG	2-deoxy-D-glucose
DTT	dithiothreitol
dsRNA	double stranded RNA
EDTA	ethylenediamine tetraacetic acid
ES	expression site
FAD	flavin adenine dinucleotide
FCS	f tal calf serum
g	gram
gRNA	guide RNA
h	hour
HAT	human African trypanosomiasis
HRP	horseradish peroxidase
IPTG	isopropylthio-β-D-galactoside
kb	kilobase
kDa	kilodalton
Km	Michaelis constant
LB	Luria-Bertani medium

M	molar
Mb	megabase
µg	microgram
mg	milligram
mM	millimolar
µl	microlitre
ml	millilitre
mRNA	messenger RNA
MSF	mitochondrial import stimulating factor
MTS	mitochondrial targetting sequence
NAD^+	nicotinamide adenine dinucleotide
nM	nano metre
ORF	open reading frame
PARP	procyclic acidic repetitive protein
PCF	procyclic form
P5C	pyrroline-5-carboxylate
PCR	polymerase chain reaction
PBS	phosphate buffered saline
PEPCK	phosphoenol pyruvate carboxykinase
Pfu	*Pyrococcus furiosus*
PRODH	proline dehydrogenase
RNA	ribonucleic acid
RNAi	RNA interference
RNA pol	RNA polymerase
rpm	revolutions per minute
SHAM	salicylhydroxamic acid
SDM	semi-defined medium
SDS	sodium dodecylsulfate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SL	spliced leader
RT-PCR	reverse transcriptase PCR
s	second
SEM	standard error around the mean

Taq	*Thermus aquaticus*
Tm	melting temperature
TM	transmembrane
UV	ultraviolet
V	volt
VSG	variable surface glycoprotein
Vmax	maximal velocity
WHO	World Health Organisation
X-Gal	5-bromo-4-chloro-3-indol-β-D-galactopyranoside
ZPFM	Zimmerman's post fusion media

Demain ne sera pas comme hier, il sera nouveau et il dépendra de nous. Il est moins à découvrir qu'à inventer.

Gaston Berger (1890-1960)

Chapter 1

Introduction

1.1- Current situation of Trypanosomiasis and Leishmaniasis

1.1.1- Human African Trypanosomiasis

Human African Trypanosomiasis, or sleeping sickness, is a disease that is resurgent in sub-Saharan Africa. The name sleeping sickness is associated with the neurological consequences connected with the second stage of the disease when the parasites have invaded the central nervous system.

In 1895 Dr David Bruce correlated the presence of trypanosomes and the disease called Nagana in cattle. He also discovered the role of tsetse flies as vectors of *Trypanosoma brucei* in animals in Zululand (Smith *et al.*, 1998). Two sub-species of trypanosome are responsible for Human African Trypanosomiasis (HAT), *T. b. gambiense* and *T. b. rhodesiense*. A chronic form of the disease is caused by *T. b. gambiense* in west and central Africa and can take from months to years to result in death. A more acute form of the disease, predominantly found in East Africa, is caused by *T. b. rhodesiense* and can be lethal in only a few weeks (Welburn *et al.*, 2001).

At the beginning of the 20th century, a large incidence of HAT was reported. One million human cases with more than 250,000 deaths were reported in Uganda alone (Seed, 2001). By the middle of the twentieth century, the incidence had been reduced as a consequence of control measures including chemotherapy, the use of insect vector control strategies, and the removal of animal reservoirs (WHO, 1998). The latter part of the twentieth century witnessed a decline in application of various control measures and HAT had begun to re-emerge. Over 300,000 patients were believed to be infected in 1998 (WHO, 1998) with a possible further 60 million people at risk (Hide, 1999). 36 countries in

sub-Saharan Africa are now affected and less than 7 % of the African population is under surveillance for the infection (Figure 1.1).

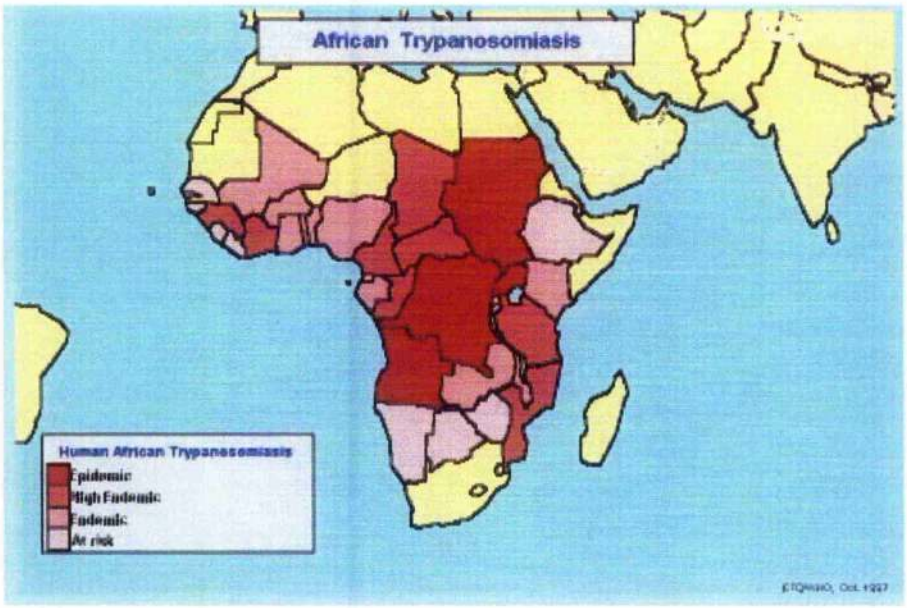


Figure 1.1: Distribution of Human African Trypanosomiasis
From http://www.who.int/emc/diseases/try

T. b. brucei, *T. congolense* and *T. vivax* cause similar diseases in cattle and other domestic animals. This has enormous economic consequences since animal trypanosomiasis is responsible for an estimated three million livestock deaths per year (Kuzoe, 1993; Hide, 1999).

1.1.2- Leishmaniasis

The Leishmaniases are a complex of diseases caused by species of the genus *Leishmania*. The genus *Leishmania*, belonging to the family Trypanosomatidae, are flagellates that occur as intracellular amastigotes in vertebrate hosts and as flagellated promastigotes in invertebrates and in *in vitro* culture at 25 °C. Leishmaniasis is endemic in areas of the tropics, subtropics and southern Europe. There are about 21 *Leishmania* species, which cause visceral leishmaniasis or Kala-azar (e.g. *L. donovani*, *L. infantum* in the old world and *L. chagasi* in the new world), cutaneous leishmaniasis (e.g. *L. major* and *L. tropica* in the old world and *L. mexicana* in the new world) and mucocutaneous leishmaniasis (e.g. *L. braziliensis*). The mode of transmission of these parasites occurs through a sandfly vector. Approximately 15 million people are infected with leishmaniasis worldwide with 350 million people at risk in over 88 endemic countries (WHO, 1998; Herwaldt, 1999) (Figure 1.2).

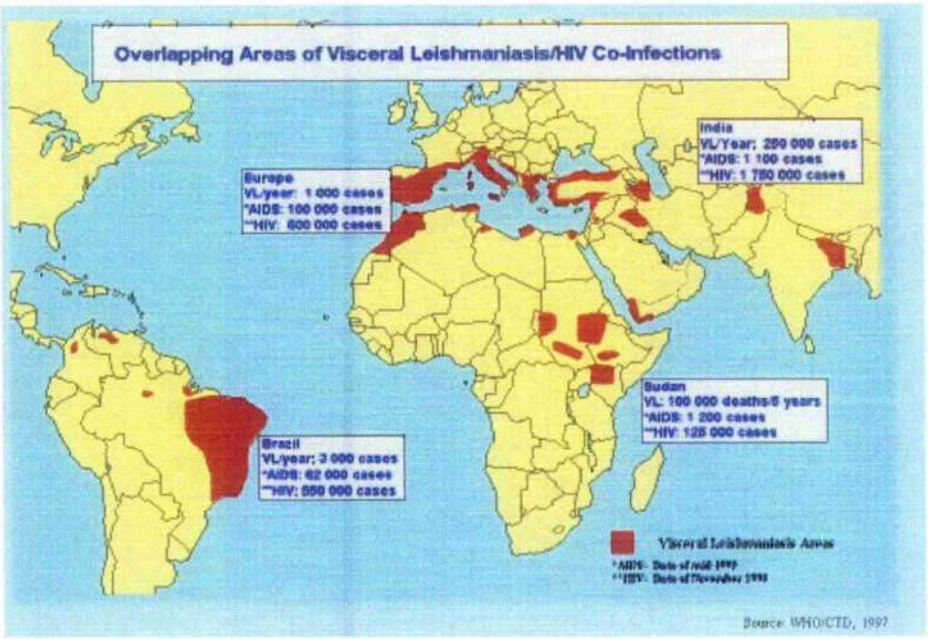


Figure 1.2: Distribution of Leishmaniasis

From http://www.who.int/emc/diseases/leish

1.2- Taxonomy

The genus *Trypanosoma* and the genus *Leishmania* both belong to the order Kinctoplastida. All Kinetoplastida possess an organelle termed the kinetoplast. It can be distinguished by light microscopy as a small, usually round or oval body, which stains similarly to a nucleus, and is situated near the base of the flagellum. Electron microscopy revealed that the kinetoplast is an unusually large mass of circular mitochondrial DNA, contained within the single mitochondrion (Shlomai, 2002). DNA is found in the mitochondria of many other organisms, but the *Kinetoplastida* have a quantity exceeding that of other cells.

Order	Suborder	Family	Genus
Kinetoplastida	Bodonina	Bodonidae	Bodo Ichtyobodo
		Cryptobiidae	Cryptobia Trypanoplasma
	Trypanosomatina	Trypanosomatidae	Trypanosoma Leishmania Endotrypanum Crithidia Blastocrithidia Leptomonas Herpetomonas Phytomonas

Figure 1.3: Taxonomy of *Kinetoplastida* (from: http://www.dbbm.fiocruz.br/www-mem/963/4213hm.html)

There are two suborders within the *Kinetoplastida*, the *Bodonina* and the *Trypanosomatina* (Figure 1.3). Many organisms of the order *Bodonina* can be distinguished from *Trypanosomatina* by having more than one flagellum (Vickerman, 1994). Among the trypanosomatids there are eight different genera, which form the family *Trypanosomatidae* and contain parasites pathogenic for humans and other mammals. The agents of sleeping sickness (*T. brucei* subspecies) and the agent of Chagas' disease (*T. cuzi*) are members of the same genus, *Trypanosoma.*

The genus *Trypanosoma* is subdivided into two groups

- *Salivaria* (where parasites develop into mammal-infectious forms in the salivary glands of the insect vector and are transmitted in saliva) including *T. b. gambiense* and *T. b. rhodesiense* responsible for the sleeping sickness in humans and *T. b. brucei*, *T. vivax*, *T. congolense*, and *T. evansi* which are responsible for different diseases in animals.
- *Stercoraria* (where parasites develop into mammal-infectious forms in the hindgut of their insect vector and are transmitted in faeces). *T. cruzi* is the principal parasite of this group and causes Chagas' disease.

There are more than 24 known species of trypanosome and evolutionary studies based on molecular biology support the idea of a common ancestor for all trypanosomes (monophyly) (Stevens and Gibson, 1999; Momen, 2001).

T. brucei includes subspecies that infect humans (*T. b. rhodesiense* and *T. b. gambiense*) and those that infect animals only, *T. b. brucei*. These three species cannot be morphologically distinguished (Vickerman, 1969), but are different at the molecular level; they can be distinguished by isoenzyme analysis (Gibson, 2002) and using selective DNA probes.

The divergence of the salivarian clade can be related to the isolation of Africa from the other continents 100 millions year ago. The *T. brucei* clade consists only of "African

mammalian tsetse transmitted species" and is unrelated to amphibian and reptile trypanosomes.

Trypanosomatids also include the genus *Leishmania*, which are responsible for human diseases such as visceral, cutaneous, mucocutaneous, and diffuse cutaneous leishmaniasis. The separation of the African continent from the American continent can also explain the divergence of old and new word *Leishmania* as well as *T. brucei* from *T. cruzi* (Stevens and Gibson, 1999).

1.3- Life cycles

1.3.1- *T. brucei*

Classical descriptions of the *T. brucei* life cycle refer to distinct morphological stages identified by shape, the position of the kinetoplast in relation to the nucleus, and the extent of the flagellar apparatus. The most striking changes occur in the mitochondrial system and in the surface membrane of the parasite (Vickerman, 1969) (Figure 1.4).

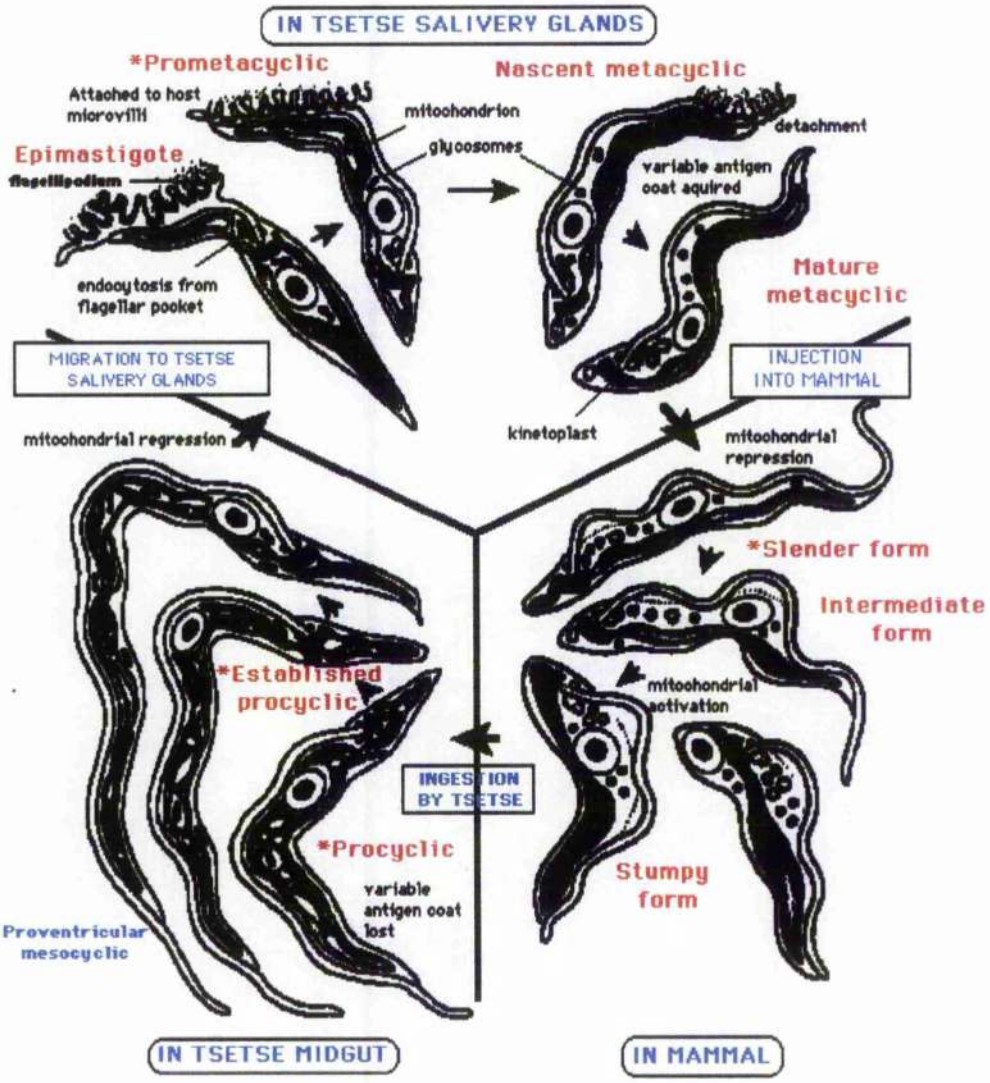


Figure 1.4: Life cycle of *Trypanosoma brucei*
From Vickerman, 1985

In the insect, the trypanosomes undergo a complex developmental cycle. Parasites ingested in blood from an infected mammal develop first in the insect midgut. They

elongate into procyclic forms (PCF) and multiply by binary fission. Procyclic development is reportedly associated with a switch from the utilisation of glucose to the utilisation of proline as the principal energy source (Vickerman, 1985). In flies, parasites invade the salivary glands. Once in the salivary glands, they transform into epimastigotes and continue to divide by binary fission. In the midgut, the parasites cover their surface with a coat of procyclin (also named PARP for Procyclic Acidic Repetitive Proteins (Roditi *et al.*, 1998), which represent the products of a family of polymorphic genes (Mowatt and Clayton, 1988). The variable surface glycoprotein, VSG, is lost. In procyclic forms, the mitochondrion is more extensively developed than in the mammalian stages (Vickerman, 1985).

Two to three weeks after the infective feed, the epimastigotes become free in the lumen of the salivary gland and undergo transformation into metacyclic trypomastigotes. These infective forms are injected with the fly saliva into the next mammal on which it feeds (Neva and Brown, 1994). The biting tsetse fly deposits metacyclic trypanosomes in the dermal connective tissue of the mammalian host, which leads to the development of the chancre. The metacyclic trypomastigotes change into the more elongated bloodstream forms (BSF) (long slender forms) and divide by binary fission with a doubling time of about 6 hours in the bloodstream (Vickerman, 1985). These forms undergo antigenic variation (Turner, 1999). This phenomenon involves the variation of the surface coat VSG, in order to avoid the immune system of the host, resulting as a fluctuation of the parasitemia in the bloodstream (Barry and McCulloch, 2001). The parasites do not enter host cells but are at first restricted to the subcutaneous tissue fluid around the site of the fly bite. During the next few days the parasites spread throughout the host's lymphatic and blood systems. Later, the trypanosomes penetrate the central nervous system (CNS). Here, too, they remain extracellular.

T. brucei in the bloodstream are pleomorphic. The trypanosomes which divide in the mammalian host are long slender individuals, with a long flagellum that extends beyond the anterior end of the cell. After a few days, shorter, more stumpy forms develop, with the flagellum extending only a little beyond the anterior end of the cell. These stumpy forms, which do not divide in the mammalian host, are those destined to continue the life cycle after being ingested by a tsetse fly, in which they elongate into the PCF, the long slender forms being destroyed in the midgut of the fly (Matthews, 1999). Stumpy forms appear to be biochemically pre-adapted to life in the fly.

1.3.2- *Leishmania*

The life cycle of *Leishmania* species alternates between a vertebrate and an insect host (Figure 1.5). The natural reservoir hosts, besides humans, include the domestic dog and a variety of wild mammals. The parasites exist in two principal forms: amastigotes in their mammalian hosts, and promastigotes in their insect vectors. The vectors are small Dipteran insects commonly known as sandflies of the genus *Phlebotomus* (Old World) and *Lutzomyia* (New World Leishmaniasis). Only the females feed on blood.

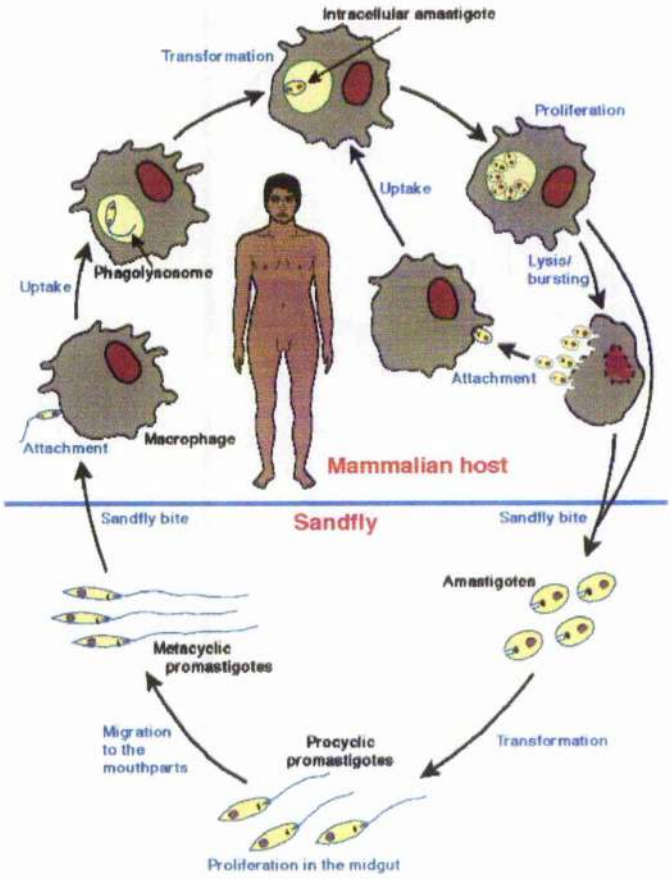


Figure 1.5: Life cycle of *Leishmania*
From http://www.wehi.edu.au/media/images/leishmania_cycle

Amastigotes are avoid forms, with virtually no flagellum. These are ingested by macrophages as part of the phagocytic activity. Instead of being destroyed, the protozoa apparently resist the lysosomal enzymes released into the phagolysosome; they then multiply by binary fission within the macrophages (Bates and Tetley, 1993). When an infected macrophage dies, the liberated protozoa are ingested by other macrophages.

Infected macrophages in the blood or skin are ingested by phlebotomine sandflies. In the midgut of the vector, the protozoa emerge from the macrophages and transform into promastigotes (Bates and Tetley, 1993). Promastigotes are elongated, flagellated forms with the kinetoplast and flagellar basal body near the anterior end. No sexual cycle is known, both forms reproducing by binary fission, though evidence suggestive of genetic exchange has been reported (Neva and Brown, 1994). Approximately ten days after first ingesting the protozoa, when the insect again attempts to feed, many metacyclic promastigotes are injected into the host's skin. Here they are phagocytosed by macrophages and, after reverting to the amastigote form; they commence dividing (Sacks, 1989).

1.4- Disease

1.4.1- African Trypanosomiasis

Human sleeping sickness is an extremely debilitating disease and is characterized by two distinct phases: early and late. Following the Tsetse fly bite, trypanosomes proliferate in the host bloodstream and undergo antigenic variation to evade the immune system.

In the early stage of the disease, after development of the chancre, infection of the blood and lymph systems results in a more or less acute febrile illness. Symptoms of this early phase (nausea, fever and lethargy) are non-specific. More serious effects result from the penetration of the parasites into the CNS. In the late phase, trypanosomes cross the blood-brain barrier and can be found in neural tissue and cerebrospinal fluid (Enanga *et al.*, 2002). Subsequent neural damage and host reactions cause the classical symptoms of sleeping sickness: disruptions of biological rhythms, inappropriate and irregular sleep

patterns, and loss of concentration and coordination. The late stage is reached within weeks with *T. b. rhodesiense* or years with *T. b. gambiense*, after the initial infection (Smith *et al.*, 1998). Once in the brain, the outcome of the inflammatory process (meningoencephalitis) is brain damage leading to somnolence, coma, and unless treated, death in almost all cases.

1.4.2- Leishmaniasis

As implied by the term "visceral leishmaniasis", macrophages infected with *L. donovani*, *L. infantum* and *L. chagasi* congregate in the viscera, notably the spleen and liver. These organs become hugely enlarged, and their functions are progressively impeded. Unless it is treated, the disease is invariably fatal (Herwaldt, 1999). *L. tropica*, *L. major* and *L. mexicana* cause only one or a few lesions at the site of the infected bite on the skin of their mammalian hosts. In humans the disease is limited to the cutaneous tissues and occasionally to the mucous membranes (Weigle and Saravia, 1996). *L. brasiliensis* is capable of causing mucocutaneous lesions in humans (Muller and Baker, 1990).

1.5- Treatment

1.5.1- African trypanosomiasis

In the early stage of African sleeping sickness, before invasion of the central nervous system, the disease is usually curable, the most commonly used drugs are Suramin or Pentamidine. In the late stage, treatment is less successful and is based on the rather toxic organic arsenical compound Melarsoprol or Difluoromethylornithine (DMFO) (Barrett *et al.*, 1999).

Suramin

This drug was introduced in 1922. It is a sulfonated naphthylamine, which is negatively charged, at physiological pH. Its mode of action is poorly understood, being an inhibitor for various enzymes (Pepin and Milord, 1994). It is generally considered to be the drug of choice for the early stages of *T. b. rhodesiense*. Due to its structure, suramin cannot pass through a lipid bilayer membrane, so it cannot pass the blood brain barrier. It is believed that suramin enters the parasites by binding to serum proteins that enter the trypanosome by endocytosis (Coppens and Courtois, 2000). Renal toxicity is the most common side effect of this drug.

Pentamidine

Pentamidine is an aromatic diamidine used to treat the early stages of *T. b. gambiense* infection for the past 50 years (Kuzoe, 1993). The drug concentrates inside the parasites through active transporters, the P2 transporter and at least two other transporter systems (de Koning, 2001). The mode of action is not fully understood but pentamidine has been shown to inhibit the synthesis of polyamines (Bitonti *et al.*, 1986) or to bind DNA (Edwards *et al.*, 1992). Hypotension, hypoglycaemia and nausea are the main side effects of this drug.

Melarsoprol

Melarsoprol was first synthesised by Friedheim in 1949. Melarsoprol is lipophilic and can cross the blood-brain barrier and cure the late stage disease of *T. b. gambiense* and *T. b. rhodesiense* (Keiser and Burri, 2000). It can also enter the trypanosomes via the active P2 transporter (Carter *et al.*, 1999). The mode of action of this drug is not known yet

(Stich *et al.*, 2003). Its inhibition of glycolytic enzymes has been suggested as a possible mode of action (Denise *et al.*, 1999). As an organic arsenical, this drug is highly toxic for humans. Adverse effects such as cutaneous reaction, diarrhoea and fever are common (Kuzoc, 1993). The most severe reactions are encephalopathic syndromes occurring in 5 to 10 % of cases (WHO, 1998), which are lethal in 1 to 5 %. Resistance to melarsoprol appears to be an increasing problem (Keiser *et al.*, 2000; Enanga *et al.*, 2002).

Eflornithine or DFMO

This drug was introduced in 1990. It inhibits ornithine decarboxylase leading to the inhibition of polyamine synthesis, which is essential for the proliferation of the cells (McCann *et al.*, 1986). This drug is used against early and late stage *T. b. gambiense*. DFMO is not effective against Rhodesian sleeping sickness (Wang, 1995). Trypanosomes treated with DFMO appeared to be non-dividing, short stumpy forms, incapable of changing their VSG coat (probably due to the decrease of putrescine and spermidine). They are then obliterated by the host immune system. This drug has few side effects but is difficult to produce and is very expensive.

1.5.2- Leishmaniasis

First line chemotherapy against leishmaniasis is mainly dependent upon pentavalent antimonials, including sodium stibogluconate or meglumine antimonate; a prolonged course may be necessary. Prolonged treatment can be associated with side effects such as body aches and fatigue (Herwaldt, 1999). The mode of action is not clear but the accumulation and retention of antimony by macrophages are important factors. Pentamidine can be used if antimonials are not effective (Amato *et al.*, 1998). In *L. donovani*, pentamidine is a competitive inhibitor of arginine transport. Resistance to

pentamidine has been described and can be due to the lack of accumulation of the drug in the mitochondrion because of a decrease in mitochondrial membrane potential (Basselin *et al.*, 2002). Amphotericin B is another drug that has been used in Europe where resistance to other drugs occurs (Olliaro and Bryceson, 1993).

Antimonial drugs are toxic and drug resistance in many areas renders parasites unresponsive to these drugs (Ouellette and Papadopoulo, 1993). Pentamidine is expensive, toxic and causes hypotension and hypoglycaemia. Amphotericin B has lower toxicity but is even more expensive. Recently a new compound, miltefosine, has been registered for use against leishmaniasis, it inhibits various enzymes of cell signalling pathways (Seifert *et al.*, 2003).

The elimination of natural reservoirs of this disease such as livestock, dogs and rodents has been used as a control strategy. Vaccines against *Leishmania* have not been successful due to the complexity of the immune response caused by the parasite (Handman, 1997; Bray *et al.*, 2003; Daneshvar *et al.*, 2003).

1.6- Morphology

1.6.1- Plasma membrane

African trypanosomes are flagellated protozoa. They measure approximately 15 × 4 μm. The plasma membrane encloses the cell and is essential to separate the cytosol from the extracellular medium. There is a flagellar pocket in all trypanosomatids, which appears as a depression of the membrane at the anterior region of the cell where the flagellum emerges. The pocket is the site where endocytic and exocytic exchanges occur with the external environment (Overath *et al.*, 1986). The plasma membrane is composed of phospholipids, cholesterol and glycolipids forming a bilayer. The membrane also contains

different proteins. In bloodstream and metacyclic forms, the parasite surface is composed largely of a single protein termed the variant surface glycoprotein (VSG), which protects the parasites from the host's immune defence (Vickerman, 1969). VSG synthesis is repressed when the parasite enters the fly (Vickerman *et al.*, 1988).

Other proteins present in the plasma membrane include transporters, which allow specific incorporation of nutrients, ions and other metabolic components into the cells. A number of different mechanisms allow a molecule to cross the plasma membrane. Non polar molecules cross the membrane by simple diffusion, whereas various polar molecules, such as ions or amino acids, use specific transport proteins or channels present in the membrane.

Passive transport is accomplished by the thermodynamic necessity of molecules to accumulate down a concentration gradient through channels and some carriers. Equilibrium is reached when the free concentration of the diffusing substance is the same on both sides of the membrane. Facilitated diffusion transport systems are saturable according to classical kinetic precepts.

Active transport enables a cell to transport substances against a concentration gradient. This phenomenon is coupled to a metabolic energy source. Most transporter protein can be characterised by two kinetics values, the V_{max} and the K_m. These parameters can be derived from experiments, in which rates of enzyme reaction are determined at different substrate concentrations, provided that the transporter obeys Michaelis-Menten rules. The values for V_{max} and K_m can be determined from a hyperbolic plot of initial velocity versus substrate concentration. K_m is the constant describing the concentration of substrate at which half maximal velocity (V_{max}) is observed. V_{max} is the maximal rate of uptake by the transporter.

1.6.2- Glycosomes

Glycosomes are membrane-bound cytoplasmic structures. They are termed glycosomes because most of the glycolytic enzymes involved in the conversion of glucose to 3-phosphoglycerate are present in this organelle (Opperdoes *et al.*, 1984; Michels *et al.*, 2000). Glycolytic enzymes isolated from the glycosome of trypanosomes frequently have an isoelectric point that is higher than the same enzymes in mammals (Souto-Padron and De Souza, 1979). *T. brucei* BSF, where the mitochondrion is poorly developed, have a large number of glycosomes (~ 250 glycosomes) (Opperdoes *et al.*, 1984). Conversely the PCF have well-developed mitochondrion but contain only ~ 50 glycosomes (Soares and De Souza, 1988). Glycosomes occupy approximately 9 % and 2.4 % of the cell volume in BSF and PCF, respectively (Bohringer and Hecker, 1975). In *Leishmania* amastigotes they comprised 1 % of the total cell volume (Coombs *et al.*, 1986).

Several pathways (other than glycolysis) can be found in glycosomes. These include enzymes involved in β-oxidation of fatty acids and pyrimidine biosynthesis (Michels *et al.*, 2000). Glycosomes do not have a genome; the proteins found in this organelle are nuclear-encoded and post-transcriptionally imported. Glycosomes are related to other microbodies such as peroxisomes and glyoxisomes in other eukaryotes.

1.6.3- Mitochondrion

Trypanosomes possess a single mitochondrion with its own genome. The DNA is condensed into a series of intercatenated circular molecules and forms the kinetoplast, representing 10 to 20 % of the total DNA of the cell. The form of the mitochondrion can change drastically between the different stages of the life cycle of the trypanosomes. The BSF of *T. brucei* possess a poorly developed mitochondrion composed of a single canal

lacking inner membrane foldings (Vickerman, 1985). In these forms, a functional Krebs cycle is absent as well as a classical respiratory chain (Vickerman, 1969). In PCF, the mitochondrial system is transformed into a network of canals consisting of inner membrane foldings. In this stage the Krebs cycle is functional (Tielens and van Hellemond, 1998) or partially functional (van Weelden *et al.*, 2003) and they possess a functional respiratory chain (Clayton and Michels, 1996). The mitochondrion occupies 3.3 % and 19.5 % of the cellular volume in bloodstream and procyclic stages, respectively (Bohringer and Hecker, 1975).

In spite of the fact that the mitochondrion has its own genome, most of the proteins present in this organelle are encoded by nuclear genes and synthesised in the cytoplasm. The majority of these are transported into the mitochondrial matrix but some of them are targeted to the intermembrane space or linked with the inner or outer membrane of this organelle. The current view is that there is a long targeting sequence on mitochondrial proteins that carries them to the correct location in the mitochondrion after their synthesis (Omura, 1998). Most proteins travelling from the cytosol to the mitochondrion share a common motif that is recognising by a mitochondrial outer membrane receptor. A group of 20-40 positively charged amino acids interspersed with some hydrophilic ones such as serine or threonine at the N-terminus of the protein constitutes the mitochondrial targeting sequence (MTS) (Nielsen et al., 1996; Nielsen *et al.*, 1997; Omura, 1998).

In the cytosol, mitochondrial proteins are bound by chaperone proteins, typically hsp70 or mitochondrial-import stimulating factor (MSF), which prevents their folding and facilitates passage through the mitochondrial pore in an unfolded state. This process is ATP-dependent (Voos *et al.*, 1999). The MTS binds to the receptor protein complex on the outer mitochondrial membrane permitting the protein to interact with the pore. This phenomenon involves a complex of proteins called Tom (Transport across the Outer

Membrane). At the level of the inner mitochondrial membrane, a similar complex of Tim (Transport across the Inner Membrane) proteins forms a pore through proteins traverse. This process is dependent on the mitochondrial proton motive force (Koehler, 2000; Voos *et al.*, 1999). Once in the right location chaperone proteins such as Hsp60 enable the folding of the mitochondrial protein and mitochondrial peptidases remove the MTS.

Hauser *et al.*, 1996, found that trypanosomatids possess typical mitochondrial targeting sequences on some proteins. Trypanosome mitochondria are clearly related to mitochondria from other eukaryotes in structure and function. The mechanism of mitochondrial protein import is also conserved, for instance, nuclear-encoded matrix proteins from trypanosomes contain cleaved N-terminal leader peptides. A classical sequence, composed of approximately 20 amino acids, similar to those found in other eukaryotes and a shorter one, composed of 9 amino acids, which alone is enough to direct protein to mitochondria have been described (Hausler *et al.*, 1997).

1.6.4- Acidocalcisomes

Acidocalcisomes are membrane-bounded structures, which possess an electron dense content. These cytoplasmic vacuoles contain high Ca^{2+} concentrations. The uptake of calcium occurs in exchange with H^+. This is done by Ca^{2+}-H^+ translocating ATPase activity. A vacuolar H^+-ATPase is found in the membrane of acidocalcisomes conferring an acidic pH (Vercesi *et al.*, 1994). Roles for acidocalcisomes are not certain, but they have been proposed to contribute to cell signalling (as a Ca^{2+} store) and possibly a store of amino acids (Docampo and Moreno, 2001).

1.7- Molecular biology

1.7.1- Kinetoplast, kinetoplastid DNA and RNA editing

Athough most mitochondrial proteins are encoded by the nuclear genome, the mitochondrion contains its own specific DNA called kinetoplastid DNA. *T. brucei* contains two types of DNA, approximately 50 maxicircles (20-40 kb), and 5,000 to 10,000 minicircles (1-3 kb), which are heterogeneous in sequence (Priest and Hajduk, 1994). These two classes of DNA form a huge intercatenated network (Simpson, 1987). Minicircles do not code for proteins, however they encode for guide RNAs (gRNAs) which have a role in the maturation of the mitochondrial mRNAs by a process called RNA editing (Blum *et al.*, 1990) (Figure 1.6).

In 1986, Benne *et al.* discovered that U residues were inserted into the mRNA after transcription and that this overcame frameshifts in the genomic sequence and created open reading frames from a non-sense sequence at the RNA level. U deletions also were found to occur at a lower frequency. Small RNA molecules, transcribed both from the maxicircle and the thousand of minicircles, contained the editing information. These gRNAs have a non-encoded 3' oligo U tail. The central portion of the gRNA contains sequence that is complementary to edited mRNA sequence.

gRNA first hybridises downstream of the editing site. Then a specific cleavage occurs at the mismatched base. Then U-residues are added to the 3' terminus of the 5' cleavage fragment. The added U residus could then extend the duplex and an RNA ligation occurs to join the cleavage fragments (Stuart *et al.*, 1997). RNA editing has been seen to be regulated during the parasite life cycle as a mechanism of regulation of mitochondrial gene expression (Read *et al.*, 1994).

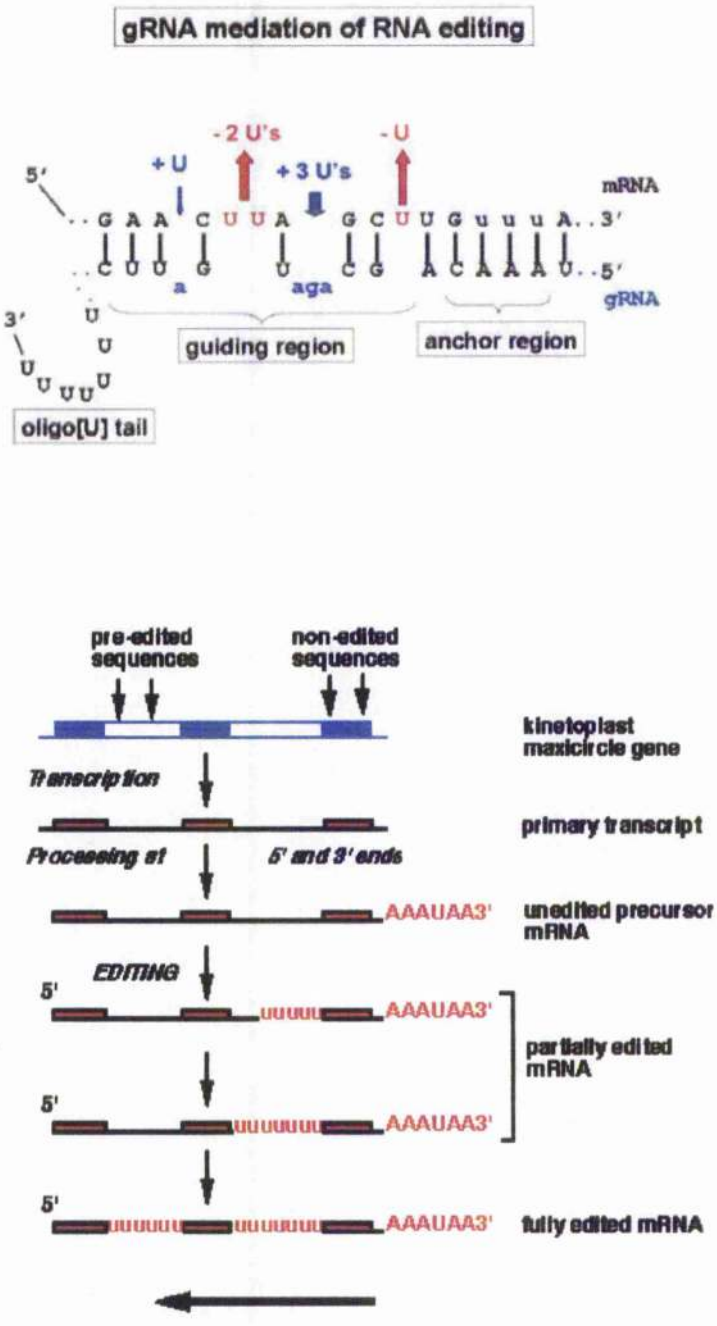


Figure 1.6: RNA editing in Trypanosomes

from: http://www.rna.ucla.edu/trypanosome/

1.7.2- Nucleus, transcription and gene expression

The size of the haploid genome of *T. brucei* is approximately 35 Mb (El-Sayed *et al.*, 2000), it is polymorphic because 25 % variation in genome content can be observed between strains. The nuclear genome of *T. brucei* consists of three chromosomes classes:

- Mega-base chromosomes (1 Mb to 6 Mb); there are 11 pairs of these chromosomes, which contain all protein-coding genes except some of the VSG genes.
- Intermediate chromosomes (200 to 900 Kb); there are between 1 to 5 intermediate chromosomes, they contain VSG expression sites.
- Mini chromosomes (50 to 150 Kb); 90 % of these chromosomes comprise a 177 bp tandem repeat with the remaining sequence consisting of other repeats and a silent telomeric VSG gene (El-Sayed *et al.*, 2000).

There are estimated to be 1000 VSG genes in the genome of *T. brucei* and only one is expressed at a time. The presence of non-transcribed VSG genes on minichromosomes might suggest that these chromosomes act as a reservoir of telomeric VSG genes, which may be transferred to expression sites by transposition.

In trypanosomatids, protein-coding genes do not, for the most part, contain introns. So far there is only a single known example of an intron, in the polyA polymerase gene of *T. brucei* and *T. cruzi* (Mair *et al.*, 2000). Genes are tightly packed on the chromosomes and are transcribed as polycistronic units of pre-mature RNA (Vanhamme and Pays, 1995; Clayton, 1999). This type of transcription could explain the lack of promoters in trypanosomatids, as only ribosomal RNA genes, procyclin and VSG gene expression sites have been shown to have specific promoters (Vanhamme and Pays, 1995).

The control of gene expression is mainly post transcriptional, using trans-splicing and polyadenylation, which require the addition of a 5' cap and a 3' poly A tail, respectively. At the 5' end of all trypanosomatid mRNAs, a spliced leader sequence (SL RNA), is added by

trans-splicing (Graham, 1995). The SL sequence is composed of a non-translated 39-41 nucleotide sequence, which is well conserved between trypanosomatid species (Agami and Shapira, 1992; Campbell *et al.*, 2000; Gibson *et al.*, 2000). The SL is derived from the 5' end of a small nuclear RNA molecule, it is about 120 nucleotides long. The mRNA of this gene is composed of the SL and what appears to be an intron. It appears that all mRNA molecules start with an intron, which is displaced by the SL. At the 3' end no specific signal for the addition of the poly A tail has yet been described although polypyrimidine tracts have been identified (Clayton, 2002) (Figure 1.7).

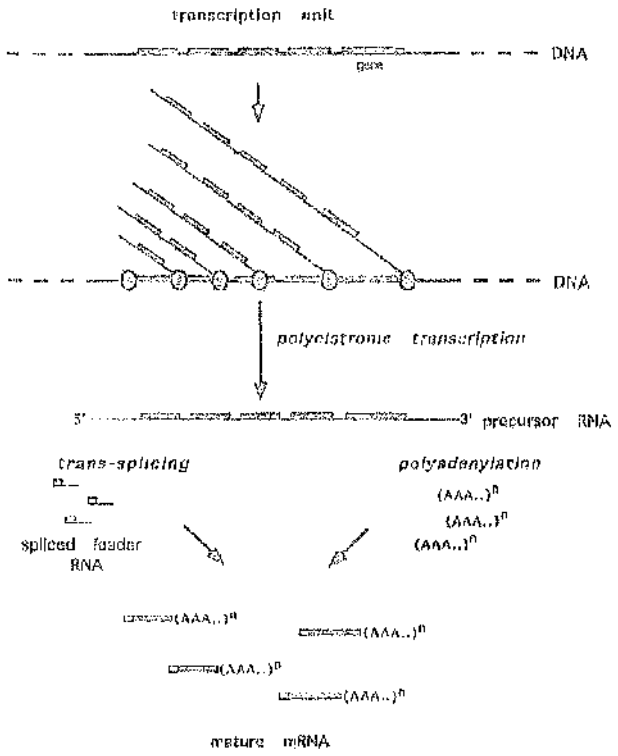


Figure 1.7: Transcription in trypanosomatids (from Graham, 1995)

The control of mRNA translation involves both the 5' UTR and 3' UTR regions. The 5' UTR contains elements for RNA-binding proteins (Day and Tuite, 1998) and the 3' UTR regulates the stability of the transcript (Boucher *et al.*, 2002).

All three classical RNA polymerases have been found in trypanosomes:

- RNA pol II is involved for the transcription of most of the protein-coding genes;
- RNA pol I seems to be involved in rRNA, VSG and procyclin gene transcription;
- RNA pol III is used for small RNAs.

1.7.3- Antigenic variation

A major consequence of polycistronic gene organisation is that regulation of expression is not possible at the level of transcription, since, having the same promoter, each gene is transcribed at the same rate. Despite this, VSG genes are expressed at different levels. The main function of the VSG is to protect against the host's immune system. One VSG gene out of an estimated 1000 genes is expressed at any given time. These genes are expressed in specific expression sites (ES) in telomeric loci. In bloodstream form *T. brucei*, the VSG genes are co-expressed with ESAGs (expression site-associated genes) in which the organisation is variable from an ES to another. VSGs and ESAGs are transcribed from a distant common 5' promoter (Borst *et al.*, 1998). Only one expression site is active at a time. Most chromosome ends possess expression sites but monoallelic expression appears to involve a unique RNA pol I body in these cells (Navarro and Gull, 2000). In addition to controlling expression of expression sites by apparently inserting them sequentially into the pol I expression body, a number of different homologous recombination events are known to bring new genes into an expression site (Figure 1.8).

The duplication of the silent gene into a ES and the deletion of the VSG resident at that site occur via duplicative transposition (if silent gene is within a chromosome) and telomere conversion (if silent gene is at the telomere). Reciprocal recombination involves the exchange of VSG sequences between two chromosomes. *In situ* activation involves silencing of the active expression site and activation of a new one.

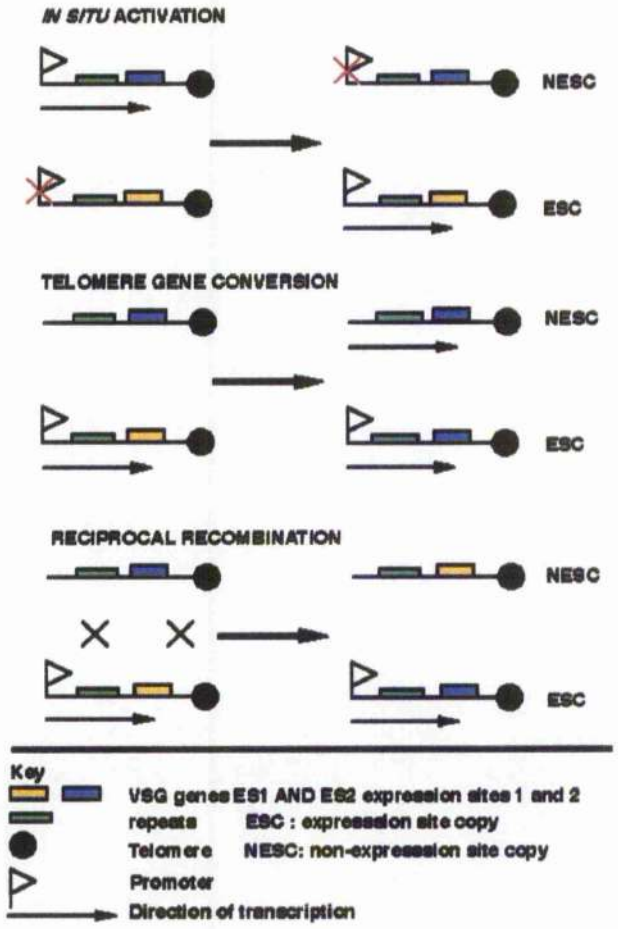


Figure 1.8: Antigenic variation mechanisms (from Vanhamme and Pays, 1995)

1.8- Energy metabolism in *T. brucei*

Trypanosoma brucei produces energy via glucose metabolism (through glycolysis) or the oxidation of amino acids (such as proline). All stages of *T. brucei* contain a mitochondrion, although the method of ATP production varies through the different life cycle stages (Tielens and van Hellmond, 1998).

In mammals, glucose is the main substrate for energy production and is metabolised via glycolysis. In bloodstream form (BSF) trypanosomes, the mitochondrion is poorly developed and lacks a functional Krebs cycle and respiratory chain (Clayton and Michels, 1996). Procyclic forms (PCF) trypanosomes present in the insect vector are thought to use predominantly amino acids as an energy source. The main energy source in tsetse flies is proline (Bursell, 1981) and this amino acid appears also to be used by the parasites for their energy production. In culture however, parasites preferentially consume glucose and threonine over proline and glutamine as judged by the rate at which these substrates are taken up from the medium (Cross *et al.*, 1975). In this life cycle stage the mitochondrion is more developed and could contain a functional respiratory chain, until the complex IV, the oxidative phosphorylation being non essential (Coustou *et al.*, 2003), as well as most enzymes of the Krebs cycle (van Weelden *et al.*, 2003).

1.8.1- Bloodstream forms

It has been reported that glycolytic activity is 10 times higher in BSF trypanosomes than in the typical cells of the mammalian host (Cazzulo, 1992). Glucose enters the parasite via a high capacity facilitative diffusion transporter (Bakker *et al.*, 1999) and then proceeds to the glycosome by an unknown mechanism (Clayton and Michels, 1996). The first seven enzymes of the glycolytic pathway (leading to two molecules of 3-phosphoglycerate per molecule of glucose) are found inside the glycosome (Figure 1.9).

The last three enzymes are in the cytosol, leading to the net production of ATP (by pyruvate kinase) (Opperdoes, 1987; Clayton and Michels, 1996) and pyruvate (Tielens and van Hellemond, 1998). The energy balance in the glycosome is zero because for two ATP molecules produced (by phosphoglycerate kinase and glycerol kinase, depending on oxygen status) two ATP molecules are consumed (by hexokinase and phosphofructokinase) for each glucose metabolised.

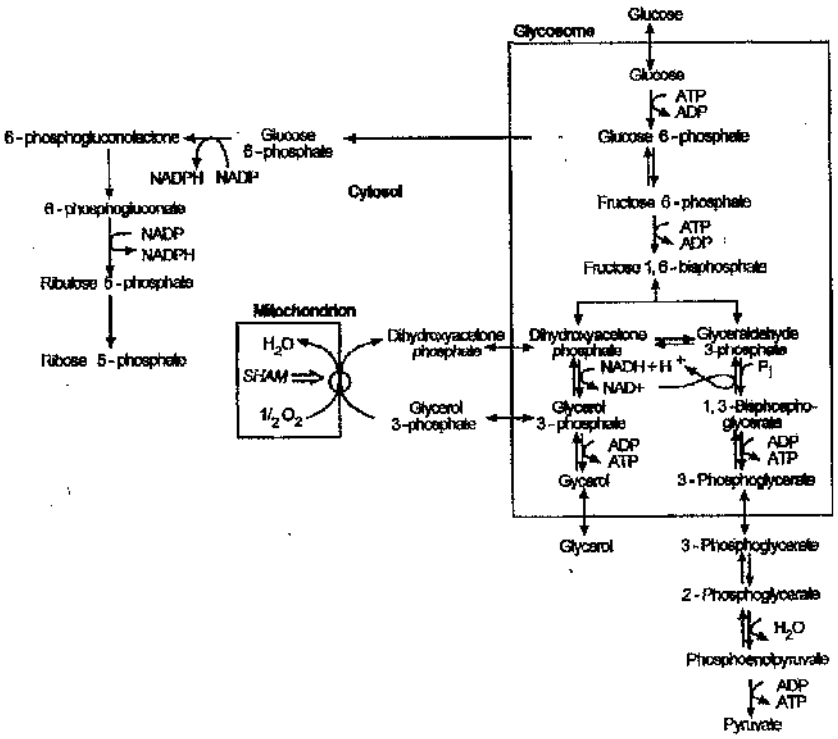


Figure 1.9: Energy metabolism in bloodstream forms of *T. brucei*, from Barrett MP, personal communication.

The high activity of glycolysis in BSF trypanosomes can be explained by the low ATP production (2 ATP molecules produced per mole of glucose consumed). The parasite therefore relies on a high consumption of glucose to produce enough energy. Glucose is present at high concentrations in mammalian blood, with its uptake rate apparently constituting the rate-limiting step of glycolysis depending on absolute concentration (Eisenthal and Cornish-Bowden, 1998; Bakker *et al.,* 1999). The glycosomal membrane constitutes a barrier to most of the glycolytic intermediates and cofactors, thus promoting high glycolytic flux (Opperdoes, 1987; Opperdoes and Michels, 1993). This compartmentalisation is essential to maintain the redox balance of NADH. NADH produced by glyceraldehyde-3-phosphate dehydrogenase is re-oxidised indirectly by mitochondrial glycerol-3-phosphate oxidase via a redox shuttle involving glycosomal glycerol-3-phosphate dehydrogenase and a transporter, which exchanges glycosomal glycerol-3-phosphate for dihydroxyacetone phosphate. In this system, involving ubiquinone, cytochrome Q and the alternative oxidase, oxygen acts as a final electron acceptor (Pollakis *et al.*, 1995). This constitutes a key role for the mitochondrion in energy metabolism of the BSF (Eisenthal and Cornish-Bowden, 1998). The BSF is also able to use glycerol to produce energy linking this metabolite to glycolysis through the glycerol kinase reaction, but it cannot produce energy by using either amino acids or fatty acids.

BSF *T. brucei* are able to survive and stay mobile in the absence of oxygen, but the ATP concentration is diminished under these conditions (Opperdoes, 1995). In this case the parasites cannot use the alternative oxidase system for NADH regeneration. As a result, glycerol-3-phosphate is converted to glycerol. Glucose metabolism continues at the same rate as in an oxygen-rich environment, but because the end products are equimolar glycerol and pyruvate, ATP production is halved. The alternative oxidase is sensitive to salicyhydroxamic acid (SHAM) and other hydroxamic acid analogues (Clarkson *et al.*,

1989), which mimics the removal of oxygen, but insensitive to cyanide (respiratory chain complex IV inhibitor) or antimycin A (complex III inhibitor).

1.8.2- Procyclic forms

The presence of the enzymes of the Krebs cycle was desmontrated in the PCF many years ago (Opperdoes, 1987). However it has also been suggested that PCF appear to lack a fully functional Krebs cycle. Recent work, involving the knock down or knock out of succinate dehydrogenase, α-ketoglutarate dehydrogenase and aconitase has also shown that the Krebs cycle is not essential for the parasite (Bochud-Allemann and Schneider, 2002). The aconitase knock out showed an accumulation of citrate (van Weelden *et al.*, 2003) but no effect on cell growth or energy production. Futhermore the activity of the Krebs cycle enzymes and the end products of metabolism vary from strain to strain.

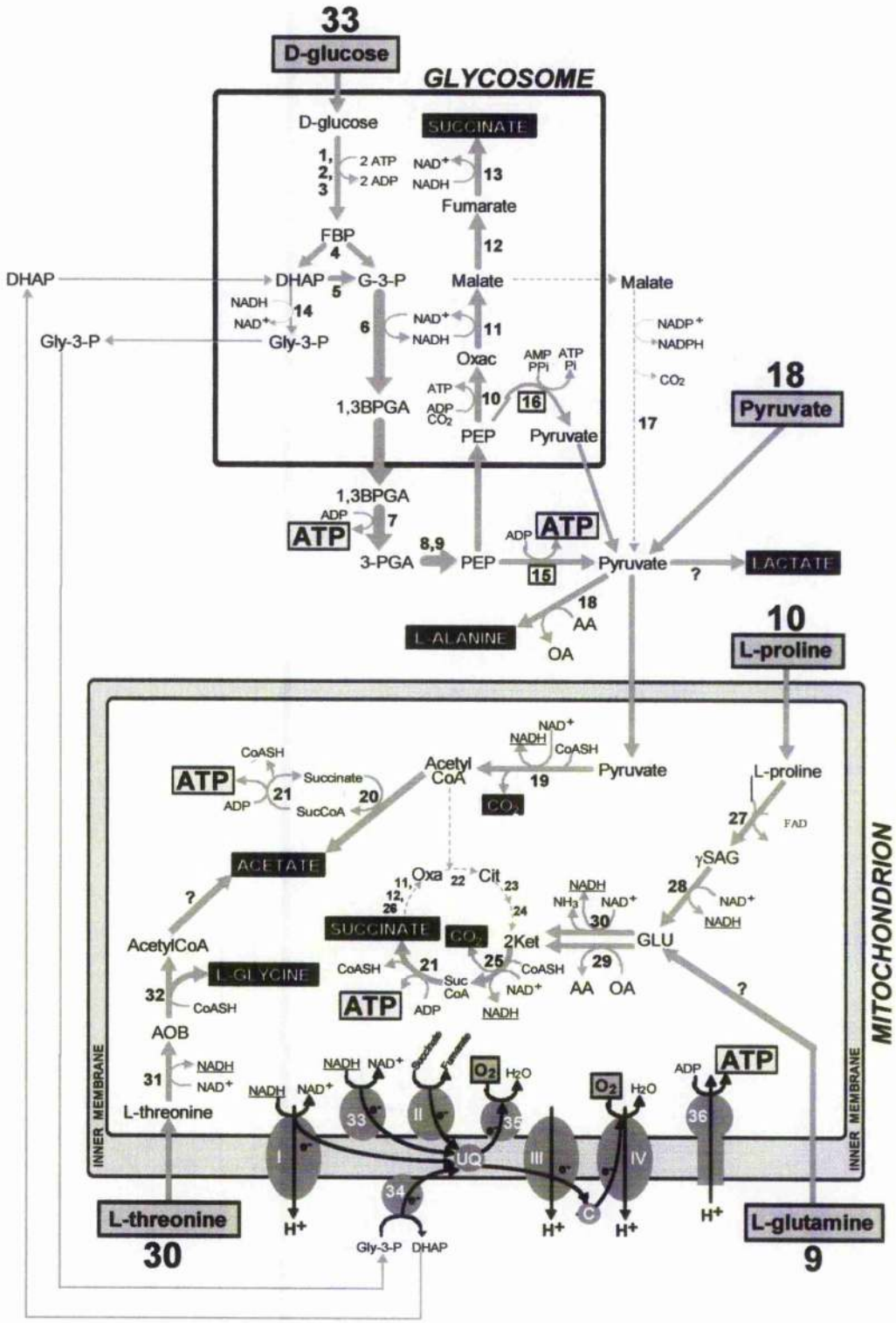


Figure 1.10: Energy metabolism in procyclic forms of *T. brucei*, from Coustou *et al.*, 2003

Legend of the figure 1.10:

Abbreviations: AA, amino acid; AOB, amino oxobutyrate; 1,3BPGA, 1,3-bisphosphoglycerate; C, cytochrome *c*; CoASH, coenzyme A; DHAP, dihydroxyacetone phosphate; FBP, fructose 1,6-bisphosphate; G-3-P, glyceraldehyde 3-phosphate; GLU, glutamate; 2Ket, 2-ketoglutarate; OA, 2-oxoacid; Oxa, oxaloacetate; PEP, phospho*enol*pyruvate; 3-PGA, 3-phosphoglycerate; Pi, inorganic phosphate; PPi, inorganic pyrophosphate; γSAG, glutamate γ-semialdehyde; SucCoA, succinate CoA; UQ, ubiquinone pool. Enzymes are: 1, hexokinase; 2, glucose-6-phosphate isomerase; 3, fructose-6-phosphate 1-kinase; 4, aldolase; 5, triose-phosphate isomerase; 6, glyceraldehyde-3-phosphate dehydrogenase; 7, phosphoglycerate kinase; 8, phosphoglycerate mutase; 9, enolase; 10, phospho*enol*pyruvate carboxykinase; 11, malate dehydrogenase; 12, fumarase; 13, NADH-dependent fumarate reductase; 14, glycerol-3-phosphate dehydrogenase; 15, pyruvate kinase; 16, pyruvate phosphate dikinase; 17, malic enzyme; 18, alanine aminotransferase; 19, pyruvate dehydrogenase complex; 20, acetate:succinylCoA transferase; 21, succinylCoA synthetase; 22, citrate synthase; 23, aconitase; 24, isocitrate dehydrogenase; 25, 2-ketoglutarate dehydrogenase complex; 26, succinate dehydrogenase; 27, proline dehydrogenase; 28, pyrroline-5 carboxylate dehydrogenase; 29, glutamate aminotransferase; 30, glutamate dehydrogenase; 31, L-threonine dehydrogenase; 32, acetyl CoA:glycine C-acetyltransferase; 33, rotenone-insensitive NADH dehydrogenase; 34, glycerol-3-phosphate oxidase; 35, alternative oxidase; 36, F_0/F_1-ATP synthase; I, II, III and IV, complexes of the respiratory chain.

ATP production in PCF appears to be possible through various pathways including oxidative phosphorylation, substrate level phosphorylation and glycosomal metabolism (Figure 10).

Recently, NMR analysis was used to study metabolic end products of proline metabolism (Coustou *et al.*, 2003). The glutamate formed during the oxidation of proline can be used as a substrate for respiration in many cell types. It is transformed into α-ketoglutarate which enters the active part of the Krebs cycle where succinate is formed as a mitochondrial end product suggesting arrest of the cycle after succinyl CoA synthetase (the enzyme producing ATP) in *T. brucei* (van Weelden *et al.*, 2003). In contrast, the end products of glucose metabolism are mitochondrial acetate, cytosolic alanine and

glycosomal succinate. Glucose therefore appears not to be metabolised through the Krebs cycle to an appreciable extent.

The activities of different enzymes of the proline degradation pathway have been defined in procyclic *T. congolense*. The glutamate derived from proline metabolism (the main amino acid present in the tsetse fly) was reported to enter the Krebs cycle or undergo transamination to form alanine (Obungu *et al.*, 1999).

In the insect, the glucose concentration is highly variable as glucose is rapidly absorbed from the blood meal within 15 min (Vickerman, 1985). As a result, it has been proposed that the parasite uses proline (whose level remains stable) as the preferred energy source (Bursell, 1973). In culture, procyclic *T. brucei,* in addition to using glucose as an energy source, can use proline. Indeed it was demonstrated that the addition of 2-deoxyglucose (2-DOG) (a substrate analogue), had no effect on trypanosome growth (Evans and Brown, 1972). When parasites were grown in high glucose concentration, more acetate than succinate was formed as an end product. When grown in high proline concentration, equal amounts of these compounds were reportedly formed (ter Kuile, 1997).

PCF *T. brucei* has been proposed to be dependent on the use of oxygen for energy production via the respiratory chain (van Weelden *et al.*, 2003). Recently there has been increased debate as to the mechanism of ATP production in procyclics. Succinate dehydrogenase has been proposed to be the only entry point of electrons into the respiratory chain, however ablation of its activity showed succinate dehydrogenase to be non-essential for the cells (Bochud-Allemann and Schneider, 2002). It has also a role in ATP production at the Acetate:Succinate CoA transferase/succinyl CoA synthetase (ASCT) reaction. Given that succinate dehydrogenase plays a role in ATP production through the ASCT cycle, the essential nature of the enzyme need not necessarily mean that

the respiratory chain is essential in ATP production. Experiments where neither KCN (an inhibitor of the respiratory chain) nor SHAM (a inhibitor of the trypanosome alternative oxidase) alone killed procyclic trypanosomes, while the two combined did, led to van Weelden *et al.*, 2003, proposing that either the respiratory chain was essential, or that KCN and SHAM together inhibit another essential pathway. However an excess of oligomycin (an inhibitor of mitochondrial ATP synthase) does not affect the intracellular steady state of ATP, suggesting that oxidative phosphorylation is not essential for procyclic trypanosomes (Coustou *et al.*, 2003). Loss of F_1F_0 ATPase, and indeed loss of respiratory chain, might be essential, but for reasons other than ATP production. For example, a proton gradient is critical to energising the mitochondrial membrane, which is essential for other features such as the accumulation of proteins and other essential macromolecules into the mitochondrial matrix.

ATP production through the electron transport chain may be non-essential, because other ways of ATP production inside the mitochondrion exist, e.g. substrate level phosphorylation. The ASCT cycle operates with acetate formed as an end product (van Hellmond *et al.*, 1998). The succinyl CoA that is produced during acetate formation by the ASCT cycle is metabolised by succinyl CoA synthetase in the mitochondrion and produces ATP from ADP. In the cytosol, substrate level phosphorylation is essential through the activity of pyruvate kinase, which is also essential for the parasite (Coustou *et al.*, 2003).

In PCF, pyruvate is the end product of glycolysis. It is not secreted but enters the mitochondrion, where it undergoes decarboxylation by pyruvate dehydrogenase into Acetyl-CoA. This is then converted into acetate by Acetate-Succinate CoA transferase, and involves a succinate/succinyl CoA cycle, with ATP also produced (van Hellmond *et al.*, 1998).

At the level of the glycosome, metabolism varies between BSF and PCF, with the former having higher levels of most glycolytic enzymes. In PCF, pyruvate kinase in the cytosol shows low activity whereas malate dehydrogenase and phosphoenolpyruvate carboxykinase (PEPCK) show high activity (Clayton and Michels, 1996; Barnard and Pedersen, 1994). Phosphoenolpyruvate re-enters the glycosome where it is ultimately metabolised to succinate. ATP is produced at this level by PEPCK and NADH is reoxidised by malate dehydrogenase. A key enzyme in this pathway, fumarate reductase has recently been found in procyclic forms. This enzyme also utilises NADH as cofactor, which it oxidises to NAD^+ thus restoring the sub-cellular redox balance (Besteiro *et al.*, 2002).

Experiments aimed at producing trypanosomes defective in glycosome biogenesis, showed that bloodstream stages cannot be selected under any conditions. PCF are also killed, but only when glucose is present. In this state parasites are not able to use glucose as an energy source, even when both glucose and proline are present in the media. Interestingly, when only proline is present the parasites are able to proliferate (Furuya *et al.*, 2002). This suggests that glucose could repress proline metabolism (or else exert its effect through an imbalance in cofactors, e.g. ATP and NAD^+ in the cytosol). With ATP production by glycolysis being zero, ATP must be produced in the mitochondrion. In this case, mitochondrial level phosphorylation is essential for *T. brucei* PCF. In the mitochondria, the ATP is produced by ASCT cycle or the succinyl CoA synthase in the Krebs cycle. Ablation of succinyl CoA synthase is lethal to the parasite, as it catalyses two essential reactions, including substrate level phosphorylation in the Krebs cycle and ASCT cycle (Bochud-Allemann and Schneider, 2002).

1.9- Proline dehydrogenase

Proline appears to be a key substrate for PCF energy metabolism. Proline utilisation requires two enzymes, proline dehydrogenase (PRODH) (EC 1.5.99.8) and pyrroline-5-carboxylate dehydrogenase (P5CDH) (EC 1.5.1.12), to convert proline into glutamate. Proline is an important source of carbon and nitrogen for the growth of many bacteria (Wood, 1981; Keuntje *et al.*, 1995). Plants can also use accumulated osmolytes including proline to counter osmotic stress. Proline accumulation is achieved by activation of the proline biosynthesis or inactivation of the proline degradation (Kiyosue *et al.*, 1996).

The first step catalysed by proline dehydrogenase yields Pyrroline-5-Carboxylate (P5C) and requires the reduction of a tightly associated cofactor, FAD (flavin adenine dinucleotide) (Brown and Wood, 1992; Becker and Thomas, 2001). Two electrons transferred from proline to FAD (via proline dehydrogenase) are then transferred to oxygen via the electron transport chain (Abrahamson *et al.*, 1983; Vinod *et al.*, 2002) (Figure 1.11).

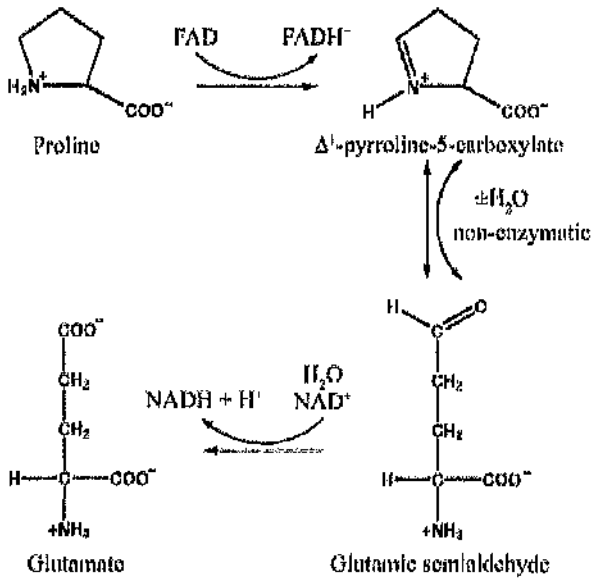


Figure 1.11: Proline degradation pathway (from Nadaraia *et al.*, 2001)

The PutA protein (comprising PRODH and P5CDH) in *E. coli* is associated with the inner cytoplasmic membrane (Brown and Wood, 1992) of these cells, because of the need for proximity of the enzyme with the electron transport chain (Menzel and Roth, 1981 a and b). In the second step P5CDH converts P5C to glutamate in an NAD^+-dependent reaction (Brown and Wood, 1993). In eukaryotes, PRODH and P5C dehydrogenase are encoded by two separate genes and are two different proteins (Peng *et al.,* 1996). PRODH is most usually associated with mitochondria in these cells, possibly because here too it transfers electrons to the electron transport chain. In bacteria both steps of proline degradation are catalysed by a single polypeptide encoded by the *putA* gene (Nadaraia *et al.,* 2001).

1.9.1- Proline dehydrogenase in other organisms

1.9.1.1- Gene organisation in prokaryotes

In enteric bacteria, the *putA* gene encodes a single polypeptide allowing both steps of proline degradation. In *E. coli* a 3,942 bp ORF encodes a protein of 1,313 amino acids (Xia *et al.,* 1995) with high homology with the same gene in *Salmonella typhimurium* and *Pseudomonas putida* (Vilchez *et al.*, 2000). It has been shown that this gene sits within an operon, which includes the *putA* gene and the *putP* gene, which encodes a proline-Na^+ symporter (Ratzkin and Roth, 1978). These two genes are divergently transcribed from an intergenic control region known as the *put* repressor (Ling *et al.,* 1994).

The expression of *putA* is induced by proline and is not affected by other amino acids. In addition, *putA* expression is autoregulated by PutA itself in *Rhodobacter capsulatus* (Keuntje *et al.,* 1995). The regulatory gene, *putR,* is located immediately upstream of *putA* and is required for the expression of proline dehydrogenase. The *putR*

gene is constituvely transcribed at a low level. In the absence of proline, *putR* activates the expression of *putA* to a low level (Keuntje *et al.,* 1995). This negative autoregulation by PutA is also found in other enteric bacteria and regulates expression of *putP* and *putA* in response to the availability of proline. The *put* operon of *S. typhimurium* is transcriptionally repressed by the PutA protein in the absence of proline. When proline and an electron acceptor are both present, PutA is prevented from binding to the *put* control region (Surber and Maloy, 1999). PutA is able to regulate its own expression. In the absence of proline, PutA remains in the cytoplasm where it can bind the *put* operator, which inhibits *put* gene expression. When proline is available PutA binds its substrate and is active (Muro-Pastor and Maloy, 1995). The genetic organisation and the mechanism of regulation of the gene cluster are different between species. In *E. coli, S. typhimurium* and *P. putida*, the *putA* gene and the *putP* gene are transcribed divergently, and they are arranged in opposite directions. In *Agrobacterium tumefaciens putA* is a monocistronic unit and *putP* is not adjacent to it. In *Vibrio vulnificus* the gene organisation is different again : the two genes *putA* and *putP* are transcribed in the same direction. In this species the operon is not repressed by the PutA protein (Lee J *et al.,* 2003).

1.9.1.2- Gene organisation in eukaryotes

The conversion of proline to glutamate takes place inside the mitochondria of *S. cerevisiae* by sequential action of two enzymes, PRODH and P5CDH, which are encoded by two different nuclear genes. The *put1* gene encodes PRODH and *put2* encodes P5CDH (Brandriss, 1983). These two genes are co-regulated by proline induction and by a control element encoded by the *put3* gene (Wang and Brandriss, 1987; Brandriss, 1983). The *put1* element is present as a single copy in the yeast genome. Expression of the gene is inducible

by proline and oxygen and is regulated at the level of RNA by proline concentration (Wang and Brandriss, 1986). It has been shown in *S. cerevisiae* that proline accumulation that occurs as a result of mutations in the proline dehydrogenase gene improves resistance to freezing and desiccation stresses (Takagi *et al.*, 2000).

In *Arabidopsis,* proline dehydrogenase is encoded by the *At-PDH* gene. Expression of this gene is down regulated during osmotic stress, allowing proline to accumulate. This amino acid can stabilise sub-cellular structures and scavenges free radicals (Mani *et al.*, 2002). Exogenous proline, in the absence of stress, is a good inducer of *At-PDH* expression in *A. thaliana* (Kiyosue *et al.*, 1996). The accumulation of proline in dehydrated plants is due to the activation of proline synthesis and the down-regulation of proline degradation. The *At-PDH* promoter is negatively regulated during dehydration and up-regulated during rehydration. An increase of *At-PDH* transcript can be seen in rehydrated plants, as well as in plants under hypo-osmotic stress (Nakashima *et al.*, 1998).

In *Drosophila melanogaster,* the *sluggish A* gene has been cloned and has homology to proline dehydrogenase of *S. cerevisiae* (Hayward *et al.*, 1993). In mutant flies, ablation of this gene induces sluggish behaviour (reduction in locomotion, phototactism problems).

A human homologue of the *D. melanogaster sluggish A* gene has been identified and is located in chromosome 22q11.2 (Harrison and Owen, 2003). This region is implicated in Type I hyperprolinaemia (CATCH 22 syndrome). Some studies have reported some neurological manifestation (schizophrenia) associated with type I hyperprolinaemia in humans (Jacquet *et al.*, 2003). A link between the human disorder and the sluggish behaviour in *Drosophila* has to be noted (i.e. both are neurological in origin) (Campbell *et al.*, 1997). Increased concentration of proline in hyperprolinemia may inhibit

synaptic decreases in the level of glutamate or γ aminobutyric acid (GABA) and plays a role in schizophrenia (Gogos *et al.*, 1999).

1.9.1.3- The enzyme

Most studies involving the PRODH protein have been carried out with products of prokaryote genes, using polypeptides that contain both PRODH and P5CDH activity.

PutA has been most extensively characterised in *E. coli* and *S. typhimurium* where it is a membrane-bound protein which catalyses the two step conversion of proline to glutamate. FAD^+ is used during the first step as a cofactor by transferring electrons from proline to an acceptor in the electron transport chain. The second step, which is the conversion of glutamic semialdehyde to glutamate, is performed by P5CDH and is coupled to NAD^+ as a cofactor. In the presence of proline, PutA is associated to the membrane and is active (Surber and Makoy, 1999). FAD^+ is reduced by proline and the FAD-redox state is implicated as a signal for this PutA-membrane association. PutA is also able to bind DNA and to act as a repressor of transcription of the *put* operon; this binding is also dependent on the FAD^+ redox state (Becker and Thomas, 2001; Zhu and Becker, 2003). The PRODH part of the PutA protein has been characterised and occupies residues 1 to 669 of the polypeptide (Vinod *et al.*, 2002). *In vitro* activity of PRODH when expressed alone is 4 times higher than in the entire PutA protein. Recently, this PRODH domain has been crystallised (Nadaraia *et al.*, 2001; Lee Y. *et al.*, 2003) and has both PRODH and DNA-binding activities. The protein is a homo dimer and each subunit contains three domains with residues essential for substrate and cofactor binding (the key residues of this enzyme will be outlined in chapter 3). The DNA-binding domain is located at the N-terminus of the protein.

By purification of the protein from membrane preparations, putA of the *E. coli* protein was characterised as a 260 kDa dimer (Scarpulla and Soffer, 1978). The monomeric and dimeric forms have also been recorded for *S. typhimurium* (Brown and Wood, 1992). The *Salmonella* protein is very similar to that of *E. coli*. A protein constituted of two equal subunits of 217 kDa was purified from membrane of *Clostridium sporogenes* (Monticello and Costilow, 1981).

In *S. typhimurium*, the putA protein shuttles between the cytoplasm and the membrane. Similar to the *E. coli* protein, in the absence of proline it acts as a repressor of transcription, while in the presence of proline it is active and is unable to bind DNA (Muro-Pastor *et al.*, 1997). Purification of the putA protein from the *S. typhimurium* membrane showed that putA-membrane interaction is due to protein-lipid interactions initiated by reduction of FAD^+; $FADH_2$ alters the conformation of putA, which increases its exposed hydrophobicity (Surber and Maloy, 1999).

When proline concentration reaches the K_m for proline binding, FAD^+ is reduced by oxidation of proline and the protein binds to the membrane. The K_m for proline was around 105 mM in *E. coli* (Graham *et al.*, 1984). A K_m of 102 mM for proline was observed with the recombinant PRODH part of the putA protein (Vinod *et al.*, 2002).

In rat liver, the location of PRODH is the inner mitochondrial membrane, (Brunner and Neupert, 1969), whereas the P5CDH can be found in the mitochondrial matrix (Small and Jones, 1990). By tagging the enzyme with the green florescent protein, Maxwell and Davis, 2000, also located proline dehydrogenase to mitochondria in humans. A putative mitochondrial targeting sequence has been found in the amino acid sequence of the *A. thaliana* PRODH (Peng *et al.*, 1996), indicating that in plants too the enzyme is mitochondrial.

1.9.2- Proline metabolism in trypanosomatids

There is evidence that non-carbohydrate compounds such as amino acids are the principal growth substrates for certain protozoa like *Leishmania* and *Trypanosoma* during their development in insect vector (section 1.8.2). Removal of carbohydrates from the medium for *Leishmania tarentolae* may be compensated by proline (Law and Ilukkada, 1979). Exponentially growing culture forms of *T. brucei* can use proline as an energy source, if an analogue of glucose (2-DOG) is included in the medium (Evans and Brown, 1972).

Proline is often the main amino acid of various insects' haemolymph with valine and alanine also frequently abundant (Balogun, 1974). Proline is present in tsetse fly hemolymph in concentrations as high as 150 mM under resting conditions and appears to be a major energy source during flight (Bursell, 1978, 1981). Many flies, including tsetse, utilize oxidation of L-proline as an energy source for flight (Bursell *et al.*, 1973). Energy for flight metabolism is derived from the partial oxidation of proline to alanine, which is then transported to the fat body where it is used in the regeneration of proline (Njagi *et al.*, 1992). The fat body represents triacylglycerol stored after digestion of the bloodmeal. It constitutes the main food reserve of tsetse. Insect stages of *Trypanosoma* have adapted to the proline-rich environment in their vectors by using this amino acid as a main metabolic source of energy (Zilberstein, 1993). Proline dehydrogenase, previously commonly called proline oxidase, is the first enzyme of the proline degradation pathway.

1.9.2.1- Proline degradation pathway

It has been demonstrated that tsetse flies depend principally on the oxidation of proline to produce energy for flying (Tritsh *et al.*, 1993; Obungu *et al.*, 1999). Proline is rapidly oxidised by PCF *T. rhodesiense*. The presence of PRODH can explain this utilisation (Ford and Bowman, 1973). The most active oxidase present in PCF of *T. rhodesiense* was shown to be L-proline oxidase (proline dehydrogenase) and the parasites can grow without glucose in the medium if L-proline is present. In this study, glucose was proposed to be used only when stationary phase was reached (Evans and Brown, 1972).

When promastigotes of *L. donovani* are incubated with radiolabelled proline, the label appears in glutamate, alanine, arginine and a number of Krebs cycle intermediates such as alpha-ketoglutarate, succinate, fumarate, malate and oxaloacetate as well as CO_2. Relatively little was incorporated into proteins (Law and Hukkada, 1979). In *T. congolense*, proline is primarily catabolised to glutamate via P5C. The P5C undergoes spontaneous hydration to form glutamate semialdehyde (Obungu *et al.*, 1999). When BSF transform to PCF, PRODH becomes active, and parasites are able to use proline (Hamm *et al.*, 1990; Overath *et al.*, 1986).

In *T. cruzi*, a proline racemase gene has been identified, which seems to be mitogenic for host B-cells (Reina-San-Martin *et al.*, 2000). Preliminary results using defined culture media indicate that epimastigote and metacyclic *T. cruzi* trypomastigotes can metabolize L or D-proline. Proline racemase is important for energy metabolism of the parasite. Such a protein has not been found in *T. brucei* and an orthologous gene is not present in the *T. brucei* genome database.

1.9.2.2- Proline transport

Proline is able to enter the trypanosomatid parasites via specific transporters. Law and Hukkada in 1979 were the first to suggest the presence of an active accumulation of proline in *L. major*. Due to the wide range of specificity of this transporter, it was postulated that proline accumulates inside these cells via a neutral amino acid transporter. In *L. donovani*, L-proline accumulates against its concentration gradient, so is dependent upon energy (Glaser and Hukkada, 1992). Proline transport is also subject to regulation by feedback inhibition and transinhibition by components of the intracellular pools in *L. tropica* promastigotes (Law and Hukkada, 1979). *L. donovani* amastigotes accumulate L-proline at a rate and extent that are much lower than those in promastigotes. Furthermore, in amastigotes the transport activity is optimal at pH 5.5, whereas in promastigotes it is optimal at pH 7. These observations suggest that promastigotes and amastigotes of *L. donovani* may possess two distinct transporters for L-proline, the expression of which may be regulated by pH of the culture medium (Zilberstein, 1993; Mazareb *et al.*, 1999). The K_m for L-proline in *L. major* promastigotes was measured at 60 μM, whereas a value of 0.65 mM was proposed for *L. donovani* (Zilberstein and Gepstein, 1993).

In *T. brucei*, glucose could be replaced by proline in the medium without any adverse effect on the growth of PCF (Evans and Brown, 1972). The characteristics of proline transport in the PCF of *T. brucei* were studied by using radioactive proline (L'Hostis *et al.*, 1993). Proline uptake was shown to be performed against its concentration gradient and to require an active transport system. This transporter was proposed to be ATP-dependent but independent of Na^+, K^+ or H^+ co-transport; apparently it is not driven by a proton motive force. The K_m value for proline uptake in *T. brucei* is somewhat lower than that found in *Leishmania* promastigotes (L'Hostis *et al.*, 1993). This transporter was of a relatively high affinity since the K_m measured was at 19 μM.

1.10 Aims of this study

The overall aim of this project was to learn more about the role of proline metabolism in trypanosomatids and the part played by proline dehydrogenase in this process. The work was divided into different two main parts:

- Identification and characterisation of genes encoding proline dehydrogenase.
- Studying the physiological role of proline dehydrogenase in *T. brucei* by knocking down expression using RNA interference.

Chapter 2
Materials and Methods

2.1- Parasites

2.1.1- Culture of *Leishmania* species

Leishmania major (MHOM/IL/80 Friedlin) and *L. mexicana* (MNYC/BZ/62/M379) promastigotes were cultured in HOMEM medium (Berens *et al.*, 1976) (Gibco, Life Technologies) with 10% (v/v) heat inactivated foetal calf serum (FCS), at 25 °C (Daneshvar *et al.*, 2003). Cultures were initiated at 2×10^5 cells/ml, with stationary phase ($1\text{-}2 \times 10^7$ cells/ml) being reached after 8 days, upon which cells were sub-passaged (Bates, 1994).

L. major metacyclics were purified from a stationary phase population of cells using the agglutination method described by Sacks *et al.*, 1985.

L. major and *L. mexicana* amastigotes were purified from infected BALB/c mice by D. Laughland (University of Glasgow) (Hart *et al.*, 1981). *L. mexicana* amastigotes were grown axenically in Schneider's Drosophila Medium, pH 5.5, containing 20 % (v/v) heat inactivated FCS at 33 °C.

2.1.2- Culture of *Trypanosoma* species

Trypanosoma brucei strain 427 procyclics were cultivated in SDM 79 medium (Brun and Schonenberger, 1979) and defined SDM medium (Appendix 1) supplemented with 10 % (v/v) heat inactivated FCS, at 25 °C. A typical culture was started at $1\text{-}5 \times 10^5$ cells/ml. After 2 or 4 days, the culture reached mid-log phase ($3\text{-}6 \times 10^6$ cells/ml) and after approximately 7 days the stationary phase ($1\text{-}2 \times 10^7$ cells/ml) was reached.

BSF *T. brucei* (strain 427) were cultured at 37 °C and 5 % (v/v) CO_2 in HMI-9 medium containing 20 % (v/v) heat inactivated FCS (Hirumi *et al.*, 1977; Hirumi, 1994).

Procyclic cultures of *T. brucei* strain 29-13 (LaCount *et al.*, 2000) were used in all RNA interference experiments. These cells were grown in SDM 79 medium containing 15 μg/ml G418 and 25 μg/ml Hygromycin B (see section 2.4).

The densities of all cultures were determined using an improved Neubauer haematocytometer (Weber Scientific).

2.1.3- Alamar blue assay

This assay was used to estimate the number of *T. brucei* cells alive in medium (Raz *et al.*, 1997). SDM79 or defined SDM media without glucose and proline and supplemented by different amino acids or other potential carbon sources were used. Parasites were grown at 25 °C in 96 wells plates (200 μl per well) from the starting density of 1-5 × 10^5 cells/ml. After a few days of growth, 20 μl of Alamar® blue dye (Trek Diagnostic System) was added to each well. After 24 hours incubation at 25 °C, to allow the reduction of Alamar® blue, the fluorescence was read (530 nm excitation, 590 nm emission) with a fluorimeter (Perkin Elmer, LS 55, Luminescence Spectrometer) and was correlated to the concentration of parasites alive in the medium by determining the density of living parasites.

2.2- Molecular biology techniques

2.2.1- Isolation of genomic DNA

2.2.1.1- L. major and T. brucei

1 x 10^8 parasites were pelleted by centrifugation at 2,000 g for 10 min at 4 °C and resuspended in 150 µl of TELT buffer (Appendix 1) (Medina-Acosta and Cross, 1993). The pellet was incubated at room temperature for 5 min. 150 µl of Phenol-Chloroform (1:1) was added and mixed gently by inversion. The samples were centrifuged at 13,000 g for 5 minutes and the aqueous layer retained. The DNA was precipitated by addition of 0.1 volume of 3 M sodium acetate and 2 volumes of absolute ethanol. After 5 min on ice, the DNA pellet was collected by centrifugation at 13,000 g for 5 min at 20 °C. The pellet was washed with 70 % ethanol, air dried and resuspended in TE (Appendix 1) buffer. Samples were then incubated with 1 mg/ml of RNAse (Sigma) for 1 hour at 37 °C and stored at 4 °C.

2.2.1.2- Isolation of Escherichia coli gDNA

100 ml of *E. coli* strain JM109 culture were centrifuged at 2,500 g for 5 min. 40 ml of TE were added to the pellet and after resuspension the mixture was centrifuged at 2,500 g for 5 min at 4 °C. The pellet was then resuspended in 3.2 ml Tris/sucrose buffer (Appendix 1) containing 100 µl lysozyme (10 mg/ml) and stored on ice 10 min (to lyse the cells). Proteinase K (0.5 mg/ml) was added and incubated for 3 hours at 51 °C in order to digest the proteins present in the sample. The DNA was extracted using phenol/chloroform and then precipitated using 0.1 volume of 3 M sodium acetate and 2 volumes of 100 % ethanol. After centrifugation at 13,000 g for 5 min, the pellet was dried and resuspended in 50 µl of sterile water and stored at 4 °C until use.

2.2.2- Total RNA isolation

2 x 10^8 cells (*T. brucei* or *L. major*) were re-suspended in 1 ml of TRIzol® reagent (Gibco, Life Technologies), comprising monophasic solutions of phenol and guanidine isothiocyanate, to maintain the integrity of RNA. The samples were incubated for 5 min at room temperature, then 0.2 ml of chloroform was added and the tubes shaken slowly and incubated at room temperature for 2 min to separate the organic and aqueous phases. After centrifugation (15 min, 12,000 g, 4 °C), the aqueous phase was transferred to a fresh Eppendorf tube and 0.5 ml of isopropyl alcohol was added. The mixture was incubated at room temperature for 10 min to precipitate the RNA. After centrifugation at 12,000 g for 10 min at 4 °C, the pellet was washed with 75 % (v/v) ethanol, air dried for 10 min and re-suspended in RNAse-free H_2O. Samples were stored at –70 °C. All equipment and reagents were made RNAse free by treatment with diethylpyrocarbonate (DEPC) at 0.01 % v/v.

2.2.3- Polymerase chain reaction (PCR)

PCR was used to amplify segments of DNA situated between two known regions. All oligonucleotides were synthesised by MWG-Biotech. The two primers used in each amplification had sequences complementary to those flanking the region for amplification. One primer was designed in the 5'-3' direction of the plus strand and the other one in the 5'-3' direction of the complementary strand of DNA.

The enzyme used for the PCR reaction was the *Pfu* DNA polymerase (Promega), which has proof-reading capability. The final concentrations of the PCR reaction were: 1X PCR buffer (Promega), 0.5 mM PCR nucleotides mix, 1 unit of *Pfu*, 100 pmol of each primer, 5 % (v/v) DMSO and 50-100 ng gDNA. Temperature and incubation time were optimised for each reaction. The PCR was performed in three stages: 94 °C, 2 min; 94 °C,

30 s; Tm, 30 s; 72 °C, 5 min; 30 cycles; 72 °C, 7 min, using the GeneAmp PCR system 2400 (Perkin Elmer). The Tm is the annealing temperature, which was chosen to be a few degrees below the Tm of each pair of primers. Primers used during this study are detailed in Table 2.1.

Name of PCR products	Primers name	*Primers sequences (5'-3')*	Tm °C	Size of product expected
Putative amplification from database	Lm1 Lm1 rev Tb1 Tb1 rev	CAC CAC TCT AAT TCT TCC AAC AAA ATG CAA GCC ACC AGA AGA CAA CGA GTC GAA GAG CAC ATG AAG GCA TTT AAA CTA GAT CGT TCG AC	52 50	2200 bp 2100 bp
RT-PCR Lm	SL1 RTlm1 SL2 RTLm1	TAA CGC TAT ATA AGT ATC AGT TTC AGC TTC AGT TTC TCG TTA AAG AAC AGT ATC AGT TTC TGT ACT TTA TTG AGC TTC AGT TTC TCG TTA AAG AAC	52 52	495 bp minimum
RT-PCR Tb	SL1 RTTb1 SL2 RTTb1	TAA CGC TAT ATA AGT ATC AGT TTC ACA TGC CTG TCA CCT TAA CAG C AGT ATC AGT TTC TGT ACT TTA TTG ACA TGC CTG TCA CCT TAA CAG C	56 56	540 bp minimum
LmPRODH	Lment Lmrev	AAA **CAT ATG** CGT CGT CTC CTC CCC CTG AAA **CTC GAG** AAA ACC GCA GCA GCG GCG	60	1683 bp
LmPRODHtruncated	Lmtrunc Lmrev	AAA **CAT ATG** CAG GAC AAG CAG CCG AAG AAA **CTC GAG** AAA ACC GCA GCA GCG GCG	60	1608 bp
LmPRODH72	LmPRODH72 Lmrev	AAA **CAT ATG** AAG AGA GTG GAA AAG ATC CTT GGC AAA **CTC GAG** AAA ACC GCA GCA GCG GCG	58	1467 bp
TbPRODH	Tbent Tbrev	AAA **CAT ATG** TTT CGC CTC CTG TGC CGA CCC **AAG CTT** CAT CCA AAA GAC GCG	60	1668 bp
TbPRODHtruncated	Tbtrunc Tbrev	AAA **CAT ATG** CAG GAC AAG CAG CCG AAG CCC **AAG CTT** CAT CCA AAA GAC GCG	60	1599 bp
TbPRODH72	TbPRODH72 Tbrev	AAA **CAT ATG** AAG CGT GCT GAG GCA ATT TTT CCC **AAG CTT** CAT CCA AAA GAC GCG	58	1473 bp
RNAi	RNAi forward RNAi reverse	ATT **TTC TAG** ACT CGG ACC CAT CCA TAT TTC G CCC **TCT AGA** ATC ACA ACA TTA CAC TCC TCC	63	606 bp
PutA669	Pox forward Pox reverse	AAA **CAT ATG** GGA ACC ACC ACC ATG AAA **CTC GAG** TTC CGC AGG GTT AAT AAC	60	2004 bp

Table 2.1: Primer sequences and PCR conditions used during this study.

Lm: *L. major*; T.b: *T. brucei*; SL: Spliced leader. Sequence in bold represents the restriction sites : *Hind* III: CTC GAG, *Nde* I: CAT ATG, *Xho* I: CTC GAG

Name of the genes amplified (Table 2.1):

LmPRODH: *L. major* proline dehydrogenase full length gene.

LmPRODHtruncated: *L. major* proline dehydrogenase truncated gene (i.e. without the putative mitochondrial targeting sequence).

LmPRODH72: *L. major* proline dehydrogenase without the N-terminal 72 amino acids (i.e. without the putative transmembrane domain).

TbPRODH: *T. brucei* proline dehydrogenase full lenth gene.

TbPRODHtruncated: *T. brucei* proline dehydrogenase truncated gene (i.e. without the putative mitochondrial targeting sequence).

TbPRODH72: *T. brucei* proline dehydrogenase without the N-terminal 72 amino acids (i.e. without the putative transmembrane domain).

2.2.4- RT-PCR

2.2.4.1- cDNA preparation

cDNA synthesis systems generally contain the reagents required for the synthesis of double stranded cDNA from RNA. The classical method of cDNA synthesis uses oligo(dT) primers. This approach was used in this project.

The enzyme SUPERScript II, RNase H⁻ Reverse transcriptase (Gibco, Life Technologies) was used to synthesize first-strand cDNA. 1 µl of oligo dT (500 µg/ml), 5 µg of total RNA, 1 µl of 10 mM dNTP mix and sterile distilled water to 12 µl were added to a nuclease-free tube. The mixture was heated to 65 °C for 5 min to remove any secondary structures and chilled on ice for 2 min. 4 µl 5X First-strand Buffer (GibcoBRL), 2 µl 0.1 M DTT and 1 µl RNaseOUT (GibcoBRL) Recombinant Ribonuclease Inhibitor (40 U/µl) were added and incubated at 42 °C for 2 min. 1 µl (200 U) of SUPERScript II

was added and the sample was incubated at 42°C for 50 min. The reaction was inactivated by heating at 70 °C for 15 min. The cDNA was used as a template for PCR amplification.

2.2.4.2- PCR

To obtain the 5' end of the putative proline dehydrogenase gene, one specific internal primer for this DNA sequence and two primers from the splice leader (SL) sequence were designed (Figure 2.1). The SL is a 39 nucleotides sequence that is added to the 5' end of all mRNA molecules in trypanosomatids (Agami and Shapira, 1992). It is well conserved between all trypanosomatids and the same SL primers were used for *T. brucei* and *L. major* amplifications.

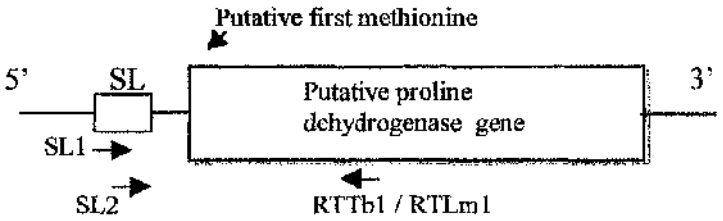


Figure 2.1: Primer positioning for the RT-PCR

A typical PCR reaction was performed using the cDNA. SL1 and RTTb1/RTLm1 were used to make the first PCR product (94 °C, 2 min; 94 °C, 30 s; Tm, 30 s; 72 °C, 5 min; 30 cycles; 72 °C, 7 min). Nested PCR was performed using the same conditions with SL2 and RTTb1/RTLm1. In this second reaction, the template used was the first PCR product. The

aim of this nested PCR was to amplify the specific sequence at the beginning of the gene (5' region) to facilitate identification of the initiation codon.

2.2.5- Cloning Strategy

DNA fragments can be inserted into vectors following different strategies. Efficient cloning is achieved when DNA insert and plasmid are digested by restriction enzymes creating "sticky ends". Another approach involves the cloning of PCR products treated by the Taq DNA polymerase. This enzyme can add a single A-residue at the 3' end of the insert, which can be cloned into specific vectors such as pGEM-T vector, which contains a single T-residue overhang. Both approaches were used during this thesis.

2.2.5.1- Isolation of PCR products

PCR products were separated on a 1% (w/v) agarose gel containing ethidium bromide, in 1X TAE buffer (Appendix 1) and calibrated using a DNA ladder (1 kb, Promega). DNA in the gel was visualised by ultraviolet light from a transilluminator (UVP Laboratory Products). Products were extracted from the gel using a QIAquick PCR purification kit (Qiagen), following the manufacturer' s instructions.

2.2.5.2- Restriction digestion of DNA

Restriction enzymes recognise specific sites of different lengths and composition. Standard restriction enzyme digests were typically performed in a volume of 20 µl containing 2 µl of 10 X buffer, 50-100 ng of DNA (plasmid or PCR product) and 5 units of the enzyme. The buffer systems used were variable, but were invariably provided by the manufacturer with the enzymes. The mixture was incubated for 1-4 hours at 37 °C to complete the digestion.

2.2.5.3- Ligation

In the first instance, PCR fragments were cloned into the pGEM-T vector (Promega). To optimise cloning efficiency, an A-tailing reaction using Taq DNA polymerase was performed on the PCR products. The reaction contained: 5 µl of purified PCR fragment; 1 µl 10X buffer; dATP (0.2 mM); 5 U Taq DNA polymerase; 3 µl distilled H_2O; incubation was at 70 °C for 30 min. Ligation of the PCR products into the pGEM-T vector was performed using the following conditions: 5 µl 2X buffer; 1 µl vector at 50 ng/µl; 3 µl PCR product (150 ng); 1 µl T4 DNA ligase at 3 U/µl; the reaction was incubated for 16 hours at 16 °C. A positive control (control insert DNA) and a background control (digested vector without insert DNA) were used to assess re-ligation of the vector. The PCR products were then cloned into the $pET21a^+$ vector (Novagen) for recombinant protein expression. $pET21a^+$ vector and the insert were initially digested by restriction enzymes and then ligated together using the same protocol as for the pGEM-T vector.

2.2.5.4- Competent cells

The *E. coli* competent cell lines JM109 or DH5α were purchased ready to use from Promega. BL21(DE3) cells were made competent (able to be transformed) using the calcium chloride protocol described by Sambrook *et al*, 1989. A single colony from a LB agar plate was picked and grown overnight at 37 °C in 5 ml of LB medium (Appendix 1). This pre-culture was then used to inoculate 100 ml of LB medium. Culture density was monitored spectrophotometrically every 30 min, until the OD_{600} was equal to 0.4-0.6. Cells were then centrifuged at 2,500 g for 10 min at 4 °C and the resulting pellet resuspended in 10 ml of 0.1 M $CaCl_2$, at 4 °C. Cells were then centrifuged at 2,500 g for 10 min at 4 °C and resuspended in 2 ml ice cold 0.1 M $CaCl_2$. Competent cells were conserved on ice and used immediately.

2.2.5.5- Transformation of competent cells

BL21(DE3) were transformed with the ligation products by a standard heat shock method (Sambrook *et al*, 1989). 10 ng of DNA was mixed with 200 µl of the competent cells and incubated on ice for 10 min. A heat shock (42 °C, 50 s) was performed and the cells were then placed on ice for 2 min. 900 µl of LB medium was then added and the mixture shaken for 1 hour at 37 °C. 100-200 µl of the transformation was then plated onto LB plates containing the relevant antibiotic selection.

Successful cloning of the insert into pGEM-T was indicated by the presence of white colonies using X-Gal colour selection. Cells were plated onto LB agar plates containing ampicillin (100 µg/ml), isopropylthio-β-D-galactoside (IPTG) (40 µg/ml) and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) (40 µg/ml). White colonies, with the vector and the insert, were picked and cultured at 37 °C overnight in LB medium containing ampicillin (100 µg/ml). The isolation of the plasmid was performed using the QIAprep Miniprep kit (Qiagen). To check for the presence of the correct insert, the plasmid was digested with different restriction enzymes and fragments separated by agarose gel electrophoresis. The vectors with cloned PCR fragments were sent to MWG-Biotech for sequencing. Electroferograms were analysed using Vector NTI (Informax).

The same protocol was used to select the cells containing the proline dehydrogenase gene inside the pET 21a^{+} vector. The selection was done in the presence of ampicillin only since β-galactosidase is not encoded within this vector.

2.2.6- Southern blot analysis

Southern blotting enables localisation of particular sequences within genomic DNA. 5 µg gDNA was digested completely by restriction enzymes and the resulting fragments separated according to their size by electrophoresis through a 1 % (w/v) agarose

gel in 1X TAE buffer. The DNA was then denatured *in situ* and transferred from the gel to a nylon membrane by capillary transfer according the method of Southern (1975) (as described in Sambrook *et al*, 1989). After having been photographed, the gel was incubated in 0.125 M HCl for 30 min under gentle agitation in order to allow the depurination of the DNA. The gel was then incubated for another 30 min in 1.5 M NaCl and 0.5 M NaOH to denature the DNA and then in 1 M Tris pH 7.5, 1.5 M NaCl to neutralise the gel. The transfer of DNA onto a nylon filter (Hybond N, Amersham) was done using 20X SSC (Appendix 1). Capillary transfer was used to transfer DNA fragments from the gel in a flow of liquid and deposited on the surface of the solid nylon support transport using standard procedures (Sambrook *et al.*, 1989). An ultraviolet cross-linking procedure was used to fix DNA to the membrane (Spectrolinker XL-1000 UV linker, Spectronics Corporation), for optimal cross-linking.

2.2.7- Northern blot analysis

RNA samples (~4 µg) were separated according their sizes, by 1 % agarose/formaldehyde gel electrophoresis for at least 4 hours at 60 V. The gel contained 1 % (w/v) agarose, 12 % (v/v) formaldehyde and 2 X Northern gel buffer (Appendix 1). RNA samples were prepared by adding the RNA to 12.5 % (v/v) formaldehyde, 50 % (v/v) formamide and H_2O to 20 µl. Samples were heated for 15 min at 55 °C to denature the RNA. The gel was then soaked in 20 X SSC for 30 min, before being placed in contact with the nitrocellulose membrane for RNA transfer to the solid support in an ascending flow of buffer (20 X SSC) similar to that used in Southern blotting. Once transferred, the membrane was UV cross-linked and stained with methylene blue in order to ensure equal amounts of total RNA in each lane.

2.2.8- Probe labelling

^{32}P-labelled probes were made by randomly priming a PCR product that had been extracted from a gel using the Prime It® II kit (Stratagene). 10 µl of random oligonucleotide primers were added to 25 ng of DNA in a total reaction volume of 34 µl. The mixture was boiled for 5 min in order to denature the probe. 10 µl 5X dATP primer buffer, 50 µCi α^{32}dATP and 1 µl of Exo Klenow enzyme (Stratagene) (5 U/µl) were added and incubated in 37 °C for 10 min. The probe was boiled 5 min before addition to the membrane.

Prehybridisation was carried out for 2 hours at 42 °C in 50 % (v/v) formamide, 5 X SSC, 10 X Dendhart's solution, 0.1 % SDS, 20 mM NaH_2PO_4, pH 6.5 and 0.2 mg/ml salmon sperm DNA. Hybridisation of the probe was effected in the same buffer overnight at 42 °C.

Blots were washed three times for 30 min at 55 °C in 0.1 X SSC and 0.1 % SDS and exposed to autoradiography (Compact X4, X-ograph, Imaging systems).

2.3- Biochemical techniques

2.3.1- Determination of protein concentrations

The Bio-Rad protein assay was used to determine protein concentrations. This system is based on the Bradford method (Bradford, 1976), which involves the addition of an acidic dye to the protein solution. It forms a complex, which absorbs at 595 nm as a function of protein concentration. A Bovine Serum Albumin calibration curve (0.05-0.5 mg/ml) was performed to quantify the samples.

2.3.2- SDS-PAGE

Denaturing polyacrylamide gels ensure dissociation of the proteins into their individual polypeptide subunits. Commonly a detergent, sodium dodecyl sulfate (SDS), is used in combination with a reducing agent and heat to dissociate the proteins before loading them onto the gel. The denatured polypeptides bind SDS and become negatively charged, the complexes migrate through the gel according their molecular mass. The method used was described by Laemmli (1970). The resolving gel contained 10 % (v/v) polyacrylamide (Acrylamide-Bis, BioRad), 25 % (v/v) resolving buffer, 0.1 % SDS, 7.5 % (v/v) ammonium persulfate (Sigma) and 0.1 % TEMED (Sigma). The stacking gel contained 4 % (v/v) polyacrylamide (Acrylamide-Bis, BioRad), 25 % (v/v) stacking buffer, 0.1 % SDS, 5 % (v/v) ammonium persulfate (Sigma) and 0.1 % TEMED (Sigma). Samples were mixed with 1 X loading buffer and boiled for 5 min prior to loading. The gel was run at 120 V for 1-2 hours in electrophoresis buffer (Appendix 1). The gel was then stained for 1 hour with 0.25 % Coomassie Brillant Blue R250 and destained with 10 % (v/v) acetic acid, 12.5 % (v/v) methanol.

2.3.3- Protein expression

Target genes cloned into the pET series of plasmids are under the control of the strong T7 bacteriophage promoter. Expression of the recombinant protein is induced by addition of isopropylthio-β-D-galactoside (IPTG) to the bacterial culture. The pET21a^{+} vector allows the addition of a Histidine tag (6 histidine residues) to the C-terminus of the protein to facilitate detection and purification.

2.3.3.1- Cloning of proline dehydrogenase into pET21a$^+$

After gel purification using a QIAquick PCR gel purification kit (Qiagen), the PCR products were cloned into the pGEM-T vector as described in section 2.2.5.3, and then into pET21a$^+$, once the quality of the insert had been checked by restriction enzymes digestion and by sequencing.

2.3.3.2- Protein expression

The positive pET21a$^+$ vectors containing the *PRODH* genes were used to transform competent *E. coli* cells (strain BL21(DE3)) by heat shock. Positive cells were selected by plating into LB plates containing ampicillin (100 µg/ml). Single colonies were picked and seeded into liquid medium. From an overnight culture of 5 ml of LB medium containing ampicillin, sample was diluted 100 fold into fresh LB medium (50 ml to 500 ml) containing 100 µg/ml of ampicillin and incubated at 37 °C until OD_{600} equalled approximately 0.6. Different concentrations of IPTG (0.1 mM-1 mM) were added to induce protein expression and the cells were shaken for another 4 hours at 37 °C, or overnight (O/N) at 16 °C or 20 °C. Cells were harvested by centrifugation (4 °C, 4500 g, 10 min) and the pellet was resuspended into appropriate buffer, depending on the experimental conditions, sonicated (20 times with 20 s pulse (frequency amplitude from 16 to 22 microns) separated by 30 s intervals between pulses) (Soniprep150, MSE) and centrifuged at 10,000 g for 30 min at 4 °C. The two subsequent fractions, pellet and supernatant, were analysed using a 10 % (v/v) SDS-PAGE gel. Soluble recombinant protein was then purified using a Ni-agarose column (Qiagen). The presence of recombinant protein was checked by Western Blot analysis using an anti His-tag antibody (see section 2.3.7).

2.3.3.3- Solubilisation of inclusion bodies

After expression of the recombinant protein, the pellet of *E. coli* was resuspended in 2.5 ml of 50 mM Tris-HCl, pH 7.5, 5 mM EDTA, 5 % (w/v) sucrose and stored for 1 hour at –20 °C, then defrosted in tap water for 10 min and refrozen for 1 hour at –20 °C. The pellet was sonicated (6 to 20 times, 10 s to 30 s pulse (frequency amplitude from 16 to 22 microns) separated by 30 s intervals between pulses) (Soniprep150, MSE) and centrifuged 10,000 g for 10 min at 4 °C. The resulting pellet was resuspended in 2.5 ml of 50 mM Tris-HCl, pH 7.5, 5 mM EDTA containing 0.1 % Triton X100 and centrifuged at 10,000 g for 10 min at 4 °C. The pellet was resuspended in 2.5 ml 50 mM Tris-HCl, pH 7.5, 5 mM EDTA containing 2 M urea and centrifuged 10,000 g for 10 min at 4 °C. The final pellet was resuspended in 5 ml of the same buffer containing 8 M urea and slowly shaken at 37 °C for 1 hour and then diluted with 5 ml 50 mM Tris-HCl, pH 7.5, 5 mM EDTA, ± 10 mM DTT and 8 M urea (Sanderson *et al.*, 2000). The recombinant protein released from the inclusion bodies was dialysed against different buffers before testing for activity.

2.3.4- Purification of the recombinant protein

2.3.4.1- Ni-Agarose column

The recombinant PRODH produced has a 6-histidine tag engineered at the C-teminus, allowing affinity purification on a nickel agarose column (Qiagen). Samples (containing soluble recombinant proteins) were added to the pre-equilibrated Ni^{2+} column and allowed to move through the matrix under gravity. Once the sample had entered the column, the Ni^{2+}-beads were washed with 10 ml of 10 mM Tris-HCl pH 7.8, 5 mM imidazole, 0.5 M NaCl, 10 % (v/v) glycerol, and then 5 ml of wash buffer (same buffer containing 30 mM imidazole). Recombinant protein was selectively eluted using 0.5 ml or

1.5 ml aliquots of the buffer containing 500 mM imidazole. The eluted fractions were dialysed against different buffers and checked for enzymatic activity. The samples eluted were also analysed by SDS-PAGE gel and / or Western blots.

2.3.4.2- Use of BioCAD

The protein containing the 6-histidine tag was also purified using a BioCAD 700 E workstation (PE Biosystems). The column used was POROS MC 4.6 mm that had a column size of 1.7 ml. The POROS beads were charged with nickel ions. Alan Scott (University of Glasgow) purified the proteins using the BioCAD system. The column was first equilibrated with Tris or phosphate buffer containing 0.5 mM imidazole. Samples were loaded onto the column with a flow rate of 5-10 ml/min. The column was washed with 20 ml of Tris or phosphate buffer (pH 7.8) containing 0.5 mM imidazole, a second wash of 15 ml of Tris or phosphate buffer containing 30-50 mM imidazole was used. The protein was eluted either by a gradient of 50-500 mM imidazole with the eluent being collected in 15 fractions of 1 ml or simply with a solution of 500 mM imidazole. The absorbance at A_{280} was followed to monitor the protein elution.

2.3.5- Antibody production

Purified recombinant PRODH protein was used to raise antiserum in rabbits, this was done by Diagnostic Scotland using standard techniques. A rabbit was immunised with 400 µl of a 1 mg/ml solution of the recombinant protein. Over the next three months, the rabbit received a further three 200 µl injections of the same solution at monthly intervals. Blood samples were taken each month before sacrificing the rabbit one month after the final injection. Samples were analysed by dot blot or by Western blot analysis to examine the antibody purity and affinity for PRODH.

2.3.6- Antibody purification

Antibodies produced were purified using AminoLink® (Pierce) kit following the manufacturer's instructions. The first step of this purification was to immobilize the recombinant PRODH on a solid support. 2 ml of the protein (1-20 mg/ml) was incubated with 1 ml of AminoLink® coupling gel overnight at 4 °C after adding 200 µl of sodium cyanoborohydride ($NaCNBH_3$) (190 mg/ml), which acts as a reductant and permits the formation of covalent bounds. After washing the column with PBS containing 0.05 % NaN_3, the sample of antiserum was applied and incubated with the column under gentle agitation at 4 °C overnight. After washing the column with PBS containing 0.05 % NaN_3, antibodies were eluted by 0.5 ml fractions of glycine buffer (100 mM, pH 2.5). Purified antibodies were aliquoted and stored at – 20 °C until use.

2.3.7- Western blotting

Samples were resuspended in 1X loading buffer (Appendix 1) and proteins separated by electrophoresis in an 10 % SDS-PAGE gel. Separated products were then electrotransfered onto a nitrocellulose membrane using a Bio-Rad electrophoretic transfer cell (100 V, 1 hour). The membrane was blocked for 3 hours at room temperature with 1X TBS (Appendix 1) containing 5 % (w/v) powdered milk and 0.2 % (v/v) Tween-20. The primary antibody, anti-histidine-tag (diluted 1:2000) or specific PRODH antibody (diluted 1:500 - 1:5000) was made up in 1X TBS, 1% (w/v) powdered milk and 0.1 % (v/v) Tween-20, before being added to the membrane and shaken overnight at 4 °C. The SuperSignal Chemiluminescent Substrate (Pierce) protocol with 1:2000 diluted anti-mouse IgG conjugated to peroxidase or with 1:500 - 1:5000 anti-rabbit horseradish peroxidase (HRP)

IgG (Sigma) allowed the detection of the proteins according to manufacturer's specifications.

2.3.8- Dot blotting

Antibodies have specific affinity for the antigen against which they were raised. Dot blotting permits antibody quantification without size fractionation (electrophoresis). Serial dilutions of 0.3-0.1-0.03 µg of the recombinant *T. brucei* PRODH (as determined by Bradford assay) were dotted onto a nitrocellulose membrane (4 µl each dot). As the same time a negative control was dotted onto the membrane. The membrane was allowed to dry and then blocked for 3 hours at room temperature by incubation with a solution containing 1 X TBS, 0.2 % Tween 20, 5 % (w/v) powdered milk. Several dilutions of the primary antibody were made (from 1:500 to 1:5000) in 1X TBS, 0.1 % Tween 20, 1 % powdered milk and incubated overnight at 4 °C. The following morning, membranes were washed 4 times for 30 min in 1X TBS, 1 % powdered milk. The secondary antibody, 1:2000 anti-rabbit HRP IgG (Sigma), allowed the detection of the proteins according to manufacturer's specifications.

2.3.9- Proline dehydrogenase assay

The cofactor flavin adenine dinucleotide (FAD) is required for the enzymatic dehydrogenation of L-proline to pyrroline-5-carboxylate (P5C). Two electrons are transferred from proline to the cofactor. In the assay used, the electrons from the reduced FAD are then transferred to an electron acceptor, which changes colour upon reduction (Becker and Thomas, 2001). PRODH activity was measured following the reduction of the electron-accepting dye, iodonitrophenyl tetrazolium (INT), by monitoring the absorbance at 520 nM. The extinction coefficient for INT is E_M= 11.5 x 10^3 $M^{-1}.cm^{-1}$ (Menzel and

Roth, 1981 a). Activity can also be determined using dichlorophenolindophenol (DCPIP) as a terminal electron acceptor (E_M= 21 x 10^3 $M^{-1}.cm^{-1}$) (Johnson and Strecker, 1962). The reduction of the dye is associated with a decrease in absorbance at 600 nm (Brown and Wood, 1993). One unit of enzyme activity was defined as the quantity of enzyme that transfers electrons from 1 µmol of proline to DCPIP in one min at 25 °C.

A log phase culture (5 x 10^6 cells/ml) of *T. brucei* procyclic cells was harvested by centrifugation (2000 g, 5 min at 4 °C). The resultant pellet (~ 2 x 10^8 cells) was washed twice with a phosphate buffered saline (PBS) (Appendix 1) (pH 7.9) before being resuspended in 500 µl of TSE buffer (Appendix 1). The mixture was sonicated 3 times on ice (for 10 s on, 30 s off) and the crude lysate tested for PRODH activity (Obungu *et al.*, 1999; Besteiro *et al.*, 2002).

The reaction mixture had a total volume of one ml and contained: 0.2 M proline; 16 % (v/v) ethylene glycol; 0.4 % (v/v) Tween 20; 0.16 M Tris-HCl, pH 8.5; 0.04 mg of gelatin and 0.5 mM INT. The reaction was initiated by addition of 1-50 µl of cell lysate and carried out at room temperature. Activity was monitored in cuvettes with a 1 cm light path at 520 nM using a Hewlett packard 8453 spectrophotometer, for 30s to 10 min.

For the second assay method, the stock solution contained: 11 mM MOPS; 11 mM $MgCl_2$; 11 % (v/v) glycerol; 0.28 mM phenazine methosulfate and 56 µM of dichlorophenolindophenol (DCPIP), pH 7.5. The solution was stored in a dark bottle to exclude light. 1-50 µl of 2 M proline or 0.2 M proline was added to 900-950 µl of the stock assay mix and the reaction was started by adding the enzyme (1 to 50 µl). Activity was monitored in cuvettes with a 1 cm light path at 600 nM. The protein concentration in the crude lysates of the parasites was determined by Bradford's method. In both assays, a negative control (no substrate) was performed to define background reduction of the dye in

the absence of proline. These assays were also used to test the activity of the recombinant protein.

2.4- RNA interference (RNAi) in *T. brucei*

RNA interference refers to the introduction of homologous double stranded RNA (dsRNA) to specifically target a gene product (mRNA), resulting in "null" phenotypes (LaCount *et al.*, 2000).

The RNAi construct used in this study places the gene or sequence of interest between opposing T7 promoters in the p2T7Ti vector (Figure 2.2) (LaCount *et al.*, 2000). This vector contains an rRNA spacer for integration into the rRNA locus of the parasite's genome, as well as a phleomycin-resistance gene for selection and a tetracycline-inducible operator. The p2T7Ti vector must be used with specially derived *T. brucei* lines, which express T7 RNA polymerase coli phage T7 and the tet repressor from bacteriophage λ. To this end, the transgenic 29-13 procyclic cell lines (which were originally derived from the EATRO 427 line) were used. The T7 RNA polymerase and the tetracycline repressor constructs are maintained within the parasite by the addition of 15 μg/ml of G418 and 25 μg/ml of Hygromycin B to the medium respectively. Without tetracycline, the tet repressor is expressed and it binds to the tet operator to inhibit the transcription. When tetracycline is present, it binds to the repressor and prevents its action on the operator and so transcription occurs.

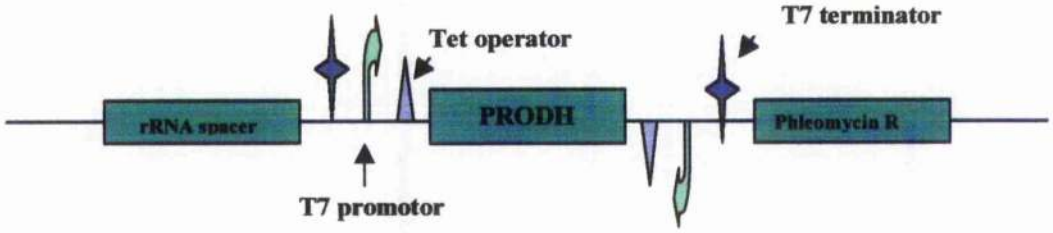


Figure 2.2: Inducible two-T7 promoter vector (p2T7Ti). PRODH represents the PCR gene fragment of proline dehydrogenase; Phleomycin R is the phleomycin-resistance gene.

2.4.1- Parasites

T. brucei 29-13 (Writz *et al.*, 1999) PCF were grown in SDM 79 medium supplemented with 10 % (v/v) heat inactivated foetal calf serum, 15 μg/ml of G418 and 25 μg/ml of Hygromycin B at 25 °C. 24 hours after transfection (see section 2.4.3), 10 μg/ml of phleomycin was added to the medium for selection of transfected clones with the PRODH construct. Induction of the RNA expression was achieved by addition of 1 to 10 μg/ml of tetracycline.

2.4.2- Plasmid construction

2.4.2.1- PCR of a fragment of the T. brucei proline dehydrogenase gene

A 606 bp fragment of the *PRODH* gene was amplified by PCR from *T. brucei* gDNA using specific primers each containing an *Xba* I linker (see Figure 2.3 and Table 2.1).

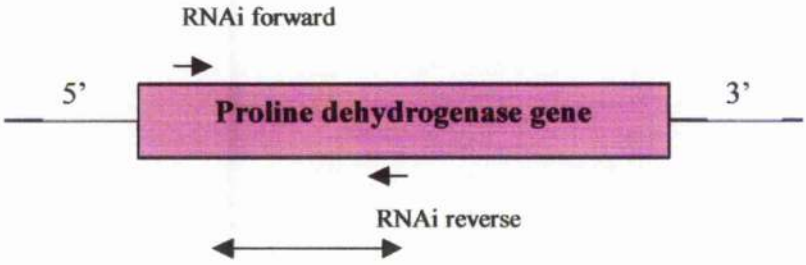


Figure 2.3: Cartoon showing primer binding sites used to amplify a fragment of proline dehydrogenase for insertion into the RNAi construct.

The final composition of the PCR reaction was: 1X PCR buffer (Promega), 0.5 mM PCR nucleotides mix, 1 unit of pfu DNA polymerase, 100 pmol of each primer, 5 % (v/v) DMSO and 50 ng *T. brucei* gDNA. The PCR cycling protocol was performed in three stages: (94 °C, 2 min); (94 °C, 15 s; 63 °C, 30 s; 72 °C, 2 min; 30 cycles); (72 °C, 7 min).

2.4.2.2- Cloning into p2T7Ti

The PCR products were run on a 1 % agarose gel and the 606 bp fragment identified. The product was extracted from the gel using QIAquick PCR purification kit (Qiagen). This PCR product and the plasmid p2T7Ti were digested for 2 hours at 37 °C with the *Xba* I restriction enzyme before being purified using the Qiagen kit. The digested insert and the vector were ligated overnight at 16 °C in the following reaction mix: 1 µl 10X ligase buffer; 1 µl p2T7Ti vector (20 ng/µl); 7 µl PCR product (130 ng); 1 µl T4 DNA ligase at 1 U/µl. *E. coli* competent cells (strain DH5 α) were transformed with 1 µl of the

ligation mixture using a standard heat shock method (42 °C for 50 s). Cells expressing the construct were selected for growth on LB plates containing ampicillin (50 μg/ml). Clones were picked and cultured at 37 °C overnight in LB medium containing ampicillin (50 μg/ml). The isolation of the plasmid was performed using the QIAprep Miniprep kit (Qiagen). To check for the presence of the correct insert, plasmids were digested with *Xba* I and fragments separated by agarose gel electrophoresis.

2.4.3- Transfection of the parasites

In order to linearise the plasmid containing the insert, 100 μg of DNA (p2T7Ti containing the proline dehydrogenase gene fragment, p2T7PRODH) was digested for 2 hours at 37 °C by *Not* I. DNA was then ethanol-precipitated and air dried in a sterile hood to sterilise the sample. It was then resuspended in 20 μl sterile water. Product size and concentration were confirmed on a 1 % agarose gel.

For transfection, ~ 1 x 10^8 29-13 *T. brucei* cells were washed twice in 4 ml of Zimmerman's post fusion media (ZPFM) (Appendix 1) before being resuspended in 2 ml of ZPFM. 0.5 ml of this suspension was mixed with ~8 μg of the sterile linearized plasmid. Transfection was performed by electroporation in a 0.4 cm cuvette, using a single pulse at a voltage of 1.5 kV and 25 μF capacitance with an infinite resistance using a BioRad Gene Pulser. After transfection, the parasites were transferred into pre-warmed SDM79 medium supplemented with 10% (v/v) heat inactivated foetal calf serum, 15 μg/ml of G418 and 25 μg/ml of Hygromycin B at 25 °C. 24 hours later 10 μg/ml of phleomycin was added to the medium to select stable expressing cell lines. To induce the expression of dsRNA, tetracycline was added to the medium at a concentration of 1 μg/ml.

2.4.4- Selection of clones

In order to assist growth of parasites during cloning, conditioned medium was prepared as follows. Procyclics were grown at 27 °C in SDM 79 medium, supplemented with 10 % (v/v) FCS to a density of approximately 1 x 10^7 cells/ml. The culture was pelleted and the supernatant filter sterilised and stored at 4 °C. The cloning was performed in 96 well plates by limiting dilution of the culture obtained after electroporation. This involved diluting the culture to 1 cell/ml and adding 0.2 ml per well (i.e. one cell per five wells on average). Cells were cultured in SDM79 medium containing 10 % (v/v) FCS, 25 % (v/v) conditioned medium complemented with selective antibiotics. Once clones were obtained, growth curves and uptake assays were performed on cells that were grown in SDM79 containing 10 % (v/v) FCS in the presence or absence of tetracycline (1 µg/ml).

2.4.5- Analysis of growth rates

A single clone (Δ*prodh*, clone F2) was selected for further analysis. Cultures were initiated using 5 × 10^5 parasites/ml in the presence or absence of tetracycline (1 µg/ml). 10 ml of complete SDM79 medium was used or 10 ml of a defined medium made without proline and glucose and then supplemented with proline and/or glucose at different concentrations as stated in the text. Parasite density was measured every 24 hours with an improved Neubauer haematocytometer (Weber Scientific). Growth was also analysed by using the Alamar® blue assay (see section 2.1.3).

2.4.6- Amino acids uptake assays

2.4.6.1- Preparation of parasites for uptake assays

Clone Δ*prodh* F2 generated using the RNAi constructs was used in the uptake analyses. It was grown in SDM79 medium in the presence of the antibiotics G418, Hygromycin B and phleomycin at the same concentrations as stated above. Experiments were carried out on parasites grown in the absence or presence of tetracycline (1 μg/ml). Parasites were harvested during the mid-log phase of growth by centrifuging at 3,000 g for 10 min at 4 °C, before being washed twice in CBSS buffer at 4 °C. The pellet was resuspended in CBSS at the density of 10^8 parasites/ml (which gives 10^7 parasites per assay).

2.4.6.2- Stock solutions

The radiolabelled amino acids proline, L-[2,3,4,5-^{3}H] (112 Ci/mmol) and glutamate, L-[3,4-^{3}H] (51 Ci/mmol) (NEN Life Science Products) were diluted in order to have 10 μl of radiolabelled compound per 1 ml of solution (10 μCi/ml). This corresponds to 190 nM of ^{3}H glutamate and 89 nM of ^{3}H proline. Other test compounds, including inhibitors or cold amino acids, were added from stock solutions for testing at the concentration detailed in the relevant Results sections.

2.4.6.3- The oil stop transport assay

All of the solutions and cells were brought to room temperature, or the temperature specified in the text, prior to initiating the uptake experiments. 100 μl of the test solution, containing the radiolabelled compounds at concentrations marked in the results sections, were added to a 0.5 ml Eppendorf tube and cushioned above 90 μl of oil mixture (di-n-butylphthalate and mineral oil at a ratio of 7:1) before being centrifuged (12 000 g, 1 min

at room temperature) to produce two distinct layers. For the uptake assay, 100 µl of cells (at 10^8 cells/ml) were added to each Eppendorf and incubated for times ranging between 3 s and 3 hours depending on the experiment. The transport assays were terminated by centrifugation (12 000 g, 1 min at room temperature) to pellet the cells beneath the oil layer, and thus separating them from the remainder of the radioactive solution. Samples in the tubes were flash frozen in liquid nitrogen before removing the pellets using a tube cutter and transferring them to scintillation vials containing 200 µl of 2 % (w/v) SDS to lyse the cells. Three ml of scintillation fluid (Ecoscint A, National Diagnostics) was added before overnight incubation. Radioactivity levels were determined by a LKB wallac 1219 Rackbeta liquid scintillation counter.

Background levels, which correspond to the non-transported radiolabelled compounds associated with the cells (in the interstitial space), were measured by performing uptake assays on ice.

2.4.6.4- Analysis of the results

The data were analysed using the PRISM software. Kinetic constants were determined by non-linear regression analysis using the Michaelis-Menten equation.

2.5- Bioinformatic Analyses

Vector NTI (v 6.0) was used to analyse DNA and protein sequences (searching for open reading frames and restriction sites, construction of cloning and expression plasmids). Sequence alignments were studied with Align X (Clustal X) and ContigExpress was used to align overlapping sequences to form contiguous sequences.

The protein alignment was re-aligned using ClustalX1.81 (Thompson *et al.*, 1994). Clustal X is an alignment program for nucleotide or amino acid sequences. It uses a progressive alignment algorithm to align sequences. It aligns more similar sequences before the distant sequences. The Phylogenetic trees for this study were generated using MEGA (Molecular Evolution Genetic Analysis) software version 2.1 (Kumar *et al.*, 2001).

The following web resources were used to analyse nucleic acid and amino acid sequences

Databases:

TIGR databases: http://www.tigr.org

The Sanger Institute: http://www.sanger.ac.uk

Parasite genome: http://www.ebi.ac.uk

NCBI: http://www.ncbi.nlm.nih.gov

GeneDB: http://www.genedb.org

Analysis of the amino acid sequences in order to predict a signal peptide were done by:

SignalP: http://www.cbs.dtu.dk/services/SignalP

Predotar: http://www.inra.fr/Internet/produits/predotar

Mitoprot II: http://bioinformer.ebi.ac.uk

Target P v 1.0: http://www.cbs.dtu.dk/Service/targetP

Some of the most popular tools for protein localisation use neural network analysis, which is a self-teaching network of computational units that learns by a process of trial and error. SignalP (Nielsen *et al.*, 1996) uses two neural networks, one to score the likelihood of a

subsequence to be a part of a signal peptide and another one to score the likelihood of a subsequence to be a cleavage site. These two scores are combined. Predotar (INRA) and TargetP (Emanuelsson *et al.*, 2000) also use neural network predicting the presence or absence of a signal sequence. TargetP can also distinguish between secretory, mitochondrial and chloroplast signals. MitoProt (Claros and Vincens, 1996) compares a sequences' discriminant function of cut-offs to predict mitochondrial localisation.

Prediction of transmembrane domains within amino acid sequences was done by:

TMpred: http://www.ch.embnet.org

The TMpred program makes predictions of membrane-spanning regions and their orientation. This program is based on the TMbase (Hofmann and Stoffel, 1992), which is a data base combining several matrices for scoring. TMbase is mainly based on Swissprot. All data are scored in different tables able to be used with a database management system.

Conserved Blocks inside the PRODH amino acid sequences were found using

MEME: http://www.meme.sdsc.edu

MEME is a software toolkit for building and using motif-based hidden Markov models of biological sequences. It produces a multiple alignment of the original set of sequences and to search a sequence database for homologues.

Chapter 3

Identification of genes encoding proline dehydrogenase in *Leishmania major* and *Trypanosoma brucei* and bioinformatic analysis of the predicted proteins

3.1- Introduction

Both *Leishmania major* and *Trypanosoma brucei* undergo a complicated developmental cycle involving an insect vector and a mammalian host. It has been demonstrated that with some strains of *T. brucei* in culture, the amino acid proline is the favoured metabolic substrate (Gutteridge and Coombs, 1977; Evans and Brown, 1972). Proline is also considered to be a major substrate for energy and carbon synthesis in promastigotes of *Leishmania* species (Zilberstein and Gepstein, 1993). These *in vitro* findings with *T. brucei* correlate well with the normal environment of procyclic *T. brucei* in the tsetse fly midgut, where the concentration of proline is high (Balogun, 1974). Proline is the major energy source of insect flight muscles (Cazzulo, 1992; Obungu *et al.*, 1999).

Proline utilisation requires two enzymes to convert proline into glutamate; proline dehydrogenase (PRODH) and pyrroline-5-carboxylate dehydrogenase (P5CDH) (Figure 3.1). In the first step, the oxidation of proline to pyrroline-5-carboxylate (P5C) is coupled with the reduction of a cofactor, flavin adenine dinucleotide (FAD). Two electrons from reduced FAD are then transferred to the electron transport chain. In the second step, P5C is hydrolysed to give glutamic semialdehyde, which is oxidised to glutamate using the cofactor nicotinamide adenine dinucleotide (NAD^+) (Becker and Thomas, 2001; Lee Y *et al.*, 2003). This proline metabolism permits oxidation of proline to glutamate, which is

further oxidised to CO_2 through the Krebs cycle, while protons are donated to the electron transport chain (Obungu *et al.*, 1999).

In all eukaryotic cells that have been studied to date (e.g. *Saccharomyces cerevisiae, Drosophila melanogaster, Homo sapiens*) the conversion of proline to glutamate takes place within the mitochondrion (Brunner and Neupert, 1969; Small and Jones, 1990; Maxwell and Davis, 2000; Deuschle *et al.*, 2001). In bacteria, such as *Escherichia coli,* proline metabolism enzymes are associated with the inner plasma membrane (Muro-Pastor *et al.*, 1997), which is functionally equivalent to the inner mitochondrial membrane of eukaryotes.

Several studies have been performed on the proline metabolism of trypanosomatids (Evans and Brown, 1972; ter Kuile and Opperdoes, 1992a and 1992b; Clayton and Michels, 1996; Obungu *et al.*, 1999). Nevertheless, the gene encoding the first enzyme of this pathway (*PRODH*) had not been cloned or studied in these organisms. This chapter presents the identification of the genes coding for *L. major* and *T. brucei* PRODH, as well as a bioinformatic analysis of the predicted proteins encoded by these genes.

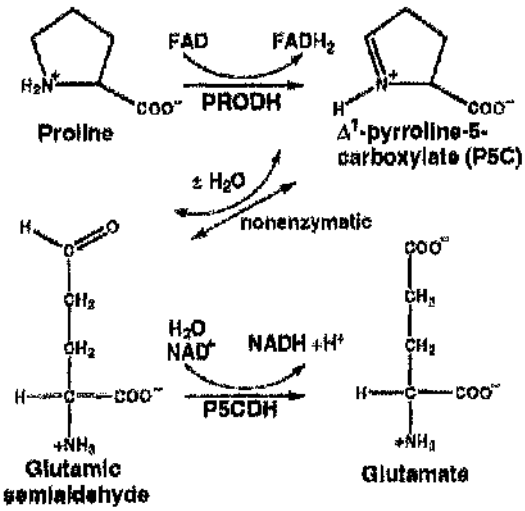


Figure 3.1: Proline degradation pathway (from Lee *et al.*, 2003)
PRODH: proline dehydrogenase; P5CDH: pyrroline-5-carboxylate dehydrogenase

3.2- Results

3.2.1- Identification of the genes

The National centre for Biotechnology Information (NCBI) and *L. major* databases were screened using the name of the enzyme "proline oxidase", now renamed proline dehydrogenase (PRODH). A peptide sequence corresponding to a putative PRODH in *L. major* (accession number CAB 97967) was found. It is present on chromosome 26. The complete sequence of this chromosome is present in the database under the accession code AL 160493 and CAB 97967 is a part of this sequence. The BLAST (Basic Local Alignment Search Tool) servers from TIGR and the parasite genome database were then searched with the *Leishmania* gene to find a "*PRODH*" homologue in *T. brucei*. The

accession numbers of the *T. brucei* sequences with homology to the *L. major* proline oxidase gene, were: AQ 944712; AQ 638820; AQ 653997; AL 495302; AL 453746. These DNA sequences were used to generate a contiguous sequence (Contig) (using the Contig Express software of Vector NTI) and a consensus sequence was produced (Figure 3.2). The contiguous nucleic acid sequence was then used to probe other databases using the BLAST algorithm, in order to verify the identity of the *PRODH* gene sequence.

This preliminary study was performed in 2000. Since then, the databases have been updated and all of the accession numbers cited above have been suppressed. The GeneDb accession number, 105.m00209, is now annotated as a putative proline oxidase present on chromosome 7 of *T. brucei*.

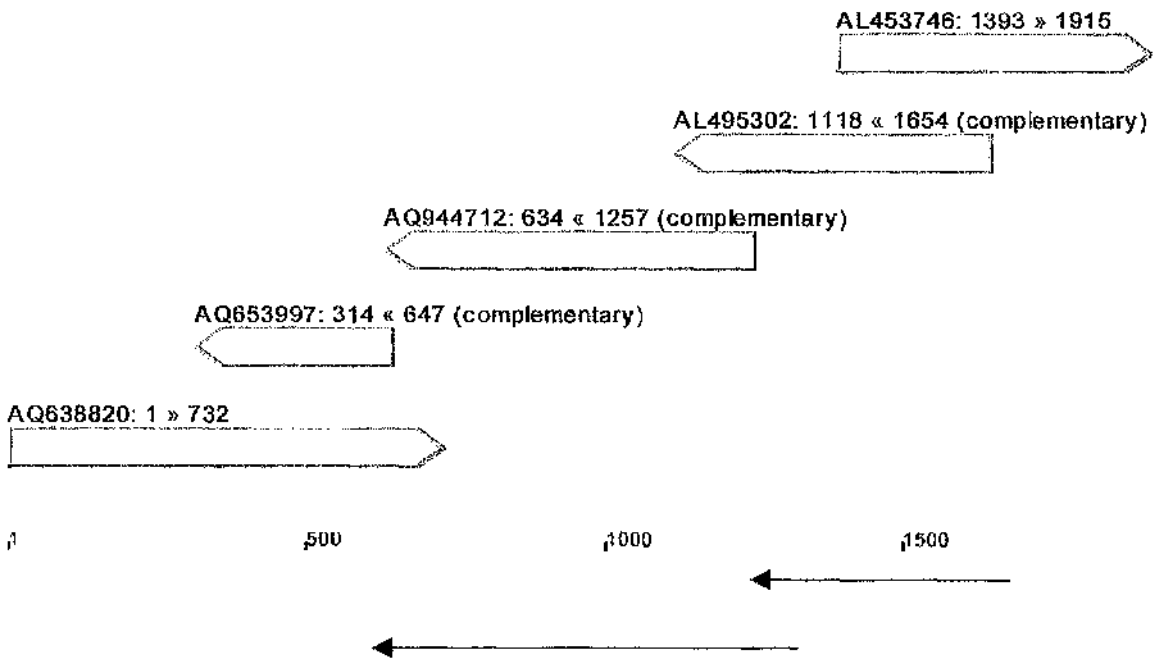
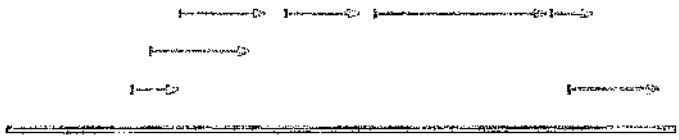


Figure 3.2: Contig assembled with the accession number references from the *T. brucei* genome project. This contig produced overlapping sequences. The proline dehydrogenase ORF was then deduced from this sequence. Arrows represent the putative ORF (with a frame shift in the raw sequence data).

The two putative *PRODH* gene sequences from *L. major* and *T. brucei* were analysed using the Vector NTI software in order to find a putative open reading frame (ORF). The nucleotide data in the *L. major* and *T. brucei* database were incomplete at the time when the contigs were first generated and there were numerous undefined bases and likely frame-shifts. This made it difficult to obtain an accurate open reading frame sequence (Figure 3.3). Therefore it was essential to re-sequence this portion of the genome in order to find the authentic first methionine and stop codon.



Putative ORFs present in the assembled contig of *L. major* (2690 bp)

Putative ORFs present in the assembled contig of *T. brucei* (1914 bp)

Figure 3.3: Apparent ORFs from DNA sequences in the databases in 2000.
The solid bars represent the *L. major* and *T. brucei* contigs containing the putative *PRODH* genes. The arrows show predicted open reading frames judged by 5' methionine and 3' termination codons.

Genomic DNA from both species was amplified using *pfu* DNA polymerase and oligonucleotides (Lm1 and Lm1 rev for *L. major* and Tb1 and Tb1 rev for *T. brucei*) shown in section 2.2.3. The primers were designed to amplify sequences upstream and downstream of the putative ORFs. PCR products obtained are shown in Figure 3.4. The amplification for *L. major* gave a PCR product of approximately 2.2 kb, and for *T. brucei* a product of approximately 2.1 kb was obtained. These were the sizes expected based on genomic sequence in the databases.

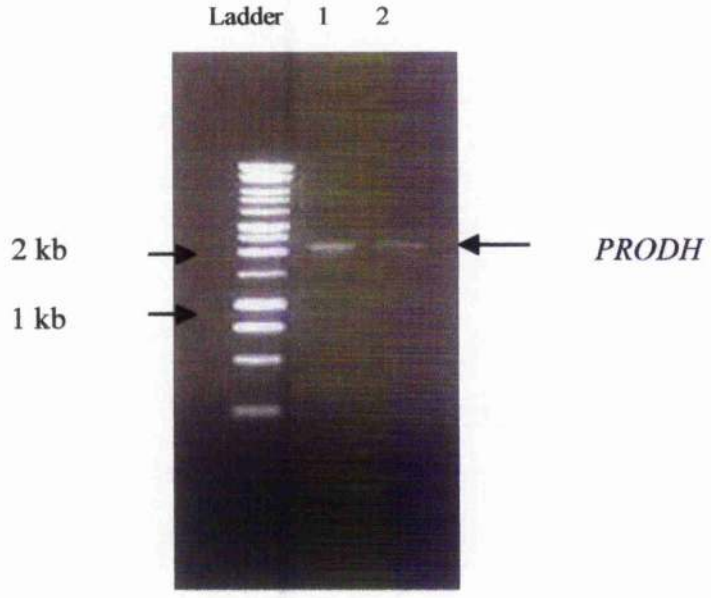


Figure 3.4: Ethidium bromide stained gel showing PCR amplification of *T. brucei* (*TbPRODH*) and *L. major* (*LmPRODH*) putative proline dehydrogenases

1: 1 kb DNA ladder

2: *T. brucei* amplification product

3: *L. major* amplification product

In order to check that the correct PCR product had been amplified, it was cloned into the pGEM-T vector and each sample was then digested using two different restriction enzymes, *Xba* I and *Nde* I for *L. major* and *BamH* I and *Apa* I for *T. brucei* (Figure 3.5).

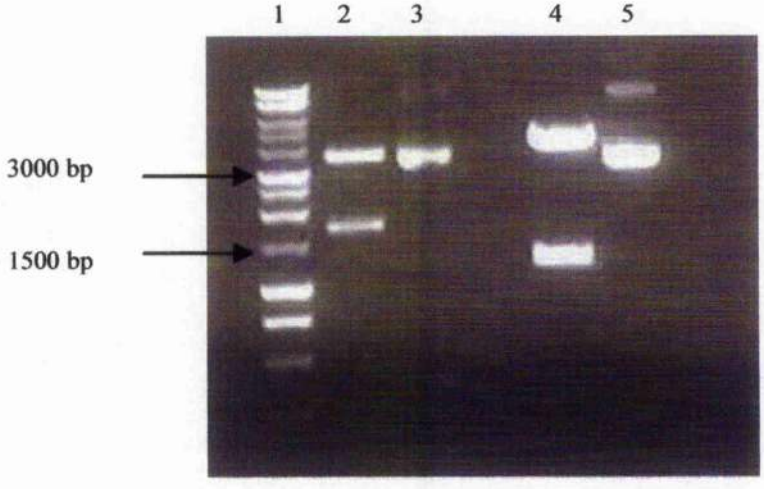


Figure 3.5: Digestion of *LmPRODH* and *TbPRODH* cloned into pGEM-T vector

1: 1 kb ladder

2: PCR product corresponding to the *L. major* putative proline dehydrogenase gene within pGEM-T vector, digested with *Xba* I and *Nde* I

3: Non-digested pGEM-T containing *L. major* proline dehydrogenase

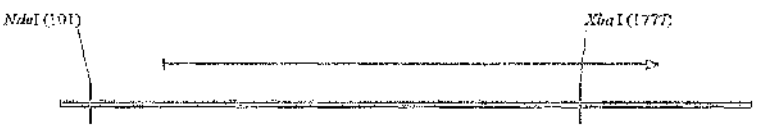
4: PCR product corresponding to the *T. brucei* putative proline dehydrogenase gene within pGEM-T vector, digested with *BamH* I and *Apa* I

5: Non-digested pGEM-T containing *T. brucei* proline dehydrogenase

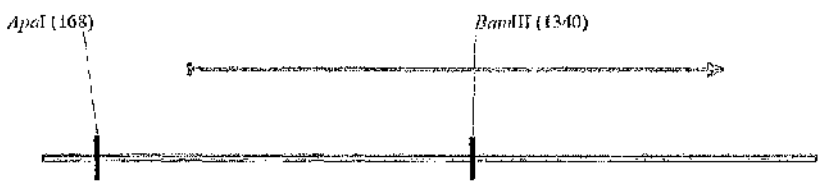
In order to check the construct, the pGEM-T vector containing the putative *PRODH* from *L. major* was digested by *Xba* I (which cuts at position 1777 in the insert) and *Nde* I, which cuts in the vector, while the pGEM-T vector containing the *T. brucei* gene was digested by

BamII I (which cuts at position 1340 in the insert) and *Apa* I, which cuts in the vector. The digested products obtained (Figure 3.5) were of the predicted sizes (Figure 3.6). A fragment of 1676 bp was obtained from the *L. major PRODII* digestion and a fragment of 1668 bp was obtained from the *T. brucei PRODH* digestion. The vectors containing the inserts were then sequenced.

Analysis of these sequences allowed the true ORFs to be defined, based on homology between *L. major*, *T. brucei* and other PRODH proteins (Figure 3.6). The ORF amplification is shown in Figure 3.7.



L. major ORF (1683 bp) map



T. brucei ORF (1668 bp) map

Figure 3.6: Restriction map of *T. brucei* and *L. major* putative proline dehydrogenase gene sequences cloned in the pGEM-T vector. The blue and green arrows represent the putative open reading frame found after amplification of the gDNA and sequencing.

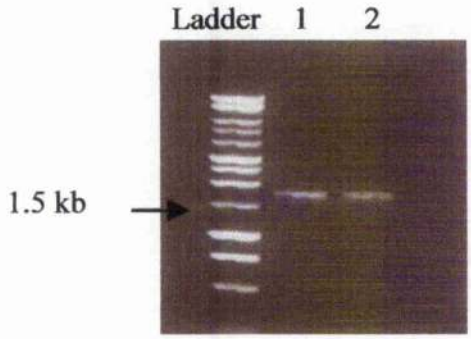


Figure 3.7: Ethidium bromide stained gel showing PCR amplification of *T. brucei* and *L. major* proline dehydrogenase genes from gDNA

1: *T. brucei* proline dehydrogenase gene

2: *L.major* proline dehydrogenase gene

Figures 3.8 and 3.9 show the nucleotide sequences translated into corresponding proteins using Vector NTI. These ORFs encode a protein of 556 amino acid residues with a calculated molecular mass of 63.8 kDa for *T. brucei* and a protein of 561 amino acid residues with a molecular mass of 63.8 kDa for *L. major*.

```
          M   F   R   L   L   C   R   R   S   A   A   N   V   K   M   Q   D   A   S   L
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
    1  ATG TTT CGC CTC CTG TGC CGA CGC TCC GCC GCG AAT GTA AAA ATG CAA GAC GCC AGT CTC
       TAC AAA GCG GAG GAC ACG GCT GCG AGG CGG CGC TTA CAT TTT TAC GTT CTG CGG TCA GAG
        R   T   M   T   K   V   D   F   S   D   P   S   I   F   R   Q   K   S   L   W
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
   61  CGC ACC ATG ACA AAA GTG GAT TTC TCG GAC CCA TCC ATA TTT CGT CAG AAA TCT CTG TGG
       GCG TGG TAC TGT TTT CAC CTA AAG AGC CTG GGT AGG TAT AAA GCA GTC TTT AGA GAC ACC
        W   L   L   R   A   L   F   V   L   R   I   C   K   F   E   F   I   S   N   N
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  121  TGG TTG TTG CGC GCA CTC TTT GTG CTC CGC ATT TGC AAG TTT GAA TTC ATT TCC AAC AAT
       ACC AAC AAC GCG CGT GAG AAA CAC GAG GCG TAA ACG TTC AAA CTT AAG TAA AGG TTG TTA
        S   V   A   L   M   K   R   A   E   A   I   F   G   P   F   L   T   Y   N   T
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  181  TCT GTG GCT CTT ATG AAG CGT GCT GAG GCA ATT TTT GGA CCC TTT TTG ACC TAT AAC ACA
       AGA CAC CGA GAA TAC TTC GCA CGA CTC CGT TAA AAA CCT GGG AAA AAC TGG ATA TTG TGT
        L   V   K   G   T   V   Y   G   H   F   C   A   G   E   S   D   R   E   V   K
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  241  CTC GTG AAG GGG ACA GTG TAC GGC CAC TTT TGT GCT GGC GAA TCC GAT CGT GAA GTT AAG
       GAG CAC TTC CCC TGT CAC ATG CCG GTG AAA ACA CGA CCG CTT AGG CTA GCA CTT CAA TTC
        N   T   V   K   S   L   E   N   L   G   I   G   S   V   L   D   Y   A   A   E
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  301  AAC ACC GTC AAA TCG CTT GAA AAT TTG GGT ATT GGA TCT GTT CTT GAC TAT GCC GCT GAA
       TTG TGG CAG TTT AGC GAA CTT TTA AAC CCA TAA CCT AGA CAA GAA CTG ATA CGG CGA CTT
        A   E   A   E   G   F   A   P   S   P   G   I   A   E   A   P   N   L   S   M
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  361  GCT GAA GCT GAG GGT TTC GCA CCC TCC CCA GGT ATT GCA GAA GCG CCG AAC TTG TCT ATG
       CGA CTT CGA CTC CCA AAG CGT GGG AGG GGT CCA TAA CGT CTT CGC GGC TTG AAC AGA TAC
        A   S   L   V   N   N   T   S   V   T   Y   L   P   H   K   Q   A   F   D   E
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  421  GCA AGT CTT GTC AAC AAC ACA TCC GTT ACC TAT CTG CCT CAC AAG CAA GCC TTT GAC GAA
       CGT TCA GAA CAG TTG TTG TGT AGG CAA TGG ATA GAC GGA GTG TTC GTT CGG AAA CTG CTT
        N   M   K   L   Y   V   M   C   V   L   H   A   A   L   H   K   P   E   G   G
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  481  AAC ATG AAA CTA TAC GTA ATG TGT GTT TTA CAT GCC GCA CTA CAT AAA CCG GAG GGC GGC
       TTG TAC TTT GAT ATG CAT TAC ACA CAA AAT GTA CGG CGT GAT GTA TTT GGC CTC CCG CCG
        V   G   L   A   A   V   K   V   T   G   M   C   D   P   Q   L   L   A   R   V
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  541  GTT GGA TTG GCT GCT GTT AAG GTG ACA GGC ATG TGT GAT CCA CAG CTT CTC GCC CGC GTC
       CAA CCT AAC CGA CGA CAA TTC CAC TGT CCG TAC ACA CTA GGT GTC GAA GAG CGG GCG CAG
        S   A   I   L   H   S   V   H   R   D   W   I   E   Y   F   T   E   E   Q   P
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  601  TCG GCA ATC CTG CAC TCC GTT CAC CGT GAC TGG ATA GAA TAC TTT ACG GAG GAG CAA CCA
       AGC CGT TAG GAC GTG AGG CAA GTG GCA CTG ACC TAT CTT ATG AAA TGC CTC CTC GTT GGT
        P   P   V   E   E   C   N   V   V   M   G   T   K   T   E   H   K   R   Y   I
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  661  CCA CCA GTG GAG GAG TGT AAT GTT GTG ATG GGA ACA AAA ACG GAG CAC AAG AGA TAT ATT
       GGT GGT CAC CTC CTC ACA TTA CAA CAC TAC CCT TGT TTT TGC CTC GTG TTC TCT ATA TAA
        T   R   D   Q   V   R   K   G   L   T   K   L   A   S   S   Q   K   Y   T   E
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  721  ACG CGT GAC CAA GTG CGG AAG GGA CTC ACA AAA TTG GCC TCT TCG CAA AAA TAC ACA GAA
       TGC GCA CTG GTT CAC GCC TTC CCT GAG TGT TTT AAC CGG AGA AGC GTT TTT ATG TGT CTT
        D   E   I   N   A   V   L   Q   V   L   D   P   N   N   E   G   K   T   N   Y
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  781  GAT GAG ATC AAT GCC GTC TTA CAA GTT CTC GAT CCA AAT AAT GAA GGA AAA ACA AAC TAC
       CTA CTC TAG TTA CGG CAG AAT GTT CAA GAG CTA GGT TTA TTA CTT CCT TTT TGT TTG ATG
        Y   K   F   K   T   V   V   S   E   A   V   L   A   L   D   P   T   P   V   Q
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  841  TAC AAG TTT AAA ACG GTT GTG TCT GAG GCC GTA CTT GCA CTG GAT CCC ACA CCA GTG CAG
       ATG TTC AAA TTT TGC CAA CAC AGA CTC CGG CAT GAA CGT GAC CTA GGG TGT GGT CAC GTC
        K   I   I   I   D   K   L   P   K   L   T   T   E   E   R   E   L   W   R   H
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  901  AAG ATA ATT ATC GAC AAA CTT CCA AAA CTT ACT ACA GAG GAG CGC GAA CTA TGG CGC CAT
       TTC TAT TAA TAG CTG TTT GAA GGT TTT GAA TGA TGT CTC CTC GCG CTT GAT ACC GCG GTA
        L   H   W   R   L   S   V   I   V   R   T   A   K   D   L   R   V   R   V   L
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
  961  TTG CAC TGG AGG TTG TCA GTC ATT GTT CGG ACT GCG AAG GAC TTG CGG GTG CGA GTT TTG
       AAC GTG ACC TCC AAC AGT CAG TAA CAA GCC TGA CGC TTC CTG AAC GCC CAC GCT CAA AAC
        F   D   A   E   Q   T   F   Y   Q   L   A   I   D   N   I   V   L   Q   F   Q
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 1021  TTT GAC GCC GAG CAA ACG TTT TAT CAA CTT GCC ATT GAT AAT ATT GTG TTG CAG TTT CAG
       AAA CTG CGG CTC GTT TGC AAA ATA GTT GAA CGG TAA CTA TTA TAA CAC AAC GTC AAA GTC
        R   Q   F   N   K   K   E   A   I   V   Y   N   T   Y   Q   C   Y   L   T   Y
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 1081  CGC CAG TTC AAT AAA AAG GAG GCT ATA GTT TAT AAC ACA TAT CAA TGC TAT TTG ACT TAC
       GCG GTC AAG TTA TTT TTC CTC CGA TAT CAA ATA TTG TGT ATA GTT ACG ATA AAC TGA ATG
        T   E   D   R   V   F   N   D   L   T   R   A   E   L   E   G   W   V   W   G
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 1141  ACC GAA GAC CGT GTT TTC AAT GAC TTA ACC CGA GCA GAG CTC GAG GGG TGG GTT TGG GGA
       TGG CTT CTG GCA CAA AAG TTA CTG AAT TGG GCT CGT CTC GAG CTC CCC ACC CAA ACC CCT
        G   K   I   V   R   G   A   Y   M   R   Q   E   R   E   T   A   E   K   Y   H
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 1201  GGG AAA ATT GTT CGA GGT GCT TAC ATG AGG CAG GAG AGG GAG ACA GCA GAA AAA TAC CAT
       CCC TTT TAA CAA GCT CCA CGA ATG TAC TCC GTC CTC TCC CTC TGT CGT CTT TTT ATG GTA
        Y   K   S   P   I   W   P   T   Y   E   E   T   N   A   C   Y   K   A   V   A
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 1261  TAC AAA AGC CCC ATT TGG CCA ACG TAC GAG GAA ACT AAC GCG TGC TAC AAA GCT GTG GCT
       ATG TTT TCG GGG TAA ACC GGT TGC ATG CTC CTT TGA TTG CGC ACG ATG TTT CGA CAC CGA
```

```
          E   R   I   L   R   E   I   A   R   L   P   E   T   R   F   E   A   L   F   G
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1321      GAG CGA ATA CTA AGA GAG ATA GCC CGG TTA CCT GAA ACA CGT TTT GAG GCC CTG TTC GGA
          CTC GCT TAT GAT TCT CTC TAT CGG GCC AAT GGA CTT TGT GCA AAA CTC CGG GAC AAG CCT
           T   H   N   Q   K   S   L   E   E   I   T   E   A   V   L   Q   L   P   P   V
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1381      ACT CAC AAT CAA AAA TCA CTT GAA GAA ATA ACT GAG GCT GTT CTT CAA CTT CCT CCT GTT
          TGA GTG TTA GTT TTT AGT GAA CTT CTT TAT TGA CTC CGA CAA GAA GTT GAA GGA GGA CAA
           K   G   Y   V   A   F   A   Q   L   Y   G   M   S   D   N   L   T   I   P   L
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1441      AAA GGT TAT GTG GCA TTT GCC CAA TTG TAT GGC ATG TCG GAC AAC TTG ACA ATT CCG CTT
          TTT CCA ATA CAC CGT AAA CGG GTT AAC ATA CCG TAC AGC CTG TTG AAC TGT TAA GGC GAA
           K   R   A   G   F   P   V   F   K   Y   V   P   Y   G   P   V   K   E   T   V
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1501      AAG AGA GCC GGT TTT CCA GTT TTT AAA TAT GTT CCG TAC GGA CCT GTC AAA GAG ACT GTT
          TTC TCT CGG CCA AAA GGT CAA AAA TTT ATA CAA GGC ATG CCT GGA CAG TTT CTC TGA CAA
           H   Y   L   G   R   R   A   M   E   N   A   S   I   L   S   N   G   G   S   R
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1561      CAC TAC CTG GGA AGG CGT GCT ATG GAG AAC GCG TCA ATT TTA TCG AAT GGG GGT AGC CGG
          GTG ATG GAC CCT TCC GCA CGA TAC CTC TTG CGC AGT TAA AAT AGC TTA CCC CCA TCG GCC
           E   V   R   L   M   R   K   E   L   R   R   R   V   F   W   M
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1621      GAA GTG CGG TTG ATG AGG AAA GAG CTT AGG CGT CGC GTC TTT TGG ATG
          CTT CAC GCC AAC TAC TCC TTT CTC GAA TCC GCA GCG CAG AAA ACC TAC
```

Figure 3.8: *T. brucei* proline dehydrogenase nucleotide gene sequence. The sequence was obtained by sequencing the putative gene cloned into pGEM-T vector. This scheme shows the DNA sequence and the corresponding encoded protein sequence.

```
          M   R   R   L   L   P   L   R   P   A   A   V   A   F   A   G   S   A   R   H
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1         ATG CGT CGT CTC CTC CCC CTG CGA CCG GCA GCT GTG GCC TTT GCC GGC TCT GCT CGT CAC
          TAC GCA GCA GAG GAG GGG GAC GCT GGC CGT CGA CAC CGG AAA CGG CCG AGA CGA GCA GTG
           S   S   L   T   M   Q   D   K   Q   P   K   L   P   N   F   N   D   D   T   T
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
61        TCC TCC TTG ACA ATG CAG GAC AAG CAG CCG AAG CTG CCG AAC TTC AAC GAC GAC ACA ACC
          AGG AGG AAC TGT TAC GTC CTG TTC GTC GGC TTC GAC GGC TTG AAG TTG CTG CTG TGT TGG
           Y   R   Q   R   S   A   W   Y   L   I   K   A   L   V   V   L   R   L   C   S
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
121       TAC CGC CAG CGC TCG GCA TGG TAC TTG ATA AAG GCG TTG GTG GTG CTC CGC CTG TGC AGT
          ATG GCG GTC GCG AGC CGT ACC ATG AAC TAT TTC CGC AAC CAC CAC GAG GCG GAC ACG TCA
           V   N   Y   L   A   M   N   S   V   P   L   M   K   R   V   E   K   I   L   G
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
181       GTG AAC TAT TTG GCA ATG AAC TCG GTG CCG CTG ATG AAG AGA GTG GAA AAG ATC CTT GGC
          CAC TTG ATA AAC CGT TAC TTG AGC CAC GGC GAC TAC TTC TCT CAC CTT TTC TAG GAA CCG
           S   K   L   T   Y   S   I   L   V   K   K   S   F   Y   N   Y   F   C   A   G
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
241       AGC AAG CTT ACC TAC AGC ATC CTC GTC AAG AAG TCC TTC TAC AAC TAC TTC TGC GCG GGC
          TCG TTC GAA TGG ATG TCG TAG GAG CAG TTC TTC AGG AAG ATG TTG ATG AAG ACG CGC CCG
           E   N   D   Q   E   L   R   D   T   V   R   K   L   S   R   N   N   I   G   A
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
301       GAA AAC GAC CAG GAG CTG CGC GAC ACG GTG CGA AAG CTT TCA CGC AAC AAC ATC GGC GCT
          CTT TTG CTG GTC CTC GAC GCG CTG TGC CAC GCT TTC GAA AGT GCG TTG TTG TAG CCG CGA
           V   L   D   Y   A   A   E   A   D   T   E   G   F   A   P   E   P   G   V   A
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
361       GTA CTC GAC TAC GCG GCG GAG GCC GAC ACG GAG GGC TTC GCA CCG GAG CCG GGT GTG GCG
          CAT GAG CTG ATG CGC CGC CTC CGG CTG TGC CTC CCG AAG CGT GGC CTC GGC CCA CAC CGC
           S   G   P   D   I   S   M   S   S   L   V   M   K   P   N   V   Q   Y   P   M
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
421       TCC GGC CCC GAT ATT TCG ATG TCT AGT CTC GTT ATG AAG CCC AAT GTT CAG TAC CCA ATG
          AGG CCG GGG CTA TAA AGC TAC AGA TCA GAG CAA TAC TTC GGG TTA CAA GTC ATG GGT TAC
           D   E   G   F   F   N   E   N   M   K   L   Y   M   M   S   I   M   H   A   S
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
481       GAC GAG GGG TTC TTT AAC GAG AAC ATG AAG CTC TAC ATG ATG AGC ATC ATG CAC GCC TCG
          CTG CTC CCC AAG AAA TTG CTC TTG TAC TTC GAG ATG TAC TAC TCG TAG TAC GTG CGG AGC
           L   Y   S   P   R   N   V   A   G   V   T   A   V   K   V   T   G   M   C   D
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
541       CTG TAC AGC CCG CGA AAC GTC GCC GGT GTA ACC GCT GTT AAG GTA ACG GGC ATG TGC GAC
          GAC ATG TCG GGC GCT TTG CAG CGG CCA CAT TGG CGA CAA TTC CAT TGC CCG TAC ACG CTG
           P   Q   L   L   A   R   V   S   A   L   L   M   S   V   H   Q   S   W   C   K
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
601       CCT CAG CTG CTT GCA CGC GTA TCG GCG CTG CTC ATG TCC GTC CAC CAG AGC TGG TGC AAG
          GGA GTC GAC GAA CGT GCG CAT AGC CGC GAC GAG TAC AGG CAG GTG GTC TCG ACC ACG TTC
           H   F   T   N   E   E   S   L   K   L   E   E   C   R   V   V   M   G   V   N
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
661       CAC TTC ACG AAC GAG GAG TCG CTG AAG CTG GAA GAG TGC CGC GTC GTC ATG GGC GTG AAC
          GTG AAG TGC TTG CTC CTC AGC GAC TTC GAC CTT CTC ACG GCG CAG CAG TAC CCG CAC TTG
           R   K   H   Q   L   F   I   T   Y   D   Q   L   R   A   G   F   E   K   Y   N
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
721       CGC AAG CAC CAG CTG TTC ATC ACC TAC GAT CAG CTA CGC GCC GGT TTC GAG AAG TAC AAC
          GCG TTC GTG GTC GAC AAG TAG TGG ATG CTA GTC GAT GCG CGG CCA AAG CTC TTC ATG TTG
           P   S   N   K   L   S   D   A   Q   F   K   E   I   T   E   A   L   D   P   R
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
781       CCC TCT AAC AAG CTG TCG GAT GCC CAG TTC AAG GAG ATT ACG GAG GCC CTG GAC CCC CGC
          GGG AGA TTG TTC GAC AGC CTA CGG GTC AAG TTC CTC TAA TGC CTC CGG GAC CTG GGG GCG
           K   T   G   K   V   N   Y   F   E   Y   K   E   M   L   T   N   A   L   I   A
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
841       AAG ACG GGC AAG GTG AAC TAC TTT GAG TAC AAG GAG ATG CTG ACG AAC GCC CTC ATC GCC
          TTC TGC CCG TTC CAC TTG ATG AAA CTC ATG TTC CTC TAC GAC TGC TTG CGG GAG TAG CGG
           V   E   P   T   P   V   Q   Q   A   L   I   E   G   L   P   Q   M   S   A   K
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
901       GTG GAG CCG ACG CCG GTG CAG CAG GCG CTG ATT GAG GGA CTG CCG CAG ATG AGT GCA AAG
          CAC CTC GGC TGC GGC CAC GTC GTC CGC GAC TAA CTC CCT GAC GGC GTC TAC TCA CGT TTC
           E   K   V   L   W   K   N   V   N   N   R   L   L   L   I   A   S   M   A   K
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
961       GAG AAG GTG CTG TGG AAA AAC GTC AAC AAC CGA CTC TTG TTG ATC GCC TCC ATG GCA AAG
          CTC TTC CAC GAC ACC TTT TTG CAG TTG TTG GCT GAG AAC AAC TAG CGG AGG TAC CGT TTC
           E   L   N   V   R   M   L   V   D   A   E   Q   T   F   Y   Q   L   A   I   D
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1021      GAG CTC AAT GTG CGC ATG CTT GTC GAC GCG GAG CAG ACC TTT TAT CAG CTG GCC ATC GAC
          CTC GAG TTA CAC GCG TAC GAA CAG CTG CGC CTC GTC TGG AAA ATA GTC GAC CGG TAG CTG
           A   I   V   A   T   L   Q   K   T   Y   N   T   E   L   P   V   V   Y   N   T
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1081      GCG ATC GTG GCG ACC CTA CAG AAG ACC TAC AAC ACG GAG CTG CCG GTG GTG TAC AAC ACT
          CGC TAG CAC CGC TGG GAT GTC TTC TGG ATG TTG TGC CTC GAC GGC CAC CAC ATG TTG TGA
           Y   Q   C   Y   L   T   Y   A   E   D   R   I   D   N   D   L   V   R   A   R
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1141      TAC CAG TGC TAC CTG ACA TAC GCA GAG GAC CGC ATT GAC AAC GAC CTC GTT CGC GCT CGC
          ATG GTC ACG ATG GAC TGT ATG CGT CTC CTG GCG TAA CTG TTG CTG GAG CAA GCG CGA GCG
           H   M   N   F   H   W   G   G   K   I   V   R   G   A   Y   I   V   Q   E   R
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1201      CAC ATG AAC TTC CAC TGG GGC GGC AAG ATC GTG CGC GGC GCC TAT ATA GTG CAA GAG CGC
          GTG TAC TTG AAG GTG ACC CCG CCG TTC TAG CAC GCG CCG CGG ATA TAT CAC GTT CTC GCG
           A   T   A   A   Q   Y   G   Y   T   S   P   I   W   S   T   Y   E   E   T   N
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1261      GCG ACG GCG GCT CAG TAC GGC TAC ACC AGC CCC ATC TGG TCT ACC TAC GAG GAG ACG AAC
          CGC TGC CGC CGA GTC ATG CCG ATG TGG TCG GGG TAG ACC AGA TGG ATG CTC CTC TGC TTG
```

```
          K   C   Y   N   A   A   A   K   R   I   F   D   T   F   E   A   Q   P   A   K
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1321      AAG TGC TAT AAT GCC GCC GCG AAG CGC ATC TTC GAC ACC TTC GAG GCG CAG CCA GCG AAA
          TTC ACG ATA TTA CGG CGG CGC TTC GCG TAG AAG CTG TGG AAG CTC CGC GTC GGT CGC TTT
           K   H   E   V   F   F   G   T   H   N   K   K   S   L   E   I   I   T   A   S
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1381      AAG CAC GAG GTC TTC TTT GGC ACT CAC AAC AAG AAG TCT CTA GAG ATT ATC ACG GCC AGC
          TTC GTG CTC CAG AAG AAA CCG TGA GTG TTG TTC TTC AGA GAT CTC TAA TAG TGC CGG TCG
           V   L   E   R   P   S   I   Q   S   R   V   S   F   G   Q   L   F   G   M   R
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1441      GTC TTG GAA CGA CCG AGC ATC CAG TCT CGC GTG TCC TTT GGG CAG CTG TTT GGG ATG CGC
          CAG AAC CTT GCT GGC TCG TAG GTC AGA GCG CAC AGG AAA CCC GTC GAC AAA CCC TAC GCG
           D   N   L   T   V   P   L   A   R   A   G   F   Q   V   Y   K   Y   V   P   Y
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1501      GAC AAC CTG ACG GTG CCC CTT GCC CGA GCC GGC TTT CAG GTC TAC AAG TAC GTG CCG TAC
          CTG TTG GAC TGC CAC GGG GAA CGG GCT CGG CCG AAA GTC CAG ATG TTC ATG CAC GGC ATG
           G   P   V   K   E   T   I   H   Y   L   G   R   R   A   V   E   N   S   S   I
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1561      GGC CCC GTG AAG GAG ACG ATC CAC TAC CTC GGC CGC CGC GCC GTG GAG AAT TCG TCG ATC
          CCG GGG CAC TTC CTC TGC TAG GTG ATG GAG CCG GCG GCG CGG CAC CTC TTA AGC AGC TAG
           L   T   T   G   D   N   E   T   V   M   M   I   K   E   L   K   R   R   C   G
          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1621      TTG ACA ACC GGT GAC AAC GAG ACG GTG ATG ATG ATC AAG GAG CTG AAG CGC CGC TGC GGT
          AAC TGT TGG CCA CTG TTG CTC TGC CAC TAC TAC TAG TTC CTC GAC TTC GCG GCG ACG CCA
           F
          ~~~~
1681      TTT
          AAA
```

Figure 3.9: *L. major* proline dehydrogenase nucleotide gene sequence. The sequence was obtained by sequencing the putative gene cloned into pGEM-T vector. This scheme shows the DNA sequence and the corresponding encoded protein sequence.

3.2.2- Gene organisation and Southern blotting

In order to determine the gene organisation and to find the number of gene copies, Southern blotting was performed using *PRODH* gene probes from *L. major* and *T. brucei*. The result for *L. major* is shown in Figure 3.10.

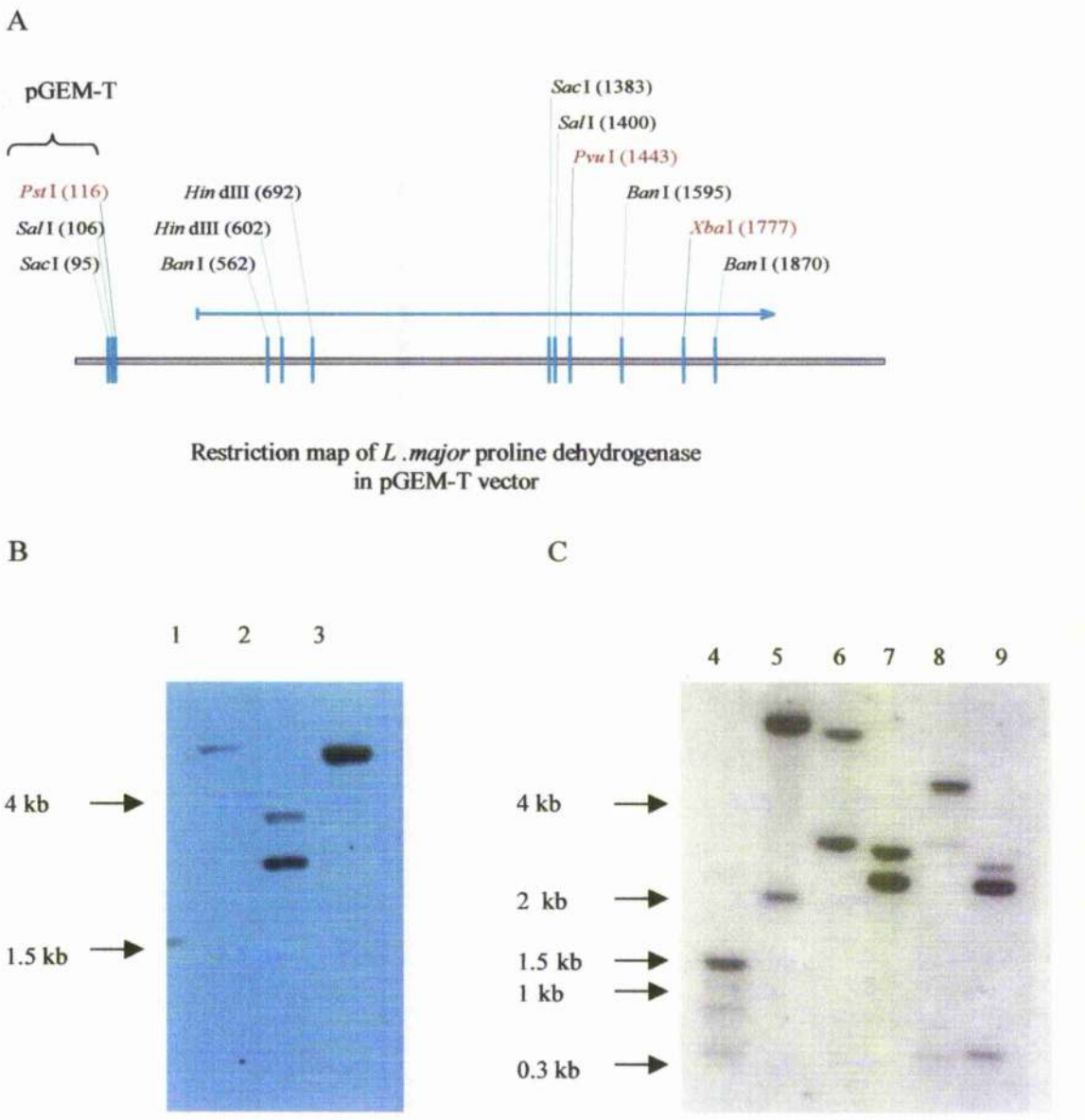


Figure 3.10: Southern blot analysis of *L. major* gDNA

A- Schematic representation of the restriction map of the *L. major* proline dehydrogenase gene sequence ORF generated using Vector NTI. The black colour for restriction enzymes means they cut more than once. The inserted *L. major* DNA is depicted by the arrow.

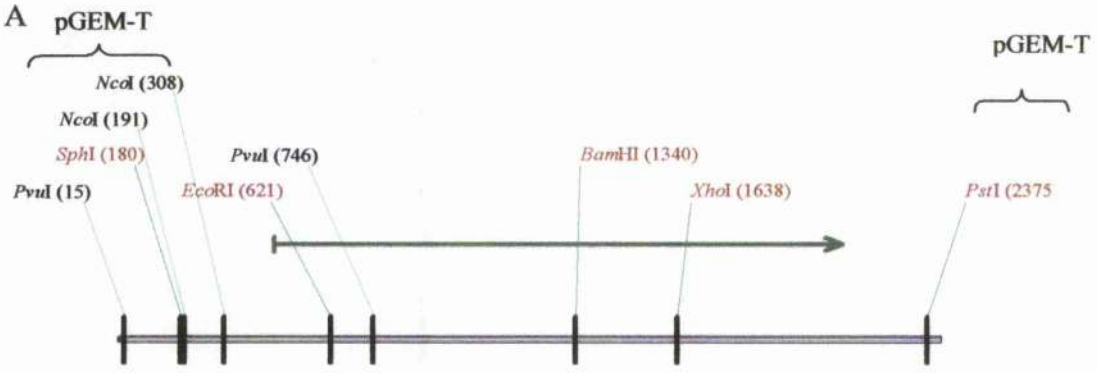
B- Southern blot analysis of *L. major* gDNA digested with *Bsa* I (lane 1), *Sal* I (lane 2) and *Pst* I (lane 3).

C- Southern blot analysis of *L. major* gDNA digested with *Ban* I (Lane 4), *Hind* III (lane 5), *Sac* I (lane 6), *Sal* I (lane 7), *Pvu* I and *Sal* I (lane 8) and *Xba* I and *Sal* I (lane 9). The two blots were probed with the *L. major* proline dehydrogenase ORF.

The complete gDNA was digested by various restriction enzymes that did not cut or cut once or several times inside the gene (see restriction map). The results indicate that the gene was present as a single copy inside the genome. For example, the *Pst* I enzyme should not cut inside the gene, and only one band was present in the Southern blot. *Sac* I cuts once inside the gene, two bands were found in the blot. Lane 8 (Figure 3.10 C) shows the result of the digestion by *Xba* I and *Sal* I, three bands were expected; one should have been situated around 377 bp, but it is too small to be seen in the gel.

The same study was performed with *T. brucei* gDNA and it was determined that the gene was also present as a single copy in this organism (Figure 3.11). For example, with the digestion with *Pvu* I and *Xho* I, three bands are expected, one should be situated around 900 bp (Figure 3.11, lane 6), the second is found at 3.1 kb and the third at 2.5 kb.

A



Restriction map of *T. brucei* proline dehydrogenase in pGEM-T vector

B

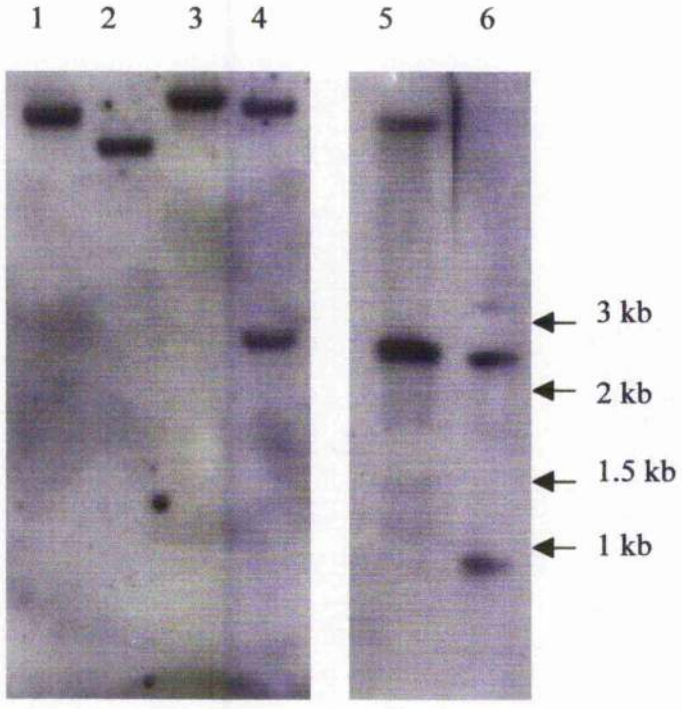


Figure 3.11: Southern blot analysis of *T. brucei* gDNA

A- schematic representation of the restriction map of the *T. brucei* proline dehydrogenase ORF. The black colour for restriction enzymes means they cut more than once. The *T. brucei* insert is depicted by the arrow. B- Southern blot analysis of *T. brucei* gDNA digested with *Nco* I (lane 1), *Hind* III (lane 2) *Pst* I (lane 3), *Xho* I (Lane 4), *BamH* I (lane 5) and Pvu I and *Xho* I (lane 6), and probed with *T. brucei* proline dehydrogenase ORF.

3.2.3- Similarity to other proline dehydrogenase amino acid sequences

The genbank Database http://www.ncbi.nih.gov was screened in order to find PRODH sequences from different organisms. Table 3.1 summarises the name of the organisms and the accession number of the protein used for the alignments.

Name of organism	**Genbank accession number**
Alpha-proteobacteria	
Agrobacterium tumefaciens	NP_356492
Bradyrhizobium japonicum	AAL 35755
Caulobacter crescentus	NP_419621
Mesorhizobium loti	NP_102812
Sinorhizobium meliloti	CAC 41903
Other bacteria	
Bacillus halodurans	NP_243606
Bacillus subtilis	AAM 27442
Brucella melitensis	NP_541542
Campylobacter jejuni	C 81 297
Escherichia coli	NP_415534
Helicobacter pylori	B 71980
Mycobacterium tuberculosis	AAK45482
Burkholderia cepacia	AF 029344
Neisseria meningitidis	NP_284781
Pyrobaculum aerophilum	NP_559919
Pseudomonas aeruginosa	NP_249473
Pseudomonas putida	AAF 73193
Rhodobacter capsulatus	CAA 55142
Neisseria meningitidis	NP_284781
Pyrobaculum aerophilum	NP_559919
Pseudomonas aeruginosa	NP_249473
Pseudomonas putida	AAF 73193
Rhodobacter capsulatus	CAA 55142
Salmonella typhimurium	AAL 20055
Vibrio vulnificus	AAL 68393
Yersinia pestis	NP_669761
Eukaryotes	
Aspergillus nidulans	CAC 18796
Arabidopsis thaliana	P 92983
Drosophila melanogaster	NP_523433
Homo sapiens	NP_057419
Mus musculus	NP_035302
Saccaromyces cerevisiae	P 09368

Table 3.1: GenBank accession numbers for proline dehydrogenase sequences used in this study

PRODH sequences from *T. brucei* and *L. major* were aligned, using Align X (Vector NTI) and showed 66 % similarity and 52 % identity (Figure 3.12).

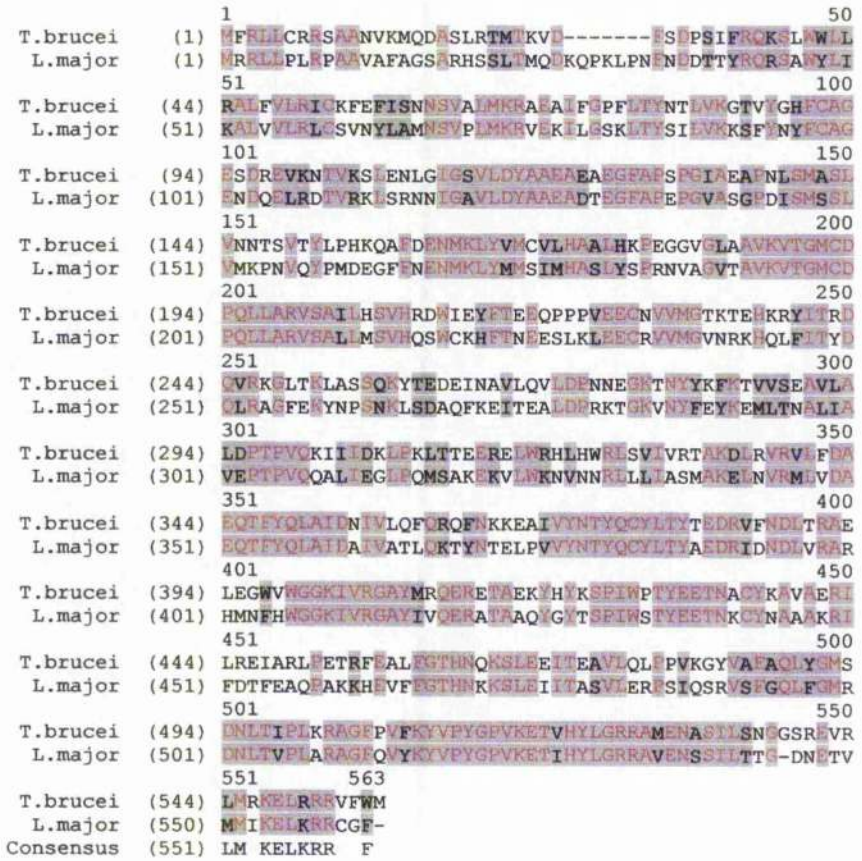


Figure 3.12: Amino acid sequence alignment of *T. brucei* and *L. major* proline dehydrogenase. The sequences were aligned using Align X (Vector NTI), Clustal X algorithm. The identical amino acids are indicated in red and grey shading, black amino acids and grey shading indicates conserved amino acids.

These two sequences were aligned with PRODH from various organisms (eukaryotes and prokaryotes) in order to find well conserved motifs. Sequences were aligned using Align X and the results are shown in Figure 3.13 and Appendix 2. The identity between all

sequences is 1.3 %. The value is very low but the alignment was done using prokaryotic and eukaryotic sequences and the absence of P5CDH in the protein of eukaryotic organisms diminishes the value of identity. However regions with high homology are found along the sequences.

In eukaryotes, PRODH and P5CDH are encoded by two independent genes, while in prokaryotes the two enzymes catalysing proline degradation are present on a single polypeptide encoded by a single gene. The N-terminal part of the prokaryotic protein possesses the PRODH activity, whereas the C-terminal part has the P5CDH activity (Straub *et al.*, 1996; Vilchez *et al.*, 2000). In *E. coli,* for example, the PRODH part is situated between residues 1-669 out of a total of 1330 amino acids (Vinod *et al.*, 2002).

In *L. major* and *T. brucei*, The PRODH and P5CDH enzymes appear to be encoded by two different genes. This was demonstrated by analysing extended contigs from the *PRODH* genes to determine whether there is any homology to P5CDH, and also by identifying the *P5CDH* genes to see if any contigs extending into PRODH homology sequence could be identified. They could not. A gene coding for a potential P5CDH was found in the chromosome 10 of *T. brucei* (Gene bank accession number, 10.0.000574). The BLAST server at the Parasite Genome web site was screened for P5CDH in *L. major*. A putative protein (id AAF31034), which is similar to P5CDH, was identified. *PRODH* and *P5CDH* genes are on different chromosomes (7 and 26 for *PRODH* in *T. brucei* and *L. major,* respectively, and chromosome 10 and 3 for *P5CDH* in *T. brucei* and *L. major,* respectively). Two regions with particularly high identity, at the amino acid level, were conserved in all proline dehydrogenase sequences examined (Figure 3.13).

A: Box 1

```
                    401                                                 450
A.tumefaciens  (334) WNGLGFVVQAYGRRCPFVLDYIIDLARRSGRRIMVRLVKGAYWDAEIKRA
       E.coli  (396) WNGIGFVIQAYQKRCPLVIDYLIDLATRSRRRLMIRLVKGAYWDSEIKRA
    H.sapiens  (321) KPLIFNTYQCYLKDAYDNVTLDVELARREGWCFGAKLVRGAYLAQERARA
      L.major  (374) LPVVYNTYQCYLTYAEDRIDNDLVRARHMNFHWGGKIVRGAYIVQERATA
  M.musculus  (302) KPFIFNTFQCYLKDAYDNVTLDMELARREGWCSGAKLVRRAYMAQERVRA
     P.putida  (393) WNGIGFVIQAYQKRCPYLINYFFDLAKRTPHRLIIRLLKGAYWDSEIKRA
 S.cerevisiae  (286) LISCVGTWQLYLRDSGDHILHELKLAQENGYKLGLKLVRGAYIHSEKNRN
S.typhimurium  (396) WNGIGFVIQAYQKRCPLVIDYLVDLASRSRRRLMIRLVKGAYWDSEIKRA
      T.brucei  (367) EAIVYNTYQCYLTYTEDRVFNDLTRAELEGWVWGGKIVRGAYMRQERETA
      Y.pestis  (397) WNGIGFVIQAYQKRCPSTIDAVIDMAQRSRRRLMIRLVKGAYWDSEIKRA
     Consensus  (401) WNGIGFVYQAYLKRCPD IDYDIDLARRSGWRLGIKLVKGAYWDSEIKRA
```

B: Box 2

```
                    551                                                 600
A.tumefaciens  (479) GTHETLLAYLVRRLLENGANSSFVNRIADPAVPVDSLLEDPVAVVKAYAV
       E.coli  (544) GTHETLLAYLVRRLLENGANTSFVNRIADTSLPLDELVADPVTAVEKLAQ
    H.sapiens  (468) GPVMEVLPYLSRRALEN-SSLMKGTHRERQLLWLELLRRLRTGNLFHRPA
      L.major  (521) GPVKETIHYLGRRAVENSSILTTG-DNETVMMIKELKRRCGF--------
  M.musculus  (449) GPVMEVLPYLSRRALEN-SSIMKGAQRERQLLWQELRRRLRTGSLFHHPA
     P.putida  (541) GTHETLLAYLVRRLLENGANTSFVNRIADHSISIQELVADRWPASIAWVP
 S.cerevisiae  (434) GPPLETKDYLLRRLQENGDAVRSDNGWPLIKAIAKSIPKRVGL-------
S.typhimurium  (544) GTHETLLAYLVRRLLENGANTSFVNRIADATLPLDELVADPVEAVEKLAQ
      T.brucei  (514) GPVKETVHYLGRRAMENASILSNGGSREVRLMRKELRRRVFWM-------
      Y.pestis  (545) GTHETLLAYLVRRLLENGANTSFVNRIADATLPLDELVADPVSAVEAMAA
     Consensus  (551) GTHETLLAYLVRRLLENGANTSFVNRIAD  L LD LVRD   AV
```

Figure 3.13: Multiple amino acid sequence alignment of portions of well conserved regions (Box 1 and Box 2) of proline dehydrogenase. The pink residues are found in all proline dehydrogenases known and are believed to be essential for the activity of the enzyme. The conserved sequences are VxGAY and YLxRRxxEN, where x is any amino acid. The number above the sequences represent the place where the two conserved boxes are in the proline dehydrogenase sequences, 1 being the first methionine.

These regions are V+GAY (where + is K or R) and YLxxRRxxEN, where x is any amino acid. These regions, which are also found in *T. brucei* and *L. major* sequences, are believed to be essential for the activity of the PRODH enzyme (Straub *et al.*, 1996). They could be a part of the active site involved in binding the substrate or the cofactor FAD (Lee Y *et al.*, 2003).

3.2.4- FAD and proline binding sites

The *E. coli* PRODH (PutA 669) is the first PRODH to have been crystallised (Lee Y *et al.*, 2003). The residues involved in binding FAD, and also the catalytic centre, have been identified in this protein. Figure 3.14 shows an alignment between the *L. major*, *T. brucei* and *E. coli* PRODHs. The similarity is 32 % and the identity 4.2 % between these three sequences. Between *L. major* and *E. coli* sequence, the identity is 6 %, and between *T. brucei* and *E. coli* the identity is 5 %.

```
E.coli      (1) MGTTTMGVKLDDATRERIKSAATRIDRTPHWLIKQAIFSYLEQLENSDTL
L.major     (1) ---MRRLLPLRPAAVAFAGSARHSSLTMQDKQPKLPNFNDDTTYRQRSAW
T.brucei    (1) ---MFRLLCRRSAANVKMQDASLRTMTKVD-------FSDPSIFRQKSLW
Consensus   (1)    M RLL LR AA  KI SAA RSLT  D   K   FSD S FRQKS W

E.coli     (51) PELPALLSGAANESDEAPTPAEEPHQPFLDFAEQILPQSVSRAAITAAYR
L.major    (48) YLIKALVVLRLCSVNYLAM----NSVPLMKRVEKILGSKLTYSILVKKSF
T.brucei   (41) WLLRALFVLRICKFEFISN----NSVALMKRAEAIFGPFLTYNTLVKGTV
Consensus  (51) WLLKALLVLRIC  DFIA     NSVPLMKRAE ILG  LTYA LVKAS

E.coli    (101) RPETEAVSMLLEQARLPQPVAEQAHKLAYQLADKLRNQKNASGRAGMVQG
L.major    (94) YNYFCAGENDQELRDTVRKLSRNNIGAVLDYAAEADTEGFAPEPGVASGP
T.brucei   (87) YGHFCAGESDREVKNTVKSLENLGIGSVLDYAAEAEAEGFAPSPGIAEAP
Consensus (101) Y HFCAGE D ELK TVK LA NAIGAVLDYAAEAD EGFAP PGIA AP

E.coli    (151) LLQEFSLSSQEGVALMCLAEALLRIPDKATRDALIRDKISNGNWQSHIGR
L.major   (144) DISMSSIVMKPNVQYPMDEGFFNENMKLYMMSIMHASLYSP---RNVAGV
T.brucei  (137) NLSMASLVNNTSVTYLPHKQAFDENMKLYVMCVLHAALHKP---EGGVGL
Consensus (151)  LSMASLV N  V YL    AF ENMKLYMM ILHAALHSP      IGL

E.coli    (201) SPSLFVNAATWGLLFTGKLVSTHNEASLSRSLNRIIGKSGEPLIRKGVDM
L.major   (191) TAVKVTGMCDPQLLARVSALLMSVHQSWCKHFTNEESL---KLEECRVVM
T.brucei  (184) AAVKVTGMCDPQLLARVSAILHSVHRDWIEYFTEEQPP---PVEECNVVM
Consensus (201) SAVKVTGMCDPQLLARVSAIL SVH SW KHFT E      PLEEC VVM

E.coli    (251) AMRLMGEQFVTGETIAEALANARKLEEKGFRYSYDMLGEAALTAADAQAY
L.major   (238) GVNRKHQLFITYDQLRAGFEKYNPSN----KLSDAQFKEITEALDPRKTG
T.brucei  (231) GTKTEHKRYITRDQVRKGLTKLASSQ----KYTEDFINAVLQVLDPNNEG
Consensus (251) GMK  H  FIT DQIR GL K   SN    KYSDD I EI   LDP N G

E.coli    (301) MVSYQQAIHAIGKASNGRGIYEGPGISIKLSALHPRYSRAQYDRVMEELY
L.major   (284) KVNYFEYKEMLTNAL----IAVEPTPVQQALIEGLPQMSAKEKVLWKNVN
T.brucei  (277) KTNYYKFKTVVSEAV----LALDPTPVQKIIIDKLPKLTTEERELWRHLH
Consensus (301) KVNYF FK MIS AL    IALDPTPVQKIIID LP LSA EK LWK LH

E.coli    (351) PRLKSLTLLARQYDIGINIDAEESDRLEISLDLLEKLCFEPELAGWNGIG
L.major   (330) NRLLLIASMAKELNVRMLVDAEQTF-YQLAIDAIVATLQKTYNTELPVVY
T.brucei  (323) WRLSVIVRTAKDLRVRVLFDAEQTF-YQLAIDNIVLQFQRQFNKKEAIVY
Consensus (351)  RL LI  LAKDL VRILIDAEQTF YQLAID IV   QK FN    IVY
```

Figure 3.14: Amino acid alignment of *E. coli* (proline dehydrogenase part), *L. major* and *T. brucei* proline dehydrogenases. The residues of the FAD binding motif is in green bold for *E. coli* and the conserved residues are in pink bold in *T. brucei* and *L. major*. The FAD motif in *E. coli* is highlighted in blue boxes. The key residues are Glu 289, Lys 329, Val 402, Arg 431, Val 433, Ala 485, Leu 513, Arg 555-556 and Glu 559. The proline binding site is constituted by Asp 370, Leu 513, Tyr 540 in *E. coli*. The blue line above the amino acid sequences represents a conservation between the 3 species at the level of 2-D structure (alpha helix).

Through personal communications with Professor D. Becker (University of Missouri) and by comparison with the crystal structure of PutA 669, potentially essential residues of proline dehydrogenases from *T. brucei* and *L. major* enzymes have been identified (see alignment Figure 3.14).

A competitive inhibitor, L-lactate, has been co-crystallised with the *E. coli* protein and gives information about the active site of the enzyme (Lee Y *et al.*, 2003). Lactate binds near the isoalloxazine ring of FAD to three basic residues, Lys 329, Arg 555 and Arg 556 (amino acids number are given according to the *E. coli* sequence). All of these residues are

conserved in both *L. major* and *T. brucei*. Glu 559, Glu 289, Asp 370 and Arg 431 stabilize these three residues while Arg 431 donates a hydrogen bond to FAD (Table 3.2).

For the *E. coli* protein, three residues (Asp 370, Tyr 540 and Leu 513) appear to interact with the substrate (Lee Y *et al.*, 2003). Leu 513 can be substituted in *L. major* and *T. brucei* proline dehydrogenases by an amino acid belonging to the same group, valine. The other two residues are identical. According the alignment illustrated in Figure 3.14 and Appendix 2, these fundamental amino acids are well conserved within all of the proline dehydrogenase sequences.

In the middle part of the three sequences, we can see a conservation in the position of alpha helix. The conservation of a number of key residues involved in binding to substrate, also the FAD binding site and the 2-D structure indicate that the trypanosomatid enzyme is almost certainly FAD-dependent and probably uses the same catalytic mechanism as the *E. coli* enzyme.

***E. coli* residues**	***T. brucei* residues**	***L. major* residues**
Lys 329 (binds lactate)	Lys	Gln
Arg555 (binds lactate)	Arg	Arg
Arg 556 (binds lactate)	Arg	Arg
Glu 559 (stabilises)	Glu	Glu
Glu 289 (stabilises)	Ala	Glu
Asp 370 (substrate binding)	Asp	Asp
Arg 431 (FAD binding)	Lys	Lys
Tyr 540 (substrate binding)	Tyr	Tyr
Leu 513 (substrate binding)	Val	Val

Table 3.2: Key residues in *E. coli* and corresponding residues in *T. brucei* and *L. major*

3.2.5- Identification of a putative mitochondrial targeting sequence

It is known that in eukaryotes, PRODH is a mitochondrial protein (Small and Jones, 1990). Nuclear-encoded mitochondrial proteins generally possess an N-terminal mitochondrial targeting peptide, which is composed by a group of 20-40 amino acids, many of which are positively charged. This signal peptide binds to specific receptors located on the mitochondrial membrane and the protein is able to cross the membrane via a specific pore. The presence of putative signal peptides at the N-termini of the *T. brucei* and *L. major* enzymes was investigated using two independent programs (Table 3.3).

TargetP v1.01 predicts the subcellular location of eukaryotic protein sequences. The subcellular location assignment is based on the predicted sequence of known N-terminal targeting sequences: chloroplast transit peptide (cTP), mitochondrial targetting peptide (mTP) or secretory pathway signal peptide (SP) (Emanuelson *et al.*, 2000).

MitoProt focuses on the N-terminal protein region that can support a mitochondrial targetting sequence and a cleavage site. The information that guides the protein to mitochondria is contained in its sequence. This program uses 47 parameters chosen by analysing a large set of mitochondrial proteins extracted from the SwissProt database. By analysing the amino acid sequence, 75-97% of the mitochondrial proteins studied have been predicted to be imported into mitochondria (Claros and Vincens, 1996).

A

TargetP v1.01 Mitoprot II

Name of organism	Probability of export to mitochondria	predicted localisation	Cleavage site	Probability of export to mitochondria	Length of the sequence (AA)
L.major	0.851	M	21	0.6	561
T. brucei	0.864	M	NP	0.8	556
E. coli	0.21	NP	NP	0.06	1320
D. melanogaster	0.961	M	42	0.98	669
H. sapiens	0.782	M	49	0.93	600
M. musculus	0.117	NP	NP	0.11	497
A. thaliana	0.913	M	18	0.9	499

B

```
T.brucei    (1) MFRLLCRRSAANVKMQDASLRT
L.major     (1) MRRLLPLRPAAVAFAGSARHS
```

Table 3.3 A: Predicted subcellular locations of different proline dehydrogenases using TargetP and Mitoprot II programmes. B: Putative MTS in *L. major* and *T. brucei*

The length of the sequence is the predicted length of the proline dehydrogenase proteins, in amino acids, entered into the programme.

M: Mitochondrial

NP: Non-predictable

These two programs both predicted that *L. major* and *T. brucei* PRODH are mitochondrially located. The MTS being the 21 first amino acids in *L. major*, and being not predicted in *T. brucei*. As controls, *E. coli*, which does not have mitochondria, gave us a non-predictable result for the location of the protein. The *D. melanogaster* (Hayward *et al.*, 1993), *A. thaliana* (Kiyosue *et al.*, 1996) and *H. sapiens* (Maxwell and Davis, 2000) enzymes, known to be present inside mitochondria, gave a positive control. The location of the *M. musculus* protein was non-predictable, which shows the limitation of the program as this protein is known to be mitochondrial. It is possible, as well, that the *M. musculus*

sequence taken from the database was not complete or that an unidentified splicing event adds a mitochondrial targeting sequence.

3.2.6- Identification of trans-membrane (TM) domains

According to the literature, PRODH is a membrane-bound protein in prokaryotes (Ling *et al.*, 1994) and it is associated with the inner mitochondrial membrane in eukaryotes (Wang and Brandriss, 1987). The TMprod program was used to establish whether potential TM domains are present within the *T. brucei* and *L. major* proteins. The TMpred program makes a prediction of membrane-spanning regions and their orientation. The prediction is made using a combination of several matrices for scoring (Hofmann and Stoffel, 1992). The program calculates the hydrophobicity of each region of the amino acid sequence and the hydrophobicity of the protein is analysed and TM segments determined.

Name of organism	**Number of Transmembrane helices**	**Preferred model**	**Location**
L. major	1	N-term outside	45-69
T. brucei	none		
T. brucei K60>S	1	N-term outside	35-65
T. brucei E62>N	none		
D. melanogaster	3	N-term inside	162-182 (i-o) 280-301(o-i) 584-605 (i-o)
E. coli	3	N-term outside	204-220 (o-i) 780-801 (i-o) 808-830 (o-i)
H. sapiens	none		
S. typhimurium	3	N-term inside	202-220 (i-o) 780-804 (o-i) 806-830 (i-o)

Table 3.4: Number of trans-membrane domains predicted using TMpred programme. The *T. brucei* sequence has been manually modified to find a potential hydrophobic domain. In the first case, K_{60} was substituted by S and in the second case E_{62} was substituted by a neutral amino acid.
i : inside, o : outside

A summary of the results is shown Table 3.4. The *L. major* sequence (Figure 3.15) has a predicted transmembrane domain situated between residues 45-69 where the hydrophobicity is the highest. The *T. brucei* sequence proved to be more difficult with respect to determining the trans-membrane domains. The TMpred programme did not predict any TM domains for the *T. brucei* sequence. However the *T. brucei* sequence did show significant homology to the *L. major* sequence within the predicted TM region of that latter protein, but also contained some charged amino acids that are given low scoring in matrices to predict TM domains. Therefore the effect of removing these charged residues *in silico* was determined. When a positive amino acid (K60) was substituted for neutral one (serine), a TM domain was predicted between residues 35-65. Sequences known to be linked to the internal membrane of the mitochondria (*D. melanogaster*) or with the plasma membrane in *S. typhimurium* (Menzel and Roth, 1981a and b) or in *E. coli* (Scarpulla and Soffer, 1978), were used as controls for predicting the TM domains.

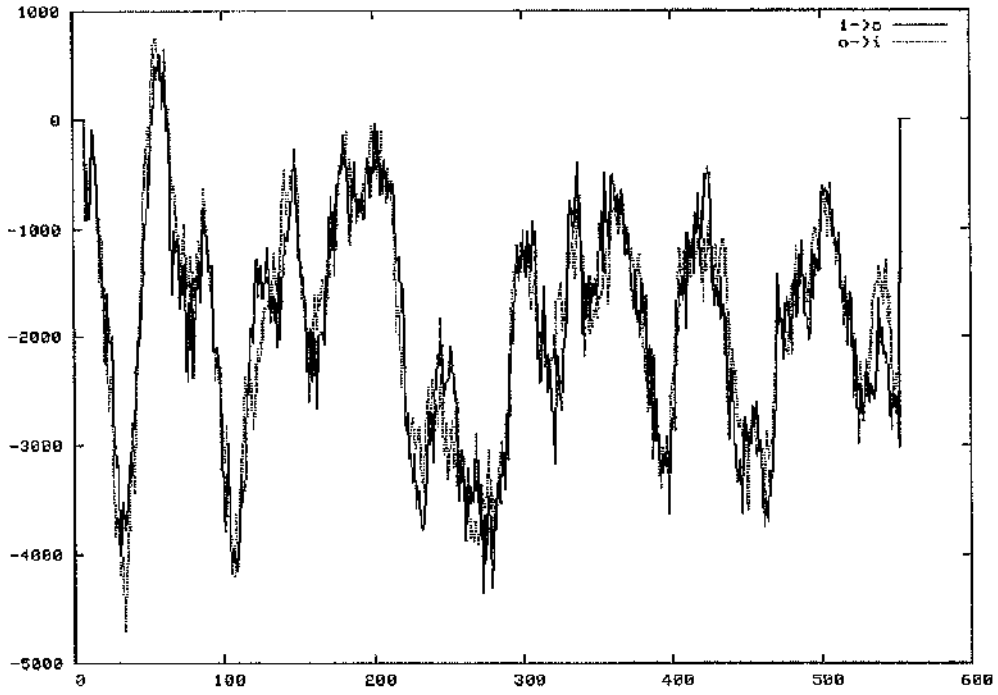


Figure 3.15: Hydrophobicity prediction by the TMpred programme for the *L. major* proline dehydrogenase. The algorithm is based on the statistical analysis of TMbase, a database of naturally occuring transmembrane proteins. The prediction is made using a combination of several weight-matrices for scoring (Hofmann and Stoffel, 1993).

The abscissa indicates amino acid numbers and the ordinate the hydrophobicity index.

A putative trans-membrane domain is situated between amino acids 45 to 69.

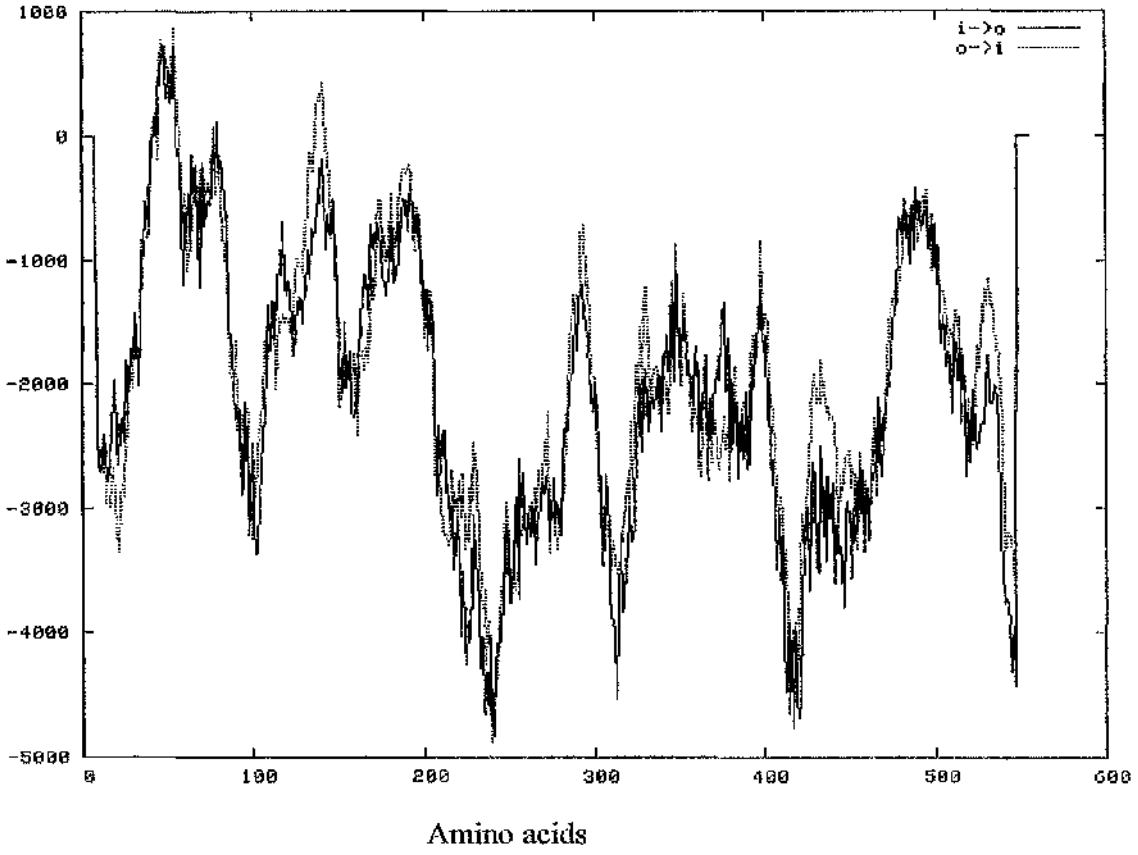


Figure 3.16: Hydrophobicity prediction by TMpred programme for *T. brucei* proline dehydrogenase mutated, K60>S.

A putative trans-membrane domain is identified as situated between residues 35 to 65

3.2.7- Phylogenetic analysis of proline dehydrogenase sequences

PRODH is an enzyme belonging to the mitochondrial proteome in eukaryotes and it is believed to be associated with the electron transport chain in prokaryotes (Muro-Pastor *et al.*, 1997). The endosymbiont hypothesis suggests that eukaryotic cells evolved from anaerobic organisms merging with α-proteobacteria when oxygen built up in the atmosphere leading to a stable endosymbiotic relationship with the α-proteobacteria using oxygen in energy generation (Cavalier-Smith, 1987; Kalberg *et al.*, 2000). Despite the fact that mitochondria have their own genomes, most of the proteins present in the mitochondrial organelle are synthesised using nuclear genes. A rudimentary phylogenetic analysis was performed using the PRODH amino acid sequences from *T. brucei* and *L. major* in order to determine their evolutionary relationships. This was to determine whether eukaryotic PRODH has an α-proteobacterial origin.

The sequences used in this study are detailed in Table 3.1. For prokaryote amino acid sequences, only the PRODH portion of the PRODH/P5CDH sequence (according to homology or annotation in the databases) was used in the analysis. Two different studies were performed to determine the evolutionary position of PRODH; the first study used entire sequences (Tree 1 to Tree 3) and the second study used the best-conserved Blocks predicted using the MEME software (Tree 4 to Tree 7). It is important to have good sequence alignment in order to produce good trees. Conserved Blocks represent the most highly conserved regions of the proteins, misaligned regions and phylogenetically uninformative regions of proteins are avoided and so trees from blocks should be more robust. Blocks chosen are shown in Appendix 2, they were selected using MEME program (Bailey and Elkan, 1994). MEME is a tool for discovering motifs, i.e. a sequence pattern that occurs repeatedly in a group of related protein sequences. Four methods of

phylogenetic analysis were used and can be divided into distance based and character based methods.

Distance-based methods use the amount of dissimilarity (the distance) between two aligned sequences to derive trees; three methods were applied.

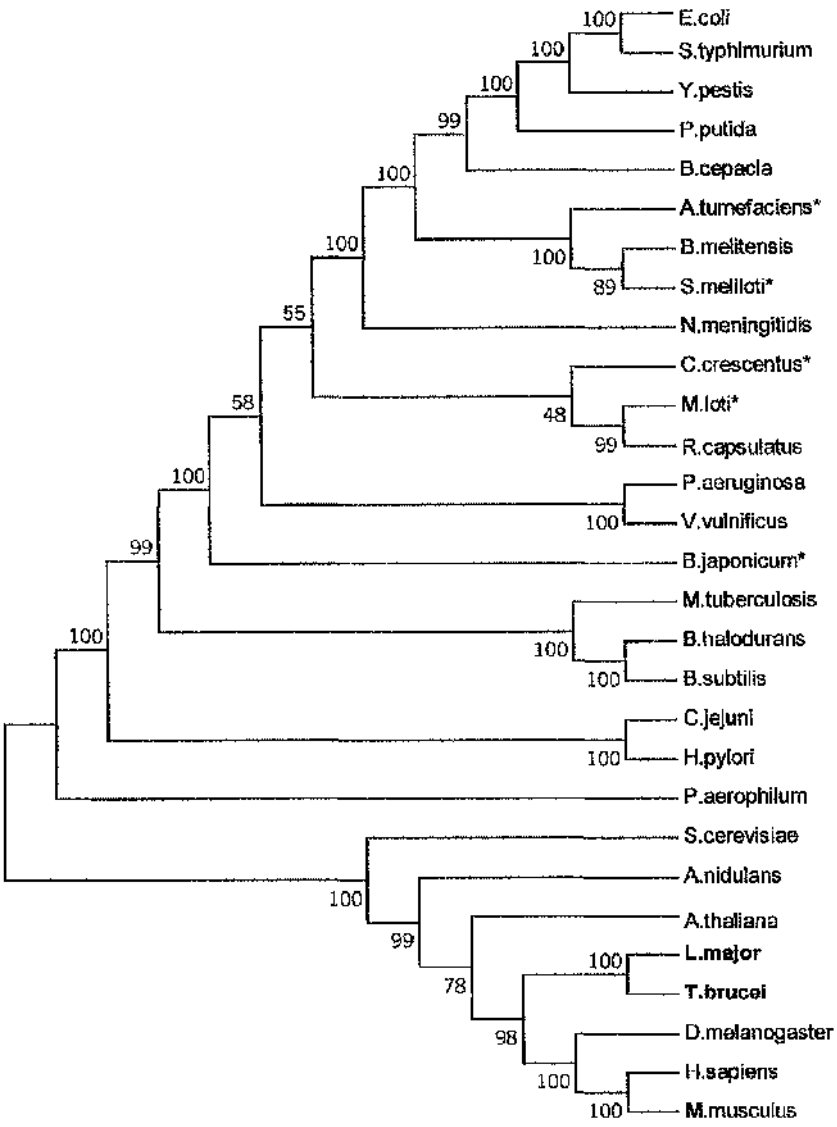
-Unweighted paired group method with arithmetic mean (UPGMA), is a phenetic algorithm. It joins tree branches based on the criterion of greatest similarity among pairs and averages of joined pairs.

-Neighbour-joining (NJ), is commonly applied with distance tree building regardless of the optimisation criterion.

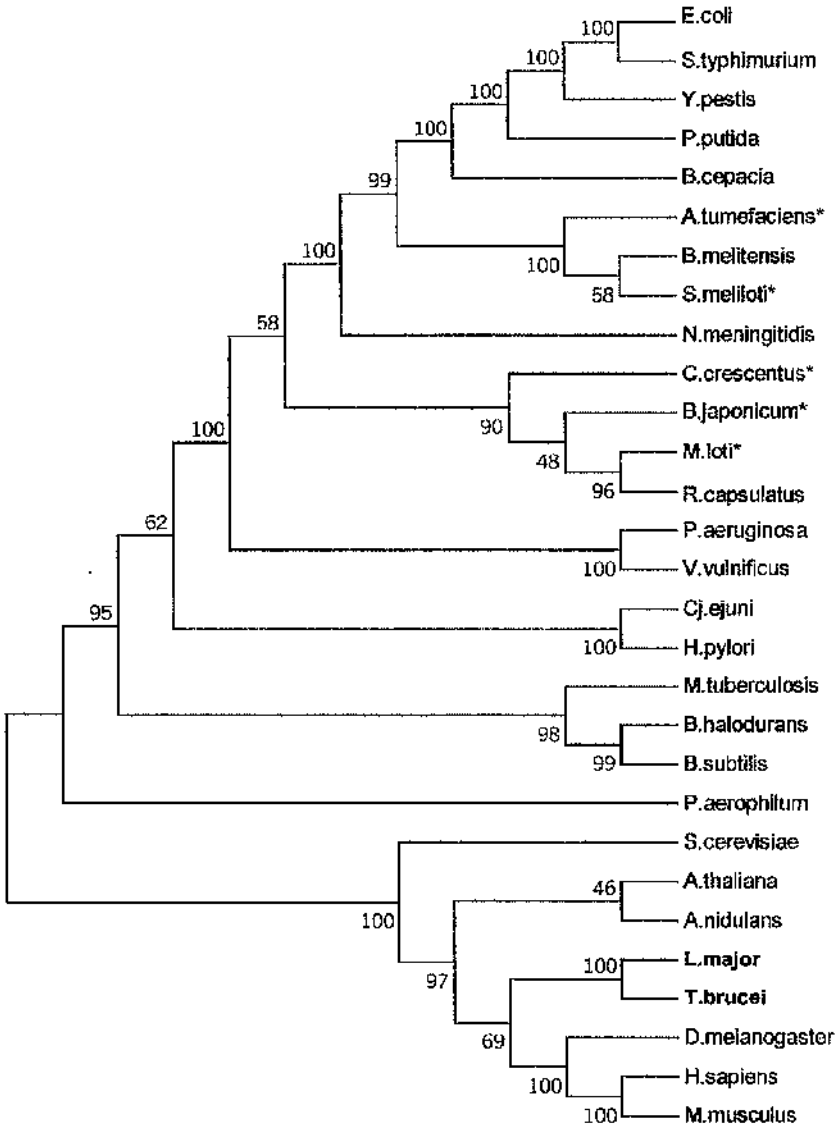
-Minimum evolution (ME), seeks to find the shortest tree that is consistent with the path lengths. It fixes the location of internal tree nodes based on the distance to an external node and then optimises the internal branch length according to the minimum measured error between these points.

Character-based methods allow the assessment of the reliability of each base position in an alignment on the basis of all other positions. Maximum parsimony (MP) is an optimisation criterion that adheres to the principle that the best explanation of the data is the simplest. The MP tree is the shortest, the one with the fewest changes.

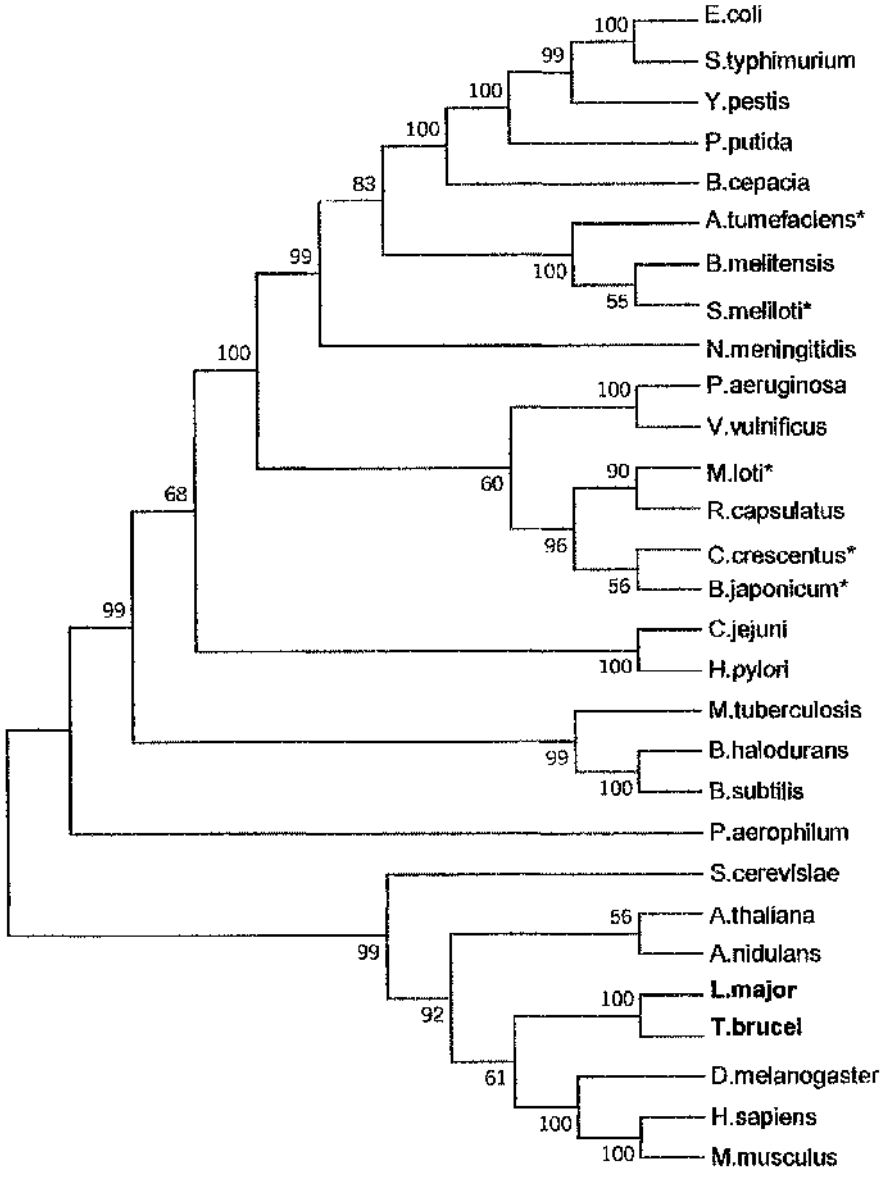
The trees show boostrap values, indicating the confidence value of the tree from 1000 random samples (indicated in %). All the trees are unrooted. All methods used gave us similar results, i.e. *L. major* and *T. brucei* PRODH are more closely related to eukaryotes than prokaryotes. Eukaryotic enzymes always cluster together, and away from the prokaryotic enzymes. The enzyme from α-proteobacteria cluster with other prokaryotes and away from eukaryotes. This indicates that the eukaryotic PRODH originated from the protoeukaryote rather than the protomitochondrial endosymbiont.



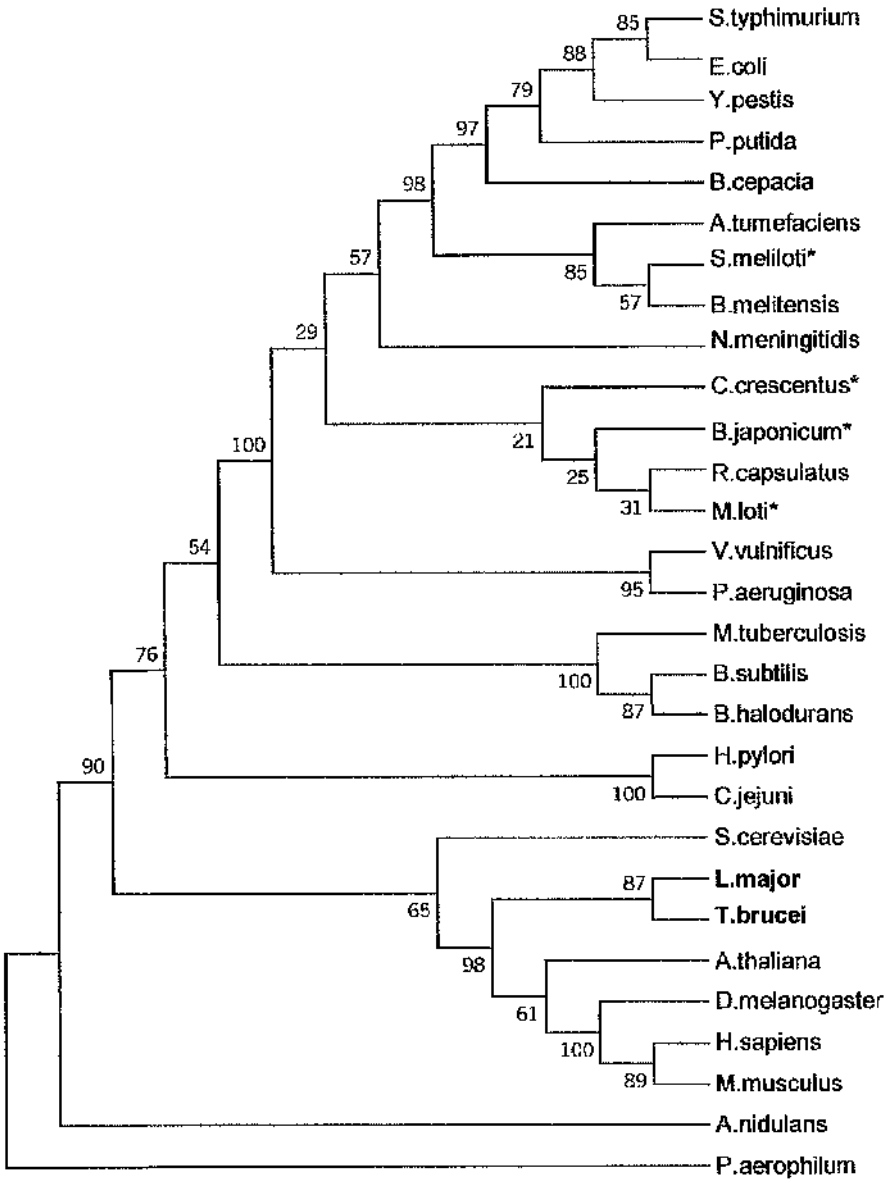
Tree 1: Phenogram tree for proteins of proline dehydrogenase amino acids sequences from different organisms. The tree was generated using MEGA, UPGMA method. The number above the branches represents bootstrap values (1000 pseudoreplicates). α-proteobacteria are highlighted with an asterisk.



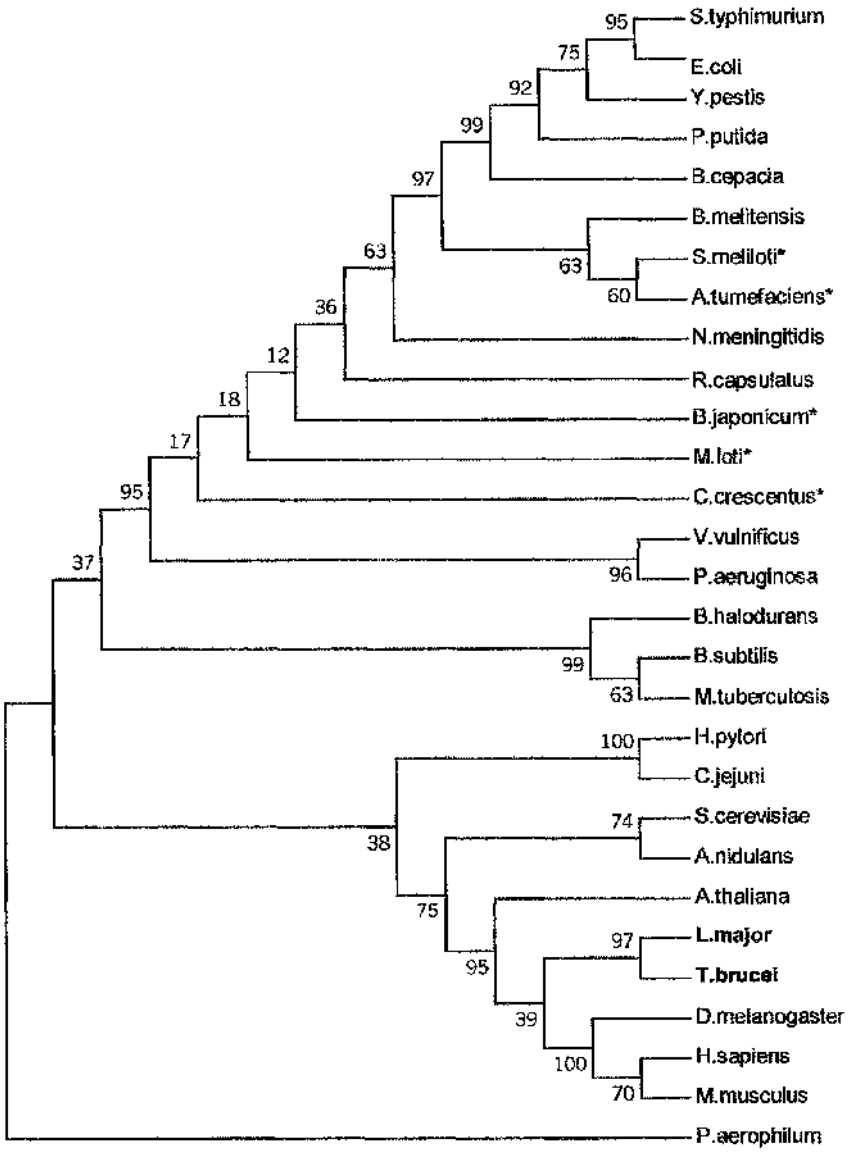
Tree 2: Phylogenetic tree for proteins of proline dehydrogenase amino acids sequences from different organisms. The tree was generated using MEGA, Neighbour-joining method. Number above the branches represents bootstrap values (1000 pseudoreplicates). α-proteobacteria are highlighted with an asterisk.



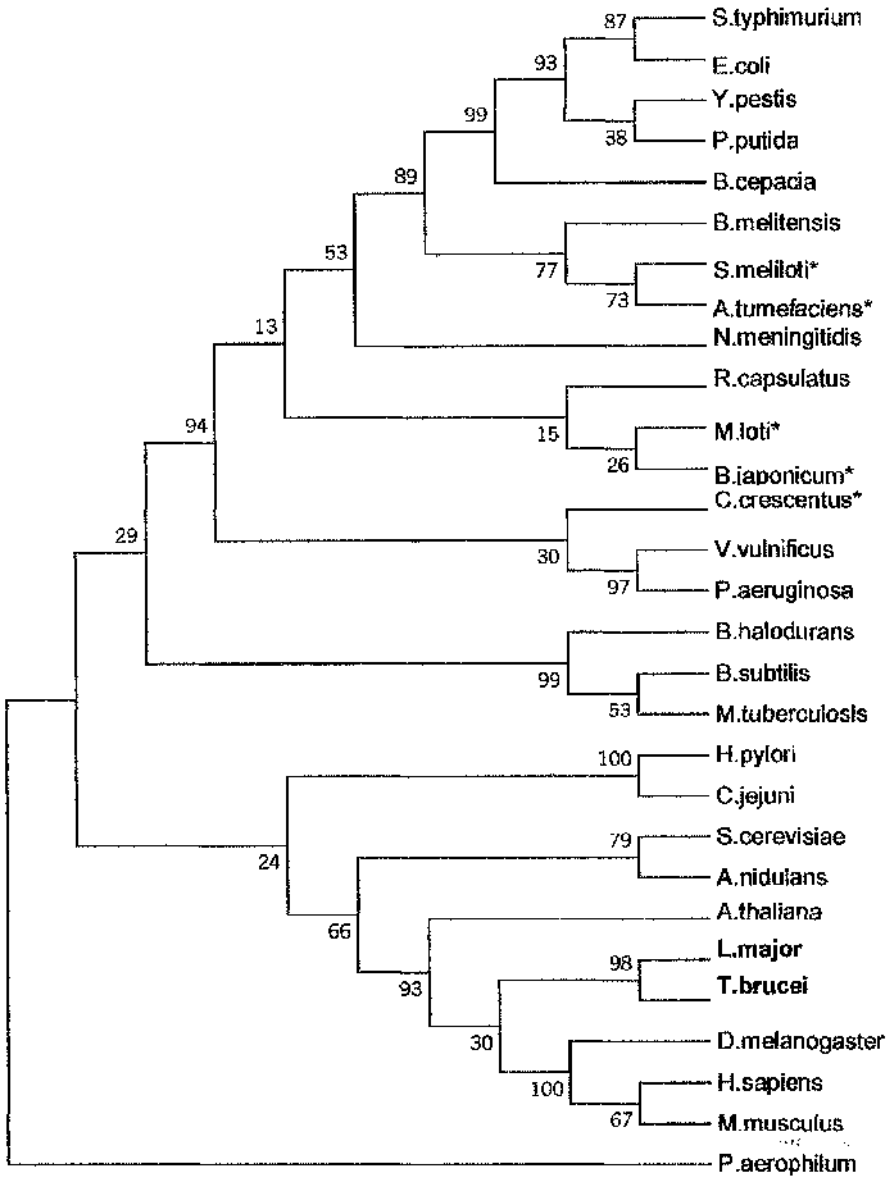
Tree 3: Phylogenetic tree for proteins of proline dehydrogenase amino acids sequences from different organisms. The tree was generated using MEGA, Minimum evolution method. The number above branches represents bootstrap values (1000 pseudoreplicates). α-proteobacteria are highlighted with an asterisk.



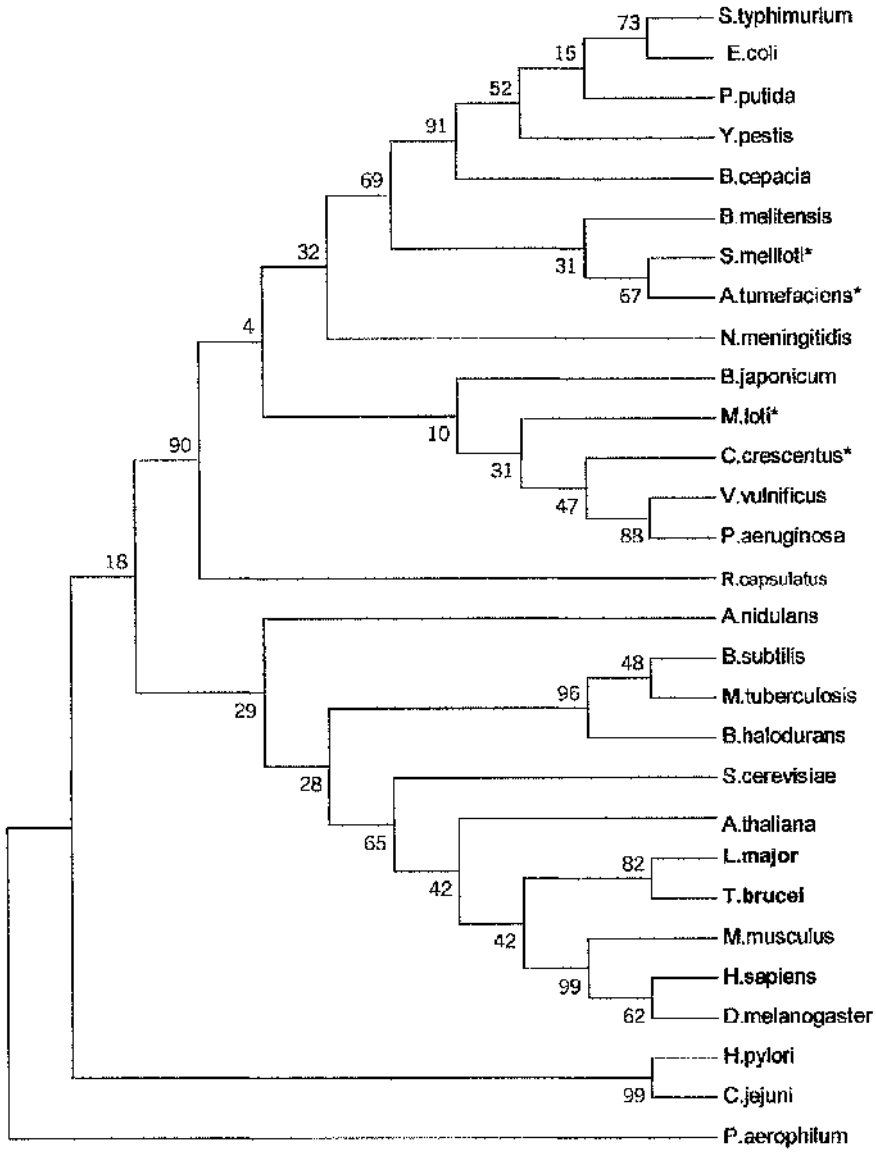
Tree 4: Phylogenetic tree for block of proline dehydrogenase amino acids sequences from different organisms. Conserved blocks were generated using MEME program. The tree was generated using MEGA, UPGMA method. The umber above the branches represents bootstrap values (1000 pseudoreplicates). α-proteobacteria are highlighted with an asterisk.



Tree 5: Phylogenetic tree for blocks of proline dehydrogenase amino acids sequences from different organisms. Conserved blocks were generated using MEME programe. The tree was generated using MEGA, Neighbour joining method. The number above the branches represents bootstrap values (1000 pseudoreplicates). α-proteobacteria are highlighted with an asterisk.



Tree 6: Phylogenetic tree for block of proline dehydrogenase amino acids sequences from different organisms. Conserved blocks were generated using MEME program. The tree was generated using MEGA, minimum evolution method. The number above the branches represents bootstrap values (1000 pseudoreplicates). α-proteobacteria are highlighted with an asterisk.



Tree 7: Phylogenetic tree for block of proline dehydrogenase amino acids sequences from different organisms. Conserved blocks were generated using MEME program. The tree was generated using MEGA, maximum parsimony method. The number above the branches represents bootstrap values (1000 pseudoreplicates). α-proteobacteria are highlighted with an asterisk.

3.3- Discussion

Genes encoding PRODH have been identified in both the *T. brucei* and *L. major* genomes. These genes encode proteins of 556 amino acids in *T. brucei* and 561 amino acids in *L. major*; in both case the molecular weight of the protein is predicted to be 63.8 kDa. The length of these sequences is comparable to other eukaryotic sequences, including *Homo sapiens* (600 amino acids) (Deuschle *et al.*, 2001) and *Arabidopsis thaliana* (499 amino acids) (Nakashima *et al.*, 1998). In both trypanosomatids, the *PRODH* gene is present as a single copy (as indicated by Southern blot analysis), which could facilitate the use of gene knock-out techniques to investigate the importance of this gene. Most trypanosomatid genes are present as a single copy (Clayton, 1999), which facilitates the knock-out technique. van Weelden *et al.*, 2003 used this technique to see the effect of deletion of aconitase in *T. brucei*. However, gene knock-out was not performed with *PRODH* as RNA interference (LaCount *et al.*, 2000) has become a preferred technique.

In prokaryotes e.g. *Escherichia coli*, *Salmonella typhimurium* or *Yersinia pestis*, PRODH and P5CDH are part of a single protein, called PutA, which is encoded by the *putA* gene (Becker and Thomas, 2001; Parkhill *et al.*, 2001; Vinod *et al.*, 2002). In these cells proline utilisation requires another gene *putP*, which encodes a proline permease, a Na^+ proline symporter. The *put* operon includes the two genes, *putA* and *putP*, which are separated by an intergenic control region (Ling *et al.*, 1994). In the absence of proline, PutA acts as a transcriptional repressor (in addition to its catalytic functions). It can regulate expression of *put* genes by binding to an operator (Muro-Pastor *et al.*, 1997). There is no evidence of similar genetic linkage between proline permease and proline dehydrogenase in eukaryotes. However this is not surprising since eukaryotes do not generally arrange their genes in operons.

In some eukaryotes including *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens*, the conversion of proline to glutamate takes place within the mitochondria (Small and Jones, 1990, Gogos *et al.*, 1999; Deuschle *et al.*, 2001). The first two steps of proline oxidation are carried out by two different polypeptides. The presence of PRODH in the inner membrane of the rat liver mitochondria shows that proline degradation is linked to this compartment, whereas the next step, the degradation of P5C, may take place within the mitochondrial matrix (Brunner and Neupert, 1969; Small and Jones, 1990). This compartmentalisation of proline degradation could also be present in *L. major* and *T. brucei*. Several types of analysis indicated that the P5CDH is separate from the *T. brucei* and *L. major* proline dehydrogenase protein. *PRODH* and *P5CDH* genes are on different chromosomes (7 and 26 for *PRODH* in *T. brucei* and *L. major*, respectively, and on chromosome 10 and 3, respectively, for *P5CDH*). The enzymes are therefore present on two different polypeptides as is the case for other eukaryotes.

The *L. major* and *T. brucei* PRODH have homology to other proline dehydogenases. Although the level of identity is relatively low, well-conserved motifs are found throughout the sequences. These regions are V+GAY (where + is K or R) and YLxxRRxxEN, where x is any amino acid. These regions, which are also found in *T. brucei* and *L. major* sequences, are believed to be essential for the activity of the PRODH enzyme (Straub *et al.*, 1996). They could be a part of the active site involved in binding the substrate or the cofactor FAD (Lee Y *et al.*, 2003). However some key *E. coli* residues, such as Leucine 513 appear to be substituted in *L. major* and *T. brucei* sequences by valine.Valine, like leucine is an aliphatic amino acid and is likely to have similar reactivity towards cofactor.

It is known that eukaryotic PRODHs are associated with the inner mitochondrial membrane (Wang and Brandriss, 1987); the oxidation of proline donates electrons to the

respiratory electron transport chain (Peng *et al.*, 1996). Wang (1987) describes PRODH as: "nuclearly encoded, cytoplasmically synthesized and mitochondrially imported". The *L. major* and *T. brucei* PRODHs are both predicted to possess a mitochondrial signal peptide, with a cleavage site occurring after amino acid 21 for *L. major*. The proteins are likely to be synthesised in the cytosol and then transported from the cytosol to the mitochondrion. Usually a group of 20 to 40 amino acids, many of which carry a positive charge, interspersed with some hydrophilic residues such as serine or threonine at the N-terminus, constitute the mitochondrial targeting sequence (MTS) (Omura, 1998; Nielsen *et al.*, 1997). With the *L. major* and *T. brucei* PRODH the predicted MTS comply with this rule, they are rich in positive amino acids and both contain serine or threonine residues (Table 3.3). The MTS is recognised by a receptor protein complex on the outer and inner mitochondrial membranes (the Tim-Tom complex) (Koehler, 2000) and this system carries the protein inside. Immunolocalisation experiments using specific PRODH antibodies and fluorescence microscopy, in order to show the exact localisation of the protein in the parasites will be of interest in futures studies. However the likelihood for a mitochondrial localisation of PRODH is high, given the presence of a mitochondrial targeting sequence and the situation in other eukaryotes.

Phylogenetic analysis performed with PRODH sequences shows that the *L. major* and *T. brucei* PRODHs appear to be evolutionarily closer to other eukaryotic versions of the enzyme than to prokaryotic enzymes. Eukaryotic mitochondria are believed to be derived from an endosymbiotic relationship between the host cell and a α-proteobacterium, close to the extant genus *Rickettsia* (Andersson *et al.*, 2003). However, approximately 50 % of the nuclear-encoded mitochondrial proteins have bacterial homologues (Karlberg *et al.*, 2000). All of the eukaryotic PRODHs, for which a subcellular localisation is known, are mitochondrial and all cluster to each other. It was considered of interest to compare the

trypanosomatids and other eukaryotic PRODHs and with those of *Rickettsia* spp (*R. cororii* and *R. prowazekii* whose genomes can be probed by FASTA server at the European Bioinformatic Institute (EBI) (www.ebi.ac.uk)). However, neither of these species appear to have PRODH (based on homology to prokaryotic and eukaryotic PRODH). Other α-proteobacteria, which do have PRODH, appear to cluster separately from the eukaryotic grouping, but interspersed within a group of other prokaryotic PRODHs. These data are consistent with PRODH in eukaryotes being derived from the non α-proteobacterial ancestor of the merger that yielded mitochondriate cells. Given that trypanosomes are evolutionary believed to be ancient, this indicates that it was a very early eukaryote that found advantage in metabolising proline in the mitochondrion (presumably by convergent evolution to ensure proximity to the electron transport chain). Given that *Rickettsia* do not have PRODH, it is possible that the proto mitochondrial endosymbiont also lacked this enzyme, which is why the protoeukaryote needed to donate this to the mitochondriate chimera.

Chapter 4

Regulation of expression of *T. brucei* and *L. major* proline dehydrogenase

4.1- Introduction

Gene organisation and gene expression are peculiar in trypanosomatids when compared with higher eukaryotes. A single intron has been found within a mRNA in *T. brucei* (Mair *et al.*, 2000). The regulation of gene expression is generally not controlled at the level of transcription since trypanosomatid genes are arranged in polycistronic units; control is mainly post-transcriptional. Genes are present as polycistronic units in the genome and individual mRNAs are post- or co-transcriptionally modified by trans-splicing and polyadenylation (Graham, 1995). All kinetoplastid mRNAs have, at their 5' end, a common sequence called the spliced leader (SL) of 39-41 nucleotides, containing a cap structure (including 7-methyl guanosine as the terminal nucleoside), which is added post-transcriptionally via the trans-splicing pathway (Campbell *et al.*, 2000). The SL sequence is well conserved between different trypanosomatids (Agami and Shapira, 1992). Primers designed based on the *L. mexicana* SL sequence are useful in RT-PCR experiments with RNA to identify the 5' extremity of stable transcript. The length of the sequence between this spliced leader and the first methionine of different genes is variable and can be important for the regulation of gene expression (Vanhamme and Pays, 1995). In addition to this 5' trans-splicing process, a 3' polyadenylation occurs for all mRNA transcripts.

The regulation of the trans-splicing and polyadenylation process may have a key role in the regulation of gene expression, by stabilising the mRNA or by inducing its degradation (Stiles *et al.*, 1999).

It was of interest to determine the levels of mRNA corresponding to the *PRODH* gene in *T. brucei* and to map the untranslated regions of the transcript. Moreover, as previously mentioned (Evans and Brown, 1972), PCF *T. brucei* is able to grow in a medium containing proline or glucose as an energy source (see also Chapter 6). It was therefore of interest to investigate whether there is regulation of PRODH activity in response to different energy sources. In prokaryotes, such as *E. coli*, glucose, if available as an energy source, is used in preference to other sugars. If this organism has, for example, both glucose and lactose in the medium, it metabolises the glucose first and represses the use of lactose. This phenomenon is called catabolite repression (Bruckner and Titgmeyer, 2002). ter Kuile (1997) suggested that a form of catabolite repression could be operative in allowing procyclic *T. brucei* to show a preference in the use of glucose or proline depending on the medium used. In this study, numerous attempts, based on work by Obungu *et al.*, 1999, with *T. congolenese*, to measure PRODH activity in crude lysates of *T. brucei*, grown in standard SDM79 medium, failed to yield as high activity. Therefore, it was decided to investigate whether proline dehydrogenase activity was somehow regulated by growth conditions.

4.2- Results

4.2.1- Identification of the spliced leader acceptor sites for proline dehydrogenase genes of *T. brucei* and *L. major*

The 5' regions of the two *PRODH* genes were identified using reverse transcriptase PCR (RT-PCR). Internal primers, specific to the *PRODH* genes and two primers from the spliced leader sequence (primers common to both the *T. brucei* and *L. major* sequences) were used for amplification of the 5' UTR of these two genes (see section 2.2.3 and Table 2.1 for methods and oligonucleotide sequences).

The result of the first amplification is shown in Figure 4.1

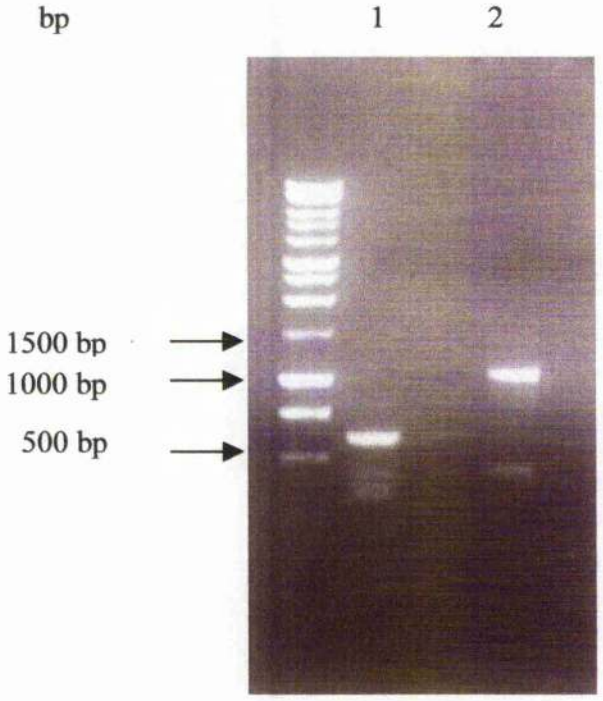


Figure 4.1: Ethidium bromide gel representing the first PCR of the RT-PCR of RNA from *T. brucei* and *L. major* amplified with primers from the SL and PRODH

1: Fragment of *T. brucei* proline dehydrogenase was amplified with primers SL1 and RTTb1

2: Fragment of *L. major* proline dehydrogenase was amplified with primers SL1 and RTLm1

For primer sequences see section 2.2.3 and Table 2.1

A ~ 600 bp fragment was obtained from *T. brucei* cDNA and a ~ 1000 bp fragment for *L. major* cDNA. These products were gel extracted and re-amplified using the SL2 primer and RTTb1 and RTLm1 for *T. brucei* and *L. major*, respectively (Figure 4.2).

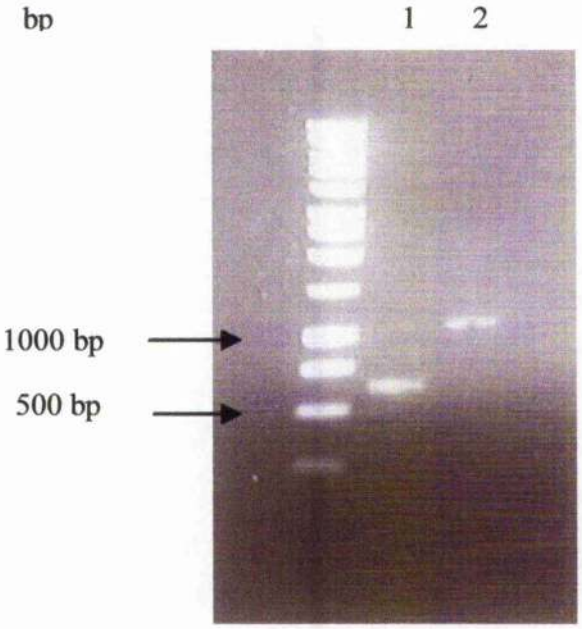


Figure 4.2: Ethidium bromide gel representing the second PCR of the RT-PCR after gel purification
1: *T. brucei* proline dehydrogenase amplified with primers SL2 and RTTb1
2: *L. major* proline dehydrogenase amplified with primers SL2 and RTLm1
For primer sequences see section 2.2.3

After purification from the gel, these two products were cloned in pGEM-T vector and sequenced. The 5' UTR regions (Figure 4.3 and Figure 4.4) were identified by translating the two sequences using Vector NTI and aligning these sequences with their respective gDNA.

```
   1    AAGCTATGCA TCCAACGCGT TGGGAGCTCT CCCATATGGT CGACCTGCAG
  51    GCGGCCGCGA ATTCACTAGT GATTAGTATC AGTTTCTGTA CTTTATTGAA
 101    TTGACGCACA CACTCTCTCG CTCTCTCCTG ATCTACTCCC CTCCCCCCTT
 151    TCCCTCCCCT CTAGCAAGCA AGCGCTTTAT ATTAGTAAAT TTTTTCTTGG
 201    TTGCTGTTGG TGCCGTGTCG TCTGTGTCTG TATCTGTGTC TGTGTCTTTA
 251    CCTTATCATC AATTTACAGG CGTGTGTGTG TGTGTGCGCG TGACTCGATC
 301    TGTCTCCCAC CAACACCACC GCCACCACCA CCACCACCGT GACCCCCTCG
 351    CTCTCTTTCA TGCCATATTC ACCACTCTAA TTCTTCCAAC AAAAAAAACG
 401    AAAACGCACG CAGAGAGGAA GTCGCTTAGT TTTAGCTAGC TACCTGCCTT
 451    GTTTTCTCTT TTCTATTGTG TGTAGTTTTG TTGGCACTCT CACCGTCTCA
 501    GCTCTCCTCC TCTTCCATCC TCCCTGCTCC CTCGGCAGCG TCTCGTCATA
 551    CCTGCTCGCC GTCAGGCCCG CTAGTCGCCC GTCCCCCAAA AGATGCGTCG
 601    TCTCCTCCCC CTGCGACCGG CAGCTGTGGC CTTTGCCGGC TCTGCTCGTC
 651    ACTCCTCCTT GACAATGCAR GACAAGCAGC CGAAGCTGCC GAACTTCAAC
 701    GACGACACAA CCTACCGCCA GCGCTCGGCA TGGTACTTGA TAAAGGCGTT
 751    GGTGGTGCTC CGCCTGTGCA GTGTGAACTA TTTGGCAATG AACTCGGTGC
 801    CGCTGATGAA GAGAGTGGAA AAGATCCTTG GCAGCAAGCT TACCTACAGC
 851    ATCCTCGTCA AGAAGTCCTT CTACAACTAC TTCTGCGCGG GCGAAAACGA
 901    CCAGGAGCTG CGCGACACGG TGCGAAAGCT TTCACGCAAC AACATCGGCG
 951    CTGTACTCGA CTACGCGGCG GAGGCCGACA CGGAGGGCTT CGCACCGGAG
1001    CCGGGTGTGG CGTCCGGCCC CGATATTTCG ATGTCTAGTC TCGTTATGAA
1051    GCCCAATGTT CAGTACCCAA TGGACGAGGG GTTCTTTAAC GAGAACATGA
1101    AGCTAATCGA ATTCCCGCGG CCGCCATGGC GGCCGGGAGC ATGCGACGTC
1151    GGGCCC
```

Figure 4.3: Nucleotide sequence of the 5' UTR of *L. major* proline dehydrogenase. Letters highlighted in blue represent the spliced leader sequence, and in pink the codon of the first methionine residue.

```
   1    AAGCTATGCA TCCAACGCGT TGGGAGCTCT CCCATATGGT CGACCTGCAG
  51    GCGGCCGCGA ATTCACTAGT GATAACTAAC GCTATTATTA GAACAGTTTC
 101    TGTACTATAT TGAGTAATTT AAGCTATCAT GTTTCGCCTC CTGTGCCGAC
 151    GCTCCGCCGC GAATGTAAAA ATGCAAGACG CCAGTCTCCG CACCATGACA
 201    AAAGTGGATT TCTCGGACCC ATCCATATTT CGTCAGAAAT CTCTGTGGTG
 251    GTTGTTGCGC GCACTCTTTG TGCTCCGCAT TTGCAAGTTT GAATTCATTT
 301    CCAACAATTC TGTGGCTCTT ATGAAGCGTG CTGAGGCAAT TTTTGGACCC
 351    TTTTTGACCT ATAACACACT CGTGAAGGGG ACAGTGTACG GCCACTTTTG
 401    TGCTGGCGAA TCCGATCGTG AAGTTAAGAA CACCGTCAAA TCGCTTGAAA
 451    ATTTGGGTAT TGGATCTGTT CTTGACTATG CCGCTGAAGC TGAAGCTGAG
 501    GGTTTCGCAC CCTCCCCAGG TATTGCAGAA GCGCCGAACT TGTCTATGGC
 551    AAGTCTTGTC AACAACACAT CCGTTACCTA TCTGCCTCAC AAGCAAGCCT
 601    TTGACGAAAA CATGAAACTA TACGTAATGT GTGTTTTACA TGCCGCACTA
 651    CATAAACCGG AGGGCGGCGT TGGATTGGCT GCTGTTAAGG TGACAGGCAT
 701    GTAATCGAAT TCCCGCGGCC GCCATGGCGG CCGGGAGCAT GCGACGTCGG
 751    GCCC
```

Figure 4.4: Nucleotide sequence of the 5'UTR of *T. brucei* proline dehydrogenase. Letters highlighted in blue represent the spliced leader sequence, and in pink the codon of the first methionine residue.

For both organisms, the first ATG codon for *PRODH* and the SL sites were determined. The untranslated region between the SL acceptor site and the start codon was 502 bp for *L. major* and 15 bp for *T. brucei*.

4.2.2- Proline dehydrogenase gene expression during the *T. brucei* and *L. major* life cycles

The level of expression of the *PRODH* gene in different life cycle stages of *T. brucei* and *L. major* was studied by Northern blot analysis. A nitrocellulose membrane containing total RNA from PCF and BSF *T. brucei* was probed with the entire proline dehydrogenase ORF.

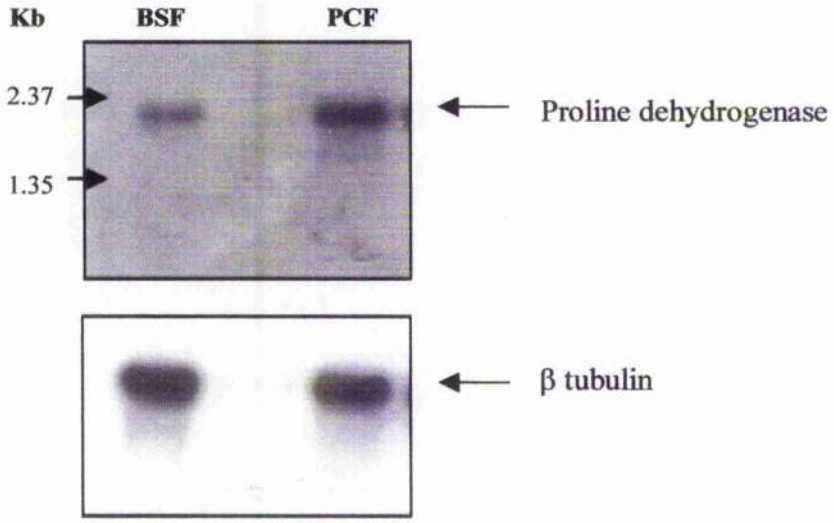


Figure 4.5: Northern blot analysis of RNA of *T. brucei* life cycles stages. The Northern blot was probed with the proline dehydrogenase gene and a β tubulin gene was used as a control. BSF is bloodstream form and PCF is procyclic form. Approximately 4 μg of RNA was loaded for each sample.

The results (Figure 4.5) revealed the *PRODH* gene was transcribed in both procyclic and bloodstream stages. The transcript was around 2 Kb in size. The RNA levels in procyclic stages, however, were higher than in bloodstream stages. The blot was also probed with the β-tubulin gene in order to control for the quantity of RNA.

The levels of steady state RNA from *PRODH* in the different life stages of *L. major* and *L. mexicana* were also studied and the results are shown in Figure 4.6. A transcript of around 3 Kb was identified.

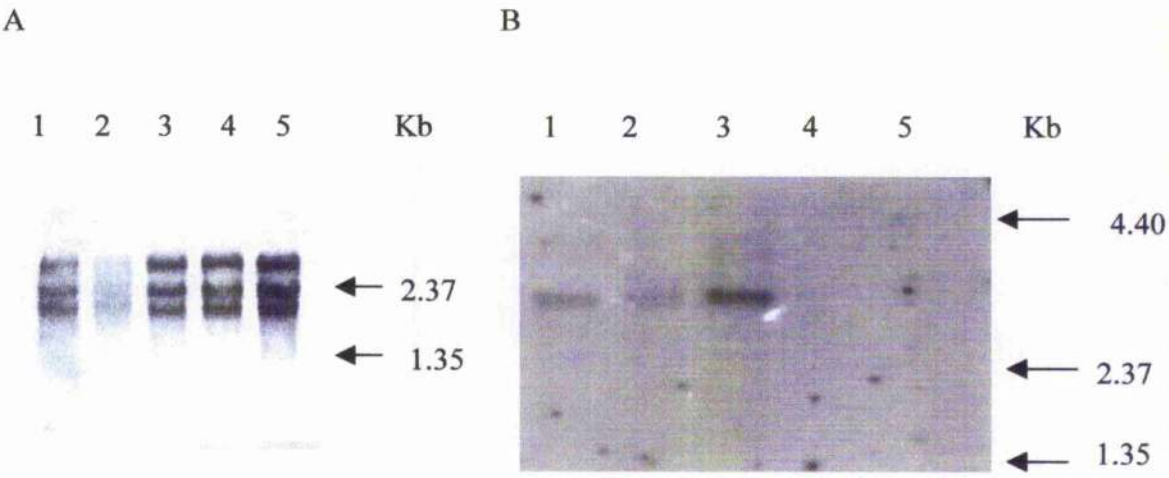


Figure 4.6: Northern blot analysis of RNA from different stages of *L. major* and *L. mexicana*.
A: Methylene blue stain of the membrane as a control of the quantity of RNA, the ribosomal RNA bands are clearly visible
B: Northern blot showing proline dehydrogenase transcript in *L. major* metacyclic promastigotes (lane 1), *L. mexicana* metacyclic promastigotes (lane 2), *L. major* promastigotes (lane 3), *L. mexicana* promastigotes (lane 4), *L. mexicana* amastigotes (lane 5).
A and B are the same membrane. Approximately 4 μg of RNA was loaded for each sample.

The Northern blot containing RNA from *L. major* and *L. mexicana* metacyclic promastigotes from stationary phase, *L. major* and *L. mexicana* promastigotes and *L. mexicana* amastigotes was probed using *L. major PRODH* ORF. A methylene blue stain of the membrane is shown (Figure 4.6A) as a control of the quantity of RNA.The stronger

signal for *PRODH* was observed for *L. major* promastigotes but the gene was also expressed in *L. major* and *L. mexicana* metacyclics. No transcript was observed in *L. mexicana* promastigote or amastigote samples. Conceivably it may be that lane 2 of the blot is a contamination from lane 3 and that *L. major PRODH* ORF recognises only *L. major PRODH*. The data suggest that *L. major PRODH* is expressed in metacyclics and promastigotes. The situation for the amastigotes of *L. major* was not investigated. These experiments were not repeated and a decision was taken to focus on the situation in *T. brucei* rather than *Leishmania*.

4.2.3- Differential expression of proline dehydrogenase in *T. brucei* grown in defined media

A modified SDM79 medium (Appendix 1) without proline and glucose was prepared in order to test the viability of trypanosomes in the absence or presence of their main predicted energy sources (see Chapter 6). This medium was supplemented by proline or glucose and it was found that *T. brucei* procyclic cells were able to grow in the presence of either glucose or proline as the energy source. It was therefore of interest to see the level of the PRODH transcript when parasites were grown in the absence or presence of glucose in the medium.

T. brucei procyclics were grown in two different media, one complete containing glucose and proline, and another lacking glucose and containing only proline as an energy source. Total RNA from these parasites was extracted and a Northern blot was probed with *T. brucei PRODH* ORF. The result (Figure 4.7) shows that *PRODH* of *T. brucei* was expressed to approximately the same level when parasites were grown in the two different conditions. The marginal decrease in signal in cells grown in medium lacking of glucose, can be due to the loading difference seen in Figure 4.7B.

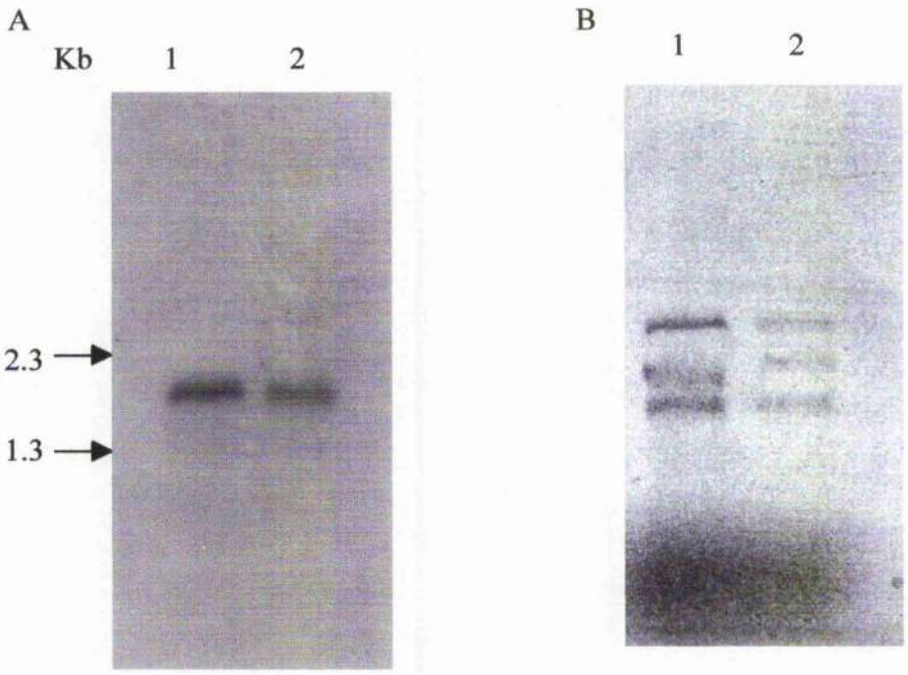


Figure 4.7: Northern blot analysis of RNA from *T. brucei* growing in specific medium.

A: Northern blot showing proline dehydrogenase transcript in *T. brucei* cells growing in modified SDM79 with proline and glucose at 10 mM (lane 1) and the same medium lacking glucose (lane 2), the blot was probed with the *T. brucei* proline dehydrogenase ORF.

B: Methylene blue stain of the same membrane as a control of the quantity of RNA, showing ribosomal RNA.

4.2.4- Western blot analysis of *T. brucei* proline dehydrogenase

To investigate the expression of PRODH protein in the *T. brucei* procyclic parasites grown in the two different media, Western blots were performed. Lysates from both *T. brucei* cultures were run in an SDS PAGE gel, antibodies used during this experiment were raised in a rabbit by immunisation with purified recombinant PRODH protein (see section 2.3.5).

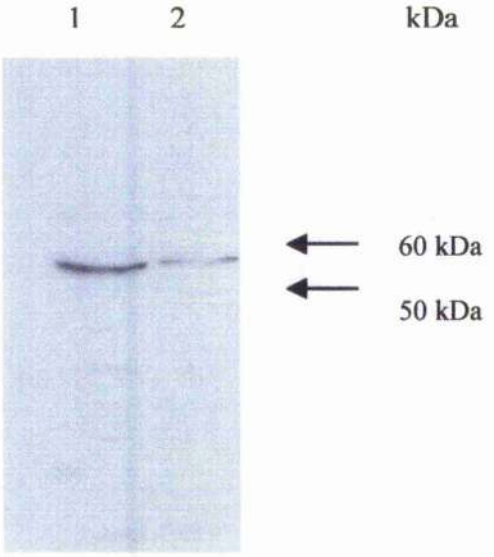


Figure 4.8: Detection of proline dehydrogenase using rabbit anti-proline dehydrogenase antibodies (1:3000 dilution). Lane 1, Lysate of *T. brucei* grown in medium containing proline but lacking glucose; lane 2, lysate of *T. brucei* grown in medium containing glucose and proline

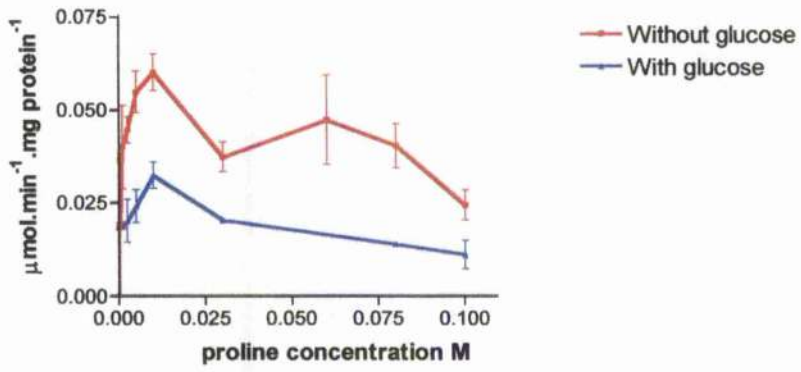
One band was present in the blot at approximately 60 kDa (Figure 4.8). This could be the PRODH from *T. brucei*. The signal was not strong, but it was present in cells from both growth culture conditions. The proteins seemed to be present even when the parasites were grown in the presence of glucose. The significance of the apparent differences in

protein levels is not clear as the experiment was not successfully repeated during the time available.

4.2.5- Proline dehydrogenase activity in *T. brucei*

The activity of PRODH in crude lysates of *T. brucei* was determined using dichlorophenolindophenol (DCPIP) as a terminal electron acceptor (section 2.3.9). The reduction of the dye is associated with a decrease in absorbance at 600 nm. The activity for PRODH was calculated by subtracting the background in the absence of proline as a substrate from the rate of dye reduction in the presence of various proline concentrations. The protein concentration in crude lysates of $2x10^8$ parasites was determined by Bradford's method to determine the specific activity. *T. brucei* procyclic parasites were grown in two different media, with or without glucose, to see the effect of this on the activity of this enzyme.

A



B

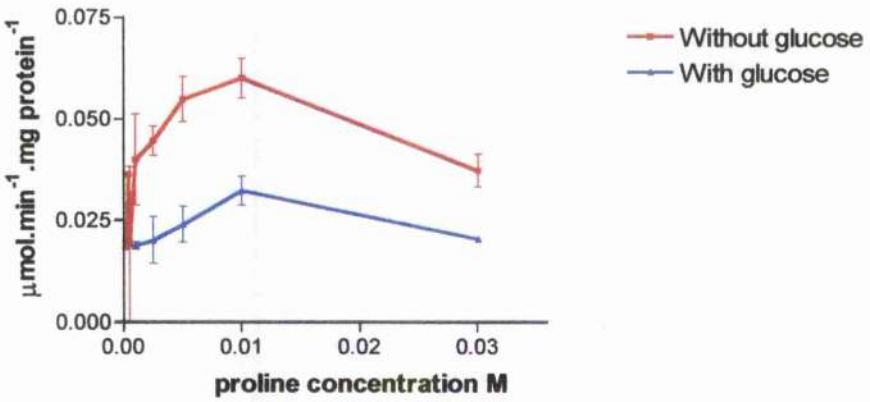


Figure 4.9: Proline dehydrogenase activity as a function of proline concentration. The red line represents the activity of *T. brucei* procyclic cells grown in a medium lacking glucose with the presence of proline (10 mM) only as an energy source. The blue line represents the activity of the *T. brucei* procyclic cells grown in the presence of glucose and proline at 10 mM each. Different concentrations of proline, as indicated, were added for the proline dehydrogenase activity assay.

Graph B shows only the 0-0.03 M proline range from graph A.

Error bars represent ± SEM, n = 6

The specific activity of *T. brucei* PRODH was found to be maximal using a proline concentration of 10 mM in the assay (Figure 4.9). The activity was higher when the parasites were grown in the absence of glucose. In these conditions the activity was 0.054 ± 0.004 μmol.min^{-1}.mg protein^{-1}, compared with 0.027 ± 0.004 μmol.min^{-1}.mg protein^{-1} when the medium was complete. A t-test, found these results to be significantly different, $p<0.05$.

In both media, the specific activity increased with proline concentration to reach a maximum at 10 mM proline and then at higher concentrations apparent substrate inhibition was observed, with a consequent decrease in specific activity. This could explain why preliminary experiments, seeking activity in cells grown with glucose, failed to yield activity as proline was used at 0.2 M in these experiments.

Interestingly, Obungu *et al.*, 1999 found in *T. congolense* a specific activity for proline dehydrogenase of 55.2 nmol.min^{-1}.mg protein^{-1} using different conditions. In their assay, 0.2 M of proline was present (Obungu *et al.*, 1999). The measurable activity of the *T. brucei* enzyme seemed to be sensitive to a variation in medium composition. Adding glucose to the medium resulted in parasites with lower PRODH activity.

4.2.6- Proline uptake in *T. brucei* procyclic forms grown in defined medium

Since PRODH activity was higher in procyclics grown in a medium lacking glucose, it was of interest to see if these growth differences could affect the kinetics of proline transport too. The kinetics of proline transport were studied using various proline concentrations for an uptake period of 30 s using an oil stop transport assay as described in section 2.4.6.

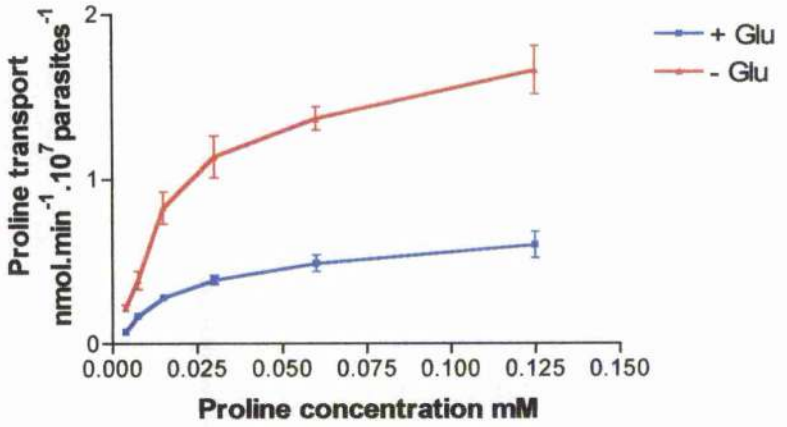


Figure 4.10: Uptake of proline into *T. brucei* procyclics grown in defined medium. The blue line represents the transport of proline into *T. brucei* grown in SDM 79 medium containing glucose and proline at 10 mM. The red line represents the transport of proline into *T. brucei* grown in medium lacking glucose. Uptake was allowed to proceed for 30s.
Error bars represent ± SEM, n = 6

Procyclics grown in both conditions (medium containing or lacking of glucose) showed a hyberbolic uptake curve (Figure 4.10) and kinetic characteristics for the proline transporter were determined. When parasites were grown in the presence of glucose, the V_{max} was 0.7 ± 0.06 nmol.min^{-1}.10^7 parasites^{-1} and the K_m was 21 ± 2.9 μM. L'Hostis *et al.* (1993), had previously reported a K_m of 19 μM and a V_{max} of 1.7 nmol.min^{-1}.10^7 parasites^{-1}, similar to the results reported here. Cells grown in the absence of glucose had a V_{max} of 1.8 ± 0.13 nmol.min^{-1}.10^7 parasites^{-1} and the K_m was 18 ± 0.0042 μM. Thus the V_{max} for proline of cells grown in the absence of glucose was much higher (significantly different by t-test, $p<0.05$) and this correlates well with the higher PRODH specific activity when cells were grown in this same medium. Without glucose in the medium, more proline entered the cells and it was more rapidly degraded (as indicated by the higher specific activity for PRODH).

The presence of glucose in the medium therefore appears to diminish proline metabolism and uptake into the cells.

4.2.7- Use of the glucose analogue 2-deoxyglucose (2-DOG) to study the growth of parasites

Proline metabolism appears to respond to the availability of glucose in procyclic *T. brucei*. Other pathways in procyclic *T. brucei*, exemplified by the expression of genes involved in production of the procyclin coat of these cells, also appear to be regulated in response to levels of glucose. In the absence of glucose, procyclics cells switch to expression of the non-glycosylated GPEET variant of the procyclin coat (Morris *et al.*, 2002). Inclusion of glycerol in the medium retards loss of expression of GPEET (Vassella *et al.*, 2000). Recently it has been shown that mitochondrial metabolic intermediates (such as pyruvate whose levels were controlled by knocking down the pyruvate dehydrogenase) influence expression of GPEET (Vassella *et al.*, 2003). It was of interest to determine whether proline metabolism responded directly to glucose, or to metabolic activity associated with glucose.

The effect of the non-metabolisable 2-DOG (which can be phosphorylated but not further used) was tested. To study the effect of the glucose analogue on *T. brucei* growth, the parasites generated by RNA interference were used (see chapter 6). Chapter 6 details the phenotype of the cells grown in the presence or absence of tetracycline, which is the inducer for the RNAi construct *Aprodh*. When tetracycline was present, and PRODH expression was switched off, the cells were unable to grow in the absence of glucose, indicating that proline can no longer support growth in these cells. The cells were grown in the SDM79-derived medium without glucose and proline and supplemented by proline,

glucose or 2-DOG at 10 mM. The Alamar blue assay was used to estimate the number of viable cells under the different conditions used.

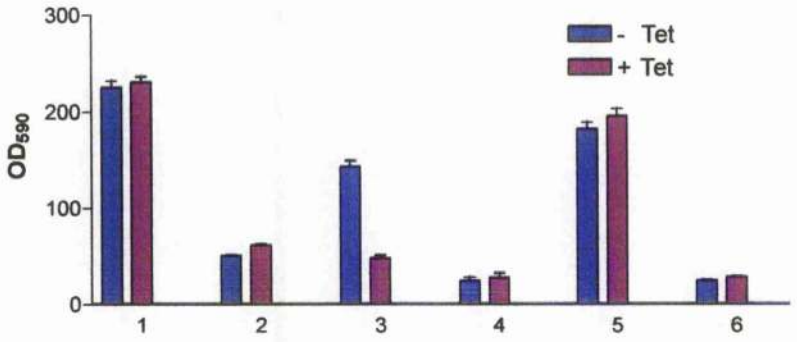


Figure 4.11: Growth of *T. brucei Δprodh* procyclic forms induced and non-induced by tetracycline in a medium lacking of proline and glucose with the following additives. Lane 1; glucose 10 mM, proline 10 mM; **lane 2**; glucose 0 mM, proline 0 mM; **lane 3**; glucose 0 mM, proline 10 mM; **lane 4**; glucose 0 mM, proline 10 mM, 2-DOG 10 mM; **lane 5**; glucose 10 mM, proline 10 mM, 2-DOG 10 mM; **lane 6**; glucose 0 mM, proline 0 mM, 2-DOG 10 mM.

The growth was deduced by the Alamar blue fluorescence estimation (section 2.1.3).

Error bars represent ± SEM, n = 3

The cells were able to grow in the presence of glucose and proline and in proline only when the PRODH was still active, in the absence of tetracycline (Figure 4.11, lane 3). *T. brucei* cannot grow on 2-DOG as an energy source, as shown in Figure 4.11, lane 6. When proline and 2-DOG were present together in the medium, the cells were not able to grow and use proline as energy source (Figure 4.11, lane 4). 2-DOG is not toxic, since cells can grow in the presence of 2-DOG and glucose (Figure 4.11, lane 5). These results indicate

that proline metabolism responds to 2-DOG and presumably to also glucose itself rather than to the cell's metabolic response to glucose. *T. brucei* parasites may recognise 2-DOG as an analogue of glucose but biochemically they cannot metabolise it.

4.3- Discussion

All trypanosome mRNAs are produced from polycistronic transcripts by the addition of a 39 nucleotides spliced leader (SL) sequence at their 5' end and a 3' poly(A) tail. The level of expression of *PRODH* genes and their regulation are detailed in this chapter.

The SL acceptor sites for *T. brucei* and *L. major* genes were identified as well as the 5' UTR. The un-translated region between the SL and the first methionine was 502 bp for *L. major* and just 15 bp for *T. brucei*. Recognition of the first ATG can be impaired when it is located close to the cap structure or next to a stop codon. Sometimes, the second ATG found in frame initiates the translation (Day and Tuite, 1998). It is the case for the *L. major PRODH*, according homology with other sequences, the second ATG was chosen as first methionine.

3' RACE has not been used in this study to identify polyA tail addition sites, however it is well known that 3' UTRs in trypanosomatids are variable in size. Frequently they are longer in *Leishmania* transcripts than those of other eukaryotes (Stiles *et al.*, 1999). They vary from several nucleotides to 2 Kb (Charest *et al.*, 1996).

The variation in length of the 3'UTR and 5'UTR regions could explain the size of the PRODH transcripts. In *T. brucei* the transcript was ~ 2 Kb and in *L. major* ~ 3 Kb. The 5' UTR was approximately 500 bp longer for the *L. major* than the *T. brucei* PRODH transcript. The 3' UTR is also probably longer in the *Leishmania* species transcript. Since

the open reading frames for *T. brucei* and *L. major* were 1668 bp and 1683 bp, respectively, it would appear that the *T. brucei* 3' UTR was around 300 bp while that of *L. major* was around 800 bp.

Gene transcription and processing was studied in the life cycle stages of these two organisms by performing Northern blot analysis to determine the quantity of RNA. Some examples of developmental regulation of *T. brucei* genes are known. This is the case for cytochrome c reductase, where the transcript is more abundant in BSF than in PCF (Priest and Hajduk, 1994b). PRODH transcripts were present in both *T. brucei* procyclic and bloodstream stages. The transcript seemed to be present in higher quantity in PCF, these forms being able to use proline as an energy source more effectively than the BFS do (Tielens and van Hellmond, 1998). It is believed that when both energy sources are present, the parasites use glucose first (van Weelden *et al.*, 2003). Given that, *T. brucei* BSF cannot use proline as an energy source, the stable transcription and processing of the gene in BSF might indicate an alternative role to energy production, such as glutamate production. In *T. cruzi* a proline racemase activity (Reina-San-Martin *et al.*, 2000) has been shown to play a critical role as a B-cell mitogen. A homologue of the *T. cruzi* proline racemase gene could not be found in the *T. brucei* genome.

The PRODH transcript is present in *L. major* promastigotes and metacyclic promastigotes. The probe used was *L. major PRODH* ORF and did not appear to detect the transcript in *L. mexicana*. Nothing could be concluded for the amastigote stages in the studies reported here.

The impact of PRODH on cell growth was also investigated. Through experiments done with the RNAi construct (Chapter 6), it was noticed that *T. brucei* PCF were able to grow in a medium lacking glucose and containing proline as an only energy source. It was therefore of interest to study *T. brucei* wild type cells when grown in different conditions.

When grown in complete medium containing glucose and proline, the parasites preferentially use glucose (Coustou *et al.*, 2003; van Weelden *et al.*, 2003). When glucose was absent from the medium, the level of transcription of PRODH RNA was similar indicating that any regulation of the PRODH levels did not occur at the level of transcription or RNA stability. In *Arabidopsis thaliana* (Nakashima *et al.*, 1998) the PRODH transcript is accumulated when the plants have been incubated under hypoosmotic conditions and when proline is present. The regulation in *Arabidopsis* occurs at the level of RNA. The *T. brucei* PRODH transcription has been studied in a medium whose glucose concentration was variable. It would have been of interest to study PRODH regulation with a medium whose the glucose concentration was stable and the proline concentration variable.

Testing the activity of the PRODH enzyme in crude lysates of *T. brucei* revealed it to be around twice as high in cells derived from a medium lacking glucose. The presence of glucose in the medium therefore leads to a decreased ability to use proline by the first enzyme of the proline catabolism pathway. In enzyme assays under the conditions tested, the highest activity observed was for a proline concentration of 10 mM, after which point substrate inhibition was apparent. Detailed kinetic analysis of the protein was not possible because the recombinant enzyme purified after expression in *E. coli* was not active (section 5.2.5).

Catabolite repression is a common phenomenon in prokaryotes (Bruckner and Titgemeyer, 2002) and yeast (Gancedo, 1998) where it is used to regulate gene expression. Catabolite repression often occurs at the transcriptional level, where metabolites bind a repressor and regulate transcription of genes. In *E. coli* when proline is present, it binds to a repressor and thus stimulates expression of the Put operon. However, when there is no proline, there is no expression of the operon. A similar phenomenon could occur in *T.*

brucei, however experiments aiming to quantitate RNA levels in cells grown, with or without proline in the medium, were not done to test this possibility. In trypanosomatids, transcriptional regulation is rare (Day and Tuite, 1998). In the case of the *T. brucei* proline dehydrogenase, Northern blot analysis did not show any difference at the level of RNA between the PCF grown in medium containing proline with, or without, glucose. Western blot experiments, performed with *T. brucei* PCF revealed the presence of the protein in both cases (parasites grown in the presence or absence of glucose). The level of transcription and translation was therefore not changed by the presence of glucose in the medium; the protein was still present regardless of conditions. Therefore, the presence of glucose in the medium must influence the activity of PRODH post-translationally.

An analogue of glucose, 2-deoxyglucose (2-DOG), seemed to inhibit proline utilisation in a fashion similar to D-glucose. Evans and Brown, 1972, concluded that 2-DOG had no effect on the parasite's growth, because the parasites were able to use proline instead. It is possible that 2-DOG, at the levels they used, was not sufficient to kill the cells (however the authors did not reveal the concentration at which they used 2-DOG in their publications (Evans and Brown, 1972)). In my RNA interference experiments, the non-induced parasites were able to grow in the presence of proline only (section 6.2.5). When 2-DOG was added to the medium, without glucose but containing proline, no growth was observed. This shows that *T. brucei* cannot use 2-DOG as an energy source and moreover that its presence inhibited the proline utilisation. This indicates that repression of proline use is a response to 2-DOG and not to metabolic consequences of glucose metabolism (such as, for example, ATP levels since 2-DOG cannot be used in energy production). It is not certain how the presence of glucose (or 2-DOG) down regulates PRODH activity. It would be of interest to determine whether this substrate had some kind of allosteric

regulatory effect on the enzyme. Alternatively, high glucose abundance could stimulate alternative, indirect mechanisms that lead to a decrease in the level of PRODH activity.

The uptake of proline transport was also studied in *T. brucei* grown in medium plus or minus glucose. Without glucose the uptake of proline was around twice as high as when the parasites were exposed to high glucose concentrations. A simple explanation is not obvious; it is possible that at 30 s the measured uptake is a combination of transport plus metabolism, hence when only proline is present, more of it is metabolised. That proline uptake is not linear at 30 s is evident in the fact that the apparent V_{max} at 3 s was substantially higher. However, even at the lower time point a substantial difference between induced and non-induced cells was apparent. It is also possible that PRODH and the transporter have a similar regulatory response to glucose with both being down-regulated. Further insight into this phenomenon is discussed in chapter 6.

It seems possible that trypanosomes have a system to down regulate proline utilisation in the presence of glucose in the tsetse fly, as flies may transiently become glucose-rich following blood meals, in which case conserving fly proline stocks could be in the parasite's interest.

Chapter 5

Cloning and heterologous expression of *T. brucei* and *L. major* proline dehydrogenase

5.1- Introduction

The proline dehydrogenase (*PRODH*) genes of *L. major* and *T. brucei* were identified as detailed in Chapter 3. A next logical step for the study of this protein was the cloning of the gene and expression of the recombinant protein in a bacterial expression system. The difficulty of obtaining sufficient quantities of pure enzyme from parasites has limited research in the past, but recent advances in molecular biology technologies mean that is now almost routine to express cloned parasite genes in heterologous systems such as *Escherichia coli*.

PRODH is a membrane-bound protein in the mitochondria of eukaryotes (Maxwell and Davis, 2000) and is plasma membrane-bound in prokaryotes (Menzel and Roth, 1981a) (equivalent of the mitochondrial membrane of eukaryotes). Several attempts were made to over-express the trypanosomal protein in a soluble form in *E. coli*. However, this was not successful and so an attempt to mutate the gene of interest was made in order to improve the solubility of the encoded protein. As described in the Chapter 3, a putative mitochondrial targeting sequence and a putative trans-membrane domain were identified in the trypanosomatid predicted proteins. These segments did not contain conserved motifs believed to be essential for the activity of the enzymes, or the motif for the binding of FAD. It was decided, therefore, to over-express the proteins without these two segments.

The *E. coli* PutA669, which is the PRODH segment of the *E. coli* polypeptide (Vinod *et al.*, 2002), was also cloned and expressed as a positive control for the procedure.

5.2- Results

5.2.1- Cloning of *T. brucei* and *L. major* proline dehydrogenase genes

PRODH genes from *L. major* and *T. brucei* were PCR amplified and cloned into pET21a^{+} vector in order to express a recombinant protein containing a hexa-His-tag at the C-terminus of each protein. For each organism, different constructs were made in an attempt to optimise the level of expression of the enzyme.

Oligonucleotides were designed to amplify the entire proline dehydrogenase gene (*LmPRODH* and *TbPRODH*) from *L. major* and *T. brucei* and a version lacking the putative mitochondrial targeting sequence (*LmPRODHtruncated* and *TbPRODHtruncated*). Figure 5.1 shows the result of the PCR amplification from genomic DNA of these four constructs.

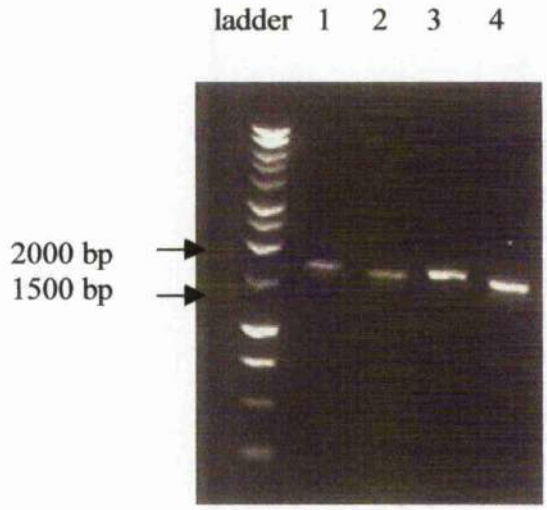


Figure 5.1: Ethidium bromide gel showing PCR amplification of proline dehydrogenase (entire and truncated genes). 1: *LmPRODH* (1683 bp); 2: *LmPRODHtruncated* (1608 bp); 3: *TbPRODH* (1668 bp); 4: *TbPRODHtruncated* (1599 bp).

Each primer contained restriction sites for cloning into the expression vector (see section 2.2.3). For both organisms, a version lacking the putative trans-membrane domain (named *LmPRODH72* and *TbPRODH72*) was also amplified (Figure 5.2).

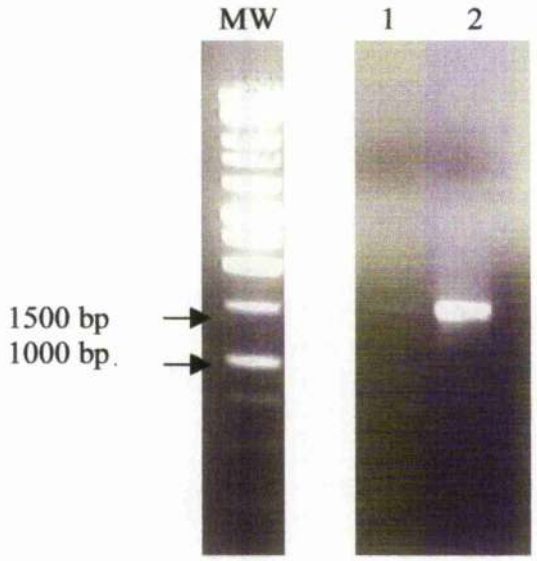


Figure 5.2: Ethidium bromide gel showing PCR amplification of proline dehydrogenase without nucleotides encoding the first 72 amino acids. *L. major* proline dehydrogenase lacking the first 72 amino acids (*LmPRODH72*) (1467bp) (lane 1) and *T. brucei* proline dehydrogenase lacking the first 72 amino acids (*TbPRODH72*) (1473 bp) (lane 2).

After gel purification, the PCR products were cloned into the pGEM-T vector and sequenced. Thus PCR products and the expression vector pET21a^{+} were both digested by restriction enzymes and ligated together. A schematic representation of the plasmids containing the various PRODH constructs can be seen in Figures 5.3-5.8.

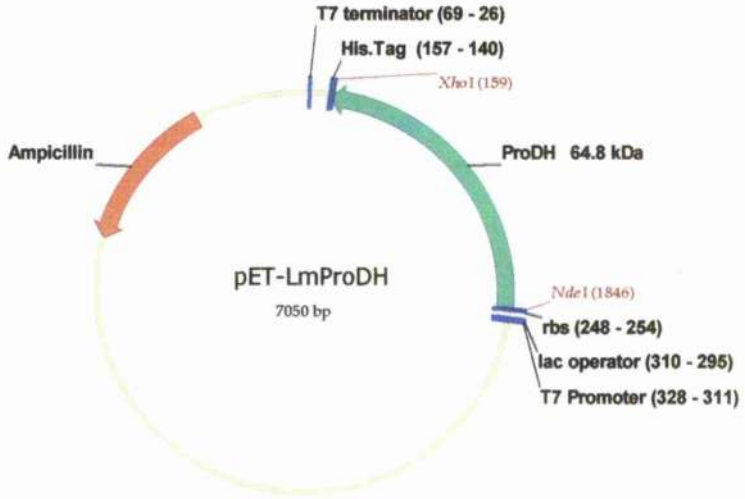


Figure 5.3: Plasmid construct for expression in *E. coli* of the *L. major* proline dehydrogenase entire gene. The gene was cloned into the pET21a^{+} vector, which generates a His-Tag at the C-terminus.

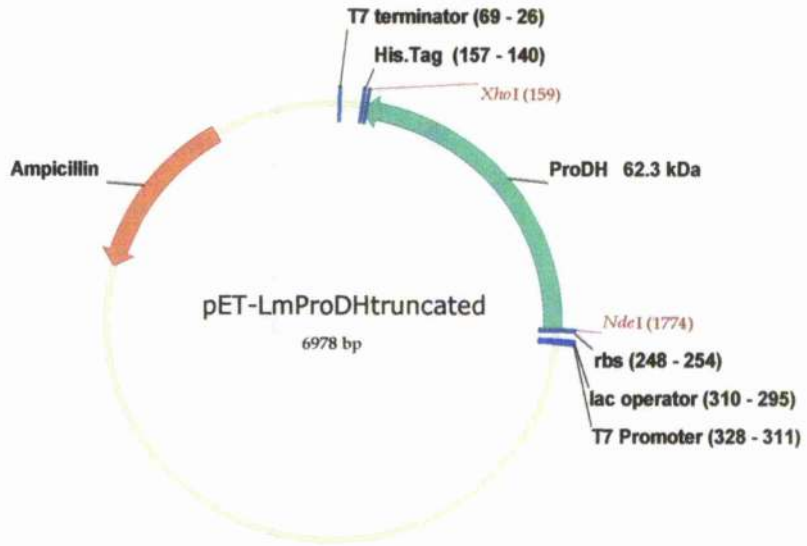


Figure 5.4: Plasmid construct for expression in *E. coli* of the *L. major* proline dehydrogenase truncated gene. The gene was cloned into the pET21a^{+} vector, which generates a His-Tag at the C-terminus.

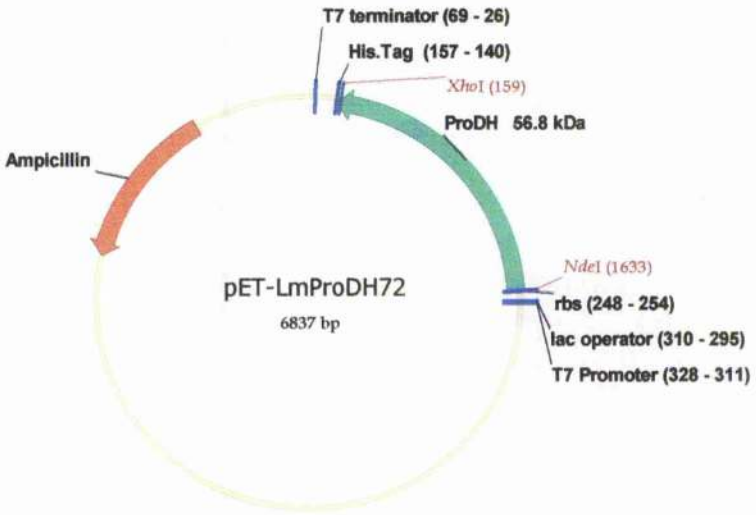


Figure 5.5: Plasmid construct for expression in *E. coli* of the *L. major* proline dehydrogenase gene without the 72 first amino acids. The gene was cloned into the pET21a^{+} vector, which generates a His-Tag at the C-terminus.

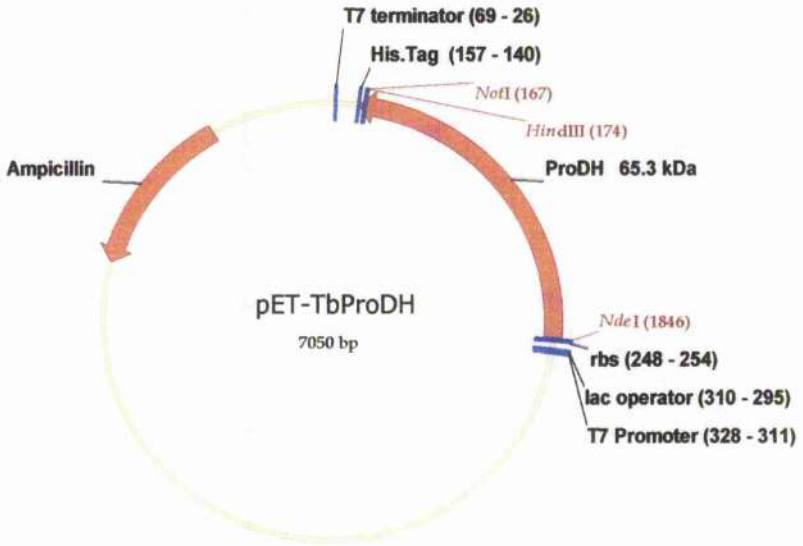


Figure 5.6: Plasmid construct for expression in *E. coli* of the *T. brucei* proline dehydrogenase gene. The gene was cloned into the pET21a^{+} vector, which generates a His-Tag at the C-terminus.

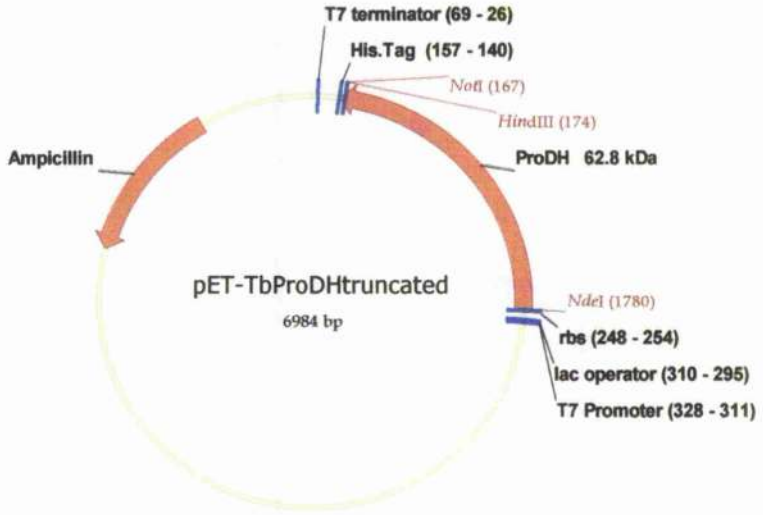


Figure 5.7: Plasmid construct for expression in *E. coli* of the *T. brucei* proline dehydrogenase truncated gene. The gene was cloned into the pET21a^{+} vector, which generates a His-Tag at the C-terminus.

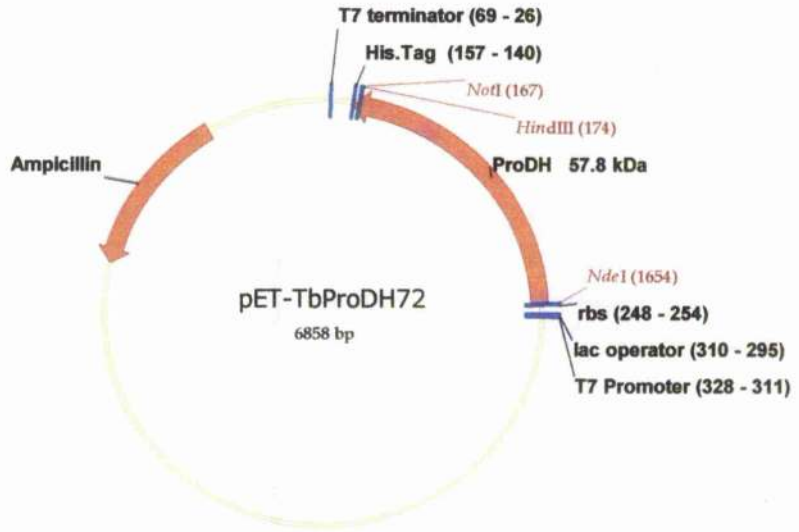


Figure 5.8: Plasmid construct for expression in *E. coli* of the *T. brucei* proline dehydrogenase gene without the 72 first amino acids. The gene was cloned into the pET21a^{+} vector which, generates a His-Tag at the C-terminus.

5.2.2- Expression of *L. major* and *T. brucei* proline dehydrogenases

E. coli BL21(DE3) containing the expression plasmids for *L. major* and *T. brucei* PRODHs was grown in 50-500 ml LB medium containing ampicillin (100 μg/ml). The six different contructs were induced under various conditions in attempts to improve the amount of expressed recombinant protein. 0.1 to 1 mM IPTG was added to the culture and the contructs were induced at 16 °C, 20 °C and 37 °C for 4 hours to 16 hours.

Expression at 37 °C with 1 mM IPTG resulted in the expression of an abundant quantity of protein at about 60 kDa corresponding with the expected size of the recombinant full length *L. major* and *T. brucei* PRODH (predicted sizes of 64.8 kDa and 65.3 kDa, respectively) and the truncated (predicted sizes of 62.3 kDa and 62.8 kDa, respectively) proteins. After harvesting the *E. coli* cells and sonicating them (20 times 20s pulse with 30s between pulses) in 10 ml of 50 mM Tris, pH 7.5, ± 0.5 M NaCl, the soluble and insoluble fractions were separated by centrifugation at 10,000 g for 30 min at 4 °C. The recombinant PRODH proteins were found in the insoluble fraction of the cells, as shown using SDS-PAGE (Figures 5.9 and 5.10). Furthermore, purification of the soluble phase by a Ni-Agarose chromatography did not yield any protein of the approximate size in the elution fractions (results not shown).

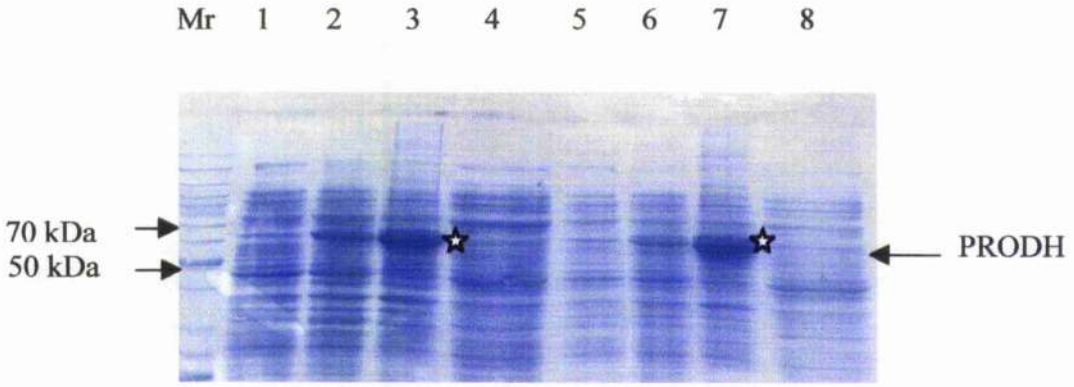


Figure 5.9: Analysis of expression of *L. major* proline dehydrogenase in BL21(DE3) using SDS-PAGE. Lanes 1-4, full length gene *L. major*, lanes 5-8, truncated gene *L. major*. **Lane Mr**, protein standards; **lane 1,** *E. coli* cell lysate prior to IPTG induction; **lane 2**, *E. coli* lysate after 4 h induction by 1 mM IPTG at 37 °C; **lanes 3 and 4,** pellet and supernatant (10 µl of total fraction, 10 ml) of the *E. coli* lysate centrifugated at 10,000 g for 30 min; **lane 5**, *E. coli* cell lysate prior to IPTG induction ; **lane 6**, *E. coli* lysate after 4 h induction by 1 mM IPTG at 37 °C; **lanes 7 and 8** pellet and supernatant (10 µl of total fraction, 10 ml) of the *E. coli* lysate centrifugated at 10,000 g for 30 min. Over-expressed protein is marked with a white star.

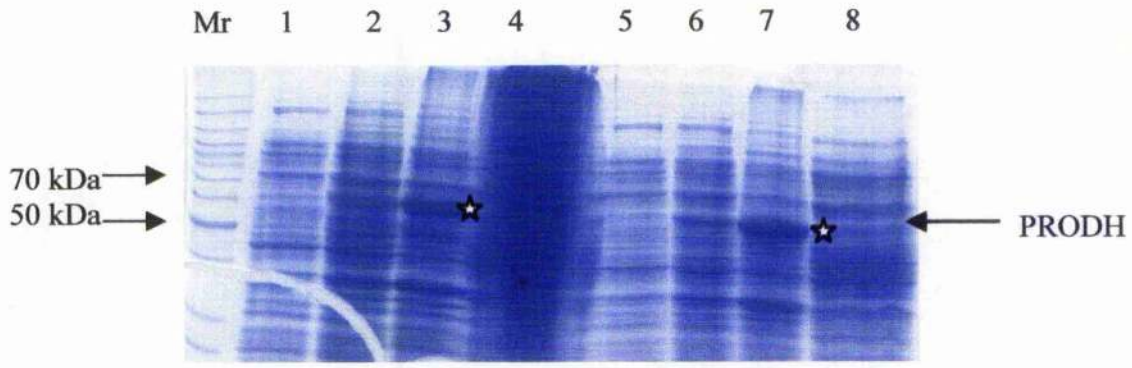


Figure 5.10: Analysis of expression of *T. brucei* proline dehydrogenase in BL21(DE3) using SDS-PAGE. Lanes 1-4 full length *T. brucei* gene; lanes 5-8 truncated *T. brucei* gene. **Lane Mr**, protein standards; **lane 1,** *E. coli* cell lysate prior to IPTG induction; **lane 2**, *E. coli* lysate after 4 h induction by 1 mM IPTG at 37 °C; **lanes 3 and 4** pellet and supernatant (10 µl of tatal fraction, 10 ml) of the *E. coli* lyasate centrifugated at 10,000 g for 30 min; **lane 5**, *E. coli* cell lysate prior to IPTG induction gene; **lane 6**, *E. coli* lysate after 4 h induction by 1 mM IPTG at 37 C; **lanes 7 and 8** pellet and supernatant (10 µl of total fraction, 10 ml) of the *E. coli* lysate centrifugated at 10,000 for g 30 min. Over-expressed protein is marked with a white star.

Different conditions were used during the sonication step in attempts to improve the solubility of the proteins. Pellets of *E. coli*, containing the recombinant protein from the constructs containing the full length and truncated *L. major* and *T. brucei* genes, were resuspended in various sonication buffers. The iso-osmotic buffers used were 10 mM Tris-HCl, pH 7.8, 0.5M NaCl, ± 10 % glycerol; and 50 mM NaH_2PO_4, pH 7.8, 0.5 M NaCl. The detergents Triton-X100 (0.1 to 1 % v/v) or Tween 20 (0.1 to 1 % v/v) were also added. Various sonication procedures (6 to 20 times, 10 s to 30 s pulses, with 30 s interval between pulses) were also tested. Freeze/thaw methods of the cells disruption were also tried in place of sonication in efforts to increase the solubility of the recombinant proteins.

However, under all of the conditions tested, the recombinant *L. major* and *T. brucei* PRODH full length and truncated proteins remained exclusively in the inclusion bodies.

A method of solubilisation of *L. major* full length PRODH from inclusion bodies was attempted, as detailed in section 2.3.3.3. After induction by 1 mM IPTG at 37 °C for 4 hours, 100 ml of cell culture was pelleted and the pellet treated with 8 M urea (Sanderson *et al.*, 2000) or 2 M thiourea, 8 M urea and 2 % (v/v) CHAPS (van Deursen *et al.*, 2003). The progress of the solubilisation is shown in Figure 5.11. The recombinant protein was solubilised, but enzyme activity could not be identified in the purified protein (see section 5.2.5).

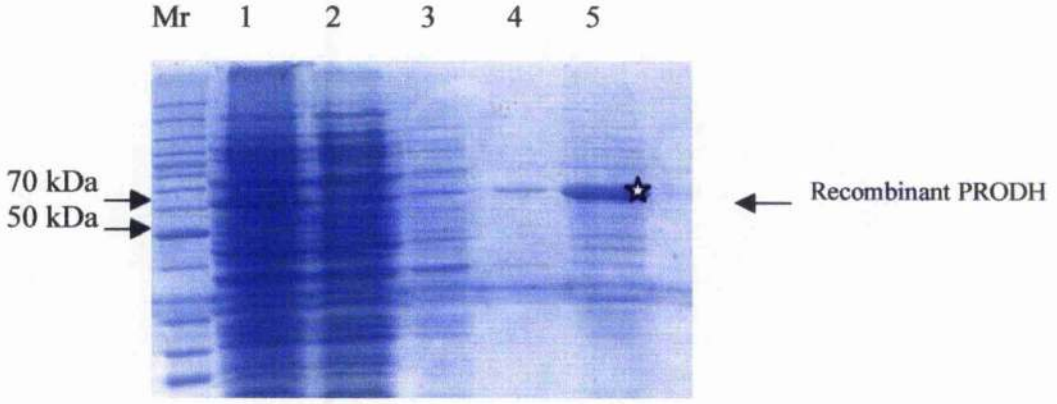


Figure 5.11: Analysis of solubilisation by urea of *L. major* entire proline dehydrogenase using SDS-PAGE. Expression was induced at 37 °C by 1 mM IPTG for 4 h using BL21(DE3) cells. The pellet of *E. coli* was resuspended in 2.5 ml of 50 mM Tris HCl, pH 7.5, 5 mM EDTA, 5 % (w/v) sucrose and stored 1 h at –20 °C, then defrosted in tap water for 10 min and re-frozen for 1 h at – 20 °C. The suspension pellet was then sonicated 6 x 20 s on, 30 s off (**lane 1**, 5 µl of total fraction) and then centrifuged at 10,000 g for 10 min at 4 °C. The resultant supernatant is shown in **lane 2** (10 µl of total fraction, 2.5 ml). The pellet was resuspended in 2.5 ml of the same buffer containing 0.1 % Triton X 100 and then centrifuged at 10,000 g for 10 min at 4 °C. The resultant supernatant is shown in **lane 3** (10 µl of total fraction, 2.5 ml). The pellet was resuspended in 2.5 ml of the buffer containing 2 M urea and centrifuged at 10,000 g for 10 min at 4 °C, and the resultant supernatant is shown in **lane 4** (10 µl of total fraction, 2.5 ml). The final pellet was resuspended in 5 ml 8 M urea and slowly shaken at 37 °C for 1 h and then diluted with 5 ml of 50 mM Tris-HCl, pH 7.5 containing 8 M urea and 5 mM EDTA (**lane 5**, 10 µl of total fraction, 10 ml). Recombinant protein is marked with a white star.

Another method used in attempts to increase the solubility of the recombinant protein involved reducing the temperature of induction. The *L. major* and *T. brucei* PRODH were induced with 1 mM IPTG at 16 °C for 16 hours. The analysis of over-expression is shown in Figure 5.12 and the presence of the recombinant PRODH was observed.

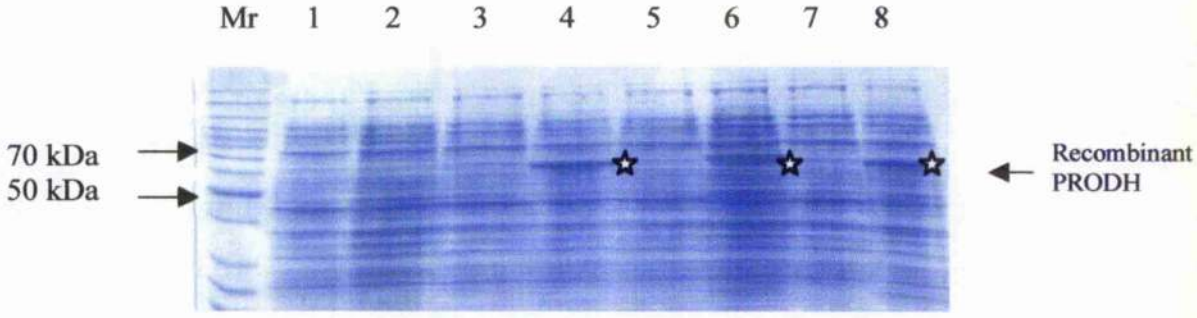


Figure 5.12: Analysis of over-expression of *L. major* and *T. brucei* proline dehydrogenase in BL21(DE3) at 16 °C using SDS-PAGE. Lanes 1-2 full length *L. major* gene; lanes 3-4, truncated *L. major* gene; lanes 5-6, full length *T. brucei* gene; lanes 7-8 truncated *T. brucei* gene. **Lane Mr**, protein standards; **lane 1,** *E. coli* cells lysate prior to IPTG induction; **lane 2**, *E. coli* lysate after overnight induction by 1 mM IPTG at 16 °C; **lane 3**, *E. coli* cells lysate prior to IPTG induction; **lane 4**, *E. coli* lysate after overnight induction by 1 mM IPTG at 16 °C; **lane 5**, *E. coli* cells lysate prior to IPTG induction ; **lane 6**, *E. coli* lysate after overnight induction by 1 mM IPTG at 16 °C; **lane 7**, *E. coli* cells lysate prior to IPTG induction; **lane 8**, *E. coli* lysate after overnight induction by 1 mM IPTG at 16 °C. Recombinant protein is marked with a white star.

At 16 °C, the expression of PRODH was considerably reduced compared with at 37 °C. It seemed that the concentration of IPTG did not have an effect so it was fixed at 1 mM. 300 ml of BL21(DE3) containing pET-TbPRODH was induced at 16 °C by 1 mM IPTG. The *E. coli* cells were harvested by centrifugation and the pellet was first sonicated in 5 ml

phosphate buffer (50 mM NaH_2PO_4, 500 mM NaCl, pH 7.8) and then centrifuged at 10,000 g for 30 min at 4 °C. The second pellet was sonicated under the same conditions in 10 ml phosphate buffer. The soluble and insoluble phases were separated by centrifugation at 10,000 g for 30 min at 4 °C. The results are shown in Figure 5.13.

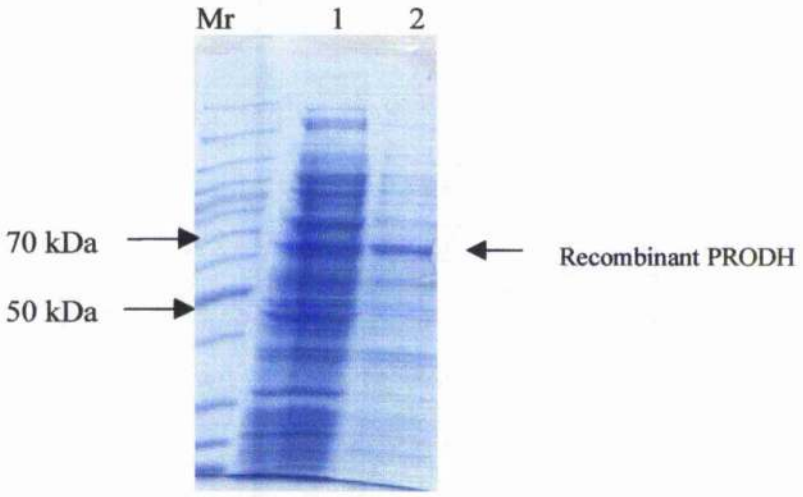


Figure 5.13: Analysis of solubilisation of expressed *T. brucei* entire proline dehydrogenase at 16 °C in BL21(DE3) using SDS-PAGE.

The *T. brucei* entire gene was induced with 1 mM IPTG overnight at 16 °C. The cells were centrifuged and the pellet resuspended in 5 ml sonication buffer (50 mM NaH_2PO_4, 500 mM NaCl, pH 7.8) before sonication (20 times 20 s on, 30 s off) and centrifugation at 10,000g for 30 min at 4 °C. The resultant supernatant is shown in **lane 1** (10 µl of the total fraction, 5 ml). The pellet was resuspended in 10 ml of sonication buffer and re-sonicated (using the same conditions) and re-centrifuged. The resultant supernatant is shown in **lane 2** (10 µl of the total fraction, 10 ml).

Approximately 20 % of the over-expressed protein was present in the soluble phase, and was purified on a Ni-Agarose column (detailed in section 5.2.3).

E. coli containing the construct encoding the *T. brucei* proline dehydrogenase without the trans-membrane domain (TbPRODH72) was induced with 1 mM IPTG at 16 °C for 16 hours. The cells were harvested and the pellet resuspended in 10 mM Tris-HCl pH 7.8, imidazole 5 mM, NaCl 0.5 M, glycerol 10 % (v/v) and sonicated. The soluble and

insoluble phases were separated by centrifugation and analysed by SDS-PAGE (Figure 5.14).

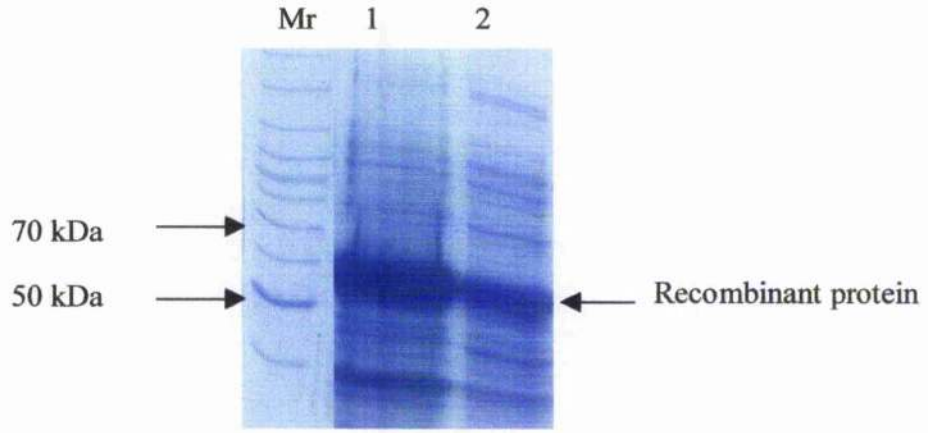


Figure 5.14: Analysis of expression in BL21(DE3) and solubilisation of *T. brucei* proline dehydrogenase lacking the 72 first amino acids using SDS-PAGE. *E. coli* cells containing the construct was induced for 16 hours at 16 °C with 1 mM IPTG. The pellet of *E. coli* cells was resuspended in 10 ml of 10 mM Tris-HCl pH 7.8, imidazole 5 mM, NaCl 0.5 M, glycerol 10 % (v/v) and sonicated (20 times 20 s on, 30 s off) before centrifugation at 10,000g for 30 min at 4 °C. The resultant pellet is shown in **lane 1** (10 µl of the total fraction 10 ml and the supernatant in **lane 2** (10 µl of the total fraction 10 ml).

Around 30 % of the recombinant protein was present in the soluble fraction and was purified using a Ni-Agarose chromatography (section 2.3.4).

5.2.3- Purification of the recombinant proline dehydrogenases

Two systems were used to purify the proteins expressed using the six different constructs of *T. brucei* and *L. major* PRODH. These systems are based on the presence of a hexa-His-tag at the C-terminus of the recombinant protein. The presence of this tag was firstly checked on Western blot analysis using anti-His-tag antibody (Figure 5.15). A single band of 60 kDa corresponding to the expected size of the recombinant protein was detected, confirming the presence of the His-tag to allow the purification.

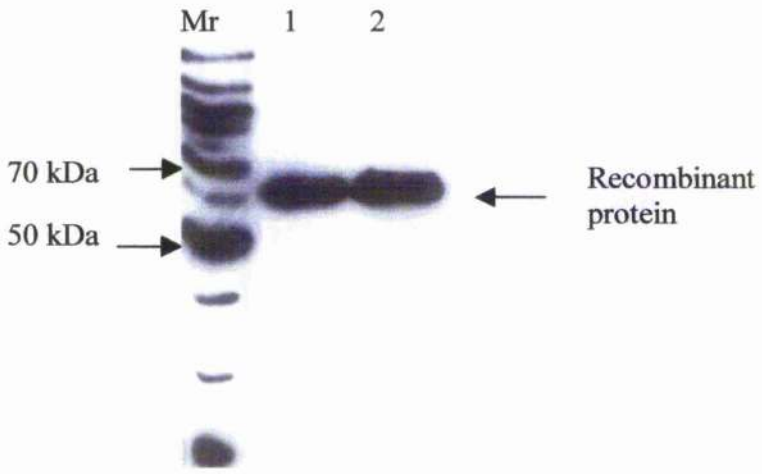


Figure 5.15: Detection of recombinant proline dehydrogenase in crude lysates of *E. coli* using anti-His tag antibody. *L. major* entire protein (lane 1) and *T. brucei* entire protein (lane 2) detected using anti-His-tag antibody (dilution 1: 2000)

Two methods were chosen to purify the recombinant protein, using a Ni-Agarose column (Qiagen) and the Ni affinity chromatography, BioCAD system (see section 2.3.4.2). 300 ml of *E. coli* BL21(DE3) containing pET21a^{+} vector containing the *T. brucei*

entire *PRODH* gene were induced by 1 mM IPTG at 16 °C for 16 hours. After sonication and centrifugation, 10 ml of the supernatant (Figure 5.13, lane 2) was loaded onto the BioCAD nickel column. After washing the column with 50 mM NaH_2PO_4, 500 mM NaCl pH 7.8, the elution of the protein was achieved by 500 mM imidazole in phosphate buffer. The results are shown in Figure 5.16.

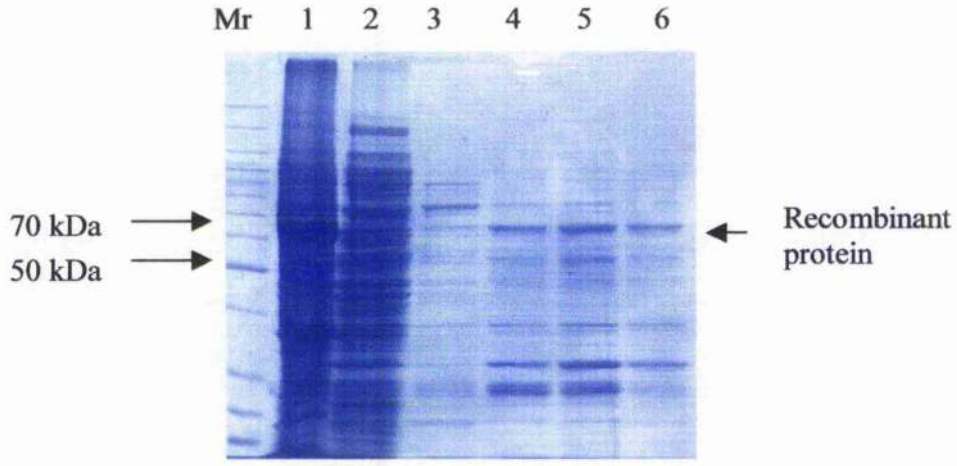


Figure 5.16: The purification profile of *T. brucei* entire gene expressed at 16 °C in BL21(DE3) analysed using SDS-PAGE. The soluble protein in the supernatant fraction was purified on a nickel chelator column (BioCAD). **Mr**, protein standards; **lane 1**, pellet of the preparation; **lane 2**, flow through (10 µl of total fraction, 20 ml); **lane 3**, wash with binding buffer containing 50 mM imidazole (10 µl of total fraction, 80 ml). Elution of the recombinant protein was performed with 1 ml fractions of phosphate buffer containing 500 mM imidazole (**lanes 4, 5 and 6**).

Only a small amount of the recombinant protein was obtained and this was quite heavily contaminated by other proteins. The eluted fractions were pooled (~ 0.2 mg/ml) and dialysed overnight into different buffers (see section 5.2.5) before testing for enzyme activity.

The *T. brucei* PRODH without the putative trans-membrane domain was purified using a Ni-Agarose column (Qiagen). 100 ml of culture was induced by 1 mM IPTG at 16 °C for

16 hours and sonicated in 10 ml of 10 mM Tris-HCl pH 7.8, imidazole 5 mM, NaCl 0.5 M, glycerol 10 % (v/v). After centrifugation, the soluble fraction (Figure 5.14, lane 2) was loaded onto the Ni-Agarose column, washed and eluted by 500 mM imidazole (3 fractions of 1.5 ml) (Figure 5.17).

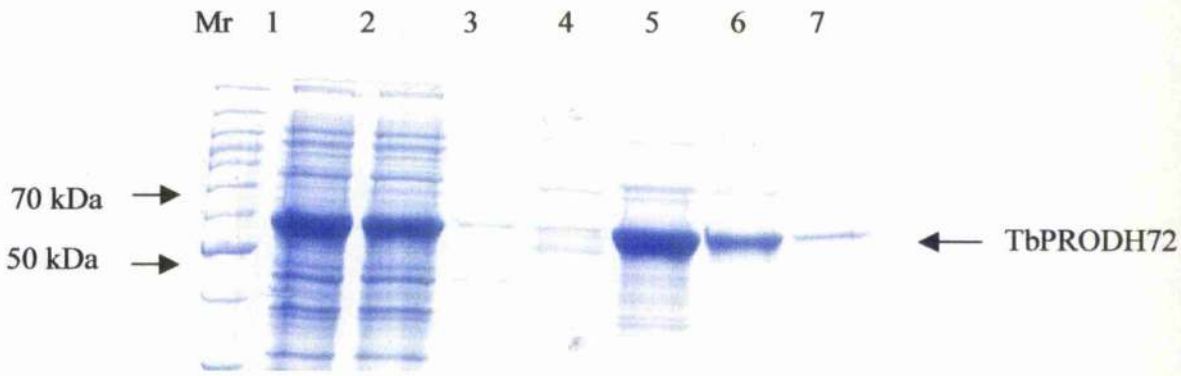


Figure 5.17: The purification profile of the soluble fraction of TbPRODH72 using Ni-Agarose chromatography and analysed using SDS-PAGE. Mr, protein standards; **lane 1**, supernatant of *E. coli* lysate (10 µl of total fraction, 10 ml); **lane 2**, flow through (10 µl of total fraction, 10 ml) ; **lane 3**, wash (10 µl of total fraction, 10 ml); **lane 4**, wash with 10 mM Tris buffer, pH 7.8, containing 30 mM imidazole (10 µl of total fraction, 5 ml); **lanes 5-7**, elution fractions 1-3 with Tris buffer containing 500 mM imidazole (10 µl of total fraction, 1.5 ml).

All the eluted fractions contained the recombinant protein. They were pooled (~ 0.5 mg/ml) and dialysed overnight at 4 °C against various buffer, before testing for enzyme activity (Section 5.2.5).

5.2.4- Cloning, expression and purification of *E. coli* PutA669

As a control, it was decided to produce recombinant *E. coli* PutA669 (Nadaraia *et al.*, 2001) containing the PRODH activity of the PutA protein (PRODH and P5CDH). *PutA669* was PCR amplified and cloned into pET21a$^+$ vector in order to express a recombinant protein containing a His-tag at the C-terminus. Oligonucleotides containing specific restriction sites were designed to amplify from gDNA of *E. coli* the *PutA669* gene. Figure 5.18 shows the result of the digestion by *Nde* I and *Xho* I of *PutA669* cloned into pET21a$^+$ (section 2.2.3).

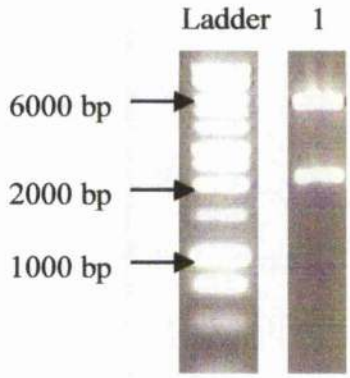


Figure 5.18: Ethidium bromide gel of digested PCR products cloned into pET21a$^+$ containing *E. coli PutA669*. The plasmid was digested by *Nde* I and *Xho* I (lane 1), the band at 2000 bp corresponds to the *E. coli* PutA669.

The resulting construct was introduced into the *E. coli* expression strain BL21(DE3). The PutA669 with the C-terminal His-tag was induced using a culture of 100 ml LB medium containing ampicillin by 1 mM IPTG at 16 °C for 16 hours. The pellet of *E. coli* cells was resuspended in 10 ml of sonication buffer (10 mM Tris-HCl pH 7.8, imidazole 5 mM, NaCl 0.5 M, glycerol 10 % (v/v)) and sonicated (20 times, 20 s on, 30 s off). Soluble and

insoluble fractions were separated by centrifugation and analysed by SDS-PAGE (Figure 5.19).

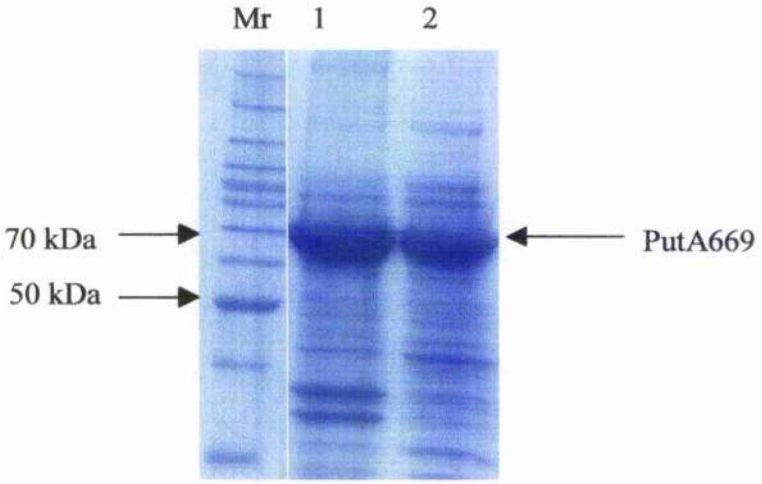


Figure 5.19: Expression of *E. coli* PutA669 in BL21(DE3) and solubilisation profile analysed using SDS-PAGE. The cells were induced at 16 °C overnight with 1 mM IPTG and then harvested by centrifugation. The pellet was resuspended in 10 ml sonication buffer (10 mM Tris-HCl pH 7.8, imidazole 5 mM, NaCl 0.5 M, glycerol 10 % (v/v)), sonicated (20 times, 20 s on, 30 s off) and centrifuged at 10,000 g for 30 min at 4 °C. The resultant pellet is shown in **lane 1** and supernatant in **lane 2** (10 µl of total extract, 10 ml).

Approximately 50 % of the recombinant protein was in the soluble fraction. The supernatant fraction (Figure 5.19, lane 2) was loaded onto a nickel affinity column (BioCAD) and the elution of the protein was achieved using 500 mM imidazole (Figure 5.20). The purification profile for the BioCAD column is shown in Figure 5.21. The *E. coli* PutA669 was mainly in fraction 5 (Figure 5.20) and was dialysed overnight into 70 mM Tris-HCl, pH 7.8, 10 % (v/v) glycerol, 2 mM EDTA.

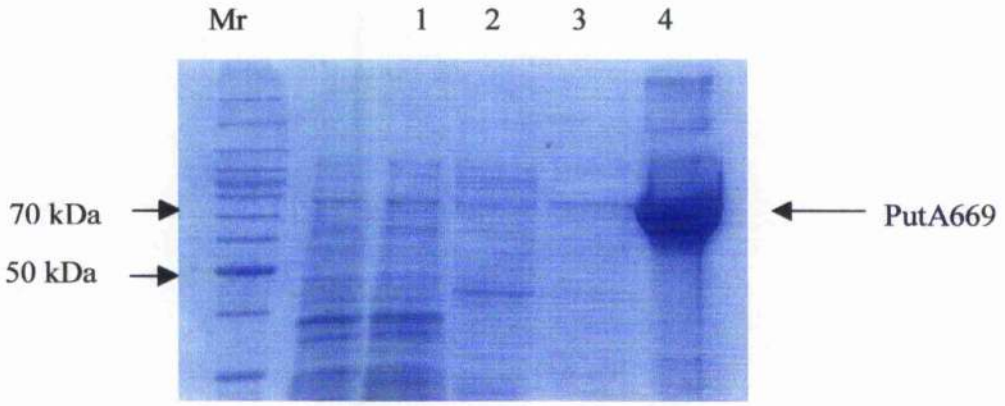


Figure 5.20: Purification of *E. coli* Put669 with BioCAD 700E column analysed using SDS-PAGE. Mr, protein standards; **lane 1**, flow through (10 µl of total fraction, 10 ml); **lane 2**, wash with 70 mM Tris buffer, pH 7.8 (10 µl of total fraction, 30 ml); **lane 3**, wash with 70 mM Tris buffer, pH 7.8 containing 50 mM imidazole (10 µl of total fraction, 10 ml); **lane 4**, fraction number 5 after elution with 500 mM imidazole, 1 ml fraction (5 µl of total fraction, 1 ml). Concentration of fraction 5: 0.65 mg protein/ml.

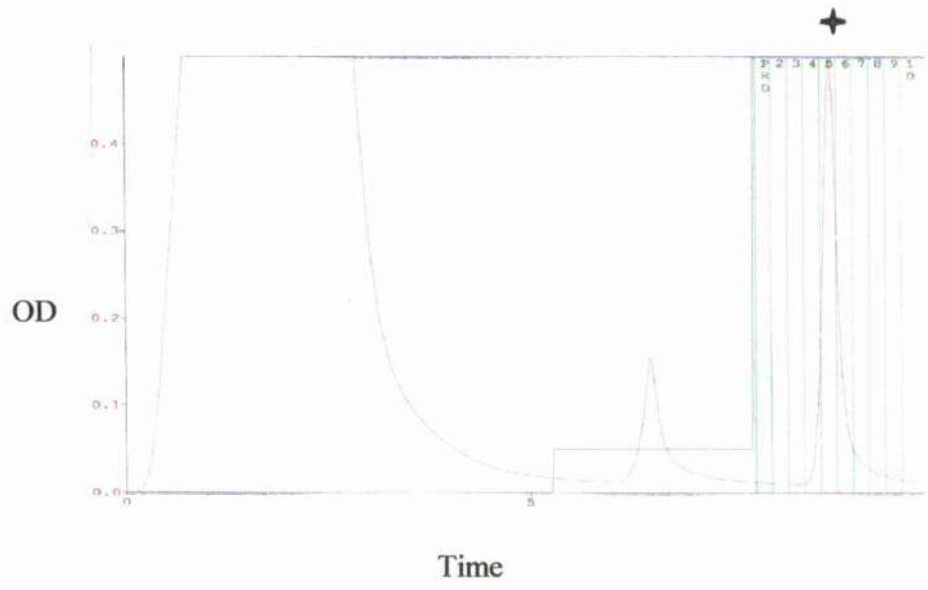


Figure 5.21: Purification profile of the *E. coli* PutA669 protein through the BioCAD 700E column. The red line represents the concentration of the proteins by monitoring absorbance at 280 nm. The blue line is the concentration of imidazole. The recombinant *E. coli* PutA669 is present mainly in fraction number 5.

5.2.5- Analysis of the activity of proline dehydrogenase recombinant proteins

Efforts to detect activity with the purified trypanosomatid PRODHs was based on the methods developed by Nadaraia *et al.*, 2001 and Vinod *et al.*, 2002, for the activation of the recombinant PutA669. According to these authors, after overexpression of the recombinant protein, sonication of the cell pellet and purification through a Ni-Agarose column, the protein lost the FAD cofactor. For this reason, the purified *E. coli* PutA669 (positive control) was incubated for 2 hours with 1 mM FAD and dialysed overnight against 70 mM Tris-HCl pH 7.8, glycerol 10 % (v/v), 2 mM EDTA. This protein was then checked for activity using DCPIP as a terminal electron acceptor as described in section 2.3.9.

The recombinant *E. coli* PutA669 had a specific activity of 0.6-2.4 $\mu mol.min^{-1}.mg\ protein^{-1}$. Vinod *et al.*, 2002, reported PRODH specific activity between 10-12 units.mg $protein^{-1}$. Therefore the specific activity of the enzyme purified here was lower than that previously published, but this control was nevertheless sufficient to allow determination of activity in different conditions. The reduced activity could relate to the fact that the construct was different in the two cases.

The same method was used for the attempted activation of the soluble PRODH from each trypanosomatid and for refolding from the inclusion bodies. No activity was detected.

Variations of this protocol were attempted by using different buffers for dialysis and different buffers during the activity testing. A summary of the conditions tested on various protein preparations is shown in Table 5.1.

Buffer	Concentration	pH variation	Cofactor	Components
Tris	10-200 mM	5-9	0.5 to 1 mM	± EDTA (2-10 mM)
HEPES		4-10	± FAD	± Glycerol (10%)
PIPES			± NAD	± $MgCl_2$
MOPS			± FMN	
TES			± Parasite extracts	
Phosphate		5-9		

Table 5.1: Summary of the activation conditions tested

No activity was detected for any of the recombinant trypanosomatid PRODHs in any of the conditions tested.

5.3- Discussion

This chapter describes efforts to express trypanosomatid PRODHs in *E. coli* for protein purification and analysis. Initial experiments failed to yield soluble protein for full length genes.

Eukaryotic PRODHs are bound to the mitochondrial membrane (section 3.2.5). The trypanosomatid PRODHs described here have a putative mitochondrial targeting sequence and also a putative trans-membrane domain located at the N-terminus of the protein. This might indicate that in these cells too, the protein is localised to the mitochondrial inner membrane. In attempts to increase the expression level and the solubility of the trypanosomatid PRODHs, different constructs were used based on the in silico study of the protein. Through homology analysis with other known PRODHs (section 3.2.3), two regions of high homology could be found in the amino acid sequence of all the PRODH

analysed. These regions correspond to the active site, which binds the substrate proline, and to the cofactor FAD binding site (Nadaraia *et al.*, 2001). These two essential regions were located well away from the N-terminus of the trypanosomatid enzymes. Therefore, it was hoped that by removing the N-terminal mitochondrial targeting sequence and/or the trans-membrane domain, the activity of the recombinant enzyme would not be affected, but that solubility might be improved. Unfortunately, removal of the mitochondrial targeting sequence had no effect on the solubility of the recombinant protein. It was still present in inclusion bodies. The *T. brucei* PRODH, without the trans-membrane domain, was found in the soluble fraction of *E. coli* expressor cell lysates. But this enzyme never gave any activity in the conditions tested. By removing the N-terminal part, the normal folding of the protein might have been affected. This segment might also contain other important residues for the activity of the *T. brucei* enzyme. However the *T. brucei* PRODH expressed in its entirety also failed to show activity after it had been refolded (section 5.2.5).

In Chapter 4, section 4.2.5, the phenomenon of substrate inhibition for the PRODH, when proline was present at a concentration superior at 10 mM was reported. The recombinant protein was tested with lower substrate concentration, but no activity was found. It cannot be ruled out that the presence of the hexa-his tag at the protein C-terminus could affect activity. For example if this region of the protein was involved in dimer formation this could affect activity. In *E. coli*, the domain I (residues 87-139) is crucial to the dimer interface. Based on the alignment between the trypanosomal and *E. coli* enzymes, this region is preserved in the mutated enzymes; hence it is unlikely that the mutated enzymes will be hindered in their ability to form dimers. Efforts to express the protein without a his-tag or with an N-terminal tag would be an interesting topic for further study, but such efforts were not made here. However, the *E. coli* recombinant PRODH was over-expressed with a C-terminal His tag, without abolishing the activity of the enzyme.

The activity for *E. coli* PRODH is lower than the one reported (Vinod *et al.*, 2002). This might be due to a less pure enzyme, or the fact than only 50 % of the recombinant protein was soluble, which might indicate that some part of the soluble protein was not refolded correctly. Since enzymatic activity is associated with membrane bound protein *in situ*, subtle changes in structure related to degree of solubility could affect activity. In Chapter 4 was shown that the presence of glucose in growth medium might impact the activity of the enzyme. If glucose in *E. coli* could affect the activity of *T. brucei* PRODH expressed in this environment, this could contribute to the lack of activity. However, since cells were grown in glucose free medium, this seems unlikely. Another possibility for the failure of obtaining active recombinant enzyme could relate to the use of the *E. coli* expression system. This system is the easiest available to obtain recombinant protein. However, it does not permit addition of secondary modifications, which could be essential for the activity of the trypanosomatid PRODH. If secondary modification is important, system such as yeast or insect cells or trypanosomatids themselves for over-expression could be tried in order to increase the probability of producing an active protein. *Pichia pastoris,* for example, can generate glycoproteins than resemble those of higher eukaryotes (high-mannose glycosylation pattern), other posttranslational modifications, such as the formation of disulfide bonds, acylation or phosphorylation could be essential for trypanosomatid PRODHs, but excluded in the *E. coli* expression system. The presence of trypanosome-specific cofactors has been tested by adding, soluble trypanosome extract to the recombinant protein, but without any success. It is also noteworthy that Western blot analysis revealed bands in *T. brucei* of two sites (section 6.2.3.1). It is conceivable that proteolytic processing of the protein (not possible in *E. coli*) could play a role in activating the enzyme.

Chapter 6

The physiological role of proline dehydrogenase in *T. brucei* determined through RNA interference

6.1- Introduction

T. brucei long slender bloodstream forms (BSF) are entirely dependent on glycolysis for the generation of ATP, with glucose as the preferred energy source (Tielens and Van Hellemond, 1998). The first seven steps of glycolysis occur in the glycosome, which is a unique organelle present in the order Kinetoplastida (Clayton and Michels, 1996). The last three steps of glycolysis occur in the cytosol leading to the production of pyruvate and ATP. In contrast to other trypanosomatids the BSF of *T. brucei* has a poorly developed mitochondrion, without a Krebs cycle or respiratory system coupled to ATP synthesis (Michels *et al.*, 2000). This apparently inefficient metabolism does not have any adverse effect since the parasites are constantly exposed to a high glucose concentration in the host's blood.

The transformation from BSF to procyclic forms (PCF) leads to striking changes in glucose metabolism (Durieux *et al.*, 1991). The level of most glycolytic enzymes is reduced in PCF compared to long slender BSF (Opperdoes *et al.*, 1984). In the PCF, pyruvate is metabolised further inside a more developed mitochondrion, where an electron transport chain is apparently coupled to oxidative phosphorylation (Tielens and Van Hellmond, 1998) although ATP production at this level appears to be non-essential (Coustou *et al.*, 2003).

Insect (including tsetse fly) haemolymph is rich in proline and/or glutamate (Balogun, 1974; Bursell, 1981), which is used as a major energy source during the initial phase of flight. In the tsetse fly, proline is first oxidised to glutamate, which serves as an important source of substrates for the Krebs cycle (Bursell, 1967). Insects appear to have a high capacity for proline synthesis. It is associated with the fat body with alanine and triglycerides representing the major substrates for proline synthesis in the tsetse fly (Bursell *et al.*, 1973; Bursell, 1978).

The insect forms of *T. brucei* have been proposed to be adapted to the environment in the vector by oxidising proline to glutamate and using it as a substrate for ATP production (Gutteridge and Coombs, 1977; Coustou *et al.*, 2003). It was proposed in an early study that during the exponential phase of growth *T. brucei* prefer proline over glucose as an energy source (Evans and Brown, 1972). Other work showed proline to be rapidly oxidized by cultured midgut forms of *T. rhodesiense* (Ford and Bowman, 1973).

Over the past few years, the understanding of PCF *T. brucei* energy metabolism has changed. The accepted dogma was that ATP was produced by oxidative phosphorylation linked to a respiratory chain oxidising products of the Krebs cycle. It was also believed that the Krebs cycle was fully functional. However, mitochondrial substrate level phosphorylation has now been shown to be essential for this stage of the parasite's life cycle (Allemann and Schneider, 2000; Bochud-Allemann and Schneider, 2002). Succinyl-CoA syntethase (SCoAS), present in both the Krebs cycle and in the Acetate:Succinate CoA transferase / SCoAS cycle (ASCT cycle), has been shown, using RNA interference, to be essential to the parasite (Bochud-Allemann and Schneider, 2002). A cytosolic pyruvate kinase has also been shown to be essential (Coustou *et al.*, 2003), although levels of this enzyme were previously reported to be negligible (Fairlamb, 1989). The importance of the Krebs cycle in procyclic stage *T. brucei* has also become controversial. Deletion of the

aconitase gene in these organisms did not change growth rate of the parasite, or the intracellular ratio of ATP/ADP (van Weelden *et al.*, 2003). Succinate dehydrogenase and α-ketoglutarate dehydrogenase genes have also been shown to be non-essential (Bochud-Alleman and Schneider, 2002). These data indicate that the Krebs cycle is not critical in the energy metabolism of these cells. Notwithstanding this, PRODH converts proline to glutamate, which was believed to enter the Krebs cycle for production of succinate and ATP (Evans and Brown, 1972; Obungu *et al.*, 1999) although it might also feed other pathways of energy metabolism.

In order to improve our understanding of PCF *T. brucei* energy metabolism in general and specifically the roles of PRODH in this process, the technique of RNA interference has been applied to functionally silence the enzyme.

RNA interference (RNAi) refers to the impact of the introduction of homologous double stranded RNA (dsRNA) to a cell to specifically target a gene product (mRNA), resulting in "null" phenotypes (LaCount *et al.*, 2000; Ullu *et al.*, 2002). The RNAi pathway is ATP-dependent at several steps (Zamore *et al.*, 2001). Cleavage of the dsRNA into small interfering RNAs (SiRNAs) involves the enzyme Dicer (RNAse III family). Then, SiRNAs complex with other proteins and, after ATP-dependent unwinding, the active RNA-induced silencing complex is able to degrade both sense and anti-sense target RNA molecules (Hutvagner and Zamore, 2002). Most studies suggest that RNAi takes place in the cytoplasm (Hannon, 2002).

RNAi was first described in *Caenorhabditis elegans*, but has since been observed in a wide range of organisms including *T. brucei* (Ngo *et al.*, 1998). A silencing mechanism has also recently been reported in *L. major* and *L. collosoma* in experiments aimed at silencing the small nuclear RNAs (Liang *et al.*, 2003), although other efforts to exploit the system in *Leishmania* have been unsuccessful (Beverley, 2003).

RNAi interferes with expression at the RNA level, to generate a "knock down" phenotype, without altering the DNA sequence, which is required for conventional gene knock out, in which the gene locus has to be deleted. Although complete abrogation of expression is rarely observed with RNAi, RNAi can efficiently "knock down" protein expression levels. One advantage of RNAi is that it can be used to simultaneously "knock down" expression of multiple gene copies without these needing to be physically linked in the genome (Wang *et al.*, 2000).

In *T. brucei*, the generation of knock down phenotypes via RNAi is achieved by introduction of a plasmid containing two tetracycline inducible T7 promoters, arranged as an inverted repeat (LaCount *et al.*, 2000). The gene of interest is placed between these two promoters and formation of dsRNA guides degradation of the target mRNA (Tschudi *et al.*, 2003). The *T. brucei* 29-13 cell line used for RNA interference experiments (section 2.4.1) has been engineered to express the Tet repressor and the T7 RNA polymerase.

This chapter describes the "knock down" of PRODH in *T. brucei* and discusses the implications of proline utilisation for energy metabolism in trypanosomes.

6.2- Results

6.2.1- Plasmid construction and clone selection

The N-terminal part of the proline dehydrogenase gene, corresponding to 606 bp of the gene (see section 2.4.2), was amplified by PCR and cloned into the p2T7Ti vector to generate the plasmid p2T7PRODH. In order to check the construct, p2T7PRODH was digested by *Xba* I to liberate the insert corresponding to the PRODH fragment (Figure 6.1).

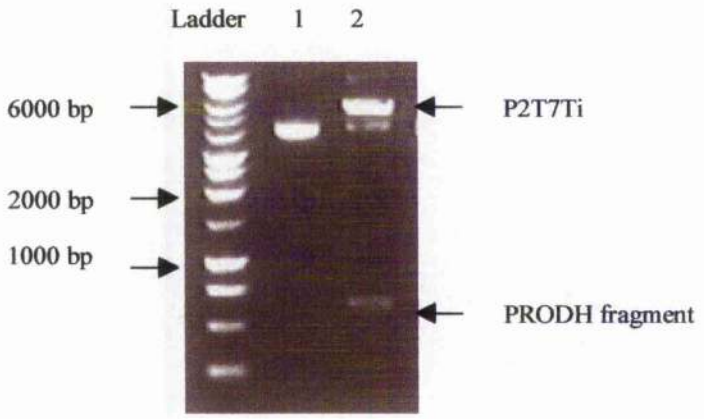


Figure 6.1: Ethidium bromide stained agarose gel showing the 606 bp proline dehydrogenase fragment in p2T7Ti. It was digested by *Xba* I (lane 2) and non-digested (lane 1).
Size of p2T7Ti: 6089 bp

p2T7PRODH was linearised by *Not* I (Figure 6.2) and transfected into *T. brucei*, 29-13, cultured in the presence of G418 and Hygromycin B, permitting the selection of the cells that had integrated the T7 RNA polymerase and the tetracycline repressor. Transfected parasites were grown in the presence of Phleomycin at a concentration of 10 μg/ml, which selects for transformants carrying the p2T7PRODH plasmid.

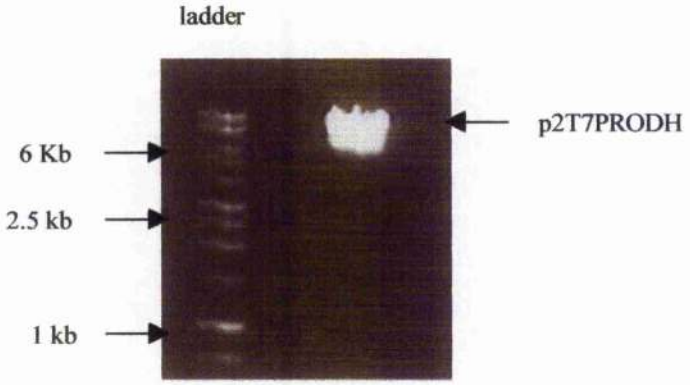


Figure 6.2: p2T7PRODH construction. 1% agarose gel showing the vector p2T7PRODH linearized by *Not* I.

Multiple clones were selected by limiting dilution (see section 2.4.4.). The clone F2 was chosen for further study.

6.2.2- RNA expression

The Δ*prodh* parasites were grown in SDM79 medium in the presence of antibiotics to maintain the construct. Tetracycline, at a concentration of 1 μg/ml, was added to induce the formation of double stranded RNA and therefore the destruction of the proline dehydrogenase transcript.

Total RNA was isolated from induced (Δ*prodh* +) and non-induced (Δ*prodh* -) cells and subjected to Northern blot analysis. The blot was probed with the *PRODH* ORF, and subsequently stripped and re-probed with a β tubulin gene as a control for the quantity and quality of RNA.

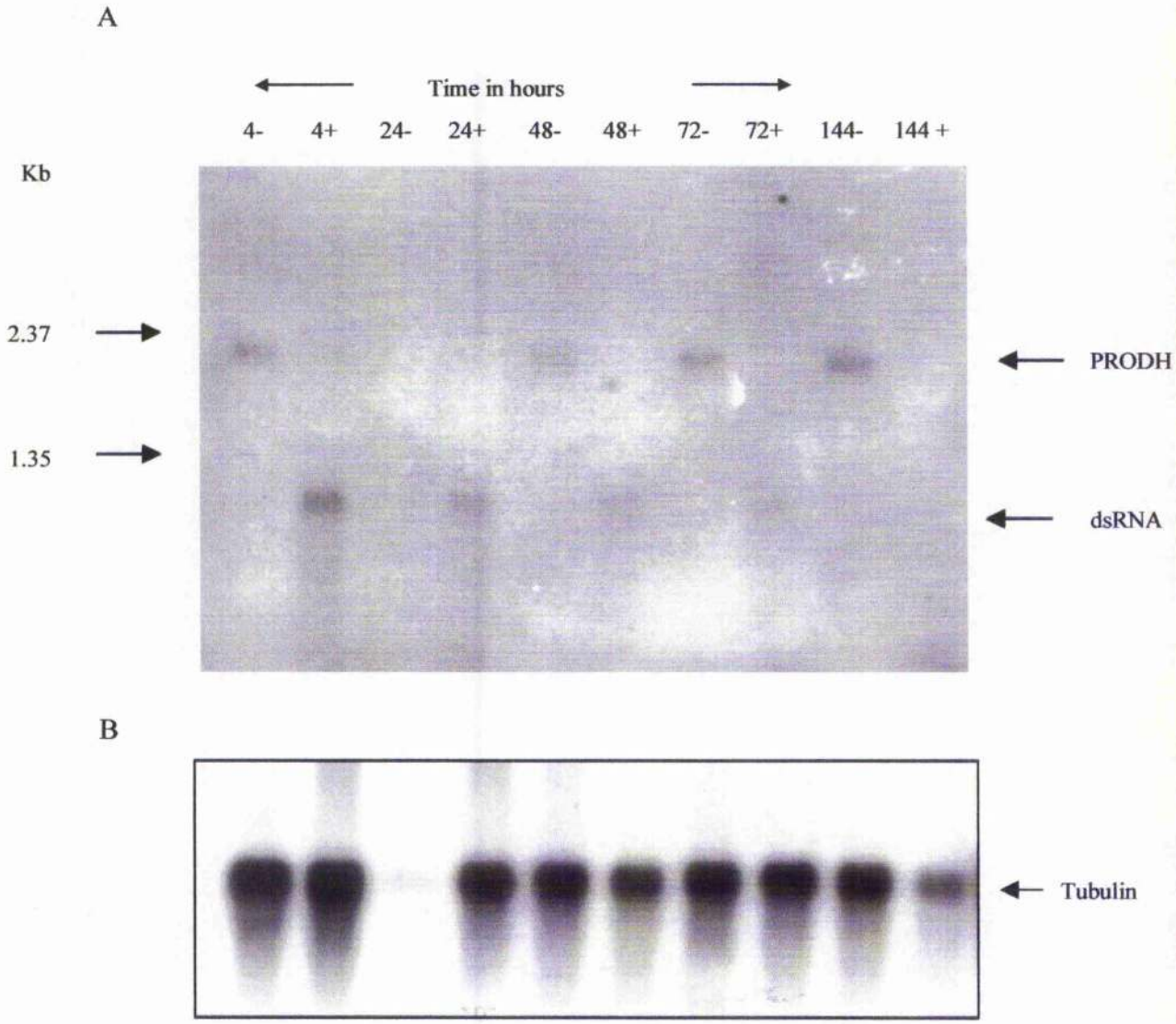


Figure 6.3: Northern blot analysis of *T. brucei* cells transfected with p2T7PRODH construct induced (+) and non-induced (-) by 1 μg/ml of tetracycline. A: The blot was probed with the proline dehydrogenase ORF. B: The blot was stripped, then re-probed with a β tubulin gene. The Northern blot shows a time course experiment with the numbers above each lane referring to the time in hours post-induction of dsRNA with tetracycline.

Figure 6.3 shows a time course experiment. The cells were induced with tetracycline and RNA was collected from 4 hours after induction to 6 days. A single hybridizing band of ~2 Kb was detected in RNA from non-induced cells at all time points, except for 24 hours. Note that no signal was observable within the non-induced cells at 24 hours, because of the loss of the RNA in this sample as witnessed by the tubulin control. Within 4 hours of induction of dsRNA expression, the 2 Kb PRODH transcript was not detectable. A smaller hybridizing transcript of ~600 bp was detected only in Δ*prodh* + cells. This band is likely to represent the 600 bp dsRNA that is transcribed from p2T7PRODH under conditions of induction.

6.2.3- Protein expression and proline dehydrogenase activity

6.2.3.1-Western blot analysis

To investigate the expression of PRODH protein in the RNAi cells, Western blots were performed with lysates from both induced and non-induced Δ*prodh* transfected *T. brucei*. The serum used during this experiment was raised in a rabbit by immunisation with purified recombinant proline dehydrogenase and subsequently purified (see section 2.3.5).

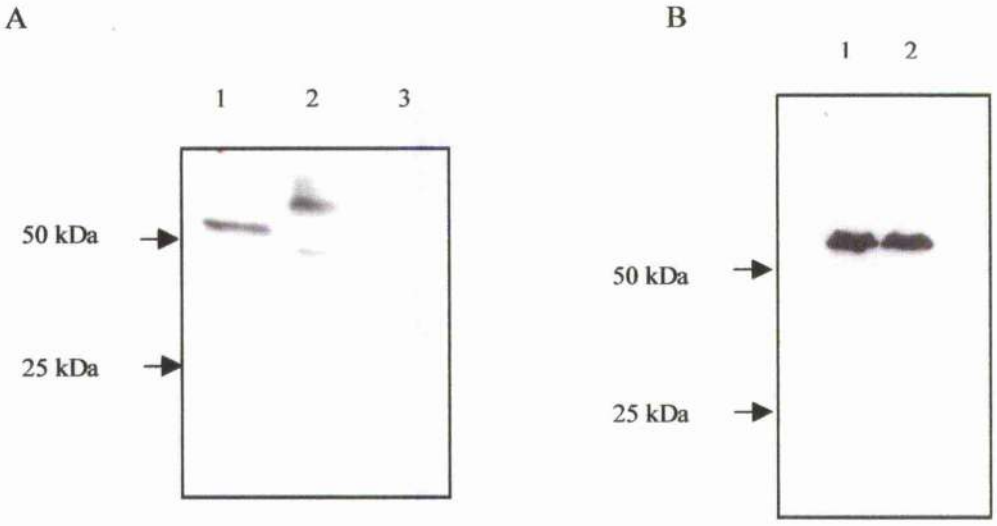


Figure 6.4: A: Detection of proline dehydrogenase using rabbit anti-proline dehydrogenase antibodies (1:3000 dilution). Lane 1, recombinant proline dehydrogenase without 72 first amino acids (10 µl of 0.3 mg/ml); lane 2, Δ*prodh* – *T. brucei* procyclic cells lysate ($2x10^7$ cells), lane 3, Δ*prodh* + *T. brucei* cells lysate ($2x10^7$ cells) after 4 days of induction.

B: Detection of tubulin using rabbit anti-tubulin antibodies (1:10,000). lane 1, Δ*prodh* – *T. brucei* procyclic cells lysate ($2x10^7$ cells), lane 2, Δ*prodh* + *T. brucei* cells lysate ($2x10^7$ cells) after 4 days of induction.

Two bands (45 kDa and 65 kDa) were detected in the non-induced RNAi cells. These bands were absent in RNAi cells that had been induced for 4 days with tetracycline (Figure 6.4). As a control, the PRODH without the 72 first amino acids recombinant protein (57.8 kDa) was run in the same gel. A control for the quantity of protein was performed using anti-tubulin antibodies. In both cell lines, a band at around 55 kDa, representing tubulin, was found. In the first gel (Figure 6.4A), the band found around 65 kDa could be the PRODH (the entire protein has a predicted size of 63 kDa). The smaller band could be a degradation product of the protein or the protein cleaved to release the putative trans-membrane domain or another domain. The protein devoid of the trans-membrane domain

should have a size of 56 kDa, although additional sequence beyond this trans-membrane domain could also be lost during this cleavage. The possibility that the band results from non-specific binding of the antibody cannot be excluded, but a BLAST search of the parasite genome and TIGR databases with the *T. brucei* PRODH as a query did not show significant homology to any other protein, which could have been recognised by the antibody. Moreover, loss of the band upon induction of the RNAi suggests that it is related to PRODH.

6.2.3.2- *Proline dehydrogenase activity in the RNAi cell line*

PRODH activity was measured using crude lysates of *T. brucei* Δ*prodh* + and Δ*prodh* – as described in section 2.3.9. For a substrate range concentration from 1 mM to 60 mM (Figure 6.5), the activity of the enzyme in non-induced parasites was comparable with that in the wild type parasites (Chapter 4). Very low activity was detectable when the construct was induced. The specific activity of PRODH in *T. brucei* was found to be maximal for a proline concentration of 10 mM. It was 0.028 ± 0.004 $\mu mol.min^{-1}.mg\ protein^{-1}$. Beyond this range, an apparent substrate inhibition was observed.

A control for this experiment was done by measuring the activity of cysteine synthase in the crude lysate of Δ*prodh* + and Δ*prodh* – cells lines (Kredich and Tomkins, 1966). This assay was performed by Dr R. Williams (Univeristy of Glasgow), and the specific activity found in both cells line was similar (results non shown).

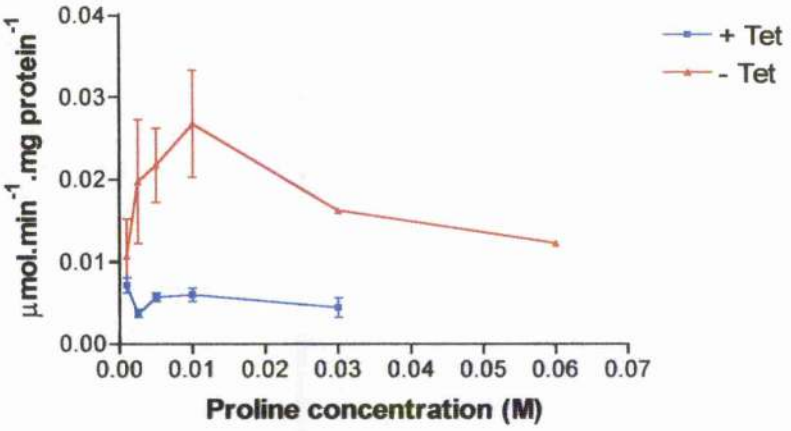


Figure 6.5: Proline dehydrogenase activity in lysate of *T. brucei* Δ*prodh* procyclic forms induced and non-induced by tetracycline. Error bars represent ± SEM, n=3

The Δ*prodh* parasites were grown in a complete SDM79 medium containing proline and glucose. PRODH specific activity, using 10 mM proline as substrate in the assay, was found to be approximately twice as high when the cells were grown with glucose absent from the medium, reaching a maximum of 0.054 ± 0.004 µmol.min^{-1}.mg protein^{-1} compared with 0.027 ± 0.004 µmol.min^{-1}.mg protein^{-1} when the medium was complete (significantly different, by t-test, $P<0.05$).

Although the PRODH activity of *T. brucei* crude lysate was much higher in cells grown in the absence of glucose, it was not possible to grow the Δ*prodh* + cells under these conditions (as cells are unable to grow when glucose is not present and when proline cannot be used due to the lack of PRODH).

6.2.4- Growth studies of RNAi cell lines

Δprodh - and *Δprodh* + parasites were also studied for a phenotype beyond those measurable biochemically. The *T. brucei* procyclic cells were grown in SDM79 medium. A typical logarithmic growth curve was observed (see Figure 6.6), reaching a stationary phase after 7-8 days at approximately $2x10^7$ cells/ml. There was no apparent difference in growth in the presence or absence of PRODH activity. Thus the presence or absence of proline degradation through PRODH does not affect the growth of *T. brucei* procyclic parasites when grown in rich medium.

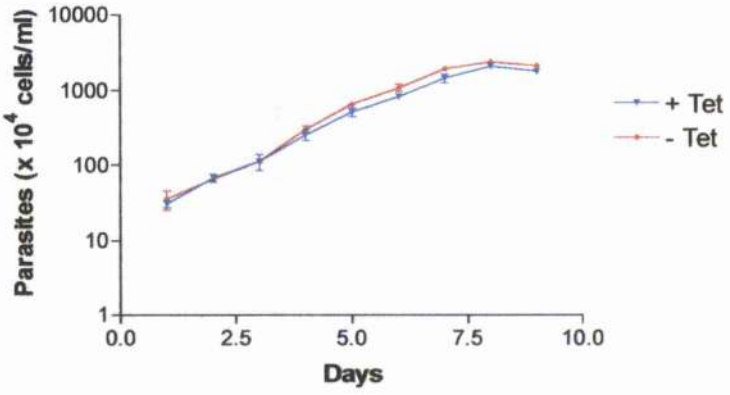


Figure 6.6: Growth of *T. brucei* *Δprodh* in SDM79 medium in the absence or presence of tetracycline (1 μg/ml). Error bars represent ± SEM, n=3

A modified SDM79 medium without proline or glucose was prepared to test the viability of trypanosomes in the absence of the main energy sources thought to be used by the parasite. This medium was supplemented with proline and/or glucose at 0.6 g/L (5.2 mM)

and 1 g/L (5.5 mM), respectively. In the presence or absence of tetracycline, growth of *Δprodh T. brucei* cells was monitored under various conditions (Figure 6.7).

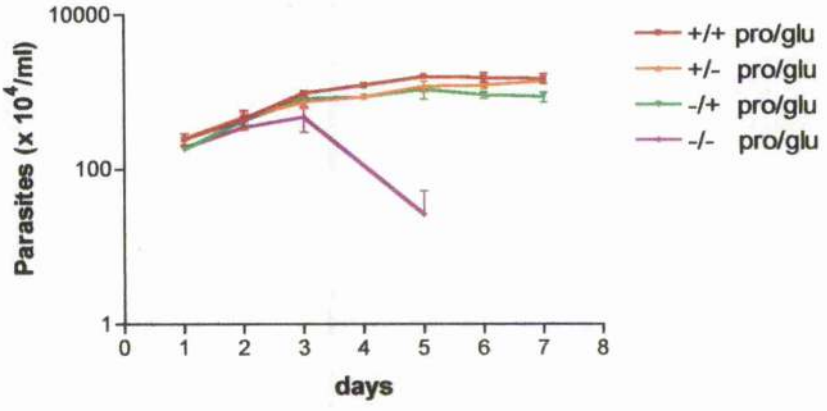


Figure 6.7A: Growth of non-induced *T. brucei* Δ*prodh* in SDM79 medium lacking proline and glucose. Non-induced cells were cultured in the absence of proline and/or glucose; these substrates were at 5 mM when present. Error bars represent ± SEM, n=3

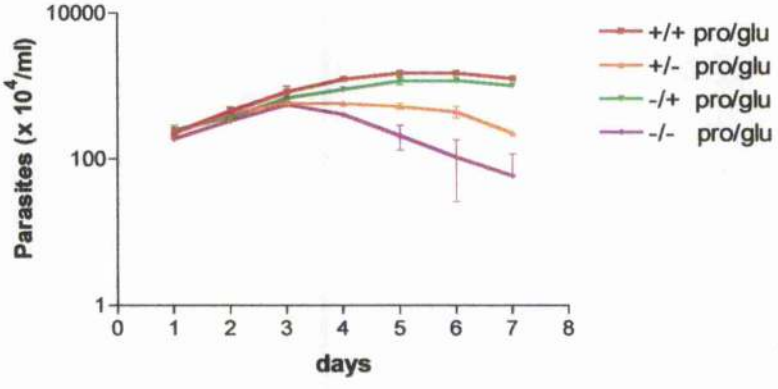


Figure 6.7B: Growth of tetracycline-induced *T. brucei* Δ*prodh* in SDM79 medium lacking proline and glucose. Induced cells were cultured in the absence of proline and/or glucose; these substrates were at 5 mM when present. Error bars represent ± SEM, n=3

In the absence of tetracycline, cells were able to grow when either proline or glucose, or both, were present in the medium. When proline and glucose were absent, growth was observed for the first two days and then the cells died. However, in the presence of proline only in the medium, the stationary phase was reached more quickly and the maximum parasite density reached was lower than when glucose was present. This may have been due to depletion of the energy source in the medium.

When tetracycline was present, and PRODH was switched off, the cells were able to grow at the same rate in a medium with or without proline. However, the cells were unable to grow in the absence of glucose over a period of 7 days, indicating that proline can no longer support growth in these cells, probably because of the loss of PRODH. Tetracycline had no lethal effect on the growth, since, when only glucose is present in this medium the number of parasites is still more abundant than when only proline is present.

6.2.5- Supplementation of growth by addition of other amino acids

In the experiments detailed in section 6.4, *T. brucei* PCF were able to grow in SDM79 only when proline or glucose was present at high concentration. Glutamate is a degradation product of proline catabolism that can be used as a substrate for the Krebs cycle. It was therefore of interest to determine whether glutamate could rescue the cells that could not longer utilise proline. Moreover, it was investigated whether other amino acids could substitute for proline.

Growth of *T. brucei* Δ*prodh* in the presence or absence of tetracycline and with different supplements for the medium was estimated using the Alamar blue assay (Figure 6.8) (see section 2.1.3).

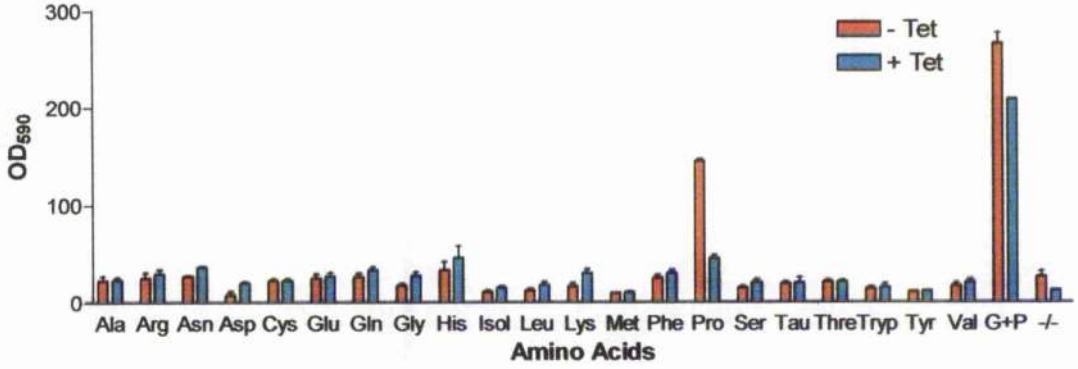


Figure 6.8: Alamar blue estimation of growth of *T. brucei* Δ*prodh* procyclic forms induced or non-induced by tetracycline in a medium lacking proline and glucose but supplemented by different amino acids at 10 mM. The three letter code for the amino acids was used. G+P is the medium supplemented by glucose and proline, -/- is the medium without glucose or proline. Error bars represent ± SEM, n=4

Parasites, induced or non-induced by tetracycline, were grown in SDM79 without glucose or proline, supplemented with various single amino acids at 10 mM. Proline and glucose were added to the medium as a positive control. The negative control (-/-) lacked any added amino acids.

None of the amino acids was able to rescue the cells from the loss of proline utilisation and the lack of glucose in the medium. Only Δ*prodh* - parasites, non-induced with tetracycline, were able to grow in the presence of just proline in the medium. In the conditions tested, proline or glucose is thus essential for parasite growth. Glutamate and other amino acids were not able to supplement for the loss of proline utilisation.

Experiments using different concentrations of glucose, proline and glutamate were also performed (Figure 6.9).

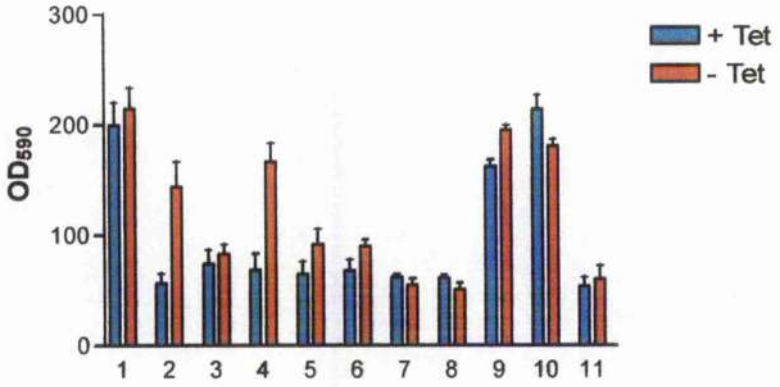


Figure 6.9: Alamar blue estimation of growth of *T. brucei* Δ*prodh* procyclic forms induced and non-induced by tetracycline in a medium lacking proline and glucose and supplemented by different concentrations of proline, glucose and glutamate. Lane 1; glucose 10 mM; proline 10 mM, glutamate 0 mM; **lane 2**; glucose 0 mM, proline 10 mM, glutamate 0 mM; **lane 3**; glucose 0 mM, proline 0 mM, glutamate 10 mM; **lane 4**; glucose 0 mM, proline 10 mM, glutamate 10 mM; **lane 5**; glucose 0 mM, proline 0.1 mM, glutamate 10 mM; **lane 6**; glucose 0 mM, proline 0.01 mM, glutamate 10 mM; **lane 7**; glucose 0 mM, proline 0.1 mM, glutamate 0 mM; **lane 8**; glucose 0 mM, proline 0.01 mM, glutamate 0 mM; **lane 9**; glucose 10 mM, proline 0 mM, glutamate 0 mM; **lane 10**; glucose 10 mM, proline 0 mM, glutamate 10 mM, **lane 11**; glucose 0 mM, proline 0 mM.

Error bars represent ± SEM, n=4

Glutamate at 10 mM was not toxic for the *T. brucei* parasites, as evidenced by addition of 10 mM glutamate to medium already containing glucose or proline. However glutamate alone could not support growth of PRODH-deficient cells (lanes 3 and 4). In the conditions tested, proline at 10 mM seems to be essential for the cells when glucose is absent, a lower concentration of proline (0.1 or 0.01 mM) does not support growth. However, in these conditions a slight increase in the number of parasites alive was observed, showing that the

parasites might need some proline, perhaps for roles other than energy production such as protein synthesis or osmoregulation.

According to various schemes (Coustou *et al.*, 2003) α–ketoglutarate, succinate or pyruvate might be able to be used in the production of ATP by *T. brucei*. However, none of these metabolites were able to support the growth of parasites in medium without glucose or proline. Pyruvate seemed to inhibit the growth of parasites when present at 10 mM (Figure 6.10, lane 5). Some intermediate growth can be observed when the medium is complemented with threonine, succinate or α-ketoglutarate, but only proline and glucose seem to be able to support robust growth.

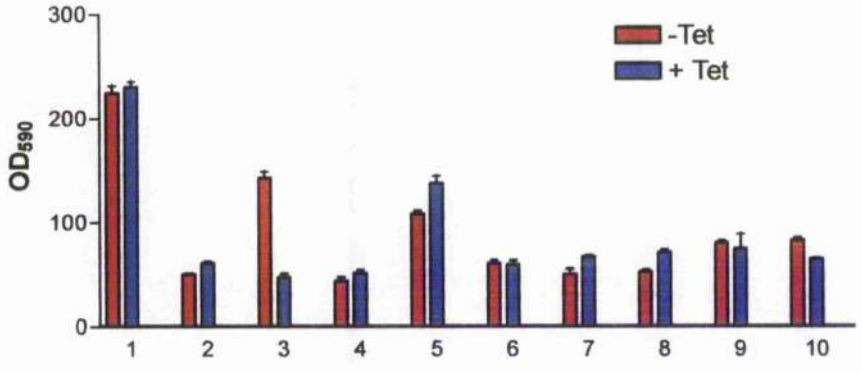


Figure 6.10: Alamar blue estimation growth of *T. brucei* Δ*prodh* procyclic forms induced and non-induced by tetracycline in a medium lacking of proline and glucose and supplemented by different components at 10 mM. Lane 1; glucose 10 mM, proline 10 mM; **lane 2**; glucose 0 mM, proline 0 mM, **lane 3**; glucose 0 mM, proline 10 mM; **lane 4**; glucose 0 mM, proline 0 mM, pyruvate 10 mM; **lane 5**; glucose 10 mM, proline 10 mM, pyruvate 10 mM; **lane 6**; glucose 0 mM, proline 0 mM, alanine 10 mM, aspartate 10 mM; **lane 7**; glucose 0 mM, proline 0 mM, succinate 10 mM; **lane 8**; glucose 0 mM, proline 0 mM, pyruvate 10 mM, succinate 10 mM; **lane 9**; glucose 0 mM, proline 0 mM, threonine 10 mM, succinate 10 mM; **lane 10**; glucose 0 mM, proline 0 mM, α-ketoglutarate 10 mM. Error bars represent ± SEM, n=4

6.2.6- Glutamate uptake

Supplementation with various amino acids was not able to rescue growth of *T. brucei* in glucose- and proline-free media. Glutamate, which is a degradation product of proline, was not able to support growth in cells that had lost the ability to use proline. One possible reason for glutamate's inability to support growth could relate to a failure of trypanosomes to accumulate this metabolite across their plasma membrane. Therefore glutamate uptake experiments were carried out in order to determine whether or not a glutamate transporter was operative in *T. brucei* PCF.

In a first experiment, uptake of 10 μM glutamate was monitored as a function of time (see section 2.4.6.3) (Figure 6.11).

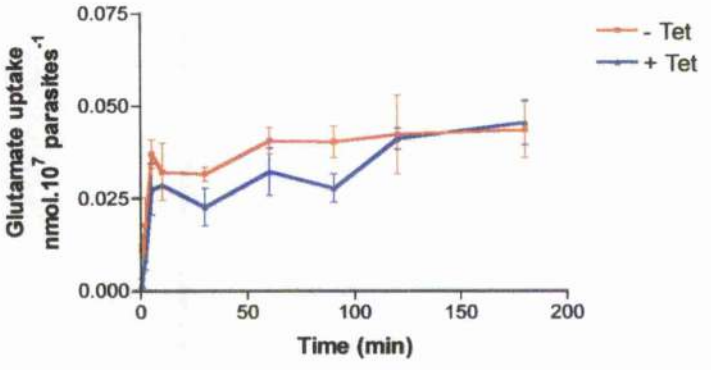


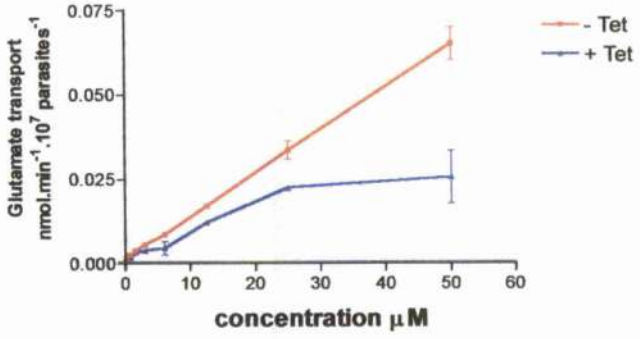
Figure 6.11: Time course of glutamate uptake into *T. brucei* Δ*prodh* procyclic forms induced and non-induced by tetracycline. Transport was measured in the presence of 10 μM of glutamate for a range of time between 30 seconds and 3 hours. Error bars represent ± SEM, n=6

There was an accumulation of glutamate with time. An increase in glutamate incorporation was observed during the first 10 minutes followed by a plateau, a maximum of 50 pmol per 10^7 cells was measured inside the cells. No significant difference (t-test, P>0.05) was observed between the two lines, Δ*prodh* + and Δ*prodh* -. For both lines, the concentration of glutamate within the cells after 10 min was determined at 63 μM compared with 10 μM outside. This was calculated assuming a volume of 5.6 μl for 1x 10^8 cells (Knodler *et al*, 1992) and using the maximal amount of glutamate in the cells.

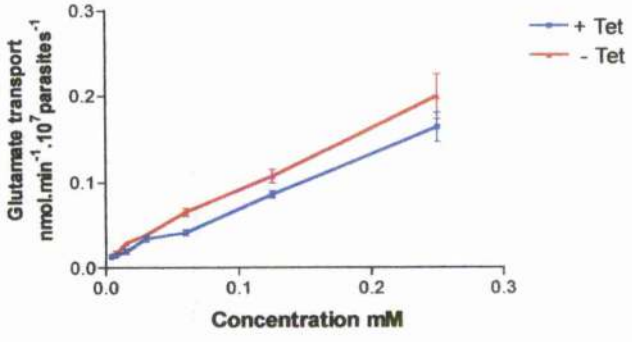
It has been previously noticed that the incorporation of glutamate into procyclic *T. brucei* proceeds much more slowly than uptake of other amino acids (Dr I. Accoceberry, University Bordeaux II, personal communications). For example, at 1 μM the rate of incorporation of glutamate was only 10 % of the rate of incorporation of alanine, glutamine or serine and 3.9 % of the incorporation of leucine.

Thus kinetics of glutamate transport were then studied using two different approaches. In the first, uptake was measured over 30 s using several concentration ranges of glutamate at doubling dilutions of labelled substrate. For both cells, induced or not induced by tetracycline, concentration ranges between 100 μM and 20 mM did not reveal any saturable transporter (Figures 6.12 B and C).

A



B



C

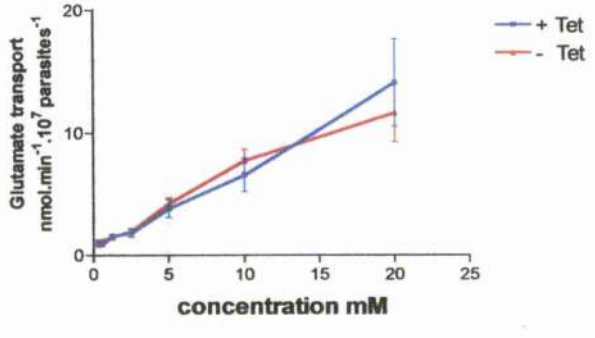
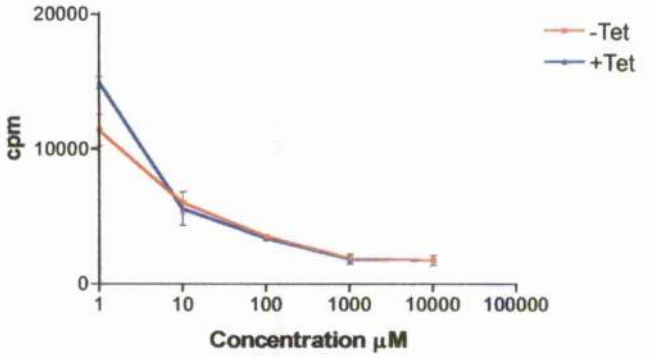


Figure 6.12: Kinetics of glutamate transport in *T. brucei* Δ*prodh* procyclic forms induced and non-induced by tetracycline. Glutamate transport was measured in the presence of various glutamate concentrations over 30 s. A: range 0.15 µM to 50 µM glutamate. B: range 3.8 µM to 0.25 mM glutamate. C: range 0.3 mM to 20 mM glutamate. Error bars represent ± SEM, n=6

Surprisingly, for a lower range up to 50 μM, the induced *T. brucei* Δ*prodh* did show a hyperbolic curve tending to saturation compared with non-induced parasites in which no saturation curve was observed although a simple explanation for this is not evident (Figure 6.12 A).

A second type of experiment was performed to try and show the presence of a specific glutamate transporter (Figure 6.13).

A



B

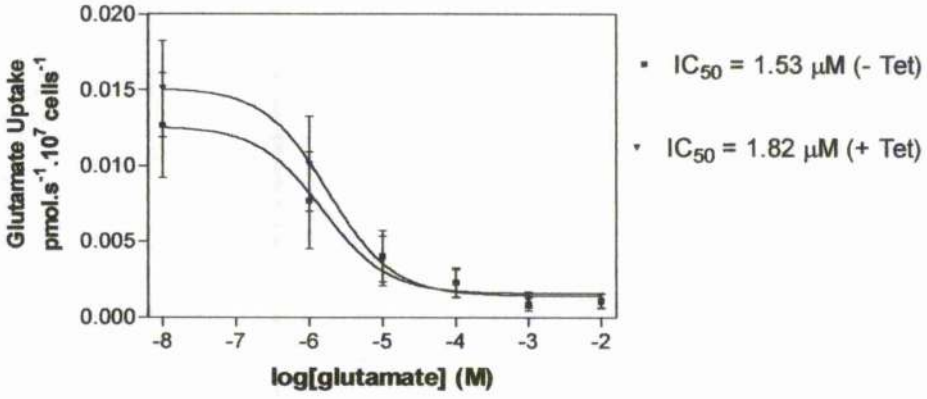


Figure 6.13: Inhibition of ^{3}H glutamate transport by non-radiolabelled glutamate. The uptake was measured over 30 s in the presence of increasing concentration of non-radiolabelled glutamate.
(A) shows the radiation (cpm: count per minute) taken into the cells. (B) shows the transformation of these cpm into moles. 10 μCi of glutamate equal to 190 nM was used. Error bars represent ± SEM, n=4

In this case, transport experiments were conducted over a period of 30 s using a constant concentration (10 μCi corresponding to 190 nM) of radiolabelled substrate and increasing the concentration of non-radiolabelled substrate. A decrease of the incorporated radiolabelled was observed with increasing concentrations of non-radiolabelled amino acid. This result indicates that a high affinity glutamate transporter is present with a K_m of around 1.5 μM. Further experiments are required to characterise this transporter. The capacity of this transporter appears to be low and the non-saturable uptake events, described above, prevented accurate assessment of the V_{max} of this transporter.

6.2.7- Proline transport

Loss of PRODH through RNAi led to a loss in the ability of trypanosomes to use proline. This could indicate that PRODH is essential in the pathway to energy production from proline. However, it is also possible that loss of PRODH had other effects on the cells that could lead to loss of viability. The first step in the metabolism of proline involves its uptake across the plasma membrane. Proline transport has previously been studied in procyclic *T. brucei* (L'Hostis *et al.*, 1993). Proline transport was here compared in cells expressing and suppressed in expression of PRODH.

In first experiments, proline transport in *T. brucei* Δ*prodh* was studied as a function of time (Figure 6.14).

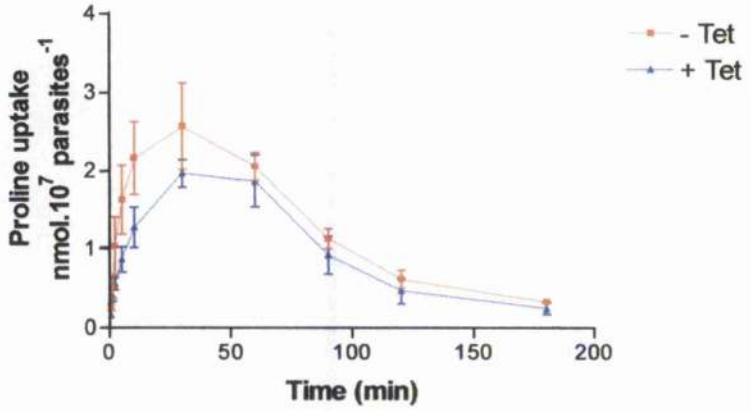


Figure 6.14: Time course of proline uptake in *T. brucei* Δ*prodh* procyclic forms induced and non-induced by tetracycline. The transport using an external concentration of 100 μM proline was measured at a range of time points between 30 seconds and 3 hours. Error bars represent ± SEM, n=6

Uptake of 100 μM proline containing ^{3}H proline was allowed to proceed for up to 3 hours. An accumulation of proline was observed for 30 min. This was followed by an apparent release of this amino acid. The non-induced cells accumulated proline 45 fold when compared to the external concentration again assuming an extracellular volume of 5.6 μl for 1x 10^{8} cells (Knodler *et al.*, 1992), whereas the induced cells accumulated the amino acid around 30 fold. A t-test showed that this difference was not significant at $P>0.05$. The absence of PRODH activity does not appear to abolish uptake. A similar pattern was seen over both short term and long-term uptake experiments. The accumulation of proline within the cells (presumed not to be metabolised due to the lack of PRODH) did not interfere with uptake. Previous studies have suggested proline uptake involves an active transport (L'Hostis *et al.*, 1993) in PCF *T. brucei*. The results presented here reinforce this

hypothesis. Using micromolar concentration of substrate, the maximum intracellular concentration of glutamate was much lower than for proline (see section 6.6) where up to 2-3 nmol per 10^7 cells could be maximally achieved (with 100 μM external proline).

Uptake of proline was investigated as a function of time between 3 s and 30 s (Figure 6.15).

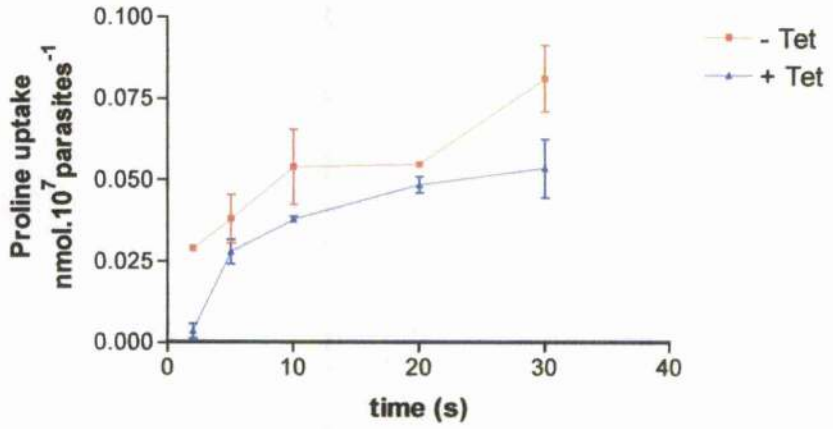
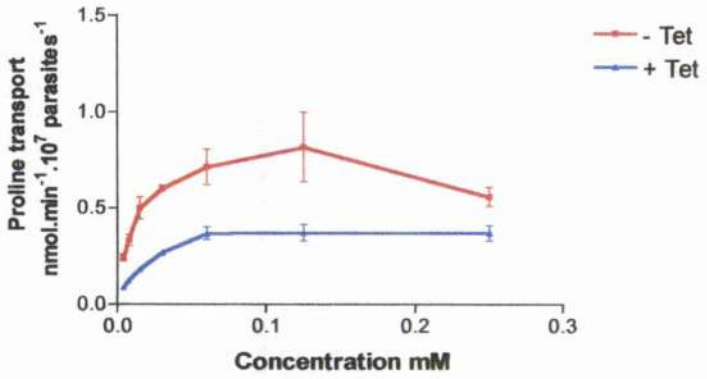


Figure 6.15: Time course of proline uptake in *T. brucei* Δ*prodh* procyclic forms induced and non-induced by tetracycline. The transport using an external concentration of 100 μM proline was measured at a range of time points between 3 seconds and 30 seconds. Error bars represent ± SEM, n=6

The uptake of proline did deviate somewhat from linearity during this period, but 30 s was considered to be the fastest time point at which reproducible data could be reliably gathered, and it was close to the time point (20 s) chosen by l'Hostis *et al.*, 1993. Thus a time point of 30 seconds was chosen for some subsequent experiments. The kinetics of proline transport was also studied using various concentrations of non-radiolabelled proline ($3x10^{-3}$ mM to 20 mM) (see section 2.4.6.3) (Figure 6.16).

A



B

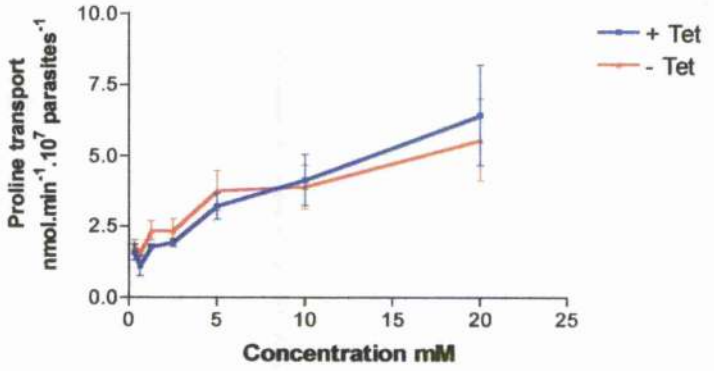


Figure 6.16: Kinetics of proline transport of *T. brucei* Δ*prodh* procyclic forms induced and non-induced by tetracycline (Δ*prodh* + and Δ*prodh* -). Proline transport was measured in the presence of various proline concentrations over 30 s. A: Concentration range 3.9 to 250 μM B: Concentration range 300 μM to 20 mM. Error bars represent ± SEM, n=6

Under high, non-physiological proline concentrations (Figure 6.16 B), non-saturable uptake of proline was observed within the cells. Using lower concentrations (Figure 6.16 A), a typical hyperbolic curve was observed for the cells under both conditions with regard to PRODH expression. The Δ*prodh* - cells have a apparent V_{max} of 0.75 ± 0.06 nmol.min^{-1}.10^{7} parasites^{-1} and apparent K_m of 11 ± 0.003 μM for proline at 20 °C. L'Hostis *et al*, (1993) found for proline transport in *T. brucei* a K_m of 19 μM and a V_{max} of 1.7 nmol.min^{-1}.10^{7} parasites^{-1} at 27 °C. Therefore the proline kinetics measured here for non-induced RNAi cells are similar to those previously reported to wild type *T. brucei*. In the presence of tetracycline, the K_m observed was 12 ± 0.0035 μM and V_{max} is 0.41 ± 0.0035 nmol.min^{-1}.10^{7} parasites^{-1}. Thus measured in the presence or absence of tetracycline the K_m is comparable although the V_{max} is significantly different (t-test, $P<0.05$). The apparent V_{max} is lower when proline metabolism is switched off.

Over 30 s it appeared that the K_m for the proline uptake process was unchanged in cells with PRODH and cells lacking this enzyme. The V_{max}, however, was reduced by around two fold in the cells without PRODH. 30 s was selected as a time point because of convenience, it could not be ruled out that the measured uptake in non-induced cells was a result of a combined effects of uptake and metabolism. The reduction in apparent V_{max} could then be explained by loss of the principal metabolic step (proline dehydrogenase). Oxidation of proline by PRODH may not be the only route to its metabolism in which case it is possible that measured uptake is the sum of transport and metabolism. Therefore it was decided to measure uptake over the shortest possible time period. The shortest point at which one can physically measure uptake using the oil stop method is around 3 s. Therefore proline uptake at 3 s was also measured for a concentration range of 3 μM to 200 μM (Figure 6.17).

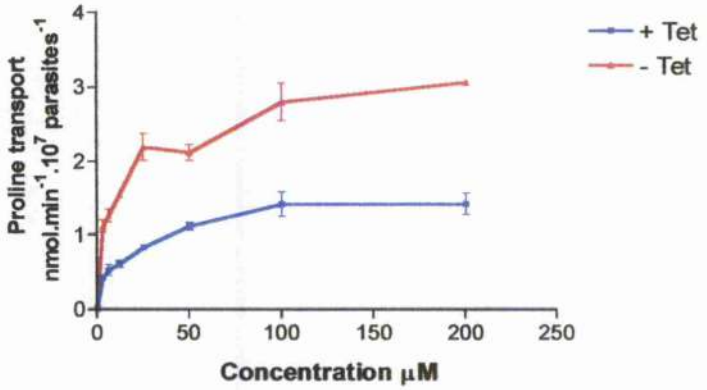


Figure 6.17: Kinetics of proline transport of *T. brucei* Δ*prodh* procyclic forms induced and non-induced by tetracycline. Proline transport was measured in the presence of various proline concentrations over 3 s. Error bars represent ± SEM, n=6

Using either 30 s or a 3 s time point to follow uptake, the pattern of the kinetics of proline uptake was similar. The V_{max} was significantly higher in the non-induced cells (t-test, $P<0.05$) while the K_m was similar in both conditions. At 3 s the apparent V_{max} was 1.5 ± 0.08 nmol.min^{-1}.10^7 parasites^{-1} and the apparent K_m was 13.2 ± 3.0 µM. For the cells grown without tetracycline the apparent V_{max} was 2.7 ± 0.13 nmol.min^{-1}.10^7 parasites^{-1} and the apparent K_m was 9.3 ± 1.5 µM. The higher apparent V_{max} at 3 s when compared with 30 seconds confirmed that the initial rate had already passed by 30 seconds. Figure 6.15, indicates that is the case.

The uptake of proline was also studied with *T. brucei* wild type cells grown in a medium lacking or containing glucose (see section 4.2.6). Similar uptake characteristics were observed for proline in these cells. When parasites were grown in the presence of glucose the apparent V_{max} was 0.70 ± 0.066 nmol.min^{-1}.10^7 parasites^{-1} and the apparent K_m

was 21 ± 2.9 µM. When grown in the absence of glucose, however, the values were apparent V_{max} 1.81 ± 0.13 nmol.min^{-1}.10^{7} parasites^{-1} and apparent K_m 18 ± 0.0042 µM. The apparent V_{max} value was thus doubled in conditions when proline metabolism appears to be essential.

6.3- Discussion

To study the physiological role of PRODH in *T. brucei* PCF, the RNA interference technique was used. Expression of a double stranded RNA molecules in cells leads to the degradation of the corresponding endogenous mRNA, by creating small (24 to 26 nucleotide) RNAs, called small interfering RNAs (Si RNAs), which interact with the target. This process is ATP-dependent (Hutvagner and Zamore, 2002; Tschudi *et al.*, 2003).

The elimination of the PRODH transcript in *T. brucei* transfected with the p2T7 vector carrying a fragment of the *PRODH* gene was evaluated by Northern blot analysis. RNA from tetracycline-induced cells and non-induced cells was compared from 4 hours up to 6 days after induction. The targeted transcript was lost from 4 hours until the parasites reached the stationary phase (6 days).

Loss of an RNA transcript is important, but the presence or the absence of the protein and its activity is critical to phenotype. The efficacy of the inducible p2T7PRODH was assessed by Western blot analysis. PRODH was detected using a specific antibody raised in a rabbit. After 4 days of induction with tetracycline, the protein could not be detected in trypanosomes. Two bands (45 kDa and 65 kDa) were detected in the non-induced RNAi cells (Figure 6.4). The band found around 65 kDa could be the PRODH (the

entire protein has a predicted size of 63 kDa). The smaller band could be a degradation product of the protein or the protein cleaved to release the putative trans-membrane domain or another domain. The protein devoid of the trans-membrane domain should have a size of 56 kDa, although additional sequence beyond this trans-membrane domain could also be lost during this cleavage. It is conceivable that only the lower molecular mass protein is active. This could explain why *E. coli* expressed protein was inactive. Further work to learn more about this apparent processing would be of interest.

PRODH activity within the Δ*prodh* cells was also evaluated. Without tetracycline, PRODH activity was detected at a similar level to the activity observed in wild type *T. brucei* (see Chapter 4). The specific activity was maximal using 10 mM proline in the assay mixture. It was 0.028 ± 0.004 $\mu mol.min^{-1}.mg\ protein^{-1}$. Higher concentrations gave lower specific activity presumably due to substrate inhibition. In the presence of tetracycline (with 4 days induction), PRODH activity was absent. It can be concluded, therefore, that the RNA interference approach successfully knocked down expression of PRODH in *T. brucei*.

When parasites were grown in complete SDM medium containing both glucose and proline, enzyme activity was substantially lower than observed in a medium lacking glucose. The specific activity was found to be twice as high when glucose was absent from the medium, reaching a maximum of 0.054 ± 0.004 $\mu mol.min^{-1}.mg\ protein^{-1}$. This compares with 0.027 ± 0.004 $\mu mol.min^{-1}.mg\ protein^{-1}$ when the medium was complete (see chapter 4). The Δ*prodh* parasites were unable to grow in the absence of glucose when PRODH was switched off by the addition of tetracycline. This result indicates that PRODH is not an essential enzyme for the parasites grown in complete medium. However, it is essential in the absence of glucose when proline is the sole energy souce.

Recently attention has focused on an understanding of metabolism in *T. brucei* PCF. A question as to the principal energy source of procyclics arises. In its natural environment, the parasite is reported to encounter high proline concentrations in the gut of the tsetse fly (Bursell, 1981; ter Kuile, 1997). However it is not clear that proline is the main energy source, Cross *et al.* (1975) reported that in rich culture medium, glucose is the main energy source and proline is used as a second substrate.

In order to analyse the growth of the parasites when PRODH was repressed, which should prevent *T. brucei* PCF from using proline as an energy source, a defined SDM79 medium without glucose and proline was used and complemented by proline and/or glucose at various concentrations. These experiments showed that without PRODH the parasites grow well in the presence of glucose but their ability to grow without glucose is abolished. PRODH proficient cells could grow using either glucose or proline.

van Weelden *et al.* (2003) showed that *T. brucei* PCF, grown in complete SDM79 medium, prefer glucose to proline as a substrate. The rate of degradation of glucose was three times higher than the rate of degradation of proline. In spite of a preference for glucose, proline could still be the main source of energy in the natural environment, as it is present at high concentrations in the tsetse fly and glucose availability may be limited. The down-regulation of PRODH reported in Chapter 4 and down-regulation of proline transport in the presence of glucose could explain the preference for glucose when present. ter Kuile, 1997, has previously reported that glucose metabolism is suppressed when proline is abundant.

van Weelden *et al.* (2003) also showed that the complete Krebs cycle is not used by *T. brucei* PCF. Some of the enzymes classically associated with the Krebs cycle are also involved in the conversion of proline to succinate. According to this scheme, energy is yielded from the Acetate-Succinate CoA transferase. According to recent data by Coustou

et al (2003), D-glucose and L-threonine are consumed in large quantities in PCF *T. brucei* grown in complete SDM79 medium; L-proline and L-glutamine were moderately consumed. The relevance of various amino acids and various metabolic intermediates, e.g. succinate, fumarate or pyruvate, that could potentially yield energy according to published schemes of metabolism on the growth of *T. brucei* PCF was assessed by growing Δ*prodh* parasites in a medium without glucose and proline supplemented with individual amino acids or metabolites. Of all substrates used, only proline and glucose seem to be able to sustain normal growth. None of the other amino acids or metabolites tested were able to be used as an energy source for the parasites. It is known than L-threonine could be used in lipid and other biosyntheses (Roberts *et al.*, 2003), which might explain why this amino acid is consumed in large quantities (Cross *et al.*, 1975; Coutou *et al.*, 2003) but apparently unable to yield energy.

Thus either glucose or proline is essential for the survival of *T. brucei* PCF. Bochud-Alleman and Schneider (2002) showed that succinyl CoA synthetase, found in the mitochondrion, is essential to the parasite. This could relate to the fact that this enzyme is crucial to ATP generation in pathways of both glucose and proline metabolism.

Glutamate is one of the degradation products of the proline metabolism pathway. The possibility that parasites, unable to use proline, may survive on glutamate was tested. Glutamate is not toxic for the cells, but it is unable to rescue the parasites from the loss of proline utilisation or the absence of glucose in the medium. A possible explanation for this could be due to lack of glutamate uptake by these cells.

In order to check if glutamate uptake occurs across the plasma membrane, uptake experiments were carried out using tritiated glutamate. Uptake of glutamate as a function of time showed that in both Δ*prodh* lines induced or non-induced by tetracycline, glutamate enters the cells. A plateau was reached after 10 min and there was accumulation

of glutamate inside the cells (6-fold increase of concentration, see Figure 6.11). Interestingly in PRODH suppressed cells, Michaelis-Menten type plotting revealed a saturable process at low concentration, indicating an apparent K_m in the order of 15 μM. An explanation for this difference is not immediately obvious. This might suggest that expression of siRNAs associated with PRODH could also have down-regulated other proteins involved in glutamate metabolism. A BLAST search using the PRODH sequence used for the RNAi construct was carried out, however no matches were found in the *T. brucei* database. Glutamate uptake, using relatively low concentrations of glutamate, was not robust when compared to uptake of a number of other amino acids. Excess of cold glutamate inhibited uptake of radiolabelled glutamate at low concentrations (190 nM) indicating a transport-mediated process. The inhibition profile suggested this was of high affinity (around 1.5 μM). However, using increasing concentrations of labelled glutamate, a plateau was not achieved indicating a substantial accumulation of glutamate via non-saturable pathways.

Rohloff *et al.*, 2003, showed that in *T. cruzi* there is an efflux of amino acids, during hypo-osmotic stress, which seems to occur through a low specificity anion channel. This channel allows the passage of several neutral and anionic amino acids including glutamate. One hypothesis to explain the observed glutamate uptake would be the presence of two or more glutamate uptake routes. One of these may allow the passage of this amino acid into the cells with high specificity. A second carrier, e.g. an anion channel as described in *T. cruzi* by Rohloff *et al.*, could allow passage of glutamate and ions with low specificity. This second carrier could then swamp the characteristics of the more specific transporter at higher concentration. Ion channel inhibitors such as DIDS (4,4'-diisothiocyanotodihydrostilbene-2,2'-disulfonic acid) at 10 μM, NPPB (5-nitro-2-(3-phenylpropylamino)benzoic acid) at 10 μM, furosemide at 1 mM, and glybenclamide at

0.5 mM were tested, but no difference in the uptake of glutamate could be observed (results non shown). This does not prove the lack of such a channel, but these compounds did not inhibit uptake and do not support the presence of such a channel.

Glutamate appeared to enter cells, via non-saturable routes, at a similar rate on a molar basis as proline when used at higher concentrations. However, even at 10 mM glutamate was unable to rescue the cells from the loss of proline utilisation, thus another explanation for this failure is required. Since glutamate metabolism occurs inside the mitochondrion, it is possible that glutamate cannot cross the mitochondrial membrane and enter the catabolic pathway leading to energy production. More experiments are needed to address this.

The transport of proline was also studied in Δ*prodh* cells in order to see if the loss of PRODH influenced the characteristics of proline uptake. Proline uptake was studied as a function of time (Figure 6.14). There was accumulation of proline over 30 min for both lines. After 30 min, accumulated proline for the non-induced cells was approximately 45 times higher than the external concentration. The intracellular concentration was also 30 times higher for the induced cells. L'Hostis *et al.*, 1993, observed a concentration of 100 fold after 1 min of transport using 0.1 mM as an external proline concentration. Law and Hukkada, 1979, observed an accumulation of 60 fold in *L. tropica* promastigotes after 10 min proline uptake at 30 °C. Proline uptake therefore appears to occur against a concentration gradient in trypanosomes and leishmania. However, after 30 min, intracellular concentrations of radiolabel diminished substantially for both lines. This was probably not due to the metabolism of proline, because proline is thought not to be metabolised in the tetracycline induced Δ*prodh* parasites (unless other pathways of metabolism not involving PRODH were present).

The kinetics of the transporter were studied using various concentrations of proline and a typical Michaelis-Menten type curve was obtained. Cells expressing PRODH or repressed in this enzyme shared the same K_m (~ 11 µM) but the V_{max} was different. It was significantly higher in the non-induced cells, 0.75 ± 0.06 nmol.min^{-1}.10^7 parasites^{-1} (- Tet) versus 0.41 ± 0.0035 nmol.min^{-1}.10^7 parasites^{-1} (+ Tet). It was not clear if this reduction related to the catabolite repression mechanism with regard to PRODH described earlier.

A difference of the volume between the parasites when RNAi was induced or non-induced by tetracycline could explain this difference in V_{max}. However, the volume of the cells was estimated to be the same, as no difference was noticed when viewed by light microscopy.

In summary, experiments reported in this chapter indicate that PCF *T. brucei* can use either glucose or proline as a carbon and energy source. They cannot, however, use other amino acids. Moreover it appears that when glucose is abundantly available, the parasites diminish their ability to utilise proline. This is apparently achieved by a post-translational down-regulation of proline dehydrogenase activity, and also down-regulation of proline transporter activity. This is interesting given that ter Kuile (1997) has previously shown that abundant proline (60 mM) stimulates a down-regulation of glucose metabolism. Thus proline and glucose usage appears to be mutually regulated in these cells. Expression of other genes, e.g. PARP (Roditi *et al.*, 1998), also appears to be sensitive to glucose concentration, although mitochondrial intermediates of metabolism also appear to impinge on these events (Vassella *et al.*, 2003). It will be of great interest to learn more about the mechanisms of metabolic regulation of protein expression in these cells.

Chapter 7

General Discussion

The overall aim of this study was to investigate proline metabolism and its role in the energy metabolism of PCF *T. brucei*. The project was comprised of two complementary approaches: a) cloning and expression of the proline dehydrogenase genes of *T. brucei* and *L. major*; b) knock down of the gene in *T. brucei* to determine the enzyme's role in energy metabolism.

Proline utilisation requires two enzymes to convert proline into glutamate, proline dehydrogenase (PRODH) and pyrroline-5-carboxylate dehydrogenase (P5CDH). In the first step, the oxidation of proline to pyrroline-5-carboxylate (P5C) is coupled with the reduction of a cofactor, flavin adenine dinucleotide (FAD). Two electrons from reduced FAD ($FADH_2$) are then transferred to the electron transport chain. In the second step, P5C is hydrolysed to give glutamic semialdehyde, which is oxidised to glutamate using the cofactor nicotinamide adenine dinucleotide (NAD^+) (Becker and Thomas, 2001; Lee Y *et al.*, 2003). Proline metabolism thus involves oxidation of proline to glutamate, which is frequently further metabolised to CO_2 through the Krebs cycle, with protons donated to the electron transport chain (Obungu *et al*, 1999). Recent work has shown that the PCF of *T. brucei* might lack a functional Krebs cycle. Knock down or knock out of succinate dehydrogenase, α-ketoglutarate dehydrogenase and aconitase showed that a complete Krebs cycle is not essential for the parasite (Bochud-Alleman and Schneider, 2002). The aconitase knock down had no effect on the growth of cells or energy production (van Weelden *et al.*, 2003). ATP production in PCF appears to be possible through various

pathways, including oxidative phosphorylation, substrate level phosphorylation and products of glycosomal metabolism.

During this study it was shown that proline metabolism is essential for the production of energy when the parasites were grown without glucose in the medium. The parasites were able to grow with only glucose or proline as an energy source in the medium, and they were able to switch metabolism of one of these substrates to the other. This may mimic what happens in the tsetse fly, where the concentration of glucose is variable according to the contents of the blood meal but the proline concentration is constant. Proline is metabolised to glutamate which is further metabolised to yield ATP, possibly through a pathway leading to the reaction catalysed by succinyl CoA synthetase (Coustou *et al.*, 2003).

In prokaryotes it is known that the *putA* gene encodes a single protein, called PutA, that contains both PRODH and P5CDH activities (Becker *et al*, 2001; Parkhill *et al*, 2001; Vinod *et al*, 2002). Proline utilisation requires another gene, *putP,* which encodes a proline permease permitting proline to enter the cell. These two genes, *putA* and *putP,* form an operon (Ling *et al*, 1994). Without proline, PutA acts as a transcriptional repressor, when proline is present the protein links to the membrane and is active (Muro-Pastor *et al.*, 1997).

In eukaryotes, the conversion of proline to glutamate takes place within the mitochondria (Small and Jones, 1990; Gogos *et al.*, 1999; Deuschle *et al.*, 2001). The first two steps of proline oxidation are carried out by two different polypeptides. PRODH is present in the inner membrane of the rat liver mitochondrion, whereas the next step, catalysed by P5CDH may take place within the mitochondrial matrix (Brunner and Neupert, 1969; Small and Jones, 1999).

The *L. major* and *T. brucei* PRODHs have homology to other PROHHs. By studying multiple alignments of many proline dehydrogenases, well-conserved motifs can be shown to be present within all sequences. The best-conserved domains are essential for the activity of the enzyme; representing the cofactor (FAD) and the substrate (proline) binding sites (Lee Y *et al.*, 2003). As in all other organisms, the *T. brucei* and *L. major* PRODH probably use FAD as a cofactor. Although activity of recombinant enzymes was not measured, extracts from *T. brucei* had PRODH activity. Moreover, key residues shown to bind FAD in the *E. coli* enzyme were conserved in the *T. brucei* enzyme (section 3.2.4). The *L. major* and *T. brucei* PRODHs were both predicted (using a computer program) to possess a mitochondrial signal peptide. The proteins are encoded by nuclear genes and synthesized in the cytosol, and then transported from the cytosol to the mitochondrion.

The mitochondrion in eukaryotes is generally believed to have originated from an endosymbiotic event (Andersson *et al.*, 2003). Thus it was of interest to study the phylogeny of proline dehydrogenase. The aim of this study was to see if this enzyme of trypanosomatids has an evolutionary link with the α-proteobacteria, the group of organisms from which the protomitochondrion is believed to have emerged (Gray *et al.*, 2001). The *L. major* and *T. brucei* PRODHs appear to be evolutionarily closer to eukaryote than to prokaryote enzymes. Moreover, in trypanosomatids, *PRODH* and *P5CDH* genes are situated in two different locations, thus these two enzymes are present on separated polypeptides. Thus supports the idea that *T. brucei* and *L. major* PRODH are closer to those of other eukaryotes, all of whom have separate PRODH and P5CDH, while all prokaryotes, including α-proteobacteria have fused enzymes. This phylogenetic analysis indicates that PRODH belongs to the 40 % of the mitochondrial proteome that was apparently contributed by the proto-eukaryotic member of the endosymbiotic union that gave rise to mitochondrion-containing eukaryotes (Kalberg *et al.*, 2000).

Having cloned genes for *T. brucei* and *L. major PRODH* orthologues, several experiments were made to overexpress the full trypanosomaatid proteins in a soluble form. As described in chapter 5, the pET21a$^+$ vector was used to overexpress these proteins. Soluble proteins were never obtained with this system. Two other systems (results not shown) were used in attempts to improve expression or solubilisation of these two proteins. The first of these was the Gateway system (GibcoBRL), where the gene of interest was cloned into an expression vector using phage-λ mediated recombination employs a phage T7 promoter to produce a recombinant protein containing a histidine tag at the N-terminus. The second system used was the Strep-tag (IBA). This system uses a tag at the C-terminus and involves the binding of biotin to streptavidin. Neither of these systems yielded an increase in the level of expression or improved the solubility of the protein. Since several vectors all failed to yield soluble, functional versions of the full length protein, we decided upon another approach, consisting of mutating the gene of interest to try to improve solubility. As described in Chapter 3, a putative mitochondrial targeting sequence and a putative trans-membrane domain were identified in the trypanosomatid predicted proteins. These fragments did not contain conserved motifs believed to be essential for the activity of the enzymes, or the motif for the binding of FAD (Lee Y *et al.*, 2003). It was hypothesised that loss of the hydrophobic putative trans-membrane domain would improve production of soluble protein, making it easier to study at the biochemical level. The *T. brucei* PRODH was expressed with or without these two segments (i.e. the putative mitochondrial targeting sequence and putative trans-membrane domain). The PRODH without the trans-membrane domain was found to be soluble in *E. coli* expressor cell lysates. However the increase in solubility was not accompanied by an activity in any of the conditions tested (Chapter 5). The reason for this is not clear. The N-terminal part of the protein might, after all, contain residues essential for activity or for the correct folding

of the protein. Another possibility relates to the use of the *E. coli* expression system, which is the easiest system with which to obtain recombinant protein. It is possible that eukaryotic-specific secondary modification could alter the protein's activity in trypanosomatids in a fashion not possible in *E. coli*. Other systems such as yeast or insect cells for over-expression could be tried in order to increase the chances of obtaining an active protein. Western blot analysis of crude lysate from *T. brucei* (shown in Chapter 4), revealed one band of around 60 kDa when the parasites were grown with or without glucose. Another blot (shown in Chapter 6) revealed that parasites, not induced by tetracycline to induce expression of double-stranded RNA had two bands of around 60 and 45 kDa. The first band could correspond to the entire protein and the second to a truncated protein corresponding to a degradation product or specific cleavage product. It is possible that this cleaved protein is essential for the activity of the enzyme and cannot be obtained with the *E. coli* recombinant system.

The *T. brucei* PRODH expressed in its entirety also failed to show activity after resolubilisation from inclusion bodies. It cannot be ruled out that the presence of the hexa-his tag at the protein C-terminus could affect activity. For example, if this region of the protein was involved in dimer formation this could affect activity. Efforts to express the protein without a his-tag or an N-terminal tag would be an interesting topic for further study, but such efforts were not made here. Thus efforts to achieve functional expression of *T. brucei* or *L. major* PRODH failed. However, the fact that the *T. brucei* gene decribed here encodes a bona fide PRODH was demonstrated by knocking down its expression by RNA interference. Since no other genes with substantial homology to that reported here could be found in the *T. brucei* genome, it is unlikely that an alternative PRODH gene was knocked down by RNAi using the constructs described in Chapter 6.

As discussed in Chapter 6, PCF *T. brucei* were able to grow in a medium containing either proline or glucose as an energy source (Evans and Brown, 1972). Moreover ter Kuile, 1997, showed that glucose is used preferentially when at high concentration and proline preferentially, when this amino acid is at high concentration. Some kind of catabolite repression thus may operate in PCF *T. brucei*. It was therefore of interest to investigate whether there is regulation of PRODH activity in response to different energy sources and at which level the regulation occurs.

Catabolite repression is a well-known phenomenon in prokaryotes (Bruckner and Titgemeyer, 2002) and yeast (Gancedo, 1998). The measurement of PRODH activity in crude lysates of *T. brucei*, grown in standard SDM79 medium, was based on assay used by Obungu *et al.*, 1999, who studied this activity in *T. congolense*. PRODH activity could not be identified in *T. brucei* using the same conditions described by Obungu *et al.*, where the parasites were resuspended in PBS containing 10 mM proline and 0.1 % Triton X-100. The reason for the failure of the detection of PRODH activity was not clear. Therefore, it was decided to investigate whether PRODH activity was somehow regulated by growth conditions. When grown in complete medium containing glucose and proline, it has previously been shown that the parasites preferentially use glucose (Coustou *et al.*, 2003; van Weelden *et al*, 2003). During my project, it was demonstrated that *T. brucei* PCF were able to grow in a medium lacking glucose but containing proline as the only energy source. By testing the activity of the PRODH enzyme in crude lysates of *T. brucei* grown in these two conditions, it was noticed that PRODH activity was around twice as high in cells grown in a medium lacking glucose. When parasites were grown in complete SDM79 medium containing both glucose and proline, the activity of PRODH was around 0.027 $\mu mol.min^{-1}.mg\ protein^{-1}$ and when the medium was lacking of glucose the specific activity was reaching a maximum of $0.054 \pm 0.004\ \mu mol.min^{-1}.mg\ protein^{-1}$. When glucose was

absent from the medium the level of transcription of *PRODII* gene was unchanged and no change in RNA stability was apparent as judged by Northern blot analysis. While transcriptional regulation of genes is unusual in trypanosomes (Vanhamme and Pays, 1995), post transcriptional control at various levels such as RNA stability (Graham, 1995) has been described.

Western blot experiments performed with *T. brucei* procyclics revealed the presence of the protein in parasites grown in the presence or absence of glucose in the medium. The presence of glucose in the medium therefore led to decreased ability to use proline by the first enzyme of the proline catabolism pathway. The level of expression however, was not notably changed by the presence of proline. The mechanism by which activity is repressed is therefore yet to be identified. Secondary modifications, protein processing, or allosteric regulation are possibilities.

The transport of proline was also studied in Δ*prodh* cells in order to see if the loss of PRODH influenced the characteristics of proline uptake. The kinetics of the transporter was studied using various concentrations of proline and a typical Michaelis-Menten type curve was obtained. The transport in cells expressing PRODH or repressed in this enzyme shared the same K_m for proline (~ 11 μM) but the V_{max} was different. It was significantly higher in the non-induced cells, 0.75 ± 0.06 nmol.min^{-1}.10^7 parasites^{-1} (- Tet) versus 0.41 ± 0.0035 nmol.min^{-1}.10^7 parasites^{-1} (+ Tet). It was not clear if this reduction related to the catabolite repression mechanism or another phenomenon occurring at the level of the transporter (such as a change of the copy number of the transporter) was involved. An earlier study on proline transport also revealed a similar K_m (19 μM) and a V_{max} of 1.7 nmol.min^{-1}.10^7 parasites^{-1} at 27 °C (L'Hostis *et al.*, 1993). Since a similar difference was identified in experiments measuring uptake over 30 s and 3 s, it seems probable that the

difference is due to changes at the transporter itself rather than at the level of uptake (involving transport and metabolism by PROD11).

Evans and Brown, 1972, using Pittman's medium (Pittam, 1970) showed that proline was used preferentially to glucose (although it has been not possible to establish concentrations of substrates in this medium). ter Kuile, 1997, showed using chemostat studies and different media that a medium containing a high concentration of proline (60 mM) led to repression of glucose metabolism, he also showed that while proline is used (when present at 0.5 mM) in medium with glucose at 5 mM, in the same medium with glucose at 40 mM, proline use was repressed. The present study showed that PRODH activity and proline transporter is higher in cells where glucose is absent, than where glucose is present at 10 mM.

This observation is consistent with that of other authors, i.e. proline catabolism is repressed as glucose is more abundant. Morover, the data of both Evans and Brown, 1972 and ter Kuile, 1997 suggest that glucose catabolism is repressed with high proline. The precise calibration of the levels of glucose that repress proline metabolism, and *vice versa* is an interesting question and wortly of further study.

In order to improve our understanding of PCF *T. brucei* energy metabolism in general and specifically the roles of proline dehydrogenase in this process, the technique of RNA interference was applied to knock down expression of the enzyme. The name given for the parasites derived of RNAi experiments is Δ*prodh*. The Δ*prodh* parasites were unable to grow in the absence of glucose when PRODH was switched off by the addition of tetracycline. Moreover, the PRODH transcript was shown by Northern blot to be selectively knocked down, the protein was absent as judged by Western blotting and activity could not be detected in the knock down cells as assessed using the PRODH assay. These results indicate that the RNAi construct is efficient at selectively knocking down

PRODH and that this enzyme is not essential for the parasites grown in complete medium. However, it is essential in the absence of glucose when proline is the sole energy source.

To further study the growth behaviour of the parasites when PRODH was repressed, a defined SDM79 medium without glucose and proline was used and complemented by proline and/or glucose at various concentrations. These experiments showed that without PRODH the parasites grow well in the presence of glucose but their ability to grow without glucose is abolished. Cells having PRODH could grow using either glucose or proline. Various amino acids and substrates were used to try to complement the loss of proline utilisation; only proline and glucose seemed to be able to sustain normal growth. None of the other amino acids, or metabolites, tested were able to be used as an energy source for the parasites. Pyruvate was tested because it is one of the end products of glucose metabolism and it enters the mitochondrion. In many cells types, pyruvate dehydrogenase then commits pyruvate to energy production via the Krebs cycle. However, the fact that the Krebs cycle is apparently not involved in energy production in these cells could explain its failure to restore growth to cells deprived of glucose and proline. Glutamate, which is one of the degradation products of the proline catabolism pathway, was tested as an amino acid with the potential to complement growth. Glutamate is unable to rescue the parasites from the loss of proline utilisation or the absence of glucose in the medium. Since one explanation of glutamate's failure to support growth would be an inability to bring this metabolite into the cell, uptake of glutamate as a function of time was studied. In both Δ*prodh* lines induced or non-induced by tetracycline, glutamate entered the cells. Glutamate transport was not robust, but at 10 mM similar levels of glutamate as proline enter cells (Chapter 6). Glutamate therefore appeared to enter cells at 10 mM, but was unable to rescue the cells from the loss of proline utilisation. A simple explanation for this failure is not evident. Glutamate metabolism occurs inside the mitochondrion. It is

possible that glutamate cannot cross the mitochondrial membrane and enter the catabolic pathway leading to energy production. However more experiments are needed to address this. Other amino acids, particularly threonine have been shown to be taken up in substantial quantities by the parasites (Cross *et al.*, 1975; Coustou *et al.*, 2003). In this work, threonine was not able to rescue the cells from the loss of proline utilisation in a medium without glucose. It is possible that threonine is not an energy substrate but a substrate for lipid and for other biosynthesis (Roberts *et al.*, 2003). PRODH seems to be essential to trypanosomes in a medium without glucose. The activity of PRODH and also the proline transporter are higher in cells grown where glucose is absent. This observation is consistent with that of ter Kuile 1997, who showed proline catabolism to be repressed, as glucose is more abundant. In this thesis, I have shown that proline dehydrogenase allows the degradation of proline to yield an energy source for PCF trypanosomes. This observation is consistent with that of van Weelden *et al.*, 2003, where NMR analysis revealed that the secreted end product of proline is predominantly succinate, and according Coustou *et al.*, 2003, proline is degraded into succinate by the succinylCoA synthetase, giving ATP to the cells.

In summary, trypanosomatid PRODH is a typical eukaryotic enzyme active in mitochondria. Its activity in trypanosomes appears to be regulated in response to glucose levels and those glucose and proline metabolisms are mutually regulated for energy production. In a medium without glucose, only proline, and not other amino acids, allowed growth of the parasites. Proline is metabolised to glutamate, which can be further metabolised to succinate via α-ketoglutarate and succinyl co-enzyme with ATP production, via succinyl coA synthetase. It would seem that this pathway is essential to ATP production in absence of glucose.

References

A

Abrahamson, J.L., Baker, L.G., Stephenson, J.T. & Wood, J.M. **1983**, "Proline dehydrogenase from *Escherichia coli* K12. Properties of the membrane-associated enzyme", *Eur. J. Biochem.*, 134, 77-82.

Agami, R. & Shapira, M. **1992**, "Nucleotide sequence of the spliced leader RNA gene from *Leishmania mexicana amazonensis*", *Nucleic Acids Res.*, 20, 1804.

Allemann, N. & Schneider, A. **2000**, "ATP production in isolated mitochondria of procyclic *Trypanosoma brucei*", *Mol. Biochem. Parasitol.*, 111, 87-94.

Amato V., Amato J., Nicodemo A., Uip D., Amato-Neto V. & Duarte M. **1998** "Treatment of mucocutaneous leishmaniasis with pentamidine isethionate", *Ann. Dermatol. Venereol.*, 125, 492-495.

Andersson, S.G., Karlberg, O., Canback, B. & Kurland, C.G. **2003**, "On the origin of mitochondria: a genomics perspective", *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 358, 165-177.

B

Bailey T.L.& Elkan C. **1994**, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers" *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 2, 28-36.

Bakker, B.M., Walsh, M.C., ter Kuile, B.H., Mensonides, F.I., Michels, P.A., Opperdoes, F.R. & Westerhoff, H.V. **1999**, "Contribution of glucose transport to the control of the glycolytic flux in *Trypanosoma brucei*", *Proc. Natl. Acad. Sci. U.S.A*, 96, 10098-10103.

Balogun R. **1974**, "Studies on the amino acids of the tsetse fly, *glossina morsitans*, maintained on *in vitro* and *in vivo* feeding systems", *Comp. Biochem. Physiol. A*, 49, 215-222.

Barnard, J.P. & Pedersen, P.L. **1994**, "Alteration of pyruvate metabolism in African trypanosomes during differentiation from bloodstream into insect forms", *Arch. Biochem. Biophys.*, 313, 77-82.

Barrett, M.P, Mottram, J.C. & Coombs G.H. **1999**," Recent advances in identifying and validating drug targets in trypanosomes and leishmanias", *Trends Microbiol.*, 7, 82-88.

Barry, J.D. & McCulloch, R. **2001**, "Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite", *Adv. Parasitol.*, 49, 1-70.

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

Bates, P.A. **1994**, "Complete developmental cycle of *Leishmania mexicana* in axenic culture", *Parasitol.*, 108, 1-9.

Bates, P.A. & Tetley, L. **1993**, "*Leishmania mexicana*: induction of metacyclogenesis by cultivation of promastigotes at acidic pH", *Exp. Parasitol.*, 76, 412-423.

Becker, D.F. & Thomas, E.A. **2001**, "Redox properties of the PutA protein from *Escherichia coli* and the influence of the flavin redox state on PutA-DNA interactions", *Biochemistry*, 40, 4714-4721.

Benne, R., Van den Burg, J., Brakenhoff, J.P., Sloof, P., Van Boom, J.H. & Tromp, M.C. **1986,** "Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA" *Cell*, 46, 819-26.

Berens R.L., Brun R. & Krassner S.M. **1976**, "A simple monophasic medium for axenic culture of hemoflagellates" *J. Parasitol.* 62, 360-365.

Besteiro, S., Biran, M., Biteau, N., Coustou, V., Baltz, T., Canioni, P. & Bringaud, F. **2002**, "Succinate secreted by *Trypanosoma brucei* is produced by a novel and unique glycosomal enzyme, NADH-dependent fumarate reductase", *J. Biol. Chem.*, 277, 38001-38012.

Beverley S.M., **2003**, "Protozomics: trypanosomatid parasite genetics comes of age." *Nat. Rev. Genet.* 4, 11-29.

Bitonti, A.J., Dumont, J.A. & McCann, P.P. **1986**, "Characterization of *Trypanosoma brucei brucei* S-adenosyl-L-methionine decarboxylase and its inhibition by Berenil, pentamidine and methylglyoxal bis(guanylhydrazone)", *Biochem. J.*, 237, 685-689.

Blum, B., Bakalara, N. & Simpson, L. **1990**, "A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information", *Cell*, 60, 189-198.

Bochud-Allemann, N. & Schneider, A. **2002**, "Mitochondrial substrate level phosphorylation is essential for growth of procyclic *Trypanosoma brucei*", *J. Biol. Chem.*, 277, 32849-32854.

Bohringer, S. & Hecker, H. **1975**, "Quantitative ultrastructural investigations of the life cycle of *Trypanosoma brucei*: a morphometric analysis", *J. Protozool.*, 22, 463-467.

Borst, P., Bitter, W., Blundell, P.A., Chaves, I., Cross, M., Gerrits, H., van Leeuwen, F., McCulloch, R., Taylor, M. & Rudenko, G. **1998**, "Control of VSG gene expression sites in *Trypanosoma brucei*", *Mol. Biochem. Parasitol.*, 91, 67-76.

Boucher, N., Wu, Y., Dumas, C., Dube, M., Sereno, D., Breton, M. & Papadopoulou, B. **2002**, "A common mechanism of stage-regulated gene expression in *Leishmania* mediated by a conserved 3'-untranslated region element", *J. Biol. Chem.*, 277, 19511-19520.

Bradford, M.M. **1976**, "A rapid and sensitive method for the quantification of microgram quantities of protein utilising the principle of protein-dye binding", *Anal. Biochem.*, 72, 248-254.

Brandriss, M.C. **1983**, "Proline utilization in *Saccharomyces cerevisiae*: analysis of the cloned PUT2 gene", *Mol. Cell Biol.*, 3, 1846-1856.

Bray, P.G., Barrett, M.P., Ward, S.A. & de Koning, H.P. **2003**, "Pentamidine uptake and resistance in pathogenic protozoa: past, present and future", *Trends Parasitol.*, 19, 232-239.

Brown, E.D. & Wood, J.M. **1992**, "Redesigned purification yields a fully functional PutA protein dimer from *Escherichia coli*", *J. Biol. Chem.*, 267, 13086-13092.

Brown, E.D. & Wood, J.M. **1993**, "Conformational change and membrane association of the PutA protein are coincident with reduction of its FAD cofactor by proline", *J. Biol. Chem.*, 268, 8972-8979.

Bruckner, R. & Titgemeyer, F. **2002**, "Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization", *FEMS Microbiol. Lett.*, 209, 141-148.

Brun, R. & Schonenberger. **1979**, "Cultivation and in vitro cloning of procyclic culture forms *Trypanosoma brucei* in a semi defined medium", *Acta Trop.*, 36, 289-292.

Brunner, G. & Neupert, W. **1969**, "Localisation of proline oxidase and Delta-pyrroline-5-carboxylic acid dehydrogenase in rat liver", *FEBS Lett.*, 3, 283-286.

Bursell, E. **1967**, "The conversion of glutamate to alanine in Tsetse fly *Glossina moritans*", *Comp. Biochem. Physiol. B*, 23, 825-829.

Bursell, E. **1973**, "Entomological aspects of the epidemiology of sleeping sickness", *Cent. Afr. J. Med.*, 19, 201-204

Bursell, E., Billing, K.J., Hargrove, J.W., McCabe, C.T. & Slack, E. **1973**, "The supply of substrates to the flight muscle of tsetse flies", *Trans. R. Soc. Trop. Med. Hyg.*, 67, 296.

Bursell, E. **1978**, "quantitative aspects of proline utilization during flight in tsetse flies", *physiol. entomology*, 3, 265-272.

Bursell, E., **1981**, " Role of proline in Energy metabolism" *plenum NY*, 135-157

C

Campbell H.D., Webb G.C. & Young I.G. **1997**, "A human homologue of the *Drosophila melanogaster* sluggish-A (proline oxidase) gene maps to 22q11.2, and is a candidate gene for type-I hyperprolinaemia", *Human Genet.*, 101, 69-74.

Campbell, D.A., Sturm, N.R. & Yu, M.C. **2000**, "Transcription of the kinetoplastid spliced leader RNA gene", *Parasitol. Today*, 16, 78-82.

Carter, N.S., Barrett, M.P. & de Koning, H.P. **1999**. "A drug resistance determinant in *Trypanosoma brucei*", *Trends Microbiol.*, 7, 469-471.

Cavalier-Smith T. **1987**, "The origin of eukaryotic and archaebacterial cells." *Ann. NY Acad. Sci.*, 503, 17-54.

Cazzulo J.J. **1992**, "Aerobic fermentation of glucose by Trypanosomatids", *The FASEB journal*, 6, 3153-3161.

Charest, H., Zhang, W.W. & Matlashewski, G. **1996**, "The developmental expression of *Leishmania donovani* A2 amastigote-specific genes is post-transcriptionally mediated and involves elements located in the 3'-untranslated region" *J. Biol. Chem.*, 271, 17081-90.

Clarkson, A.B. Jr, Bienen, E.J., Pollakis, G. & Grady, R.W. **1989**. "Respiration of bloodstream forms of the parasite *Trypanosoma brucei brucei* is dependent on a plant-like alternative oxidase" *J. Biol. Chem.*, 264, 17770-17776.

Claros, M.G. & Vincens, P. **1996**, "Computational method to predict mitochondrially imported proteins and their targeting sequences" *Eur. J. Biochem.*, 241, 779-86.

Clayton, C.E. & Michels, P. **1996**, "Metabolic compartmentation in African trypanosomes", *Parasitol. Today*, 12, 465-471.

Clayton C.E. **1999**, "Genetic manipulation of kinetoplastida", *Parasitol. Today*, 15, 372-378.

Clayton, C.E. **2002**, "Life without transcriptional control? From fly to man and back again", *EMBO J.*, 21, 1881-1888.

Coppens, I. & Courtoy, P.J. **2000**, "The adaptative mechanisms of *Trypanosoma brucei* for sterol homeostasis in its different life-cycle environments", *Annu. Rev. Microbiol.*, 54, 129-156.

Coombs, G.H., Tetley, L., Moss, V.A. & Vickerman, K. **1986**, "Three dimensional structure of the *Leishmania* amastigote as revealed by computer-aided reconstruction from serial sections", *Parasitology*, 92, 13-23.

Coustou, V., Besteiro, S., Biran, M., Diolez, P., Bouchaud, V., Voisin, P., Michels, P.A., Canioni, P., Baltz, T. & Bringaud F. **2003** "ATP generation in the *Trypanosoma brucei* procyclic form: Cytosolic substrate level phosphorylation is essential, but not oxidative phosphorylation." *J. Biol. Chem.*, 278, 49625-35.

Cross, G.A., Klein, R.A. & Linstead, D.J. **1975**, "Utilization of amino acids by *Trypanosoma brucei* in culture: L-threonine as a precursor for acetate", *Parasitology*, 71, 311-326.

D

Daneshvar, H., Coombs, G.H., Hagan, P. & Phillips, R.S. **2003**, "*Leishmania mexicana* and *Leishmania major*: attenuation of wild-type parasites and vaccination with the attenuated lines" *J. Infect. Dis.*, 187, 1662-1668.

Day, D.A. & Tuite, M.F. **1998**, "Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview", *J. Endocrinol.*, 157, 361-371.

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.

Denise, H., Giroud, C., Barrett, M.P. & 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.

Deuschle, K., Funck, D., Hellmann, H., Daschner, K., Binder, S. & Frommer, W.B. **2001**, "A nuclear gene encoding mitochondrial Delta-pyrroline-5-carboxylate dehydrogenase and its potential role in protection from proline toxicity", *Plant J.*, 27, 345-356.

Docampo, R. & Moreno, S.N. **2001**, "The acidocalcisome." *Mol. Biochem. Parasitol.*, 114, 151-159.

Durieux, P.O., Schutz, P., Brun, R. & Kohler P., **1991**, "Alterations in Krebs cycle enzyme activities and carbohydrate catabolism in two strains of

Trypanosoma brucei during in vitro differentiation of their bloodstream to procyclic stages", *Mol. Biochem. Parasitol.*, 45, 19-27.

E

Edwards, K.J., Jenkins, T.C. & Neidle, S. **1992**, "Crystal structure of a pentamidine-oligonucleotide complex: implications for DNA-binding properties", *Biochemistry*, 31, 7104-7109.

Eisenthal, R. & Cornish-Bowden, A. **1998**, "Prospects for antiparasitic drugs. The case of *Trypanosoma brucei*, the causative agent of African sleeping sickness", *J. Biol. Chem.*, 273, 5500-5505.

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

Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. **2000**, "Predicting subcellular localization of proteins based on their N-terminal amino acid sequence", *J. Mol. Biol.*, 300, 1005-1016.

Enanga, B., Burchmore, R.J., Stewart, M.L. & Barrett, M.P. **2002**, "Sleeping sickness and the brain", *Cell Mol. Life Sci.*, 59, 845-858.

Evans, D.A. & Brown, R.C. **1972**, "The utilization of glucose and proline by culture forms of *Trypanosoma brucei*", *J. Protozool.*, 19, 686-690.

F

Fairlamb, A.H. **1989**, "Novel biochemical pathways in parasitic protozoa", *Parasitology*, 99, 93-112.

Ford, W.C. & Bowman, I.B. **1973**, "Metabolism of proline by the culture midgut form of *Trypanosoma rhodesiense*", *Trans. R. Soc. Trop. Med. Hyg.*, 67, 257.

Furuya, T., Kessler, P., Jardim, A., Schnaufer, A., Crudder, C. & Parsons, M. **2002**, "Glucose is toxic to glycosome-deficient trypanosomes", *Proc. Natl. Acad. Sci. U.S.A*, 99, 14177-14182.

G

Gancedo, J. **1998**, " Yeast carbon catabolite repression" *Microbiol. Mol. Biol. Rev.*, 62, 334-361.

Gibson, W. **2002**, "Epidemiology and diagnosis of African trypanosomiasis using DNA probes", *Trans. R. Soc. Trop. Med. Hyg.*, 96, 141-143.

Glaser, T.A. & Hukkada, A.J. **1992**, "Proline transport in *Leishmania donovani* amastigotes: dependence on pH gradients and membrane potential", *Mol. Biochem. Parasitol.*, 51, 1-8.

Gogos, J.A., Santha, M., Tokac, Z., Beck, K.D., Luine, V., Lucas, L.R., Nadler, J.V. & Karayiorgou, M. **1999**, "The gene encoding proline dehydrogenase modulates sensimotor gating in mice", *Nat. genet.*, 21, 434-439.

Graham, S.B., Stephensen, J.T. & Wood, J.M. **1984**, "Proline dehydrogenase from *Escherichia coli* K 12", *J. Biol. Chem.*, 259, 2656-2661.

Graham, S.V. **1995**, "Mechanisms of stage regulated gene expression in Kinetoplastida", *Parasitol. Today*, 11, 217-223.

Gray, M.W., Burger, G. & Lang, F.B. **2001**, "The origin and early evolution of mitochondria", *Genome Biology*, 2, 1018.1-1018.5.

Gutteridge, W.E. & Coombs, G.H. **1977**, Biochemistry of parasitic protozoa, *The Macmillan press LTD edn.*

H

Hamm, B., Schindler, A., Mecke, D. & Duszenka, M. **1990**, "Differentiation of *Trypanosoma brucei* bloodstream trypomastigotes from long slender to short stumpy like forms in axenic culture", *Mol. Biochem. Parasitol.*, 40, 13-22.

Handman, E. **1997**, "*Leishmania* vaccines: old and new", *Parasitol., Today*, 13, 236-238.

Hannon, G.J. **2002**, "RNA interference", *Nature,* 418, 244-51.

Harrison, P.J. & Owen, M.J. **2003**, "Genes for schizophrenia? Recent findings and their pathophysiological implications", *Lancet*, 361, 417-149.

Hart, D.T., Vickerman, K. & Coombs, G.H., **1981**, "A quick, simple method for purifying *Leishmania mexicana* amastigotes in large number." *Parasitology*, 82, 345-355.

Hauser, R., Pypaert, M., Hausler, T., Horn, E. K. & Schneider, A. **1996**, "*In vitro* import of proteins into mitochondria of *Trypanosoma brucei* and *Leishmania tarentolae*", *J. Cell Sci.*, 109, 517-523.

Hausler, T., Stierhof, Y.D., Blattner, J. & Clayton, C. **1997**, "Conservation of mitochondrial targeting sequence function in mitochondrial and hydrogenosomal proteins from the early-branching eukaryotes *Crithidia*, *Trypanosoma* and *Trichomonas*", *Eur. J. Cell Biol.*, 73, 240-251.

Hayward, D.C., Delaney, S.J., Campbell, H.D., Ghysen, A., benzer, S., Kasprzak, A.B., Cotsell, J.N., Young, I.G. & Miklos, G.L. **1993**, "The sluggish A gene of *Drosophila melanogaster* is expressed in the nervous system and encodes proline oxidase, a mitochondrial enzyme involved in glutamate biosybthesis", *Proc. Natl. Acad. Sci. U.S.A*, 90, 2979-2983.

Herwaldt, B.L. **1999**, "Leishmaniasis", *Lancet*, 354, 1191-1199.

Hide, G. **1999**, "History of sleeping sickness in East Africa", *Clin. Microbiol. Rev.*, 12, 112-125.

Hirumi, H., Doyle, J.J. & Hirumi, K. **1977**, "African trypanosomes: cultivation of animal-infective *Trypanosoma brucei in vitro*", *Science*, 196, 992-994.

Hirumi, H. & Hirumi, K. **1994**, "Axenic culture of African Trypanosome bloodstream forms", *Parasitol. Today*, 10,

Hofmann, K. & Stoffel, W. **1992**, "PROFILEGRAPH: an interactive graphical tool for protein sequence analysis." *Comput. Appl. Biosci.*, 8, 331-337.

Hutvagner, G. & Zamore, P.D. **2002**, "RNAi: nature abhors a double-strand", *Curr. Opin. Genet. Dev.*, 12, 225-232.

J

Jakson, P. & Fisher, F. **1977**, " Carbohydrate effects on amino acid transport by Trypanosoma equiperdum", *J. Protozool.*, 24, 345-353.

Jacquet, H., Berthelot, J., Bonnemains, C., Simard, G., Saugier-Veber, P., Raux, G., Campion, D., Bonneau, D. & Frebourg, T. **2003**, "The severe form of type I hyperprolinaemia results from homozygous inactivation of the PRODH gene", *J. Med. Genet.*, 40, 7.

Johnson, A.B. & Strecter, H.J. **1962**, "The interconversion of glutamic acid and proline", *J. Biol. Chem.*, 237,

K

Karlberg, O., Canback, B., Kurland, C.G. & Andersson, S.G. **2000**, "The dual origin of the yeast mitochondrial proteome", *Yeast*, 17, 170-187.

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

Keuntje, B., Masepohl, B. & Klipp, W. **1995**, "Expression of the putA gene encoding proline dehydrogenase from *Rhodobacter capsulatus* is independent of NtrC regulation but requires an Lrp-like activator protein", *J. Bacteriol.*, 177, 6432-6439.

Kiyosue, T., Yoshiba, Y., Yamaguchi-Shinozaki, K. & Shinozaki, K. **1996**, "A nuclear gene encoding mitochondrial proline dehydrogenase, an enzyme involved in proline metabolism, is upregulated by proline but downregulated by dehydration in *Arabidopsis*", *Plant Cell*, 8, 1323-1335.

Knodler, L.A., Schofield, P.J. & Edwards, M.R. **1992**, "Glucose transport in *Crithidia luciliae*", *Mol. Biochem. Parasitol.*, 56, 1-13.

koehlor, C.M. **2000**, "protein translocation pathways of the mitochondria", *FEBS letters*, 476, 27-31.

Kredich, N.M. & Tomkins, G.M. **1966**, " The enzymatic synthesis of L-cysteine in *E. coli* and *S. typhimurium*", *J. Biol. Chem.*, 241, 4955-4965.

Kumar, S., Tamura, K., Jakobsen, I.B. & Nei, M. **2001**, "MEGA2: molecular evolutionary genetics analysis software", *Bioinformatics*, 17, 1244-5.

Kuzoe, F.A.S. **1993**, "Current situation of African trypanosomiasis", *Acta Trop.*, 54, 153-162.

L

Laemmli, U.K. **1970**, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4" *Nature*, 227, 572-574.

LaCount, D.J., Bruse, S., Hill, K.L. & Donelson, J.E. **2000**, "Double-stranded RNA interference in *Trypanosoma brucei* using head-to-head promoters", *Mol. Biochem. Parasitol.*, 111, 67-76.

Law, S.S. & Hukkada, A.J. **1979**, "Transport of L-proline and its regulation in *Leishmania tropica* promastigotes", *J. Protozool.*, 26, 295-301.

Lee, J.H., Park, N.Y., Lee, M.H. & Choi, S.H. **2003**, "Characterization of the *Vibrio vulnificus* putAP operon, encoding proline dehydrogenase and proline permease, and its differential expression in response to osmotic stress", *J. Bacteriol.*, 185, 3842-3852.

Lee, Y.H., Nadaraia, S., Gu, D., Becker, D.F. & Tanner, J.J. **2003**, "Structure of the proline dehydrogenase domain of the multifunctional PutA flavoprotein", *Nat. Struct. Biol.*, 10, 109-114.

L'Hostis, C., Geindre, M., & Deshusses, J. **1993**, "Active transport of L-proline in the protozoan parasite *Trypanosoma brucei brucei*", *Biochem. J.*, 291, 297-301.

Liang, X.H., Liu, Q. & Michaeli S. **2003**, "Small nucleolar RNA interference induced by antisense or double-stranded RNA in trypanosomatids", *Proc. Natl. Acad. Sci. U.S.A.*, 100, 7521-7526.

Ling, M., Allen, S.W. & Wood, J.M. **1994**, "Sequence analysis identifies the proline dehydrogenase and pyrroline-5-caboxylate dehydrogenase domains of the multifunctional *Escherichia coli* Put A protein", *J. Mol. Biol.*, 243, 950-956.

M

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. & Tschudi, C. **2000**, "A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA" *RNA*, 6, 163-169.

Mani, S., Van De, C.B., Van Montagu, M. & Verbruggen, N. **2002**, "Altered levels of proline dehydrogenase cause hypersensitivity to proline and its analogs in *Arabidopsis*", *Plant Physiol.*, 128, 73-83.

Matthews, K.R. **1999**, "Developments in the differentiation of *Trypanosoma brucei*", *Parasitol. Today*, 15, 76-80.

Mazareb, S., Fu, Z.Y. & Zilberstein, D. **1999**, "Developmental regulation of proline transport in *Leishmania donovani*", *Exp. Parasitol.*, 91, 341-348.

Maxwell, S.A. & Davis, G.E. **2000**, "Differential gene expression in p53-mediated apoptosis-resistant vs. apoptosis-sensitive tumor cell lines", *Proc. Natl. Acad. Sci. U.S.A*, 97, 13009-13014.

McCann, P.P., Bacchi, C.J., Clarkson, A.B., Jr., Bey, P., Sjoerdsma, A., Schecter, P.J., Walzer, P.D., & Barlow, J.L. **1986**, "Inhibition of polyamine biosynthesis by alpha-difluoromethylornithine in African trypanosomes and *Pneumocystis carinii* as a basis of chemotherapy: biochemical and clinical aspects", *Am. J. Trop. Med. Hyg.*, 35, 1153-1156.

Medina-Acosta, E. & Cross, G.A. **1993**, "Rapid isolation of DNA from trypanosomatid protozoa using a simple 'mini-prep' procedure", *Mol. Biochem. Parasitol.*, 59, 327-329.

Menzel, R. & Roth, J. **1981a**, "Enzymatic properties of the purified putA protein from *Salmonella typhimurium*", *J. Biol. Chem.*, 256, 9762-9766.

Menzel, R. & Roth, J. **1981b**, "Purification of the putA gene product. A bifunctional membrane-bound protein from *Salmonella typhimurium* responsible for the two-step oxidation of proline to glutamate", *J. Biol. Chem.*, 256, 9755-9761.

Michels, P.A.M., Hannaert, V. & Bringaud, F. **2000**, "Metabolic aspects of glycosomes in *Trypanosomatidae* - New data and views", *Parasitol. Today*, 16, 482-489.

Momen, H. **2001**, "Some current problems in the systematics of *Trypanosomatids*", *Int. J. Parasitol.*, 31, 640-642.

Monticello, D.J. & Costilow, R.N. **1981**, "Purification and partial characterization of proline dehydrogenase from *Clostridium sporogenes*", *Can. J. Microbiol.*, 27, 942-948.

Morris, J.C., Wang, Z., Drew, M.E. & Englund, P.T., **2002**, "Glycolysis modulates trypanosome glycoprotein expression as revealed by an RNAi library", *EMBO J.* 21, 4429-4438.

Mowatt, M.R. & Clayton, C.E. **1988**, "Polymorphism in the procyclic acidic repetitive protein gene family of *Trypanosoma brucei*", *Mol. Cell Biol.*, 8, 4055-4062.

Muller, R. & Baker, J.R. **1990**, *Medical Parasitology*, Grower Medical Publishing ed.

Muro-Pastor, A.M. & Maloy S. **1995**, "Proline dehydrogenase activity of the transcriptional repressor Put A is required for induction of the Put operon by proline", *J. Biol. Chem.*, 270, 9819-9827.

Muro-Pastor, A.M., Ostrovsky, P. & Maloy, S. **1997**, "Regulation of gene expression by repressor localization: biochemical evidence that membrane and DNA binding by the PutA protein are mutually exclusive", *J. Bacteriol.*, 179, 2788-2791.

N

Nadaraia, S., Lee, Y.H., Becker, D.F. & Tanner, J.J. **2001**, "Crystallization and preliminary crystallographic analysis of the proline dehydrogenase domain of the multifunctional PutA flavoprotein from *Escherichia coli*", *Acta Crystallogr. D. Biol. Crystallogr.*, 57, 1925-1927.

Nakashima, K., Satoh, R., Kiyosue, T., Yamaguchi-Shinozaki, K. & Shinozaki, K. **1998**, "A gene encoding proline dehydrogenase is not only induced by proline and hypoosmolarity, but is also developmentally regulated in the reproductive organs of *Arabidopsis*", *Plant Physiol.*, 118, 1233-1241.

Navarro, M. & Gull, K. **2000**, "A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*" *Nature*, 414, 759-763.

Neva, F.A. & Brown, H.W. **1994**, *Basic Clinical Parasitology*, Prentice Hall International ed.

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

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. **1997**, "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites", *Protein Eng.*, 10, 1-6.

Nielsen, H., Engelbrecht, J., von Heijne, G. & Brunak, S. **1996**, "Defining a similarity threshold for a functional protein sequence pattern: the signal peptide cleavage site", *Proteins*, 24, 165-177.

Njagi, E.N., Olembo, N.K. & Pearson, D.J. **1992**, "Proline transport by tsetse fly *Glossina morsitans* flight muscle mitochondria", *Comp. Biochem. Physiol. B*, 102, 579-584.

O

Obungu, V.H., Kiaiva, J.K., Njogu, M. & Olembo, N.K. **1999**, "Catabolism of proline by procyclic culture forms of *Trypanosoma congolense*", *Comp. Biochem. Physiol. B*, 123, 59-65.

Olliaro, P.L. & Bryceson, A.D.M. **1993**, "Practical progress and new drugs for changing patterns of Leishmaniasis", *Parasitol. Today*, 9, 323-328.

Omura, T. **1998**, "Mitochondria-targeting sequence, a multi-role sorting sequence recognized at all steps of protein import into mitochondria", *J. Biochem.*, 123, 1010-1016.

Opperdoes, F.R., **1987**, "Compartmentation of carbohydrate metabolism in trypanosomes." *Annu. Rev. Microbiol.*, 41, 127-151.

Opperdoes, F.R. **1995**, Biochemical and Molecular biology of parasites. Carbohydrate and energy metabolism in aerobic protozoa, Academic Press.

Opperdoes, F.R. & Michels, P.A. **1993**, "The glycosomes of the *Kinetoplastida*", *Biochimie J.*, 75, 231-234.

Opperdoes, F.R., Baudhuin, P., Coppens, I., De Roe, C., Edwards, S.W., Weijers, P.J. & Misset, O. **1984**, "Purification, morphometric analysis, and characterization of the glycosomes (microbodies) of the protozoan hemoflagellate *Trypanosoma brucei*", *J. Cell Biol.*, 98, 1178-1184.

Ouellette, M. & Papadopoulo, B. **1993**, "Mechanism of drug resistance in *Leishmania*", *Parasitol. Today*, 9, 150-153.

Overath, P., Czichos, J. & Hass, C. **1986**, "The effect of citrate and cis-aconitate on oxidative metabolism during transformation of *Trypanosoma brucei*", *Eur. J. Bioch.* , 160, 175-182.

P

Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdeno-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, A.V., Leather, S., Moule, S., Oyston, P.C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. & Barrell B.G., **2001**, "Genome sequence of *Yersinia pestis*, the causative agent of plague", *Nature*, 413, 523-527.

Peng, Z., Lu, Q. & Verma, D.P. **1996**, "Reciprocal regulation of delta 1-pyrroline-5-carboxylate synthetase and proline dehydrogenase genes controls proline levels during and after osmotic stress in plants", *Mol. Gen. Genet.*, 253, 334-341.

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

Pittam, M.D. **1970**, "Medium for in vitro culture of *Trypanosoma rhodesiense* and *Trypanosoma brucei*" *Comp. Biochem. Physiol.*, 33, 127-128.

Pollakis, G., Grady, R.W., Dieck, H.A., Clarkson, A.B. Jr., **1995**, "Competition between inhibitors of the trypanosome alternative oxidase (TAO) and reduced coenzyme Q9" *Biochem. Pharmacol.*, 12, 1207-1210.

Priest, J.W. & Hajduk, S.L. **1994**, "Developmental regulation of mitochondrial biogenesis in *Trypanosoma brucei*", *J. Bioenerg. Biomembr.*, 26, 179-191.

Priest, J.W. & Hajduk, S.L.**1994 b**, "Developmental regulation of *Trypanosoma brucei* cytochrome c reductase during bloodstream to procyclic differentiation", *Mol. Biochem. Parasitol.*, 65, 291-304.

R

Ratzkin, B. & Roth, J. **1978**, "Cluster of genes controlling proline degradation in *Salmonella typhimurium*", *J. Bacteriol.*, 133, 744-754.

Raz, B., Iten, M., Grether-Buhler, Y., Kaminsky, R., & Brun, R. **1997**, "The Alamar Blue assay to determine drug sensitivity of African trypanosomes (*T.b. rhodesiense* and *T.b. gambiense*) in vitro", *Acta Trop.*, 68, 139-147.

Read, L.K., Stankey, K.A., Fish, W.R., Muthiani, A.M. & Stuart, K. **1994**, "Developmental regulation of RNA editing and polyadenylation in four life cycle stages of *Trypanosoma congolense*", *Mol. Biochem. Parasitol.*, 68, 297-306.

Reina-San-Martin, B., Degrave, W., Rougeot, C., Cosson, A., Chamond, N., Cordeiro-Da-Silva, A., Arala-Chaves, M., Coutinho, A. & Minoprio, P. **2000**, "A B-cell mitogen from a pathogenic trypanosome is a eukaryotic proline racemase", *Nat. Med.*, 6, 890-897.

Roberts, C.W., McLeod, R., Rice, D.W., Ginger, M., Chance, M.L. & Goad, L.J., **2003**, "Fatty acid and sterol metabolism: potential antimicrobial targets in apicomplexan and trypanosomatid parasitic protozoa" *Mol. Biochem. Parasitol.* 126, 129-142.

Roditi, I., Furger, A., Ruepp, S., Schurch, N. & Butikofer, P. **1998**, "Unravelling the procyclin coat of *Trypanosoma brucei*", *Mol. Biochem. Parasitol.*, 91, 117-130.

Rohloff, P., Rodrigues, C.O. & Docampo, R., **2003**, "Regulatory volume decrease in *Trypanosoma cruzi* involves amino acid efflux and changes in intracellular calcium", *Mol. Biochem. Parasitol.*, 126, 219-230.

S

Sacks, D.L, Hieny, S. & Sher, A., **1985**, "Identification of cell surface carbohydrate and antigenic changes between noninfective and infective developmental stages of *Leishmania major* promastigotes" *J. Immunol.*,135, 564-569.

Sacks, D.L. **1989**, "Metacyclogenesis in *Leishmania* promastigotes", *Exp. Parasitol.*, 69, 100-103.

Sambrook, J., Tristsch, E.F. & Maniatis, T., **1989** Molecular cloning: A Laboratory Manual, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY,.

Sanderson, S.J., Pollock, K.G., Hilley, J.D., Meldal, M., Hilaire, P.S., Juliano, M.A., Juliano, L., Mottram, J.C. & Coombs, G.H., **2000**, "Expression and characterization of a recombinant cysteine proteinase of *Leishmania mexicana*." *Biochem. J.*, 15, 383-388.

Scarpulla, R.C. & Soffer, R.L. **1978**, "Membrane-bound proline dehydrogenase from *Escherichia coli*. Solubilization, purification, and characterization", *J. Biol. Chem.*, 253, 5997-6001.

Seed, J.R. **2001**, "African trypanosomiasis research: 100 years of progress, but questions and problems still remain", *Int. J. Parasitol.*, 31, 434-442.

Seifert, K., Matu, S., Javier Perez-Victoria, F., Castanys, S., Gamarro, F. & Croft, S.L., **2003**," Characterisation of *Leishmania donovani* promastigotes resistant to hexadecylphosphocholine (miltefosine)." *Int. J. Antimicrob. Agents.*, 22, 380-387.

Shlomai, J., **2002**, "Specific recognition of the replication origins of the kinetoplast DNA", *Acta Microbiol.*, 49, 455-567.

Simpson, L., Neckelmann, N., de la Cruz, V.F., Simpson, A.M., Feagin, J.E., Jasmer, D.P. & Stuart, J.E. **1987**, "Comparison of the maxicircle (mitochondrial) genomes of *Leishmania tarentolae* and *Trypanosoma brucei* at the level of nucleotide sequence" *J. Biol. Chem.*, 262, 6182-6196.

Small, W.C. & Jones, M.E. **1990**, "Pyrroline 5-carboxylate dehydrogenase of the mitochondrial matrix of rat liver. Purification, physical and kinetic characteristics", *J. Biol. Chem.*, 265, 18668-18672.

Smith, D.H., Pepin, J. & Stich, A.H.R. **1998**, "Human African Trypanosomiasis: an emerging public health crisis", *Br. Med. Bull.*, 54, 341-355.

Soares, M.J. & de Souza, W. **1988**, "Cytoplasmic organelles of trypanosomatids: a cytochemical and stereological study", *J. Submicrosc. Cytol. Pathol.*, 20, 349-361.

Souto-Padron, T. & de Souza, W. **1979**, "Cytochemical analysis at the fine-structural level of trypanosomatids stained with phosphotungstic acid", *J. Protozool.*, 26, 551-557.

Stevens, J.R. & Gibson, W.C. **1999**, "The evolution of pathogenic trypanosomes", *Cad. Saude Publica*, 15, 673-684.

Stich, A., Barrett, M.P. & Krishna S., **2003**, "Waking up to sleeping sickness", *Trends Parasitol.*, 19, 195-197.

Stiles, J.K., Hicock, P.L., Shah, P.H. & Meade, J.C. **1999**, "Genomic organization, transcription, splicing and gene regulation in *Leishmania*", *Ann. Trop Med. Parasitol.*, 93, 781-807.

Straub, P.F., Reynolds, P.H., Althomsons, S., Mett, V., Zhu, Y., Shearer, G. & Kohl, D.H. **1996**, "Isolation, DNA sequence analysis, and mutagenesis of a proline dehydrogenase gene (putA) from *Bradyrhizobium japonicum*", *Appl. Environ. Microbiol.*, 62, 221-229.

Stuart, K., Allen, T.E., Heidmann, S. & Seiwert, S.D. **1997**, "RNA editing in kinetoplastid protozoa", *Microbiol. Mol. Biol. Rev.*, 61, 105-120.

Surber, M.W. & Maloy, S. **1999**, "Regulation of flavin dehydrogenase compartmentalization: requirements for PutA-membrane association in *Salmonella typhimurium*", *Biochim. Biophys. Acta*, 1421, 5-18.

T

Takagi, H., Sakai, K., Morida, K. & Nakamori, S. **2000**, "Proline accumulation by mutation or disruption of the proline oxidase gene improves resistance to freezing and desiccation stresses in *Saccharomyces cerevisiae*", *FEMS Microbiol. Letter*, 184, 103-108.

ter Kuile, B.H. & Opperdoes, F.R. **1992a**, " Comparative physiology of two protozoan parasites, Leishmania donovani and Trypanosoma brucei, grown in chemostats", *J. Bacteriol.*, 174, 2929-2934.

ter Kuile, B.H. & Opperdoes, F.R. **1992b** "A chemostat study on proline uptake and metabolism of *Leishmania donovani*", *J. Protozool.*, 39, 555-558.

ter Kuile, B.H., **1997**, "Adaptation of metabolic enzyme activities of *Trypanosoma brucei* promastigotes to growth rate and carbon regimen." *J. Bacteriol.*, 179, 4699-4705.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. **1994**," CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice", *Nucleic Acids Res.*, 22, 4673-4680.

Tielens, A.G.M. & Van Hellemond, J.J. **1998**, "Differences in energy metabolism between *Trypanosomatidae*", *Parasitol. Today*, 14, 265-271.

Tritsch, D., Mawlawi, H. & Biellman, J.F. **1993**, "Mechanism based inhibitors of proline dehydrogenase by proline analogues", *Biochim. Biophys. Acta*, 1202, 77-81.

Tschudi, C., Djikeng, A., Shi, H. & Ullu, E. **2003**, "*In vivo* analysis of the RNA interference mechanism in *Trypanosoma brucei*", *Methods*, 30, 304-312.

Turner, C.M. **1999**, "Antigenic variation in *Trypanosoma brucei* infections: an holistic view", *J. Cell Sci.*, 112, 3187-3192.

U

Ullu, E., Djikeng, A., Shi, H. & Tschudi, C. **2002**, "RNA interference: advances and questions", *Philos. Trans. R. Soc. Lond B Biol. Sci.*, 357, 65-70.

V

van Deursen, F.J., Thornton, D.J. & 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.

van Hellemond, J.J., Opperdoes, F.R. & Tielens, A.G. **1998**, "*Trypanosomatidae* produce acetate via a mitochondrial acetate:succinate CoA transferase", *Proc. Natl. Acad. Sci. U.S.A*, 95, 3036-3041.

van Weelden, S.W., Fast, B., Vogt, A., van der, M.P., Saas, J., van Hellemond, J.J., Tielens, A.G. & Boshart, M. **2003**, "Procyclic *Trypanosoma brucei* do not use Krebs cycle activity for energy generation", *J. Biol. Chem.*, 278, 12854-12863.

Vanhamme, L. & Pays, E. **1995**, "Control of gene expression in trypanosomes", *Microbiol. Rev.*, 59, 223-240.

Vassella, E., Den Abbeele, J.V., Butikofer, P., Renggli, C.K., Furger, A., Brun, R. & Roditi, I., **2000**, "A major surface glycoprotein of *Trypanosoma brucei* is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia", *Genes Dev.*, 14, 615-626.

Vassella, E., Studer, E., Schneider, A., Boshart, M. & Roditi, I. **2003**, "Expression of GPEET procyclin on the surface of procyclic forms of *T. brucei* is controlled by the activity of mitochondrial enzymes" *Molecular Parasitology Meeting*, Woods Hole, MA.

Vercesi, A.E., Moreno, S.N. & Docampo, R. **1994**, "Ca^{2+}/H^{+} exchange in acidic vacuoles of *Trypanosoma brucei*", *Biochem. J.*, 304, 227-233.

Vickerman, K., **1969**, " On the surface coat and flagellar adhesion in trypanosomes." *J. Cell Science*, 163-193.

Vickerman, K. **1985**, "Developmental cycles and biology of pathogenic trypanosomes", *Br. Med. bull.*, 41, 105-114.

Vickerman, K., Tetley, L., Hendry, K.A. & Turner, C.M. **1988**, "Biology of African trypanosomes in the tsetse fly", *Biol. Cell*, 64, 109-119.

Vickerman, K., **1994**, "The evolutionary expansion of the trypanosomatid flagellates" *Int. J. Parasitol.*, 24, 1317-1331.

Vilchez, S., Molina, L., Ramos, C. & Ramos, J.L. **2000**, "Proline catabolism by *Pseudomonas putida*: cloning, characterization, and expression of the put genes in the presence of root exudates", *J. Bacteriol.*, 182, 91-99.

Vinod, M.P., Bellur, P. & Becker, D.F. **2002**, "Electrochemical and functional characterization of the proline dehydrogenase domain of the PutA flavoprotein from *Escherichia coli*", *Biochem. J.*, 41, 6525-6532.

Voorheis, H.P. & Martin, B.R. **1980**, "'Swell dialysis' demonstrates that adenylate cyclase in *Trypanosoma brucei* is regulated by calcium ions", *Eur. J. Biochem.*, 113, 223-227.

Voos, W., Martin, H., Krimmer, T. & Pfanner, N. **1999,** "Mechanisms of protein translocation into mitochondria", *Biochim. Biophys. Acta*, 1422, 235-254.

W

Wang, C.C. **1995**, "Molecular Mechanisms an therapeutic approaches to the treatment of African trypanosomiasis", *Ann. Rev. Pharmocol. Toxicol.*, 35, 93-127.

Wang, S. & Brandriss, M.C. **1986**, "Proline utilization *in Saccharomyces cerevisiae*: analysis of the cloned PUT1 gene", *Mol. Cell. Biol.*, 6, 2638-2645.

Wang, S. & Brandriss, M.C. **1987**, "Proline utilizatin in *Saccharomyces cerevisiae*: sequence, regulation and mitochondrial localization of Put1 gene product", *Mol. Cell. Biol.*, 7, 4431-4440.

Wang, Z., Morris, J.C., Drew, M.E. & Englund, P.T. **2000**, "Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an integratable vector with opposing T7 promoters", *J. Biol. Chem.*, 275, 40174-40179.

Weigle, K. & Saravia, N.G. **1996,** "Natural history, clinical evolution, and the host-parasite interaction in New World cutaneous Leishmaniasis", *Clin. Dermatol.*, 14, 433-450.

Welburn, S.C., Fevre, E.M.F., Coleman, P.G., Odiit M. & Maudlin I. **2001**, "Sleeping sickness: a tale of two diseases", *Trends Parasitol.*, 17, 19-24.

Wiemer, E.A., ter Kuile, B.H., Michels, P.A, & Opperdoes, F.R. **1992**, " Pyruvate transport across the plasma membrane of bloodstream form of *T. brucei* is mediated by a facilitated diffusion carrier", *Biochem. Biophys. Res. Commun.*, 184, 1028-1034.

WHO, Report of WHO expert committee, **1998,** Control and surveillance of African trypanosomes, WHO technical report series 881.

Wirtz, E., Leal, S., Ochatt, C. & Cross, G.A. **1999**, "A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*", *Mol. Biochem. Parasitol*, 99, 89-101.

Wood, J.M. **1981**, "Genetics of L-proline utilization in *Escherichia coli*", *J. Bacteriol.*, 146, 895-901.

X

Xia, M., Zhu, Y., Cao, X., You, L. & Chen, Z. **1995**, "Cloning, sequencing and analysis of a gene encoding *Escherichia coli* proline dehydrogenase", *FEMS Microbiol. Lett.*, 127, 235-242.

Zamore, P.D. **2001**, "RNA interference: listening to the sound of silence", *Nat. Struct. Biol.*, 8, 746-750.

Zhu, W. & Becker, D.F. **2003**, "Flavin redox state triggers conformational changes in the PutA protein from *Escherichia coli*", *Biochemistry*, 42, 5469-5477.

Zilbertein, D. **1993**, "Transport of nutrients and irons across membranes of trypanosomatid parasites", *Adv. Parasitol.*, 32, 261-291.

Zilbertein, D. & Gepstein, A. **1993**, "Regulation of L-proline transport in *Leishmania donovani* by extracellular pH", *Mol. Biochem. Parasitol.*, 61, 197-206.

Appendix 1

Composition of media and solutions

Defined Medium

SDM 79 without glucose and proline (according to Dr F. Bringaud, University Bordeaux, France, personal communication)

NaH_2PO_4	157 mg
NaCl	6.8 g
$MgSO_4$	100 mg
KCl	400 mg
$CaCl_2$	200 mg
L-Arginine	100 mg
L-Methionine	70 mg
L-Phenylalanine	80 mg
L-Threonine	350 mg
L-Tyrosine	100 mg
Taurine	160 mg
L-Alanine	200 mg
L-Asparagine	13.2 mg
L-Aspartate	13.3 mg
L-Glutamate	14.7 mg
L-Glutamine	200 mg
Glycine	7.5 mg
L-Serine	60 mg
HEPES	8 g
MOPS	5 g
$NaHCO_3$	2.2 g
Pyruvate	220 mg
Mercaptoethanol 0.1 M	2 ml
Hypoxanthine	14 mg
Thymidine	4 mg
Vitamins 100 X (Sigma, M6895)	10 ml
Essential amino acids 50 X (Gibco BRL, 1130-036)	20 ml
Phenol Red	4 ml
Hemin (2.5 mg/ml)	2 ml

Make up to 1 L. Filter sterilise. Store at 4 °C

Solutions for molecular biology

TELT buffer

Tris-HCl pH 8	50 mM
EDTA pH 9	62.5 mM
LiCl	2.5 M
Triton X-100	4% v/v

TE buffer

Tris-HCl pH 8	10 mM
EDTA pH 8	1 mM

Tris/sucrose buffer

Tri-HCl pH 8	50 mM
Sucrose	0.7 M

1X TAE buffer

Tris acetate pH 8.5	0.04 M
EDTA	0.001 M

LB medium (Luria Bertani medium)

Bacto-tryptone	10 g
Bacto yeast extract	5 g
NaCl	10 g

Make up to 1 L. Sterilise by autoclaving.

Tris-Sucrose buffer

Tris-HCl pH 8	25 mM
EDTA	1 mM
Sucrose	0.25 M

20 X SSC

Tri-sodium citrate	88.23 g
NaCl	175.32 g

Make up to 1 L. Check that the pH is 7-8, and store at room temperature.

20 X Northern Gel Buffer

Na_2HPO_4	0.36 M
NaH_2PO_4	0.04 M

100 X Denhardt's solution

Bovine serum albumin	2 g
Ficoll	400.2 g
Polyvinylpyrolidone	2 g

Make up to a final volume of 100 ml and store at -20 °C

PBS

Sodium phosphate	20 mM
NaCl	150 mM

Solutions for biochemical studies

Resolving buffer

Tris HCl pH 8.8	1.5 M

Stacking buffer

Tris HCl pH 6.8	0.5 M

1 X loading buffer

Tris HCl pH 6.8	50 mM
SDS	2 % (v/v)
Glycerol	10 % (v/v)
β-mercaptoethanol	1 % (v/v)
bromophenol blue	0.1 %

Electrophoresis buffer

Tris	25 mM
Glycine	250 mM
SDS	0.1 %

10 X TBS

Tris-HCl	0.2 M
NaCl	137 mM

TSE buffer

Sucrose	250 mM,
Tris HCl pH 7.7	25 mM
EDTA	1 mM

Solutions for RNA interference studies

ZPFM Media

NaCl	132 mM
KCl	8 mM
$Na_2H_2PO_3^{2-}$	8 mM
KH_2PO_3	1.5 mM
$Mg(C_2H_3O_2)$	1.5 mM
$C_4H_6CaO_4$	90 μM

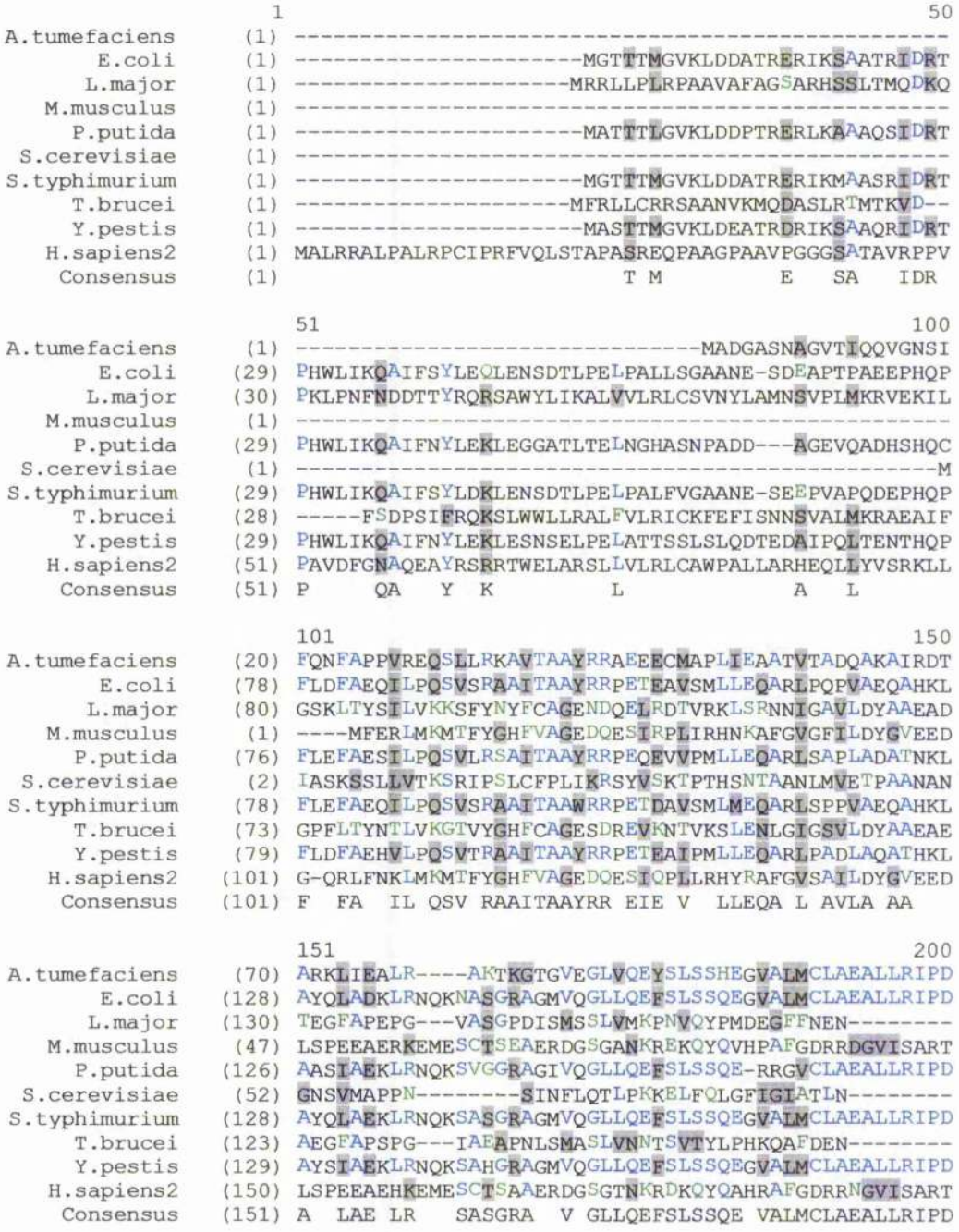
CBSS Buffer

Hepes	25 mM
NaCl	120 mM
$CaCl_2$	0.55 mM
$MgSO_4$	0.4 mM
Na_2HPO_4	5.6 mM
D-glucose	11.1 mM

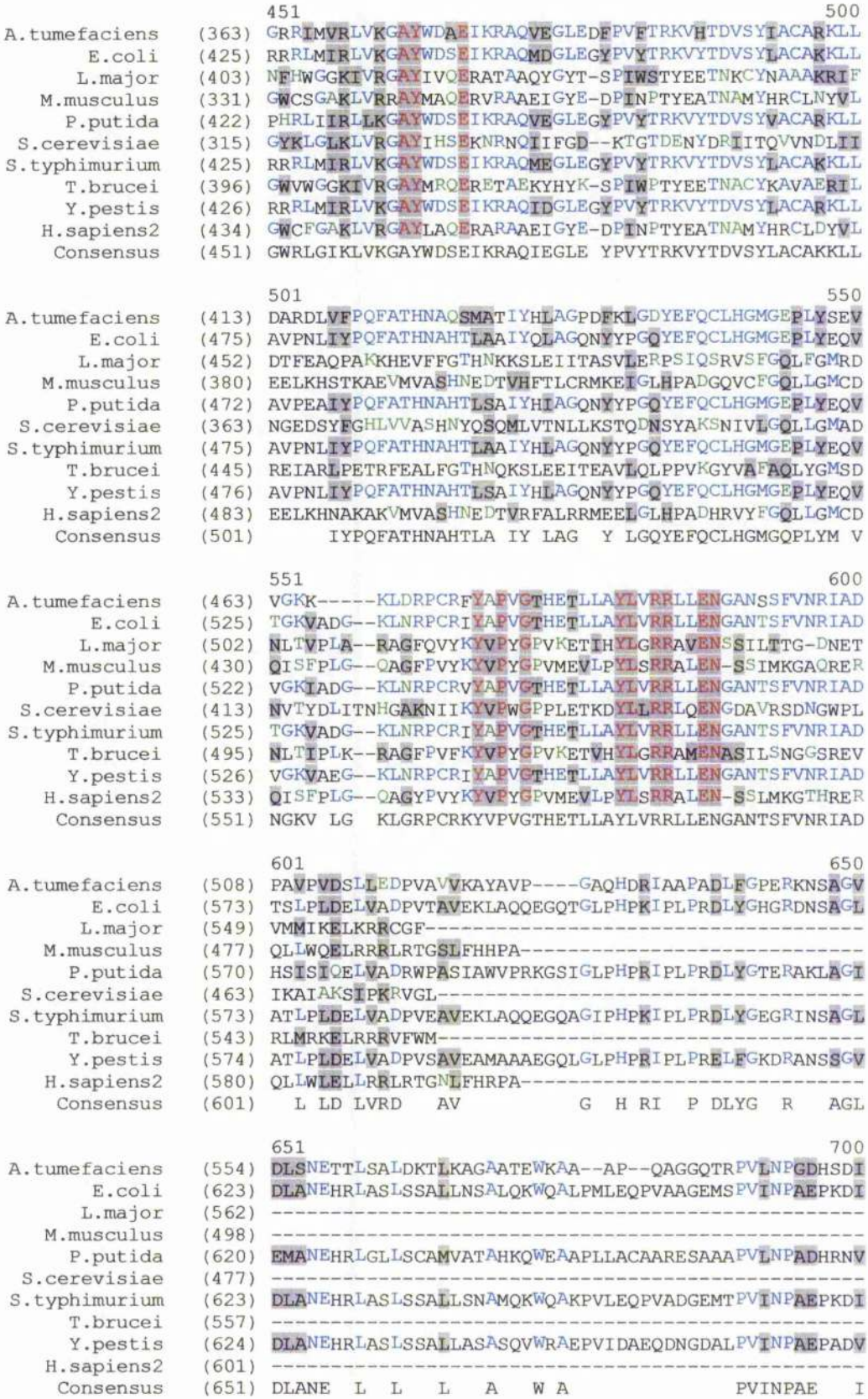
Adjust to pH 7.4 and store at - 20 °C.

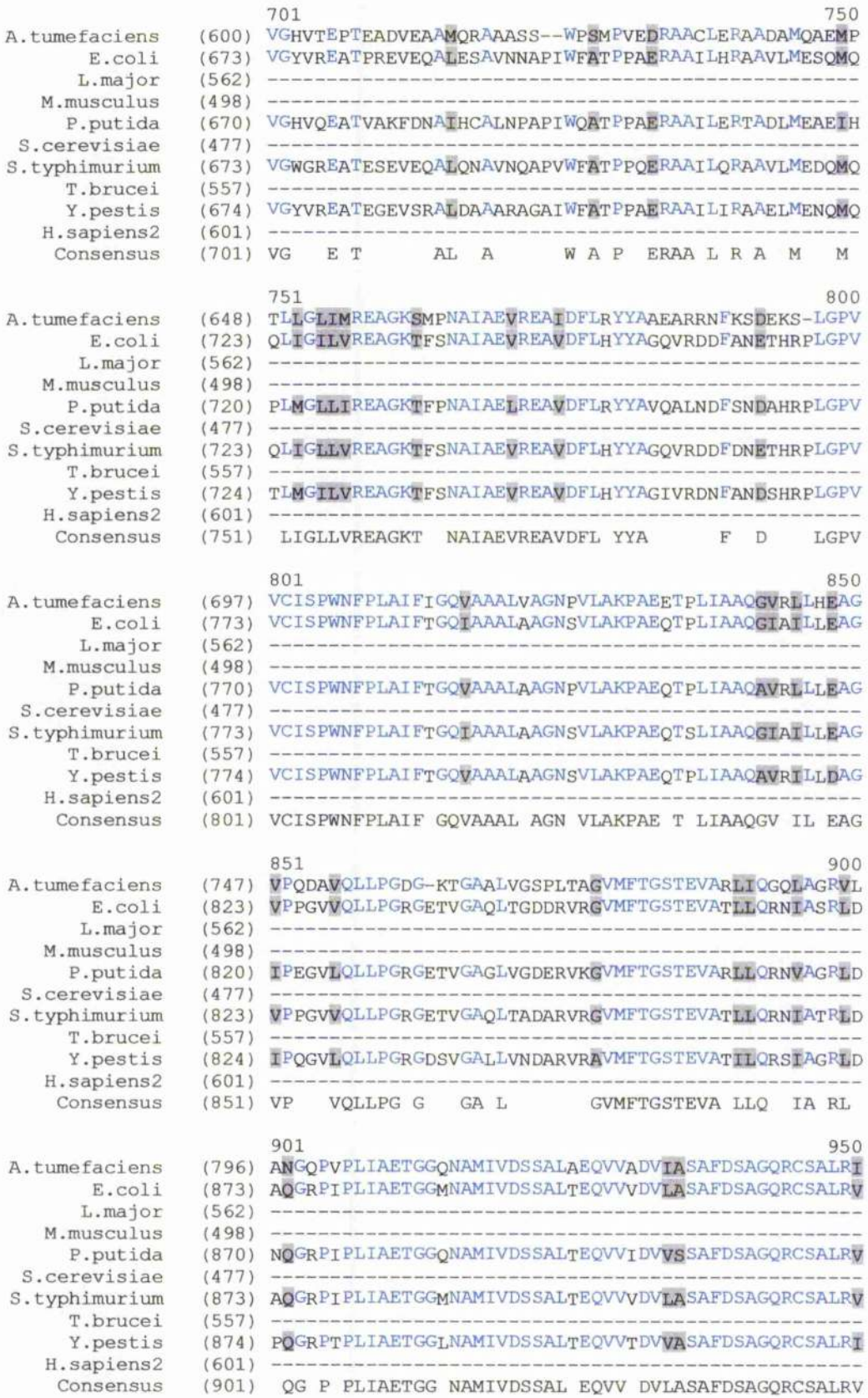
Appendix 2

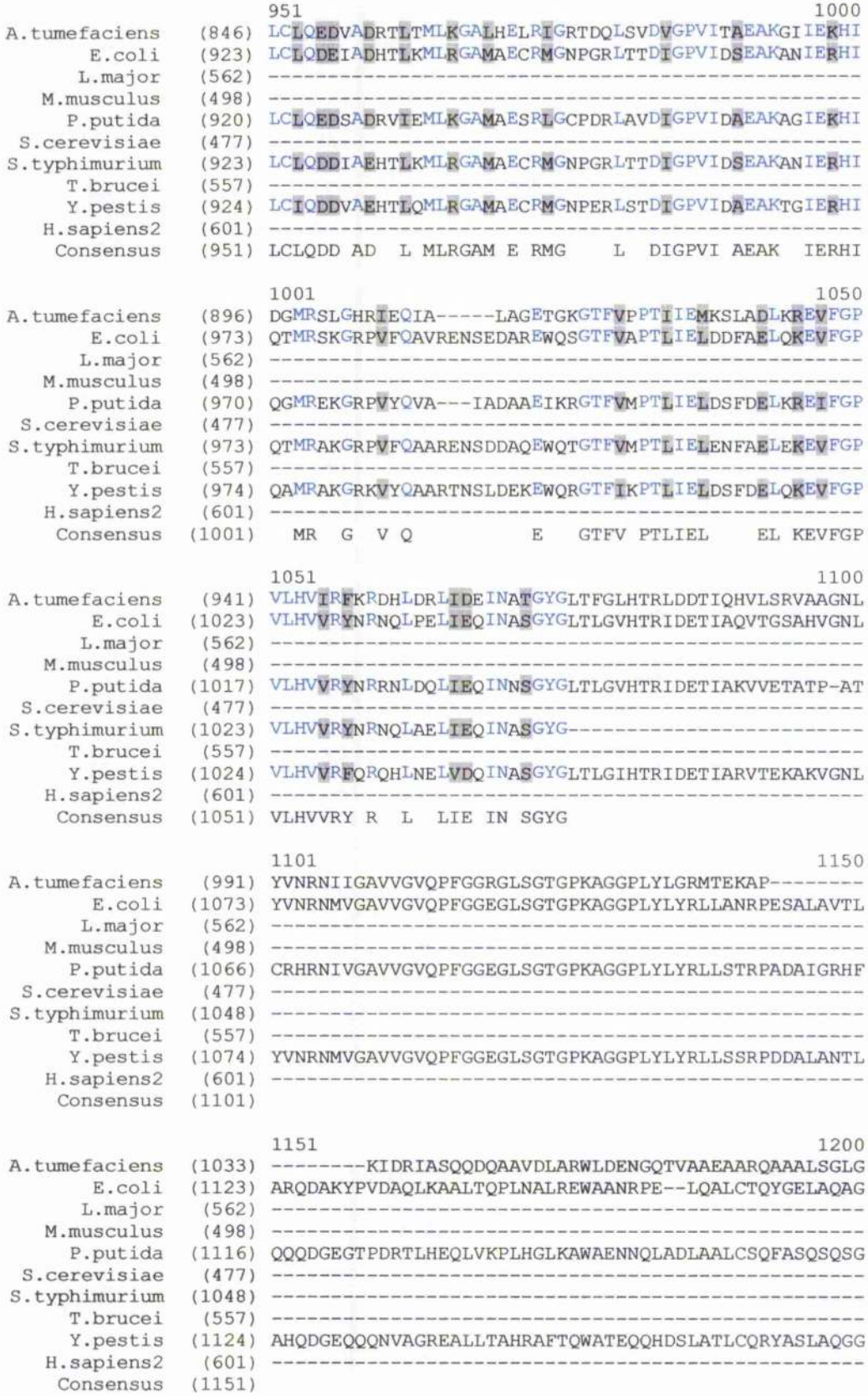
Multiple amino acid sequence alignment of various proline dehydrogenase



```
                      201                                                 250
A.tumefaciens  (116) TATRDALIRDKIARGDWKSHIGGGRSLFVNAATWGLVITGKLTSTVNDSG
E.coli         (178) KATRDALIRDKISNGNWQSHIGRSPSLFVNAATWGLLFTGKLVSTHNEAS
L.major        (169) --------MKLYMMSIMHASLYS-PRNVAGVTAVKVTGMCDPQLLARVSA
M.musculus      (97) YFYANEAKCDNYMENLLQCIKASGGASDGGFSAIKLTALGRPQFLLQFSD
P.putida       (175) KGTRDALIRDKISTGNWQPHLGNSPSLFVNAATWGLLLTGKLVSTHNETG
S.cerevisiae    (86) ----------SFFLNTIIKLFPYIPIPVIKFFVSSLYCGGE-----NFKE
S.typhimurium  (178) KATRDALIRDKISNGNWQSHIGRSPSLFVNAATWGLLFTGRLVSTHNEAN
T.brucei       (162) --------MKLYVMCVLHAALHK-PEGGVGLAAVKVTGMCDPQLLARVSA
Y.pestis       (179) KPTRDALIRDKISNGNWHSHLGRSPSMFVNAATWGLLFTGRLVSTHNEAK
H.sapiens2     (200) YFYANEAKCDSHMETFLRCIEASGRVSDDGFIAIKLTALGRPQFLLQFSE
Consensus      (201)   TRDALIRDKIS G W SHLG  PSLFVNAATWGLL TGRL ST NES

                      251                                                 300
A.tumefaciens  (166) LSAALTKLIARAGEPVIRRGVDMAMRMMGEQFVTGETIGEAIKRSKPLEE
E.coli         (228) LSRSLNRIIGKSGEPLIRKGVDMAMRLMGEQFVTGETIAEALANARKLEE
L.major        (210) LLMSVHQSWCKHFTNEESLKLEECRVVMGVNRKHQLFITYDQLRAGFEKY
M.musculus     (147) VLTRWRRFFHQMAAEQG-----QAGRAAVDTKLEVAVLQDSIAKMGIASR
P.putida       (225) LTSSLTRIIGKSGEPMIRKGVDMAMRLMGEQFVTGETIAEALANASRFEA
S.cerevisiae   (121) VIECGKRLQKRGISNMML------LTIENS-EGTKSLSSTPVDQIVKETI
S.typhimurium  (228) LSRSLNRIIGKSGEPLIRKGVDMAMRLMGEQFVTGETIAQALANARKLEE
T.brucei       (203) ILHSVHRDWIEYFTEEQPPPVEECNVVMGTKTEHKRYITRDQVRKGLTKL
Y.pestis       (229) LSGSLNRIIGKGGEPLIRKGVDMAMRLMGEQFVTGETISEALANARKLED
H.sapiens2     (250) VLAKWRCFFHQMAVEQG-----QAGLAAMDTKLEVAVLQESVAKLGIASR
Consensus      (251) LS SL RIIGK GEPLIRKGVDMAMRLMGEQFVTGETIAEALARA K E

                      301                                                 350
A.tumefaciens  (216) QGFQYSYDMLGEAATTAKDAERYYKDYENAIHAIGKASAGRGIYGGPGIS
E.coli         (278) KGFRYSYDMLGEAALTAADAQAYMVSYQQAIHAIGKASNGRGIYEGPGIS
L.major        (260) NPSNKLSDAQFKEITEALDPRKTGKVNYFEYKEMLTNA---LIAVEPTPV
M.musculus     (192) AEIEGWFTPETLGVSGTVDLLDWN----SLIDSRTRLS---RHLVVPNVQ
P.putida       (275) KGFRYSYDMLGEAALTEHDAQKYLASYEQAIHSIGKASHGRGIYEGPGIS
S.cerevisiae   (165) SSVHNILLPNIIGQLESKPINDIAPGYIALKPSALVDNPHEVLYNFSNPA
S.typhimurium  (278) KGFRYSYDMLGEAALTAADAQAYMVSYQQAIHAIGKASNGRGIYEGPGIS
T.brucei       (253) ASSQKYTEDEINAVLQVLDPNNEGKTNYYKFKTVVSEA---VLALDPTPV
Y.pestis       (279) KGFRYSYDMLGEAALTEADAQAYLLSYQQAIHAIGKASNGRGIYEGPGIS
H.sapiens2     (295) AEIEDWFTAETLGVSGTMDLLDWS----SLIDSRTKLS---KHLVVPNAQ
Consensus      (301)  GF YSYDMLGEAALTA DAQ Y  SY QAIHAIGKAS GRGIY GPGIS

                      351                                                 400
A.tumefaciens  (266) IKLSALHPRYARAQAERVMAELLPRVKSLMLLSKTYDIGLNIDAEEADRL
E.coli         (328) IKLSALHPRYSRAQYDRVMEELYPRLKSLTLLARQYDIGINIDAEESDRL
L.major        (307) QQALIEGLPQMSAKEKVLWKNVNNRLLLIASMAKELNVRMLVDAEQTFYQ
M.musculus     (235) TGQLEPLLSRFTEEEEQQMKRMLQRMDVLAKKAKEAGVRLMIDAEQSYFQ
P.putida       (325) IKLSALHPRYSRAQYERVMEELYPRLLSLTLLAKQYDIGLNIDAEEADRL
S.cerevisiae   (215) YKAQRDQLIENCSKITKEIFELNQSLLKKYPERKAPFMVSTIDAEKYDLQ
S.typhimurium  (328) IKLSALHPRYSRAQYDRVMEELYPRLKSLTLLARQYDIGLNIDAEEADRL
T.brucei       (300) QKIIIDKLPKLTTEERELWRHLHWRLSVIVRTAKDLRVRVLFDAEQTFYQ
Y.pestis       (329) IKLSALHPRYSRAQYERVMDELYPRLLSLTLQARQYDIGINIDAEEADRL
H.sapiens2     (338) TGQLEPLLSRFTEEEELQMTRMLQRMDVLAKKATEMGVRIMVDAEQTYFQ
Consensus      (351) IKLSALHPRYSRAQ ERVM ELYPRL SL LLAK YDIGLNIDAEEADRQ

                      401                                                 450
A.tumefaciens  (316) ELSLDLLEELALDKDLAG---WNGLGFVVQAYGRRCPFVLDYIIDLARRS
E.coli         (378) EISLDLLEKLCFEPELAG---WNGIGFVIQAYQKRCPLVIDYLIDLATRS
L.major        (357) LAIDAIVATLQKTYNTE----LPVVYNTYQCYLTYAEDRIDNDLVRARHM
M.musculus     (285) PAISRLTLEMQRRFNVD----KPFIFNTFQCYLKDAYDNVTLDMELARRE
P.putida       (375) ELSLDLLERLCFEPSLAG---WNGIGFVIQAYQKRCPYLINYFFDLAKRT
S.cerevisiae   (265) ENGVYELQRILFQKFNPTSSKLISCVGTWQLYLRDSGDHILHELKLAQEN
S.typhimurium  (378) EISLDLLEKLCFEPELAG---WNGIGFVIQAYQKRCPLVIDYLVDLASRS
T.brucei       (350) LAIDNIVLQFQRQFNKK----EAIVYNTYQCYLTYTEDRVFNDLTRAELE
Y.pestis       (379) EISLDLLERLCFEPQLAG---WNGIGFVIQAYQKRCPSTIDAVIDMAQRS
H.sapiens2     (388) PAISRLTLEMQRKFNVE----KPLIFNTYQCYLKDAYDNVTLDVELARRE
Consensus      (401) EISLDLLERL FE NLAG   WNGIGFVYQAYLKRCPD IDYDIDLARRS
```


451 500
A.tumefaciens (363) GRRIMVRLVKGAYWDAEIKRAQVEGLEDFPVFTRKVHTDVSYIACARKLL
E.coli (425) RRRLMIRLVKGAYWDSEIKRAQMDGLEGYPVYTRKVYTDVSYLACAKKLL
L.major (403) NFHWGGKIVRGAYIVQERATAAQYGYT-SPIWSTYEETNKCYNAAAKRIF
M.musculus (331) GWCSGAKLVRRAYMAQERVRAAEIGYE-DPINPTYEATNAMYHRCLNYVL
P.putida (422) PHRLIIRLLKGAYWDSEIKRAQVEGLEGYPVYTRKVYTDVSYVACARKLL
S.cerevisiae (315) GYKLGLKLVRGAYIHSEKNRNQIIFGD--KTGTDENYDRIITQVVNDLII
S.typhimurium (425) RRRLMIRLVKGAYWDSEIKRAQMEGLEGYPVYTRKVYTDVSYLACAKKLL
T.brucei (396) GWVWGGKIVRGAYMRQERETAEKYHYK-SPIWPTYEETNACYKAVAERIL
Y.pestis (426) RRRLMIRLVKGAYWDSEIKRAQIDGLEGYPVYTRKVYTDVSYLACARKLL
H.sapiens2 (434) GWCFGAKLVRGAYLAQERARAAEIGYE-DPINPTYEATNAMYHRCLDYVL
Consensus (451) GWRLGIKLVKGAYWDSEIKRAQIEGLE YPVYTRKVYTDVSYLACAKKLL
501 550
A.tumefaciens (413) DARDLVFPQFATHNAQSMATIYHLAGPDFKLGDYEFQCLHGMGEPLYSEV
E.coli (475) AVPNLIYPQFATHNAHTLAAIYQLAGQNYYPGQYEFQCLHGMGEPLYEQV
L.major (452) DTFEAQPAKKHEVFFGTHNKKSLEIITASVLERPSIQSRVSFGQLFGMRD
M.musculus (380) EELKHSTKAEVMVASHNEDTVHFTLCRMKEIGLHPADGQVCFGQLLGMCD
P.putida (472) AVPEAIYPQFATHNAHTLSAIYHIAGQNYYPGQYEFQCLHGMGEPLYEQV
S.cerevisiae (363) NGEDSYFGHLVVASHNYQSQMLVTNLLKSTQDNSYAKSNIVLGQLLGMAD
S.typhimurium (475) AVPNLIYPQFATHNAHTLAAIYHLAGQNYYPGQYEFQCLHGMGEPLYEQV
T.brucei (445) REIARLPETRFEALFGTHNQKSLEEITEAVLQLPPVKGYVAFAQLYGMSD
Y.pestis (476) AVPNLIYPQFATHNAHTLSAIYHLAGQNYYPGQYEFQCLHGMGEPLYEQV
H.sapiens2 (483) EELKHNAKAKVMVASHNEDTVRFALRRMEELGLHPADHRVYFGQLLGMCD
Consensus (501) IYPQFATHNAHTLA IY LAG Y LGQYEFQCLHGMGQPLYM V
551 600
A.tumefaciens (463) VGKK-----KLDRPCRFYAPVGTHETLLAYLVRRLLENGANSSFVNRIAD
E.coli (525) TGKVADG--KLNRPCRIYAPVGTHETLLAYLVRRLLENGANTSFVNRIAD
L.major (502) NLTVPLA--RAGFQVYKYVPYGPVKETIHYLGRRAVENSSILTTG-DNET
M.musculus (430) QISFPLG--QAGFPVYKYVPYGPVMEVLPYLSRRALEN-SSIMKGAQRER
P.putida (522) VGKIADG--KLNRPCRVYAPVGTHETLLAYLVRRLLENGANTSFVNRIAD
S.cerevisiae (413) NVTYDLITNHGAKNIIKYVPWGPPLETKDYLLRRLQENGDAVRSDNGWPL
S.typhimurium (525) TGKVADG--KLNRPCRIYAPVGTHETLLAYLVRRLLENGANTSFVNRIAD
T.brucei (495) NLTIPLK--RAGFPVFKYVPYGPVKETVHYLGRRAMENASILSNGGSREV
Y.pestis (526) VGKVAEG--KLNRPCRIYAPVGTHETLLAYLVRRLLENGANTSFVNRIAD
H.sapiens2 (533) QISFPLG--QAGYPVYKYVPYGPVMEVLPYLSRRALEN-SSLMKGTHRER
Consensus (551) NGKV LG KLGRPCRKYVPVGTHETLLAYLVRRLLENGANTSFVNRIAD
601 650
A.tumefaciens (508) PAVPVDSLLEDPVAVVKAYAVP----GAQHDRIAAPADLFGPERKNSAGV
E.coli (573) TSLPLDELVADPVTAVEKLAQQEGQTGLPHPKIPLPRDLYGHGRDNSAGL
L.major (549) VMMIKELKRRCGF----------
M.musculus (477) QLLWQELRRRLRTGSLFHHPA----------
P.putida (570) HSISIQELVADRWPASIAWVPRKGSIGLPHPRIPLPRDLYGTERAKLAGI
S.cerevisiae (463) IKAIAKSIPKRVGL----------
S.typhimurium (573) ATLPLDELVADPVEAVEKLAQQEGQAGIPHPKIPLPRDLYGEGRINSAGL
T.brucei (543) RLMRKELRRRVFWM----------
Y.pestis (574) ATLPLDELVADPVSAVEAMAAAEGQLGLPHPRIPLPRELFGKDRANSSGV
H.sapiens2 (580) QLLWLELLRRLRTGNLFHRPA----------
Consensus (601) L LD LVRD AV G H RI P DLYG R AGL
651 700
A.tumefaciens (554) DLSNETTLSALDKTLKAGAATEWKAA--AP--QAGGQTRPVLNPGDHSDI
E.coli (623) DLANEHRLASLSSALLNSALQKWQALPMLEQPVAAGEMSPVINPAEPKDI
L.major (562) ----------
M.musculus (498) ----------
P.putida (620) EMANEHRLGLLSCAMVATAHKQWEAAPLLACAARESAAAPVLNPADHRNV
S.cerevisiae (477) ----------
S.typhimurium (623) DLANEHRLASLSSALLSNAMQKWQAKPVLEQPVADGEMTPVINPAEPKDI
T.brucei (557) ----------
Y.pestis (624) DLANEHRLASLSSALLASASQVWRAEPVIDAEQDNGDALPVINPAEPADV
H.sapiens2 (601) ----------
Consensus (651) DLANE L L L A W A PVINPAE I


701 750
A.tumefaciens (600) VGHVTEPTEADVEAAMQRAAASS--WPSMPVEDRAACLERAADAMQAEMP
E.coli (673) VGYVREATPREVEQALESAVNNAPIWFATPPAERAAILHRAAVLMESQMQ
L.major (562) ----------
M.musculus (498) ----------
P.putida (670) VGHVQEATVAKFDNAIHCALNPAPIWQATPPAERAAILERTADLMEAEIH
S.cerevisiae (477) ----------
S.typhimurium (673) VGWGREATESEVEQALQNAVNQAPVWFATPPQERAAILQRAAVLMEDQMQ
T.brucei (557) ----------
Y.pestis (674) VGYVREATEGEVSRALDAAARAGAIWFATPPAERAAILIRAAELMENQMQ
H.sapiens2 (601) ----------
Consensus (701) VG E T AL A W A P ERAA L R A M M
751 800
A.tumefaciens (648) TLLGLIMREAGKSMPNAIAEVREAIDFLRYYAAEARRNFKSDEKS-LGPV
E.coli (723) QLIGILVREAGKTFSNAIAEVREAVDFLHYYAGQVRDDFANETHRPLGPV
L.major (562) ----------
M.musculus (498) ----------
P.putida (720) PLMGLLIREAGKTFPNAIAELREAVDFLRYYAVQALNDFSNDAHRPLGPV
S.cerevisiae (477) ----------
S.typhimurium (723) QLIGLLVREAGKTFSNAIAEVREAVDFLHYYAGQVRDDFDNETHRPLGPV
T.brucei (557) ----------
Y.pestis (724) TLMGILVREAGKTFSNAIAEVREAVDFLHYYAGIVRDNFANDSHRPLGPV
H.sapiens2 (601) ----------
Consensus (751) LIGLLVREAGKT NAIAEVREAVDFL YYA F D LGPV
801 850
A.tumefaciens (697) VCISPWNFPLAIFIGQVAAALVAGNPVLAKPAEETPLIAAQGVRLLHEAG
E.coli (773) VCISPWNFPLAIFTGQIAAALAAGNSVLAKPAEQTPLIAAQGIAILLEAG
L.major (562) ----------
M.musculus (498) ----------
P.putida (770) VCISPWNFPLAIFTGQVAAALAAGNPVLAKPAEQTPLIAAQAVRLLLEAG
S.cerevisiae (477) ----------
S.typhimurium (773) VCISPWNFPLAIFTGQIAAALAAGNSVLAKPAEQTSLIAAQGIAILLEAG
T.brucei (557) ----------
Y.pestis (774) VCISPWNFPLAIFTGQVAAALAAGNSVLAKPAEQTPLIAAQAVRILLDAG
H.sapiens2 (601) ----------
Consensus (801) VCISPWNFPLAIF GQVAAAL AGN VLAKPAE T LIAAQGV IL EAG
851 900
A.tumefaciens (747) VPQDAVQLLPGDG-KTGAALVGSPLTAGVMFTGSTEVARLIQGQLAGRVL
E.coli (823) VPPGVVQLLPGRGETVGAQLTGDDRVRGVMFTGSTEVATLLQRNIASRLD
L.major (562) ----------
M.musculus (498) ----------
P.putida (820) IPEGVLQLLPGRGETVGAGLVGDERVKGVMFTGSTEVARLLQRNVAGRLD
S.cerevisiae (477) ----------
S.typhimurium (823) VPPGVVQLLPGRGETVGAQLTADARVRGVMFTGSTEVATLLQRNIATRLD
T.brucei (557) ----------
Y.pestis (824) IPQGVLQLLPGRGDSVGALLVNDARVRAVMFTGSTEVATILQRSIAGRLD
H.sapiens2 (601) ----------
Consensus (851) VP VQLLPG G GA L GVMFTGSTEVA LLQ IA RL
901 950
A.tumefaciens (796) ANGQPVPLIAETGGQNAMIVDSSALAEQVVADVIASAFDSAGQRCSALRI
E.coli (873) AQGRPIPLIAETGGMNAMIVDSSALTEQVVVDVLASAFDSAGQRCSALRV
L.major (562) ----------
M.musculus (498) ----------
P.putida (870) NQGRPIPLIAETGGQNAMIVDSSALTEQVVIDVVSSAFDSAGQRCSALRV
S.cerevisiae (477) ----------
S.typhimurium (873) AQGRPIPLIAETGGMNAMIVDSSALTEQVVVDVLASAFDSAGQRCSALRV
T.brucei (557) ----------
Y.pestis (874) PQGRPTPLIAETGGLNAMIVDSSALTEQVVTDVVASAFDSAGQRCSALRI
H.sapiens2 (601) ----------
Consensus (901) QG P PLIAETGG NAMIVDSSAL EQVV DVLASAFDSAGQRCSALRV


951 1000
A.tumefaciens (846) LCLQEDVADRTLTMLKGALHELRIGRTDQLSVDVGPVITAEAKGIIEKHI
E.coli (923) LCLQDEIADHTLKMLRGAMAECRMGNPGRLTTDIGPVIDSEAKANIERHI
L.major (562) ----------
M.musculus (498) ----------
P.putida (920) LCLQEDSADRVIEMLKGAMAESRLGCPDRLAVDIGPVIDAEAKAGIEKHI
S.cerevisiae (477) ----------
S.typhimurium (923) LCLQDDIAEHTLKMLRGAMAECRMGNPGRLTTDIGPVIDSEAKANIERHI
T.brucei (557) ----------
Y.pestis (924) LCIQDDVAEHTLQMLRGAMAECRMGNPERLSTDIGPVIDAEAKTGIERHI
H.sapiens2 (601) ----------
Consensus (951) LCLQDD AD L MLRGAM E RMG L DIGPVI AEAK IERHI
1001 1050
A.tumefaciens (896) DGMRSLGHRIEQIA-----LAGETGKGTFVPPTIIEMKSLADLKREVFGP
E.coli (973) QTMRSKGRPVFQAVRENSEDAREWQSGTFVAPTLIELDDFAELQKEVFGP
L.major (562) ----------
M.musculus (498) ----------
P.putida (970) QGMREKGRPVYQVA---IADAAEIKRGTFVMPTLIELDSFDELKREIFGP
S.cerevisiae (477) ----------
S.typhimurium (973) QTMRAKGRPVFQAARENSDDAQEWQTGTFVMPTLIELENFAELEKEVFGP
T.brucei (557) ----------
Y.pestis (974) QAMRAKGRKVYQAARTNSLDEKEWQRGTFIKPTLIELDSFDELQKEVFGP
H.sapiens2 (601) ----------
Consensus (1001) MR G V Q E GTFV PTLIEL EL KEVFGP
1051 1100
A.tumefaciens (941) VLHVIRFKRDHLDRLIDEINATGYGLTFGLHTRLDDTIQHVLSRVAAGNL
E.coli (1023) VLHVVRYNRNQLPELIEQINASGYGLTLGVHTRIDETIAQVTGSAHVGNL
L.major (562) ----------
M.musculus (498) ----------
P.putida (1017) VLHVVRYNRRNLDQLIEQINNSGYGLTLGVHTRIDETIAKVVETATP-AT
S.cerevisiae (477) ----------
S.typhimurium (1023) VLHVVRYNRNQLAELIEQINASGYG----------
T.brucei (557) ----------
Y.pestis (1024) VLHVVRFQRQHLNELVDQINASGYGLTLGIHTRIDETIARVTEKAKVGNL
H.sapiens2 (601) ----------
Consensus (1051) VLHVVRY R L LIE IN SGYG
1101 1150
A.tumefaciens (991) YVNRNIIGAVVGVQPFGGRGLSGTGPKAGGPLYLGRMTEKAP--------
E.coli (1073) YVNRNMVGAVVGVQPFGGEGLSGTGPKAGGPLYLYRLLANRPESALAVTL
L.major (562) ----------
M.musculus (498) ----------
P.putida (1066) CRHRNIVGAVVGVQPFGGEGLSGTGPKAGGPLYLYRLLSTRPADAIGRHF
S.cerevisiae (477) ----------
S.typhimurium (1048) ----------
T.brucei (557) ----------
Y.pestis (1074) YVNRNMVGAVVGVQPFGGEGLSGTGPKAGGPLYLYRLLSSRPDDALANTL
H.sapiens2 (601) ----------
Consensus (1101)
1151 1200
A.tumefaciens (1033) --------KIDRIASQQDQAAVDLARWLDENGQTVAAEAARQAAALSGLG
E.coli (1123) ARQDAKYPVDAQLKAALTQPLNALREWAANRPE--LQALCTQYGELAQAG
L.major (562) ----------
M.musculus (498) ----------
P.putida (1116) QQQDGEGTPDRTLHEQLVKPLHGLKAWAENNQLADLAALCSQFASQSQSG
S.cerevisiae (477) ----------
S.typhimurium (1048) ----------
T.brucei (557) ----------
Y.pestis (1124) AHQDGEQQQNVAGREALLTAHRAFTQWATEQQHDSLATLCQRYASLAQGG
H.sapiens2 (601) ----------
Consensus (1151)

```
                        1201                                                1250
A.tumefaciens   (1075)  FETELAGPVGERNVYALHPRGKILLVPATEQGLYRQLAAALSTGNSVVID
       E.coli   (1171)  TQRLLPGPTGERNTWTLLPRERVLCIADDEQDALTQLAAVLAVGSQVLWP
      L.major    (562)  --------------------------------------------------
    M.musculus   (498)  --------------------------------------------------
      P.putida  (1166)  IARLLPGPTGERNSYTILPREHVLCLADNETDLLAQFAAVLAVGSSAVWV
  S.cerevisiae   (477)  --------------------------------------------------
 S.typhimurium  (1048)  --------------------------------------------------
      T.brucei   (557)  --------------------------------------------------
      Y.pestis  (1174)  TVRLLPGPTGERNTYALLPRERVLCLADTESDTLTQLAAVLATGSQVLWP
    H.sapiens2   (601)  --------------------------------------------------
     Consensus  (1201)

                        1251                                                1300
A.tumefaciens   (1125)  NASGLEKAIYGLPASVTSRIVWADDWEKS-APFAGALVEGDAERVVAINR
       E.coli   (1221)  DDALHRQLVKALPSAVSERIQLAKAENITAQPFDAVIFHGDSDQLRALCE
      L.major    (562)  --------------------------------------------------
    M.musculus   (498)  --------------------------------------------------
      P.putida  (1216)  DGEPGKALRARLPRELQAKVKLVADWNKDEVAFDAVIHHGDSDQLRGVCQ
  S.cerevisiae   (477)  --------------------------------------------------
 S.typhimurium  (1048)  --------------------------------------------------
      T.brucei   (557)  --------------------------------------------------
      Y.pestis  (1224)  ENDVQKALLPQLPTEVQSRITLTHDWQTANITFDAVIYHGDADQLRTLCE
    H.sapiens2   (601)  --------------------------------------------------
     Consensus  (1251)

                        1301                                                1350
A.tumefaciens   (1174)  KIAALPGPLVLFQAATTDALSGESQPYTLDWLVEEVSVSVNTTAAGGNAS
       E.coli   (1271)  AVAARDGTIVSVQGFARGESN-----ILLERLYIERSLSVNTAAAGGNAS
      L.major    (562)  --------------------------------------------------
    M.musculus   (498)  --------------------------------------------------
      P.putida  (1266)  QVAKRAGAIVGVHGLSSGDHQ-----IALERLVIERAVSVNTAAAGGNAS
  S.cerevisiae   (477)  --------------------------------------------------
 S.typhimurium  (1048)  --------------------------------------------------
      T.brucei   (557)  --------------------------------------------------
      Y.pestis  (1274)  QVAQIDGPIVSVQGFARGETN-----ILLERLLIEHSLSVNTAAAGGNAS
    H.sapiens2   (601)  --------------------------------------------------
     Consensus  (1301)

                        1351
A.tumefaciens   (1224)  LMSIG
       E.coli   (1316)  LMTIG
      L.major    (562)  -----
    M.musculus   (498)  -----
      P.putida  (1311)  LMTIG
  S.cerevisiae   (477)  -----
 S.typhimurium  (1048)  -----
      T.brucei   (557)  -----
      Y.pestis  (1319)  LMTIG
    H.sapiens2   (601)  -----
     Consensus  (1351)
```

Multiple amino acid sequence alignment of various proline dehydrogenase.

The eukaryotic proteins contain only proline dehydrogenase while prokaryotic proteins (including those of α-proteobacteria) contain proline dehydrogenase and pyrroline-5-carboxylate dehydrogenase domains.

% of positivity : 54 % (in blue)

% of identity : 1.3 % (in red)

Appendix 3

The conserved blocks

>Vvulnificus
LPYLVRRLLENGANSSFVHR-RLVKGAYWDSEVKWS-GLGVFVCISPWNFPLAIF
LGQ-SDYEFQRLHGMGDSLYNHVMEQ-RQPVRIYAPVG-QFVRGRTIEEAQKNGRP
MRDKG
>Smeliloti
LAYLVRRLLENGANSSFVHR-RLVKGAYWDAEIKRA-PLGPIVCISPWNFPLAIF
TGQ-GKYEFQCLHGMGEPLYEEVVGR-DRPCRIYAPVG-QFVTGETIREALKRSKE
LEEKG
>Bmelitensis
LAYLVRRLLENGANSSFVNR-RLVKGAYWDAEIKRA-ALGPVVCISPWNFPLAIF
TGQ-GKFEFQCLHGMGEPLYDEVVGP-GRPARIYAPVG-QFVTGETIDEALKRAKE
LEERG
>Atumefaciens
LAYLVRRLLENGANSSFVNR-RLVKGAYWDAEIKRA-SLGPVVCISPWNFPLAIF
IGQ-GDYEFQCLHGMGEPLYSEVVGK-DRPCRFYAPVG-QFVTGETIGEAIKRSKPL
EEQG
>Ypestis
LAYLVRRLLENGANTSFVNR-RLVKGAYWDSEIKRA-PLGPVVCISPWNFPLAIF
TGQ-GQYEFQCLHGMGEPLYEQVVGK-NRPCRIYAPVG-QFVTGETISEALANAR
KLEDKG
>Styphimurium
LAYLVRRLLENGANTSFVNR-RLVKGAYWDSEIKRA-PLGPVVCISPWNFPLAIF
TGQ-GQYEFQCLHGMGEPLYEQVTGK-NRPCRIYAPVG-QFVTGETIAQALANAR
KLEEKG
>Pputida
LAYLVRRLLENGANTSFVNR-RLLKGAYWDSEIKRA-PLGPVVCISPWNFPLAIF
TGQ-GQYEFQCLHGMGEPLYEQVVGK-NRPCRVYAPVG-QFVTGETIAEALANAS
RFEAKG
>Ecoli
LAYLVRRLLENGANTSFVNR-RLVKGAYWDSEIKRA-PLGPVVCISPWNFPLAIF
TGQ-GQYEFQCLHGMGEPLYEQVTGK-NRPCRIYAPVG-QFVTGETIAEALANAR
KLEEKG
>Bcepacia
LAYLVRRLLENGANTSFVNR-RLVKGAYWDSEIKRA-PLGPVVCISPWNFPLAIF
VGQ-GQYEFQCLHGMGEPLYEQVVGS-GRPCRIYAPVG-QFVTGETIKEALDHAR
GLEAQG
>Rcapsulatus
LAYLVRRLLENGANSSFVNQ-RLVKGAYWDAEMKRA-PRGAVVAISPWNFPLA
IFTGQ-RPFEFQRLHGMGARLHDIVLRE-GGRCRIYAPVG-QFVLGETIEKALERA
EKREAEG
>Nmeningidis
LAYLVRRLLENGANSSFVNQ-RLVKGAYWDSEIKWA-AVGAIVAISPWNFPLAIF
TGE-KDFEHQCLHGMGETLYDQVVGP-GRRVRVYAPVG-QFVTGQTIEEALQNGK
EREKMG
>Mloti
LAYLVRRLLENGANSSFVHQ-RLVKGAYWDTEIKRA-ARGAIVCISPWNFPLAIF
TGQ-DSFEFQRLHGMGEALHETVRQA-GTRCRIYAPVG-QFVLGRTIAEAVKRGR
PMTQKG
>Paeruginosa
LPYLVRRLLENGANSSFVHK-RLVKGAYWDSEIKQC-GRGIFACVSPWNFPLAIF
LGQ-RDFEFQRLHGMGDALYDTVIEK-RRNVRIYAPVG-QFVLGRTISEALKNGRP
CREQG
>Ccrescentus
LPYLVRRLLENGANTSFVHA-RLVKGAYWDSEIKRA-GRGVFVCISPWNFPLAIF
TGQ-VKIEHQRLHGMGEALYKAADDL-GITLRAYAPVGMGEQFVLGRTIEAAIKR
AAAEG

>Bjaponicum
LAYLVRRLLENGANSSFVAQ-RLVKGAYWDTEIKRA-GRGAFVAISPWNFPLAIF
LGQ-GSFEFQRLHGMGEALYEQLAKD-DIAYRTYAPVG-HFVLGETIEQALERGKP
RSGQK
>Hpylori
IAYLVRRLDENTSSDNFMKA-RFVKGANMESEETIA-PKGVGVVIAPWNFPVGIS
VGT-EHFSFEMLEGMSLQASQELKEM-AGNRVIYKPSS-TFSSKQDTDSNYNKMLD
FVLEG
>Hsapiens
LPYLSRRALENSSLMKGTHR-KLVRGAYLAQERARA-ARKIGIMVASHNEDTVR
FAIQ-HQVYFGQLLGMCDQISFPLGQA-GYPVYKYVPYG-HFVAGEDQESIQPLLR
HYRAFG
>Mmusculus
LPYLSRRALENSSIMKGAQR-KLVRRAYMAQERVRA-STKAEVMVASHNEDTV
HFTLC-GQVCFGQLLGMCDQISFPLGQA-GFPVYKYVPYG-HFVAGEDQESIRPLIR
HNKAFG
>Cjejuni
IAYLVRRLDENTSEDNFMRY-RFVKGANMESEETIA-PKGIGVTIAPWNFPIGISV
GT-DSFTFEMLEGMSLQCSYELSKM-AGNVVIYKPSS-DFILAQNRVWAKEIKTKY
ENLK
>Bsubtilis
FSYFMRRIAERPANAAFVLK-RLVKGAYKESAAVAF-LSGNYTAVATHDDDIIK
FTKQ-SQFEFQMLYGIRPERQKELAKE-GYRMRVYVPYG-RFVAGDTIESAVKTVK
RLNRSG
>Mtuberculosis
YGYFLRRLAERPANLAFFLR-RLCKGAYDEPASVAY-RVGAWVTVDAEDHTTT
DSTLS-GDFEYQMLYGVRDDEQRRLTGA-GNHVRVYVPFG-RFVPGDTLDDVVDI
VTALRDSG
>Bhalodurans
FGYFMRRLAERPQNVAFALR-RLVKGAYKESPEVAY-ELGIFVRIDMEDYGHCQ
QTLD-SQFEFQMLYGFRTNLQEELVKE-GYNVRIYVPFG-QVVAGVTISEAIEKVK
ELNEKG
>Dmelanogaster
LPYLSRRAQENKGVLKKIKK-KLVRGAYMDQERDRA-ARKIGIMVASHNEDTVR
FAIQ-KVICFGQLLGMCDYITFPLGQA-GYSAYKYIPYG-HFVAGEDQIKIIPTLER
LRSFG
>Lmajor
IHYLGRRAVENSSILTTGDN-KIVRGAYIVQERATA-TYSILVKKSFYNYFCAGE
NDQ-SRVSFGQLFGMRDNLTVPLARA-GFQVYKYVPYG-QTFYQLAIDAIVATLQ
KTYNTE
>Tbrucei
VHYLGRRAMENASILSNGGS-KIVRGAYMRQERETA-DPSIFRQKSLWWLLRAL
FVLR-GYVAFAQLYGMSDNLTIPLKRA-GFPVFKYVPYG-HFCAGESDREVKNTVK
SLENLG
>Athaliana
IPYLLRRAYENRGMMATGAH-KLVRGAYMSSEASLA-GSGFGVVLATHNADSG
RLASR-GKIEFAQLYGMSDALSFGLKRA-GFNVSKYMPFG-HFCAGEDADAAAER
VRSVYEAT
>Scerevisiae
KDYLLRRLQENGDAVRSDNG-KLVRGAYIHSEKNRN-SSKLISCVGTWQLYLRD
SGDH-SNIVLGQLLGMADNVTYDLITN-AKNIIKYVPWG-LYCGGENFKEVIECGK
RLQKRG
>Anidulans
MGFLLRRAVENTEAVGRTKQ-KLVRGAYLKTEPRHL-EQAVQPGIEEWATMYQ
KYCNS-GDFVALKFTGMGIQALEYLQNQ-LVKHTIYKQFN-AGENKLEVQRSINAI
KELGYRG
>Paerophilum
FTSWINRVLAKTSIDFYREV-ILPKGVYMRPEPGEG-ARNVILATGAWTERLMDA
LGF-AAYHIRRLSPHSKVLVVDQNPG-GLADREYGPMF-PAVEGRVPQNAWAGHY
DENVVD

The conserved blocks generated by the MEME program.
These blocks were used in the phylogenetic analysis (Trees 3 to 7)