https://theses.gla.ac.uk/

Theses Digitisation:

https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

The study of small GTP - binding proteins in *Trypanosoma brucei*

Mercy Patricia Kordai Sowa

PhD

Wellcome Unit of Molecular Parasitology,

University of Glasgow.

3RD February, 1997

ProQuest Number: 10992110

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.

ProQuest 10992110

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

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

Theis
10694
Copy 1

GLASGOW
UNIVERSITY
LIBRARY

Abstract

Very little is known about cellular signalling pathways involved in the regulation of growth in trypanosomes. In higher eukaryotes, members of the *ras* GTPase superfamily of proteins are involved in signalling pathways which regulate cellular growth and differentiation. *Ras* genes have been well conserved through evolution and can be found in a range of organisms, such as the slime moulds, yeasts and man. The aims of this thesis are to investigate *ras* subfamily GTPases in *T. brucei* and to develop tools necessary for the study of the role of these monomeric GTP binding proteins in trypanosomal growth regulation.

Two approaches were taken to identify monomeric GTP binding proteins in *T. brucei*. The first approach involved the identification of GTP binding proteins in trypanosomal detergent extracts, and the second, the cloning of a *ras* subfamily gene from *T. brucei*. Monomeric GTP binding proteins are normally found within a molecular weight range of 20-29 kDa and, using the *in situ* GTP binding assay of Coulter and Hide (1995), GTP binding proteins within the same molecular weight range were identified in trypanosomal detergent extracts. A 24 kDa GTP binding protein was partially purified from trypanosome extracts by HPLC fractionation, although this protein was in low abundance. A 681 bp gene encoding a *ras* like GTP binding protein (*tbrlp*) was cloned using degenerate primer PCR. It was expressed in both bloodstream and procyclic forms of *T. brucei*. Southern blot analysis indicated that *tbrlp* is a single copy gene, however, preliminary sequence data derived from a 1.4 Kb PCR product amplified from *T. brucei* genomic DNA using *tbrlp* specific 3' and 5' oligonucleotide primers, indicated the presence of at least one other tandemly linked copy of *tbrlp*. A restriction enzyme map constructed from Southern blot data is consistent with the other copy being located upstream of the originally identified *tbrlp*. *Tbrlp* has all the essential features of a *ras* subfamily GTPase, namely, a GTP binding domain containing three phosphate/magnesium (PM 1, 2 & 3) and three guanine nucleotide (G1, 2 & 3) binding sites, a C-terminal extension domain which is the variable region of a *ras* gene, a C-terminal CAAX box (where C - cysteine, A - aliphatic amino acid and X - any amino acid) which is essential for post translational protein modification and a conserved C-terminal cAMP dependent protein kinase phosphorylation site. *Tbrlp* has 45 and 41 % amino acid identity to human *rap* 2 and H-*ras* respectively, but the regions of identity are interspersed throughout the gene, suggesting that it may be an ancestral *ras/rap* gene. Evolutionary analysis of *tbrlp* confirmed that this is indeed the case.

Tools and reagents such as *tbrlp* specific anti-peptide sera and a fusion protein were developed for future studies on the role of the GTPase in trypanosomal growth regulation. The

ultimate goals of this project are to acquire a better understanding of signal transduction in *T. brucei* and explore new avenues of drug discovery.

Dedication

To my family.

Thanks for everything.

Acknowledgements

First and foremost, I would like to thank Dr Geoff Hide, my supervisor, for his guidance, support and encouragement over the last three years. I would also like to thank Professor Andy Tait for his advice and helpful discussions. I would like to acknowledge the University of Glasgow for their financial support during the course of this work. I would also like to acknowledge members of the Wellcome Unit of Parasitology with whom I have enjoyed working. Special thanks to Ms Annette Macleod for her technical advice and companionship during the course of my stay in Glasgow. I would also like to thank all my friends for their encouragement. Last of all, but not the least, I would like to express my gratitude to my family, (the Sowas and the Colliers) to whom I dedicate this thesis, for their love and support both financial and spiritual over the last three years, I do not think I could have undertaken this PhD without their encouragement.

Declaration

This thesis and the results presented in it are entirely my own work,

except where indicated.

Table of contents

Chapter 2

Materials and Methods

Chapter 3

Detection of monomeric GTP binding proteins in *Trypanosoma brucei*

Chapter 4

Cloning sequencing and identification of *tbrlp*, a *ras* like small GTPase gene

Chapter 5

Evolutionary analysis of *tbrlp* I

Chapter 6

Development of tools for the functional analyses of *tbrlp* I

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Tables

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Abbreviations

bp	base pair
BSA	bovine serum albumin
CAAX	isoprenylation signal
cAMP	cyclic AMP
cDNA	complimentary deoxyribonucleic acid
cm	centimetre
DEAE	diethylaminoethyl
DFMO	difluoromethylornithine
DNA	deoxyribonucleic acid
EATRO	East African Trypanosomiasis Research Organisation
EDTA	ethylenediamine tetra-acetic acid (disodium salt)
GDP	guanosine diphosphate
GSP	gene specific primer
GTP	guanosine triphosphate
GTPBD	guanosine triphosphate binding domain
HPLC	high performance liquid chromatography
IL-1	interleukin-1
Kb	kilobase
kDa	kilodalton
M	Molar
MAPK	mitogen activated protein kinase
MCS	multiple cloning site
min	minute
ml	millilitre
mM	Millimolar
mm	millimetre
µm	micrometre
µl	microlitre
µg	microgram
mRNA	messenger RNA
ng	nanogram
NP-40	Nonidet P-40

OD	optical density
ODC	ornithine decarboxylase
ORF	open reading frame
OTU	operational taxonomic unit
PARP	procyclic acidic repetitive protein
PBSG	phosphate buffered saline containing 1 % glucose
PCR	polymerase chain reaction
Pfu	*Pyrococcus furiosus*
poly(A)+	polyadenylated
PVDF	polyvinylidene difluoride
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
SDE	spun detergent extracts
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SL	spliced leader
SSC	Saline sodium citrate
STIB	Swiss Tropical Institute Basel
Taq	*Thermus aquaticus*
TEMED	tetra-methyl-1,2-diaminoethane
TREU	Trypanosomiasis Research Edinburgh University
Tris	Tris (hydroxymethyl) amino methane
UV	Ultraviolet
V	volt
VSG	Variant surface glycoprotein
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Chapter 1

Introduction

CHAPTER 1

THE STUDY OF SMALL GTP BINDING PROTEINS FROM *Trypanosoma brucei*

1.0 Introduction

Trypanosoma brucei, a digenetic parasite belonging to the group Salivaria, is responsible for the morbidity and mortality observed in both man and animals in a number of countries found within the tsetse belt of sub-Saharan Africa. *T. brucei* is transmitted by haematophagous tsetse flies belonging to the genus *Glossina* and is responsible for either the West and Central (Gambian), or East and Southern (Rhodesian) African form of human sleeping sickness and Nagana in cattle. Trypanosomes fall into two different groups: the Stercoraria and the Salivaria. There are two main differences between these groups. Members of the group Stercoraria, which are responsible for South American Trypanosomiasis, develop in the hind gut of their vector host and transmission occurs via faecal contamination while members of the group Salivaria, develop in the mid gut of the vector and transmission occurs via the inoculation of the parasites into the mammalian host by the vector during a blood meal.

Figure 1.0 is an illustration of the classification of members of the group Salivaria (Hoare, 1972), with a few modifications, due to the isolation of one novel subspecies within the subgenus *Nannomonas* (Majiwa *et al*, 1993; McNamara *et al*, 1994; Garside and Gibson, 1995). There is currently a controversy on the status of *T. b. gambiense* (Gibson, 1986) and *T. b. rhodesiense* (Hide *et al*, 1996). This controversy has not yet been resolved and therefore only the classification as proposed by Hoare, 1972, will be considered. Members of this group, with a few exceptions, are transmitted via the saliva of the tsetse fly. Transmission normally occurs when the vector has a blood meal from a mammalian host. The exceptions are *T. equiperdum*, where the disease is sexually transmitted, and *T. evansi* and *T. vivax*, both of which can be transmitted mechanically by biting flies, such as tabanids, which play no role in parasite development and differentiation (Dirie *et al*, 1992).

FIGURE 1.0

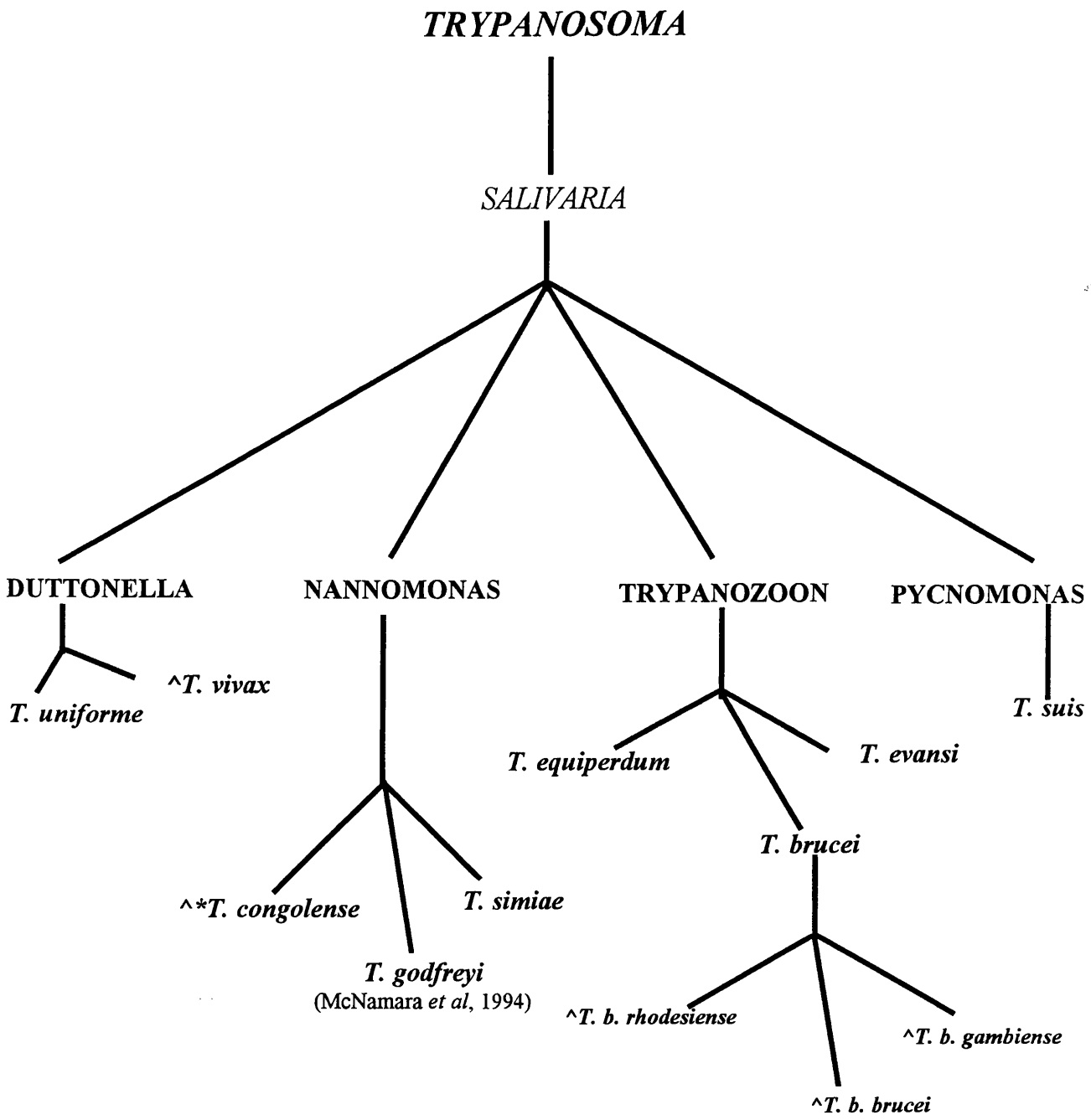


Figure 1.0 is an illustration of the classification of members of the Group Salivaria (Hoare, 1972), with a few modifications. The subspecies belonging to the subgenus *Trypanozoon* are responsible for East and West African Trypanosomiasis in humans and Nagana in cattle. * indicates that the subspecies *T. congolense* can be divided into 4 subgroups (Majiwa *et al*, 1993; Garside and Gibson, 1995 and all references therein). ^ - species or subspecies that cause Nagana in cattle.

Along with the other trypanosomatids, members of the species *T. brucei* possess a unique structure, the kinetoplast, an organelle which houses their mitochondrial DNA; this feature accounts for their classification as members of the Order Kinetoplastida. The three members of this species (*T. brucei*), *T. b. rhodesiense*, *T. b. gambiense* and *T. b. brucei* are the causative agents of the East and Southern, West and Central African human trypanosomiasis and Nagana in cattle respectively. The three subspecies are morphologically indistinguishable: *T. b. rhodesiense* and *T. b. gambiense* can be distinguished from each other by means of isoenzyme electrophoresis (Gibson *et al*, 1980; Tait *et al*, 1984) and molecular markers (Hide *et al*, 1990; reviewed by Hide and Tait, 1991). *T. b. brucei*, which is also known to infect goats and sheep, is susceptible to destruction by cytotoxic factors present in normal human serum. Both *T. b. rhodesiense* and *T. b. gambiense* are resistant to destruction by these components (Hajduk *et al*, 1994) although there is a controversy as to whether the main lytic factors are high density lipoproteins (Rifkin, 1978) or less well defined, high molecular weight, non-lipoprotein factors (Raper *et al*, 1996; Tomlinson and Raper, 1996).

Members of the species *T. brucei* undergo a number of host dependent adaptive changes which ensure both survival and proliferation of the parasites within the host. The life cycle of *T. brucei*, together with some of the adaptive changes used by the parasite in the evasion of the hosts immune response, are discussed below (Section 1.1).

1.1 The life cycle

Changes in the genetic, physiological and physical structure of the trypanosome are a result of some of the adaptations of trypanosomes to either the mammalian or insect host e.g the repression and derepression of the mitochondria when the trypanosome migrates from the tsetse fly to the mammalian host and vice-versa (Vickerman, 1985; Vickerman *et al*, 1988; Hajduk *et al* 1992). Some of these changes are found to be linked to the ability of the parasite to successfully evade immune destruction within its mammalian host. Figure 1.1 is a simplified illustration of the life cycle of *T. brucei*.

FIGURE 1.1

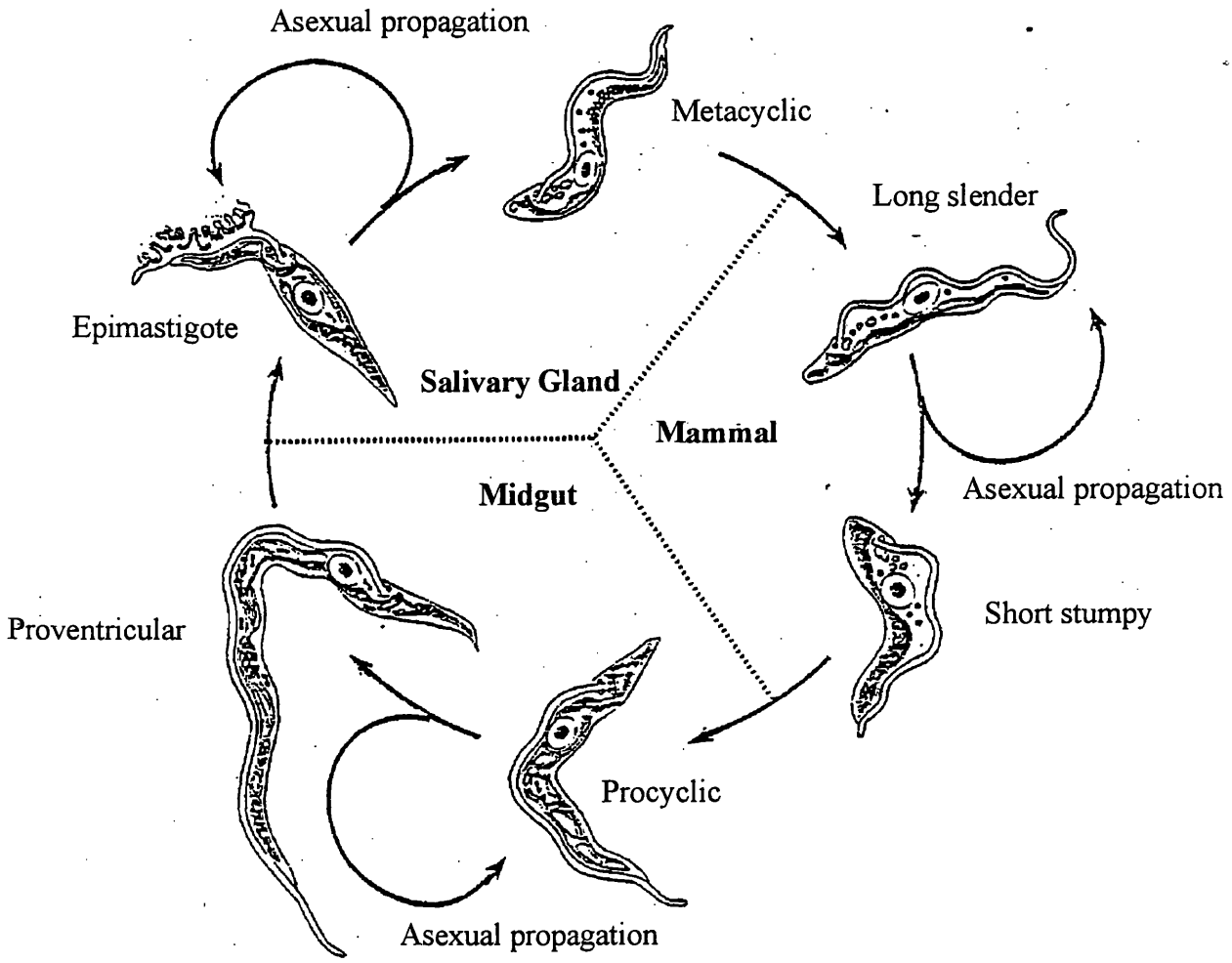


Figure 1.1 is an illustration of the different life cycle stages of *T.brucei*. Each stage is accompanied by physical and metabolic changes which enhance the parasites chances of survival in the host. The final life cycle stages in either host are preadapted for the anticipated environmental switch (adapted from Vickerman, 1985).

1.11 *Developmental changes within the tsetse fly*

The duration of the life cycle of the trypanosome within its vector host (i.e the tsetse fly) is approximately 3 - 5 weeks. Results from experiments in which flies were fed on trypanosome infected blood indicated that only 2 - 5 % of flies that ingest short stumpy trypanosomes become infected (Vickerman, 1985). The life cycle of the trypanosome within the tsetse fly can be divided into two stages: the midgut and the salivary gland stages.

Flies of the genus *Glossina*, specifically (*G. morsitans*), are highly susceptible to infection by trypanosomes inspite of the presence of a lectin defence system in their midguts (Maudlin and Welburn, 1987). Chitinases produced by rickettsia - like symbionts in the midgut release a sugar which appears to bind and, in effect, neutralise the action of lectin. This results in a breakdown in the defence system and the establishment of the infection within the fly (Maudlin and Welburn, 1987; Vickerman *et al*, 1988).

On entering the fly the trypanosome undergoes a number of morphological, genetic and metabolic changes. The short stumpy forms almost immediately shed their Variable Surface Glycoprotein (VSG) coat and replace it with a procyclic acidic repetitive protein (PARP) during the transformation to procyclic forms. The trypanosome uses its VSGs for the evasion of the mammalian hosts immune response (discussed in section 1.41), but does not require the use of antigenic variation for survival in the tsetse fly. The actual function of PARP is presently unknown (Hajduk *et al*, 1992). The following changes to the trypanosome occur in the midgut:

1) An overall increase in body length caused by the elongation of the post kinetoplastic portion of the body (Vickerman, 1985; Vickerman *et al*, 1988).

2) Termination of receptor mediated endocytosis in the flagellar pocket (Vickerman, 1985).

3) Activation of the mitochondrion and reduction in the use of glycosomes (Vickerman, 1985; Vickerman *et al*, 1988). These changes occur concurrently with a change in energy source from glucose to proline. The tubular cristae of the mitochondrion are replaced by a network of discoid cristae in order to accommodate the energy source switch. This results in an overall increase in volume of 5 - 25 %. The marked reduction in the use of glycosomes results in their morphology altering to yield bacilliform like structures.

The procyclic forms in the midgut penetrate the peritrophic membrane, migrate to the ectoperitrophic space and then into the salivary glands, where an increase in the rate of division is

observed. There are four stages in the life cycle of the trypanosome on migration to the salivary gland:

1) Epimastigote stage:
The trypanosome is attached to the epithelial lining of the gland by flagellipodia. At this stage it has no variable surface glycoprotein coat and undergoes an elevated level of cell division.

2) Trypomastigote stage:
This stage is also referred to as the pre-metacyclic stage. Proline is still used as an energy source and they are still attached to the epithelial cell lining.

3) Nascent metacyclics:
The trypanosomes acquire a surface coat, cease to divide and the mitochondrion and glycosomes become preadapted to the anticipated environmental switch (i.e.unbranched and spherical respectively). They, however, remain attached to the epithelial cell lining (Vickerman, 1985).

4) Mature or "free - swimming" metacyclics:
At this stage the trypanosomes are fully preadapted to life in the mammalian host. The trypanosome acquires a surface coat with a limited number of metacyclic variant antigenic types (M - VATs). These M - VATs appear to be predetermined at an earlier stage in the salivary gland stage in the life cycle of the trypanosome (Barry, 1989; Graham and Barry, 1995) and serve as a first line of defence on entering the mammalian host. No cell division occurs at this stage.

1.12 *Developmental changes in the mammalian host.*

Tsetse flies are haematophagous arthropods. It normally takes approximately 4 minutes for coagulation to occur at the site of tissue damage and less than 4 minutes for a fly to complete the uptake of a blood meal. The possible coagulation of the blood during the blood meal is, therefore, not a problem. However, an anti - coagulant is necessary, in order to ensure complete digestion of the blood meal and an unblocked proboscis for subsequent meals. An anti - thrombin has recently been isolated from the salivary glands of flies of the species *G. morsitans* and *G. austeni* (Stark and James, 1996). Therefore, while feeding, the tsetse fly inoculates the mammalian host with saliva containing preadapted trypanosomes. These trypanosomes enter into

the draining lymphatics and migrate into the bloodstream. The bloodstream stage is referred to as "early stage" trypanosomiasis. From the bloodstream they migrate to the brain and cerebro-spinal fluid (CSF), via the blood and lymph capillary walls, connective tissue and the choriod plexus. Late stage trypanosomiasis is the final stage of the disease. At this stage trypanosomes can be isolated from the CSF of infected patients.

In the mammalian host, the long slender trypanosomes have a doubling time of 6 hours and divide by binary fission. The levels of parasitaemia in an infected patient fluctuate over a time period. This is found to be due to the transformation of the long slender types to non - dividing short stumpy forms which are preadapted to life in the vector host (Pays, 1988). The trypanosome is covered by a "surface coat" composed of a homogeneous layer of a single type of glycoprotein - the VSG. The ability to change the VSG type coating the surface of the trypanosome results in a phenomenon known as VSG switching or antigenic variation. There are approximately 10^{-2} - 10^{-3} switches/trypanosome generation. The waves of parasitaemia observed are caused by the rapid doubling time, VSG switching and subsequent destruction of previous antigenic types by the host immune system. There is also the successive replacement of the original VSG by a series of immunologically distinct VSGs. This allows the trypanosome to be one step ahead of the host immune response. By the time the host is prepared to eliminate one generation of parasites, the next generation has already been established.

In the mammalian host, the mitochondria are repressed and the energy source is switched from proline to glucose. At this stage the glycosome, a single membrane bound organelle which contains the first nine enzymes of the glycolytic cycle, is activated and supplies the trypanosome with energy (Opperdoes and Borst, 1977).

All forms of trypanosomes within the mammalian host utilise receptor mediated endocytosis via the flagellar pocket for the uptake of larger molecules (e.g low density lipoproteins and transferrin).

1.2 Human African Trypanosomiasis or HAT

Within the mammalian host, trypanosomes are the causative agents of Human African Trypanosomiasis (HAT) or African sleeping sickness and Nagana in cattle. There are two forms of HAT; the chronic, slow developing, West and Central African form caused by *Trypanosoma brucei gambiense* and the more acute, East and Southern African form, caused by *Trypanosoma brucei rhodesiense.* The geographical distribution of these subspecies is not as rigid as the names of the different forms of the disease would have it seem. A number of cases of *gambiense*

- like origin have been found in East and Southern Africa and similarly cases of *rhodesiense* - like disease have been found in West and Central Africa (Kuzoe, 1993; Pepin and Milord, 1994). The West African form of the disease appears to have two distinct phases. The initial phase is characterised by symptoms such as headaches, fever and lymphadenopathy and is known as the haemolymphatic stage (Jernigan & Pearson 1993). This phase produces symptoms characteristic of a number of tropical diseases such as malaria and dysentery and, therefore, patients usually only seek medical assistance after the onset of the second/late phase. The late phase is characterised by severe neurological disturbance, irregular sleep patterns and finally, a deep coma which leads to death. *T. b. gambiense* infections can last for several years before resulting in the death of the patient (TDR 12th Programme Report). *T. b. rhodesiense* infections, however, have a much shorter course and the patient often dies a few months to a year after the initial infection.

Approximately 55 million people in 36 sub-Saharan countries (Figure 1.2) are at risk from HAT (Kuzoe, 1993) but only 25,000 cases are reported annually. Massive epidemics are known to occur, and during such periods a large number of cases are recorded. An example, is the epidemic that occurred in the Busoga region of Uganda, which spans an area of approximately 130,000 Km^2 (Abaru, 1985). During 1980, as many as 8000 cases were officially recorded in this region (Koerner *et al*, 1995), with a monthly incidence of approximately 1000 cases being reported at the peak of the epidemic (Abaru, 1985). The number of cases reported, however, appears to be a gross underestimation as many infected cases go undiagnosed and untreated. The disease almost exclusively affects inhabitants of rural areas, and as there are only occasional outbreaks, HAT is not considered to be a disease that requires the expense of regular surveillance (TDR news, 1994).

The lack of regular surveillance and, consequently, early diagnosis and treatment of the disease are only a few of the numerous problems encountered in the bid to control the spread of African Trypanosomiasis. The parasite has the ability to effectively evade the hosts immune response and it would appear that the use of chemotherapy is the only practical way to treat and control the spread of the disease. There are currently no easy methods for the diagnosis of a current infection of HAT or a previously cured infection and, when finally diagnosed, the drugs available for the treatment of the disease are either extremely toxic with several unpleasant side effects or too expensive to be used regularly (e.g DFMO costs approximately $500.00 per patient). Even in the case of DFMO, which has been referred to in the past as the resurrection drug, not only are there a few unpleasant side effects but it has also been found to be ineffective in the treatment of trypanosomiasis caused by *T. b. rhodesiense* (Kuzoe, 1993; Pepin and Milord, 1994; Chan and Fong, 1994). The following section is a summary of diagnostic kits and chemotherapy presently

FIGURE 1.2

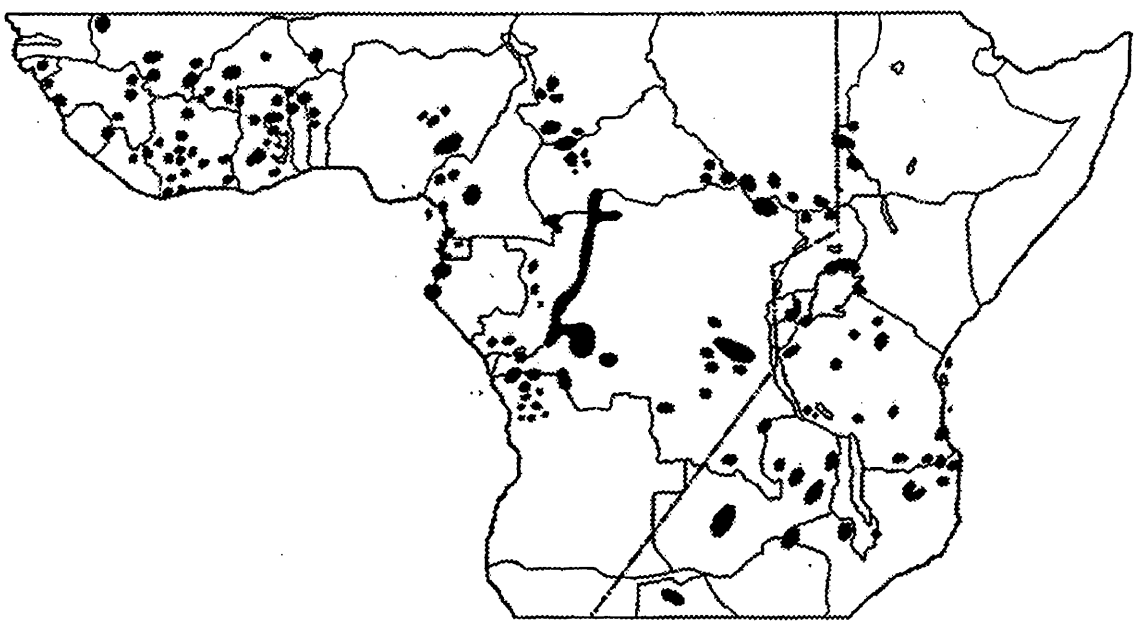


Figure 1.2 is a map showing the distribution of trypanosomiasis foci in subsaharan Africa. *Trypanosoma brucei gambiense* infected areas are to the left of the dotted line and *Trypanosoma brucei rhodesiense* infected areas, to the right. The geographical boundary is not rigid, patients infected with *T. b. gambiense* have been found in *T. b. rhodesiense* infected areas and vice versa. (Photocopied from Manson's Tropical Diseases, 1996)

1.3 Diagnosis, treatment and control of the spread of HAT.

The link between the incidence of HAT and its transmission by the tsetse fly, was first recorded in 1897 by Sir David Bruce. Approximately 30 years after, a number of drugs with trypanocidal activity were discovered and are presently still in use (Wang, 1995). The drugs used in the treatment of HAT are extremely toxic and produce very unpleasant side effects (Kuzoe, 1993; Pepin and Milord, 1994). Problems are encountered in the diagnosis of the disease as the initial symptoms of HAT can easily be confused with those of other tropical diseases (WHO, 1994). Initially the diagnosis of trypanosomiasis was dependent on the detection of parasites in wet blood films. Parasites are more readily detectable in stained thick blood films but can go undetected due to the low parasite numbers in the blood. New techniques involving the concentration of the parasites by either centrifugation or ion-exchange chromatography of the infected blood have since been developed. These techniques are more sensitive and are currently being used in the diagnosis of HAT, but there is still the need to develop more sensitive diagnostic kits that can be used effectively in the field (reviewed by Nantulya, 1991).
The only reliable form of control appears to be the use of tsetse fly traps and avoidance of known HAT foci (WHO, 1994).

1.31 *Kits currently available for the diagnosis of HAT*

The key to proper management and control of the spread of a disease lies in making a proper diagnosis. Diagnostic kits should therefore be reliable, extremely sensitive, capable of being used in the field and, in the case of HAT, should be able to determine which species is causing the infection. In addition to these requirements, a kit should be simple, produce results promptly and, most of all, be cheap. These kits should be able to supply medical staff with enough information to make a proper diagnosis at the onset of the disease. Unfortunately, there are presently no kits with all these specifications for the diagnosis of HAT (reviewed by Nantulya, 1991).

Members of the species *T. brucei*, especially *T. b. gambiense*, are basically tissue dwellers. On inoculation into the mammalian host, the parasites are sequestered in various host tissues such as the liver, spleen, heart, kidneys and lymph nodes (Nantulya *et al*, 1992). The sequestration of the parasites in the tissues, therefore, leaves the blood with very few or no parasites in the early stages of the disease; normal procedures whereby the presence of trypanosomes in the blood could be used in confirming a diagnosis cannot therefore be used.

Most of the kits presently available rely on the ability to either show the presence of trypanosomes within blood, lymph or the CSF or to detect antibodies generated against an infection. Examples of such kits are the card agglutination test or CATT which is routinely used in the diagnosis of trypanosomiasis; the mini anion exchange column or MAEC; the microhaematocrit centrifugation technique or HCT; the quantitative buffy coat technique (QBC) and the use of ELISA in the detection of *T. b. rhodesiense* procyclic trypomastigote antigens. All these kits do detect the presence of trypanosomes in the blood but are prone to producing false negative results, due to the fluctuating aparasitaemic phases observed in both chronic and acute infections (Molyneux *et al*, 1996). The antibody based kits cannot be used to differentiate a current infection from one that has already been treated and cured (Nantulya *et al*, 1992; Komba *et al*, 1992; Nantulya *et al*, 1992; Bailey and Smith, 1992).

KIVI or kit for isolating African trypanosomes *in vitro* is one of the kits used for the isolation of trypanosomes in the field and can also be used to confirm a diagnosis obtained by other methods. The kit is known to have produced positive results, in several cases, where other routinely used diagnostic kits have produced negative results e.g CATT (Truc *et al*, 1992). Unfortunately one of the main limitations of the use of KIVI routinely is the length of time needed to establish the absence or presence of trypanosomes in a patient. The average time taken for the test to show a positive result is approximately 7 - 14 days. Therefore a patient in the early stages of the disease showing negative results with the other tests, might not be treated, thus giving the parasites the opportunity to establish themselves within their new host. Another limitation is the lack of adequate culture facilities.

In view of these limitations it appears as though there is a need for rapid, more sensitive, methods of detection of trypanosomiasis during the early stages and aparasitaemic periods of the infection. This would be quite useful as early detection could help to control the morbidity and spread of the disease.

1.32 *Treatment currently available for HAT*

Four anti - HAT drugs are available, three of which have been used for over 60 years. The actual mode of action of these drugs, with the exception of DFMO (the most recently discovered), is not unequivocally known. Most of them are found to be extremely toxic to the mammalian host and produce several side effect ranging from vomiting to severe neurological disorders and, in 1 - 5 % of melarsoprol treated patients, death (Kuzoe, 1993; Pepin and Milord, 1994). A number of strains of *T. brucei* have also developed resistance to the currently used anti

- trypanosomal drugs. Due to these problems, there is a need for the production of new and less toxic drugs for the treatment of HAT. The following sections provide a brief summary of the drugs currently available and their limitations.

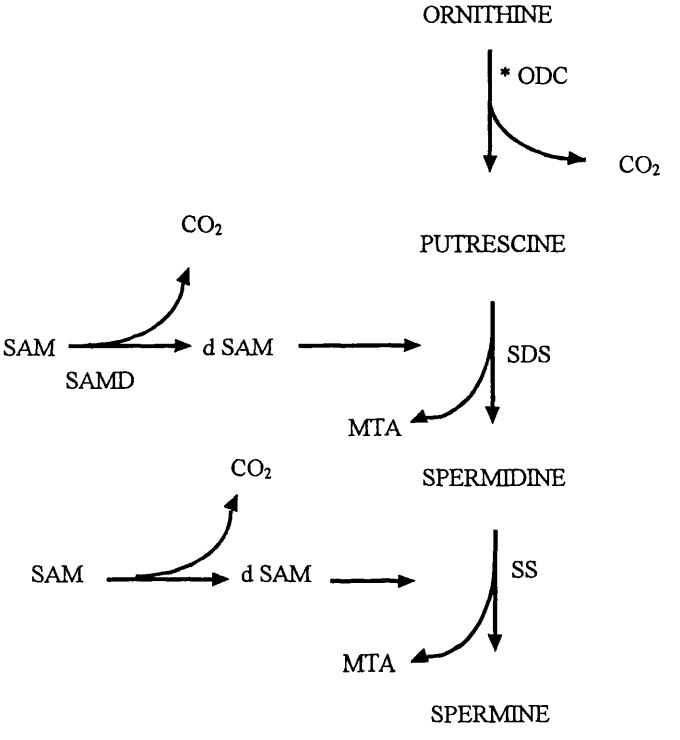
Suramin

This drug has been in use since 1922. It is a polyanionic naphthylamine sulphonate which is highly negatively charged at physiological pH. It is soluble in water but is poorly absorbed and is therefore administered by either intravenous or intramuscular injection. Suramin is only effective during the haemolymphatic/early stage of HAT before the sequestration of the parasites in the mammalian host organs. It cannot be used in the treatment of the late stage of the disease as it is unable to cross the blood-brain barrier. Suramin is used as a prophylactic and a curative drug in the treatment of both East and West African HAT. Very few strains of *T. brucei* are known to have developed a resistance to suramin (Wang, 1995). The actual mode of action of suramin is unknown. It appears to enter the trypanosome, bound to a variety of serum proteins, by receptor - mediated endocytosis via the flagellar pocket. On entering the trypanosome, its anti - trypanosomal activity seems to be directed against a number of enzymes e.g dihydrofolate reductase and thymidine kinase. Suramin inhibits their activity by binding to their active sites (Fairlamb and Bowman, 1980; Wang, 1995). Suramin also acts on enzymes of the glycolytic pathway resulting in a gradual decrease in ATP production by the parasites and, eventually, parasite death (Fairlamb and Bowman, 1980). A number of studies have reported a fairly high relapse rate after treatment with suramin.

Melarsoprol

An arsenical, introduced in 1949, is insoluble in water and, therefore, is dissolved in propylene glycol. Until recently, it was the only drug used in the treatment of advanced/late - stage trypanosomiasis. It is administered intravenously over a period of four weeks. It is sometimes impossible to administer the drug due to solvent induced phlebitis. Other side effects include vomiting, encephalopathy and, in 1 - 5 % of treated cases, death (Wang, 1995). The exact mode of action of melarsoprol is currently not known. It however, appears to bind irreversibly to trypanothione (Simarro and Asumu, 1996). Studies have also shown that it blocks glycolysis by inhibiting trypanosomal phosphofructo-kinase and this results in cell lysis (Wang, 1995). There are presently several melarsoprol resistant strains of both *T. b. gambiense* and *T.*

FIGURE 1.3



A schematic illustration of the pathways involved in the biosynthesis of the polyamines putrescine, spermidine and spermine. These polyamines have been implicated in a number of functions in most eukaryotic cells including *T. brucei*. The polyamines are essential for normal cell growth and differentiation (Bacchi *et al*, 1983) * indicates the point of inhibition of the spermine biosynthesis pathway; ODC - ornithine decarboxylase; SDS - spermidine synthase; SS - spermine synthase; SAM - S - adenosylmethionine; d SAM - decarboxylated S - adenosylmethionine; SAMD - S - adenosylmethionine decarboxylase; MTA - methylthioadenosine.

The decarboxylation of ornithine in the presence of ornithine decarboxylase is the rate - limiting step in the production of putrescine from ornithine. An aminopropyl group produced by the decarboxylation of S - adenosylmethionine in the presence of S - adenosylmethionine decarboxylase, binds to putrescine to produce spermidine. This reaction is catalysed by spermidine synthase. Spermine is produced by the addition of another aminopropyl group to spermidine in the presence of spermine synthase. DFMO inhibits the initial step in the pathway i.e the decarboxylation of ornithine, thus decreasing the levels of polyamines produced (Bacchi *et al*, 1983).

Elevated levels of polyamines in cancer patients, have been implicated in tumour formation. Unfortunately the drug was ineffective in the reduction of tumours, but was found to have a trypanostatic effect on trypanosomes. Its action is found to induce the transformation of trypanosomes from long slender to short stumpy forms which can then be eliminated by the host immune response (Bacchi *et al*, 1983; Bitonti *et al*, 1986).

The mild toxicity of DFMO to the mammalian host is exactly what accounts for its failure as an anti -tumourigenic drug (Schechter and Sjoerdsma, 1986). Trypanosomes have a slow *in vivo* ornithine decarboxylase (ODC) turnover rate and, therefore, the irreversible inhibition of the enzyme by DFMO lasts for a longer period of time. However, the high turnover rate of mammalian ornithine decarboxylase leads to all the DFMO being rapidly bound to the enzyme, thus resulting in the lack of effective inhibition of the action of ornithine decarboxylase.

1.33 *Current control measures against the spread of HAT*

There are currently 23 known species of the genus *Glossina*, which are capable of serving as vectors of parasites responsible for African trypanosomiasis. Two of these are commonly known to be capable of transmitting the disease. *Trypanosoma brucei gambiense* is transmitted by *Glossina palpalis* and *T. b. rhodesiense* by *Glossina morsitans* (Molyneux *et al*, 1996). Flies of the group *G. palpalis* are normally found along rivers, lakes and forest regions, whereas those of the group *G. morsitans* are normally inhabit the savanna regions. One of the strategies employed for the control of the transmission of trypanosomiasis is by the use of fly traps. The earliest fly traps were actually designed for collecting flies for entomological studies. It was, however, observed that the reduction in the numbers of flies was closely linked to a reduction in the chances of being bitten by a fly. Based on this observation, several traps were designed for capturing tsetse flies, e.g the insecticide impregnated biconical trap produced in Burkina Faso and the use of attractants such as the natural odours of cattle in Zimbabwe. Trials

in Cote d'Ivoire and Burkina Faso showed that impregnated traps were highly successful in capturing flies.

The measures being used presently are the safest and cheapest methods available for the control of the spread of HAT. Traps designed for attracting and killing the vector host at a cost of $ 6.00 / trap are currently being used. These traps have been found to be quite effective in capturing flies and seropositivity levels decreased from 8.5 % to 1.7 % in one case study (TDR twelfth Programme Report, 1995). There is the need, however, for regular surveillance, which in addition to a cost of approximately $ 550.00 / sq km results in a fairly expensive process. If the traps are properly maintained, in addition to regular surveillance, it appears as though it would be a safer and cheaper mode of control of the spread of HAT compared to any of the alternatives.

Dipping of herds of cattle into insecticides, such as deltamethrin, is an effective and cheap way of killing flies (reviewed by Hide, 1994). However dipping might have residual effects on the consumers and, therefore, may not be the safest way to eliminate the vector host.

There is a need for drugs which are, cheaper, less toxic and have greater bioavailability for the treatment of HAT. Unfortunately, as the disease is predominant in some of the poorest countries in the developing world, the production of new drugs would need to be a non -profit making venture for pharmaceutical companies, and therefore, it is not likely that any money will be invested in research into new drugs solely for the treatment of HAT. Merrell Dow, the manufacturer of DFMO, is no longer interested in its production due to its failure as an anti - cancer agent and in the treatment of AIDS - related pneumocystis infections (Pepin and Milord, 1994; TDR Twelfth Programme Report, 1995). They are currently supplying WHO with the drug free of charge. Current trials are aimed at shorter durations of treatment (from 14 to 7 days) in order to reduce the cost of the drug.

The only way money will be invested in research into new drugs for the treatment of HAT is by finding a drug that can be used in the treatment of some other diseases, which are more "commercially viable". An example is pentamidine which was used in the treatment of early - stage HAT but has recently been used for the treatment of AIDS patients with opportunistic parasitic infections. As a result, the price was increased from $1 to $300 / 300mg, but is now supplied free where HAT is endemic.

In the long term there is a need to develop new drugs. One way of achieving this aim is by gaining a greater understanding of the mechanisms of immune evasion used by different pathogens and the signalling pathways involved in the regulation of growth and differentiation of the parasites.

1.4 Mechanisms for the evasion of the host immune response

The mammalian immune system has evolved to control levels of infecting pathogens. The suppression or modulation of the host immune response by the pathogen therefore enhances the establishment of an infection. Human pathogens have evolved a number of mechanisms designed to enhance their chances of survival in their host and thus ensuring their propagation. An in depth understanding of these mechanisms of immune evasion is the first step towards finding an effective means to control the disease.

A number of pathogens evade complete immune destruction by the manipulation of the hosts immune response. An example of such a pathogen is the schistosome, a parasitic worm of the genus *Schistosoma*. Adult worms and larval stages have the ability to employ the use of antigenic mimicry by covering their surface coat with host antigens, thereby protecting themselves from the host immune response (Abath and Werkhauser, 1996). Parasitic nematodes, some of which are responsible for a number of severe intestinal infections in humans e.g *Ascaris lumbricoides* and *Trichuris trichura*, are known to enhance their chances of survival in the host by modulating the hosts cytokine response (Grencis, 1996). Kinetoplastids such as parasites of the genus *Trypanosoma* and *Leishmania* also use the modulation of the host immune response to enhance growth (Barcinski and Costa-Moreira, 1994; Reiner, 1994). Infections by the causative agents of both HAT and Chagas disease result in the suppression of the host immune response (Sztein and Kierszenbaum, 1993). There appears to be an overall decrease in IL-2 in both cases but the mechanism used to achieve this, seems to depend on the infecting trypanosome (Sztein and Kierszenbaum, 1993). Other parasites known to use more elaborate methods of immune evasion are, e.g *Plasmodium falciparum* and *Babesia bovis*. These parasites evade the immune response by hiding within host erythrocytes during their reproductive stages. This ensures both propagation and evasion of the immune response by the parasites (reviewed by Allred, 1995). Members of the genus *Leishmania* are also known to employ this strategy. They reside in macrophages instead of erythrocytes (Jernigan and Pearson, 1993; Brittingham and Mosser, 1996).

Strategies which result in structural changes in the parasite, can leave the immune response one step behind. Examples of structural changes used in the evasion of the immune response include: shedding of the surface coat (e.g *Schistosoma mansoni*, Pearce *et al*, 1986) and antigenic variation, a process which involves the successive replacement of the surface coat of the parasite by immunologically distinct VSGs (e.g *T. brucei*, Vickerman, 1985; Barry, 1989). Antigenic variation or switching of the Variant Surface Glycoprotein (VSG) is the main mechanism for immune evasion in trypanosomes. This process appears to be mediated by a

number of possible mechanisms including gene rearrangements. A number of pathogens are known to exhibit antigenic diversity, in addition to several other evasive tactics. Antigenic diversity occurs as a result of an accumulation of point mutations in the genes that code for proteins recognised by the immune system (Borst, 1991). The reverse transcriptase of HIV has a high error rate, which results in the rapid accumulation of point mutations and, as a consequence of this, a large number of variants are found within a single host. The influenza viruses (Marrack and Kappler, 1994) employ the same strategy. Occasionally, the influenza virus evades the host immune response by mutation which occurs as a result of gene reassortments (Marrack and Kappler, 1994).

1.5 Putative drug targets

To date, most drugs have been discovered by large scale screens. There are currently 3 major screening centres for the discovery and development of new drugs for the treatment of leishmaniasis and the trypanosomiases (TDR, 1995). However, a better understanding of the biological mechanisms unique to the parasite and the aspects of the parasite that regulate its growth could lead to the discovery of new drug targets which could be used in the treatment of disease. A number of such novel mechanisms or features are known in trypanosomes (e.g antigenic variation, glycosomes and RNA editing). A number of aspects of parasite growth regulation which could be also be targeted (e.g cell signalling, proteases, cell cycle) are currently known. The following sections highlight some of the mechanisms used by trypanosomes in the evasion of the host immune response and some biological features which could be used as drug targets. Gene rearrangements in surface antigen genes, which occur as a result of gene conversion and recombination, are known to play an important role in enhancing a parasites chances of survival in the mammalian host (Tobin *et al*, 1991; Weiden *et al*, 1991). These rearrangements result in the ability of the trypanosome to undergo VSG switching to generate variable antigenic types. This VSG switching allows the trypanosome to avoid immune destruction via complement mediated lysis. VSGs are first synthesised during the final life cycle stages of the trypanosome in the salivary gland (Cross, 1990), are homogeneous within a single trypanosome and the surface coat is made up of approximately 10^7 glycoprotein molecules (Hadjuk *et al*, 1992; Hadjuk *et al*, 1994). They are attached to the plasma membrane by a glycosyphosphatidylinositol (GPI) anchor. The genome has a 1000 or more VSG genes (Van der Ploeg, 1984; Hadjuk, 1994; Graham, 1995; Vanhamme and Pays, 1995). The large number of

VSGs, the randomness of VSG switching and a doubling time of 6 hours (Barry, 1989) discounts the use of a cocktail of VSGs as a vaccination strategy.

Some important aspects of trypanosome biology lend themselves to further investigation in relation to developing methods of controlling trypanosome growth and developing new drugs. Below is a list of some of these aspects:

1) Procyclic acidic repetitive protein or PARP: this is the invariant glycoprotein surface layer which replaces the VSG coat after the insect is ingested into the vector host (Roditi *et al.*, 1989; Roditi and Pearson, 1990; Graham, 1995). The expression of this glycoprotein layer appears to be linked to the change in temperature from 37° to 26° C. The function of PARP is presently unknown but an understanding of the role of PARP in the tsetse/trypanosome interaction could lead to new methods of control. Current knowledge of the structure of PARP indicates that it could be used as a potentially good transmission blocking vaccine. Antibodies raised against PARP, could be ingested into the fly and thus block subsequent transmission by the destruction of procyclic forms. A transmission blocking vaccine will not prevent a primary trypanosome infection but in the long term could lead to a decrease in the transmission of the disease.

2) Differentiation: environmental changes may act as triggers for the initiation of differentiation of the parasite on ingestion into the vector or inoculation into the mammalian host. Factors which contribute to the initiation of differentiation whilst within the mammalian host (long slender to short stumpy forms) or in the vector (short stumpy to mature metacyclic forms), if any, remain to be discovered.

3) Molecular mechanisms involved in trypanosome differentiation.

4) Factors controlling gene rearrangements of VSGs and expression site selection.

5) The mechanism by which the trypanosome carries out receptor - mediated endocytosis via the flagellar pocket without evoking an immune response.

These aspects cover broad areas which need to be investigated, however, more specific investigations are currently underway for the development of drugs for the treatment of HAT. One of the promising areas is the use of the GPI anchor as a target for the production of chemotherapeutic drugs due to the presence of myristate in the anchor. The plasma membrane GPI anchor is found in both lower and higher eukaryotes but trypanosomes differ in that myristate is essential for trypanosome GPI anchor formation but it is not required for GPI anchor

formation in mammalian cells. Bloodstream form trypanosomes must acquire myristate from the host serum (Buxbaum *et al*, 1994 and all references therein). Analogues of myristate have been found to kill bloodstream but not procyclic trypanosomes (McConville and Ferguson, 1993; Buxbaum *et al*, 1996; Werbovetz *et al*, 1996). As myristate is a rarely used fatty acid in mammals, a drug designed to interfere with its synthesis within mammals could act specifically on bloodstream trypanosomes (Hadjuk, 1992).

A number of metabolic pathways are also being investigated to determine whether or not they can be used in the production of novel drugs for the treatment of trypanosomiasis. One of the pathways being investigated is the glycosomal pathway required for the conversion of host glucose to energy which is utilised by the trypanosome. The compartmentation of the glycolytic enzymes was initially attributed to the need to maintain a high rate of glycolysis for the production of energy required by bloodstream trypanosomes. This theory has since then been discarded as glycosomes have been discovered in both monogenetic, (e.g the *Bodonidae*) and other digenetic (e.g *Phytomonas*) kinetoplastids, the majority of which do not rely solely on glycolysis for their supply of energy (reviewed by Clayton *et al*, 1995; Clayton and Michels, 1996). Experiments have shown that the compartmentation of glycolysis is essential for normal growth of bloodstream trypanosomes, and, has a different pattern of regulation from its mammalian host. It therefore, could be exploited in the search for chemotherapeutic drug targets (reviewed by Clayton *et al*, 1995; Clayton and Michels, 1996).

The ornithine pathway for the biosynthesis of polyamines (such as spermine and spermidine) is one of the metabolic pathways which has currently provided the greatest breakthrough, in the quest for trypanocidal drugs. Difluoromethylornithine or DFMO, one of the drugs presently being used in the treatment of HAT, irreversibly inhibits the action of the enzyme ornithine decarboxylase (Schechter and Sjoerdsma, 1986; Bellafatto *et al*, 1987). This enzyme is essential for the decarboxylation of ornithine which is essential for the synthesis of polyamines. Polyamines are thought to play an important role in cellular proliferation and differentiation in most eukaryotic cells (Schechter and Sjoerdsma, 1986; Henderson and Fairlamb, 1987).

Another pathway which is presently being studied is the trypanothione pathway. Most prokaryotic and eukaryotic cells have glutathione (GSH) which has an important role in a number of physiological processes in the cell e.g amino acid transport, disulphide reduction and peroxide metabolism (Henderson and Fairlamb, 1987) just to name a few. Mammalian cells have glutathione reductase which is involved in the reduction of glutathione. This is an important step in several physiological pathways. Trypanosomes have the ability to reduce glutathione but the reduction is carried out by the trypanosome equivalent of the mammalian glutathione reductase

known as trypanothione reductase. Studies have shown that the two reductases have different substrate specificities and, therefore, a trypanocidal drug, which could inhibit one of the substrates of the trypanothione pathway in trypanosomes, would not interfere with the glutathione pathway in the mammalian host. Exactly what trypanothione does within the trypanosome is presently unknown but, if it plays an essential role in cell growth and differentiation, it might be a useful chemotherapeutic target to exploit.

Information gathered from such studies emphasise the need to acquire a greater knowledge of the metabolic pathways used by these pathogens. Inspite of the fact that trypanosomes appear to be the most studied protozoan parasites there is still not extensive knowledge of their metabolic pathways. The knowledge acquired from two of the pathways (the trypanothione pathway and S-adenosylmethionine pathway for the polyamine synthesis) seems to have provided investigators with enough information to enable them to investigate the possibility of producing trypanocidal drugs which could interfere with the parasites normal cellular function (Henderson and Fairlamb, 1987; Bacchi *et al*, 1996). Further studies of their metabolic pathways, could shed light on strategic points at which the pathways could be blocked and result in the inhibition of normal cellular growth and differentiation as has been shown by DFMO. This might lead to the discovery of novel ways of controlling the spread of HAT or at least the discovery of new and cheaper drugs with less toxic side effects.

1.6 Evolutionary relationships

One of the oldest problems, encountered in the study of the evolution of organisms, is choosing the correct homologous characters for comparative purposes. Until quite recently, evolutionary classifications were based on morphological characters and behavioural patterns. This led to many cases where the classification of organisms has been incorrectly assigned. New molecular methods have revolutionised these issues as the phylogenetic "history" of organisms can be reconstructed from their DNA and amino acid sequences. The use of molecular data as a measure of divergence reflects the genetic relationships between organisms (inherited features) rather than the epigenetic relationships (features inherited as a result of morphological constraint, environment or behaviour). Genetic variability between organisms occurs as a result of one or a combination of a number of changes that take place at the DNA level. The four basic types of changes (mutations) that can occur are 1) substitutions; 2) deletions; 3) insertions and 4) inversions.

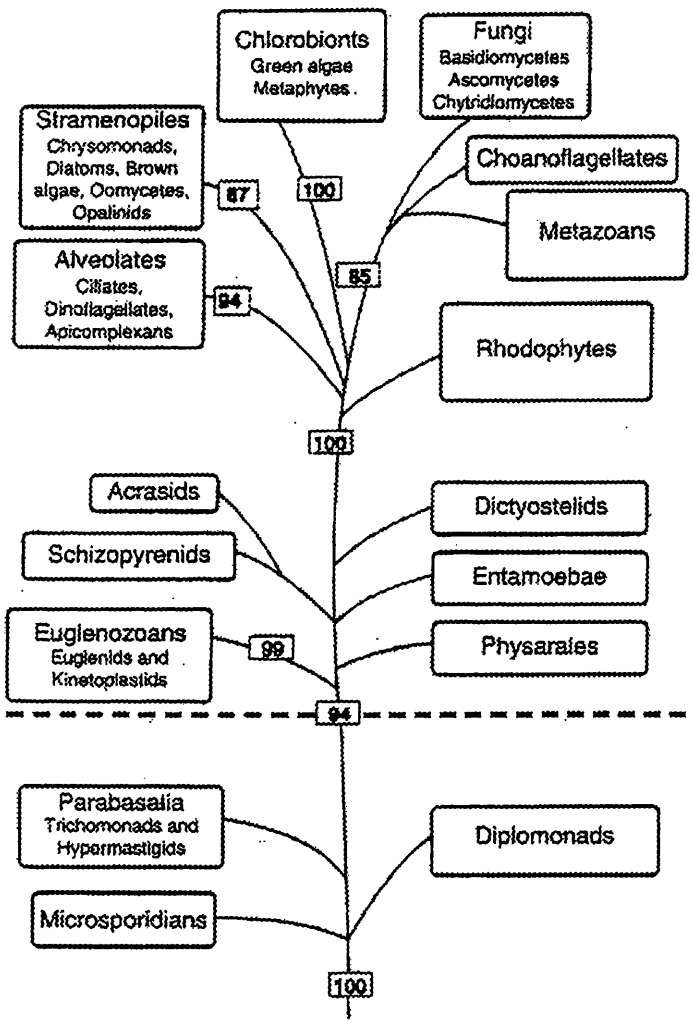
Classical morphological studies have shown that organisms can be defined in terms of three fundamental groups: Eubacteria, Archaebacteria and the Eukaryota (Schlegel, 1994). The use of molecular methods such as DNA sequencing have substantiated this finding and are now able to provide evolutionary links between the three groups. However, some discrepancies occur when molecular data are used to construct a Universal tree of life. Several different molecular methods have been used to construct the universal tree (e.g the use sequences from large subunit rRNA (LSU rRNA); small subunit rRNA (SSU rRNA); actin and tubulin) and, while the overall shape of the tree is consistent, some individual branches show inconsistencies. These data, therefore, need to be treated with a healthy degree of caution.

The SSU rRNA is ubiquitous and is found in all protein synthesizing systems. The additional functional constancy of this subunit makes it an ideal candidate for use as a molecular clock (De Rijk *et al*, 1995 and all references therein). SSU rRNA has both conserved and variable regions of sequence and can therefore be used in the study of both ancient relationships and recent divergences amongst different life forms (Van de Peer *et al*, 1993).

Sequences from LSU rRNA have a longer chain length and are less prone to variation than SSU rRNA sequences. Data from studies carried out on closely related species (after computer based corrective measures) show a higher level of consistency and higher bootstrap values than trees based on SSU rRNA sequences. The trees show larger differences in branch lengths but still show the same major clusters as trees derived from SSU rRNA sequences. There is currently no LSU rRNA universal tree because of the unavailability of sequences from some taxonomic groups.

Sequences from actin (Doolittle, 1992; Drouin *et al*, 1995) and tubulin have also been used in the construction of a universal tree. These trees basically agree with the SSU and LSU rRNA derived trees and particularly in relation to the divergence of the kinetoplastids. The kinetoplastids appear to have diverged soon after the appearance of endosymbionts which have resulted in the emergence of life forms with mitochondria and other organelles. Figure 1.4 is an illustration of a currently accepted SSU rRNA universal tree.

FIGURE 1.4



An illustration of a universal tree of life showing the order of divergence of Eukaryota, inferred from SSU (16s) rRNA. The life forms above the dotted line are descendents of mitochondriate eukaryotes. The life forms below the dotted line, do not possess mitochondria. Trees like this one, have been used as evidence for the acquisition and subsequent loss of mitochondria in some a mitochondriate protozoa. The numbers are the confidence levels of the points of divergence, i.e bootstrap values (adapted from Schlegel, 1994).

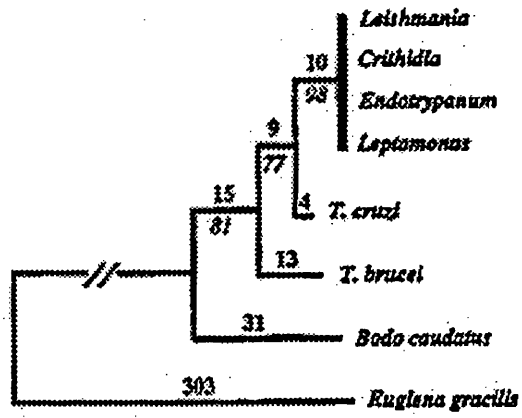
1.61 *Evolutionary relationships within the order kinetoplastida*

All trees to date agree on the point of divergence of the kinetoplastida i.e close to the evolutionary endosymbiotic inclusion event(s) which led to the acquisition of mitochondria in higher eukaryotes. Molecular data suggest that all parasitic trypanosomatids have evolved from free - living bodo - like flagellates (Maslov and Simpson, 1995). Using sequences from LSU and SSU rRNA, in conjunction with a suitable outgroup (*Bodo caudatus*, a kinetoplastid from the family Bodonidae), it has been shown that *T. brucei* was one of the first kinetoplastids to diverge after the acquisition of mitochondria (Fernandes *et al*, 1993; Vickerman, 1994; Maslov and Simpson, 1995). A tree constructed by Maslov *et al*, (1994), based on 18S nuclear ribosomal RNA sequences with *Euglena gracilis* as the outgroup agrees with that of Fernandes *et al*. The order of divergence of the kinetoplastids therefore appears to be as follows; *Bodonidae, T. brucei, T. cruzi, Blastocrithidia, Herpetomonas, Phytomonas, Leptomonas, Crithidia, Leishmania* and *Endotrypanum*. The general order of divergence is shown in Figure 1.5, with either *E. gracilis* or *B. caudatus* as the outgroup.

Data acquired from the study of RNA editing of three genes namely, cytochrome c oxidase subunit III, NADH dehydrogenase subunit 7 and an unidentified maxicircle open reading frame from different trypanosomatids, suggest that RNA editing is a primitive or ancestral system found in trypanosomes (Maslov *et al*, 1994). The hypothesis put forward by Maslov *et al* is that, the phenomenon of total RNA editing may have been replaced by the partial editing observed in monogenetic trypanosomatids and *Leishmania*, which are less primitive. This hypothesis appears to support the findings of Fernandes *et al*, (1993), with regards to the point of divergence of *T. brucei*.

FIGURE 1.5

A



B

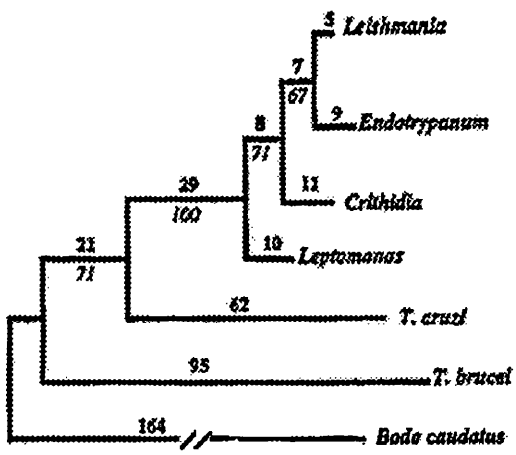


Figure 1.5 is an illustration of the general order of divergence of Eukaryota, based on sequences from small subunit rRNA, using two different outgroups, namely *Euglena gracilis* (A) and *Bodo caudatus* (B). The numbers on top of the branches are the calculated branch lengths and the ones in italics (beneath the branches) are bootstrap values, which give an indication of the confidence levels at the points of divergence (adapted from Fernandes *et al*, 1993).

1.7 Cellular signalling pathways in trypanosomatids

Currently little is known about cellular signalling pathways of trypanosomatids. The ability of trypanosomatids to undergo transformation (i.e from long slender trypanosomes to short stumpy forms or short stumpy forms to procyclics) in response to extracellular or environmental stimuli, indicates the presence, however primitive, of a system of cellular signalling. In higher eukaryotes such signals can be transduced by G proteins, protein phosphorylation and a variety of other mechanisms. The primitive origins of the trypanosomatids may therefore provide information on the origins of cellular signalling in eukaryotes and highlight some important differences, which could be exploited in the search for novel drug targets.
The following section is a review of what is known about cellular signalling in trypanosomatids.

1.71 *Factors involved in the recognition and response of trypanosomatids to external stimuli*

The ability of organisms to respond to extracellular stimuli is the main factor that contributes to their survival in the environment to which they are adapted. In trypanosomatids, such an ability could enhance not only their survival rate but also the assurance of the growth and propagation of the species. The trypanosome is in direct contact with host tissues and under normal circumstances, their presence elicits a powerful destructive immune response from the host. Antigenic variation has overcome that negative growth regulatory effect but it seems that more subtle host-parasite interactions could result in the regulation of growth and differentiation of the parasite (Pays *et al*, 1997).

Trypanosomes go through phases of elevated levels of growth and growth arrest at different stages of their life cycle. The long slender bloodstream and procyclic forms undergo rapid cell division, whereas the infective forms, i.e the mature metacyclic and short stumpy forms are always in a state of growth arrest (Matthews and Gull, 1994). The different stages may be subject to regulation by extracellular factors (host or parasite) which interact with the trypanosome cell surface. These interactions could activate intracellular signalling cascades. Knowledge of these signalling cascades could lead to the discovery of novel chemotherapeutic agents. In the following sections, I will review those factors involved in host-parasite interactions, which may play a role in cell growth and differentiation.

Epidermal growth factor receptor (EGF-R)

In higher eukaryotes, the EGF-R is an important component of growth factor mediated growth regulatory pathways. The stimulation of this receptor results in a cascade of events, which results in the initiation or termination of cell proliferation and/or differentiation. Hide *et al*, (1989), found evidence for the presence of an EGF-like receptor in both bloodstream and procyclic life cycle stages of trypanosomes. The receptor binds to EGF and is found to influence the rate of growth of both procyclic (Hide *et al*, 1990) and bloodstream forms of *T. brucei* (Sternberg and McGuigan, 1994). Though no-one has been able to isolate an EGF-like receptor gene from *T. brucei* to date, the evidence suggests that, there could be a signalling pathway in trypanosomes similar to that found in mammalian cells. Such a pathway could be involved in cellular proliferation and differentiation of the parasite. The isolation of the receptor, its gene and a functional analysis need to be carried out to determine the role of this putative receptor in trypanosome growth and differentiation.

Low density lipoprotein receptor (LDL-R)

Studies on mammalian host cell-trypanosome surface interactions have shown evidence for the presence of receptor mediated uptake of LDL in both bloodstream and procyclic forms of the trypanosome (Coppens *et al*, 1988; Coppens *et al*, 1991; Lee *et al* 1990a). Trypanosomes lack the machinery for the *de novo* synthesis of sterols, which are essential nutrients for normal cell growth in trypanosomes (Bastin *et al*, 1996 and all references therein), and are, therefore, dependent on receptor mediated endocytosis of low density lipoproteins (LDLs) for the provision of sterols. The LDLs are endocytosed via an LDL-like receptor (Hadjuk, 1992) found on the flagellum and within the flagellar pocket. Studies are currently being carried out on a 145 kDa LDL receptor isolated from *T.brucei* (Bastin *et al*, 1994). The receptor has the ability to bind LDLs and is also conserved in other kinetoplastids (Bastin *et al*, 1996).

The transferrin receptor (TR)

The flagellar pocket is the site of sequestration of receptors involved in the uptake of essential nutrients (Balber, 1990; Webster and Russell, 1993). Transferrin receptors are also found in the flagellar pocket and have the ability to bind transferrin (Schell *et al*, 1990).

The genes encoding a putative transferrin receptor have been cloned and sequenced. It is currently the only *T. brucei* receptor that has been characterised at the molecular level. The receptor differs both structurally (heterodimeric) and by its mode of membrane attachment from the human transferrin receptor (Ligtenberg *et al*, 1994; Salmon *et al*, 1994; Steverding *et al*, 1994). It is a heterodimer, encoded by the expression site-associated genes (ESAGs) 7 and 6, which binds transferrin. The receptor is attached to the membrane by a glycosyl phosphatidylinostol anchored carboxyl tail (reviewed by Borst, 1991; Pays *et al*, 1997) and shows some sequence homology to VSGs.

The interesting aspect of the localisation of receptors involved in receptor-mediated endo/exocytosis is the fact that despite being in contact with host serum, they are not destroyed by the host immune response. The exact mechanism whereby endo/exocytosis occurs without the destruction of the parasite by the host immune response is yet to be determined.

Some cell surface receptors are used in the modulation of the host immune response to enhance cell proliferation and differentiation. Indirect evidence, i.e the response of trypanosomes to IFN-γ, predicts the involvement of an IFN-γ specific receptor (Olsson *et al*, 1993). A TNF-α specific receptor has been found on trypanosome cell surface (Lucas *et al*, 1993). The TNF-α receptor is localised in the lumen of the flagellar pocket where the transferrin receptor is also found (Salmon *et al*, 1994; Steverding *et al*, 1994).

These cytokines appear to be involved in the regulation of cell proliferation and most possibly differentiation. When trypanosome cultures are treated with IFN-γ the rate of proliferation increases but when incubated with TNF-α they undergo lysis. It is postulated that a balance between these two processes ensures the survival of both the parasite and its host. Factors secreted by the trypanosomes, which activate the production of IFN-γ and TNF-α from CD8+ cells and macrophages respectively (Olsson, 1993; Tachado and Schofield, 1994), allow trypanosomes to modulate the hosts cytokine response for their own benefit. The regulation of growth appears to be life cycle stage dependent (Pays *et al*, 1997).

Apart from the ability of the trypanosomatids to induce and make use of various host components in the regulation of cell growth, the trypanosome produces components which have been linked to cell growth. A few examples of such cell surface components are:

1) Glycans of the GPI anchor of the VSG are directly involved in the activation of the production of TNF-α and IL-1 from macrophages (Tachado and Schofield, 1994). The released form of VSG has also been implicated as a strong inducer of TNF-α (Pays *et al*, 1997). The mode of activation and the components involved in the stimulation of IFN-γ are yet to be determined for CD8+ lymphocytes.

2) The role of cell surface lectin binding glycoproteins in the regulation of the developmental cycle of the trypanosome within its arthropod host. The glycoproteins are expressed on the outer membrane of both life cycle forms, i.e bloodstream and procyclic trypanosomes. They, however, only appear to be of obvious importance to the procyclic life cycle stage of the trypanosome. It has been proposed that the binding of lectins to the receptors could be involved in either cell growth or death, depending on which part of the surface receptor they bind (Maudlin and Welburn, 1994; Welburn *et al*, 1996). Growth of procyclic *T. b. rhodesiense* in the tsetse fly is maintained at a constant level which implies that trypanosomes maintain a fine balance between cell growth and death. It appears as though the procyclics have the ability to undergo apoptosis to maintain this balance. In order to carry out this process, host components are once again, used to the advantage of the parasite population (Welburn *et al*, 1996).

3) Trypanosomes are incapable of the *de novo* synthesis of sialic acid. To compensate for this, trypanosomes have cell surface trans-sialidases. These are enzymes which are needed for the transfer of sialic acid from the mammalian host to the parasite (Cross and Takle, 1993; Schenkman and Eichinger, 1993). In *T. cruzi*, sialic acid has been found to be essential for cell invasion and development. In *T. brucei*, cell surface trans-sialidases appear to be stage specific and have been observed on procyclic but not bloodstream forms. This would suggest a specific role for sialic acid in host-parasite interactions required solely for the development of procyclic trypanosomes. The exact role of the trans-sialidases in the development of procyclic trypanosomes is yet to be determined (Pays *et al*, 1997).

In higher eukaryotes, extracellular signals are known to be essential for the regulation of gene expression, cell proliferation and differentiation. This is made possible by the presence of membrane bound signal transducers (Karin, 1991). A limited amount of information is available on the role of such signal transducers in trypanosomatids, despite the considerable amount of data showing evidence of their presence.

1.72 *Membrane bound signal transducers*

The following section is a review on membrane bound transducers in trypanosomatids and, where known, their role in the development of the parasite.

Adenylate cyclase receptors and factors involved in their activation

In mammalian cells the stimulation of adenylyl cyclase receptors results in the initiation of the cAMP cascade (Tang *et al*, 1992). The cascade in turn activates several target molecules which control a number of events e.g gene transcription and cell growth. Mammalian adenylyl cyclases undergo type specific stimulation or inhibition by the α and $\beta\gamma$ subunits of heterotrimeric G proteins (Milligan, 1995; Neer, 1995) and Ca^{2+}/calmodulin (Bourne *et al*, 1992). In *Saccharomyces cerevisiae*, adenylyl cyclase is stimulated by G proteins in a similar manner to mammalian adenylyl cyclases. This might lead one to conclude that the stimulation of the adenylyl cyclases by G proteins, has been conserved through evolution. However, adenylyl cyclase in yeast, is stimulated by monomeric GTP-binding proteins, whereas the mammalian adenylyl cyclases are stimulated by the heterotrimeric G proteins (Ives, 1991; Neer, 1995). Therefore one cannot draw any valid conclusions about the evolutionary significance of the stimulation of adenylyl cyclase until further studies have been carried out on the molecules involved in its stimulation in lower eukaryotes (Kataoka *et al*, 1985).

As in mammalian cells, a number of factors are essential for the activation of adenylate cyclase in trypanosomes. Despite the evidence available on the existence of heterotrimeric GTP-binding proteins in *T. brucei* (Coulter and Hide, 1995), *T.cruzi* (Eisenschlos *et al*, 1986b; Coso *et al*, 1992; Oz *et al*, 1994) and *Leishmania* (Cassel *et al*, 1991), there is no evidence for the activation of adenylate cyclase by this class of proteins. Activity of the cyclases in trypanosomes can still be induced even in the presence of non-hydrolysable analogs of GTP (Martin *et al*, 1978; Torruella *et al*, 1986; Eisenschlos *et al*, 1986a). These results suggest the absence of a G-protein specific binding site on trypanosomal adenylate cyclase (Chen *et al*, 1995). In *Sacchromyces cerevisiae*, and other yeasts, activation of the adenylate cyclases occurs after stimulation by monomeric GTP-binding proteins (Verzotti *et al*, 1994). The lack of involvement of G proteins in the activation of trypanosomal adenylate cyclase, could therefore be due to the existence of a more primitive mode of activation which has not yet been discovered.

Evidence is accumulating for the presence of small monomeric GTP-binding proteins in *T. brucei* (Coulter and Hide, 1995; Field *et al*, 1995; El-Sayed *et al*, 1995; Sowa and Hide,

1997). A number of monomeric G proteins have been cloned e.g *ran* (Field *et al*, 1995), a number of *rab* GTPases (El-Sayed *et al*, 1995; Field and Boothroyd, 1995) and a member of the *ras* GTPase subfamily (Sowa and Hide, unpublished data). As yet the exact role of these proteins in the development of the trypanosome is not yet known. One can only speculate as to what role these proteins might have in the trypanosome, from studies in higher eukaryotes. It would be particularly interesting to determine the role played by the *ras* GTPase subfamily member in *T. brucei*, as studies in higher eukaryotes have shown that members of this subfamily are directly involved in cellular signalling pathways involved in growth regulation.

In *T. cruzi*, the adenylate cyclase receptor is activated by peptides present in the hind gut of the insect vector (Fraidenraich *et al*, 1993). Elevated levels of cAMP were observed in *T. cruzi* epimastigotes after the release of peptides by the proteolysis of fibronectin. This suggests that, the rise in cAMP concentration occurred as a result of the activation of adenylate cyclase. In *T. brucei*, these cyclases can be activated by tissue specific membrane extracts of the tsetse fly (Van Den Abbeele *et al*, 1995). There is currently no evidence for the direct involvement of peptides in the activation of *T. brucei* adenylate cyclase. *T. brucei* adenylate cyclase is encoded by ESAG 4 (Paindavoine, 1992). It is, therefore perhaps, not surprising that the activation of adenylate cyclase is linked to VSG release (Voorheis and Martin, 1982; Rolin *et al*, 1996). The initiation or inhibition of the two processes appear to be linked to varying concentrations of Ca^{2+} (Rolin *et al*, 1990). The adenylate cyclase gene family is also found in *T. equiperdum* (Ross *et al*, 1991) and in *Leishmania donovani* (Sanchez *et al*, 1995).

The inhibition of *T. brucei* protein kinase C (PKC), is found to activate the adenylate cyclase receptor and cause VSG release. The inhibition of PKC was also found to block the differentiation of short stumpy trypanosomes to procyclic forms. The close relationship between the activation of adenylate cyclase and VSG release indicates that the activation of this cyclase could play an important role in the differentiation of bloodstream trypanosomes (Pays *et al*, 1997).

Studies on adenylate cyclases in *Dictyostelium* (Pitt *et al*, 1992), have shown that the cyclases might be involved in developmental regulation. This finding suggests that *T. brucei* ESAG 4 (i.e adenylate cyclase), which is bloodstream specific, might play a role in the developmental regulation of *T. brucei*.

1.73 *Intracellular signal transduction*

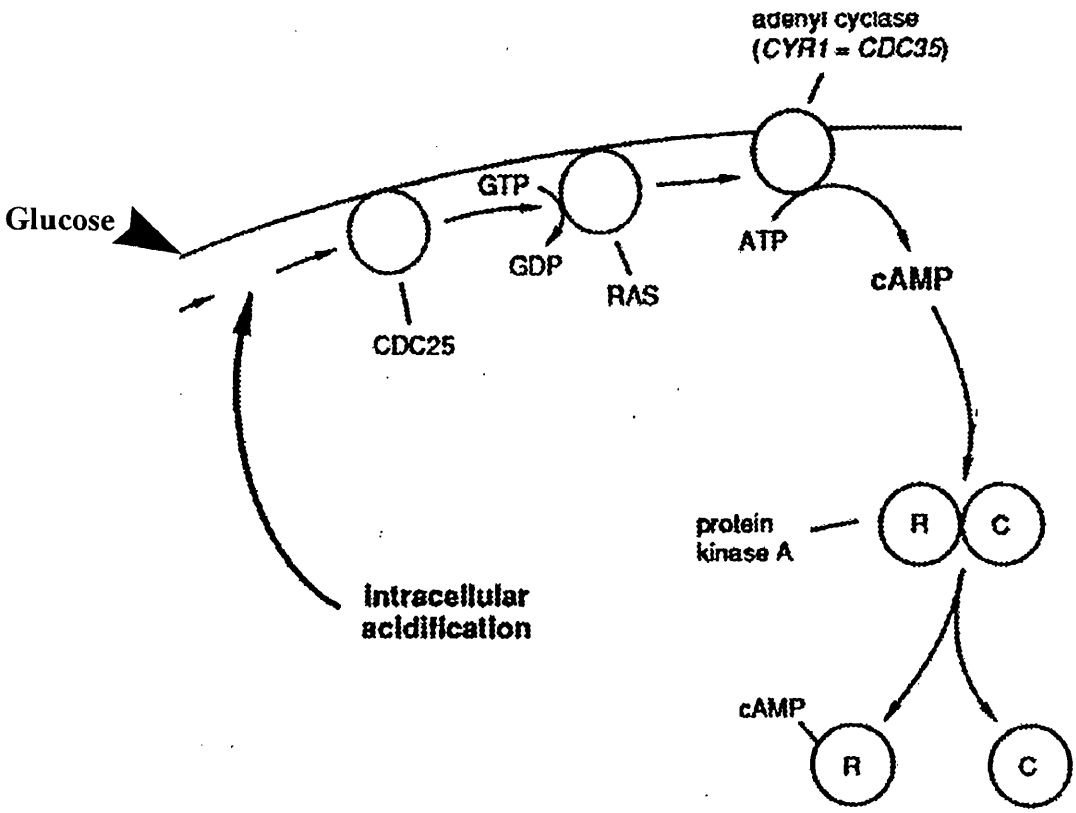
The initiation of the adenylate cyclase cascade appears to be linked to a number of physiological changes in different organisms. In *Saccharomyces cerevisiae*, the stimulation of this cascade is essential for the progression and completion of the G1 phase of the cell cycle (Tokiwa *et al*, 1994; Morishita *et al*, 1995). In both yeast and mammalian cells, cAMP in conjunction with *ras*, is implicated in the regulation of the transcription of heat shock genes. The function of heat shock proteins has been well conserved through evolution and are known to be involved in the regulation of normal cellular function (Engelberg *et al*, 1994). Cyclic AMP is also known to be involved in the regulation of cell growth and differentiation in *T. brucei*. Levels of intracellular cAMP are increased immediately before the transformation of bloodstream long slender to short stumpy forms (Mancini and Patton, 1981; Reed *et al*, 1985) and before the transformation of short stumpy to procyclic forms (Rolin *et al*, 1993). Increased levels of cAMP, therefore, appear to be associated with growth arrest, a pattern that has also been observed in mammalian cells (Marx, 1993). One of the modes of activation of the adenylate cyclase cascade in *T. brucei* is calcium induced, but the exact mechanism is not understood (Rolin *et al*, 1990; Paindavoine *et al*, 1992). There is also evidence for the initiation of the adenylate cyclase cascade (and therefore increases in cAMP levels) and the onset of differentiation from epimastigote to mature metacyclic forms of *T. cruzi* (Gonzales-Perdomo *et al*, 1988; Rangel-Aldao *et al*, 1988; Fraidenraich *et al*, 1993).

Based on the information available on the role of cAMP in yeast and mammalian cells and the evidence for the presence of a cAMP pathway in *T. brucei*, one could conclude that a cAMP intracellular pathway exists in trypanosomatids. This pathway is most likely to be involved in the mediation of cellular signals essential for the developmental changes that the trypanosomatids undergo. Figure 1.6 is an illustration of a model of the glucose and acidification-induced signalling pathway in yeast. The pathway is involved in the regulation of the G1 phase of the yeast cell cycle. The activation of this pathway by extracellular glucose or intracellular acidification leads to the activation of *ras* which in turn activates adenylate cyclase (and hence increases levels of cAMP) and finally Protein kinase A. The pathway is regulated by feedback inhibition of Protein kinase A on the production of cAMP. The activation of Protein kinase A induces the cell to enter the GI phase of the cell cycle. A similar pathway may exist in trypanosomes, though the extracellular stimulus might not be the same as in yeast cells.

The inhibition of the cAMP pathway by the phosphoinositide pathway in *T. cruzi* (Docampo and Pignataro, 1991; Racagni *et al*, 1992) suggests the inverse regulation of the two pathways. There is currently no evidence for the presence of this pathway in *T. brucei*.

However, it is possible that it has not yet been discovered. The actual function and regulation of this pathway is currently unknown.

FIGURE 1.6



An illustration of the signalling pathway involved in the regulation of the G1 phase of the yeast cell cycle. The activation of this pathway leads to the activation of adenylate cyclase and cAMP. The activation of the cAMP cascade results in the activation of Protein kinase A (PKA). When activated the pathway induces the cell to enter the G1 phase of the cell cycle. R and C - regulatory and catalytic subunits of PKA; CDC 25 - guanine nucleotide exchange factor. (Adapted from Thevelein *et al*, 1993)

Intracellular protein kinases

Data is accumulating on the presence of protein serine and tyrosine kinases involved in developmental regulation in trypanosomatids (Pays *et al*, 1997; Boshart and Mottram, 1997). Developmentally regulated autophosphorylation of protein kinases has also been shown to occur in *T. brucei* (Parsons *et al*, 1993; Hide *et al*, 1994). Parsons *et al*, (1993), have identified at least 10 different threonine kinases in *T. brucei*, six of which have been found to be developmentally regulated. Preliminary evidence suggests the presence of protein kinase A (Boshart and Mottram, 1997), protein kinase C (Keith *et al*, 1990; Boshart and Mottram, 1997) and calmodulin dependent protein kinases (Ogueta *et al*, 1994). The ability of protein kinase C inhibitors to block *in vitro* transformation of bloodstream to procyclic forms of *T. brucei* (Pays *et al*, 1997) substantiate the view that it is involved in developmental regulation. Parsons *et al*, (1991), showed evidence for the presence of life cycle regulated tyrosine phosphorylation of *T. brucei*. The presence of these kinases indicate a possible association with cellular differentiation. A fairly large number of protein kinase and phosphatase genes have been cloned from *T. brucei* (Boshart and Mottram, 1997). Their functions are presently unknown but, as homologues of these proteins are involved in developmental regulation in higher eukaryotes, they could have the same role in *T. brucei* (Parsons *et al*, 1993).

Mitogen activated protein kinase or MAPK

A putative MAP kinase, with 37 % identity to rat MAPK, has been cloned from *T. brucei* (Hua and Wang, 1994). This is an extremely interesting finding, because the stimulation of the MAPK pathway in mammalian cells has been linked to external activation via a number of membrane bound receptors e.g the heterotrimeric G-protein coupled and epidermal growth factor receptors (Leevers and Marshall, 1992) of which there is preliminary evidence for both in *T. brucei*. The mitogen activated protein kinase, or MAPK cascade, plays a significant role in the transduction of extracellular signals into intracellular responses in mammalian cells (Cano and Mahadevan, 1995) and the pathway is highly conserved in eukaryotic evolution (Pelech and Sanghera, 1992; Guan, 1994). A large number of factors are involved in the activation of this highly conserved pathway (Nishida and Gotoh, 1993) and homologues of molecules involved in the cascade have been isolated in *Caenorhabditis elegans* and *Drosophila melanogaster* (Cano and Mahadevan, 1995).

Studies carried out on mammalian cells present strong evidence for ligand-stimulated MAP kinase activation by $p21^{ras}$, a *ras* subfamily member. The *ras* protein lies upstream of *raf* kinase which is a known MAP kinase kinase in mammalian cells. *Raf* proteins are a family of cytosolic serine/threonine protein kinases (Moodie and Wolfman, 1993) involved in the transduction of signals from the cell surface to the nucleus (Daum *et al*, 1994). The activation of *raf* is known to be involved in the regulation of cell proliferation and differentiation (Avruch *et al*, 1994; Marais, *et al*, 1995). An interesting relationship exists between the stimulation of *raf* kinase and members of the *ras* and *rap* subgroups. In mammalian cells an antagonistic relationship has been observed between some members of the *ras* and *rap* subfamilies (see section 1.7 for a review of the *ras* and *rap* subfamilies). The relationship entails the competition for the *ras* binding site on *raf* by *ras* and *rap* proteins (Sprang, 1995; Block *et al*, 1996). When either *ras* or *rap* bind to the *ras* binding domain of *raf* in mammalian cells, it results in the stimulation or termination respectively of downstream signals from *raf* (e.g the MAP kinase cascade) (Leevers and Marshall, 1992; Minato *et al*, 1994; Burgering and Bos, 1995). There is evidence for the stimulation of protein kinase C by $p21^{ras}$. This is consistent with both kinases being downstream of the *ras* protein. However, there is evidence for PKC (indirect evidence found in *T. brucei*) being upstream of $p21^{ras}$ in some cell types (Leevers and Marshall, 1992). Some MAPK related protein homologues have also been isolated from a number of trypanosomatids, for example from *Leishmania chagasi* and *Leishmania major* (Boshart and Mottram, 1997). These MAPK related proteins are known intracellular signal transducers involved in the mating pheromone response in *Schizosacchromyces pombe* and *Saccharomyces cerevisiae* respectively.

With the exception of the lack of evidence for the presence of a *raf* homologue in *T.brucei*, (which could be yet to be identified in *T. brucei*, see Boshart and Mottram, 1997), most of the machinery required for the mammalian equivalent of a MAP kinase pathway in trypanosomatids, has been identified. What remains to be discovered is whether these homologues in the trypanosomatids have the same role in cellular signalling as do their mammalian equivalents. Figure 1.7 is an illustration of the complex relationship which exists between *ras*, *rap* and *raf* in mammalian cells and the link between two major signalling pathways, i.e the MAP kinase and cAMP pathways, which result in the regulation of cellular proliferation and differentiation.

Levels of calmodulin-binding proteins appear to be life cycle dependent and are thought to be linked to the regulation of gene expression in the parasites. A number of immunophilins have been characterised in *T. brucei* and *T. cruzi*. These molecules are involved in the calcineurin signalling pathways of higher eukaryotes. Their presence in trypanosomatids suggest the presence of an equivalent pathway.

FIGURE 1.7

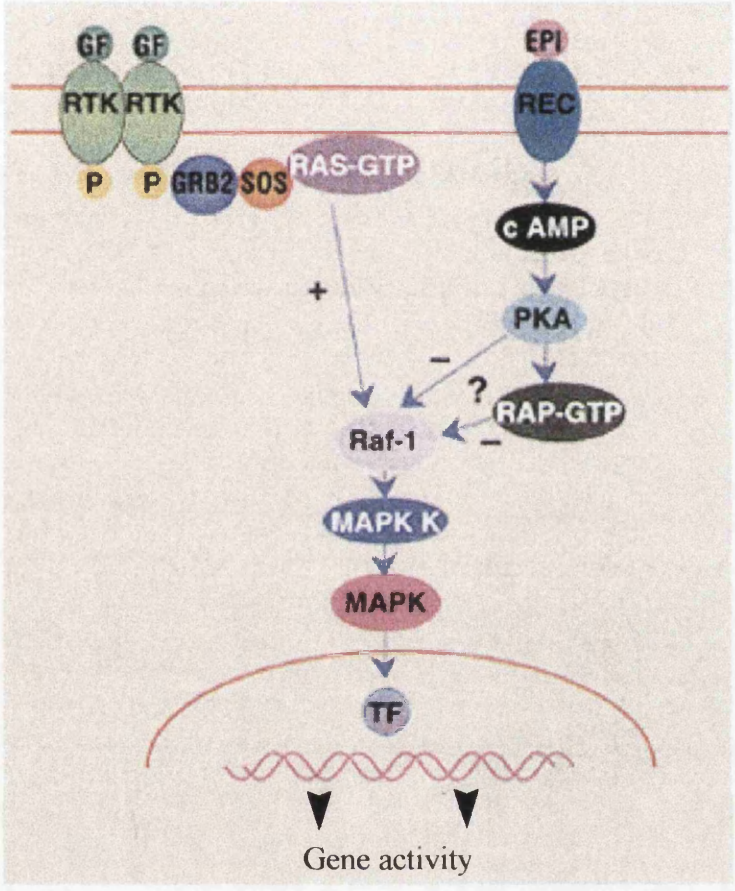


Figure 1.7 is an illustration of the two pathways which are linked by the interactions of *ras*, *rap* and *raf* in mammalian cells. The two pathways are the MAP kinase and the cAMP pathways. Both of the pathways are involved in the regulation of growth and differentiation in higher eukaryotes. GF - growth factor; RTK- receptor tyrosine kinase; GRB2 - adaptor protein; SOS- guanine nucleotide releasing factor; PKA - Protein kinase A; TF - transcription factors; REC - hormone receptor; EPI - hormone. In some cells the regulation of growth and differentiation is achieved by the activation of the cAMP pathway. This blocks the activation of *raf* by *ras*, thus preventing the initiation of the MAP kinase pathway (copied from Marx,1993).

Summary of cellular signalling in trypanosomatids

Studies on the life and cell cycles of trypanosomatids have shown that the two cycles are closely linked. The different stages are found to be dependent on their extracellular surroundings. The presence of host molecules may serve as extracellular signals thus mediating host-parasite interactions. These interactions are found to have a direct effect on the growth regulation patterns of the parasites.

The extensive regulation of parasite development, by protein kinases and phosphatases, suggests that these molecules may play an important role in the regulation of cell growth and differentiation as has been observed in higher eukaryotes.

A number of pathways lead to the stimulation of the adenylate cyclase cascade (via cell surface receptors) in different organisms. In yeast a direct link has been established between the activation of the *ras* signalling pathway and the initiation of the adenylate cyclase cascade (Matsuura *et al*, 1994; Neuman-Silberberg *et al*, 1995; Hurwitz *et al*, 1995; Yin *et al*, 1996). *Ras* proteins are thought to be involved in the mediation of cellular signals which result in normal cellular growth and differentiation (Krengel *et al*, 1990; McCormick, 1993; Feig and Schaffhausen, 1994; Bokoch, 1996). The *ras* pathway appears to be an important link between the intracellular environment and its external surroundings i.e, the activation of molecules downstream of *ras* result in phenotypic changes. A number of *ras* - like genes have been cloned from *T. brucei* (Field *et al*, 1995; Field and Boothroyd, 1995; El-Sayed *et al*, 1995). Their presence indicates the possibility of the existence of signalling pathways which involve members of the *ras* superfamily.

Based on the current knowledge of cellular signalling in trypanosomatids, it appears that there are signalling pathways equivalent to those in higher eukaryotes (e.g mammals) which are involved in the regulation of cellular growth and differentiation. One of the key sets of molecules regulating the transduction of extracellular signals to the various intracellular kinase cascades in higher eukaryotes is the *ras* GTPase superfamily. As one of the approaches to finding new drug targets is by gaining a greater understanding of the regulatory mechanisms involved in cell growth, this project was designed with the aim to investigate the role of *ras* subfamily GTPases in the development of *T. brucei*.

The following section is a review of the members of the *ras* superfamily, with particular emphasis on the *ras* subfamily and their role in cellular signalling. The review is based on studies in mammalian cells unless otherwise stated.

1.8 The *Ras* superfamily

The *ras* gene family encodes small monomeric GTP - binding proteins that belong to a large superfamily of GTPases. There are six subfamily members within this superfamily, and each subfamily is made up of a number of distinct groups (Zerial and Huber, 1995). The small GTPases form a family of molecular switches which have been implicated in the control of a number of physiological functions such as cytoskeletal organisation, proliferation and intracellular vesicle transport (Hall, 1990). The *ras* GTPase superfamily is made up six subfamilies: *ras*, *rap*, *rho*, *rab*, *ran*, *arf* and *sar*. Members of the superfamily have evolutionarily conserved GTP binding domains. They are involved in the cyclical regulation of the transmission of signals to downstream effectors, the final result of which is the regulation of cellular growth, differentiation or secretion. The cyclical regulation is achieved through the activation of the proteins by binding GTP, or the termination of the transmission of the signal, brought about by the hydrolysis of GTP i.e in the GDP bound state. Members of the *ras* GTPase superfamily are found across a wide range of organisms and studies have shown that their function has also been well conserved through evolution.

Figure 1.8 is an illustration of the *ras* GTPase superfamily, showing in particular, the members of the *ras* subfamily which are involved in cellular signalling. A number of small GTP - binding proteins have been cloned from *T. brucei* namely, *ran* (Field *et al*, 1995) and several *rab* GTP -binding proteins (El Sayed *et al*, 1995). To date no members of the *ras* subfamily of GTP - binding proteins have been cloned from *T. brucei* despite the ubiquitous distribution of these proteins (Hall, 1990).

FIGURE 1.8

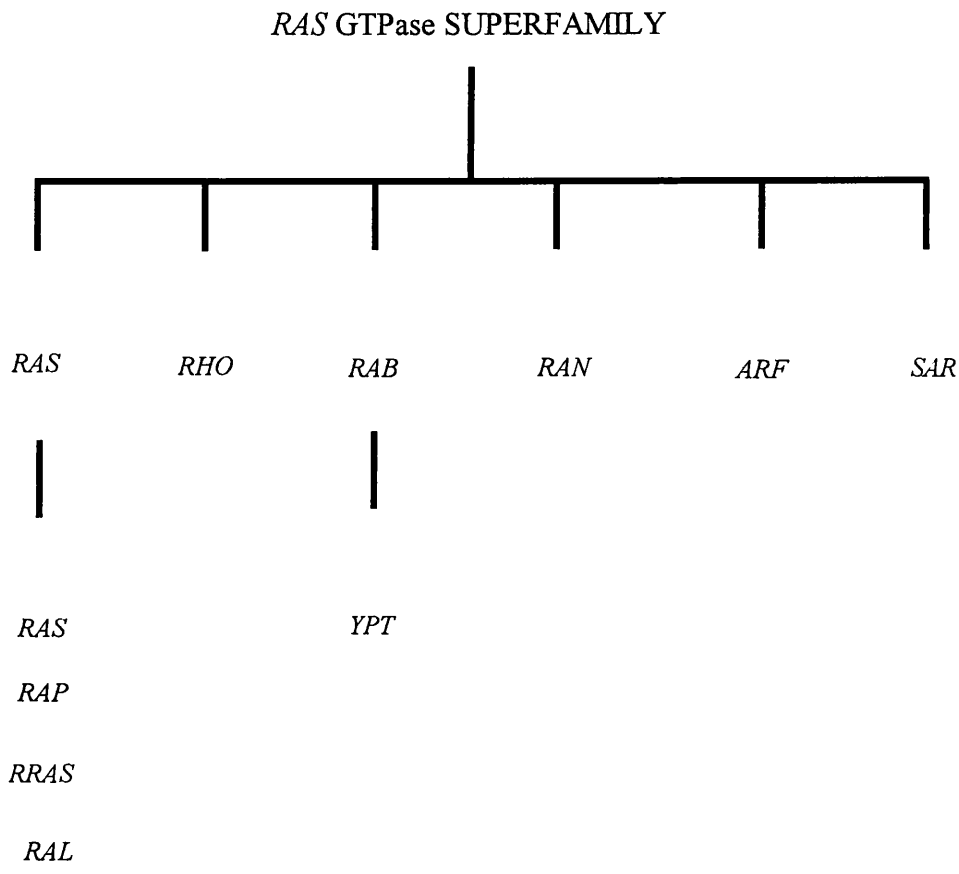


Figure 1.8 is a family tree showing the main subfamily members of the *Ras* GTPase superfamily namely, *ras*. *rho*, *rab*, *ran*, *arf* and *sar*. Group members of the *ras* subfamily, namely, *ras*, *rap* , *rras* and *ral* are shown as the review is mainly focused on two member of the *ras* subfamily known to be involved in cellular signalling. *Ypt* is a subgroup of the *rab* subfamily.

1.81 *Nomenclature of small GTP - binding proteins*

In 1992, Kahn proposed a method for naming small GTP - binding proteins, due to the increasing numbers of the genes encoding these proteins being cloned. Using these guidelines, a small GTP - binding protein is one that falls within a molecular weight range of 20 - 29 kDa (Bokoch and Der, 1993) and has a conserved, well defined GTP -binding domain. If a number of proteins share approximately 85 % or more identity, the same name with a different letter is assigned to the new protein. If the identity is less than 85 % but more than 35 % the new protein is given the same name but with a different number (Kahn *et al*, 1992; Zerial and Huber, 1995). The guidelines are, however, not very rigid, as illustrated by yeast *ras* 1 and 2, which have molecular sizes of 40 and 41 kDa respectively (Barbacid, 1987).

1.82 *General structure of a ras GTP - binding protein*

There are over 50 members of the *ras* GTPase superfamily which are known to regulate a wide range of cellular activities. The activities which are influenced by the presence of these small GTP binding proteins include cellular growth and differentiation, vesicular transport and cytoskeletal control, just to name a few (Bokoch and Der, 1993). *Ras* genes are well conserved in evolution and can be found in many eukaryotes, from yeast to man. It is the structure of a *ras* protein that defines it as a member of the *ras* GTPase super family. A protein belonging to the *ras* superfamily has four basic domains, 1) the amino - terminus; 2) the GTP binding/catalytic domain; 3) the C - terminal extension region and 4) the carboxyl terminus. Figure 1.9 is an schematic diagram of the general structure of a mammalian *ras* GTP binding protein, showing the four basic domains and their conserved sequence motifs.

FIGURE 1.9

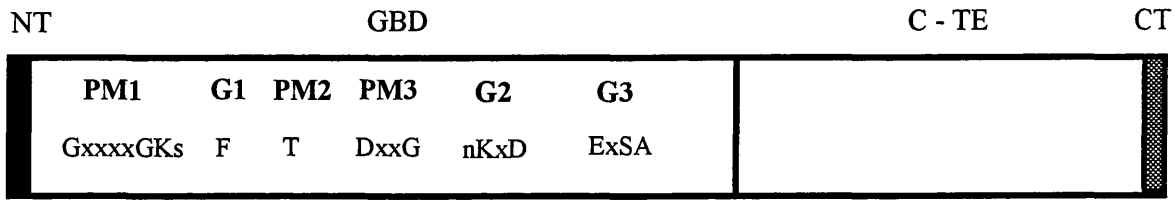


Figure 1.9 is a schematic diagram of the general structure of a mammalian *ras* GTP -binding protein depicting the four main domains: NT - the amino terminus; GBD - the GTP - binding or catalytic domain;
C - TE - the carboxyl terminal extension region; CT - the carboxyl terminus; G - Glycine, K - Lysine. F - Phenylalanine, T - Threonine, D - Aspartic acid, n - Asparagine (a small letter is used to show that it is not conserved in all *ras* genes), E - Glutamic acid and x - any amino acid. Within the GBD there are six regions of highly conserved sequence motifs, which are essential for activation of the *ras* signalling pathway. PM1, 2 & 3 denote the three phosphate/magnesium binding sites and G1, 2 & 3, the three guanine nucleotide binding sites of the GTP-binding domain.

The amino terminus

The length of the amino terminus (i.e the region immediately preceding the first phosphate/magnesium binding site) varies from 4 to more than 30 amino acids. The actual role of the amino terminus is unknown. It may be involved in interactions with other proteins but there is currently no evidence to this effect (Valencia *et al*, 1991).

The GTP - binding/catalytic domain (GBD)

The GTP - binding domain is the most highly conserved domain within a *ras* protein. There are six regions of conserved sequence motifs which are essential for the folding of the protein and its activity (the binding and hydrolysis of GTP). The six regions are involved in binding phosphate/magnesium and the guanine base of GTP and GDP. There are three phosphate/magnesium (PM 1, 2 and 3) and three guanine nucleotide (G 1, 2 and 3) binding sites within the GTP - binding domain (Valencia *et al*, 1991; Lowy and Willumsen, 1993). Conserved sequence motifs GxxxxGKs, T and DxxG correspond to PM 1, 2 and 3. The guanine base binding sites consist of the following conserved sequence motifs F, nKxD and ExSA (where G - glycine, T - threonine, n/s - asparagine/serine, the use of lower case letters indicate that the amino acid is conserved in most but not all *ras* genes. K - lysine, x - any amino acid, D - aspartic acid, E - glutamic acid, S - serine and A - alanine).

GxxxxGKs or PM 1

The sequence is highly conserved in all proteins belonging to the *ras* GTPase superfamily. The lower case "s" denotes the variability of the sequence at this position. In all members of the *ras* superfamily, the amino acid found at that position is either serine or threonine. Most members of the *ras* subfamily have the amino acid glycine at position 12 (based on mammalian ras gene amino acid positions, the general consensus is as follows: GxGxxGKs). Replacement of this glycine with any other amino acid (with the exception of proline), results in the formation of a mutant protein with transforming potential. *Ras* subfamily genes also tend to have Glycine at position 13, therefore the general consensus for a *ras* subfamily gene is GxGGxGKs (Valencia *et al*, 1990).

PM 2

There is only one conserved amino acid within this site, threonine. Threonine (T) plays a regulatory role in the structural transition from the activated to the inactivated form of the *ras* protein (Valencia *et al*, 1991).

PM 3

This site contains the motif DTAGQE. In *rap* proteins, the glutamine (Q) is replaced by threonine (T). It appears as though this difference might account for different mechanisms of activation (Valencia *et al*, 1991).

Guanine base binding sites 1, 2 and 3 or G1, 2 and 3

The sequence motif essential for the binding of the *ras* GTPase to the guanine base of GTP (i.e the G1 binding site) is phenylalanine (F) (Valencia *et al*, 1991). This residue is replaced by tyrosine in *rab* GTPases. A considerable amount of variation is observed in the G2 region and this depends on the subfamily to which the GTPase belongs. Mutants produced by the substitution of asparagine (N) have an increased rate of dissociation and an alteration in nucleotide specificity (Valencia *et al*, 1991). G 3 which has the motif ExSA is also highly conserved in all members of the *ras* superfamily. It does not bind directly to the guanine base, but appears to be essential for the maintenance of the stability of the protein (Valencia *et al*, 1991). A considerable degree of amino acid variability is observed in other regions of the catalytic domain in the various subfamilies.

The carboxyl - terminal extension or C - TE region

This region, containing over 40 amino acids, appears to be the most variable region of the *ras* GTPase genes. Yeast *ras* genes have uncharacteristically large molecular sizes and are known to have as many as 120 amino acids in the carboxyl terminal extension region. As yet a definite function for this region has not been determined.

The carboxyl terminus or CT

The carboxyl terminus of members of the *ras* GTPase superfamily, is essential for post translational modification and subsequent subcellular localisation of the protein. The sequence found in this region is characteristic of the *ras* GTPase subfamily. Members of the *ras* subfamily have a CAAX box, where C - cysteine; A - aliphatic amino acid and X - any residue (Valencia *et al*, 1991; Hancock *et al*, 1991; Lerosy *et al*, 1991; Bokoch, 1993; Zerial and Huber, 1995). *Rho* subfamily members also have a CAAX box which undergoes isoprenylation (Zerial and Huber, 1995). Members of the *rab* subfamily have at least six alternative terminal sequences; CXXX, CXC, CC, CCX, CCXX and CCXXX (Zerial and Huber, 1995). The carboxyl terminus sequence is involved in determining which transport pathway the *rab* proteins are involved in (Chavrier *et al* 1990).

1. 83 *Three dimensional structure of a mammalian ras GTPase*

The three dimensional structures of mammalian $p21^{ras}$ in both the active (bound to GTP) and inactive (bound to GDP) forms are now known (Grand and Owen, 1991). The structure is being used to study the mode of activation of members of the *ras* superfamily and the residues involved in each of the interactions. Figure 1.10 is a three dimensional diagram of mammalian $p21^{ras}$, showing the α helices and β sheets. A number of conformational changes occur in the structure of the protein when bound to GTP. In the GTP bound form, the hydrophobic core of the protein consists of six β sheets connected by hydrophilic loops and α helices (Bourne *et al*, 1991). The regions of the protein involved in the exchange of GTP for GDP and vice versa are the three phosphate/magnesium and guanine base binding sites.

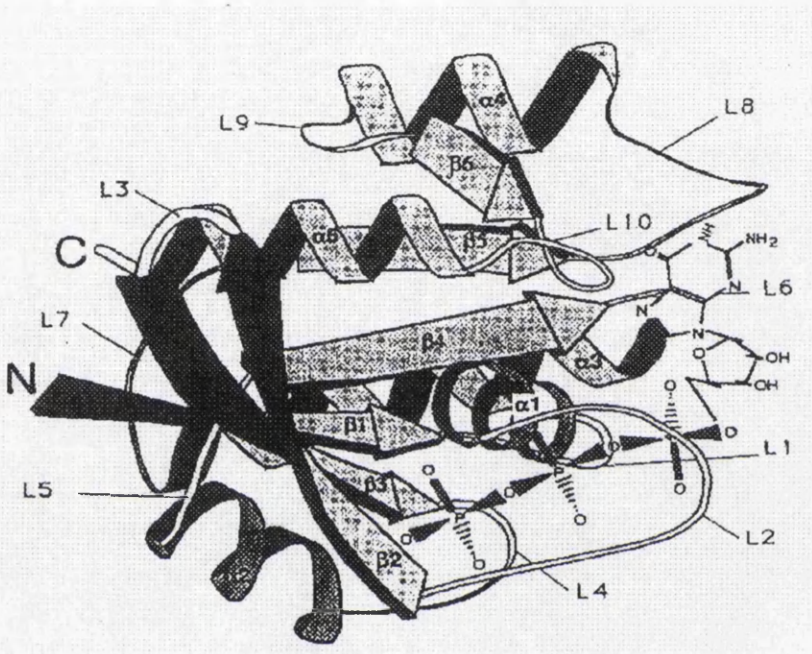
PM 1 forms a loop which is closed by a non-covalent interaction between lysine (K) and the carbonyl groups of two residues in the main chain, namely glycine 10 and alanine 11 (where glycine 10 and alanine 11 are the two x's found in the PM sequence motif GxxxxGKs, the first glycine is glycine 9) (Pai *et al*, 1989). Lysine, which also forms a side chain, is thought to be involved in catalysis as it is in contact with the β and γ phosphate oxygens (Reinstein, 1990; Pai *et al*, 1991). The conformation of this region is the same in both the active and inactive forms. This is due to bonds formed between the ε amino group of lysine and the α and β phosphates of GTP or GDP. The loop formed in this region connects the first β sheet (β 1) to an amphipathic α helix referred to as α 1 (Valencia *et al*, 1991).

PM 2 is involved in the stabilisation of the γ phosphate of GTP (Marshall, 1993). A conformational change occurs in the hydrophilic loop, which precedes the second hydrophobic β sheet (β 2) on binding GTP. There is evidence that this region might be involved in the hydrolysis of GTP by positioning the water necessary for this process. It is also involved in the co-ordination of a magnesium ion to the β and γ phosphate oxygens of GTP. The magnesium ion is essential for the inactivation of the protein or GTP hydrolysis.

The second α helix (α 2) and the third hydrophobic β (β 3) sheet are found within the PM 3 region. The conformation of the helix is dependent on the state of activation of the protein. Aspartate (D) binds to the magnesium ion via a molecule of water, while Glycine (G), binds to the γ phosphate of GTP by hydrogen bonds (Bourne *et al*, 1991). Glutamine (Q) in this region, is presumed to be involved in the catalytic reaction, due to its proximity to a water molecule which is thought to react with the γ phosphate of GTP (Pai *et al*, 1990).

The most important of the three guanine base binding regions is G 2. The fifth hydrophobic β sheet (β5) is located just before G 2. A link between three of the nucleotide loops can be found in this region and serves to some extent as a stabilising factor for the whole protein (Valencia *et al*, 1991 and all references therein). G 3 contributes indirectly to the overall stability of the protein in both active and inactive states. The sixth β sheet (β 6) is found within this region. G 1 also contributes to the stability of the protein by binding to the guanine base of both the GTP and the GDP forms of the protein.

FIGURE 1.10



Three dimensional representation of the structure of mammalian *ras* p21, showing residues1-166 (Pai et al, 1990). Helices are represented by alpha 1-5 and beta sheets by beta 1-6. Loops are represented by L1-10. The highly conserved sequence motifs of the GTP-binding domain are found in loops L1, L4, L6 and L10 (on the right hand side of the 3-D figure). The GTP nucleotide position and conformation have been approximated (photocopied from Valencia *et al*, 1991).

1.84 *Post-translational modification and subcellular localisation*

All proteins of the *ras* and *rho* subfamilies have a C - terminal CAAX motif which serves as a signal for post translational modification (Hancock *et al*, 1991). The post translational modifications of members of the *ras* and *rho* subfamilies are found to be essential for membrane targeting, lipid binding (Willumsen *et al*, 1984; Willumsen, Christensen *et al*, 1984) and function (Bokoch and Der, 1993). In other words the process of post translational modification is necessary for the subcellular localisation and normal function of the protein. The phenomenon of post translational modification is not only confined to members of the *ras* and *rho* subfamilies. Other proteins which have the C - terminal modifications that determine their subcellular location, include nuclear lamins, fungal mating factors and the γ subunit of retinal transducin (Moores *et al*, 1991 and all references therein). Some *rab* subfamily members are also known to undergo post translational modification (Bokoch and Der, 1993).

The process of post translational modification occurs in three distinct stages:

1) the isoprenylation of the cysteine of the CAAX motif (Hancock *et al*, 1989; Hancock *et al* 1991).

2) The peptide bond between the cysteine (of the CAAX motif) and the adjacent alanine is proteolytically cleaved (Gutierriez *et al*, 1989; Valencia *et al*, 1991; Lowy and Willumsen, 1993).

3)The α carboxyl group of the terminal isoprenylated cysteine (from the CAAX box), undergoes methylation (Gutierriez *et al*, 1989; Valencia *et al*, 1991; Lowy and Willumsen, 1993; Bokoch, 1993). All three processes are essential for the subsequent membrane association, localisation and function of the protein. The exact order of these modifications is yet to be determined.

The isoprenylation of the protein may occur by the addition of a farnesyl (C15) or a geranylgeranyl group (C20). The reaction occurs in the presence of either farnesyl transferase (for farnesylation) or geranylgeranyl transferase (for geranylgeranylation). In general, a protein will undergo farnesylation if the X of the CAAX motif is any one of the following amino acids; S, C, M or Q. If the X is L or F the protein will be geranylgeranylated (Moores *et al*, 1991; Bokoch and Der, 1993). Post translational modification results in the distribution of different subfamily members in different subcellular fractions and members of the subgroups *ras* and *rap* are never found in the same subcellular fractions (Beranger *et al* 1991; Kim *et al* 1990). Thus members of the *ras* subfamily are found on the inner surface of the plasma membrane (Lowy and Willumsen, 1993), while *rap* subfamily members can be found on the endoplasmic reticulum and endosomal/lysosomal compartments (Zerial and Huber, 1995).

1.85 *Biochemical activation and function of members of the ras subfamily*

Activation of the *ras* protein occurs in conjunction with a number of proteins. These proteins are regulatory proteins, essential for the dissociation of GDP (the inactive form) and its replacement by GTP. These proteins are known as Guanine nucleotide exchange factors or GEFs (Boguski and McCormick, 1993)

In the GTP - bound state, the *ras* GTPase protein serves as a mediator of cellular signals. It has an intrinsic GTPase activity (Lowy and Willumsen, 1991) and therefore can be classified as an autoregulatory protein. When activated (in mammalian cells), members of the *ras* subgroup activate the mitogen activated protein kinase (MAPK) cascade via *raf* protein kinase (Zhang *et al*, 1993; Warne *et al*, 1993; Burgering and Bos, 1995; Farrar *et al*, 1996). In yeasts there is evidence for the activation of the adenylate cyclase pathway which is essential for normal cellular differentiation (Thevelein, 1991). There is evidence that shows that members of the *rap* subgroup when in the activated state (i.e bound to GTP) act in opposition to members of the *ras* subgroup (Lerosy, 1991; Nassar *et al*, 1995; Sprang, 1995).

The hydrolysis of GTP and the concomitant loss of a molecule of inorganic phosphate, returns the protein to its inactive state. A number of regulatory proteins are also involved in this process e.g the GTPase activating protein otherwise referred to as GAP (Gibbs *et al*, 1990; McCormick, 1990). The catalysis of the *ras* - GTP to *ras* - GDP reaction by GAPs have been observed in both mammalian and yeast cells. Figure 1.11 is a schematic diagram showing the cycling of a *ras* protein between its active (GTP-bound) and inactive (GDP-bound) states.

FIGURE 1.11

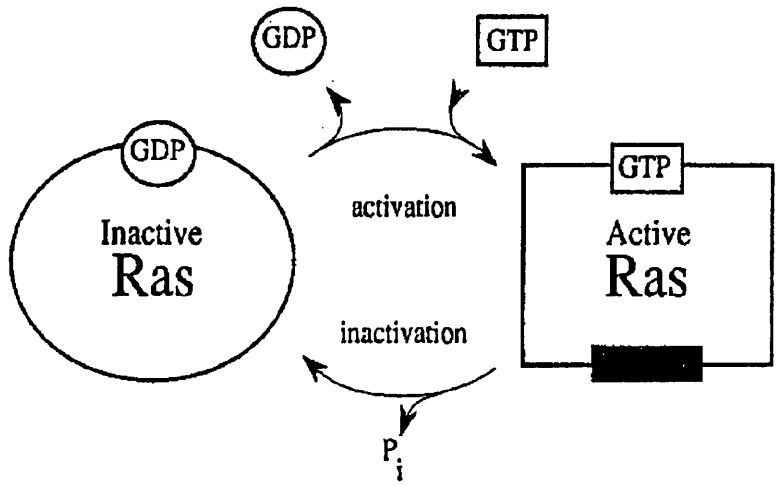


Figure 1.11 is a model of the activation and inactivation of *ras*. Extracellular signals induce the guanine nucleotide exchange factors (GEFs) and/or the GTPase activating proteins (GAPs) to initiate or terminate the binding of GTP to *ras*. In the activated form, GDP is physically replaced by GTP. The inherent GTPase activity immediately starts to function and in conjunction with the GAPs, GTP is hydrolysed with the concomitant loss of a molecule of inorganic phosphate (Pi). The protein at this stage returns to its inactive state, i.e it binds GDP.

1.9 The evolution of the *ras* gene family

The *ras* superfamily has evolved by a succession of gene duplication and divergence events over evolutionary time. Estimates suggest that the gene duplication rate of the *ras* subfamily is very slow, at 0.33 duplications per 100 million years (Iwabe *et al*, 1996). This highlights the extent to which the subfamily has been conserved throughout evolution. The subfamily is known to have been in existence, prior to the divergence of vertebrates and invertebrates (Valencia *et al*, 1991; Iwabe *et al* 1996). Experiments carried out on *E. coli* showed that no *ras* superfamily homologues exist. This suggests that the superfamily evolved subsequent to the divergence of the eukaryotes from the Eubacteria or (possibly) the Archaebacteria.

Ras genes are known to exist in higher eukaryotes. Recently a number of them have been cloned in the lower eukaryotes. *Dictyostelium* has 7 *ras* and 1 *rap* gene; two *ras* and two *rap* genes have been cloned from *Entamoeba histolytica*; two *ras* and one *rap* from *Physarum polycephalum*. Only a few *ras* superfamily genes have been cloned from the kinetoplastids. A *ypt* gene has been isolated in *Leishmania* (Cappai *et al*, 1993) and, one *ran* gene and nine *rab* genes have been cloned from *T. brucei* (Field and Boothroyd, 1995; Field *et al*, 1995; El Sayed *et al*, 1995). To date no *ras* genes from the *ras* GTPase subfamily have been cloned from *T. brucei* or any other kinetoplastids.

1.10 Aims and Objectives

There is evidence for the presence of membrane bound signal transducers but a very limited amount of information is available on their role in trypanosomatid development e.g the adenylate cyclase receptors. The factors involved in the activation of these receptors are yet to be determined. In higher eukaryotes the adenylate cyclase cascade is activated by heterotrimeric GTP-binding proteins. There is, however, no evidence for its activation by heterotrimeric G proteins in trypanosomes, even though there is evidence for their presence in the parasites. In yeasts, monomeric GTP-binding proteins have been implicated in the activation of the adenylate cyclase cascade. There is mounting evidence for the presence of monomeric GTP-binding proteins in the trypanosomatids. A number of small GTP-binding proteins have been cloned from *Leishmania* and *T. brucei*.

To date none of the monomeric GTP-binding proteins that have been cloned from *T. brucei* are from the *ras* subfamily. Members in this subfamily are known (from studies in higher

eukaryotes) to be directly involved in the transduction of extracellular signals into intracellular responses. Coulter and Hide, (1995) found evidence for the presence of small GTP binding proteins in *T. brucei*, using an *in situ* GTP binding assay. The identities of the monomeric GTPases were unknown, but as they have been highly conserved through evolution, the project was designed to characterise the monomeric GTPase subfamilies present in *T. brucei*. This was undertaken with the aim of isolating a *ras* subfamily GTP binding protein. The main objective of the project was to determine the role of *ras* subfamily GTP binding proteins in the regulation of trypanosomal growth, by isolating, cloning and fully sequencing a *ras* GTPase subfamily gene from *T. brucei* and developing tools and reagents for the analysis of the function of the *ras* gene. A secondary objective was to determine the relationship between present day *ras* GTPases (of eukaryotes higher up the evolutionary scale than the kinetoplastids) and *ras* GTPases from *T. brucei*, as *T. brucei* belongs to the Order Kinetoplastida, one of the first groups of eukaryotes to diverge after the acquisition of mitochondria.

1.11 Overview of the results chapters

Chapter 3: Detection of monomeric GTP binding proteins in *Trypanosoma brucei*

In this chapter, small GTP binding proteins were identified in *T. brucei* using the GTP binding assay of Coulter and Hide, 1995. A slot blot GTP binding assay was developed and used for screening large numbers of HPLC fractions. Results from the use of anion exchange chromatography in the purification of a *ras* like GTP binding protein from spun detergent extracts of bloodstream *T. brucei* (strain TREU 869) showed the presence of a 24 kDa peptide capable of binding GTP but, due to its low abundance, there was a need to considerably scale up the amount of starting material. The use of heterologous *ras* antibodies for the detection of small GTP binding proteins in *T. brucei* proved inconclusive due to the excessive cross reactivity of the antibodies.

Chapter 4: Cloning, sequencing and identification of *tbrlp*, a *ras* like GTPase gene

This chapter concerns the isolation of a *ras* like GTPase gene, *tbrlp* from *T. brucei*. *Trypanosoma brucei ras* like protein or *tbrlp* has an open reading frame of 681 bp (227 amino acids) and all the highly conserved sequence motifs of a member of the *ras* GTPase superfamily. PCR amplification of *tbrlp* using a specific set of 3' and 5' *tbrlp* primers suggests that there is an identical copy of *tbrlp*, with the second copy located upstream of and tandemly linked to *tbrlp*. The first gene was therefore renamed *tbrlp* I and the second copy, *tbrlp* II. Low stringency hybridisation provided no evidence for related genes.

Chapter 5: Evolutionary analysis of *tbrlp* I

In addition to the highly conserved sequence motifs characteristic of members of the *ras* GTPase superfamily, *tbrlp* has a number of unusual features. The GTP binding domain has 45 % amino acid identity to human *rap* 2 and 41 % identity to human *ras* h, both of which are members of the *ras* subfamily. The mixed *ras* and *rap* features of the GTP binding domain of *tbrlp* I, suggest that it may have an interesting evolutionary relationship with the *ras* and *rap* genes of higher eukaryotes. In Chapter 5, six possible hypotheses explaining the evolutionary origin of *tbrlp* I are proposed and tested.

Chapter 6: Development of tools for functional analyses of *tbrlp* I

From the nucleotide and amino acid sequence of *tbrlp* I, it is difficult to determine the exact subgroup of the *ras* subfamily to which it belongs. Studies of the function of *ras* and *rap* GTPases in mammalian cells, indicate that the two proteins function in opposition to each other i.e *ras* function is activated when *rap* function is inhibited and vice versa. Both of these proteins are known to be involved in cellular proliferation and differentiation. From the sequence of *tbrlp* I, it is therefore difficult to determine the possible function of *tbrlp* I i.e whether the function of *tbrlp* I is predominantly *ras* like or *rap* like. It was therefore necessary to develop tools for functional analyses of *tbrlp* I, which would, in the long run, give an insight into its role in the development of *T. brucei*. Chapter 6 describes the approaches taken in the development of tools for the functional analyses of *tbrlp* I.

Chapter 2

Materials and Methods

CHAPTER 2

MATERIALS AND METHODS

Materials and methods used in protein manipulations

The following materials and methods, were used for all protein manipulations described in chapters 3 and 6.

2.0 Preparation of *T. brucei* bloodstream and procyclic spun detergent extracts

Sprague-Dawley or Wistar rats were inoculated with *T. b. brucei* cloned monomorphic stock TREU 869. Bloodstream form trypanosomes were harvested at maximum parasitaemia (approximately 4 days after inoculation), separated from host blood by anion exchange chromatography (Lanham and Godfrey, 1970), washed twice in phosphate buffered saline at pH 8.0 containing 1 % glucose (PBSG) and used in the preparation of bloodstream spun detergent extracts. Procyclic *T. brucei* of the stock EATRO 1125 were grown in culture using SDM-79 medium (Brun and Schonenberger, 1979). The cultured procyclic trypanosomes were also washed twice in PBSG and frozen as packed cell pellets at -70°C until required for the preparation of spun detergent extracts.

Spun detergent extracts of both bloodstream and procyclic trypanosomes were prepared in the following way:

(a) 1 ml of packed whole trypanosomes were thawed homogenised with a glass homogeniser in an equal volume of the following on ice: ice cold HEPES/sucrose/EDTA pH 7.4 (HSE) buffer (10mM HEPES, pH 7.4, 0.25M sucrose and 1mM EDTA) containing the following protease inhibitors: aprotinin and leupeptin added to a final concentration of 25μg ml $^{-1}$ and 1mM benzamidine. Triton X-100 and Nonidet P40 were added to a final concentration of 0.1 %.

(b) The homogenate was centrifuged at 38K for 45 minutes at 4°C in a precooled Beckman 70.1 Ti rotor.

(c) The supernatant was removed and stored at -70°C. The pellet was resuspended in 1 ml of HSE buffer (or the equivalent amount for a lower volume of procyclics) and also stored at -70°C

2.1 HPLC purification

Purification by anion exchange chromatography was performed using a pre-packed (Protein-Pak 8HR resin) Waters AP-1 glass column with a diameter of 1 cm. The column was set up with a 0.8 ml min^{-1} flow rate and samples were collected every thirty seconds. Separate bottles containing the following reagents: 100 mM Tris-HCl, 100 mM Tris-base, 1 M NaCl and distilled water, were mixed in the HPLC fluid handling system as required for each of the purification procedures. After use, the column was regenerated by elution in 0.5 M NaCl followed by washing through with distilled water.

2.2 *In situ* GTP-BINDING ASSAY

The *in situ* GTP-binding assay was carried out as described in Coulter and Hide, (1995). Briefly, spun detergent extracts of both bloodstream and procyclic *T.brucei* were separated on 15% SDS polyacrylamide gels in Bio-Rad Mini Protean II apparatus for 2 hours at 100v. The proteins were either visualised by silver staining or the gel was pre-soaked in Tris/Glycerol buffer (50mM Tris-HCl - pH7.5 and 20% Glycerol) for an hour before setting up the transfer onto PVDF membranes. The PVDF membranes were rehydrated in the following way: the membranes were presoaked in methanol for 10-15 minutes, distilled water for 10-15 minutes and then in transfer buffer (10mM $NaHCO_3$ and 3mM Na_2CO_3 pH 9.8) for 10 minutes. The proteins were then transferred onto the membranes using 10mM $NaHCO_3$ and 3mM Na_2CO_3 pH 9.8 as the transfer buffer. The transfer was carried out at 50 volts for 2 hours at 4°C in a Bio-Rad Mini Protean II apparatus.

After the transfer, the membranes were blocked for either 1 hour at room temperature or overnight in a blocking buffer made up of the following: 50mM NaH_2PO_4, 10mM $MgCl_2$, 2mM dithiothreitol (DTT) and 0.3% Tween 20 pH 7.5. The blocked membranes were then incubated in hybridisation bottles containing 10mls of blocking buffer and 30μCi of [α^{32}-P]GTP (specific activity: 3000 Ci $mmole^{-1}$, NEN Research Products) for 2 hours at room temperature. After the incubation period the membrane was washed 3 times for 10 minutes in blocking buffer and exposed to X-Omat XS-1 film at -70°C. The autoradiograph was developed after 1 hour.

2.21 *Slot blot GTP binding assay*

PVDF membranes were cut to fit into a slot blot apparatus (Gibco BRL 24-well filtration manifold). The membranes were rehydrated as previously described (section 2.2). The rehydrated membranes were placed on the lower half of the apparatus and the upper half was screwed down into place on top of the membrane. The slots were then loaded with 100 µl aliquots of the HPLC fractions and the samples were transferred onto the PVDF membranes by briefly applying a vacuum to the slot blot apparatus. After the transfer, the membranes were blocked in blocking buffer (50mM NaH_2PO_4, 10mM $MgCl_2$, 2mM dithiothreitol and 0.3% Tween 20 pH 7.5) for 1 hour at room temperature. The membranes were assayed for GTP binding by incubation in hybridisation bottles containing 10mls of blocking buffer and 30µCi of [α^{32}-P]GTP (specific activity: 3000 Ci mmole $^{-1}$, NEN Research Products) for 2 hours at room temperature. After incubation, the membranes were washed 3 times for 10 mins and exposed to X-Omat XS-1 film at -70°C. The autoradiograph was developed after 1 hour.

2.3 Polyacrylamide gel electrophoresis and immunoblotting

Polyacrylamide gel electrophoresis was carried out using standard procedures (Laemmli, 1970). 12 or 15 % gels with a thickness of 1.5 mm, were prepared in Bio-rad Mini Protean II apparatus. 12 % resolving gels were prepared as follows: 3.35 ml of distilled water, 2.5 ml of 1.5 M Tris-HCl (Sigma) pH 8.8, 100 µl of 10 % w/v SDS (Sigma), 4.0 ml of Acrylamide/Bisacrylamide (30 % stock, Sigma), 50 µl of fresh 10 % ammonium persulfate (Sigma) and 5 µl of tetra-methyl-1,2-diaminoethane (TEMED, Sigma). The resolving gels were overlaid with 1 ml of distilled water and allowed to set for 1 hour. 4 % stacking gels were used routinely and were prepared as follows: 3.05 of distilled water, 1.25 ml of 0.5 M Tris-HCl, pH 6.8, 50 µl of 10 % w/v SDS, 1.3 ml of Acrylamide/Bisacrylamide, 50 µl of fresh 10 % ammonium persulfate and 10 µl of TEMED, making a total volume of 5 ml. The protein samples were prepared for separation by the addition of SDS-PAGE sample buffer (without β-mercaptoethanol for samples to be assayed for GTP binding), boiled for 5 minutes and loaded into the wells. Protein molecular weight markers were either from Sigma or New England Biolaboratories (NEB) and used according to the manufacturers instructions. Gel electrophoresis was undertaken in a buffer consisting of the following: 50 mM Tris-base, 384 mM glycine, 2 mM EDTA and 0.1 % SDS at room temperature for 2 hours at 100 volts.

15 % gels were prepared in the following manner:

2.3 ml of distilled water, 2.5 ml of 1.5 M Tris-HCl (Sigma) pH 8.8, 100 µl of 10 % w/v SDS (Sigma), 5.0 ml of Acrylamide/Bisacrylamide (30 % stock, Sigma), 100 µl of fresh 10 % ammonium persulfate (Sigma) and 4 µl of tetra-methyl-1,2-diaminoethane (TEMED, Sigma). The resolving gels were overlaid with 1 ml of distilled water and allowed to set for 1 hour. 4 % stacking gels were used routinely and were prepared as follows: 3.05 of distilled water, 1.25 ml of 0.5 M Tris-HCl, pH 6.8, 50 µl of 10 % w/v SDS, 1.3 ml of Acrylamide/Bisacrylamide, 50 µl of fresh 10 % ammonium persulfate and 10 µl of TEMED, making a total volume of 5 ml. Protein samples were prepared as described for 12 % gels.

The resolved proteins were visualised by staining in Coomassie blue or transferred by Western blotting onto rehydrated PVDF membranes at 4°C, for 2 hours at 100 volts. The membranes were blocked for 1 hour at room temperature (or 4°C overnight), using a buffer containing the following reagents: 20mM Tris, 137mM NaCl (TBS) pH 7.6, containing 0.1 % Tween 20, 10 % dried milk and 10 % horse serum. Filters were washed 2 times for 15 minutes in TBS/0.1 % Tween 20 and incubated in a 1:200 dilution of anti-serum for 1 hour. The filters were washed 6 times for 5 minutes in TBS/0.1 % Tween 20 after 1 hour and incubated in a 1:400 dilution of goat anti-mouse IgG peroxidase conjugate (Sigma) for 1 hour. The membranes were washed 6 times for 5 minutes in TBS/0.1 % Tween 20 and, the anti-mouse IgG peroxidase conjugate was detected by using the enhanced chemiluminescence (ECL) kit, (Amersham) as described in the manufacturers manual.

2.31 *Modifications used in peptide binding studies in Chapter 6*

Peptide blocking studies were carried out in the following way: total bloodstream trypanosome spun detergent extracts were separated in triplicate on 12 % SDS polyacrylamide gels. The samples were transferred onto PVDF membranes and the membranes were cut into three separate blots after the transfer, resulting in a lane of the separated SDE of bloodstream trypanosomes on each blot. The same procedure as described in section 2.3 was used with the following modifications:

1) Blocking buffer - 20 mM Tris, 137 mM NaCl (TBS), pH 7.6 containing, 0.5 % Tween 20, 10 % dried milk and 10 % horse serum.

2) The blots were blocked for 1 hour at room temperature (or 4°C overnight), using a buffer containing the following reagents: 20 mM Tris, 137 mM NaCl (TBS), pH 7.6 containing, 0.5 % Tween 20, 10 % dried milk and 10 % horse serum. Blots were washed 2 times for 15 minutes in

blocking buffer and incubated in the following solutions: Blot A - a 1:200 dilution of preimmune serum; blot B - 1:200 dilution of immune serum; blot C - 1:200 dilution of immune serum, preincubated for 1 hour at 4°C with 12 μg ml^{-1} of tbrlp-pep. After the incubation period, the membranes were washed in blocking buffer, 6 times for 5 minutes.

3) The membranes were subsequently incubated in a 1:800 dilution of anti-rabbit IgG peroxidase conjugate (Sigma) for 1 hour, washed 6 times for 5 mins in TBS/0.5 % Tween 20 and detected using the ECL (Amersham) method.

Materials and methods used for molecular cloning techniques

The following materials and methods were used in the molecular cloning techniques described in Chapters 4 and 6.

2.4 Preparation of *T. brucei* genomic DNA and RNA

Standard procedures were used in the preparation of bloodstream and procyclic *T. brucei* genomic DNA (Borst *et al*, 1980). RNA was extracted from TREU 869 and EATRO 1125 bloodstream and procyclic trypanosomes using the single step acid guanidinium thiocynate phenol chloroform extraction method of Chomczynski and Sacchi, 1987.

2.5 Growth and transformation of *Escherichia coli*

The *E. coli* strains used routinely were either XL-1 Blue, INVαF' (Invitrogen) or supercompetent cells (SURE) from Stratagene and 25-100 ng of ligated DNA. Transformed cells were grown on LB agar plates supplemented with 0.004 % X-gal and 34 μM IPTG which enabled blue/white colony selection in *E.coli* strains with a *lacZΔM15* mutation (α-complementation of the *lacZDM15* mutation results in blue/white colony selection, where white colonies contain plasmids with inserts). Transformants/overnight cultures were grown in LB (or 2 YT/super broth, to enhance rapid growth of the transformed cells) medium (Sambrook *et al*, 1989), with a final concentration of 100 μg ml^{-1} ampicillin (or 50 μg ml^{-1} kanamycin, when required). Transformations were carried out according to the manufacturers instructions. *E. coli*

strain Y1090(r^-) was transfected with phage and grown on LB plates supplemented with 10 mM $MgSO_4$.

2.6 Plasmid vectors

The following plasmid vectors were used for cloning DNA into plasmids, according to the manufacturers instructions: pCRII (Invitrogen), pTAG (R&D Systems), Bluescript II SK (+) (Stratagene), pTZ18R (Stratagene), pGEX-5X-1, 2 and 3 (Pharmacia Biotech), pMAL-c2 (NEB) and pQE 30 (Qiagen). DNA from cells containing these plasmids was prepared according to the manufacturers instructions.

In cases where there were no instructions, DNA was extracted by using either the Wizard miniprep kit (Promega) or the appropriate Qiagen miniprep kit.

2.7 Plasmid DNA

Plasmid DNA was routinely calf intestinal phosphatase treated (as described in Sambrook *et al*, 1989) before ligating, to prevent self ligation. Ligations were carried out overnight at 16°C using T4 ligase and ligase buffer from either Gibco BRL or NEB. Restriction enzymes used in subsequent manipulations of plasmid DNA, were either from Gibco BRL or NEB. All reactions were set up according to the manufacturers instructions.

2.8 Electrophoresis

0.6-2.0 % agarose gels were prepared from Seakem agarose and run in 40 mM Tris acetate and 1 mM EDTA (1x TAE) containing 0.5 μg ml^{-1} ethidium bromide. Mini or midi gel tanks from Pharmacia or Gibco BRL were used in running the gels. The 1 Kb ladder molecular marker (Gibco BRL) was used routinely, in the determination of molecular sizes on gels. Bands were visualised under UV light and photographed as required.

2.81 *Southern and Northern blots*

Standard procedures were used for the preparation of agarose gels for Southern and Northern blotting. Briefly, for Southern blot analysis, 2-5 μg of genomic DNA from trypanosome strains TREU 869, 927 or STIB 247 and 386, was digested with the appropriate enzymes and fractionated by agarose gel electrophoresis (0.8 %). The gels were washed three times (fifteen minutes per wash) in denaturation buffer and three times (fifteen minutes per wash) in neutralisation buffer (see section 2.15). The DNA was transferred onto nylon filters (same size and placed on top of the gels) overnight by placing the gels on a platform covered with a layer of 3 MM filter paper and with the ends of the filter paper immersed in 20x SSC. The Hybond-N membranes were in turn covered with 3x SSC soaked 3 MM paper, a stack of paper towels and a weight. The DNA was cross-linked to the nylon membranes by placing on a UV transilluminator (DNA side down) for 7 minutes and then was hybridised to *T. brucei* specific ^{32}P labelled probes. Probes (PCR products or double stranded DNA) were prepared using random primed labelling (Prime-it kit, Stratagene), using the manufacturers instructions. Oligonucleotides (Strathclyde University or Cruachem) were end labelled with T4 polynucleotide kinase (Gibco BRL) as described by Sambrook *et al*, (1989). DNA hybridisations were performed as described by Sambrook *et al* (1989). The hybridisations were performed overnight at 65°C (hybridisations were also performed at 60°C and 55°C). The hybridisation solution (Church and Gilbert) contained the following reagents: 1 mM EDTA, 7 % SDS and 0.5 M Phosphate buffer pH 7.0. Membranes were washed 3 times for 15 minutes in 0.1 % SDS, 2x SSC and at 65°C

For Northern blot analysis, 10 μg of the extracted RNA was fractionated on 1 % agarose formaldehyde gels, transferred onto nylon filters, baked for 2 hours and cross-linked in the ultraviolet cross-linker (UVP) for 30 seconds. The RNA was hybridised to ^{32}P labelled probes at 68°C for 24 hours. RNA hybridisations (as described by Sambrook *et al*, 1989) were carried out in the following buffer: 6x SSC, 2x Denhardts reagent and 0.1 % SDS. The blots were washed 3 times for 15 minutes in 0.1 % SDS and 2x SSC. Blots were sealed in polythene bags and exposed to X-Omat XS-1 film at -70°C.

2.9 Polymerase chain reaction

All PCR amplifications were performed using the 11.1 x PCR buffer, which is made up of the following: 45 mM Tris-HCl (pH 8.8), 11 mM ammonium sulphate, 4.5 mM magnesium chloride, 6.7 mM 2-mercaptoethanol, 4.4 μM EDTA (pH 8.0), 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 1 mM dTTP and 113 μg ml^{-1} BSA. 0.9 μl of the 11.1 x PCR buffer was used per 10 μl reaction. The following conditions were routinely used (unless otherwise stated) for each 10 μl reaction: Hot start - 95°C, oligonucleotides - 0.5 μM, genomic DNA - 5-10 ng. Annealing temperatures used were chosen using the Tm of the oligonucleotides. Annealing temperature titrations were performed to determine the optimum annealing temperature required for each reaction. *Taq* or *pfu* polymerase were routinely used in amplification reactions. An example of the conditions used in the amplification of *tbrlp* from genomic DNA is as follows:

Hot start:
95°C - 3 minutes
No. of cycles: 1

Main run:
95°C - 30 seconds
70°C - 30 seconds
72°C - 2 minutes
No. of cycles: 25

Final extension:
72°C - 5 minutes for *Taq* (7 minutes for *Pfu*)
No. of cycles: 1

5' and 3' RACE techniques (Gibco BRL) were performed according to the manufacturers protocol. A slight modification in the use of the 5' RACE method involved the use of a 35-mer spliced leader primer in the PCR amplification steps after cDNA synthesis with an anti sense gene specific primer. The conditions used for this reaction were as follows:

Hot start:
94°C - 5 minutes

80°C - 10 minutes

No. of cycles: 1

Main run:

95°C - 1 minute

44° - 58°C - 1 minute

70°C - 2 minutes

No. of cycles: 30

Final extension:

70°C - 10 minutes

No. of cycles: 1

2.91 *Modifications for long range PCR*

A 15:1 mixture of *Taq:pfu* was used in the amplification reactions. The reactions were performed exactly as described by Barnes, 1994.

2.92 *Chromosomal walking*

The 3' end DNA sequence was obtained by chromosomal walking, and was performed exactly as described by Screaton, Bangham and Bell, 1993. Briefly, genomic DNA was deliberately misprimed by using a single gene specific primer, low annealing temperature and several PCR cycles. The sequence of the gene was obtained by sequencing, using a second gene specific primer downstream of the first gene specific primer. Sequencing reactions were prepared using the fmol sequencing kit (Promega) according to the manufacturers instructions.

2.10 DNA Sequencing

DNA sequencing was performed either manually, using the Sequenase version 2.0 reaction kit (USB) and [α-^{35}S]dATP (Dupont), which makes use of the dideoxynucleotide chain termination method of Sanger *et al* (1977) or, by automated sequencing, using the ABI automated

sequencer. Reactions were performed using double stranded DNA as the template and either vector specific or gene specific sequencing primers. The reactions were prepared according to the manufacturers instructions in both cases. The following gels were used routinely (for manual sequencing): in order to improve resolution at the top half of gels, 4 % gels run in 1 M sodium acetate in 1x TBE (in lower sequencing apparatus tray), or gradient gels were used. 6 % (or 8 %) gels were run in either 1x TBE or 1x TTE. All gels were run for a minimum of 2 or maximum of 6 hours, depending on the amount of sequence required. Gels were fixed for 20 minutes in 10 % methanol and 10 % acetic acid, the fixing solution drained and then, transferred onto Whatman 3 MM filter paper. Gels were dried at 80°C in a Bio-Rad gel dryer and exposed to Kodak X-Omat S film overnight at room temperature.

2.11 cDNA and genomic library screening

Standard methods were used as described in Sambrook *et al*, 1989 with no modifications. Briefly, a 50 ml culture of *E. coli* strain Y 1090, grown to an optical density (650 nm) of 0.5 in LB medium and 50 μg ml^{-1} of ampicillin was prepared for transfection and plating as follows: the pelleted cells were resuspended in 5 ml SM buffer (5.8 g NaCl, 2 g $MgSO_4.7H_2O$, 50 ml 1M Tris-HCL pH 7.5 and 2 % gelatin made up to 1 litre with distilled water). Serial dilutions of the phage stocks were prepared and 100 μl aliquots of each dilution were added to an equal volume of plating cells and incubated in a water bath at 37°C for 15 minutes. 2.5 ml of top agarose (containing 50 μg ml^{-1} of ampicillin) at a temperature of 55°C were added to each tube and immediately poured onto 8.5 cm (diameter) plates containing bottom agar. After overnight incubation at 37°C the plaques were counted to titre the phage stocks. The plaques were blotted onto nylon filters (Hybond-N, 0.45microns, 82 mm, Amersham) in duplicate, labelled, cross-linked in a ultraviolet cross-linker for 30 seconds and hybridised overnight at 65°C as previously described. Filters were washed 3 times for 15 minutes in 1 x SSC and 5 % SDS. The filters were sealed in polythene bags (DNA face up) and exposed to X-Omat XS-1 film at -70°C. A total of 20,000 plaques were screened in each case. The following libraries were screened for the presence of a *ras*-like subfamily gene in *T. brucei*: *Sau* 3A partial genomic DNA libraries of; *Trypanosoma brucei rhodesiense* EATRO 795 ILTat 1.2, *Trypanosoma brucei rhodesiense* EATRO 2340 Pop 339, *Trypanosoma brucei rhodesiense* EATRO 2340 Gug 340 and *Trypanosoma brucei rhodesiense* Eco20Eco complete genomic DNA library, all of which were gifts from Mr V.S. Graham of the Wellcome Unit of Molecular Parasitology. The *T. brucei* strain EATRO 1125 *Sau* 3A cDNA library was a gift from Professor E. Pays.

2.12 Preparation of anti peptide sera

A desalted and lyophilised keyhole limpet haemocyanin (KLH) conjugated peptide (Genosys Biotechnologies), was used in raising anti peptide sera, as recommended by the manufacturer. The conjugated peptide was dissolved in PBS at a concentration of 1 mg ml^{-1}, an equal volume of a suspension of Alum (Pierce Ltd) was added dropwise with stirring and the suspension stirred for 30 minutes. Two New Zealand white rabbits were each immunised with 500 μg of the peptide in alum by multiple site subcutaneous injections. After four weeks, boosts of 250 μg peptide suspension in alum were given to each rabbit at intervals of three weeks and test bleeds made 8-10 days after each boost, until a satisfactory immune response was observed.

The following expression systems were used in the production of fusion proteins from *tbrlp*: The GST fusion protein Expression System (Pharmacia), The Maltose Binding Protein Purification System (NEB) and the Qiaexpressionist Fusion Protein System (Qiagen). All fusion proteins were prepared as described in the manufacturers protocol.

2.13 Computer programs used in sequence and phylogenetic analysis

Nucleotide and peptide sequence analysis was carried out using the GCG and Phylip 3.5 suites of programs. Briefly, the GCG program MAP, was used to translate nucleotide sequences into peptide sequences and for checking that they were in frame. FASTA and BLAST searches were conducted to find related sequences from GENEMBL and SWISSPROT databases. Sequences from different clones were assembled using the GELASSEMBLE program and visualised using PRETTYOUT. In this way the sequences from different clones could be compared and corrected. MOTIFS was used to determine the significant motifs present in the sequence of *tbrlp*. PROFILE was used to determine the predicted molecular size and isoelectric point.

For phylogenetic analysis, sequences were aligned using the GCG PILEUP program and manually corrected using the GCG LINEUP program. TOPHYLIP was used in formatting the aligned sequences into input files for PHYLIP 3.5 phylogeny inference programs. SEQBOOT was used to bootstrap the input file from TOPHYLIP. Both bootstrapped and non bootstrapped trees were used in the generation of distance matrix trees from PROTDIST using the Dayhoff PAM option, maximum parsimony using PROTPARS and maximum likelihood trees using the PUZZLE program. Trees were constructed from the output file of PROTDIST using the NEIGHBOR program. All three options from the NEIGHBOR program were used: Neighbor

Ethidium Bromide: 10 mg ml^{-1} stock in distilled water. Final working concentration 0.5 μg ml^{-1}

IPTG: 0.5 M stock in sterile distilled water. Stored at -20°C.

Herrings testes DNA: 10 mg ml-1 stock in sterile distilled water and sheared with a hypodermic needle and syringe (19 gauge). Stored at -20°C.

Kanamycin: 50 mg ml^{-1} in sterile distilled water. Stored at -20°C.

L-Broth: 1 % Bactotryptone (Difco), 0.5 % yeast extract (Difco), 0.5 % NaCl in distilled water. Stored at room temperature.

L-Broth agar: 1 % Bactotryptone (Difco), 0.5 % yeast extract (Difco), 0.5 % NaCl, 1.5g 100ml-1 of bacto-agar (Difco) in distilled water. The agar is sterilised by autoclaving, and antibiotics added if required, after cooling to 55°C. Plates stored at 4°C.

Neutralisation buffer: 1.5 M NaCl, 0.5 M Tris-HCl pH 8.0 in distilled water. Stored at room temperature.

SDS-PAGE Sample Buffer (4x): 0.5 M Tris-HCl pH 6.5, 0.08 % SDS, 0.2 M EDTA, 0.4 % glycerol in distilled water.

SDS: 10 and 14 % stock solutions in sterile distilled. Stored at room temperature.

SM Buffer: 50mM Tris-HCl (pH 7.5), 0.1 M NaCl, 8mM $MgSO_4$, 0.01 % gelatin in distilled water. Autoclaved and stored at room temperature.

SSC (20x): 3 M NaCl, 0.3 M tri-sodium citrate pH 7.0 (using NaOH) in distilled water. Stored at room temperature.

Super Broth: 3.5 % Bacto-tryptone, 2 % Bacto-yeast, 0.5 % NaCl pH 7.5 in distilled water. Autoclaved and stored at room temperature.

TAE (50x): 40 mM Tris-acetate, 1 mM EDTA in distilled water.

TBE (10x): 0.9 M Tris-HCl, 0.9 M Boric acid, 25 mM Na_2EDTA in distilled water. Stored at room temperature.

TE Buffer: 1 mM EDTA, 10 mM Tris-HCl, pH 7.5, in distilled water. Autoclaved and stored at room temperature.

Tetracycline: 12 mg ml^{-1} in 50 % sterile distilled water and 50 % ethanol. Stored at -20°C.

TTE (20x): 1.8 M Tris-HCl, 0.6 M Taurine, 0.01 M Na_2EDTA in distilled water. Stored at room temperature.

X-gal (100x): 2 % dimethylformamide (DMF). Wrapped in aluminium foil and stored at -20°C.

2xYT Broth: 1.6 % tryptone (Difco), 1 % NaCl, 1 % yeast extract (Difco), pH 7.6 with NaOH. Autoclaved and stored at room temperature.

Chapter 3

Detection of monomeric GTP binding proteins in *Trypanosoma brucei*

CHAPTER 3

DETECTION OF MONOMERIC GTP-BINDING PROTEINS IN *Trypanosoma brucei*

3.0 Introduction

A large number of members of the *ras* superfamily of monomeric GTP-binding proteins have been identified in a wide range of eukaryotic organisms, including some as widely divergent as *Dictyostelium*, yeast and man (Lowy and Willumsen, 1993). A *ras* GTPase is defined as a small molecular weight GTP binding protein (i.e within a molecular weight range of 20-29 kDa) which is capable of binding to and hydrolysing GTP. The general structure of *ras* GTPases can be roughly divided into three domains, the GTP binding or catalytic domain, the C-terminal extension domain and the carboxyl terminal isoprenylation domain. Two of these domains are essential for normal *ras* activity, i.e the GTP binding domain and the carboxyl terminal isoprenylation site. *ras* GTPases are involved in a range of cellular functions which include; cytoskeletal organisation, intracellular vesicular transport, protein translocation (Hall, 1990; Bokoch and Der, 1993) and the control of cell growth and differentiation (Barbacid, 1987). They are also known to serve as central regulatory points in a number of signalling pathways which are essential for normal cellular function (Hall, 1990). Utilising both GTP hydrolysis and GDP/GTP exchange reactions, monomeric GTP-binding proteins act as two-way molecular switches (see introduction) which are active in the GTP bound state and inactive in the GDP bound state.

A number of studies on *Trypanosoma brucei* have provided evidence for the presence of both heterotrimeric and small monomeric GTP-binding proteins. Using a GTP-binding assay, Coulter and Hide (1995) identified a number of small GTP-binding proteins in *T. brucei*. Furthermore, both full length and partial cDNA sequences for members of the *rab* (El Sayed *et al*, 1995; Field and Boothroyd, 1995) and *ran* (Field *et al*, 1995) GTPase subfamilies have been isolated from *T. brucei*.

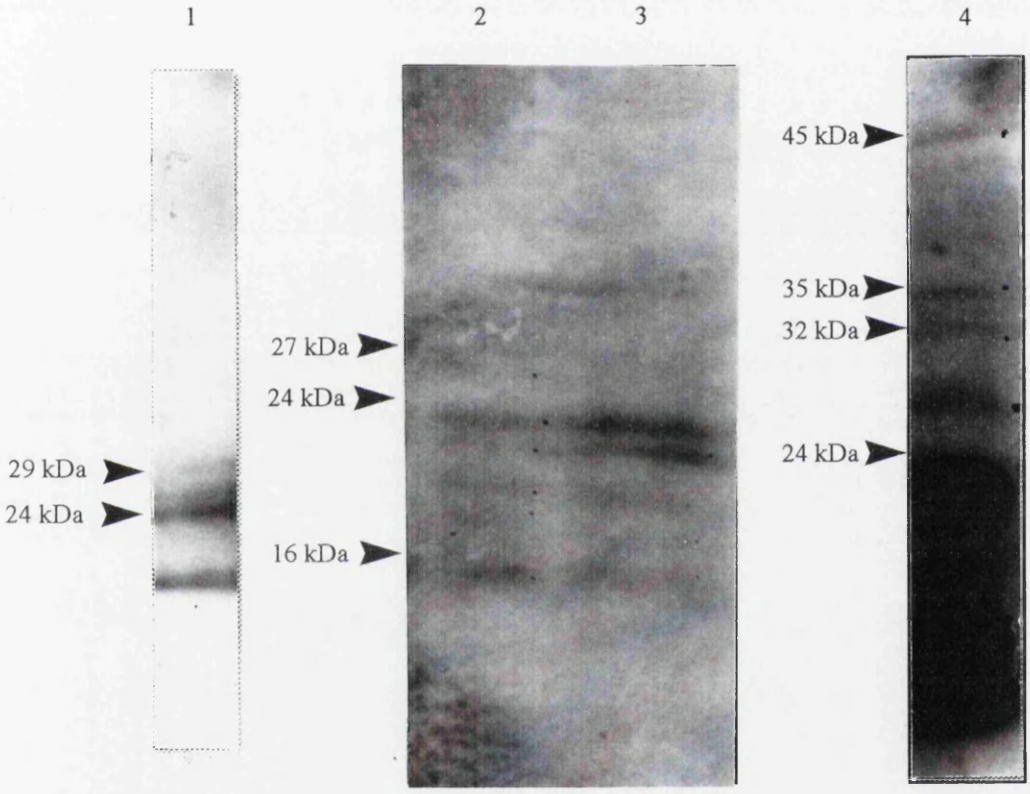
To date, GTPases belonging to the *ras* subfamily have not yet been identified in *T. brucei*. Using biochemical and immunological methods, the experiments described in this chapter, were designed to determine whether evidence could be obtained for the existence of *ras*-like GTP-binding proteins in *T. brucei*.

3.1 Detection of GTP-binding proteins in *T. brucei*

Using an *in situ* GTP-binding assay, Coulter and Hide (1995) identified trypanosome polypeptides which bind GTP. The authors postulated that these polypeptides, of molecular mass 14 - 29 kDa, could represent members of the *ras* GTPase superfamily of GTP-binding proteins. As the sizes of mammalian $p21^{ras}$ and *ras* proteins from several other organisms (Valencia *et al*, 1991) are consistent with the sizes of the polypeptides identified by Coulter and Hide (1995), such an hypothesis is plausible. To confirm these results, the GTP-binding assay of Coulter and Hide was repeated to investigate whether GTP-binding polypeptides of the 14 - 29 kDa molecular size range could be detected in *T.brucei*.

Detergent extracts (see Materials and Methods) were made from both bloodstream and procyclic stage trypanosomes. The extracts (supernatant and pellet) were fractionated by SDS polyacrylamide gel electrophoresis (SDS-PAGE), transferred onto PVDF membrane and incubated with [α^{32}-P]GTP (as described in Materials and Methods). Due to the ability of *ras*-like proteins to catalyse GTP hydrolysis, [α^{32}-P]GTP is used for this assay instead of the more commonly used [γ^{32}-P] to prevent the loss of radioactive phosphate by hydrolysis. Figure 3.0 shows an example of a typical set of results. The following radioactively labelled bands were detected: In lane 1 (procyclic stage pellet) 22 kDa, 24 kDa, 26 kDa, 29 kDa; lane 2 (procyclic stage, supernatant), 16 kDa, 20 kDa, 22 kDa and 24 kDa; lane 3 (bloodstream stage pellet), 16 kDa, 18 kDa, 20 kDa, 22 kDa, 24 kDa, and 27 kDa while the following bands were observed in lane 4 (bloodstream stage supernatant), 18 kDa, 20 kDa, 22 kDa, 24 kDa, 26 kDa, 28 kDa, 32 kDa, 35 kDa and 45 kDa. Replicate experiments produced similar results. The GTP-binding polypeptides identified in a series of experiments are summarised in Table 3.0.

IGURE 3.0



Detection of GTP binding proteins in *T. brucei*.

Figure 3.0 shows the results compiled from 3 *in situ* GTP binding assays performed on SDS-PAGE fractionated *T. brucei* spun detergent extracts. The following *T. brucei* extracts were fractionated and assayed: in lane 1, procyclic stage pellet; lane 2, procyclic stage supernatant, lane 3, bloodstream pellet and in lane 4, bloodstream supernatant. Sigma molecular markers were used in the determination of the molecular sizes of the observed bands.

TABLE 3.0

Summary of sizes of GTP-binding proteins in T. brucei identified by in situ binding assay

	BLOODSTREAM		PROCYCLIC	
MOLECULAR SIZE (kDa)	SUPERNATANT	PELLET	SUPERNATANT	PELLET
16✠		◆	◆	
18♦	◆	◆		
20✠	◆	◆	◆	
22✠	◆	◆	◆	◆
24✠	◆	◆	◆	◆
26✠	◆			◆
27♦			◆	
28♦	◆			
29♦				◆
32♦	◆			
35♦	◆			

♦ STAGE SPECIFIC

✠ NON-STAGE SPECIFIC

◆ GTP-BINDING POLYPEPTIDE PRESENT

Most of the bands listed in Table 3.0 were observed in all replicate experiments although occasionally some were not observed in the same extracts in all experiments. Their absence may have been due to either the lack of good separation of the polypeptides on SDS polyacrylamide gels, in which case two bands may have been mistaken for a single band, a lack of intensity of signal, which may have resulted in the band being overlooked or a slight difference in sample preparation.

These results confirm those of Coulter and Hide (1995) and the observed molecular sizes (18 kDa-35 kDa) are similar to those of higher eukaryotic *ras*-like proteins (Valencia *et al*, 1991 and Kahn *et al*, 1992).

While many cellular proteins are capable of binding both ATP and GTP, the monomeric GTPases of the *ras* subfamily only bind GTP. It was therefore important to confirm the

specificity of the GTP-binding reaction. If the GTP-binding proteins had sole specificity for GTP, binding would be competed out by excess unlabelled GTP but not by excess unlabelled ATP. To test this the following GTP-binding assays were carried out: spun detergent extracts of bloodstream and procyclic stage trypanosomes were run on 15% SDS polyacrylamide gels and transferred to PVDF membranes as previously described. The samples were run in triplicate thus producing three identical sets of immobilised extracts and each membrane was incubated under different conditions as follows: Panel A, 3μCi [α^{32}-P]GTP + 10 μM unlabelled ATP; Panel B, 3μCi/ml (0.1 μM) [α^{32}-P]GTP; and Panel C, 3μCi [α^{32}-P]GTP + 10μM unlabelled GTP. Figure 3.1 shows an autoradiograph obtained from one of these experiments. All the bands on panel A (assay in the presence of 10 μM unlabelled ATP), can be seen on panel B (radiolabelled GTP alone). No bands were observed on panel C (GTP-binding assay in the presence of 10 μM unlabelled GTP), which shows that radiolabelled GTP-binding was competed out by excess unlabelled GTP. However, approximately eight of the radiolabelled GTP-binding bands were competed out by excess unlabelled ATP (compare the intensities of the bands present in panel A to those on panel B) and therefore not all the bands observed on panel B are present on panel A. Bands of the following sizes were competed out by the unlabelled ATP: 43, 55, 17 and 19 kDa from the bloodstream supernatant and pellet respectively and 18, 20, 17 and 19 kDa from the procyclic supernatant and pellet respectively. Therefore not all the aforementioned bands, summarised in Table 3.0, are strictly GTP binding proteins due to their additional ability to bind ATP. The remaining polypeptides, ranging from 20 - 36 kDa, however, bind specifically to GTP. This experiment, therefore confirms that GTP-binding proteins with high specificity for GTP are found in *T. brucei*.

FIGURE 3.1

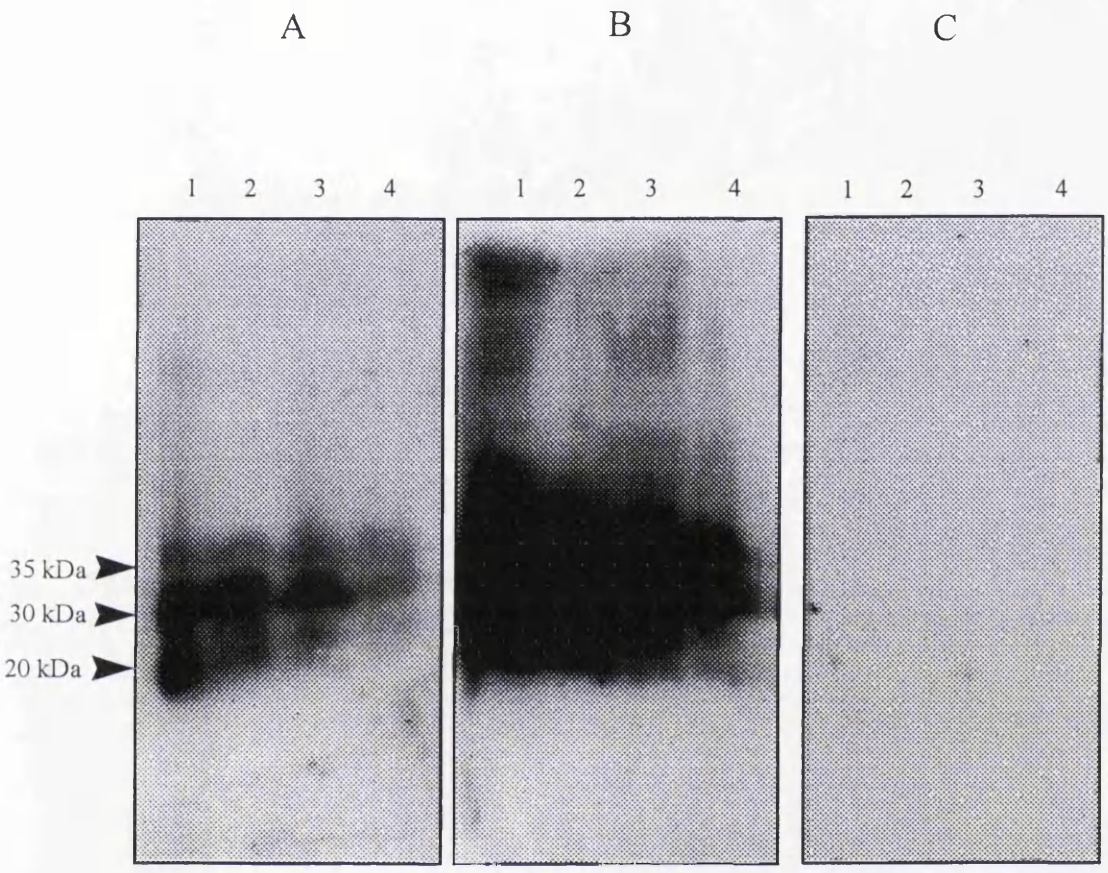


Figure 3.1 is an autoradiograph of GTP binding proteins separated by SDS-PAGE, showing 3 membranes, A , B and C with immobilised, bloodstream supernatant (1), procyclic supernatant (2), bloodstream pellet (3) and procyclic pellet (4). Panel A was incubated in 3 μCi of ^{32}P labelled GTP + 10 μM unlabelled ATP. Panel B was incubated in 3 μCi of ^{32}P labelled GTP. Panel C was incubated in 3 μCi of ^{32}P labelled GTP + 10 μM of unlabelled GTP.

3.2 Purification of GTP-binding proteins

To further identify and characterise the properties of the monomeric GTP binding proteins from *T. brucei*, it is necessary to generate reagents, such as gene clones or antibodies. A first step towards obtaining antibodies is to purify a monomeric GTP-binding protein (the gene cloning approach is the subject of Chapter 4). Two basic requirements are necessary for the purification of GTP-binding proteins: (i) a GTP-binding assay system suitable for analysing a large number of chromatography fractions and (ii) the development of a suitable fractionation system.

3.21 *Development of a routine GTP-binding assay*

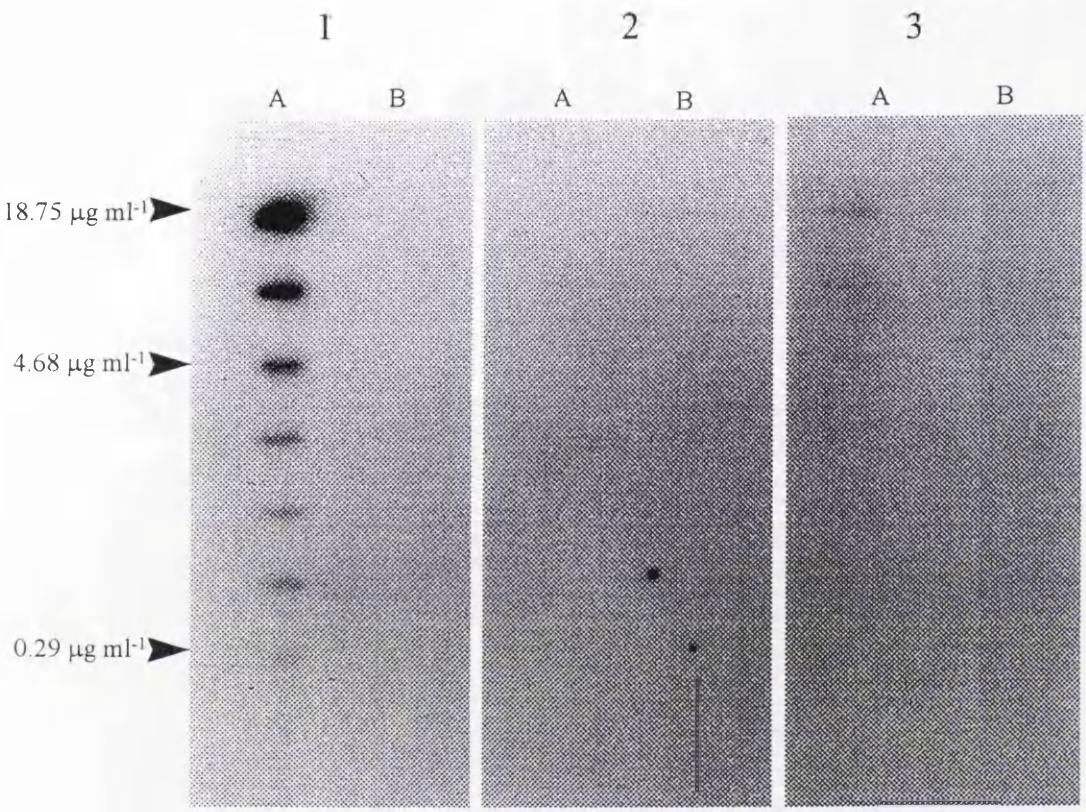
An ideal assay for HPLC purified GTP-binding proteins, would be one which could be used to concentrate and screen large numbers of fractions simultaneously, as the system previously described (Coulter and Hide 1995 - see Chapter 3, section 3.1) may not have sufficient sensitivity to identify GTP-binding proteins in the dilute fractions produced by HPLC fractionation. Furthermore, the large number of fractions generated would entail running and blotting a large number of gels which would be time consuming, cumbersome and require the use of considerable amounts of radioactivity.

Based on these requirements, an assay was developed which made use of the slot blot technique. PVDF protein binding membranes were cut to size and prepared as described in the Materials and Methods section, placed (individually) on the lower half of the HibriSlot manifold [Gibco BRL 24-well filtration manifold] and the upper half of the manifold was screwed down onto the lower half. The 24-well manifold allows samples of up to 200 µl to be drawn through the PVDF membrane under vacuum. Wells were loaded with 100 µl from each HPLC fraction and connected to a vacuum pump. The shape of the wells is such that proteins contained within each sample are concentrated onto a 5 mm^2 area of the membrane after brief application of a vacuum. To test this approach, 100µl aliquots of decreasing concentrations of BSA (which does not bind GTP) were loaded into the first row of 12 wells and serial dilutions of a bloodstream supernatant loaded into the second row of 12 wells. After the vacuum was applied, the proteins were concentrated and immobilised onto the PVDF membrane. Decreasing concentrations of bloodstream supernatant and BSA on PVDF membrane were made in triplicate and incubated in [α^{32}-P]GTP (Figure 3.2 panel 1), [α^{32}-P]GTP + 10µM unlabelled GTP (Figure 3.2, panel 2) and [α^{32}-P]GTP + 10µM ATP (Figure 3.2, panel 3). In this way both the sensitivity of the system

and the specificity of the GTP-binding reaction were determined. Figure 3.2 shows that GTP-binding activity can be detected using this system, and furthermore the unlabelled GTP completely abolished the α-^{32}P[GTP] binding activity of the test sample, while the unlabelled ATP removed a large amount of activity. The results show that in the presence of unlabelled ATP, a minimum protein concentration of 9.37 μg/ml is required for the detection of specific GTP-binding. BSA showed no GTP-binding activity.

A total protein content of 20 μg/well is required for the routine detection of GTP-binding proteins by SDS PAGE and blotting (Coulter and Hide 1995), whereas, only 0.937 μg is required for the detection of specific GTP-binding using the slot blot technique. The results suggest that the slot-blot GTP-binding assay is suitable for the specific assay of GTP-binding activity. In addition to this, the system can be used to concentrate and screen large numbers of dilute HPLC samples, making it ideal for GTP-binding assays of HPLC fractionated trypanosome extracts.

FIGURE 3.2



A slot blot assay was used to test HPLC fractions for the ability to bind GTP. The panels had two rows (A and B), of decreasing protein content. The sample in row A is a bloodstream spun detergent extract (SDE) supernatant. 100 μl aliquots of decreasing concentrations equivalent to final protein concentrations of 18.75, 9.37, 4.68, 2.34, 1.17, 0.58 and 0.29 μg ml^{-1} respectively were loaded into the wells. Row B contained the negative control, bovine serum albumin (BSA). The highest protein content of the negative control was 60 μg and the lowest 0.23 μg ml^{-1}. Panel 1 was incubated in 3 μCi/ml of ^{32}P labelled GTP; panel 2 in 3 μCi/ml of the radiolabelled GTP + 10 μM unlabelled GTP and panel 3 in 3 μCi/ml (0.1 μM) of radiolabelled GTP + 10 μM ATP.

3.22 *Development of a fractionation system.*

A number of different types of HPLC fractionation systems are available for protein purification such as, anion exchange, cation exchange, size exclusion, hydrophobic interaction and affinity based methods. As little is known about the properties of *ras*-like GTP-binding proteins in *T. brucei* (size and hydrophobicity etc), it was decided to use anion exchange chromatography as a first approach. A number of features of anion exchange chromatography make it suitable for a first purification step: (1) it has a high binding capacity; (2) all charged proteins can bind to it; (3) specific elution can be achieved by using appropriate pH or ionic conditions and (4) a relatively high concentration can be maintained in eluted fractions.

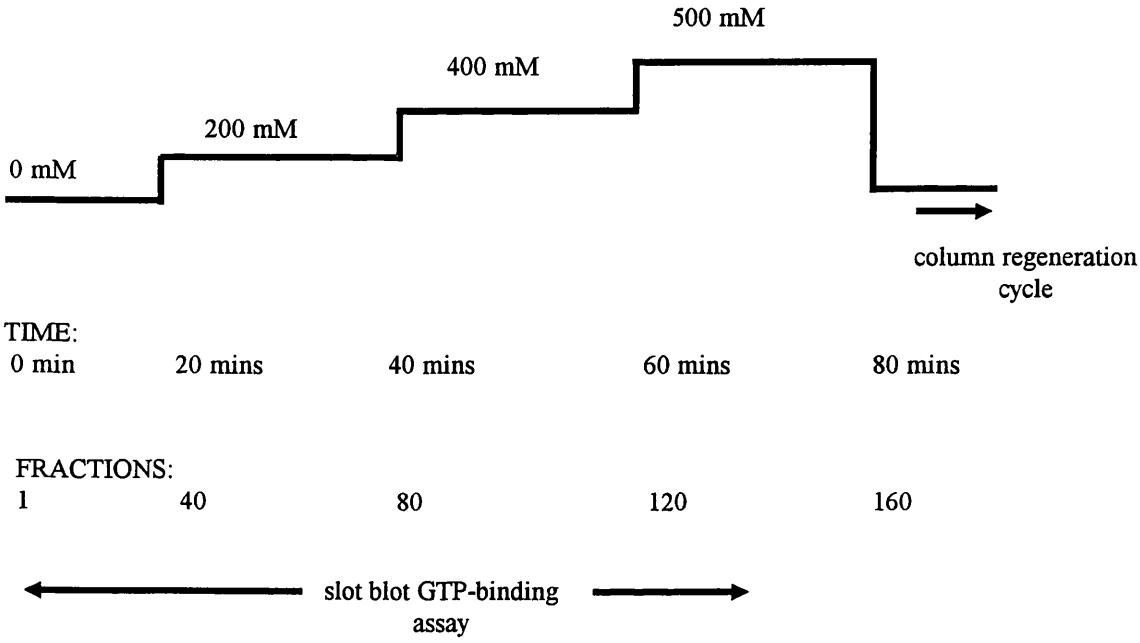
To determine conditions suitable for the elution of GTP-binding proteins, an initial HPLC run was carried out using a DEAE-Cellulose anion exchange column [Waters AP-1; prepacked with Protein-Pak DEAE 8HR resin; internal diameter 1 cm]. In the first instance, a NaCl step gradient elution strategy was employed to determine approximate elution conditions. The column was equilibrated with 20 mM Tris - HCl at pH 7.6 and elution was achieved by addition of NaCl (Figure 3.3). 7.5 mg of the supernatant from a spun detergent extract (bloodstream) was applied to the column. A buffer flow rate of 0.8 ml/min was maintained throughout the run and the sodium chloride concentration increased in a step-wise manner at 20 min intervals over a 1.3 hour period from 0 mM to 200 mM, 400 mM and finally 500 mM at 20, 40 and 60 min respectively (Figure 3.3). The column was regenerated as described (see Materials and Methods). Fractions were collected at 0.5 min intervals into eppendorfs containing 1 µl aliquots of a cocktail of protease inhibitors, consisting of 2.5 µg/ml aprotinin, 1 mM benzamidine and 2.5 µg/ml leupeptin (see Materials and Methods). A total of 160 fractions were collected and fractions 1-132 (see Figure 3.3 for details) were assayed for GTP-binding activity, using the slot blot system immediately after collection. Fractions 133-160 were stored at -70°C.

Following autoradiography, the radioactive signal from each individual slot, from the slot blot assay, was scored on a scale ranging from - to ++++, where - represents no signal and + to ++++ indicate increasing degrees of signal intensity. Only fractions with labelling intensities greater than ++ were used in the subsequent GTP-binding assays. Fractions 59 - 67 were eluted within a NaCl concentration range of 200 mM - 400 mM and had a labelling intensity of ++++. The fractions obtained after the step-wise increase of the NaCl concentration from 200 mM to 400 mM showed labelling intensities less than ++ and were, therefore, stored at -70°C for subsequent analysis. Results from this HPLC run showed that most proteins with high specificities for GTP-binding were eluted within a NaCl concentration range of 200 mM - 400 mM (see figure 3.3).

A linear NaCl gradient elution regime was carried out to further improve the separation of GTP-binding proteins. A gradient of 0 mM - 250 mM was chosen as results from the step gradient showed that most GTP-binding proteins were eluted following the 200 mM NaCl step. 7.25 mg of bloodstream stage supernatant spun detergent extract was applied to the column, and the salt concentration linearly increased from 0 mM - 250 mM, after which the concentration was maintained at 250 mM for 30 min to ensure complete elution of the GTP binding proteins. A final salt concentration of 500 mM NaCl (Figure 3.4) was maintained for 20 minutes to ensure that all proteins were eluted from the column in preparation for the regeneration of the column. Fractions were collected as previously described. Fractions 1-96 corresponded to the fractions eluted during the linear increase in salt concentration (Figure 3.4) and were assayed for GTP-binding activity.

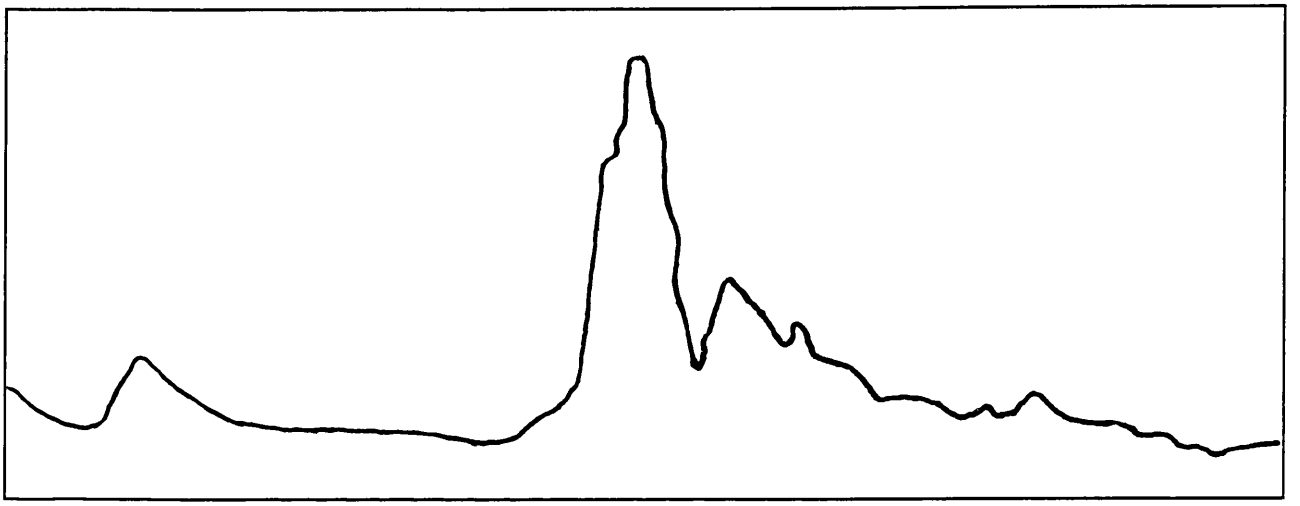
Figure 3.5 shows the fractions positive for GTP-binding, B 2-5, E 8-12, F 1-12, G 1-12 and H 1-10 obtained from the linear HPLC run. The positive fractions correspond to HPLC fractions 14 - 17 (B 2 -5) and 56 - 93 (E 8 - 12, F 1 - 12, G 1 - 12 and H 1 - 9), which were eluted within a NaCl concentration range of 150 - 200 mM. However only fractions 59 - 81 (which correspond to fractions E 11 - 12, F 1 - 12 and G 1 - 9), were subsequently separated by SDS polyacrylamide electrophoresis and reanalysed for GTP-binding activity, to assess the level of purity of the fractions, as they were observed to have labelling intensities greater than ++. The fractions positive for GTP-binding and the corresponding HPLC fraction numbers are summarised in Table 3.1.

FIGURE 3.3



A profile of the step NaCl gradient used in the separation of GTP-binding proteins from the spun detergent lysates of bloodstream trypanosomes. The concentration was increased from 0 mM to 200 mM, 400 mM and finally equilibrated at 500 mM at time intervals of 20, 40, 60 and 80 minutes respectively. The column was put through a regeneration cycle after 80 mins. Samples 1 - 132 were immediately assayed for GTP-binding, while samples 133 - 160 were stored at -70°C.

FIGURE 3.4



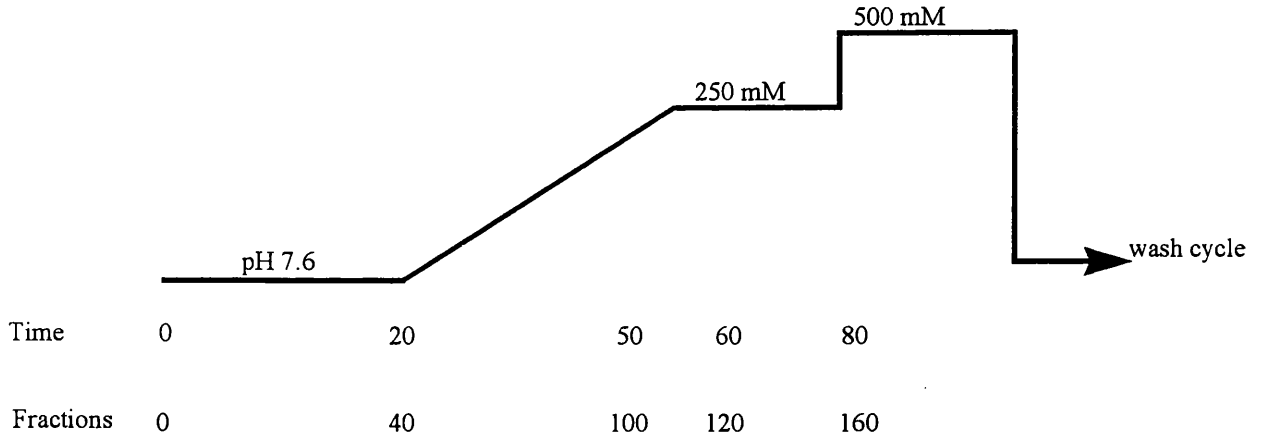


Figure 3.4 is an illustration of the relationship between the fractions derived from the linear gradient HPLC run (profile) and the peaks at which they were eluted from the bloodstream spun detergent extract The NaCl concentration was maintained at 250 mM and then 500 mM to ensure the complete elution of the proteins. Fractions 1-96 were assayed for GTP-binding using the slot blot technique. Time - in minutes. After the purification the column was regenerated (wash cycle).

FIGURE 3.5

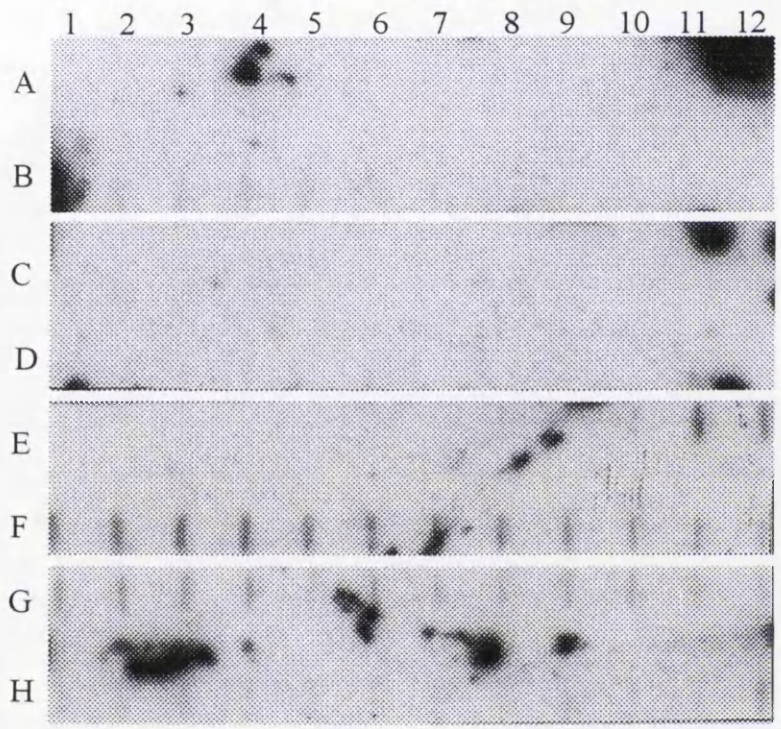


Figure 3.5 shows the results from a slot blot GTP-binding assay of the HPLC purified fractions. 4 panels of PVDF membrane were used in the concentration and immobilisation of the fractions. Each panel had two rows of concentrated protein, labelled A-H. Each row had 12 slots, numbered 1-12 and each slot was loaded with 200 µl of the HPLC derived GTP-binding protein fractions. Fractions E11-12 (HPLC fractions 56-60), F1-12 (HPLC fractions 61-72) and G1-12 (HPLC fractions 73-84) were subsequently assayed for GTP binding activity after separation on 12 % SDS polyacrylamide gels.

Table 3.1

Summary of positive GTP-binding fractions and their corresponding HPLC (linear gradient) fraction numbers.

POSITIVE GTP-BINDING FRACTIONS	CORRESPONDING LINEAR HPLC FRACTIONS
B 2 - 5	14 - 17
• E11 - 12	56 - 60
• F 1 - 12	61 - 72
• G 1 - 12	73 - 84
H 1 - 9	85 - 93

- POSITIVE SLOT BLOT GTP-BINDING FRACTIONS, WHICH WERE SUBSEQUENTLY FRACTIONATED ON SDS POLYACRYLAMIDE GELS AND ASSAYED IN ORDER TO CONFIRM THEIR DEGREE OF PURITY.

In order to confirm that the fractions, eluted by HPLC anion exchange chromatography, contain GTP-binding proteins, the positive fractions identified by the slot blot technique were assayed using the *in situ* GTP binding assay. 30 µl of each of the HPLC positive fractions E 11 - 12, F 1 - 12, G 1 - 12, were loaded on SDS polyacrylamide gels and transferred to PVDF membrane and incubated with α -^{32}P labelled GTP. Autoradiography was carried out overnight and the data analysed. Linear HPLC gradient fractions 73-81, which correspond to positive GTP-binding fractions G 1-9 (Figure 3.5), were each observed to contain a 24 kDa band (Figure 3.6A). All other fractions contained no radiolabelled bands. The positive fractions were also analysed by SDS polyacrylamide gel electrophoresis followed by silver staining, to determine the profile of proteins present in each fraction, in order to assess the level of purity of the eluted 24 kDa polypeptide. Figure 3.6B shows that a minimum of six other polypeptide bands were in each of the fractions. No silver stained bands were observed at the expected position of the 24 kDa polypeptide. As silver staining detects polypeptides at levels as low as 10 ng (Guillmette and Lewis, 1983), this suggests that the amount of the 24 kDa polypeptide present in each fraction is less than 10 ng.

Despite the fact that several other proteins are detected in the various HPLC derived fractions, the 24 kDa small GTP-binding protein is not present in high enough concentrations to be detected by silver staining. However, partial purification of one of the GTP-binding

polypeptides was achieved and this was detected by the *in situ* GTP-binding assay. The low amounts of protein present in each fraction indicated that large amounts of trypanosome material would be required for a full scale purification and it was, therefore, decided at this point to investigate other possible means of purification.

FIGURE 3.6

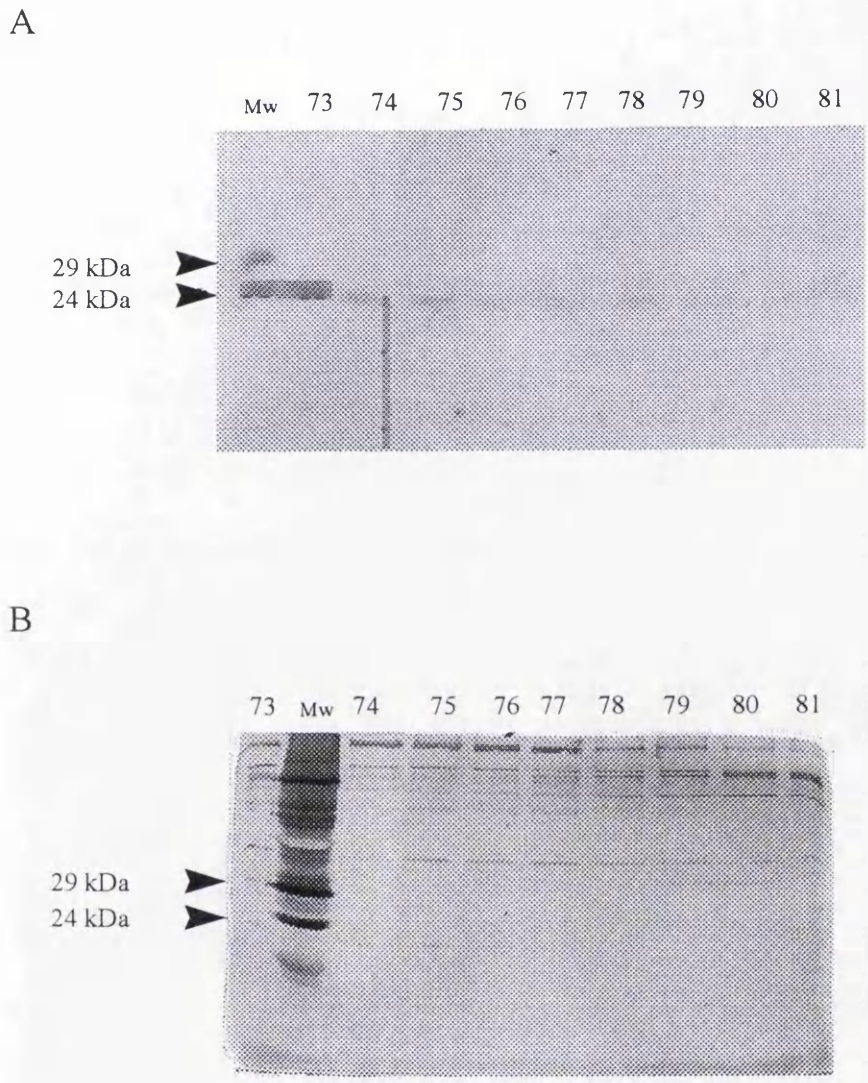


Figure 3.6 (A), shows the results of an *in situ* GTP-binding assay of anion exchange HPLC purification of *T. brucei* spun detergent extracts. Fractions 73 - 81 were positive for GTP-binding, each of the fractions show a radiolabelled band with a molecular weight (Mw) of approximately 24 kDa. Figure 3.6 (B) is a silver stained gel of fractions 73 - 81. The 24 kDa GTP-binding protein observed in (A), is not observed on the silver stained gel.

3.3 Detection of *ras*-like GTP-binding proteins in *T. brucei* by immunoblotting with heterologous antibodies.

Many reagents are available for the analysis of *ras*-like GTP-binding proteins in higher eukaryotes. These reagents include a number of antibodies (both monoclonal and polyclonal) which recognise higher eukaryotic *ras*-like proteins. Immunoblots with heterologous antibodies have previously been used to identify other signalling molecules in *T. brucei* e.g Protein kinase C (Keith *et al*, 1990) and the $G_{o\alpha}$ subunits of G-proteins (Coulter and Hide, 1995). If such an approach could be used with antibodies to *ras*-like proteins, it would permit the identification of *ras*-like proteins in *T. brucei* crude extracts and also determine whether the 24 kDa GTP-binding protein cross reacts with *ras*-like polypeptides. A number of considerations need to be taken into account when choosing a suitable antibody: (1) The need to distinguish *ras* subfamily members from other members of the *ras* superfamily e.g. *rab, ran, rho* etc and (2) The need to maximise the chances of the antibody cross - reacting with a homologue (which may be evolutionarily divergent) in *T. brucei*. Based on these criteria, the ideal antibody would be one which recognises a highly conserved region of *ras* (an enzymatic/catalytic active site), which is absent or altered in other *ras* family proteins. On this basis, the monoclonal antibody which was chosen, was highly specific for mouse H- K- and N-*ras* oncogene proteins (anti-*ras* pan, Boehringer Mannheim Biochemica). Goat anti-mouse IgG peroxidase conjugate (Sigma Immunochemicals) was used as a second antibody for the immunoblotting experiments.

To determine whether any of the HPLC anion-exchange fractions contained *ras*-like polypeptides, fractions 1 - 96 from the HPLC run (step-wise NaCl elution gradient, 0 - 400 mM; see section 3.22, Figure 3.3) were examined by immunoblotting with anti- *ras* pan. All fractions derived from the initial HPLC purification were concentrated and immobilised on PVDF membranes by slot blotting and incubated in blocking buffer overnight (see Materials and Methods). After blocking, the membranes were incubated in a 1:400 dilution of the primary antibody in blocking buffer for 1 hour at room temperature. After two 15 minute washes in TBS-Tween, the membranes were incubated in a 1:400 dilution of the second antibody in blocking buffer for 1 hour at room temperature. The membranes were washed, as described above, and the ECL detection method (see Materials and Methods) was used to assay whether the anti-*ras* pan antibody bound to any of the fractions. The results, shown in Figure 3.7, indicate that fractions A5-10, C4-12, D4-10, F3-10, G4-5, 7-8 &10-11, F1-8, G5-11 and H6-10 cross reacted with anti-*ras* pan antibodies, suggesting that these fractions contain *ras*-like polypeptides. In many cases it has been observed that heterologous antibodies can sometimes have cross-

reactivities with many polypeptides in *T. brucei* (Coulter and Hide, personal communication). It was, therefore, important to determine the specificity of the antibody.

Results from the immunoblot analysis of the HPLC fractions, indicated the presence of *ras* like GTPases in fractions which had previously been shown to lack GTP binding proteins by both the slot blot and *in situ* (Coulter and Hide 1995) GTP binding assays e.g fractions A5-9 (refer section 3.3, Figure 3.7). This was a clear indication that the monoclonal antibody may have crossed reacted with polypeptides other than *ras* GTPases in *T. brucei*. As anti-*ras* pan normally recognises a 21 kDa polypeptide, the following experiment was performed in order to determine (a) the molecular size(s) of the polypeptide(s) in *T. brucei* which cross react with anti-*ras* pan; and, (b) the trypanosomal polypeptides which are specifically recognised by the antibody, using preimmune serum as the negative control.

Bloodstream spun detergent extracts (SDEs), were separated by SDS polyacrylamide gel electrophoresis and transferred onto nitrocellulose membranes. This procedure resulted in three sets of identical panels of immobilised bloodstream detergent extracts: Panel A 1, from the first set of membranes, was incubated in a 1:400 dilution of the first antibody (anti-*ras* pan) and the other panel (A 2), was incubated in preimmune serum, and therefore served as a "negative" control. Both panels (A 1 & 2) were subsequently incubated in a 1:1600 dilution of the second antibody . Panels numbered 1 from the second (B) and third (C) sets were also incubated in a 1:400 dilution of the primary antibody, whilst Panels B 2 and C 2 were incubated in preimmune serum, and therefore served as "negative" control membranes. Panels B (1 & 2) and C (1 & 2) were incubated in 1:4000 and 1:10,000 dilutions of the second antibody respectively. In a separate experiment, performed with the aim of determining whether the observed cross reactivity was due to the second antibody recognising epitopes other than that of anti-*ras* pan, bloodstream SDEs were immobilised on three separate membranes (D 1, D 2 & D 3) and incubated in the following dilutions of the second antibody, 1:10,000, 1:4000 and 1:1600 respectively, without prior incubation in the first antibody/preimmune serum. The resulting bands were visualised by the ECL detection method (see Figure 3.8). After a 1 minute exposure, polypeptides with molecular sizes greater than 21 kDa, were observed on almost all the membranes. The observed bands, however, did not correspond to the expected molecular size of small GTP binding proteins, i.e 20 - 29 kDa. Even though no positive bands were observed on the membrane incubated in the 1:10,000 dilution of the second antibody in the absence of the first antibody (D 1), the results showed that the reactions were not specific for *ras* GTPases in *T. brucei*. At dilutions lower than 1:10,000, the second antibody appeared to cross react with a component of the nitrocellulose membrane. Further manipulation of the dilutions of the antibodies i.e. using lower dilutions of the first and second antibodies and washing the membranes with the blocking buffer instead of

TBS -Tween (to ensure continuous blocking of non-specific binding sites), in between antibody binding incubation periods did not solve the problem of cross-reactivity.

The results obtained from immunoblots of SDS-PAGE separated bloodstream SDEs, indicated that non-specific binding was responsible for the cross reactivity observed between the positive HPLC derived GTP-binding fractions, and the anti-*ras* pan antibody. Based on these results, it was not possible to determine whether the cross-reactivity on the slot blot was due to the presence of *ras* like GTP-binding proteins or not. Further purification or characterisation of *ras* like GTP-binding proteins from *T.brucei* could therefore not be achieved by this approach.

FIGURE 3.7

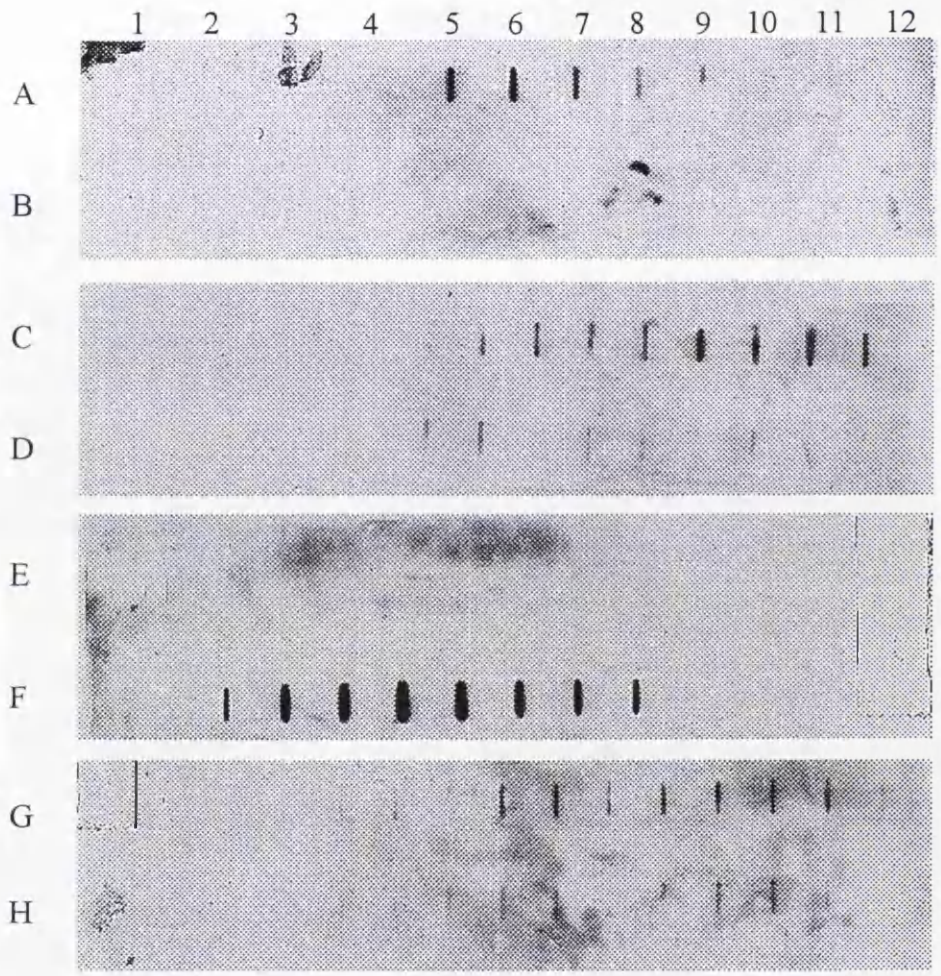


Figure 3.7 is an autoradiograph, showing the cross-reactivity of the anti- *ras* pan antibody with HPLC purified slot blot samples. 96 fractions from the step-wise NaCl gradient were loaded and immobilised on 5 different PVDF membranes (see section 3.22 for details). Each panel had two rows (A-H) with 12 wells, numbered 1-12. 200 µl of fractions 1-96 were loaded into the wells in rows A1 to H12.

FIGURE 3.8

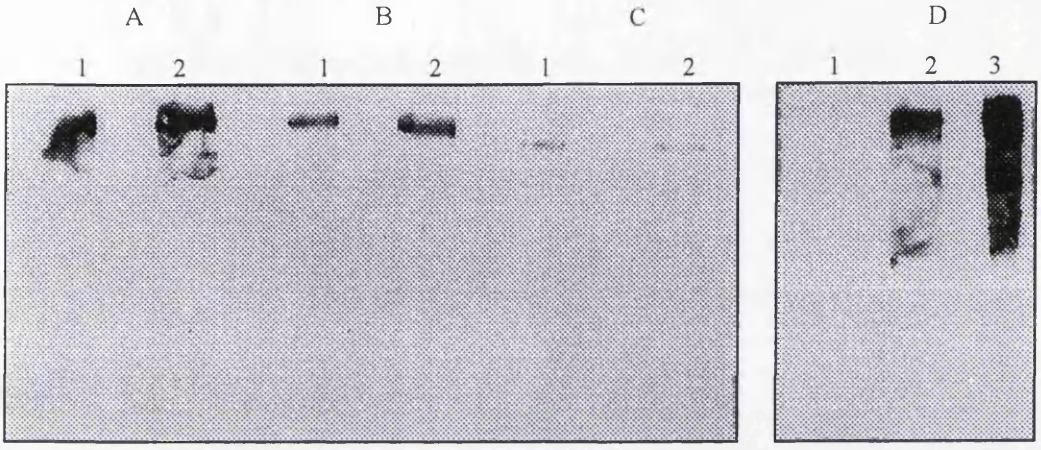


Figure 3.8 shows the results from the immunoblot analysis of bloodstream trypanosome spun detergent extracts (SDEs) using anti- *ras* pan as the first antibody. Panel A (second antibody dilution - 1:600) lane 1: bloodstream SDE incubated in a 1:400 dilution of preimmune serum; lane 2: bloodstream SDE incubated in a 1:400 dilution of anti-*ras* pan. Panel B (second antibody dilution - 1:4000) lane 1: bloodstream SDE incubated in a 1:400 dilution of preimmune serum; lane 2 bloodstream SDE incubated in a 1:400 dilution of anti-*ras* pan. Panel C (second antibody - 1:10 000) lane 1: bloodstream SDE, incubated in a 1:400 dilution of preimmune serum; lane 2, bloodstream SDE incubated in a 1:400 dilution of anti-*ras* pan. Panels D1, D2 and D3: bloodstream SDEs incubated in 1:10 000, 1:4000 and 1:600 dilutions of the second antibody respectively, without prior incubation in anti-*ras* pan.

3.4 Discussion

The presence of low molecular weight monomeric GTP-binding proteins in *T. brucei* was confirmed by means of the *in situ* GTP-binding assay of Coulter and Hide (1995). The proteins fall within a molecular weight range of 16 kDa - 35 kDa which is consistent with members of the *ras* GTPase superfamily (Kahn, Der and Bokoch, 1992). A number of polypeptides larger than 29 kDa were observed in some but not all gels. These higher molecular weight polypeptides of 32 and 35 kDa were observed in the bloodstream spun detergent extract supernatant. Even though the general guidelines on nomenclature state that *ras* -like GTP-binding proteins must fall within a molecular weight range of 20 - 29 kDa, allowance is made for proteins which may fall outwith this range but still possess other small GTP-binding properties, e.g. yeast *ras* which has a molecular weight of 40 kDa. Over thirty small molecular weight GTP binding proteins of different classes and similar molecular masses are thought to exist within the 20 - 30 kDa molecular weight range (Field *et al*, 1996). Results from the GTP binding assays on bloodstream and procyclic trypanosomes showed that some of these polypeptides are stage specific, with the following polypeptides being bloodstream stage specific, 18, 28, 32 and 35 kDa, whilst the 27 and 29 kDa polypeptides appeared to be procyclic stage specific. Polypeptides of the following sizes were competed out by the presence of unlabelled ATP: 55 kDa, 43 kDa, 17 kDa, and 19 kDa from bloodstream supernatant and pellet respectively and, 20, 18, 17 and 19 kDa from the procyclic supernatant and pellet respectively, indicating that not all of the observed polypeptides bind exclusively to GTP. Results from the experiment also showed that polypeptides within the range of 20-36 kDa, bind specifically to GTP.

A routine GTP-binding assay, capable of handling and concentrating large sample numbers, was designed to complement the *in situ* GTP-binding assay. This was shown to be suitable for identifying HPLC fractions positive for GTP-binding. Anion-exchange chromatography was used in an initial attempt to isolate and purify small GTP-binding proteins from *T. brucei* and a series of active fractions identified by the *in situ* GTP-binding assay were found to contain a 24 kDa polypeptide. Despite the presence of other polypeptides (visualised on Coomassie and silver stained SDS polyacrylamide gels), the 24 kDa polypeptide was not observed on silver stained SDS polyacrylamide gels. This result suggests that the 24 kDa polypeptide is not highly abundant in the *T. brucei* spun detergent extracts. In order to visualise the 24 kDa polypeptide on a silver stained gel, there would be the need to considerably scale up the amount of protein used for purification.

A number of studies, in which small GTP-binding proteins have been purified to near homogeneity from human placenta and bovine brain extracts (Evans *et al*, 1986 and Waldo *et al*,

1987), show that a reasonably high initial total protein concentration is required in order to ensure that the final concentrations of the eluted fractions of small GTP-binding proteins are reasonably high. Bovine brain contains 5 - 10 fold more *ras* GTP-binding protein then any other mammalian tissue (Yamashita *et al,* 1988) and therefore following the successive purification steps, detectable amounts of *ras* GTP-binding proteins were obtained from the HPLC eluted fractions. Kawata *et al*, 1988 and Yamashita *et al*, 1988, both initiated purification with 858 mg of crude membrane protein from bovine brain and purified amounts of 11 and 10 μg of GTP-binding proteins respectively.

The initial total protein content of the bloodstream spun detergent extract used in the purification steps was approximately 14 mg obtained from 1 ml of packed bloodstream trypanosomes. Therefore, based on the results of Kawata *et al,* approximately 60 ml of packed bloodstream trypanosomes would be required to produce extracts with a high enough initial protein content from which a detectable amount of small GTP-binding proteins could be purified. Furthermore, this could be a minimum estimate as further material would have been required to develop the purification protocol. Given this considerable requirement for trypanosome material and the limited resources available, it was decided not to continue with the purification approach. However, had this approach had been continued, slightly modified versions of the purification strategies used by Kawata *et al*, 1988 and Yamashita *et al*, 1988 would have increased the chances of isolating small GTP-binding proteins by HPLC purification. They both made use of sequential chromatography steps using concentrated fractions capable of binding [^{35}S] GTPγS. A number of HPLC resins are commercially available and therefore anion exchange chromatography which was used in the partial purification of small GTP-binding proteins from bloodstream spun detergent extracts would have been the first in a series of purification rounds. The fractions positive for GTP-binding would have been concentrated and then further purified by means of a size exclusion column. Bearing in mind that all fractions positive for GTP-binding would not necessarily be small GTP-binding proteins (i.e. they could be heterotrimeric GTP-binding proteins or EGF etc), the size exclusion column would serve not only as a second purification step, but also as a means of estimating the molecular size of GTP-binding proteins present in the starting material. Further purification by means of hydrophobic interaction chromatography would separate the membrane bound small GTP-binding proteins from the soluble ones (found in the cytosol). A fourth chromatography resin could be used to further purify the eluant. The eluant at this point, is likely to be highly enriched with small GTP-binding proteins, which could then be further characterised by means of immunoblot analysis using commercially prepared antibodies, or for the preparation of specific antibodies.

An antibody against a member of the *ras* subfamily, was used to determine the presence or absence of the subfamily in *T. brucei*. Preliminary results of Western blotted polypeptides, probed with different dilutions of the antibody, anti-*ras* pan, showed a high degree of cross reactivity. The observed cross-reactivity, therefore, made it difficult to interpret the results with any certainty. The cross reactivity may have been due to non-specific binding of the first antibody to other polypeptides or the second antibody to a component of the nitrocellulose membrane. This could have been due to the use of ineffective blocking buffers. Future work would be aimed at exploring different blocking buffers which may have been more effective e.g the use of BSA instead of horse serum. With cells known to have small GTP-binding proteins of the *ras* subfamily (such as bovine brain or human placenta) a method for immunoblotting could be developed and used to test the system. This system could then be used as a model for the development of a suitable assay to help in the interpretation of the results obtained from *T. brucei*.

Chapter 4

Cloning, sequencing and identification of *tbrlp*, a *ras* like small GTPase

CHAPTER 4

CLONING, SEQUENCING AND IDENTIFICATION OF *tbrlp*, A *RAS*-LIKE SMALL GTPase GENE

4.0 Introduction

The *ras* GTPase superfamily is made up of six subfamilies namely *arf, rab, ran, ras, rho* and *sar* (see introduction). The six subfamilies have four features in common: (1) The ability to bind and hydrolyse GTP; (2) Their genes encode proteins with a molecular size range of 20 - 29 kDa; (3) They all possess six highly conserved non-contiguous sequence motifs in the GTP-binding domain and (4) various subfamilies have conserved sequence motifs that are characteristic for members of the subfamily (Kahn, Der and Bokoch, 1992). In order to characterise the small GTP binding proteins of *T. brucei*, given the lack of feasibility of using a purification approach (see Chapter 3), a gene cloning approach was investigated. This approach was considered to be a good option for the isolation of a specific *ras* like protein, unlike the HPLC approach, in which the isolated low molecular weight GTP binding fractions could be contaminated with a number of proteins other than *ras* subfamily GTPases, e.g members of the *rab*, *rho*, *ran* and *arf* subfamilies, and low molecular weight heterotrimeric GTP binding protein α subunits. Three commonly used approaches are available for gene cloning, these are: the probing of genomic and/or cDNA libraries with a related/homologous gene from a different organism; the probing of expression libraries with heterologous antibodies, or, the use of PCR based methods for the amplification of genes from either genomic or cDNA using degenerate and/or gene specific primers.

As already stated, *ras* genes have been well conserved throughout evolution and are found in a range of organisms, making heterologous probing a possible method for isolating a related gene in *T. brucei*. The disadvantage with this method is that if the homology is very low, a *T. brucei ras* gene homologue may go undetected.

The screening of expression libraries with heterologous *ras* antibodies could also be used to clone a *ras* subfamily member from *T. brucei*. Problems could occur with this approach. Firstly, it was shown (Chapter 3) that heterologous *ras* antibodies are prone to extensive non-specific cross-reactivity. Secondly, success of this approach is dependent on the abundance of a suitable clone in the cDNA expression library. As *ras* proteins are typically not very abundant, suitable clones may be significantly under represented in such a library. Probing of genomic

expression libraries would be a more promising approach as the problem of the lack of abundance of the protein, would not be an issue.

The third approach is the amplification of DNA (either genomic or cDNA), by means of the polymerase chain reaction (PCR), using degenerate oligonucleotide primers to the conserved sequence motifs found in the GTP-binding domain. This method has been successfully used in the isolation of low molecular weight GTP binding proteins (*rab* and *ran*) from *T. brucei* previously (Field and Boothroyd, 1995; Field *et al*, 1995), and was, therefore, considered to be an appropriate approach for the isolation of a *ras* subfamily GTPase from *T. brucei*. Trypanosomes are evolutionarily distinct from higher eukaryotes and, therefore, their *ras* genes (if present) could have divergent sequences. The use of degenerate oligonucleotide primers and low annealing temperatures for the reaction, would result in the amplification of unrelated genes, but would also maximise the chances of isolating a *ras* subfamily member from this organism. Again, there is the possibility that the gene might not be highly abundant in the cDNA, but as PCR is good for the amplification of genes of low abundance such an approach may be a suitable method of isolating a *ras*-like gene from *T. brucei*. The problem of low abundance genes can also be overcome by the use of *T. brucei* genomic DNA as a template for the PCR amplification of a *ras* like gene.

A degenerate primer PCR approach was adopted, after considering all the methods available for the isolation of a member of the *ras* subfamily, for the following reasons: (1) Amplification by means of the polymerase chain reaction is a highly sensitive and powerful technique and, therefore, would be ideal for the amplification of genes which are expressed at low levels in cDNA; (2) The use of degenerate primers would increase the chances of isolating a gene of related sequence; (3) It is a rapid and relatively easy method which has been previously used successfully in the isolation of *ras* superfamily genes (Field *et al*, 1995; Field and Boothroyd, 1995); (4) It is a relatively more environmentally friendly process as it does not require the use of radiolabels.

4.1 Amplification of *T. brucei* cDNA using degenerate oligonucleotides

Genes that encode members of the *ras* GTPase superfamily have five well conserved sequence motifs which are extremely important for the proper folding and functioning of the proteins they encode. The conserved motifs offer good opportunities for designing degenerate oligonucleotides for PCR amplification of a *T. brucei ras* gene. These motifs are shown in Figure 4.0.

Using the highly conserved sequence motifs found in *ras* subfamily genes, four degenerate oligonucleotide primers were designed for use in the amplification of a *ras*-like gene from *T. brucei*, by the polymerase chain reaction. Degenerate primers and low annealing temperatures were used in the amplification, so as to enhance the chances of isolating a *ras* subfamily member. Most *ras* subfamily members have glycine residues at positions 12 and 13 of the PM 1 binding site (Valencia *et al*, 1991). The PM 1 binding site therefore has a general consensus of GxGGVGKS (where G is glycine, K is lysine, S is serine and V is valine). As the aim of this project was to isolate and study the role of a *ras* subfamily GTPase in trypanosomal growth regulation, a primer (Primer 1) incorporating glycine residues 12 and 13 (GGVGKSALT), was designed so as to enhance the chances of isolating a *ras* subfamily member. Primer 2, was also designed to specifically amplify a *ras* subfamily gene. All members of the *ras* subfamily have a general consensus sequence of YDPTIEDSY in this region (see Figures 4.0 and 4.1). The amino acid threonine (T) is highly conserved in all members of the *ras* GTPase superfamily, therefore in order to specifically amplify a *ras* subfamily gene from the *T. brucei* DNA template, it was essential to design a primer with the general consensus sequence of a *ras* subfamily member. Phenylalanine (F, see Figure 4.0), a single amino acid which is highly conserved in all members of the *ras* superfamily, could therefore, not be used in the specific amplification of *ras* subfamily genes. The two other primers, (i.e 3 and 4) were also designed to specifically amplify *ras* subfamily genes. The amino acid sequences from which the oligonucleotides were designed are shown in Figure 4.1 and the oligonucleotide primers and their degeneracies are summarised in Table 4.0.

TABLE 4.0

Summary of the oligonucleotide primer nucleotide and amino acid sequences and their degeneracies

OLIGONUCLEOTIDE	AMINO ACID SEQUENCE	DEGENERACY
1	GGVGKSALT	16 - FOLD
2	YDPTIEDSY	192 -FOLD
3	LDTAGQE	8 - FOLD
4	VGNKC/ADL*	128 - FOLD

* C/A - implies that the nucleotide sequence codes for either a cysteine or an alanine residue at that position.

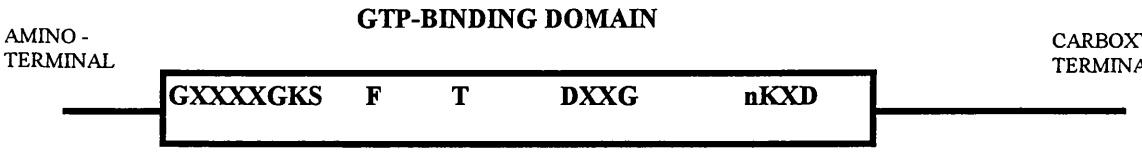
The nucleotide sequence for the oligonucleotide primers were as follows (listed as 5' to 3'; I - Inosine):

1. CCC AAG CTT GGI GGI GTI GGI AAA_G T_AC_GI GCI C_TTI AC.
2. CCC AAG CTT TAT_C GAT_C CCI ACI $ATT_{C/A}$ GAA_G GAT_C T_AC_GI TA.
3. GGC CTC GAG T_CTC T_CTG ICC IGC IGT A_GTC IA.
4. GGC CTC GAG AG_AA_G TCI $G_CC_AT_C$ TTA_G TTI CCI AC.

The letters in subscript are the alternative bases that could bind to the complementary genomic/ cDNA sequence and account for the different degrees of degeneracy displayed by the primers. The orientation of the 4 oligonucleotide primers are shown in Figure 4.1. 5' *Hind* III and 3' *Xho* I restriction sites were added to the ends of the oligonucleotide primers to facilitate the subsequent ligation and cloning steps.

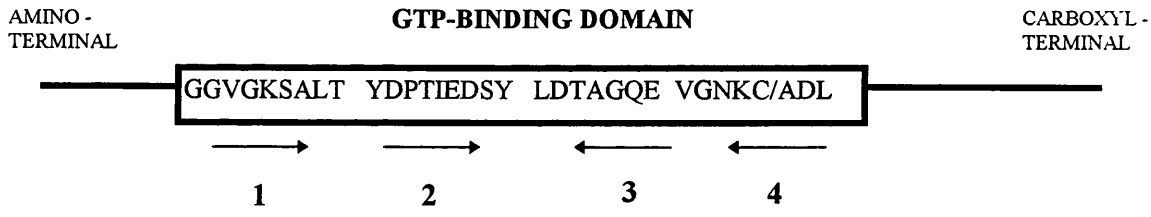
To maximise the chances of amplifying a *T. brucei ras* gene, four different combinations of the oligonucleotides were used in the amplification from *T. brucei* cDNA. These combinations are shown in Figure 4.2, together with the expected (approximate) molecular sizes of the amplified fragments, calculated from mammalian *ras* gene sequences (summarised in Table 4.1).

FIGURE 4.0



Schematic diagram of the conserved motifs of the GTP-binding domain of a mammalian *ras* gene. The diagram also shows the general consensus of the conserved non-contiguous regions of the gene. The single amino acid code is used, where D - aspartic acid, F - phenylalanine, G - glycine, K - lysine, n - asparagine (a small letter is used because it is conserved in most but not all *ras* genes), T- threonine (this represents the effector binding domain and was used in designing primer 2, see Figure 4.1). X - any amino acid.

FIGURE 4.1



Schematic diagram of the GTP-binding domain of a mammalian *ras* gene showing the four non-contiguous conserved sequence motifs/degenerate oligonucleotides (1-4) and their orientations, used in the amplification of *T. brucei* cDNA by the polymerase chain reaction.

FIGURE 4.2

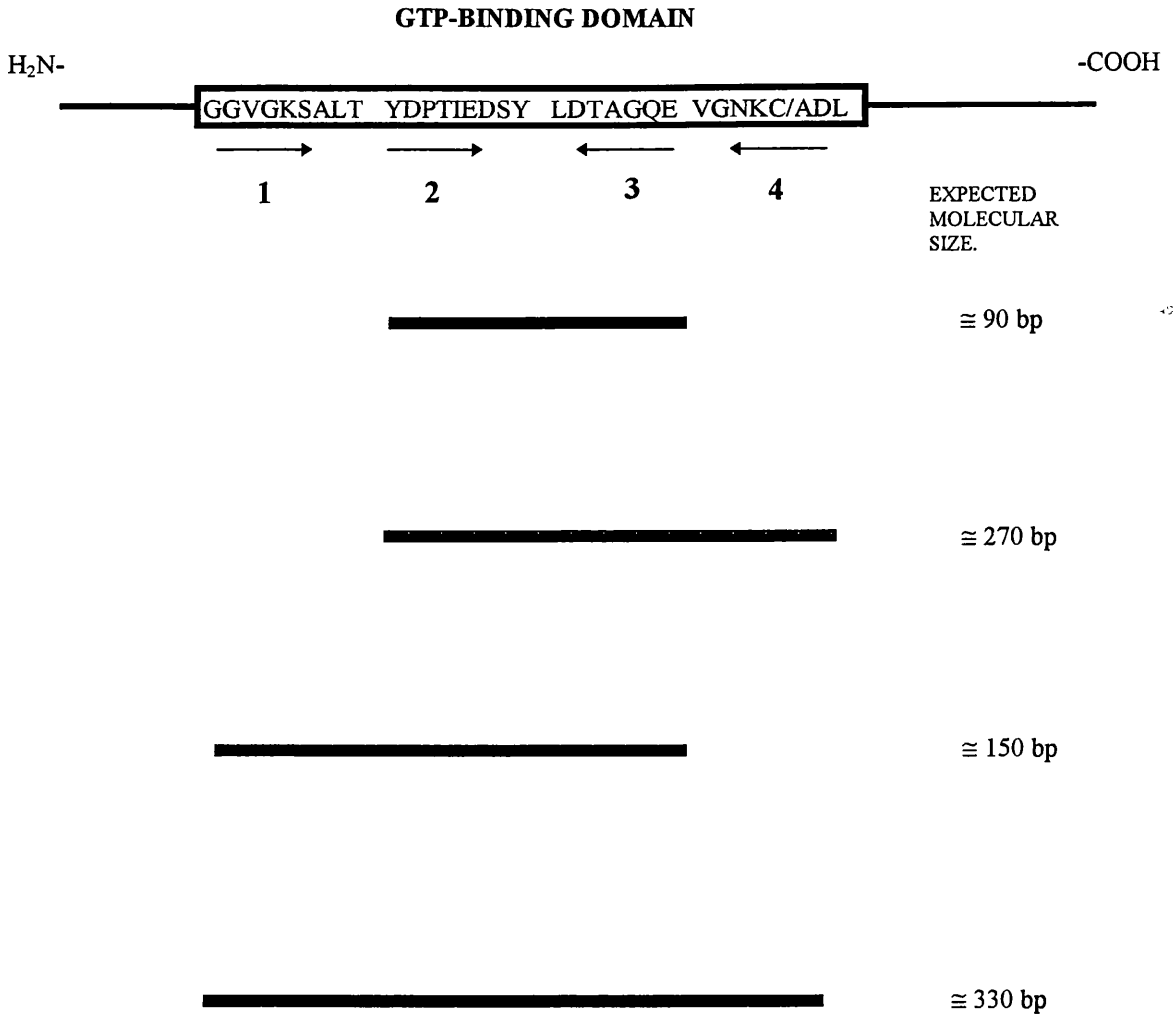


Figure 4.2 is an illustration of the oligonucleotide primers (1-4), together with the primer combinations used in the amplification of *T.brucei* cDNA and the expected PCR product sizes, based on conserved domains in mammalian H-*ras* genes. -H_2N - amino terminal, -COOH - carboxyl terminal

FIGURE 4.3

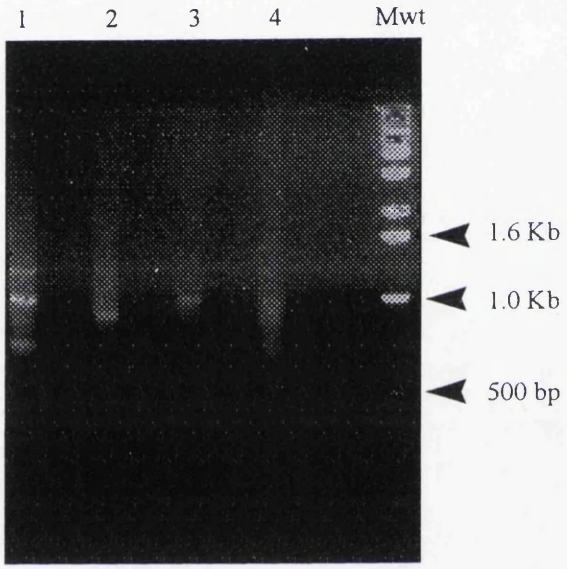


Figure 4.3 shows the results from the separation of four different PCR amplifications on an agarose gel. The PCR products (molecular sizes in the range of 500-1.6 Kb) were excised and gel purified. The products in lane 1 were amplified using degenerate oligonucleotide primers 1 and 4; in lane 2 degenerate oligonucleotide primers 2 and 4 were used in the amplification. Oligonucleotide primer combinations 1 & 3 and 2 & 3 were used in the amplification of products in lanes 3 and 4. Mwt - molecular weight marker (Gibco).

TABLE 4.1

Summary of the oligonucleotide combinations used in the amplification of T. brucei cDNA and the expected PCR product sizes

OLIGONUCLEOTIDE COMBINATION	EXPECTED PCR PRODUCT SIZE (APPROXIMATE SIZE)
1 + 3	150 bp
2 + 3	90 bp
1 + 4	330 bp
2 + 4	270 bp
1 + 2 + 3 + 4	150 bp, 90 bp, 330 bp, 270 bp

Using PCR conditions described in the Materials and Methods section, the degenerate oligonucleotides were used in the amplification of *T. brucei* bloodstream cDNA (prepared by the reverse transcription of poly (A+) RNA). Briefly, the PCR conditions were derived by varying the number of cycles (for the main run only) and annealing temperatures, to determine the optimum annealing temperature and number of cycles necessary for the amplification of PCR products with the minimum of smearing. These conditions were used and the resulting PCR products were separated on 1.5 % low melting point agarose gels by electrophoresis; the results are shown in Figure 4.3. Bands of the expected sizes could not be visualised i.e 150-300 bp (see Table 4.1), which was probably due to the low abundance of the gene in the cDNA used in the amplification. The visualised bands, from 500 bp - 1.6 Kb (see Figure 4.3) were excised, gel purified and digested using *Hind* III and *Xho* I. The digested PCR products were then ligated into pTZ18R pre-digested with the same enzymes, and transformed into XL1-blue *E. coli*, to generate a series of clones for sequencing analysis.

No transformants were recovered despite control ligations and transformations producing colonies. This could have been due to a number of reasons: 1). The loss of PCR products during gel purification; 2) The presence of inhibitors in the agarose used in the purification step, which could have been carried over into the ligation reactions and thus inhibited the ligations from taking place; or 3). The digests were not complete and therefore due to the absence of the necessary restriction sites the ligations could not occur. In order to improve the chances of cloning the PCR products different brands of low melting point agarose were tested. Longer

restriction digest periods and the use of different methods for the isolation/cleaning of the PCR products did not produce clones with inserts . The vector was therefore replaced by one which does not require gel purification or digestion of PCR products before ligation.

4.11 *Screening of clones derived from ligation into the TA cloning vector*

The TA cloning vector (Invitrogen) has a single 3' T - overhang which is required for the ligation of PCR products. The vector relies on the "A overhang" which is added to 3' ends of PCR products produced by *Taq* DNA polymerase. Ligations can, therefore, be carried out without a PCR product purification step, which might result in the loss of the product, or the introduction of inhibitors, which may be present in the agarose used in gel purification (section 4.1). The disadvantage of using the TA cloning vector is that no prior size selection of the PCR product is made, so that a range of products will be cloned. As degenerate primers and low annealing temperatures were used in the amplification step, many PCR products were generated, some of which could have been due to mispriming. The use of the TA cloning vector therefore led to the production of a large number of clones with different insert sizes which had to be screened for appropriate inserts (see Table 4.1). Using the TA cloning system approximately 150 recombinant clones were recovered.

More than 80 DNA minipreps, each from different clones, were digested with *Eco* RI and the products separated on 1.5% agarose gels. Table 4.2 is a summary of the oligonucleotide primer combinations used in the amplification reactions and the observed insert sizes of 40 clones derived after ligating aliquots of the PCR mixes into the TA cloning vector.

TABLE 4.2

Oligonucleotide primer combinations used in PCR mixes and the resulting insert sizes derived after ligation into the TA cloning vector

PCR PRIMERS	OBSERVED INSERT SIZES
1+3	200, 250, 344, 396, 400, 500, 600, 800, 900 bp, 2.0 Kb
1+4	200, 298, 344, 500, 600, 700 bp, 2.0, 3.0 Kb
2+3	200, 250, 344, 396, 400, 600, 900 bp, 2.0 Kb
2+4	298, 344, 396, 400, 500, 600 bp, 1.0, 1.2 Kb
1+2+3+4	200, 250, 344, 700, 900 bp, 2.0 Kb

Based on the knowledge available on the sizes of small molecular weight GTPases in *T. brucei* (see Chapter 3, section 3.1), and bearing in mind that *ras* genes in yeasts are known to have molecular sizes that fall outwith the range stipulated by the guidelines to nomenclature of low molecular weight GTPases (Kahn and Der, 1992), insert sizes above the expected (150-300 bp, see Table 4.1) could not be discarded without prior screening.

Initially several DNA minipreps were prepared and screened by manually sequencing the ends, this was a slow and laborious task. In order to speed up the process a new approach was taken. The TA cloning vector was used, as it contains a LacZα gene, which allows for blue-white selection when grown in the presence of X-gal and IPTG (white colonies represent clones with plasmids containing inserts). White colonies were picked, grown up overnight and digested to check insert sizes. Plasmids with insert sizes of up to 500 bp were separated by agarose gel electrophoresis, transferred to nylon membranes by Southern blotting and hybridised sequentially with ^{32}P end labelled degenerate oligonucleotides (as used in the amplification step). Clones positive for the correct combinations (see Table 4.1) were sequenced in order to determine whether or not they were *ras*-like. In summary, if after probing with all 4 labelled oligonucleotides, a clone was positive for oligonucleotides 3 and 4, but not 1 or 2 (see Figure 4.2) then, the clone would not be sequenced. The blots were stripped of radioactivity by boiling in 0.1% SDS for 3 x 15 mins before rehybridisation. Photographs showing the results derived from hybridising 37 clones with oligonucleotides 2 and 3 are shown in Figures 4.4a and b and summarised in Table 4.3.

TABLE 4.3

Summary of the results obtained from hybridisations with oligonucleotides 2 and 3

CLONE Nº	PCR PRIMER COMBINATION	SIGNAL INTENSITY AFTER 1 HR	
		OLIGO 2	OLIGO 3
1	1 + 4	–	–
2	1 + 4	–	–
3	1 + 4	–	–
4	1 + 4	–	–
5	1 + 4	–	–
6	1 + 4	–	–
7	1 + 4	–	–
8	1 + 4	–	–
9	1 + 4	–	–
10	1 + 4	–	–
11	1 + 4	–	–
12	1 + 4	–	–
13	1 + 4	–	–
14	2 + 3	+++	+++
15	2 + 3	+++	+++
16	2 + 3	–	+++
17	2 + 3	–	+++
18	2 + 3	–	+++
19	2 + 3	–	–
20	2 + 3	–	–
21	2 + 3	–	+++
22	2 + 3	–	+++
23	2 + 3	–	++
24	2 + 4	–	+++
25	2 + 4	+++	+
26	2 + 4	–	–
27	2 + 4	–	–
28	2 + 4	–	–
29	2 + 4	+++	–
30	2 + 4	–	–
31	2 + 4	–	–
32	2 + 4	–	+++
33	2 + 3	–	+++
34	1, 2, 3 & 4	–	++
35	1, 2, 3 & 4	–	–
36	1, 2, 3 & 4	–	–
37	1, 2, 3 & 4	–	+++

+, ++ & +++ - increasing signal intensity.

- no signal.

40 clones were found to be positive for appropriate combinations of oligonucleotide hybridisations and these were manually sequenced using the USB sequenase reaction (see Materials and Methods). Of the 40 clones only one had the DNA sequence of a *ras*-like GTP-binding protein (labelled clone number 411). One reason for the low recovery of appropriate clones, is probably due to the fact that degenerate oligonucleotides were used in the amplification process, together with a low annealing temperature. This may have greatly reduced the

specificity of the reaction and therefore the oligonucleotides probably led to the amplification of non *ras*-like genes by mispriming.

FIGURE 4.4A

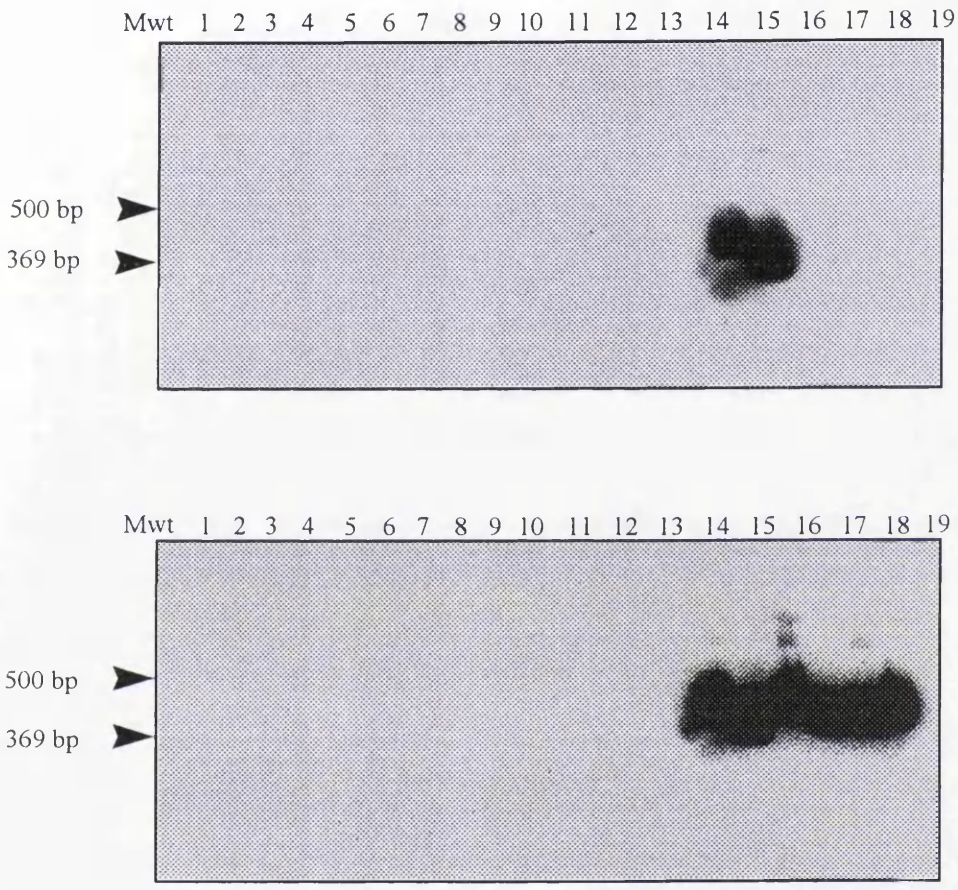


Figure 4.4A shows an autoradiograph of the results derived from hybridising 19 out of 37 *Eco* RI digested clones, using ^{32}P labelled oligonucleotide numbers 2 (for the top gel) and 3 (for the lower gel) (refer section 4.1, figure 4.1). Mwt - molecular weight markers.

FIGURE 4.4B

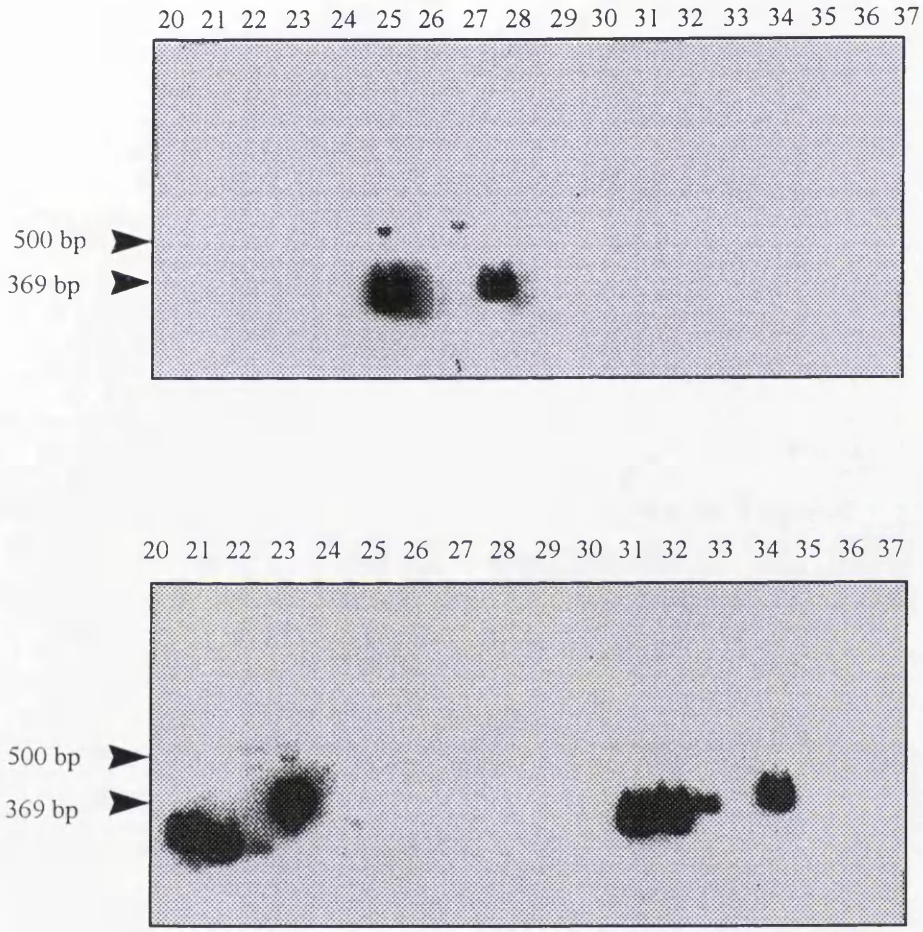


Figure 4.4B shows an autoradiograph with the results from the hybridisation of *Eco* RI digested clones 20 -37. Radiolabelled oligonucleotide numbers 2 and 3 (refer section 4.1, figure 4.1), were used in probing the samples (2 was used in probing the top gel and 3 the lower gel). Clone number 411 is equivalent to sample number 25. Mwt - molecular weight markers.

4.12 *Clone 411 - partial sequence of a ras-like GTPase gene*

Clone 411 was amplified from *T. brucei* cDNA with degenerate oligonucleotide primers 2 and 4 (see Figure 4.2 and number 25 in Figure 4.4 b). Sequencing showed that it has a 280 bp (or 93 amino acid) open reading frame (ORF). Fasta searches confirmed that 411 has 37 % and 39 % amino acid identity to the same region within mammalian N-*ras* and *Dictyostelium discoideum ras* b genes respectively. Further examination showed that 411 did not have a significant level of amino acid identity to any of the small GTP-binding proteins that had been previously cloned from *T. brucei* (El Sayed *et al* 1995 and Field *et al*, 1995; Swissprot and Genembl databases, 1996).

Based on the full length of mammalian H-*ras* gene, 411 appears to account for a third of the full length sequence. To characterise a full length copy of 411, it was necessary to determine the 5' and 3' flanking sequence of clone 411. There were two options 1) to determine the 5' and 3' flanking sequence of clone 411 by using methods designed for the amplification of unknown cDNA ends with gene specific primers, design gene specific primers for the 5' and 3' ends and then amplify up the full length gene from either genomic or cDNA; or 2) To isolate the full length gene by screening genomic or cDNA libraries with the partial *ras*-like GTPase gene, clone 411. The following sections discuss the theory of the methods used and the results achieved in the isolation of a full length gene.

4.2 Amplification of the 5' flanking sequence of 411 using the 5' random amplification of cDNA ends (RACE) method.

5' RACE involves the synthesis of cDNA from poly A+ or total messenger RNA and an antisense gene specific primer (5' GSP 1). Poly A+ RNA, was used in the isolation of the 5' - end of the gene containing clone 411, as it is less likely to be contaminated by genomic DNA than total trypanosome RNA. The synthesised cDNA was purified and a poly dT tail added using terminal transferase. The cDNA was then used as a template in a PCR reaction with a second gene specific primer, 5' GSP 2, located further towards the 3' end of 5' GSP 1 and an anchor primer complementary to the poly dT tail. A nested reaction using a third gene specific primer 5' GSP 3 and the anchor primer was then carried out to complete the process. The procedure should have produced a PCR fragment with the 5' sequence, which, even if not complete, could be used to design more gene specific primers, and used repeatedly until the full 5' end sequence was achieved.

The oligonucleotide sequences of the three gene specific primers and the anchor primer are shown below (listed 5' to 3'):

1. 5' GSP 1 - CGC GAC AAA GTT GCG TGT GAA TCG CC
2. 5' GSP 2 - GGA GAA GCT TTC AGC GTC TAT CAC GC
3. 5' GSP 3 - CAC CGA ATA CAT CCT GCC CGG AAG TG
4. Anchor primer - AGG CCA CGC GTC GAC TAG TAC GGG

Products of the 5' RACE method were analysed by gel electrophoresis and despite several attempts using a range of annealing temperatures, no products were obtained. As a result of this lack of success, another approach was taken to obtain the 5' end sequence of 411. A unique feature of trypanosomes is the presence of a spliced leader at the 5' ends of all mature mRNAs. A slightly modified version of the 5' RACE method was used in conjunction with a primer for the spliced leader in the second approach used to obtain the 5' end sequence of 411.

4.21 *Amplification of the 5' flanking sequence of 411 using the 5' RACE principle and a spliced leader primer*

Almost all mature messenger RNAs in *T. brucei* are attached to a spliced leader sequence (Agabian, 1990). Bearing this in mind, *T. brucei* cDNA was synthesised using the antisense oligonucleotide 5' GSP 1. A PCR amplification was then carried out using 5' GSP 2 and a 35-mer primer representing the spliced leader, with the previously synthesised cDNA as a template. A nested PCR reaction was then carried out using 5' GSP 3 and the spliced leader primer to obtain specific products for 411. A second set of reactions with a 1:10 dilution of the spliced leader primer was run alongside the first set, this was done in order to increase the specificity of the reaction by decreasing the chances of mispriming due to the presence of a low concentration of the spliced leader primer. PCR reactions were carried out using an annealing temperature range of 44 - 58° C to determine the optimum annealing temperature required for the amplification steps. Figure 4.5A and B are negative images of photographs of the different PCR products obtained after the second amplification step. Aliquots of the PCR products from each of the annealing temperatures were transformed into XL 1 - Blue *E. coli* after ligation in the TA cloning vector. Glycerol stocks of clones from reactions 46°C, 48°C and 50° C (products from the transformation of the ligation mixes of the products from the 1:10 dilution of the spliced leader /5' GSP 3 amplification reaction and the TA cloning vector) were made and stored at -70^0 C. Several colonies from the 46° C reaction (produced by the transformation of the ligation mix

of the product of the 1:10 dilution of the spliced leader primer and 5'GSP 3 reaction, into the TA cloning vector) were picked and grown up overnight in LB medium. DNA was prepared from the overnight culture according to the Wizard DNA miniprep protocol (Promega), digested with *Eco* RI and checked for the expected insert size i.e, approximately 200 bp, by separating the restriction digest products on 1.5 % agarose gels. All the plasmids had the expected insert size and four of these were randomly picked and sequenced. The sequence from clone 46-2 (hereafter referred to as clone 462) showed a region of overlap at the 3' end which corresponded to the 5' end of 411. Figure 4.6 is an illustration of the different gene specific primers used in the 5' RACE/spliced leader PCR amplification protocols, which resulted in the amplification of 462, together with the region of overlap between clone 411 and 462.

FIGURE 4.5

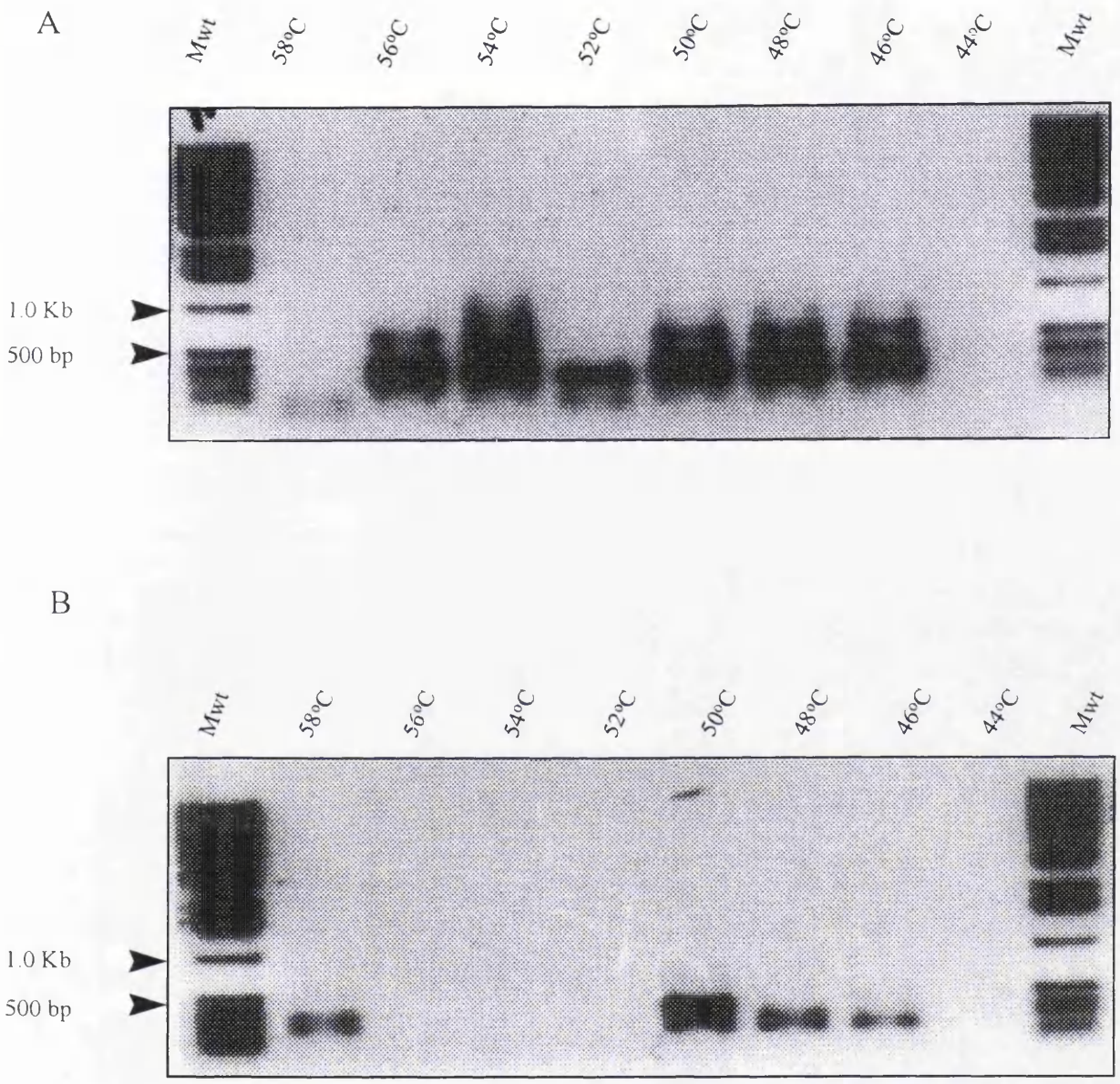
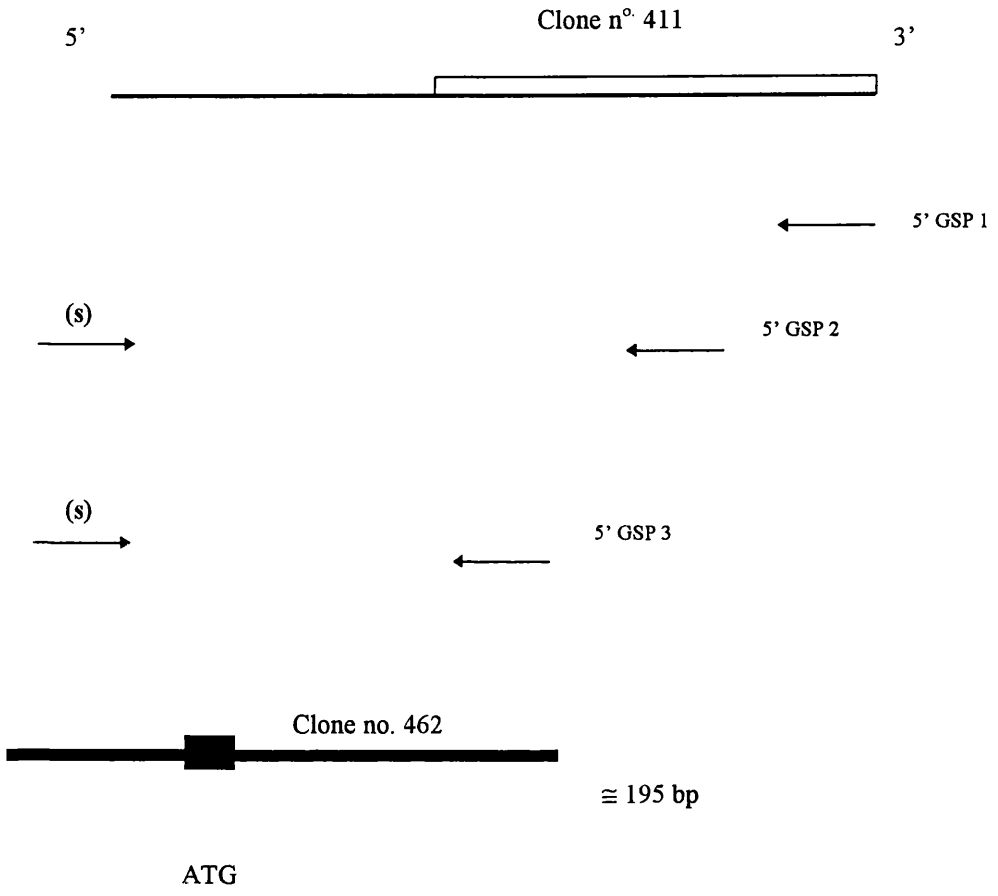


Figure 4.5 shows an ethidium bromide stained gel of two sets of PCR amplifications carried out at different annealing temperatures (44-58° C). Gels A and B show the different PCR products obtained after the nested reaction. The reactions in Gel A were carried out with neat spliced leader primer and the appropriate 5' gene specific primers (GSP). A 1/10 dilution of the spliced leader primer used in A, was used in the amplification of samples in B (see materials and methods for details).

FIGURE 4.6



An illustration of the different gene specific primers (5' GSPs 1, 2 & 3) and the spliced leader primer (s), used in the production of clone 462. 5' GSP 1 (an anti-sense oligonucleotide) was used in the synthesis of cDNA, which served as a template in the subsequent amplification steps with the spliced leader primer and 5' GSPs 2 and 3.

4.21a *Identification of T. brucei clone nº 462*

DNA sequencing was carried out to further characterise the insert of *T. brucei* DNA in clone 462. The sequence showed that clone 462 had an open reading frame of 195 bp and a start codon, ATG at the beginning of the sequence. A short untranslated region of 56 bp was observed between the spliced leader and ATG start codon. The 3' end of clone 462 has an overlap of 86 bp with the 5' end of 411. Having obtained the 5' end of the gene, it was necessary to recover the 3' end. A number of approaches are currently available for the determination of unknown 3' end sequences. The first one that was attempted was the 3' RACE method.

4.3 Amplification of the 3' end of clone number 411 by the 3' RACE method

The 3' RACE method relies on the presence of a poly dA tail at the 3' end of the mRNA and was used in the attempt to obtain the 3' end nucleotide sequence of the full cDNA containing the sequence of clone 411. cDNA was synthesised using total trypanosome RNA and the anchor primer (according to the manufacturers instructions). The sequence of the anchor primer was as follows: 5'CGG TGG CAG CAG CCA ACT TTT TTT TTT TT 3'. The cDNA was then used as a template for the subsequent rounds of PCR, together with a universal anchor primer with the following sequence: 5' CGA GAA TTC GGT GGC AGC AGC CAA CT 3' and the first 3' gene specific primer, 3' GSP 1. Two nested reactions were carried out using the PCR product from the first round as the template. Gene specific primers 3' GSP 2 and 3' GSP 3 were used with the universal primer in each of the nested reactions. The gene specific primer sequences were as follows (listed 5' to 3'):

1. 3' GSP 1 - ACT TCC GGG CAG GAT GTA TTC GGT GG
2. 3' GSP 2 - TCA TGG CGT TAT TCT CGT GTA CAG CG
3. 3' GSP 3 - GCG ATT CAC ACG CAA CTT TGT CGC GC

Several bands, none of which were the expected size i.e approximately 400 bp, were observed after the PCR products had been run out on 1.2 % agarose gels. After several months of manipulating the conditions required for the technique, this line of action was abandoned. A new approach was taken, which involved the screening of genomic and cDNA libraries with clones 411 and 462.

4.4 Screening of genomic and cDNA libraries using clones 411 and 462

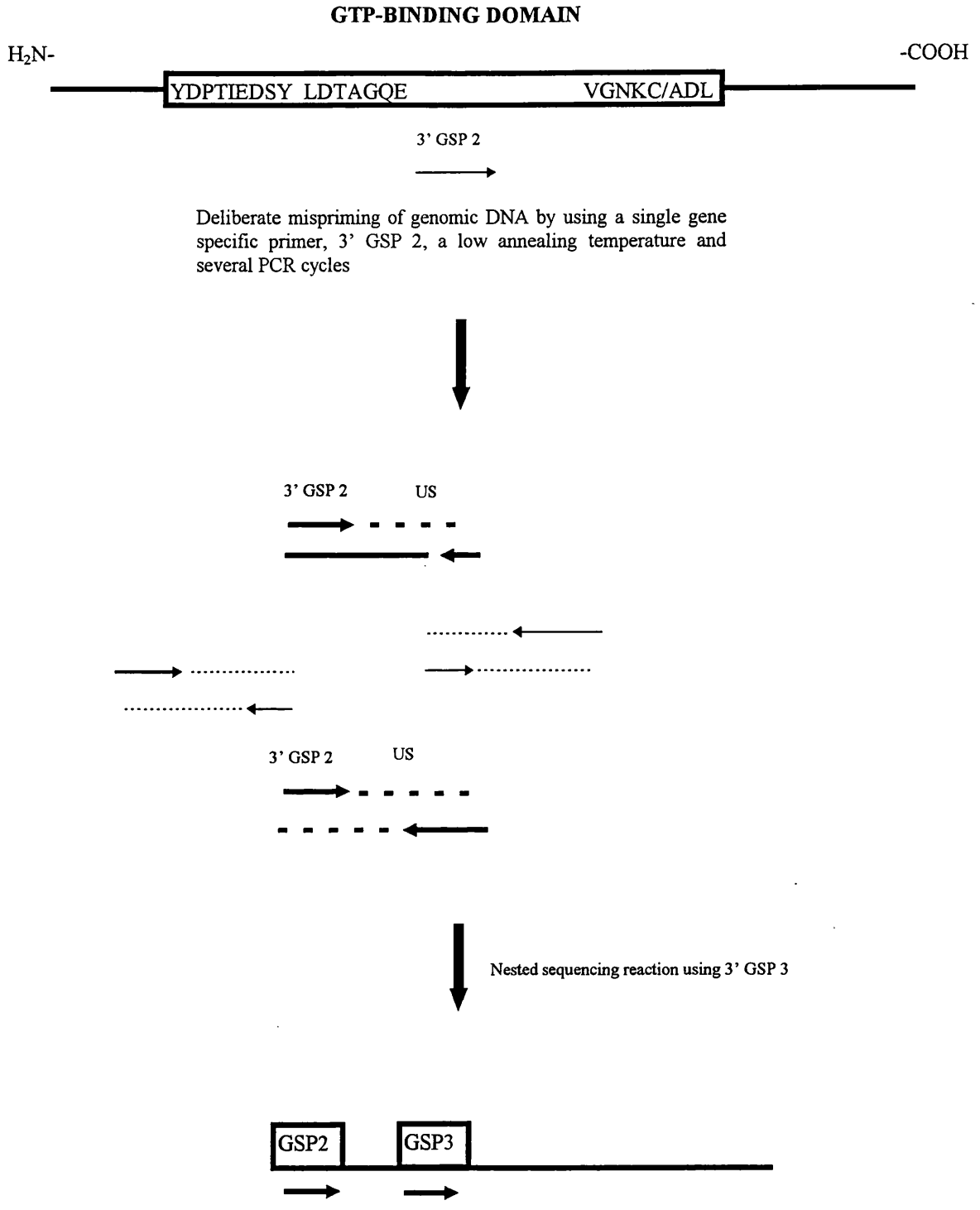
Inserts from the two clones (411 and 462) were digested using *Eco* RI, gel purified and ^{32}P labelled by random priming, using the Stratagene Prime-it kit. The EATRO 795 ILTat 1.2 genomic DNA library, a gift from Mr V.S Graham, had an average insert size of 18 Kb. The library was constructed from fragments of partial digests of *Sau* 3A ligated into *Bam* H1 digested arms of λDASH II (Stratagene). An EATRO 1125 procyclic cDNA library, a gift from Professor E. Pays, was also screened. Duplicate lifts were made and hybridised separately with each labelled insert. Approximately 20,000 plaques were screened in each case. Using this approach no recombinant phage hybridised to either of the probes. This could have been due to a number of reasons; 1) the low message level of 411/462 in the procyclic cDNA library; 2) the possibility that the expression of 411 and 462 is stage specific, i.e, bloodstream, not procyclic stage specific and 3) their absence in the genomic DNA library. The screening process was abandoned after 3 unsuccessful attempts were made. Two more PCR based methods were attempted concurrently in order to determine the 3' end sequence of 411, these were: 1) Short chromosomal walks using gene specific primers and 2) Long range PCR.

4.5 Amplification and sequencing of the 3' end of 411 by means of short chromosomal walks with gene specific primers

Chromosomal walking, a technique described by Screaton, Bangham and Bell, 1993, involves the deliberate mispriming of genomic DNA by using a single gene specific primer, a low annealing temperature and several PCR cycles. As the conditions for this technique encourage a decrease in the overall specificity of the amplification step, many PCR fragments are produced. The sequence for the gene is finally isolated by means of an increase in specificity which requires the use of a second gene specific primer as a sequencing primer. A single chromosomal walk can produce at least 200 base pairs of very good sequence. Based on the open reading frame (ORF) sizes of mammalian *ras* genes, it could be postulated with some degree of certainty, that the combined nucleotide sequences of clone number 462 and 411 represent approximately half of the full length ORF of the *T. brucei ras*-like GTPase. Therefore at least two chromosomal walks would be required in order to obtain the rest of the 3' end ORF nucleotide sequence, the first one would provide sequence data which would be used to design an additional gene specific primer. This primer would be used as a sequencing primer for the second chromosomal walk, which should cover the region up to the stop codon.

Two 5' → 3' facing primers from 411 (the same primers as those used in the 3' RACE technique) were used: 3' GSP 2 was used in the first mispriming PCR amplification using TREU 927 DNA as a template and 3' GSP 3 was used for the specific sequencing from the amplified products. Approximately 200 bp of good sequence were derived from the initial chromosomal walk, and a gene specific primer downstream of 3' GSP 3 was synthesised (3'untbrlp) based on this sequence (the 5' to 3' sequence of 3'untbrlp is as follows: CCG TGC GGT GAG TAG TGA AGA GGC AAG CAA). In a second amplification designed to extend the sequence further, 3' GSP 3 was used as the primer for the mispriming phase of the chromosomal walk (also using TREU 927 DNA), and 3' untbrlp was used as the specific sequencing primer. The rest of the gene sequence 3' of the clone 411 sequence was obtained by using this strategy. Figure 4.7 shows an illustration of the principles of chromosomal walking and gives details of the 3' gene specific primers 2 and 3. Two gene specific primers which corresponded to the beginning of the *ras* like gene, i.e 5' of 411 (5' end of clone 462), where the start codon (methionine) is situated, and the 3' end of 411 (sequence derived from the chromosomal walks), where the sequence terminates (indicated by the presence of a stop codon and a large untranslated region), were synthesised and used in an amplification reaction which involved the use of TREU 927 DNA as the PCR template and resulted in a product of approximately 700 bp. The product was ligated into pBluescript, and subsequently transformed into XL-1 Blue *E. coli*. Positive colonies (i.e white colonies) were picked and DNA was prepared from overnight cultures according to the Qiagen DNA miniprep protocol. The DNA was digested with *Eco* RI and *Xho* I (the same enzymes were used in the ligation reaction) and checked for plasmids with inserts of the expected size (i.e 700 bp). DNA from two of the colonies with inserts of the expected size was prepared (as previously described) and sequenced using the ABI automated Sequencer and two complementary pBluescript primers. The sequence from both strands of the insert showed that the *ras* like gene has an open reading frame of 681 bp, and the six highly conserved sequence motifs of a *ras* subfamily gene (see Chapter 4, section 4.7 for details).

FIGURE 4.7



An illustration of the principle of chromosomal walking as described by Screaton, Bangham and Bell, 1993. It also shows the gene specific primers, 3' GSP 2 and 3' GSP 3, used in the mispriming of *T. brucei* genomic DNA, the various PCR products including some with the unknown sequence. (US) and in the subsequent "nested sequencing" of the misprimed PCR product. NH_2 - amino terminus; -COOH - carboxyl terminus.

4.6 Amplification of the 3' end of 411 from *T. brucei* by means of long range PCR

As a complementary approach to obtaining the 3' sequence, long range PCR was employed to specifically amplify the 3' end of 411 from a genomic DNA library. This procedure was carried out concurrently with the chromosomal walk (see section 4.5). For a period of time, the ability to amplify large fragments of DNA of about 35 Kb, was limited by the low fidelity of *Taq* (*Thermus aquaticus*) DNA polymerase. Barnes, 1994 showed that fragments of up to 35 Kb in size could be accurately produced by using *Pfu* (*Pyrococcus furiosus*) DNA polymerase instead of *Taq*. *Pfu* is a thermostable enzyme with high fidelity and the additional ability to proof-read. It is found to decrease the rate of mutations per base from 10^{-4} to 10^{-5} per cycle. This technique was therefore used in the amplification of DNA from a *Trypanosoma brucei rhodesiense* (EATRO 795 ILTat 1.2) λDASH II, *Sau* 3A partial digest genomic DNA library, (average insert sizes of 18 Kb).

Four *T. b. rhodesiense* size selected libraries in λDASH II (Stratagene) namely, EATRO (EA) 795 ILTat 1.2, EA2340 Pop 339, EA2340 Gug 339 (all of which were *Sau* 3A partial genomic DNA libraries) and EA2340 Gug 339 (an *Eco*20*Eco* complete genomic DNA library, which is a library construct from a size selected subset of *Eco* RI restriction fragments cloned into Embl 4) with different average insert sizes were initially screened for the presence of 411 by means of PCR. 5' GSP 2 and 3' GSP 2 were used in the preliminary amplification step. The PCR products were separated on a 2% agarose gel and a band of the expected size, (80 bp) was observed in one of the libraries. Figure 4.8 is a photograph of the gel with the separated PCR products. Based on these results this *Trypanosoma brucei rhodesiense* strain EATRO 795 ILTat 1.2 (in λDASH II) genomic DNA library was used as a template in subsequent PCR reactions. A number of PCR reactions were set up using two λDASH II primers, T3 and T7. This initial reaction was used to determine the optimum conditions necessary for the amplification of a DNA fragment of approximately 18 Kb. A second set of PCR reactions was carried out, as the orientation of 411 within the library was not known, using the following PCR primer combinations:

1. 3' GSP 2 + T7
2. 3' GSP 2 + T3

A nested reaction was carried out after the preliminary PCR amplification. 1 µl of the PCR product was used as a template and a primer downstream of 3' GSP 2 was used in the second amplification step. The combinations used were as follows:

3. 3' GSP 3 + T7 and

4. 3' GSP 3 + T3.

Combination 3 was used in the amplification of the PCR product from 1 and combination 4 in the amplification of the PCR product from 2.

After determining the orientation of 411 within the library, the experiment was repeated, using vector primers as positive controls. The results from the long range amplification of *T. b. rhodesiense* are shown in Figure 4.9. Two PCR products were observed one of approximately 7 Kb and the other 2 Kb (products from the use of 3' GSP 2 + T3 and 3' GSP 3 + T3), no bands were visualised in the lane with the negative control (no DNA). Direct ligation of the PCR products using the TA cloning vector system produced only one clone with an insert size of 2 Kb. The 2 Kb insert contained the 3' end sequence of the *ras*-like gene from *T. brucei* and, other DNA sequence from the *T. brucei* genome 3' of the *ras*-like gene and for which no sequence data was obtained. The sequence of this clone was determined and found to have a region of sequence overlap which was almost identical to that derived from the chromosomal walks, thus confirming the 3' end sequence. The observed difference which is a conservative amino acid change (Serine to Threonine) appears to have been due to the use of different strains of *T. brucei*. A number of sequencing errors, one of which caused a visible frame shift, were not resolved as the full length *ras* gene was amplified from TREU 927 DNA, using primers from the 5' end of 462 and the 3' end of the long range PCR/chromosomal walk sequence. Figure 4.10 is an alignment of the 5' and 3' flanking sequence of 411 derived from 1) Clone 462, which was amplified from *T. brucei* cDNA with a gene specific primer from 411 and a primer for the spliced leader; 2) A 2.0 Kb long range PCR product, which was amplified from a *T. b. rhodesiense Sau* 3A partial genomic DNA library with gene specific primers from 411 and T7 (vector primer); 3) two chromosomal walks using gene specific primers from 411 for the "mispriming reaction" and for the sequencing reaction (the second sequencing primer was derived from the sequence of the first chromosomal walk); 4) the full length sequence of the *ras*-like gene in two different cloning vectors amplified from *T. brucei* strain TREU 927 genomic DNA, with primers from the 5' end of 462, and the 3' end of the long range PCR product/chromosomal walk sequence. Figure 4.10 also shows the regions of overlap of 462 and 411; 411, the long range PCR product and the sequence from the two chromosomal walks.

FIGURE 4.8

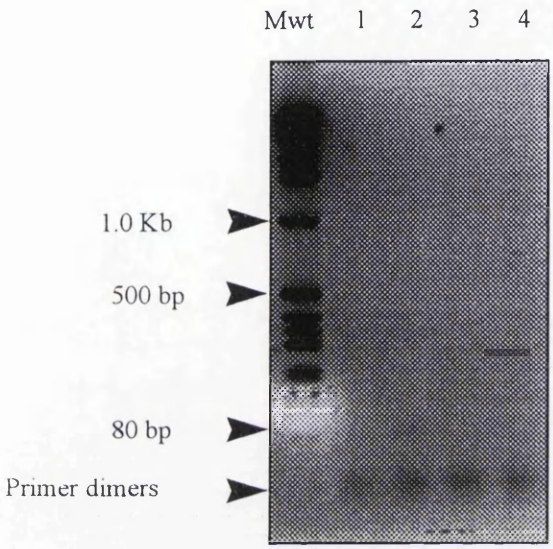


Figure 4.8 shows an ethidium bromide stained gel of PCR products separated on a 2 % agarose gel. The following templates were used in the amplification reactions, in lane 1, *T. b. r.* EATRO 2340 Pop 339; lane 2, *T. b. r.* EATRO 795 ILTat 1.2; lane 3, *T. b. r.* EATRO 2340 Gug 339; and in lane 4, *T. b. r.* EATRO Gug 340 (see Materials and Methods for details). A band of the expected size, 80 bp can be visualized in lane 2, between the 500 bp marker and the primer dimers. No bands were observed in the other lanes. Mwt - molecular weight marker.

FIGURE 4.9

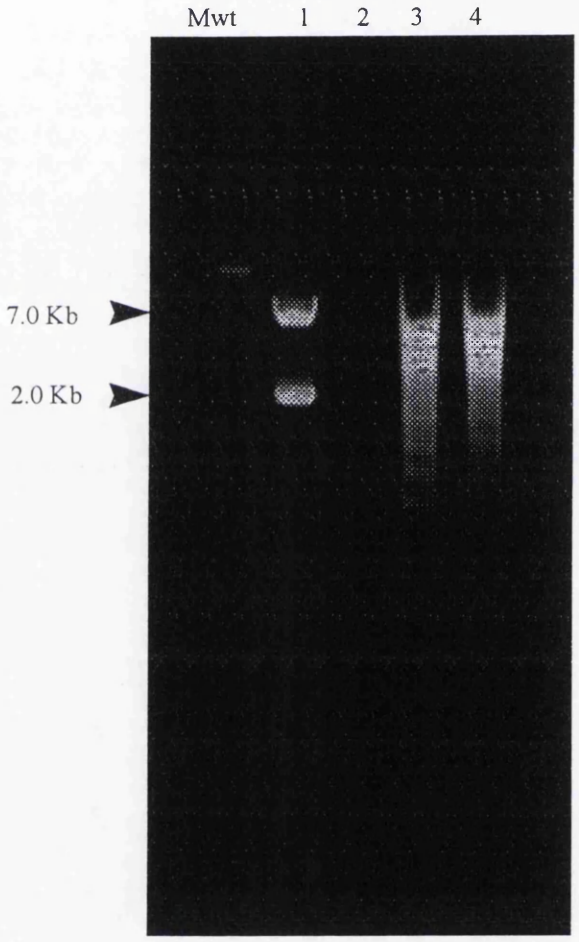


Figure 4.9 shows the PCR products from the long range PCR amplification separated on a 1.0 % agarose gel in lane 1; There were no bands in lane 2 which was the negative control (no DNA). Primers for the vector were used in the amplification of PCR products in lanes 4 and 5. Mwt - molecular weight marker.

FIGURE 4.10

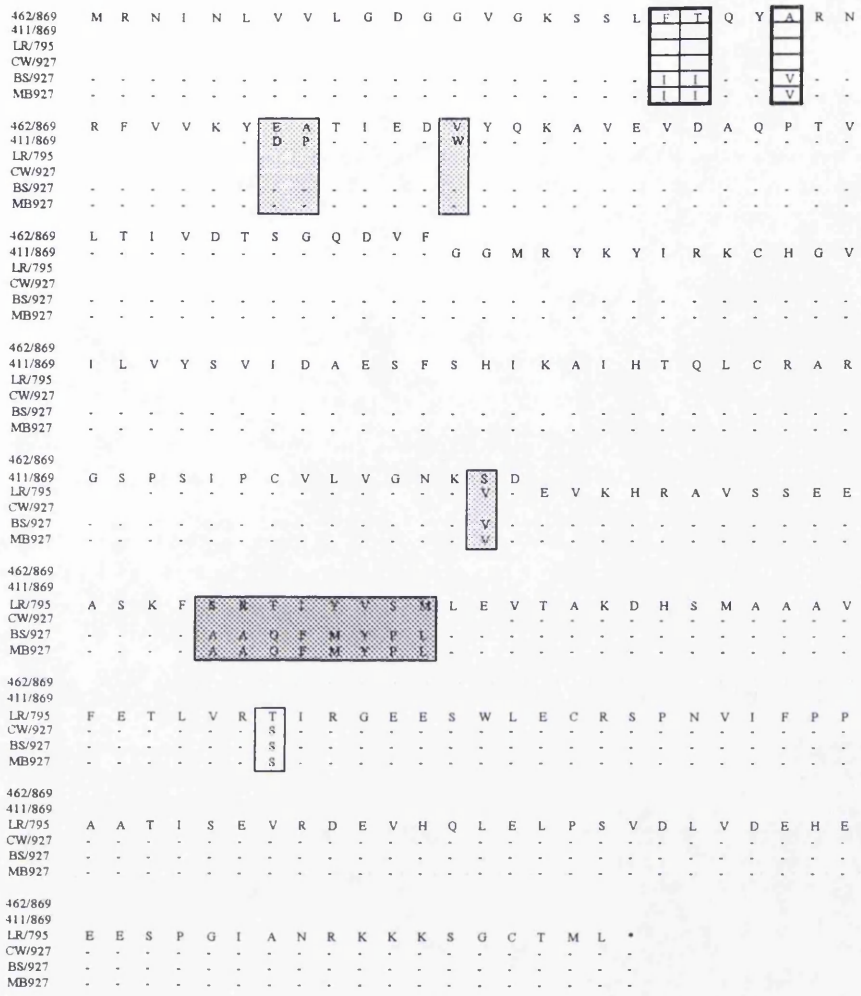


Figure 4.10 shows an alignment of the 5' and 3' flanking sequence of clone 411. The blank spaces indicate regions where sequence data are unavailable, the dashes represent identical amino acids (compared to the topmost sequence in each set of sequences) and * represents the stop codon. The bordered areas (with grids) represent changes due to sequencing errors in the 462/869 sequence. The bordered, lightly shaded regions highlight the sequence differences between the different PCR products generated by inclusion of the degenerate primers in the 411/869 sequence. The bordered, dark grey regions represent amino acid differences which were generated by a sequencing error causing a short frame shift in the LR/795 sequence. The bordered unshaded region represents a conservative Serine to Threonine amino acid change which was attributed to strain difference. Clone 462 was amplified from *T. brucei* strain TREU 869 cDNA with a gene specific primer from clone 411 and a primer for the spliced leader. Clone 411 was also amplified from *T. brucei* strain TREU 869 cDNA, with degenerate *ras* primers. LR/795, was the product derived from the long range PCR amplification of *T. brucei. rhodesiense* EATRO 795 IL Tat 1.2 *Sau* 3A partial genomic library. CW/927 (sequence derived from two chromosomal walks), BS/927 and MB/927, full length *tbrlp* subcloned into pBluescript (BS) and the maltose binding protein vector pMAL-c2 (MB) were all amplified from *T. brucei* strain TREU 927.

Using the sequences determined for the 5' and 3' ends of the gene, two PCR primers, 5' tbrlp and 3' utbrlp (the full length *ras*-like gene was named *tbrlp* i.e *Trypanosoma brucei ras*-like GTP binding protein, hence the use of the names 5' tbrlp and 3' "ultimate" tbrlp), were designed to amplify the full length gene from *T. brucei* strain TREU 927 genomic DNA. This PCR product was cloned into the Maltose binding protein and pBluescript cloning vectors and sequenced. The sequence for the full length gene, *tbrlp*, had a high percentage of amino acid identity to that of the combination of the fragments i.e clones 462, 411 and the sequence derived from the chromosomal walk and long range PCR. The gene was amplified and resequenced from a different source of TREU 927 *T. brucei* genomic DNA; the sequence from this stock was identical to the sequence from the full length gene (refer Figure 4.10). This was done so as to ensure that the full *ras*-like gene was not a contaminant peculiar to one source of TREU 927. Figure 4.11 is a photograph showing the presence of the amplified gene in three different aliquots of TREU 927 (from different sources) and a different strain of *T. brucei* genomic DNA i.e. STIB 247. This was carried out so as to determine whether the *ras* like gene could be amplified from other strains of *T. brucei*. The sequences of the primers used in the amplification of the full length gene were as follows (listed 5' to 3'):

1. 5' tbrlp - ATG AGG AAT ATT AAC CTC GTC GTT TTG GGT
2 3' utbrlp - TCA GAG CAT AGT GCA TCC CGA TTT CTT TTT CCT A

Figure 4.12 is a summary of the different fragments of the *ras* gene from *T. brucei,* the primers and the techniques used in the isolation of the full length open reading frame of the *ras*-like gene.

FIGURE 4.11

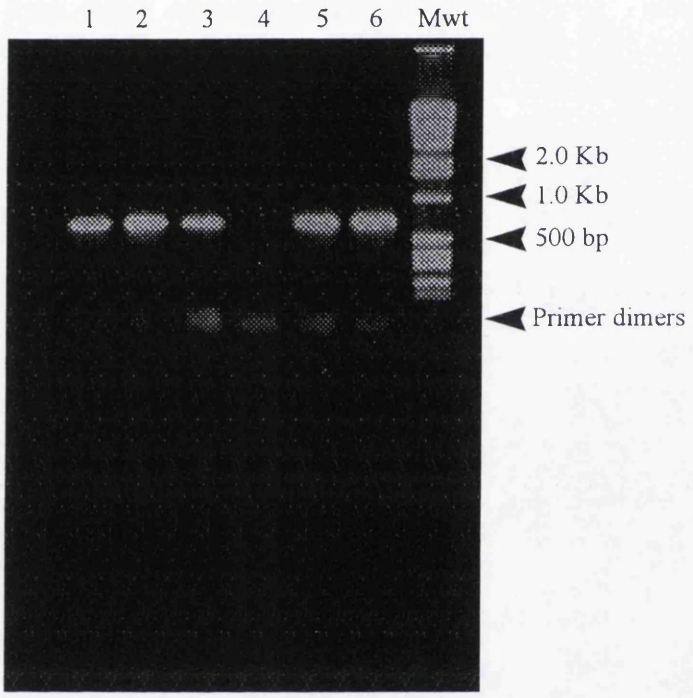


Figure 4.11 shows the presence of *tbrlp* in three different aliquots of 927 (lanes 1-3), the negative control (no DNA) was in lane 4, and two different aliquots of 247 (lanes 5-6). The bands visualized in the lower region of the gel are as a result of dimerization of the primers.

FIGURE 4.12

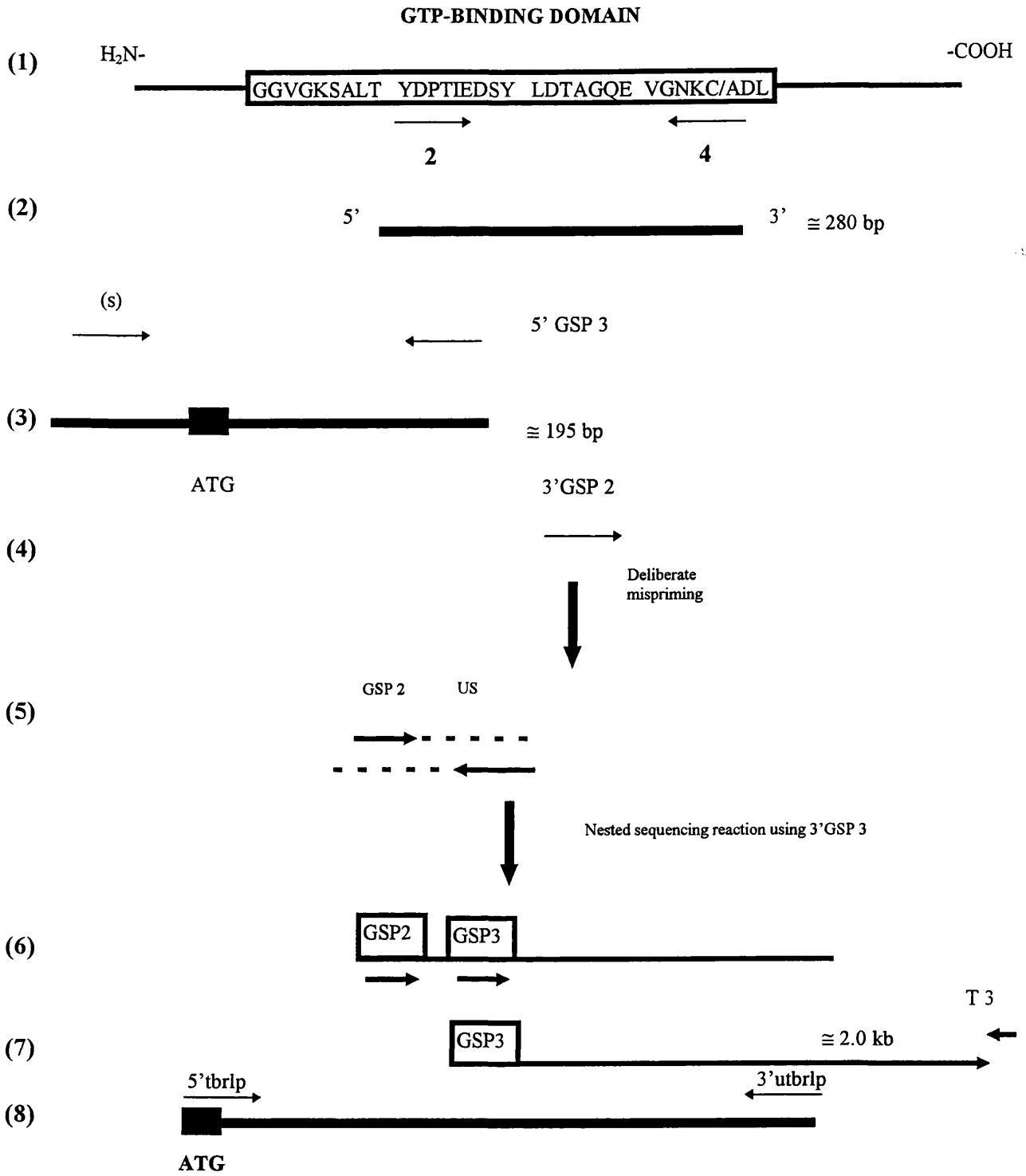


Figure 4.12 (not drawn to scale) shows a summary of the primers and techniques used in the amplification of *tbrlp*. (1) shows the degenerate primers used in the initial amplification of clone 411 (280 bp) from *T. brucei* strain TREU 869 cDNA (2). (3) represents the amplification reaction in which a primer for the spliced leader (s) and a gene specific primer from clone 411 (5' GSP 3) were used in the production of a PCR fragment (clone 462, approximately 195 bp), which contained the 5' end sequence of 411. (4) - (6) represent the steps taken to produce sequence 3' of 411 by means of chromosomal walking with gene specific primers (GSPs) from 411 using TREU 927 DNA as a template and long range PCR (7), using a gene specific primer from 411 (GSP 3) and a λDASH vector primer, T 3. (8) represents full length *tbrlp*, amplified from *T. brucei* strain TREU 927, with gene specific primers derived from the 5' end of clone 462 (5' tbrlp) and the 3' end of the long range PCR product/the chromosomal walk sequence (3' utbrlp). ATG - methionine (start codon).

4.7 *Trypanosoma brucei ras*-like GTP-binding protein

Based on the general structural features of mammalian H-*ras* (refer Figure 4.0) the *ras* like gene from *T. brucei* appears to be a member of the *ras* GTPase superfamily. It has an open reading frame of 681 bp or 227 amino acids and an estimated protein size of 25 kDa which classifies it as a low molecular weight GTP-binding protein. The *ras*-like GTP-binding protein was named *Trypanosoma brucei ras* like GTP binding protein or *tbrlp*. Figure 4.13 shows the full length amino acid sequence of *tbrlp*, in alignment with 2 *ras* and 3 *rap* genes from three higher eukaryotes in which the general consensus of the conserved non-contiguous regions are highlighted.

In addition, *tbrlp* has 45 % amino acid sequence identity to human *rap* 2 genes and 41 % amino acid sequence identity to human H-*ras* genes. Several structural features within the GTP-binding domain, define *tbrlp* as a *ras*-like gene and are discussed below.

4.71 *The phosphate/magnesium ion binding sites*

All members of the *ras* GTPase superfamily have three highly conserved phosphate/magnesium ion binding sites within the GTP-binding domain. Each site has well conserved amino acid residues which are essential for the proper folding and function of the encoded proteins.

The first phosphate/magnesium ion binding site (PM 1), also referred to as the P-loop, has a consensus sequence of GxxxxGK S/T (S/T - either an S or a T is found in that position). The PM 1 site of *tbrlp* has a sequence that corresponds with the consensus sequence, GDGGVGKS. This site is involved in the maintenance of the GDP conformation of the protein, its inactive state. Lysine (K), is involved in binding tightly to the β - phosphate of GDP, while the serine residue binds to the cofactor in this case a magnesium ion. This binding site is therefore extremely important in enhancing the stability of the GDP or inactive conformation of the protein. Almost all members of the *ras* subfamily have Gly-Gly in positions 12 and 13 (based on positions in $p21^{ras}$, Valencia *et al*, 1991). *tbrlp*, has the same sequence arrangement and therefore can be classified as a *ras* subfamily member (positions 12 and 13 underlined).

Phosphate/magnesium ion binding sites 2 (PM 2) and 3 (PM 3), bind the γ - phosphate of GTP and the co-factor. They are necessary for the maintenance of the stability of the GTP bound conformation of the protein. PM 2 is made up of only one conserved amino acid residue,

threonine (T), this residue is also conserved in *tbrlp*. PM 3 however has a consensus sequence of DxxG Q/T E (where Q/T means either a Q or a T at that position). Members of the *ras* subgroup have a glutamine (Q) at that position whereas those belonging to the *rap* subgroup have a threonine (T). These residues are found to be involved in the GTPase activity of the proteins. All of the conserved amino acids are found within the sequence of *tbrlp*. The sequence that corresponds to PM 3 in *tbrlp* is DTSG_Q D, which suggests, that *tbrlp* belongs to the *ras* subgroup (see Figure 4.13).

Figure 4.13

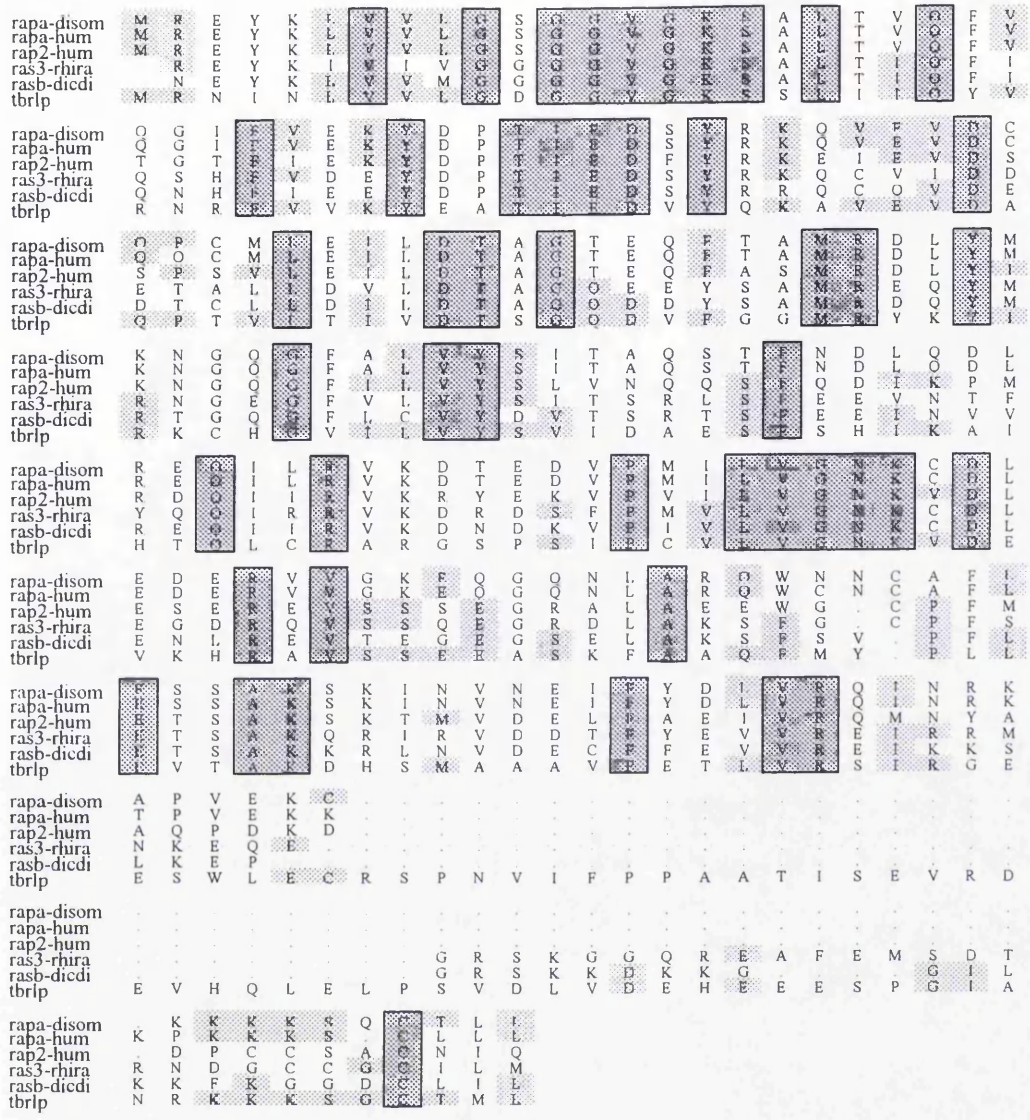


Figure 4.13 shows an alignment of the full length sequences of *tbrlp* and *ras* (or *rap*) genes from Dicdi - *Dictyostelium discoideum*, disom - *Discopyge ommata*, hum - *Homo sapiens* and rhira - *Rhizomucor racemosus*. The highlighted and bordered areas are the conserved sequence motifs found in *tbrlp* and both *ras* and *rap* genes. Sequence motifs found in either ras or *rap* genes and *tbrlp* are highlighted but not bordered.

4.72 *The guanine nucleotide binding sites*

There are three guanine nucleotide binding sites within the GTP-binding domains of members of the *ras* GTPase superfamily, G1, G2 and G3. The conserved sequence motifs within these domains are responsible for the maintenance and stability of the bonds between the protein and the guanine base of the nucleotide. Only one amino acid residue is conserved in the first guanine nucleotide binding site or G1. The conserved amino acid is phenylalanine (F) and this residue is found in the same position in *tbrlp*. The conserved residues within G2 are as follows: nKxD - where asparagine (n) is conserved in most (not all) *ras* genes and is therefore denoted by a lower case letter, while lysine (K) and aspartic acid (D) are highly conserved and found in all *ras* genes and x is any amino acid (see Figure 4.12). The G2 site of *tbrlp* has the following sequence which contains the two important conserved amino acid residues, lysine and aspartic acid - NKXD. G3 is not directly involved in the binding of the base to the protein, but appears to be involved in the maintenance of the stability of the bonds formed by residues in G1 and G2 and the protein. The consensus sequence in this region is ExSA, where E is glutamic acid, x - any amino acid, S - serine and A is alanine. The sequence of *tbrlp* in this region of the GTP-binding domain is as follows: EVTA. The presence of a threonine residue where the serine residue should be is a conservative amino acid change which may have occurred during the evolution of the gene. Based on the conserved sequences motifs in the GTP-binding domain alone, *tbrlp* can be classified as a member of the *ras* GTPase superfamily and the *ras* GTPase subgroup.

4.73 *C - terminal extension region*

This region of the gene, in members of the *ras* GTPase superfamily, is the variable region. There can be as many as 180 or as few as 20 amino acid residues in this region in different *ras* family members and a very low level of conservation is observed. In *tbrlp* this region is 79 amino acid residues long and is found to be completely unrelated to the C - terminal regions of other *ras*-like genes. However, a consensus sequence for a cyclic AMP dependent phosphorylation site is found just before the C-terminal CAAX box. The consensus sequence is BBxS/T, where B is a basic amino acid, x is any amino acid, and either a serine (S) or a threonine (T) is found in position 4 (Lapetina *et al* 1993). This motif is found to be conserved in human *rap*1b proteins (Bokoch, 1993). The sequence found at the C-terminal end of *tbrlp* is KKKS. The presence of this motif suggests that *tbrlp* is a member of the *rap* subgroup.

4.74 *The carboxyl terminal CAAX box*

The CAAX box (where C - cysteine, A - aliphatic residue and X - any amino acid), is found at the carboxyl terminal of all *ras* subfamily genes and is the site for post translational modification, essential for the subcellular localisation of the protein. Generally, the presence of a methionine or serine at the X position results in the farnesylation of the protein. This is normally found in members of the *ras* subgroup. The presence of leucine at this same position suggests that the protein undergoes geranylgeranylation and has been observed in members of the *rap* subfamily. The difference in the post translational modification processes account for the different subcellular locations of *ras* and *rap* proteins. *Ras* GTPases are normally found on the inner surface of the plasma membrane (Lowy and Willumsen, 1991), whereas *rap* genes are found on the endoplasmic reticulum and the endosomal/lysosomal compartments (Zerial and Huber, 1995).

The CAAX box of *tbrlp* has the following sequence - CTML. The fact that the residue at the X position is leucine suggests that the protein undergoes geranylgeranylation in the post translational modification process and should, therefore, be a member of the *rap* and not *ras* subgroup.

It appears as though *tbrlp* has amino acid sequence characteristics of members of both *ras* and *rap* subgroup members (see section 4.7, Figure 4.13). Due to these mixed features, it is difficult to determine the actual subgroup to which *tbrlp* belongs.

4.8 Characterisation of *tbrlp*

In order to determine the copy number of *tbrlp*, single and double restriction digests were carried out on *T. brucei* strain STIB 247 genomic DNA with the following enzymes: *Hind* III, *Xho* I, *Bam* HI, *Eco* RI and *Pst* I. The digests were separated on a 0.8 % agarose gel and transferred onto a nylon membrane by Southern blotting. Clone 411 representing the inner third portion of *tbrlp* was labelled by random priming using the "Prime-it" random labelling kit (Stratagene) and hybridised to the blot at 65° C. Results from the hybridisation of the Southern blot suggested that *tbrlp* is a single copy gene (Figure 4.14).

In order to confirm that *tbrlp* is a single copy gene and to determine whether or not there are any other genes to which it is closely related, *Hind* III digests of *T. brucei* strain TREU 927 genomic DNA were separated in triplicate on a 0.8 % agarose gel and transferred onto a nylon membrane. The membrane was cut into three panels (after the transfer) each of which had a lane of *Hind* III digested genomic DNA as shown in Figure 4.15. The full length *tbrlp* was radiolabelled and used as a probe on the three membranes which were then hybridised at 65°, 55° and 50° C . After overnight incubation the three membranes were washed 3 times for 15 minutes, with a low stringency wash buffer (2x SSC + 0.1 % SDS) at each of the hybridisation temperatures, i.e 65°, 55° and 50° C. The results, (Figure 4.16), indicate that no related genes are detected at low stringency. Four bands were detected on each of the membranes, these were: 1.3 Kb, 12 Kb, and two other bands which had molecular sizes greater than 12 Kb. The two topmost bands are consistent with the residue of the digested DNA in the wells and partially digested DNA respectively. Another Southern blot was carried out using single and double restriction enzyme digests of *T. brucei* strains TREU 927 and STIB 386. The DNA fragments were separated on a 0.8 % agarose gel after digestion with the following restriction enzymes, *Hind* III, *Pst* I and *Hind* III + *Pst* I, transferred onto a nylon membrane, and probed with labelled full length *tbrlp*. The pattern for the *Hind* III digests was consistent with the previous experiment, whilst that of *Pst* I, highlighted the presence of a *Pst* I polymorphism in the three different strains, upstream of *tbrlp* I.

FIGURE 4.14

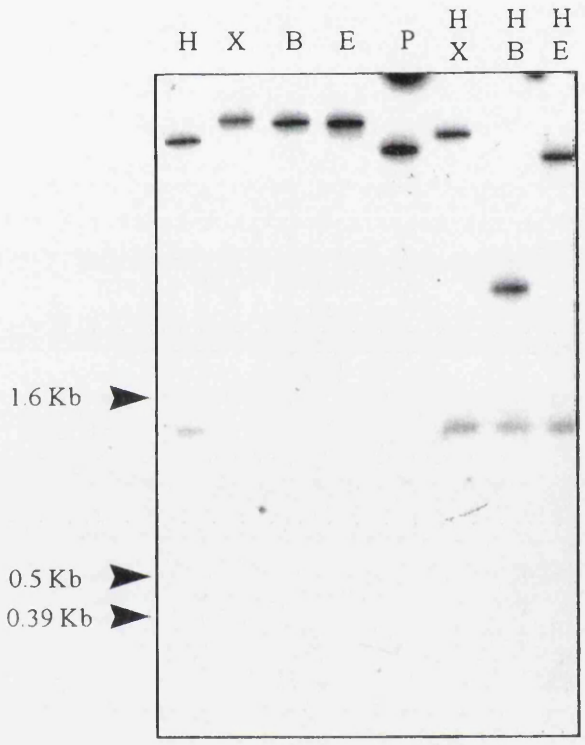


Figure 4.14 shows Southern blot of single and double digests *T. brucei* genomic DNA. Results from the hybridisation using radiolabelled 411 as a probe, indicated that *tbrlp* is a single copy gene. The hybridisation was carried out at 65° C. The enzymes used were as follows: *Hind* III (H), *Xho* I (X), *Bam*HI (B), *Eco* RI (E), *Pst* I (P) and double digests of the afore mentioned.

FIGURE 4.15

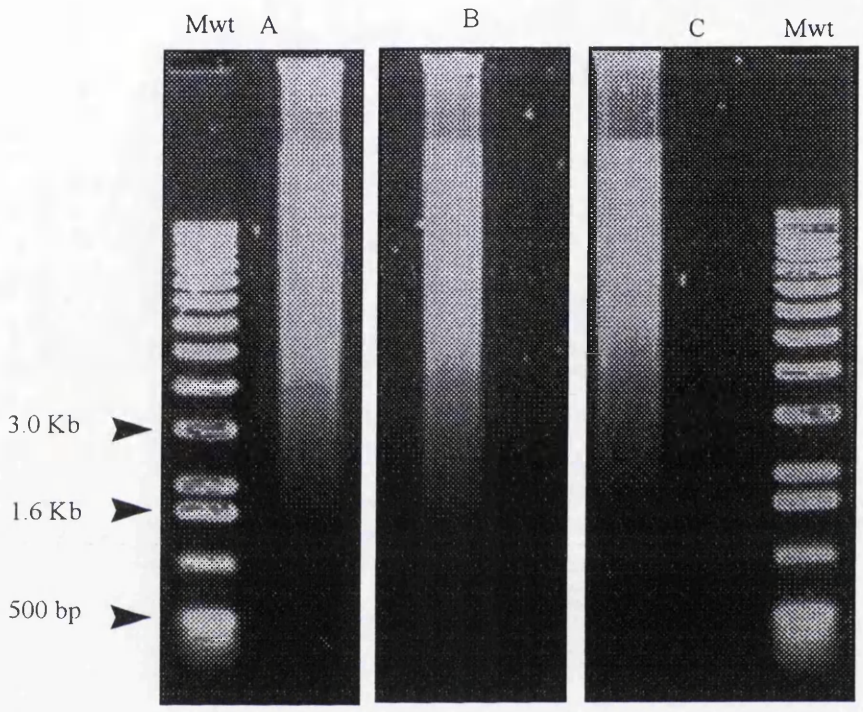


Figure 4.15 shows a 0.8 % ethidium bromide stained agarose gel of *Hind* III digested *T. brucei* genomic DNA separated in triplicate (A, B and C). The DNA was transferred onto nitrocellulose membranes and probed with *tbrlp*. Mwt - molecular weight marker. The results of the hybridisation are shown in Figure 4.16

FIGURE 4.16

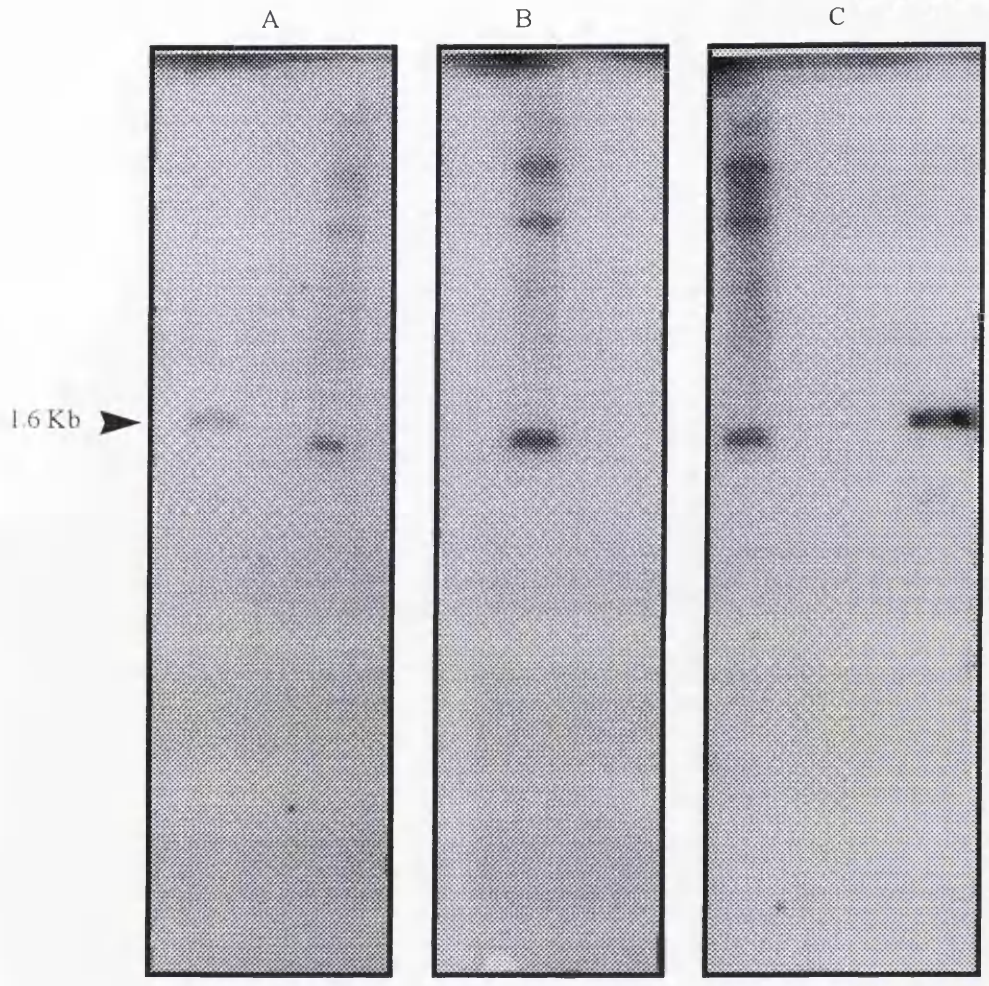


Figure 4.16 shows an autoradiograph of *Hind* III digested *T. brucei* genomic DNA probed with radiolabelled *tbrlp* . Hybridisations were carried out at 50, 55 and 65° C (from left to right). Lanes same as Figure 4.14. A number of bands (greater than 10 Kb), can be visualized at the top of the gel and the molecular sizes are consistent with partially digested genomic DNA and DNA residue in the wells. The 1.6 Kb band is a Gibco molecular weight marker, which was used in the determination of molecular sizes of the visualized bands.

The results from these Southern blots suggest that *tbrlp* is a single copy gene, but do not eliminate the possibility that there may be more than one copy of the gene, as some of the fragments are very large and could contain a tandemly linked copy. Figure 4.17 shows an autoradiograph of a Southern with *Hind* III and *Pst* I single and double digests of *T. brucei* strains TREU 927 and STIB 386. Restriction maps from the three strains of *T.brucei* are presented in Figure 4.18.

Under a specific set of PCR conditions, the amplification of *tbrlp* from genomic DNA, using primers designed for the 3' and 5' ends of the gene, resulted in the production of a second band of 1.4 Kb in addition to the expected 681 bp band, i.e *tbrlp* (Figure 4.19). Reamplification of the excised and gel purified 1.4 Kb PCR product, using the same primers, produced both the 1.4 Kb and 681 bp fragments. The simplest interpretation of this observation is that two tandemly arranged copies of *tbrlp* exist. These two copies will be referred to as *tbrlp* I (the original, sequenced copy) and *tbrlp* II (the new gene). A *Hind* III restriction digest of the 1.4 Kb PCR product, produced two predominant fragments of 1 Kb and 400 bp respectively, with the 400 bp fragment being consistent with the size of the *Hind* III fragment predicted from the sequence of *tbrlp* I. The sizes of these fragments suggest that the 400 bp fragment must contain the sequence from the *Hind* III site to the 3' end of *tbrlp* I and also show that *tbrlp* II (part of the 1 Kb fragment) is missing this *Hind* III site. The results from the Southern blots indicate that this hypothesis (that there are two tandemly linked copies of *tbrlp*) is correct so long as the second copy (*tbrlp* II) is upstream of the original (see Figure 4.18). As would be expected the amplification of *tbrlp* I and *tbrlp* II separately, occurs more readily than the two genes together, thus accounting for the high intensity of the 681 bp band relative to the 1.4 Kb band. Direct sequencing of short sections from each end of the 1.4 Kb PCR product confirmed the presence of two tandemly linked genes and showed that *tbrlp* II has an identical sequence to that of *tbrlp* I over the sequenced region.

FIGURE 4.17

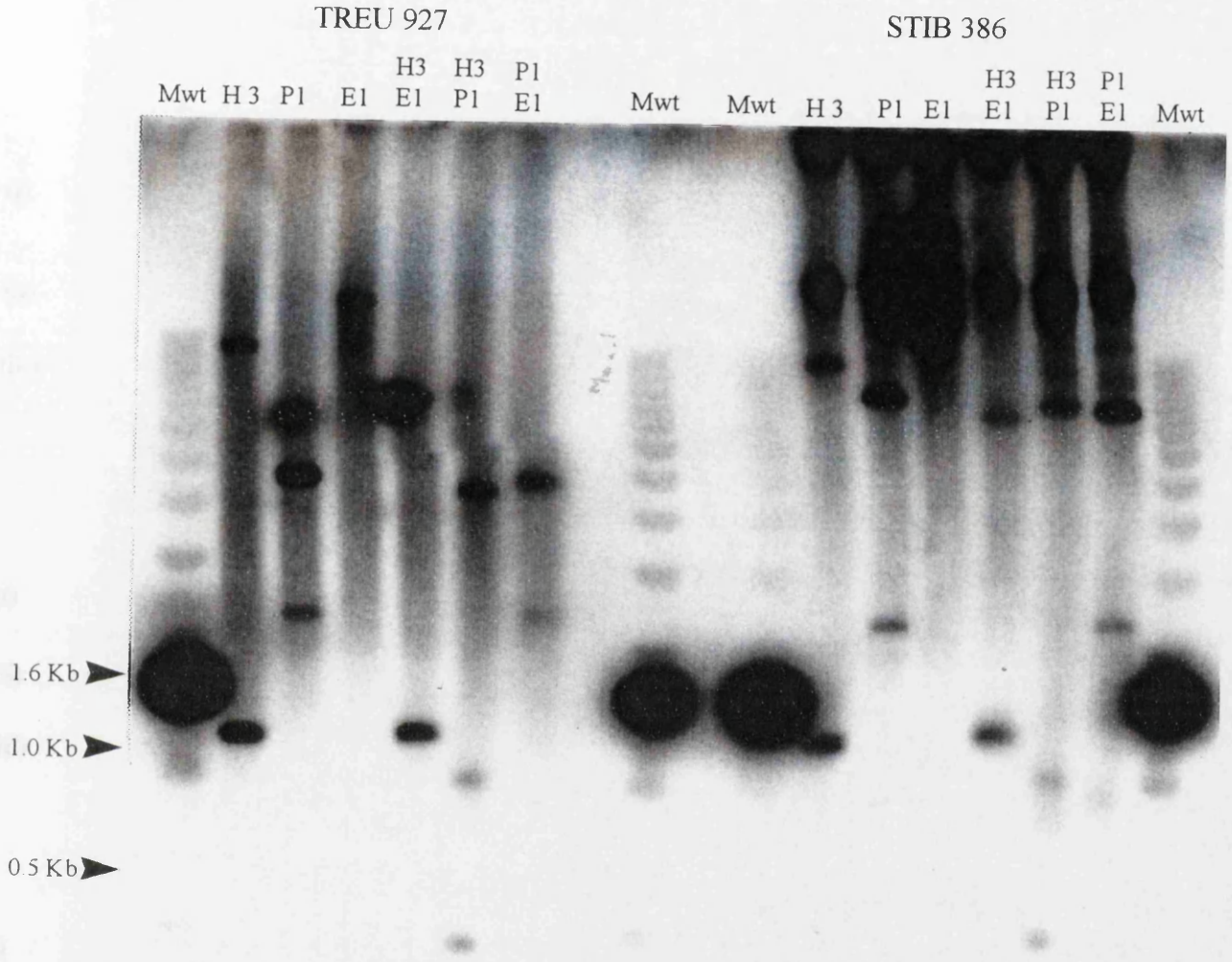


Figure 4.17 shows an autoradiograph of a Southern of single and double digests of two different strains of *T. brucei,* STIB 386 and TREU 927. H 3- *Hind* III, P1 - *Pst* I, EI - *Eco* RI, Mwt - molecular weight markers (Gibco)

FIGURE 4.18

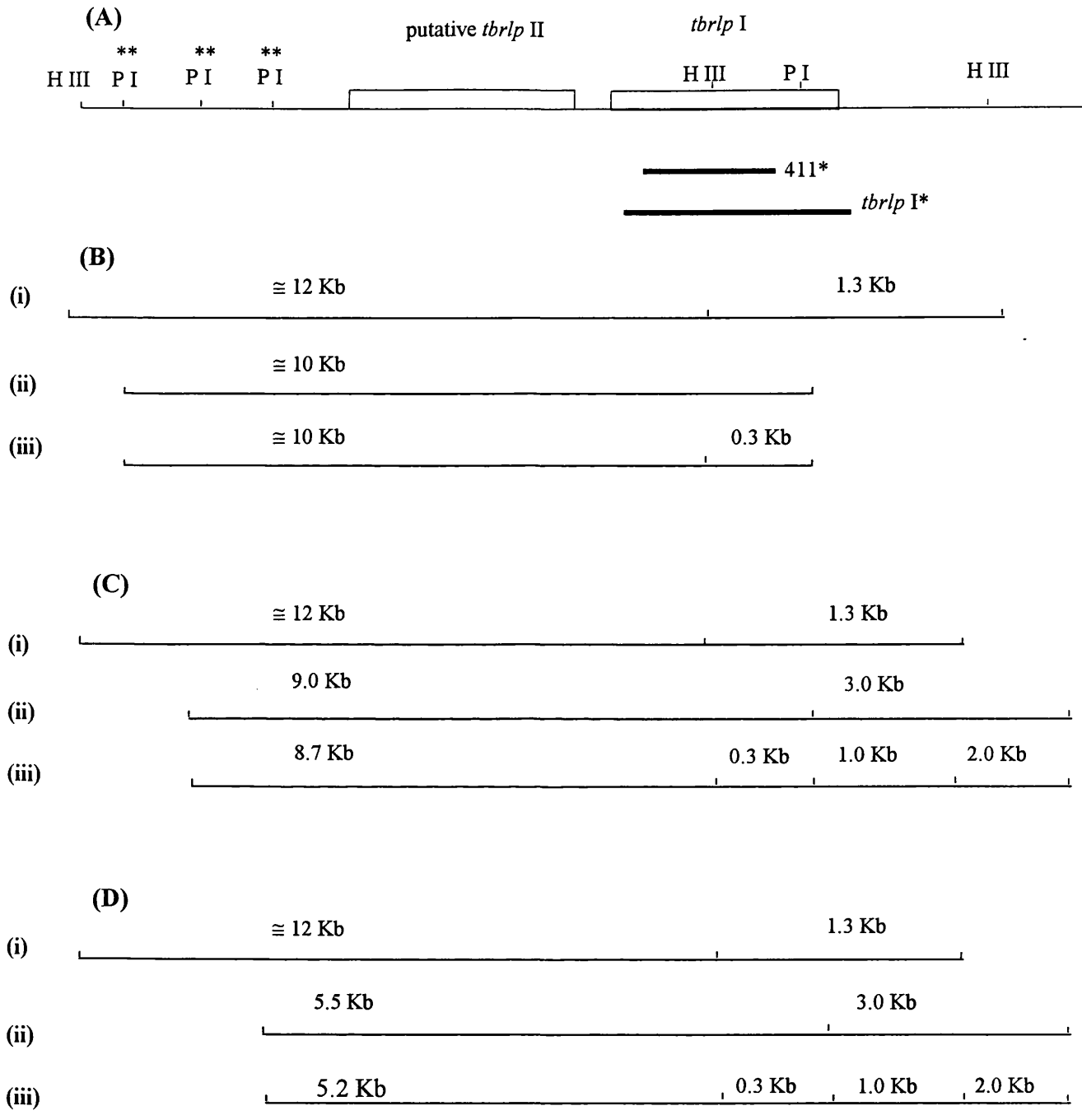


Figure 4.18 is an illustration of *Hind* III (H III) and *Pst* I(P I) restriction digest maps of *tbrlp* I and II using three different strains of *T. brucei,* namely, STIB 247 (B), STIB 386 (C) and TREU 927 (D), based on the autoradiographs shown in Figures 4.14 and 4.17. Radiolabelled 411* (STIB 247 map) and *tbrlp* I* (STIB 386 and TREU 927 maps) were used as probes. The lines next to the probes show the points of hybridisation to *tbrlp* I and II. (i), (ii) and (iii) are the results from the hybridisation of the probes to the *Hind* III, *Pst* I and a double digests of *Hind* III and *Pst* I respectively. The numbers above the respective lines (not to scale) denote the observed band sizes. (A) is the consensus restriction map showing the presence of *Pst* I (PI) polymorphisms (**).

FIGURE 4.19

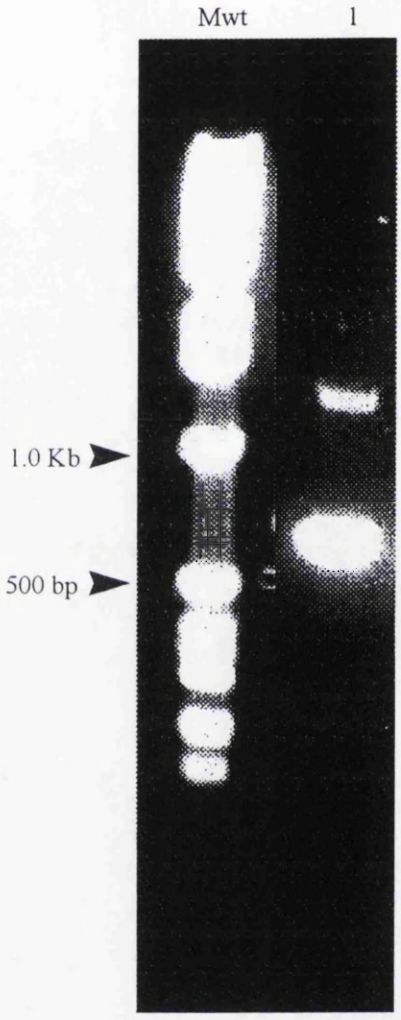


Figure 4.19 shows an ethidium bromide stained gel of the separated PCR products *tbrlp* I (681 bp) and both *tbrlp* I and the putative second copy *tbrlp* II (1.4 Kb) in lane 1. Mwt - 1 Kb ladder (Gibco), used in the estimation of molecular sizes.

4.9 Expression of *tbrlp* I

To investigate the expression of *tbrlp*, *T. brucei* bloodstream and procyclic total RNA were separated on an agarose gel, transferred onto nylon membranes and hybridised using ^{32} P labelled 411 as the probe. Figure 4.20 is an autoradiograph of a Northern blot, which shows that *tbrlp* I has a 1.4 Kb message and is expressed in both bloodstream and procyclic life cycle stages of *T. brucei*. The size of the message, when compared to the size of the open reading frame of *tbrlp* I, is indicative of a rather long 3' untranslated region. This phenomenon has been observed in a number of small GTPases but the actual significance of the length of this region is currently unknown.

FIGURE 4.20

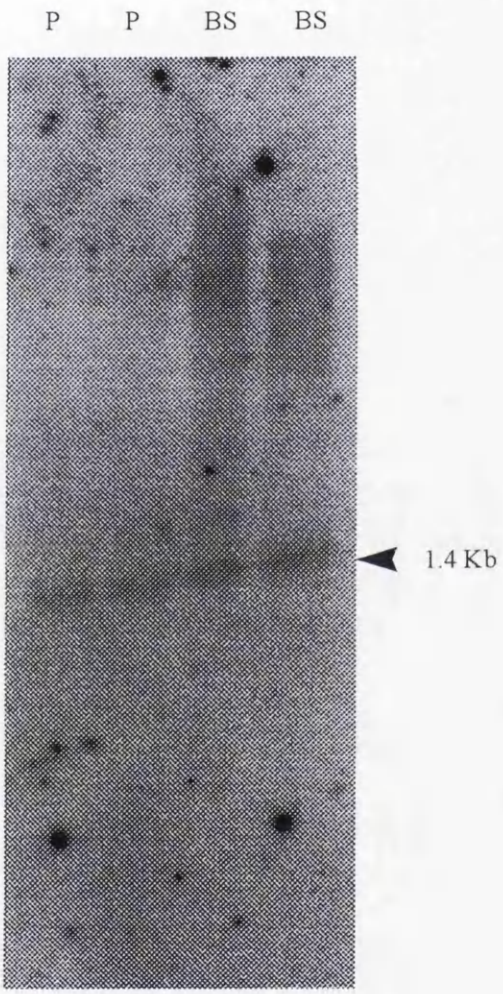


Figure 4.20 shows the results from a Northern blot of procyclic and bloodstream trypanosomes, where P - procyclic and BS - bloodstream. Clone 411 was used in probing the blot. Results from the northern show that *tbrlp* is expressed in almost equal amounts in both life cycle stages. RNA molecular weight markers (Sigma) were used in the determination of molecular sizes.

4.10 Summary and Discussion

Using degenerate PCR primers a *ras* subfamily gene was cloned from *T. brucei*. Initially a fragment (411) was sequenced and then this was complemented with further sequence of the 5' end (462) and the full length gene (*tbrlp*).

Using degenerate primer PCR, a product of 280 bp (93 amino acids) was amplified from *T. brucei* specific cDNA. Based on mammalian *ras* gene sizes (Barbacid, 1987), 411 appeared to account for approximately a third of the full length gene minus the 5' and 3' ends. The 5' end sequence was determined by using the 5' RACE technique and a primer for the spliced leader. PCR products from the reaction were cloned into the TA cloning vector and screened for the expected insert size, i.e approximately 200 bp. Using this method a clone, 462, with the expected insert size was isolated. Clone 462 has an open reading frame (ORF) of 195 bp (65 amino acids) and a short untranslated region of 56 bp precedes the start codon (ATG). The 3' end of 462 has an overlap of 86 bp with the 5' end of 411. A number of methods were used in the determination of the 3' end sequence of 411. The final sequence was derived from two PCR-based methods: 1) Chromosomal walking and 2) Long range PCR. The 3' sequence obtained by these methods were almost identical. The observed differences may have been due to the use of different strains, or, sequencing errors. Primers for the 5' and 3' ends of 411 were designed and used in the amplification of the full length gene from *T. brucei* strain TREU 927 genomic DNA, which was named *Trypanosoma brucei ras*-like GTP-binding protein or *tbrlp*.

tbrlp has an open reading frame of 681 bp (227 amino acids) which contains all the of the conserved sequence motifs characteristic of the *ras* GTPase superfamily (i.e 3 phosphate/magnesium and 3 guanine nucleotide binding sites) and a predicted molecular size of 25 kDa. The GTP-binding domain has 45 % and 41 % amino acid sequence identity to human *rap* 2 and H-*ras* subgroup genes respectively. The percentage amino acid identity of *tbrlp* to other members of the *ras* superfamily is less than 35 %. This is to be expected as the *ras* superfamily is made up a divergent group of genes with an interfamily sequence similarity of less than 50 % (Zerial and Huber, 1995). *tbrlp* is a novel *ras*-like GTPase, that has not previously been cloned from *T. brucei*. It has several unique features; the presence of glutamine (Q) in the PM 3 binding domain classifies it as a *ras* subgroup member. The C-terminal extension region has a sequence preceding the CAAX box, which has the consensus sequence for a cAMP dependent protein kinase A phosphorylation site i.e BBxS/T (where B is a basic amino acid, x is any amino acid , S is Serine and T is Threonine) (Lapetina *et al*, 1993). This phosphorylation site is characteristic of members of the *rap* subgroup e.g human *rap*1b (Bokoch, 1993). Three mammalian *ras* subgroup members are known to undergo phosphorylation by both Protein

kinases A and C. *rap* GTPases are phosphorylated solely by Protein kinase A, therefore the presence of the consensus sequence for a cAMP dependent phosphorylation site indicates that *tbrlp* has *rap* subgroup features. The CAAX box is essential for the subcellular localisation of members of the *ras* subfamily members. Members of the mammalian *ras* and *rap* subgroups are known to exist in an antagonistic relationship (Kitayama *et al*, 1989; Cook *et al*, 1993) and are never found within the same subcellular fractions (Zerial and Huber, 1995). A general rule to determine the exact mode of post translational modification exists for both *ras* and *rap* subgroup members. If the X of the CAAX box is leucine (L) or phenylalanine (F), the protein undergoes geranylgeranylation, which is characteristic of members of the *rap* subgroup. On the other hand, if X is serine (S), cysteine (C), methionine (M) or glutamine (Q), the protein will be farnesylated (Moores *et al*, 1991; Bokoch and Der, 1993). *Ras* subgroup members normally undergo farnesylation. The X of the CAAX box of *tbrlp* is leucine, which is characteristic of *rap* subgroup members. Thus there are characteristics of both *ras* and *rap* genes interspersed throughout the sequence (refer section 4.7, Figure 4.13).

A number of approaches were taken in order to investigate the presence of closely related genes and other copies of *tbrlp* in the *T. brucei* genome. High stringency hybridisation using clone n^{o} 411 as a probe indicated that *tbrlp* could be a single copy gene on the basis that multiple bands generated from other identical genes were not observed on the autoradiograph (see Chapter 4, section 4.8, Figure 4.14). However it did not eliminate the possibility that there could be more than one copy tandemly linked. Low stringency Southern blotting was used as a second approach and the results indicated that there are no other genes related to *tbrlp* in *T. brucei*. Southern blots of three different strains of *T. brucei* genomic DNA, namely, STIB 247, STIB 386 and TREU 927, identified a *Pst* I polymorphism. Surprisingly, PCR amplification of *tbrlp*, using a 5' and 3' primer, generated two fragments, a 681 bp fragment corresponding to tbrlp and a 1.4 Kb fragment. Re-amplification of the gel purified 1.4 Kb bands, using the same primers, also resulted in two products, the 681 bp and the 1.4 Kb bands. The reamplification of the 681 bp band from the gel purified 1.4 Kb fragment suggested the presence of a second tandemly linked copy of *tbrlp*. The putative second copy was named *tbrlp* II, so as to distinguish it from the original gene, *tbrlp* I. Re-examination of the restriction fragment patterns showed that they were consistent with the presence of a second copy as long as *tbrlp* II was upstream of *tbrlp* I. The DNA sequence of a short region at the 5' and 3' end of the 1.4 Kb fragment confirmed the presence of an identical second copy. Northern blot analysis was used to determine the stage specificity of the expression of *tbrlp* I and showed that the gene is expressed in both bloodstream and procyclic life cycle stages. This suggests that the role of *tbrlp* I is not confined to the bloodstream stage and therefore is unlikely to be involved in host specific signalling systems. The

result from the Northern blot also suggests that *tbrlp* might be involved in a basic process which is common to both bloodstream and procyclic stages of *T. brucei*. The ambiguous *ras* and *rap*-like features indicate that *tbrlp* is an unusual member of the *ras* subfamily. Its evolutionary significance is examined in Chapter 5.

Chapter 5

Evolutionary analysis of *tbrlp* I

CHAPTER 5

THE EVOLUTIONARY ANALYSIS OF *tbrlp* I

5.0 Introduction

A *ras*-like gene, *tbrlp* I, with an open reading frame of 681 bp, was cloned and sequenced from *T. brucei*. It has all the conserved sequence motifs characteristic of a member of the *ras* GTPase superfamily in addition to a number of unusual features. The GTP-binding domain of *tbrlp* I has respectively 45 and 41 % amino acid identity to the GTP-binding domains of human *rap* 2 and H-*ras* subgroup genes.

Analysis of the nucleotide and peptide sequences of *tbrlp* I showed that it was difficult to determine the exact subgroup to which it belonged. There is one main distinguishing feature between *ras* and *rap* genes, i.e the presence of a threonine in *rap* genes instead of a glutamine in the PM 3 domain, as is found in *ras* subgroup genes. This feature is thought to be involved in the modulation of the GTPase activity of *ras* and *rap* subgroup GTPases (Der *et al*, 1986a). In addition to the difference in the conserved *ras*-like or *rap*-like amino acid in the PM 3 domain, there are a number of conserved amino acids found in either *ras* or *rap* GTPases but not both, which are essential for normal *ras* or *rap* activity. Figure 5.0 is a schematic diagram showing the conserved amino acids in the GTP binding domains common to both *ras* and *rap*, the amino acids essential for normal functioning of *ras* and *rap* respectively and the conserved amino acids which are either *ras*-like or *rap*-like in *tbrlp* I. Due to the *rap*-like features of *tbrlp* I, e.g ^{9}L and ^{44}VE^{45} (numbers in superscript indicate the amino acid positions in *tbrlp* I's sequence), it could be placed within the *rap* subgroup, but equally, some *ras*-like features e.g ^{61} QD^{62} and some single amino acids that are conserved in *ras* subgroup genes, e.g ^{73}R and ^{138}F, indicate that it could be placed within the *ras* subgroup. This suggests that it may have an interesting evolutionary relationship with the *ras* and *rap* genes of higher eukaryotes. The evolution of *tbrlp* I could have occurred in a number of ways. To examine the evolutionary origin of *tbrlp* I, the possible hypotheses explaining the evolution of a *ras* subfamily member with mixed features, will be presented, tested and discussed in the following sections. As functional data is not yet available for *tbrlp* I, only DNA and amino acid sequence analysis will be considered.

FIGURE 5.0

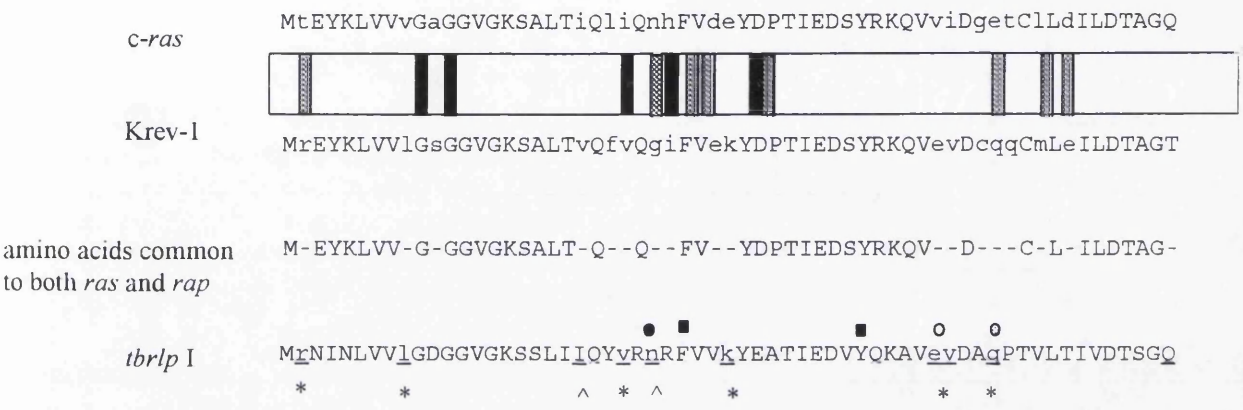


Figure 5.0 is a schematic diagram showing the amino terminus of one *ras* (c-ras) and one *rap* gene (Krev-1). These genes represent typical sequences of *ras* and *rap* genes. On the bar diagram the white areas represent identical amino acids in both *ras* and *rap* genes, the lightly shaded regions represent amino acid differences between *ras* and *rap* and the regions shown in black represent amino acids that have undergone conservative changes. A consensus sequence showing the amino acids common to both *ras* and *rap* genes is presented below the Krev-1 sequence. The dashes represent the amino acid differences between the two genes. The sequence of *tbrlp* is presented below the consensus sequence showing the amino acids common to both *ras* and *rap* genes. The underlined amino acids represent the presence of either *ras* like (^) or *rap* like (*) residues in *tbrlp* and highlights the alternation between *ras* and *rap* residues essential *ras*/*rap*-effector interactions. ● represents a *ras* like amino acid which is essential for normal *ras*-effector interactions; ■ residues common to both *ras* and *rap* GTP binding proteins and essential for normal *ras*/*rap* effector interactions; o *rap* like residue, essential for normal *rap*-effector interactions; o the amino acid found in this position in both *ras* and *rap* GTP binding proteins is also essential (different amino acids are conserved in both cases) for *ras*/rap-effector interactions. The amino acid at this position in *tbrlp* is neither *ras* like nor *rap* like. Adapted from Zhang *et al*, 1990, Valencia *et al*, 1991 and Marshall, 1993. c-ras and Krev-1 are from mammalian cells.

5.1 Small GTP-binding proteins in the trypanosomatids

A total of eleven small GTP-binding proteins have been cloned and fully sequenced from two members of the order Kinetoplastida. One *ypt*, has been cloned from *Leishmania* (Cappai *et al*, 1993), and one *ran* and nine *rab* GTPase genes from *Trypanosoma brucei* (Field and Boothroyd, 1995; Field *et al*, 1995; El Sayed *et al*, 1995). To date no genes for small GTP-binding proteins of the *ras* subfamily have been cloned from any of the trypanosomatids. Figure 5.1 shows an alignment, created by the GCG PILEUP program, of the protein sequence of the GTP-binding domains of four trypanosomatid small GTPases and that of *tbrlp* I. A list of the accession numbers of the genes used in the alignment are shown below (Table 5.0). The alignments clearly show that *tbrlp* I is a novel small GTPase, which is unrelated to those previously cloned from *T. brucei* and the other trypanosomatids. Table 5.1 shows the level of amino acid identity between the cloned trypanosomatid genes and *tbrlp* I. A number of other partial sequences (Expressed Sequence Tag sequences) are available for *T. brucei* GTPase-like sequences and *tbrlp* is clearly not represented by any of them.

TABLE 5.0

List of accession numbers for trypanosomatid monomeric GTPases

ORGANISM	GENE	ACCESSION Nº
Leishmania major	*ypt*	L12031
Trypanosoma brucei	*ran*	U17085
Trypanosoma brucei	*rab* 4	T26881
Trypanosoma brucei	*rab* 5	T26882

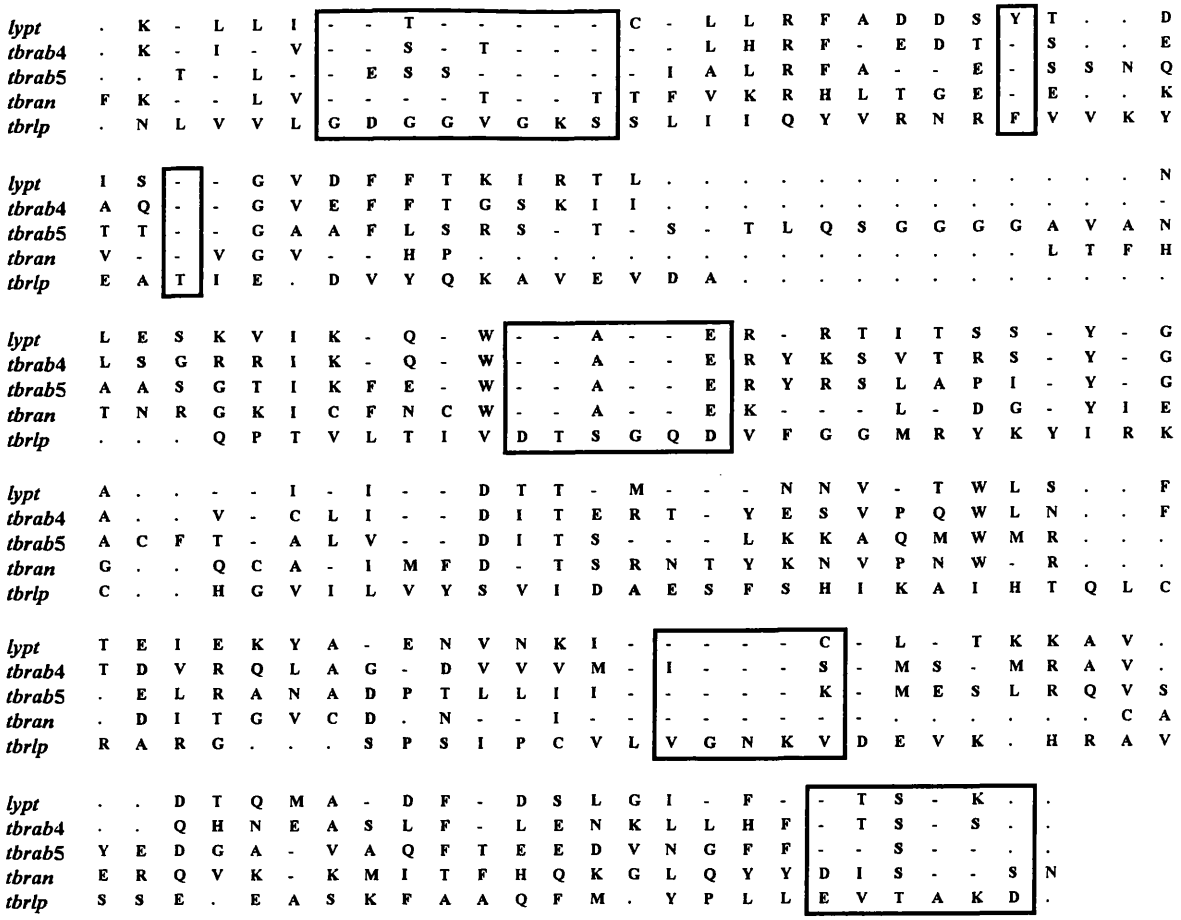


Figure 5.1 shows an alignment of the GTP binding domains of *tbrlp* I and four other trypanosomatid *ras* like genes, *Leishmania ypt*, and *T. brucei ran*, *rab*4 and *rab*5. The bordered areas correspond to the six highly conserved regions of the GTP binding domain, characteristic of members of the *ras* GTPase superfamily. The dashes indicate identical amino acids and the dots indicate gaps in the sequence.

TABLE 5.1

Level of amino acid identity of the GTP binding domains of cloned trypanosomatid genes to tbrlp I

ORGANISM	GENE	AMINO ACID IDENTITY (%)
Leishmania major	*ypt*	26 %
Trypanosoma brucei	*ran*	20 %
Trypanosoma brucei	*rab* 4	23 %
Trypanosoma brucei	*rab* 5	22 %

5.2 Hypotheses explaining the origins of *tbrlp* I

A number of possible hypotheses could explain the evolutionary relationship of *tbrlp* I with other *ras*-like GTP-binding proteins. These hypotheses are presented and discussed in the following sections.

5.21 *Genetic recombination*

T. brucei has been shown to be capable of undergoing genetic exchange (Tait, 1983; Tait and Turner, 1990; Sternberg and Tait, 1990; Gibson, 1995; Swindle and Tait, 1995). Genetic recombination (i.e a recent event in evolutionary terms), therefore, could be a plausible explanation of the evolutionary origins of *tbrlp* I. If the mixed characteristics of *tbrlp* I occurred as a result of genetic recombination, then the following predictions can be made: (1) Different stocks of *T. brucei* would not necessarily possess *tbrlp*, but may have *ras* and *rap* homologues i.e *tbrlp* I would be isolated from some stocks but not others; (2) The GTP-binding domain would not have interspersed *ras* and *rap*-like features throughout, i.e one part of the domain would be *ras*-like and the other part *rap*-like. Figure 5.2 is an illustration of the results that would be expected if the origins of *tbrlp* I could be accounted for by genetic recombination.

FIGURE 5.2

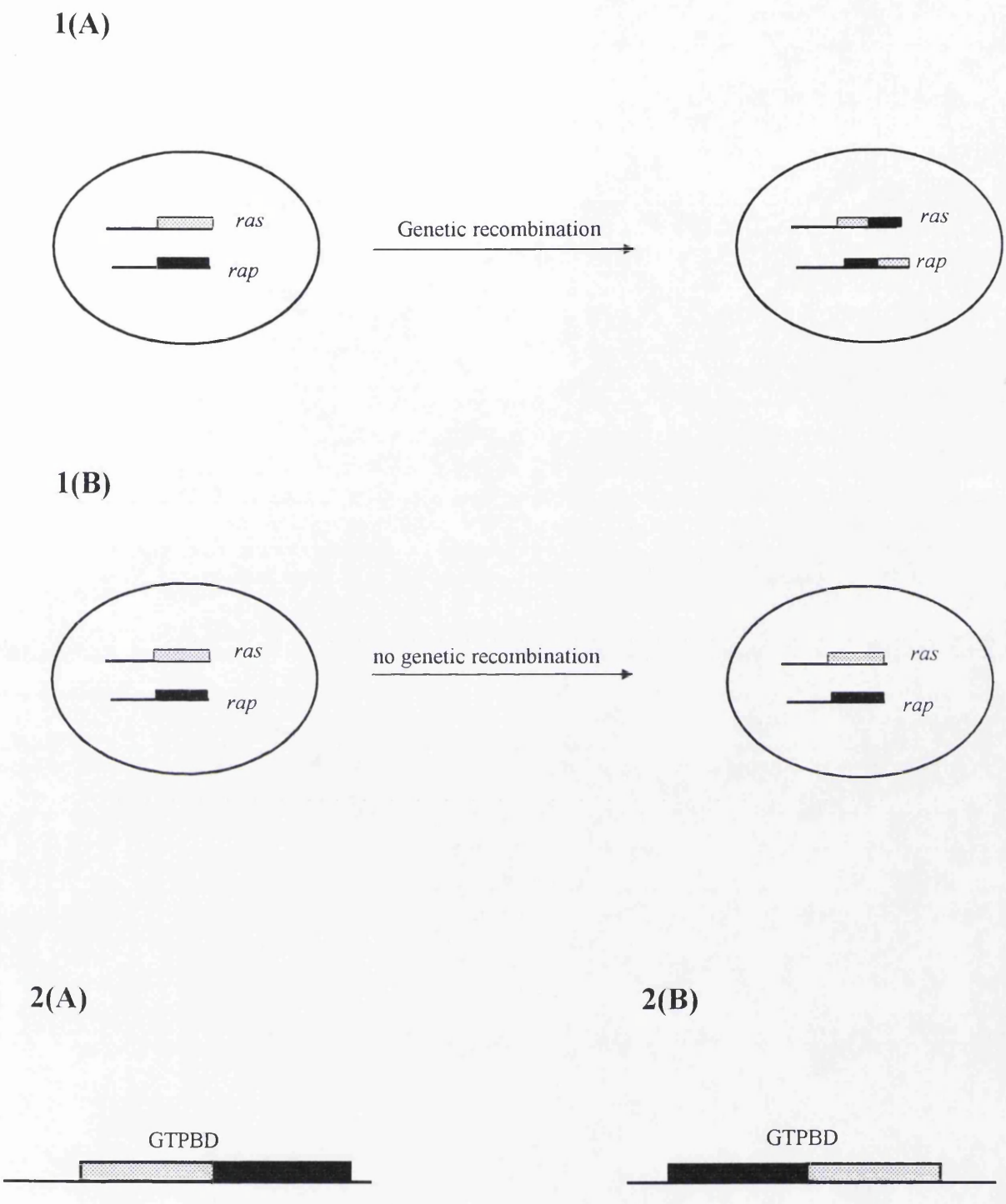


Figure 5.2 is an illustration of the expected results, which could be used in to explain the evolutionary origins of *tbrlp* I. If *tbrlp* I originated as a result of genetic recombination, then one would expect to find it in one stock (1(A)), but not the other, (1(B)). Genetic recombination would most likely result in a gene with a GTP-binding domain arranged in blocks rather than interspersed. as illustrated in 2 (A) and (B). ▭ - denotes *ras* like features and ▬ - *rap* like features. GTPBD - GTP binding domain.

5.22 *Recombination as a result of an ancient evolutionary event*

The evolutionary origins of *tbrlp* I could have been as a result of a recombination event in an ancestral kinetoplastid. If this hypothesis is true then three predictions can be made, (1) all kinetoplastids should have a *tbrlp* I like gene; (2) other ancient eukaryotes, i.e more ancient than the kinetoplastids, (e.g *Giardia*) should have both *ras* and *rap* genes; and (3) the GTP-binding domain should have *ras* and *rap* features in blocks rather than being interspersed (see Figure 5.2).

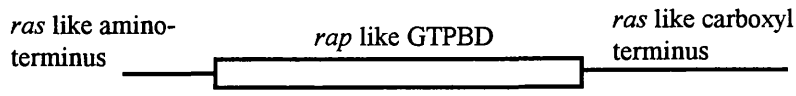
5.23 *Ancestral insertion of a GTP-binding domain*

Evolution occurs as a result of the modification of genes. The modification of a gene usually occurs as a result of a mutation (or a number of mutations). At the DNA level, four main types of mutations occur, transitional, transversional and the insertion or deletion of individual nucleotides (or a whole block of nucleotides). Insertions and deletions are found to occur frequently in evolutionary time.

The origins of *tbrlp* I could therefore be explained by an ancient evolutionary event, which could have resulted in the deletion and subsequent insertion of a *ras* GTP-binding domain into a *rap* gene or vice versa. The prediction from this hypothesis is that, a gene with mixed features should, for example, have a *ras*-like amino terminus, a *rap*-like GTP-binding domain and a *ras*-like carboxyl terminus. Figure 5.3 is an illustration of the result of an ancestral insertion of a *ras*-like or *rap*-like GTP-binding domain into a *rap* and *ras* gene respectively.

FIGURE 5.3

(A)



(B)

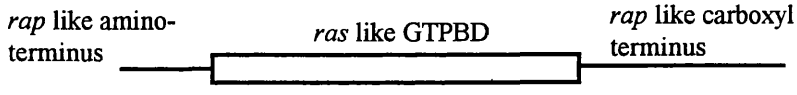


Figure 5.3 (A) is a schematic diagram of a typical *ras* gene, with a *rap* like GTP-binding domain (GTPBD). (B) is a schematic diagram of a typical *rap* gene with a *ras* GTP-binding domain (GTPBD). Both of which could have been as a result of an ancestral insertion of a GTP-binding domain into either a *ras* or a *rap* gene characteristic of the higher eukaryotes.

5.24 *New member of the ras GTPase superfamily*

tbrlp I could be a new member of the *ras* GTPase superfamily and co-exist with other *ras* and *rap* subgroup members. The prediction of this hypothesis is that, in addition to *tbrlp* I, *ras* and *rap* genes would also be detected in *T. brucei*.

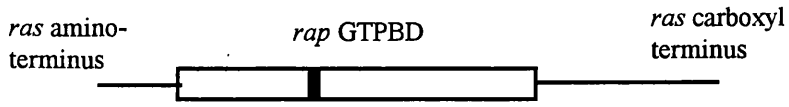
5.25 *The evolutionary loss of either ras or rap*

If *T. brucei* initially had both *ras* and *rap* genes, but lost one during the course of evolution, then *tbrlp* I may have evolved to compensate for the loss of one of the genes. In such a case, the sequence of the gene should be predominantly *ras* or *rap*-like, with the addition of important functional domains of either *rap* or *ras* respectively. The PM 3 binding domain of members of the *ras* subfamily is essential for interactions with downstream effectors. Therefore if this hypothesis is correct, the resulting gene should have a *rap*-like sequence with key *ras* residues. The alternative hypothesis would be the loss of a *rap gene* and the subsequent addition of *rap* functional residues to a *ras* gene. The result of such a process would lead to a gene product with a dual function and should be detected from the gene sequence.

Figure 5.4 (A) is an illustration of a *rap* gene specific PM 3 binding domain (*rap* genes have a threonine (T) in the PM 3 binding domain, whilst *ras* genes have a glutamine at the same position) with a *ras*-like carboxyl terminus. 5.4 (B) is an illustration of a *ras* gene with a *ras* GTP-binding domain, but with the addition of a Protein kinase A specific phosphorylation site (characteristic of some *rap* GTPases), at the carboxyl end.

FIGURE 5.4

(A)



(B)

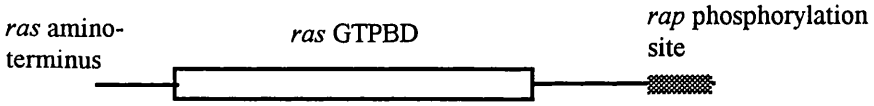


Figure 5.4 A is an illustration of *ras* gene with a *rap* specific threonine in the PM 3 binding site of the GTP-binding domain (GTPBD) and an additional *ras* like carboxyl terminus. 5.3 B is an illustration of a typical *ras* gene with a *ras* specific glutamine in the effector binding domain and a *rap* specific phosphorylation site at the carboxyl terminus. ▌ - *rap* specific PM 3 binding site; ▒▒ - *rap* specific (i.e Protein kinase A) phosphorylation site.

5.26 *An ancestral ras subfamily gene*

Another hypothesis is that *tbrlp* I could be a member of an ancestral *ras* gene subfamily. The gene could have been retained in *T. brucei*, but subsequently evolved by gene duplication and divergence into the *ras* and *rap* genes of higher eukaryotes. Assuming that, after duplication of a primeval GTP-binding domain, the duplicated genes diverged into the *ran*, *rab* and *ras* subfamily GTPases, and *T. brucei* evolved prior to the duplication and divergence of a *ras/rap* ancestor into *ras* and *rap*, then *T. brucei* would possess the ancestral gene in addition to *ran* and *rab* GTPases.

The predictions of this hypothesis are: (1) the sequence of the GTP binding domain of the ancestral *ras/rap* gene, should have *ras* and *rap*-like features interspersed throughout rather than grouped into blocks of *ras* or *rap* sequence; (2) Phylogenetic analysis of *tbrlp* I and the *ras* and *rap* genes of other eukaryotes, should show a point of divergence intermediate to the *ras* and *rap* genes and (3) Phylogenetic analysis using *rab*, *ran*, *ras* and *rap* from higher eukaryotes, should show that the ancestral gene diverges from the main evolutionary branch from the *ras* subfamily members to the *rab* and *ran* GTPase gene families.

Figure 5.5 is a schematic diagram of the proposed evolutionary pathway of the divergence of an ancestral *ras/rap* gene from a primeval GTP-binding domain, to the *ras* and *rap* genes of higher eukaryotes. An illustration of the expected results from the phylogenetic analysis of *tbrlp* I (an ancestral *ras/rap* gene) is shown in Figure 5.6.

FIGURE 5.5

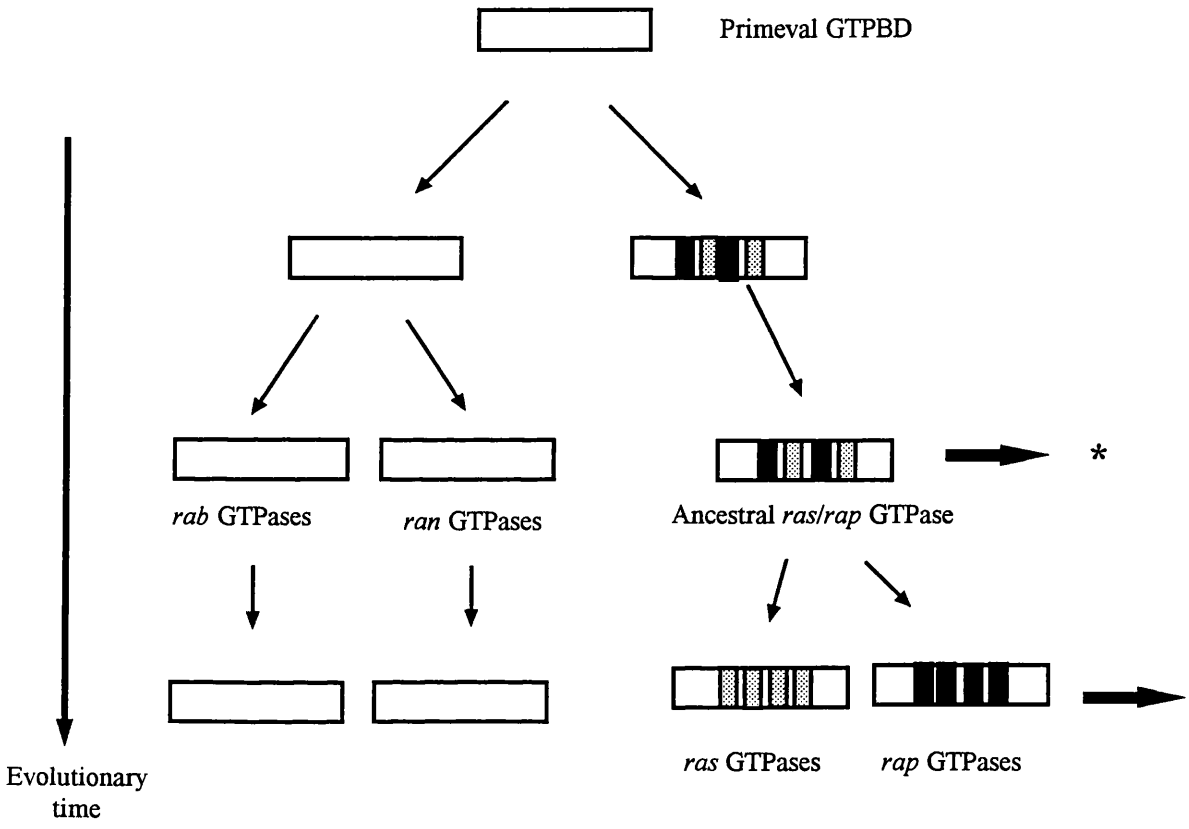


Figure 5.5 is a schematic diagram of the evolutionary pathway of an ancestral *ras/rap* gene, from a primeval GTP binding domain (GTPBD) to the *rap* and *ras* genes of higher eukaryotes. It shows the point of divergence of the *ran* and *rab* genes of higher eukaryotes, and the points of evolution of the kinetoplastids and other eukaryotes higher up the evolutionary tree e.g *Physarum polycephalum*. * - evolution of the kinetoplastids; ** - evolution of other eukaryotes; ▒ - *ras* like features; ▉ -*rap* like features.

FIGURE 5.6

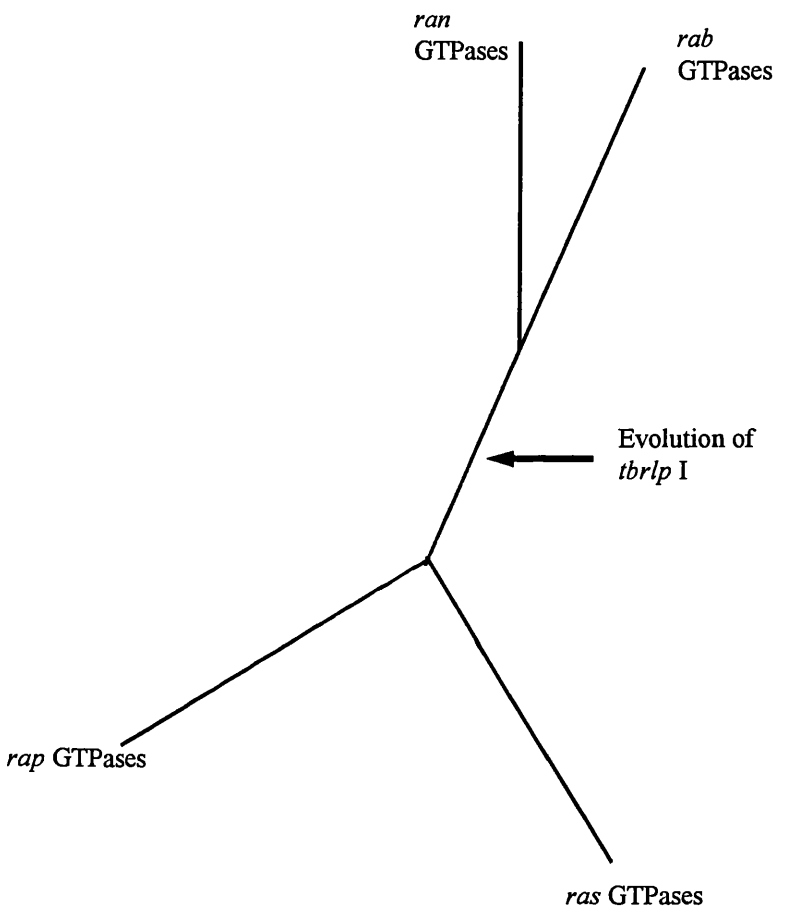


Figure 5.6 is a schematic diagram of the expected results from the phylogenetic analysis of *tbrlp* I and other members of the *ras* GTPase superfamily. If *tbrlp* I is an ancestral *ras/rap* GTPase gene, then the topology of the tree should show that *tbrlp* I, diverged from the main evolutionary branch from the *ras* subfamily members to the *ran* and *rab* GTPase gene families. The point of divergence of *tbrlp* I (indicated by the arrow), should precede the divergence of *ras* and *rap* GTPases.

5.3 Methods used in phylogenetic analysis

The ordered sequence of nucleotide bases, codons and amino acids in genes can be used to recover information about the evolution of genes or organisms i.e the phylogeny of those genes or organisms. Using phylogenetic methods, the degree of relatedness and difference between different genes or organisms can be measured using the basic assumption that closely related organisms or genes (based on DNA/protein sequence) are closely related in evolutionary terms. The levels of relatedness can be expressed as trees where the branch lengths indicate the degree of relatedness and the branch points (or nodes) indicate the possible point of divergence of the two branches. Two types of phylogenetic trees can be distinguished: species trees and gene trees. Species trees can only be constructed from data originating from functionally homologous genes (i.e orthologous genes, Schlegel, 1994). Species trees are meaningless if the data is taken from genes which have no proven functional homology. Gene trees, on the other hand, map the evolution of a gene (or gene family) and can include other genes with different functions, as long as it is clear that those genes have a common ancestor (e.g the GTP binding domain of *ras*-like proteins, Schlegel, 1994). The comparison of the DNA and amino acid sequences of the GTP binding domains of *tbrlp* I and other *ras* superfamily members, suggests that the genes have a common ancestor, i.e the GTP binding domain. Due to this observation, and the current lack of knowledge of the function of *tbrlp* I, only gene trees will be considered for the phylogenetic analysis of *tbrlp* I.

5.31 *The process of phylogenetic tree construction*

A number of steps must be taken so as to ensure that the results from the analysis will produce the maximum amount of information possible. First and foremost, there should be a specific hypothesis to test. Based on the hypothesis, one has to determine whether or not, (a) there are any suitable sequences available for the analysis, that is to say, are there any closely related sequences (i.e with a common ancestor) that can be used in the analysis? (b) a suitable outgroup must be picked for the analysis. The use of outgroups in phylogenetic analysis is important as it provides both the direction and time scale of evolution on a tree. The divergence of the outgroup must antedate the divergence of the branch point in question (Brown and Doolittle, 1995). Last of all, but not the least, (c) one must choose a suitable tree construction method to ensure that the maximum amount of information possible is obtained. To ensure this, the following steps must therefore be taken:

(a) Choice of sequences: The sequences to be used in the analysis should be closely related and be able to be unambiguously aligned. With highly divergent genes where considerable levels of insertions and deletions have occurred, it is important to use the most highly conserved region of a set of closely related genes for phylogenetic analysis (e.g the GTP binding domain of the *ras* superfamily) in order to ensure unambiguous alignment.

(b) Choice of an outgroup: Brown and Doolittle (1995), state that the most suitable outgroup would be one from a gene family in which duplication antedates divergence. For the analysis of a *ras* subfamily member, the most suitable outgroup would be any one of the other *ras* superfamily members, e.g. *ran* or *rab*.

(c) Choice of a suitable tree construction method: There are three categories of methods currently available for tree construction. The first category is made up of a set of methods that determine the most parsimonious tree based on the smallest number of mutational changes that have occurred between two given extant sequences e.g maximum parsimony (Nei, 1987). The second category is made up of the distance matrix methods which are based on a pairwise estimation of genetic distance between two or more extant sequences. The pairwise estimation normally results in the clustering of related genes based on similar amino acid or DNA sequences. The basic assumption of these methods is that all closely related sequences are closely related in evolutionary terms. A number of distance matrix dependent tree construction methods are currently available e.g UPGMA, neighbor-joining, and Fitch-Margoliash (Nei, 1987). The third category is based on a statistical approach. The most commonly used statistical method for phylogenetic analysis is the maximum likelihood method. This method determines statistically, the tree with the highest maximum likelihood values, based on the minimum number of changes required to move from one sequence to another (Schoeniger and von Haeseler, 1993). These methods will be discussed in the following sections.

Maximum Parsimony

The maximum parsimony method produces a tree which requires the least number of mutational changes (i.e the most parsimonious tree). The most parsimonious tree, is that which requires a minimum number of evolutionary changes to explain the observed differences between amino acid or DNA sequences. More often than not, several trees are chosen as the most parsimonious, this is due to the fact that there are often several ways of going from one sequence

to another with the minimum number of changes. Maximum Parsimony is a phylogenetic method based on one major assumption, which is that evolution occurs in a parsimonious way. A number of computer programs are available for maximum parsimony tree construction these include; PROTPARS (for protein sequences) and DNAPARS (for DNA sequences). Both programs are part of Felsenstein's PHYLIP package version 3.5c, 1993.

Distance matrix methods:

A number of distance matrix methods are available for tree construction. In the following section, three distance matrix methods will be described and the differences discussed.

Neighbour-Joining method

The neighbour-joining method is related conceptually to the UPGMA method (see later). The trees are constructed by the successive clustering of lineages and branch lengths are formed as the lineages cluster (Saitou and Nei, 1987). However, this method produces unrooted trees unless an outgroup is specified and does not have the requirement that all branches evolve at the same rate, as does the UPGMA method. A transformed distance matrix is constructed at each step of the analysis. The matrix is used to calculate the branch lengths between each pair of nodes, based on the nearest neighbour distance from all other nodes. A number of methods are available for the computation of distance matrices for the NJ method, such as Jukes-Cantor, Kimura two-parameter distance method, Dayhoff PAM or gamma distances. These computational methods are needed to correct the observed distances to take into account the reversion rate of base or amino acid changes. Any of these distance methods can be used in the construction of phylogenetic trees, provided the number of amino acid or nucleotide substitutions per site is low (Tateno *et al*, 1994). The NJ method is found to be more efficient in the construction of trees, with the correct topologies, when varying rates of gene substitution occur in different branches (Kuhner and Felsenstein, 1994). In computational terms, it is faster than the Fitch-Margoliash method.

Fitch-Margoliash method

This is an additive tree model in which three groups of sequences (also referred to as operational taxonomic units or OTUs) are compared at a time. Two OTUs with the least evolutionary distance between them are found and designated the letters A and B. The third OTU is designated the letter C, which is actually made up of the rest of the OTUs being used in the analysis. A and B are then clustered together and subsequently used as a single OTU. A new set of three OTUs is picked based on the least evolutionary distance and clustered as previously described. This process is continued until all OTUs have been clustered. The Fitch-Margoliash method also produces unrooted trees unless an outgroup is specified and can be used in cases of varying gene substitution rates (Kuhner and Felsenstein, 1994).

Unweighted pair-group method with arithmetic mean (UPGMA)

Trees constructed using the UPGMA method of Sneath and Sokal, (1973) are rooted, distance trees. The rate of nucleotide/amino acid substitution is assumed to be constant and therefore the distance measure is proportional to evolutionary time. The evolutionary distance is calculated for all pairs of sequences, (which are referred to as operational taxonomic units or OTUs). A distance matrix is formed using an average linkage method of clustering, starting from the smallest distance. The branch lengths between two clusters is expressed as the average distance between all members of the first cluster and all members of the second cluster. A major limitation of this method is the lack of compensation for the varying rates of evolution which is often observed along different branches.

All of these distance matrix methods are part of the PROTDIST and NEIGHBOR programs (for protein sequence) of the PHYLIP package (DNADIST is available for nucleotide sequences) (Felsenstein, 1993).

Maximum likelihood method

This a statistical method based on the maximum likelihood of obtaining the actual minimum number of amino acid or DNA sequence substitutions required to change from one sequence to another. The program chooses the tree with the highest maximum likelihood value.

Both rooted and unrooted tree can be constructed using this method. The basic assumption of this method is that all closely related sequences have unequal rates of evolution. The different rates of evolution are therefore taken into consideration when the method is being used. The maximum likelihood method for tree construction can be carried out using the PUZZLE program (Schoeniger and von Haeseler, 1993; Strimmer and von Haeseler, 1996). The program computes maximum likelihood distances in a pairwise manner for amino acid sequences using correction methods such as, Dayhoff PAM, Dayhoff (1978).

5.4 Resampling methods

In order to ensure the statistical accuracy of the constructed phylogenetic trees a number of resampling methods were considered. Two main methods available for resampling are: Bootstrapping and Jackknifing. Both methods involve the random reassortment of input data with the production of trees for each data set. Bootstrapping (Efron and Gong, 1983; Efron and Tibshirani, 1991) is the most commonly used method for the statistical analysis of phylogenetic trees derived from protein sequences. The method reassorts the input data and produces a tree for each set of data . The percentage of the number of times that a particular point of divergence occurs in each data set is calculated. High bootstrap values indicate high confidence levels at the points of divergence.

5.5 Procedure for the phylogenetic analysis of *tbrlp* I

The main concern was to ensure that the results from the phylogenetic analysis were as accurate as possible. After considering all the methods available for the inference of phylogenies, five of the most commonly used methods for tree construction, and a resampling method were chosen for the analysis. Different methods are occasionally known to give different results, I therefore decided to use all the methods to ensure that consistent tree topologies and high bootstrap values were obtained in all cases.

150 amino acids within the GTP-binding domain were used for the analyses, starting from the fifth amino acid after the start codon (methionine) and ending at amino acid number 160. This region included all six highly conserved sites within the GTP binding domain. The C-terminal extension region of *tbrlp* I could not be unambiguously aligned with other *ras* genes, and was, therefore, not used in the analysis. Sequences from the GTP-binding domain of *ras* and *rap* genes from a number of sources were aligned using the PILEUP program. All sequence alignments were subsequently checked and manually aligned, if necessary, using the LINEUP program of the GCG package. After alignment of the sequences the following steps were taken:

1) TOPHYLIP, an EGCG program was used to format the sequence files into a suitable input file for the PHYLIP phylogeny inference programs.

2a) SEQBOOT was used to bootstrap the input file from TOPHYLIP.

2b) Trees were also constructed without bootstrapping and the topologies were compared to the bootstrapped ones.

3a) PROTDIST, using the Dayhoff PAM option, was used to generate a distance matrix based on amino acid differences.

3b) PROTPARS was used to generate a maximum parsimony tree based on the minimum number of amino acid changes between sequences.

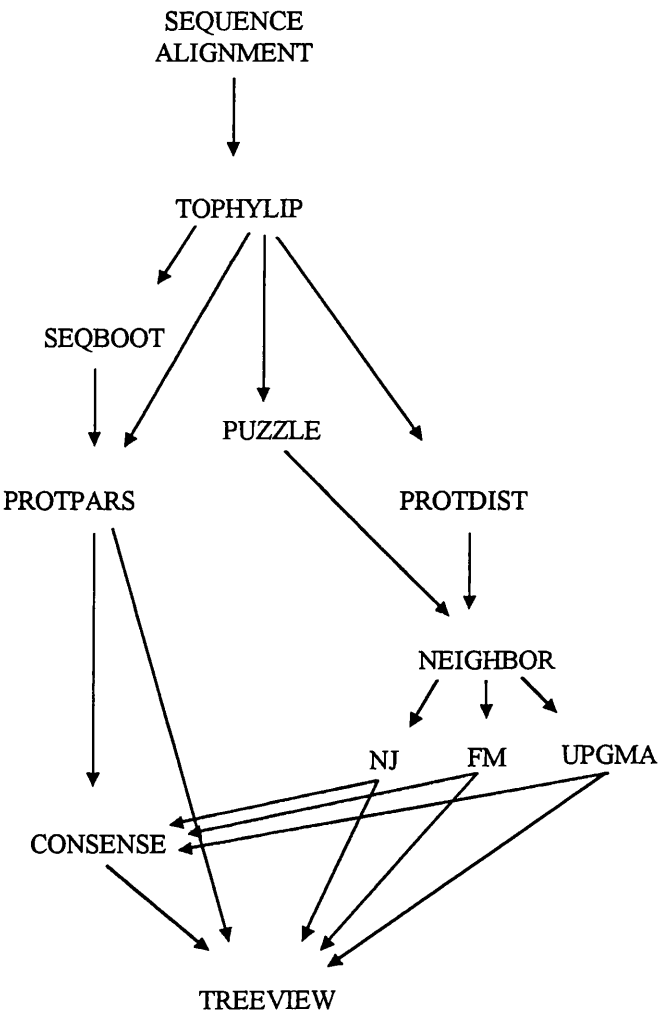
3c) PUZZLE was used to generate a maximum likelihood tree based on the number of amino acid changes.

4) The output files from 3a) and 3c) were analysed using the NEIGHBOR program. Three of the options from the NEIGHBOR program were used : NJ, FM and UPGMA. (NJ was used in the construction of the maximum likelihood trees).

5) CONSENSE was used to generate consensus trees from the bootstrapped output files of 3b) and 4.

6) TREEVIEW, (Page, 1995), was used to redraw the trees to produce slanted cladograms, phylograms and radial trees. The data from 3b, 4 and 5 were used to generate trees from TREEVIEW. Figure 5.7 is a flow chart of the steps taken in the phylogenetic analysis of *tbrlp* I.

FIGURE 5.7



A flow chart of showing the steps taken in the phylogenetic analysis of *tbrlp* I.

5.6 Testing the hypotheses of the evolutionary origins of *tbrlp* I.

A number of hypotheses explaining the possible origins of *tbrlp* I, were presented in section 5.2. In order to determine the origins of *tbrlp* I, these hypotheses were tested to determine whether they explain the evolutionary origins of *tbrlp* I. The results from the tests of each of these hypotheses are presented in the following sections.

5.61 *Genetic recombination*

The hypothesis in this case is that *tbrlp* I originated as a result of genetic recombination. If this is correct then one would expect to find *tbrlp* I in some *T. brucei* stocks but not others. *Tbrlp* I has however been isolated and sequenced from two different stocks of *T. brucei* i.e cDNA from strain TREU 869 and genomic DNA from *T. brucei* strain TREU 927. A PCR product of the same molecular size has also been amplified from *T. brucei* strain STIB 247 genomic DNA. One would also expect that the GTP-binding domain of *tbrlp* I would be made up of a *ras*-like region confined to one end and a *rap*-like region confined to the other, i.e a hybrid. The sequence of *tbrlp* I does not show such features. The *ras*-like and *rap*-like features are interspersed throughout the GTP-binding domain. Figure 5.8 is an alignment of the protein sequences of *ras* and *rap* genes of other eukaryotes, showing the mixed features of the GTP-binding domain of *tbrlp* I.

The results from the test clearly show that the evolutionary origins of *tbrlp* I are unlikely to be attributed to recent genetic recombination.

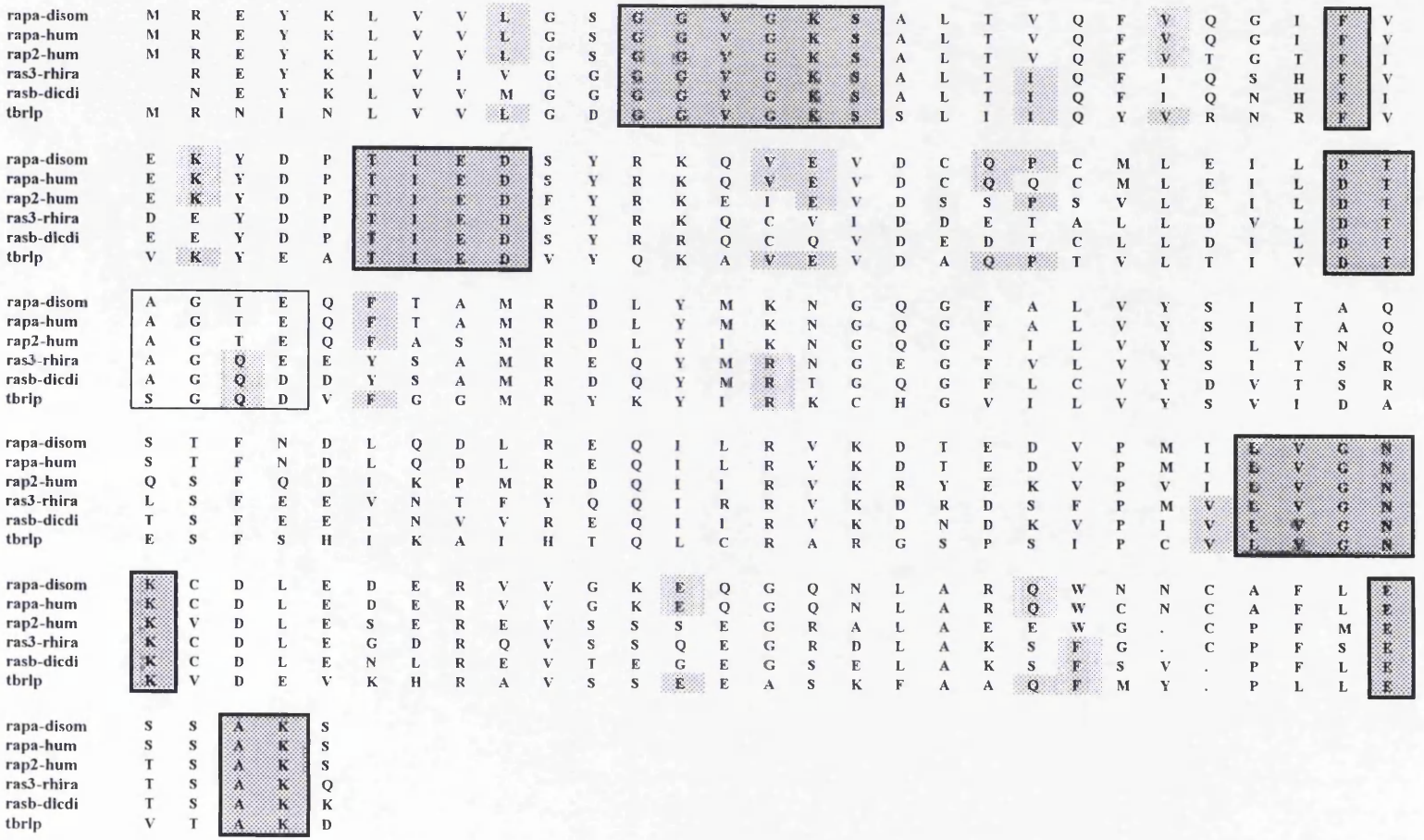


Figure 5.8 is an alignment of the GTP binding domains of *tbrlp* I and some *ras* and *rap* genes from higher eukaryotes. This figure shows the interspersed *ras* and *rap* features throughout *tbrlp* I. The bordered areas show the sequence motifs that have been highly conserved through evolution i.e PM1, 2 & 3 and G1, 2 & 3. The bordered and partially shaded region is the PM 3 binding domain in which glutamine (Q) is conserved in all *ras* subfamily members. The lightly shaded areas show the regions of *tbrlp* I (either blocks of amino acids or single residues) that are identical to either *ras* or *rap*.

5.62 *Recombination as a result of an ancient evolutionary event*

If *tbrlp* I originated as the result of an ancient evolutionary recombination event, the following predictions can be made. First, that all kinetoplastids should possess a *tbrlp* I homologue. This is currently not answerable as the presence of *tbrlp* I has not been examined in the other kinetoplastids. The second, that more "ancient" eukaryotes e.g *Giardia* and *Trichomonas*, should have retained their *ras* and *rap* genes. This cannot be currently answered as *ras* family genes have not yet been examined in these organisms. The third prediction is that the ancestral recombination event will be recognised as a remnant in the sequence. Thus part of the GTP binding domain will have *ras*-like motifs while the other part will have *rap*-like ones. The sequence of *tbrlp* I shows that both *ras* and *rap*-like features are interspersed, making this hypothesis unlikely.

5.63 *Ancestral insertion of a GTP-binding domain*

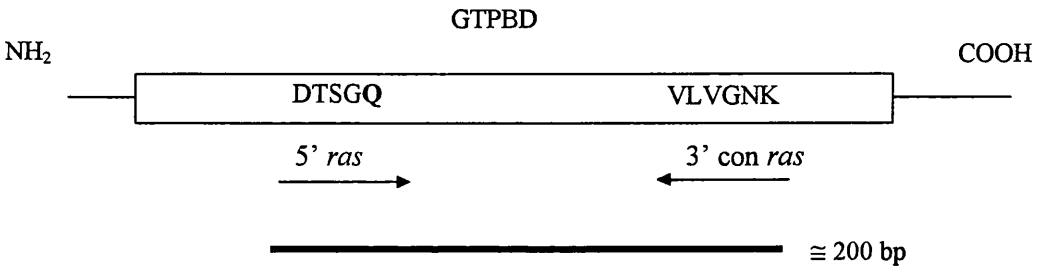
The insertion of a *ras* GTP binding domain into a *rap* gene or vice versa would be detected from the amino acid sequence. The sequence of such a gene would show a *ras* (or *rap*) GTP-binding domain flanked by *ras* (or *rap*) specific sequence. The sequence of *tbrlp* I does not show such features. The mosaic *ras/rap* features are interspersed throughout the sequence of the GTP binding domain, suggesting that *tbrlp* I did not originate from an ancestral insertion of a *ras* GTP binding domain into a *rap* gene, or vice versa.

5.64 *New member of the ras GTPase superfamily*

By implication, for *tbrlp* I to be a new member of the *ras* superfamily in *T. brucei*, *ras* and *rap* homologues would be detectable. Despite the sequencing of 40 PCR products, amplified using conserved GTP binding domain primers (see Chapter 4), no other *ras*-like genes were detected. This suggests that no other *ras* subfamily members were present in *T. brucei*. The possibility exists, however, that the PCR approach may not have detected a *ras* or *rap* homologue due to small sequence changes. To investigate this a pair of primers were designed which would identify and distinguish putative *ras* and *rap* homologues. Both primers were designed against the PM 3 binding site of *tbrlp* I. The *ras* specific primer (named 5' *ras* - GAC ACT TCC GGA CA) contained an amino acid sequence (CA) which could encode the glutamine residue

characteristic of *ras* genes at its 3' end. The *rap* specific primer, (5' *rap* - GAC ACT TTC GGA AC) contained an amino acid sequence which could encode the threonine residue characteristic of *rap* at its 3' end. A third primer (3' con *ras* - TTT ATT ACC GAC AAG AAC) a conserved *ras* primer, was designed using the highly conserved guanine nucleotide binding site, G2 of the *ras* GTPase superfamily GTP binding domain. Two separate sets of PCR amplifications were carried out using 3' con *ras* and 5' *ras* primers in one reaction and 5' *rap* and 3' con *ras* primers in the second reaction. The distance between the PM 3 and G2 domains is fairly constant amongst members of the *ras* GTPase superfamily and the expected band size from the PCR amplification was 200 bp. Figure 5.9 is a diagram showing the primers that were used in the amplification and the expected band size which would be produced from *tbrlp* I, using the *ras* specific primers. In the absence of *rap* genes in *T. brucei*. bands would not be expected from the PCR amplifications with *rap* specific primers.

FIGURE 5.9



The two primers based on the effector domain were the 5'*ras* shown above and 5' *rap*. 5' *rap* primer had almost the same sequence as the *ras* specific primer with the exception of the last amino acid, which was replaced by threonine, the sequence was therefore DTSGT. The sequence of 5'con *ras* was VLVGNK. The expected band size from both amplifications was 200 bp. NH_2 - amino terminus; COOH - carboxyl terminus and GTPBD - GTP binding domain

Twelve sets of amplifications were carried out over an annealing temperature range of 35 - 46° C. Each set was made up of one *ras* specific and one *rap* specific reaction. The results are shown in Figure 5.10. Using the *ras* primer, a 200 bp fragment, representing *tbrlp* I, was observed as expected. The *rap* primer did not produce bands of the expected size showing that a sequence corresponding to the *rap* gene was not present in *T. brucei*. A high molecular weight fragment of approximately 700 bp was observed in a number of tracks when a low annealing temperature was used. The fragment was gel purified and sequenced and a subsequent FASTA search indicated that it was unrelated to the *ras* GTPase superfamily and was probably a product of mispriming. The reactions which produced the 700 bp fragment were all carried out at annealing temperatures of 43° C or below, and therefore, the low annealing temperatures which were used could have encouraged the mispriming reaction. This result suggests that *tbrlp* I is probably not a new GTPase subfamily, due to the lack of evidence for the presence of other *ras* or *rap* genes. However, this cannot be completely ruled out until the entire *T. brucei* genome is sequenced.

5.65 *The evolutionary loss of either ras or rap*

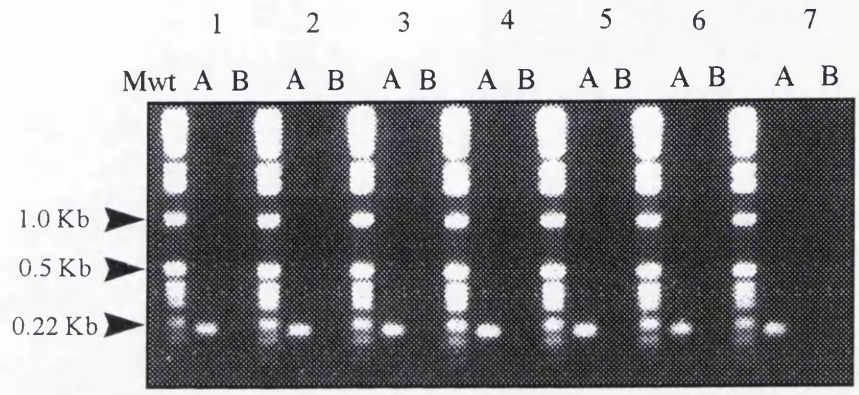
The predictions of this hypothesis are that *tbrlp* I would have the basic structure of either *ras* or *rap* but would have evolved additional domains. The sequence of *tbrlp* I does not show the characteristics of a typical *ras* gene with a Protein kinase A specific phosphorylation site at the carboxyl terminal, nor does it show the features of a typical *rap* gene with *ras*-like carboxyl terminus. Furthermore the sequence of *tbrlp* I clearly shows that the *ras*-like or *rap*-like features are interspersed throughout the GTP binding domain. A number of insertions or deletions would result in the introduction or removal of blocks of nucleotides and not just single amino acids (see Figure 5.0). This hypothesis is therefore unlikely to explain the origins of the mosaic nature of *tbrlp* I.

5.66 *An ancestral ras subfamily gene*

The last hypothesis considered is that *tbrlp* I could be an ancestral *ras* subfamily gene. The predictions from this hypothesis are that the phylogenetic analysis of *tbrlp* I and other related sequences should show that, *tbrlp* I diverges from a point intermediate to *ras* and *rap*, and from the

FIGURE 5.10

A



B

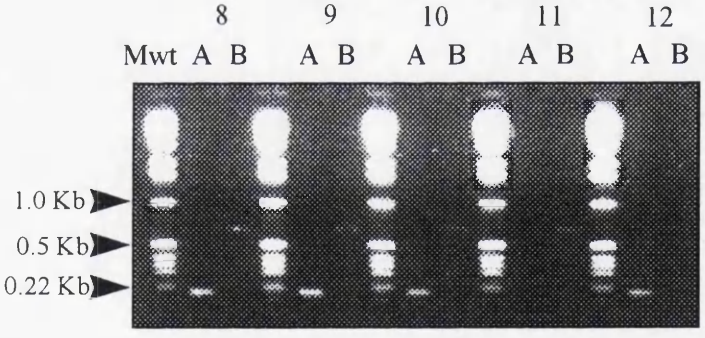


Figure 5.10 is an ethidium bromide stained gel showing the results from PCR reactions carried out to determine the presence or absence of *ras* and *rap* genes from *T. brucei*. Two sets of reactions were carried out at decreasing annealing temperatures. The annealing temperature range was from 46°C - 40°C i.e 1 - 7 (Figure 5.10 A) and 39°C -35°C i.e 8-12 (Figure 5.10 B) . Reactions with the *ras* specific primers - lanes marked A; and *rap* - lanes marked B (for each set of reactions). Mwt - molecular weight marker.

main evolutionary branch to *rab* and *ran* GTPases. The final prediction is that the sequence of the GTP binding domain of *tbrlp* I should be interspersed with both *ras* and *rap* features. As has already been shown, the GTP-binding domain of *tbrlp* I has interspersed features of both *ras* and *rap* GTPase genes. The results from the phylogenetic analysis of *tbrlp* I are presented in the following section.

5.66a *Phylogenetic analysis of tbrlp I*

Ras and *rap* genes, characteristic of those found in higher eukaryotes have been cloned and sequenced from a number of lower eukaryotes such as *Physarum polycephalum*, *Entamoeba histolytica* and *Dictyostelium discoideum*. These *ras* and *rap* genes were used in the phylogenetic analysis of *tbrlp* I. Neither *ras* nor *rap* genes have been cloned from any of the kinetoplastids with the exception of *T. brucei*, i.e *tbrlp* I. The analysis, therefore, could not include *ras* or *rap* genes from the kinetoplastids. As *ras* and *rap* are also found in higher eukaryotes, a number of *ras* and *rap* genes together with a few other *ras* superfamily members from humans were used in a separate analysis of *tbrlp* I. Table 5.2 is a list of the genes and accession numbers used in the phylogenetic analysis of *tbrlp* I.

The GTP-binding domain of *ras* and *rap* genes (and other *ras* superfamily members for the analysis of *tbrlp* I and human *ras* and *rap* GTPases) from each organism and *tbrlp* I were aligned and used in the construction of four bootstrapped, Neighbor-joining, Fitch-Margoliash, UPGMA and Maximum Parsimony trees and four non bootstrapped NJ, FM, UPGMA and MP trees for each of the different groups of organisms used in the analysis (trees were also constructed using alignments of *ras* and *rap* sequences from all the lower eukaryotes used in the analyses and *tbrlp* I). Figures 5.11 - 5.21 show trees constructed by the NJ (non bootstrapped) and bootstrapped MP tree building methods. Trees constructed using the FM, UPGMA and bootstrapped NJ are not shown. The branch lengths from the trees constructed with the FM and UPGMA methods are indicated by non-italics and asterisks respectively at the major points of divergence on the NJ trees. The shape of the trees derived from the non bootstrapped trees were identical to the bootstrapped trees using the FM, UPGMA and NJ methods and are therefore not shown.

Table 5.2

List of the genes and accession numbers used in the phylogenetic analysis of tbrlp I

ORGANISM	GENE	ACCESSION Nº
Entamoeba histolytica	*rap* 1	U01051
Entamoeba histolytica	*rap* 2	U01052
Entamoeba histolytica	*ras* 1	U01053
Entamoeba histolytica	*ras* 2	U01054
Physarum polycephalum	*rap*	U15594
Physarum polycephalum	*ras* 1	P34729
Physarum polycephalum	*ras* 2	P34726
Dictyostelium	*rap* 1	P18613
Dictyostelium	*ras*	M62610
Dictyostelium	*ras* a	P03967
Dictyostelium	*ras* ab	P03967
Dictyostelium	*ras* b	P32252
Dictyostelium	*ras* c	P32253
Dictyostelium	*ras* g	P15064
Dictyostelium	*ras* s	P32254
Human	*rab* 1a	gi131786
Human	*ral*	A34387
Human	*ran*	P17080
Human	*rho* A	P06749
Human	*rho* G	S25722
Human	*ypt* 3	X79780
Human	*ras* h	P01112
Human	*ras* k	P01116
Human	*ras* l	P01118
Human	*ras* n	P01111
Human	*rras*	P10301
Human	*rap* a	P10113
Human	*rap* b	P09526
Human	*rap* 2	P10114
Human	*rap* 3	P17964

Figures 5.11 and 5.12 show the results of the phylogenetic analysis of *tbrlp* I and two *ras* and one *rap* gene from *Physarum polycephalum*. The tree in Figure 5.11 was constructed using the neighbor joining method but other methods generated an identical topology. The numbers on each of the branches are the branch lengths derived from trees constructed with (1) the neighbor joining method (numbers in italics); (2) the Fitch-Margoliash method (non-italics) and (3) the UPGMA method (*) The branch lengths derived from the different methods are almost identical. The topology of the tree suggests that *tbrlp* I diverged at a point which precedes the divergence of *ras* and *rap*. The maximum parsimony tree construction method, produced a tree which shows the same topology as the distance matrix methods. The data were bootstrapped before being used in the construction of the tree. The bootstrap values are shown at the points of divergence on the tree, (Figure 5.12) and are high, thus indicating a high degree of confidence in branch positions is being inferred.

Figures 5.13 and 5.14, are phylogenetic trees derived from using the sequences of the GTP-binding domains of *tbrlp* I and two *rap* and *ras* genes from *Entamoeba histolytica*. The branch lengths from the three distance matrix trees i.e NJ, FM and UPGMA (FM and UPGMA superimposed on NJ), show approximately the same value for all major branches. The topologies of the three distance matrix trees were the same in all cases. Results from each of the trees suggest that the divergence of *tbrlp* I occurred before the formation of the *ras* and *rap* subgroups. Figure 5.14 is a bootstrapped tree, constructed with the maximum parsimony method. The topology of this tree is the same as those of the distance matrix trees. The bootstrap value at the point of divergence of *tbrlp* I into *ras* and *rap* is very high. Once again this emphasises a high degree of confidence in the trees.

Dictyostelium discoideum has one *rap* and six *ras* GTPases. Figures 5.15 and 5.16 show the results from the phylogenetic analysis of all *ras* and *rap* GTPases from *D. discoideum* and *tbrlp* I. The neighbor joining method was used to construct the tree in Figure 5.15 and branch length values derived from the FM - (bold numbers) and UPGMA - (*) methods are shown together with the values from the NJ (italics) method. The branch length values are very similar and the topologies of the trees are consistent using the different tree construction methods. As previously observed, there are two "clusters", one of *ras* and the other of *rap* on either side of *tbrlp* I. This is consistent with *tbrlp* I being an ancestral *ras* subfamily gene from which present day *ras* and *rap* genes diverged. The topology of the tree produced from the maximum parsimony analysis of *Dictyostelium ras* and *rap* is also consistent with *tbrlp* being an ancestral *ras/rap* gene. Even though only one *rap* gene has been cloned from *Dictyostelium*, the analysis shows that it is distinct from the *ras* genes and therefore clusters away from the *ras* genes, on the opposite side of *tbrlp* . Bootstrap values of the maximum parsimony analysis of *tbrlp* I and

Dictyostelium ras and *rap* (Figure 5.16) are high and thus provide confidence in the topology of the tree.

Trees were constructed by combining data from all the three lower eukaryotes, *P. polycephalum*, *E. histolytica* and *D. discoideum*. The results are shown in Figures 5.17 and 5.18. The branch length values from the FM and UPGMA methods are similar to those from the NJ method. The topologies of the trees show a cluster of *ras* and *rap* genes on either side of *tbrlp* I and the different tree construction methods gave the same tree topology. The bootstrap values are lower than the values on trees obtained when trees are constructed with each of *P. polycephalum*, *E. histolytica* and *D. discoideum* separately. This is probably due to the super-imposition of "gene trees" and "species trees" in this diagram.

The analysis of *tbrlp* I and human *ras* and *rap* was carried out with a number of outgroups from the *ras* superfamily. Figures 5.19 - 5.21 show the results of the analysis. The NJ method was used in the construction of the tree (Figure 5.19). The branch lengths from the FM and UPGMA methods of tree construction were similar to those from the NJ method, but are not shown. The topology of the tree is identical using all methods and shows a cluster of *ras* and *rap* genes to either side of *tbrlp* I and it also shows that *tbrlp* I diverged from the main evolutionary branch leading to the divergence of other *ras* superfamily members. Bootstrap values are shown on the maximum parsimony tree (Figure 5.20). Although these values are lower, this may be due to the considerable evolutionary divergence of *tbrlp* I, *rab* and *ran* from the human *ras* subfamily members. Figure 5.21 shows the results from the maximum likelihood analysis of *tbrlp* I, showing high bootstrap values and the same topology as has been consistently observed. The branch lengths from the NJ method are almost identical to those derived from the maximum likelihood method.

The results from the phylogenetic analysis of *tbrlp* I and *ras* and *rap* genes from both lower and higher eukaryotes, suggest that the divergence of *tbrlp* I occurred before the duplication and divergence of the ancestral *ras/rap* gene into *ras* and *rap*. The divergence of the *ran* and *rab* GTPases (and other *ras* GTPase superfamily members) from an ancestral GTP binding domain, occurred before the divergence of *tbrlp* I from the main evolutionary branch. The sequence of *tbrlp* I shows interspersed features of *ras* and *rap* subgroup members. These features taken together, satisfy the predictions of the hypothesis that *tbrlp* I is an ancestral gene of *ras* and *rap*.

Thus the most likely hypothesis which can be used to explain the origins of *tbrlp* I is that *tbrlp* I diverged from an ancestral *ras/rap* gene, prior to its duplication and divergence to the *ras* and *rap* genes of higher eukaryotes.

FIGURE 5.11

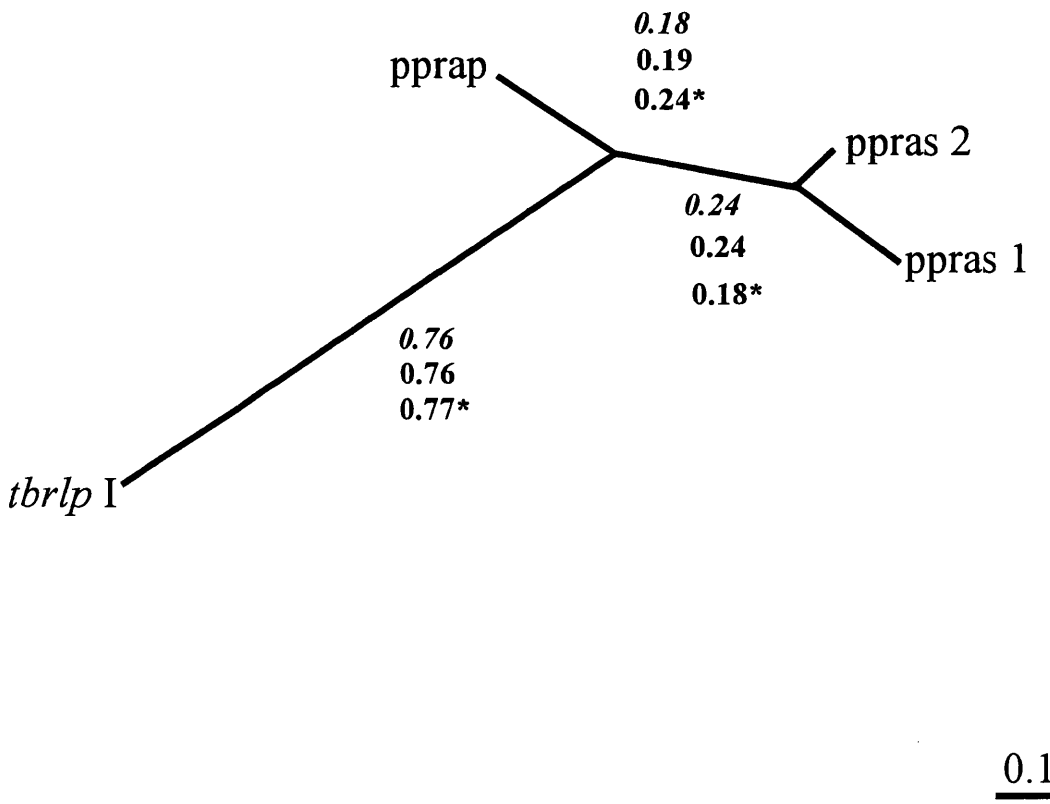


Figure 5.11 shows the results from the phylogenetic analysis of the GTP- binding domains of *tbrlp* I and all *ras* and *rap* subgroup members of *Physarum polycephalum*. The neighbor joining method was used in the construction of the tree. The figures shown are branch lengths derived from the analysis of the data, using three different tree construction methods. Branch lengths derived from the neighbor joining method (in italics); Fitch-Margoliash method (non-italics); and from the UPGMA method (*). ppras 1and 2 - *P. polycephalum ras* 1 and 2; pprap 1 and 2 - *P. polycephalum rap* 1 and 2.

FIGURE 5.12

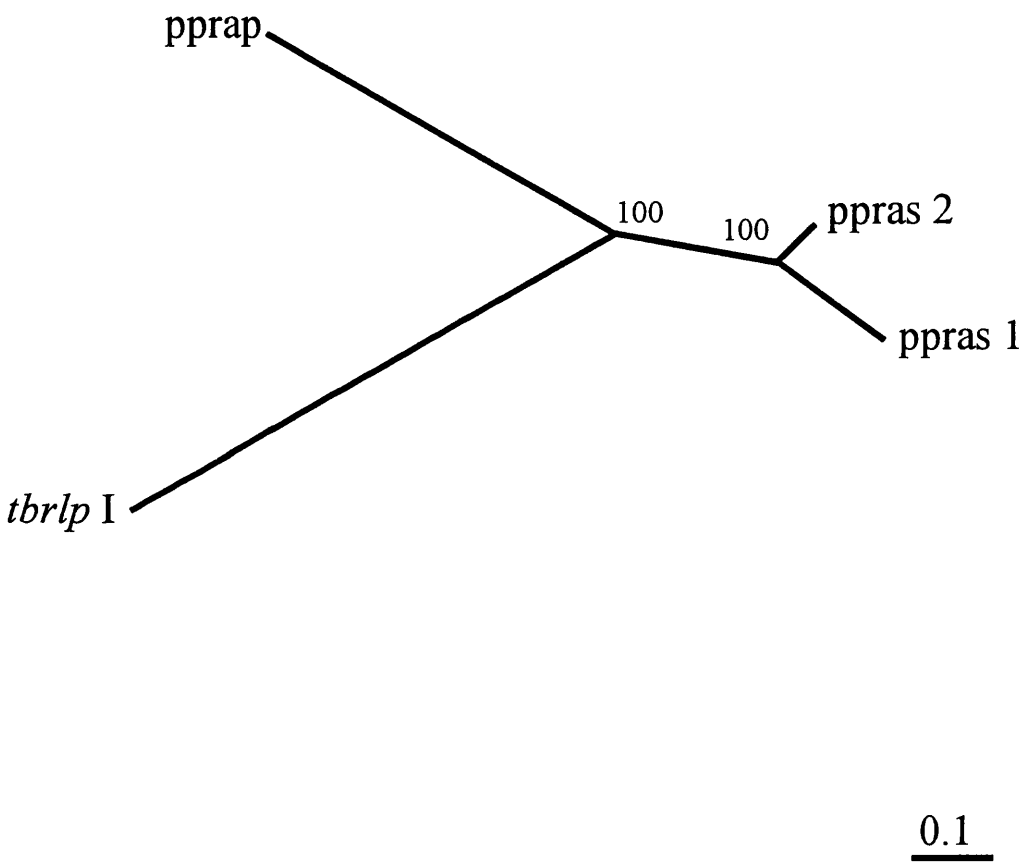


Figure 5.12 shows the results of the bootstrapped maximum parsimony analysis of the GTP binding domains of *tbrlp* I and all members of the *ras* and *rap* subgroups, cloned from *Physarum polycephalum*. The bootstrap values are shown at the major points of divergence. Where ppras and pprap refer to *P. polycephalum ras* and *rap*.

FIGURE 5.13

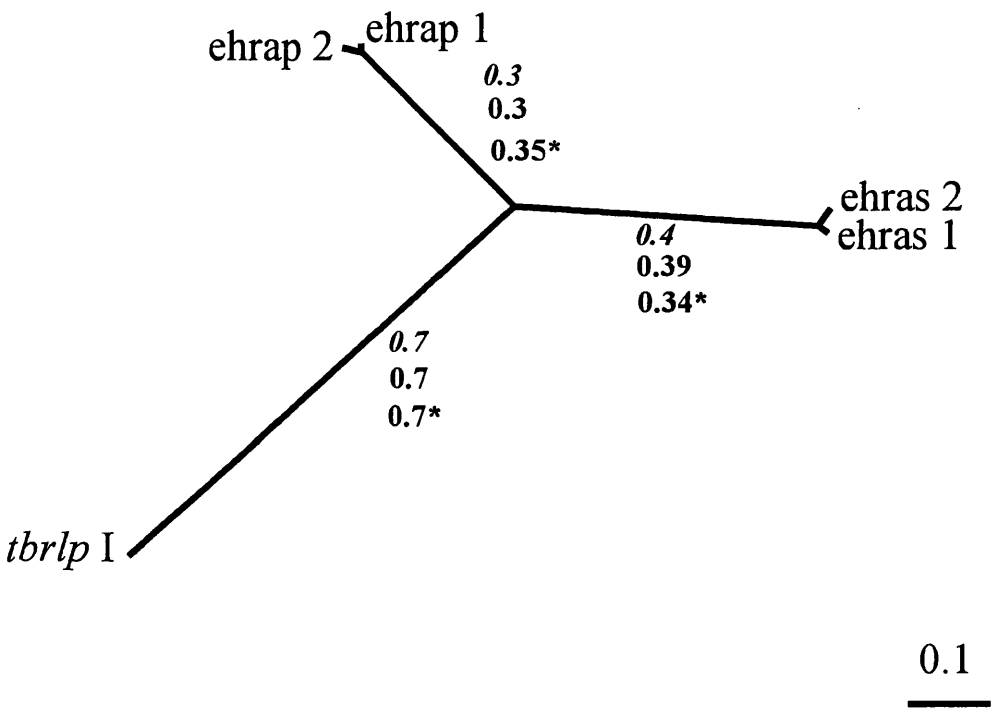


Figure 5.13 shows the results from the phylogenetic analysis of the GTP-binding domains of *tbrlp* I and all members of the *ras* and *rap* subgroups of *Entamoeba histolytica* (ehras and ehrap, respectively). The neighbor-joining method was used in the construction of this tree. The figures indicate branch lengths derived from 1) the neighbor joining method (numbers in italics); 2) the Fitch-Margoliash methods (non-italics); and 3) the UPGMA method (*).

FIGURE 5.14

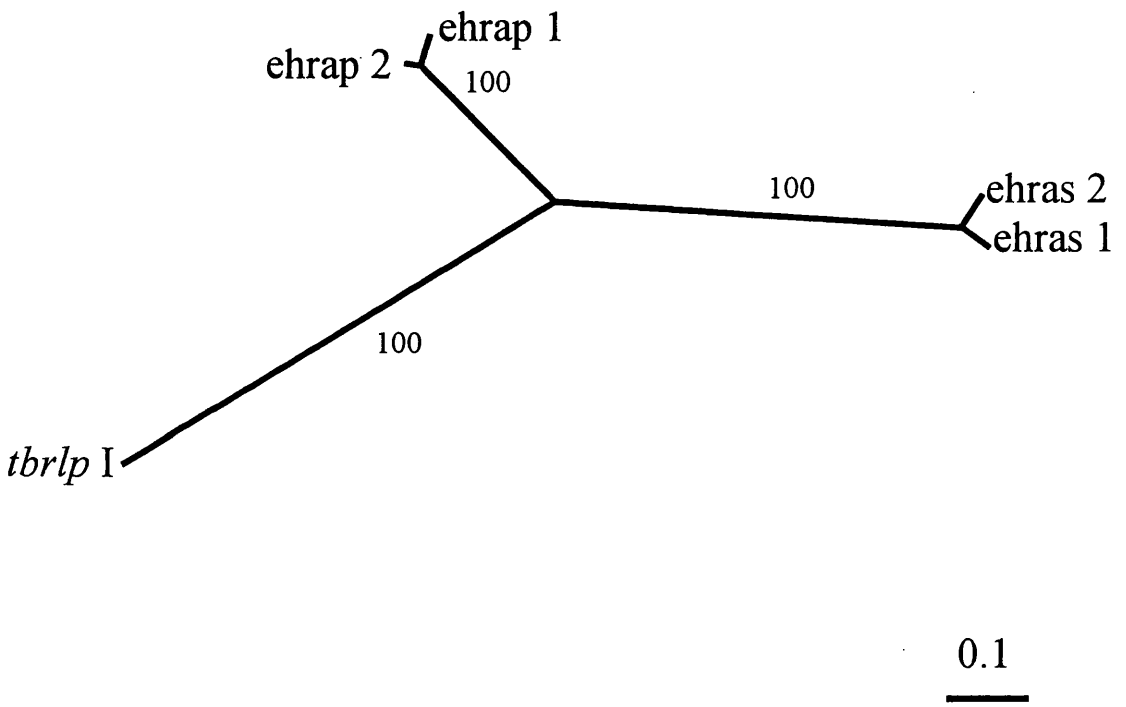


Figure 5.14 shows the results derived from the phylogenetic analysis of the GTP-binding domains of *tbrlp* I and all members of the *ras* and *rap* subgroups of *Entamoeba histolytica* (ehras and ehrap, respectively) using the maximum parsimony method. Data was bootstrapped before the analysis and the bootstrap values are shown at the points of divergence.

FIGURE 5.15

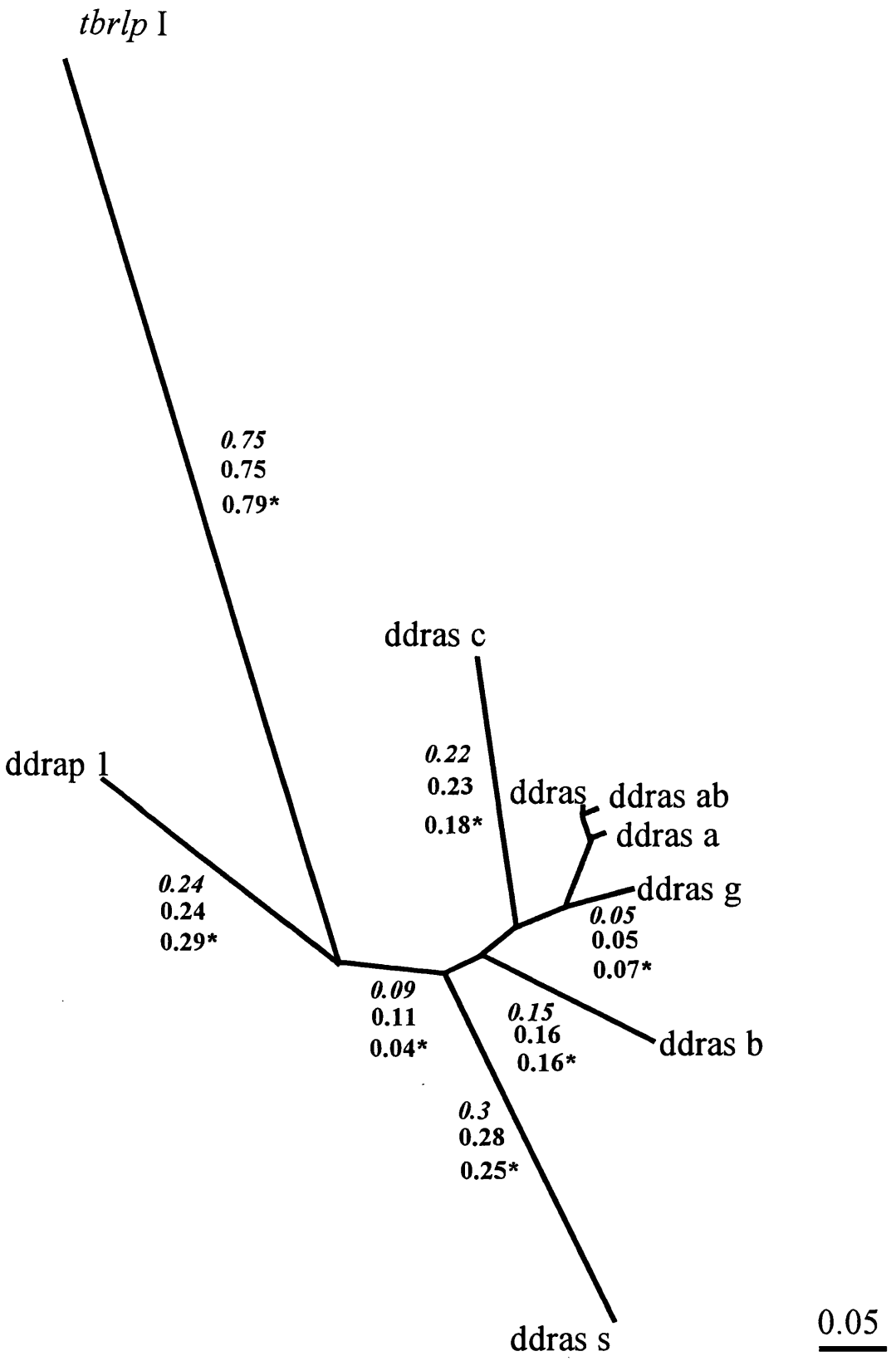


Figure 5.15 shows the results from the phylogenetic analysis of the GTP-binding domains of *tbrlp* I and all members of the *ras* and *rap* subgroups of *Dictyostelium discoideum* (ddras and ddrap refer to *D. discoideum ras* and *rap*). The tree was constructed using the neighbor-joining method. The figures indicate branch lengths 1) from the neighbor joining method (numbers in italics); 2) branch lengths from the Fitch-Margoliash method (non-italics) and 3) branch lengths derived from the tree constructed with the UPGMA method (*).

FIGURE 5.16

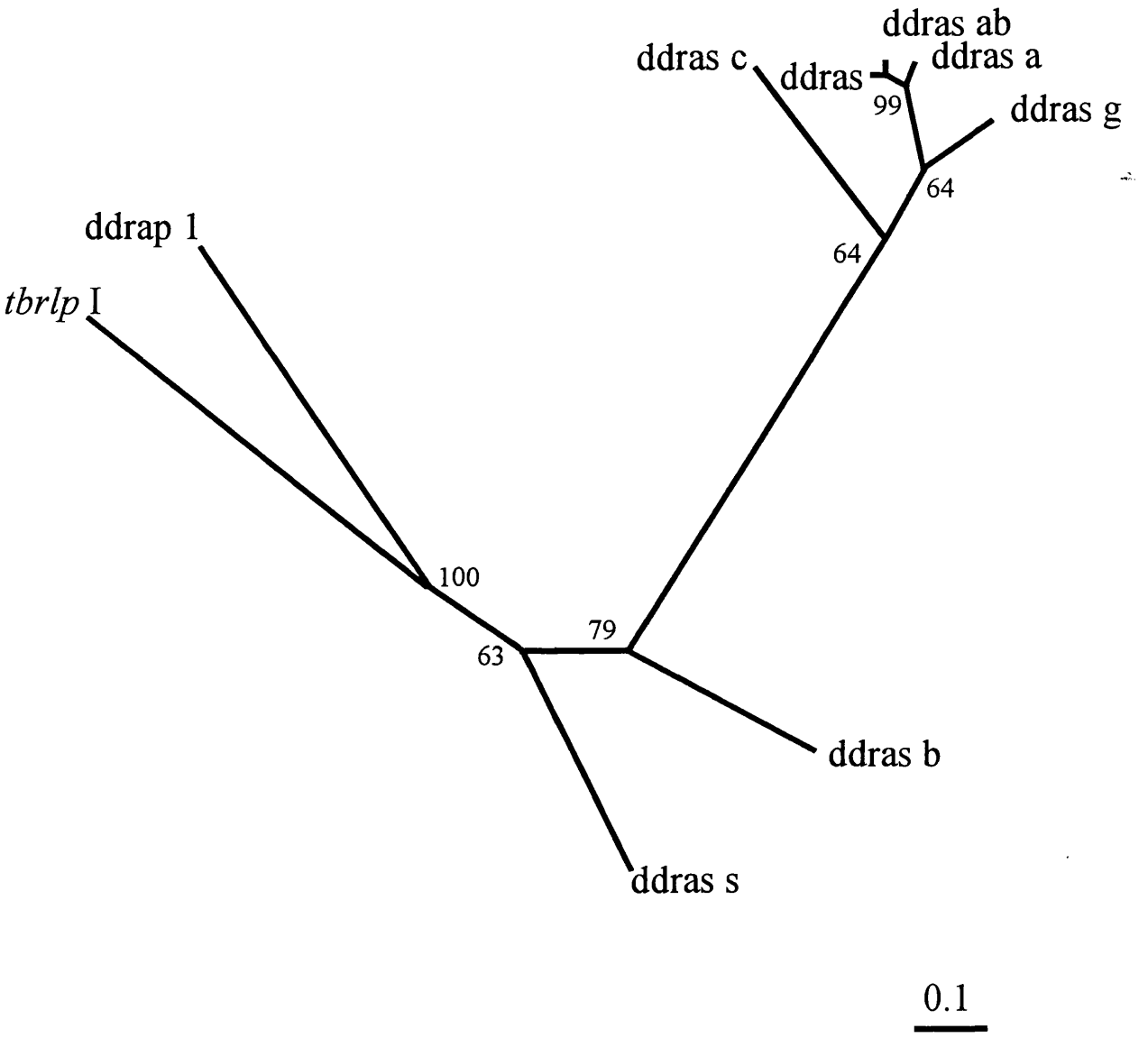


Figure 5.16 shows the results from the phylogenetic analysis of the GTP binding domains of *tbrlp* I and all members of the *ras* and *rap* subgroups of *Dictyostelium discoideum* (ddras and ddrap refer to *D. discoideum ras* and *rap*) using the maximum parsimony method. The figures at the different points of divergence are the bootstrap values.

FIGURE 5.17

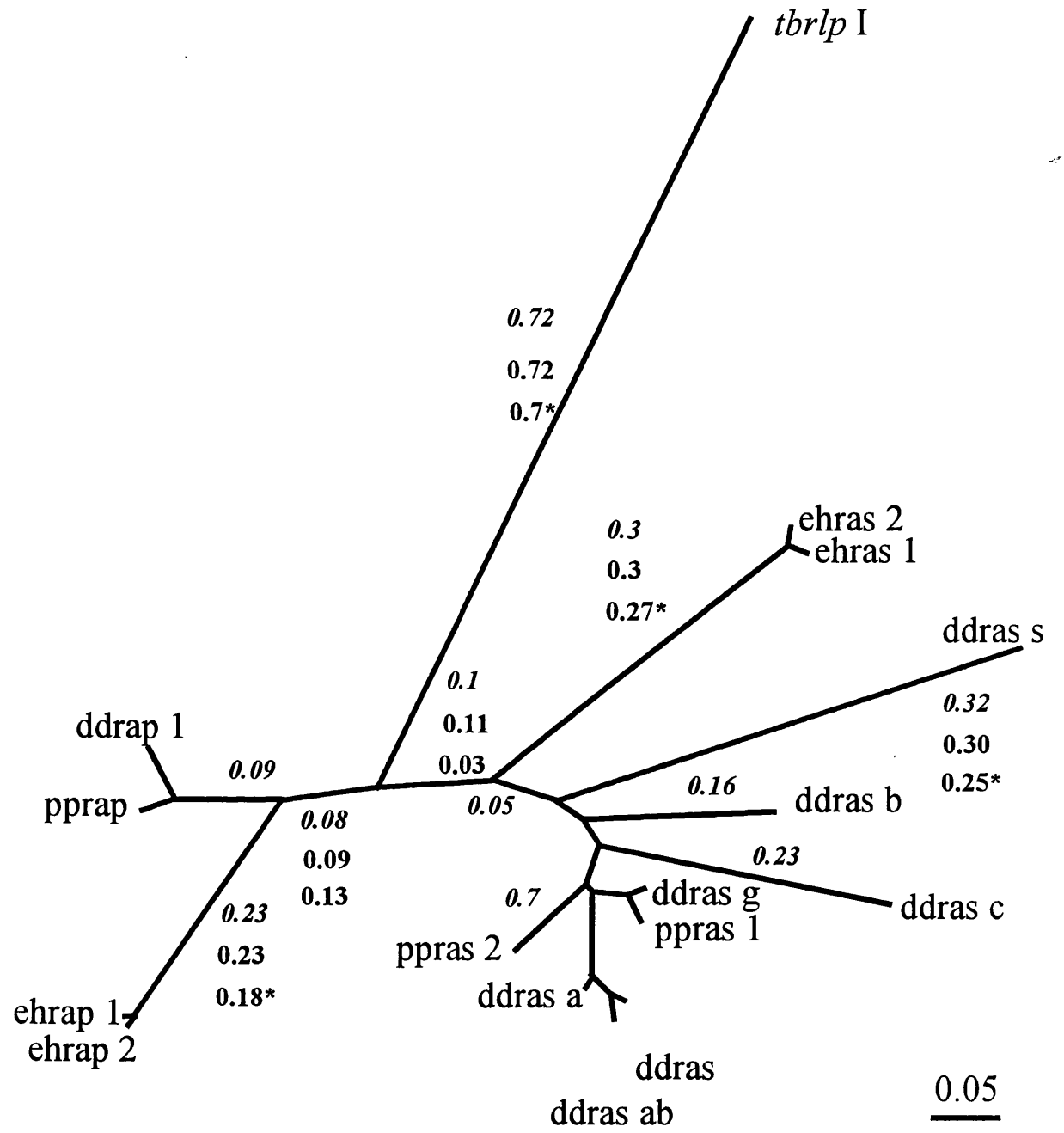


Figure 5.17 shows the results from the analysis of all *ras* and *rap* subgroup GTP binding domains of *tbrlp* I and three lower eukaryotes, *P. polycephalum* (pp), *E. histolytica* (eh) and *D. discoideum* (dd). The branch lengths derived from the use of different tree construction methods are shown on major branches. Branch lengths from the neighbor joining method (numbers in italics); Fitch-Margoliash method (non-italics); and branch lengths from the UPGMA method (*).

FIGURE 5.18

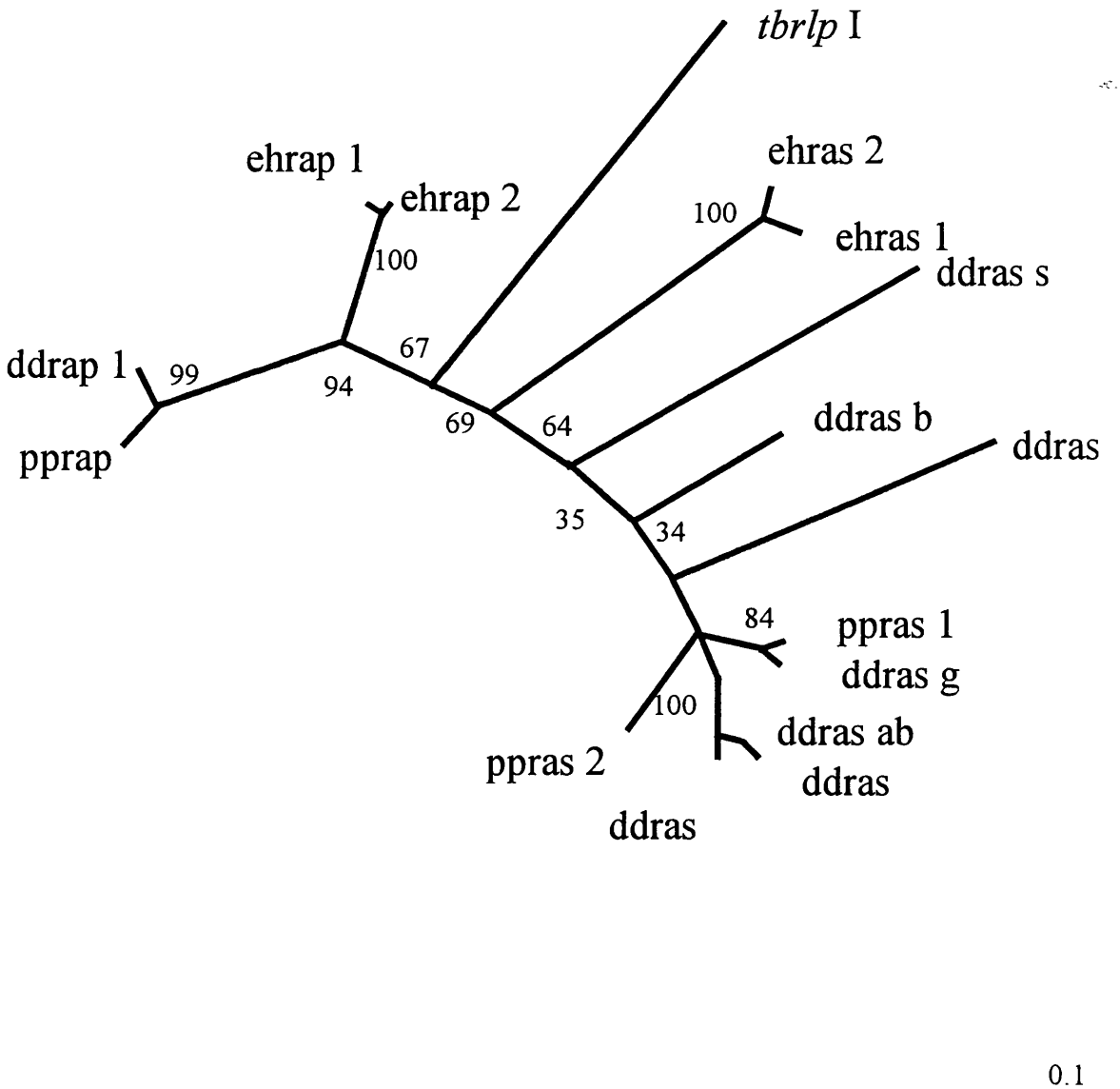


Figure 5.18 shows the results from the maximum parsimony analysis of the GTP binding domains of *tbrlp* I and all *ras* and *rap* subgroup members from three lower eukaryotes, namely, *P. polycephalum* (pp), *E. histolytica* (eh) and *D. discoideum* (dd). Figures at the major points of divergence are the bootstrap values derived from the analysis.

FIGURE 5.19

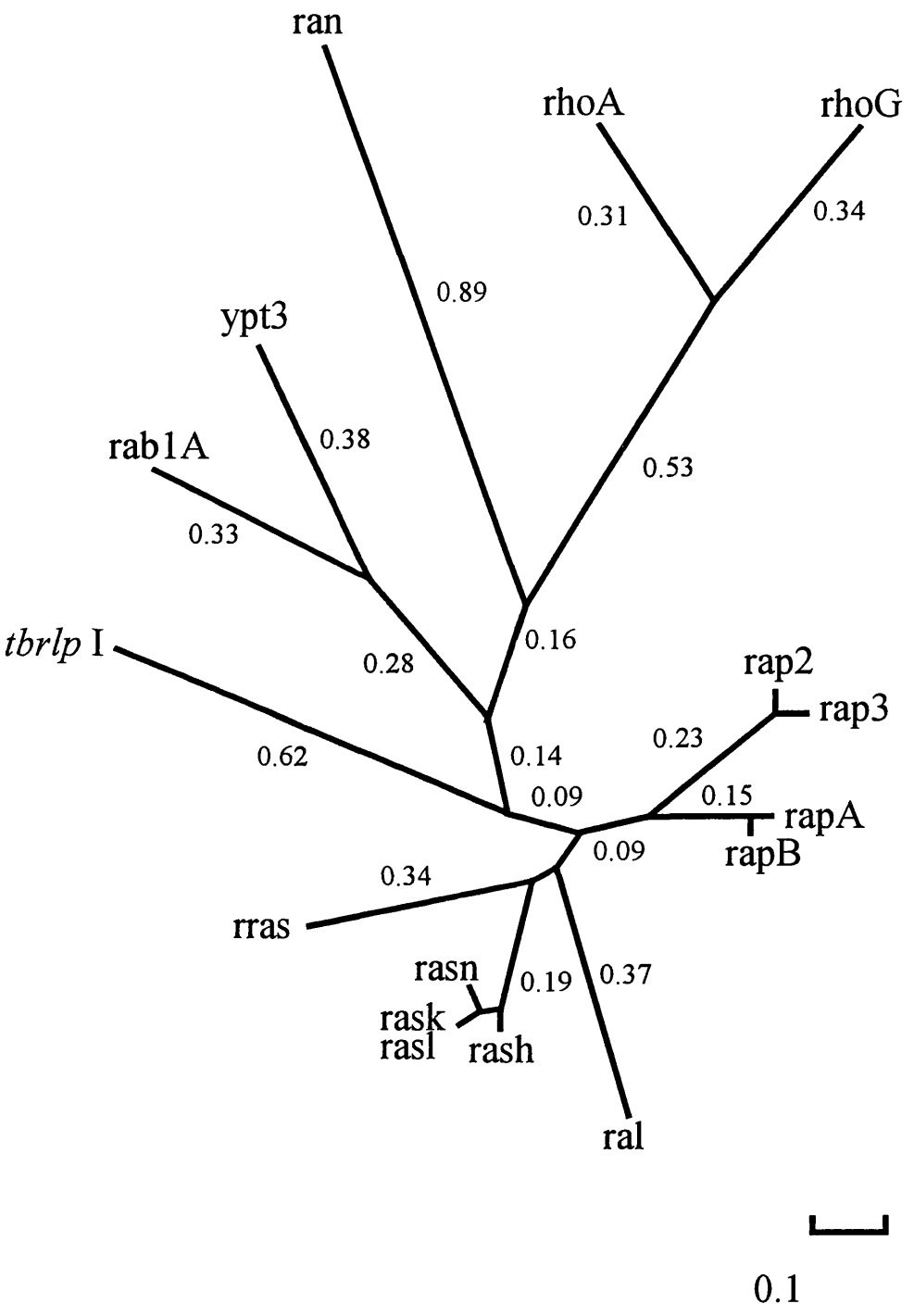


Figure 5.19 show the results derived from the neighbor joining method of tree construction. The GTP binding domains of *tbrlp* I and 15 members of the human *ras* superfamily (10 of which belong to the *ras* GTPase subfamily), were used in the analysis. Branch lengths preceding major points of divergence are shown.

FIGURE 5.20

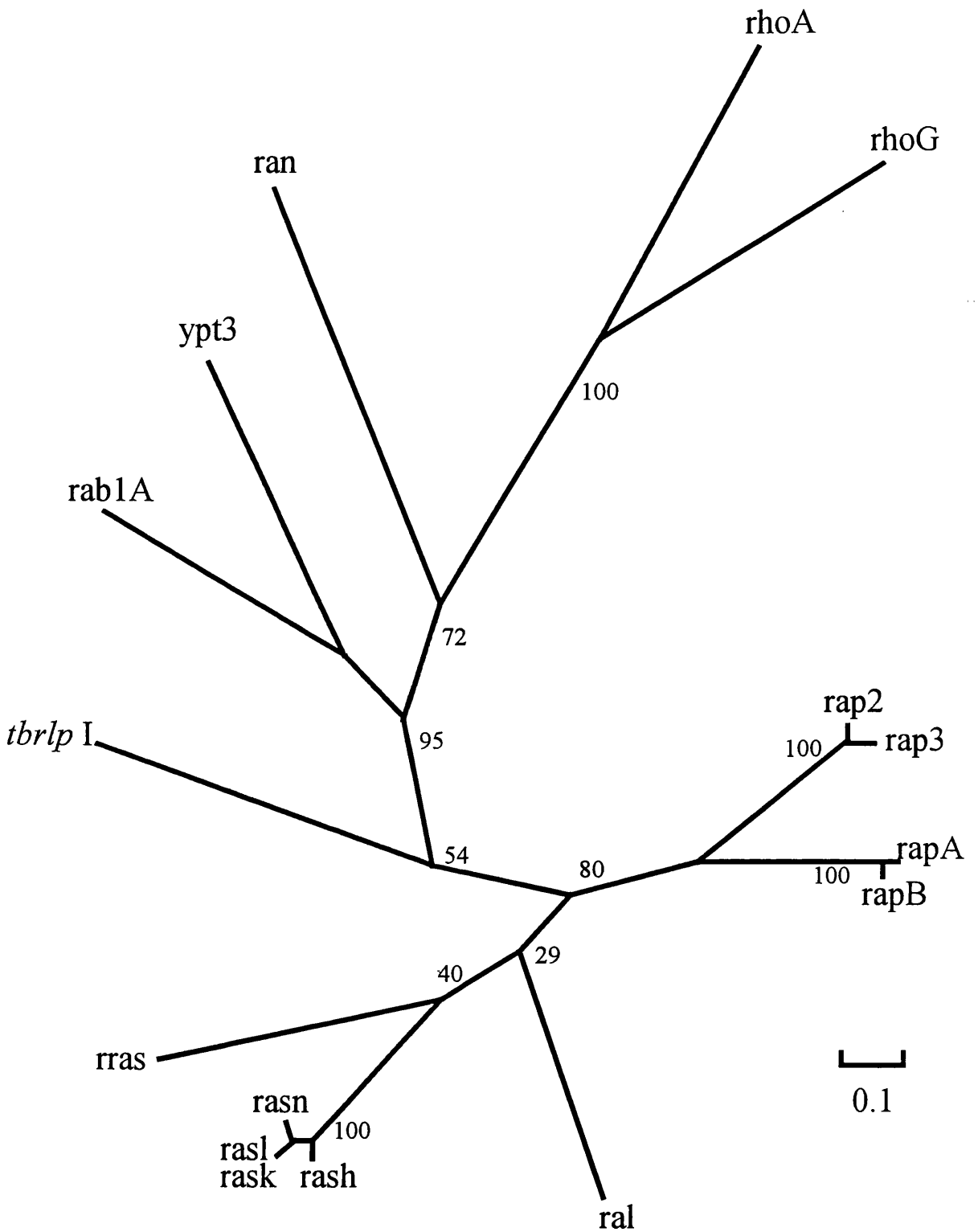


Figure 5.20 shows the results from the maximum parsimony analysis of the GTP binding domains of *tbrlp* I and 15 members of the human *ras* GTPase superfamily (10 of which belong to the *ras* GTPase subfamily). Bootstrap values are shown at the major points of divergence on the tree.

FIGURE 5.21

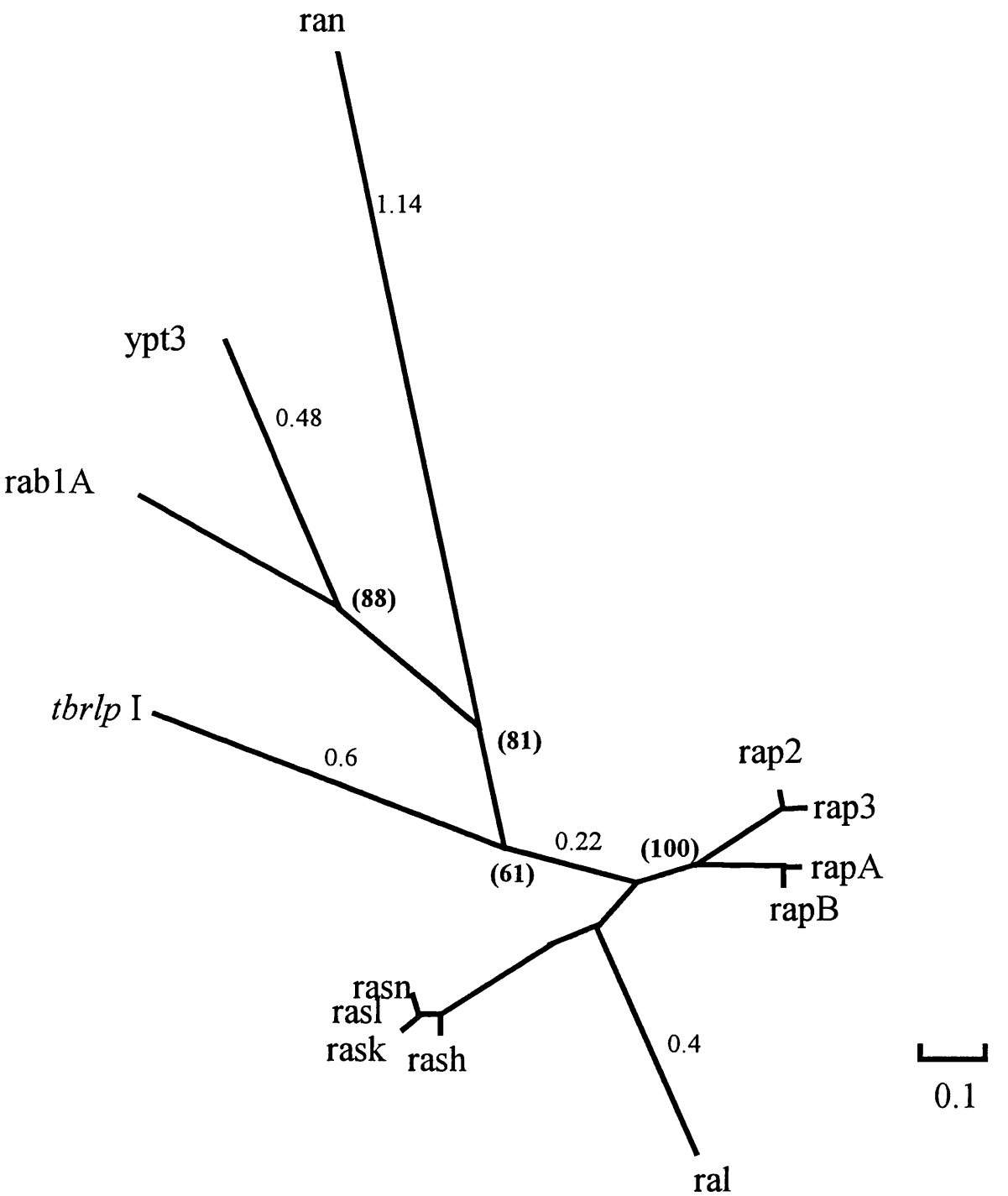


Figure 5.21 shows results from the maximum likelihood analysis of the GTP binding domains of *tbrlp* I and 12 members of the human *ras* GTPase superfamily (9 of which belong to the *ras* GTPase subfamily). Bootstrap values (in brackets) are shown at the major points of divergence of the tree. Figures for the major branch lengths are also shown.

5.7 Discussion

Six hypotheses were put forward to explain the possible evolutionary origins of *tbrlp* I. Results from the tests on the different hypotheses suggested that the data support only one hypothesis for the evolutionary origin of *tbrlp* I, namely that, it is an ancestral *ras* subfamily gene. Results from the phylogenetic analysis of *tbrlp* I and *ras* and *rap* GTPases from four eukaryotes, *P. polycephalum*, *E. histolytica*, *D. discoideum* and humans, confirm that *tbrlp* I is related to an ancestral *ras/rap* gene. The branch lengths obtained for all sets of trees used in the analysis i.e, FM, UPGMA, NJ and ML were almost identical in all cases. The bootstrap values on these trees were very high, especially in the case of *E. histolytica* and *P. polycephalum*. All three predictions of the results from testing this hypothesis were upheld. In all cases, *tbrlp* I joins the tree at a node which designates the divergence point of *ras* and *rap*. This suggests that *tbrlp* I diverged from the ancestral *ras/rap* gene. The trees from the analysis of members of the human *ras* superfamily and *tbrlp* I showed that *tbrlp* I diverged from the main evolutionary branch to *ran* and *rab* GTPase genes (and other *ras* GTPase superfamily members). The sequence of the GTP binding domain of *tbrlp* I is also consistent with that of an ancestral *ras/rap* GTP binding domain. It has features of *ras* and *rap* interspersed throughout the sequence. In addition to this unique feature, it also retained all of the sequence motifs of a GTP binding domain, which are necessary for its classification as a *ras* GTPase superfamily member and its function as a GTPase protein.

Results from the evolutionary analyses of *tbrlp* I, therefore suggest that after the duplication of a primeval GTP binding domain, the genes diverged into present day *rab*, *ran* and *rho* GTPases. Subsequent to that, further duplication and divergence events resulted in the generation of *ras* and *rap* from a *tbrlp* I like ancestor. The uptake of endosymbionts (which resulted in life forms with mitochondria) possibly also occurred at this point in time, as most universal trees show that the kinetoplastids were one of the first organisms to diverge after the acquisition of mitochondria (based on the rRNA tree of life, Fernandes *et al*, 1993; Schlegel, 1994; Vickerman, 1994; Maslov and Simpson, 1995). *T. brucei* probably evolved from the main lineage before the divergence of the ancestral *ras/rap* gene into *ras* and *rap*. This would account for the absence of *ras* and *rap* GTPase genes in *T. brucei*, as they had not diverged at that time, and based on this we would predict that they would also be absent from organisms more ancient then *T. brucei*. The higher eukaryotes from the Physarales upwards, therefore must have evolved after the divergence of the ancestral *ras/rap* gene into the current *ras* and *rap* genes of higher eukaryotes. Both *ran* and *rab* GTPases genes have been isolated from *T. brucei* suggesting that both *ran* and *rab* GTPase genes had already diverged from a primeval GTP binding domain at

the time *T. brucei* evolved. If this hypothesis is correct, *tbrlp* I homologues should be present in all of the kinetoplastids and organisms more ancient than the kinetoplastids, i.e *Giardia*, and the *Trichomonids*. Based on this hypothesis, one could conclude that *ras* and *rap* genes are also absent from other kinetoplastids, however, if *ras* and *rap* genes are isolated from these organisms in the future, it would imply that the hypothesis explaining the origins of *tbrlp* is invalid.

There is reason to believe that *tbrlp* I diverged from the ancestral *ras/rap* gene, but, its function still remains unknown. It would be interesting to determine the role of *tbrlp* I in the development of *T. brucei*. Both *ras* and *rap* are known to be involved in growth regulation and differentiation of cells in other eukaryotes (Barbacid, 1987). There is also documented evidence of the antagonistic relationship between human *ras* and *rap* GTPase proteins (Lerosy, 1991; Nassar *et al*, 1995; Sprang, 1995). It would be interesting to find out whether *tbrlp* I has an autoregulatory function (due to the presence of the cAMP dependent protein kinase A phosphorylation site at the carboxyl terminus) which results in the initiation or termination of *ras* induced reactions and the physiological processes of *T. brucei* that it might affect. The sequence of the PM 3 binding site, suggests that *tbrlp* I might interact with downstream effectors in the same manner as a "true" *ras* subfamily member. However, if subsequent biochemical analysis of the expressed protein, show that it is phosphorylated solely by Protein kinase A, it would suggest that *tbrlp* I is activated in a *rap*-like manner. The X of the CAAX box of *tbrlp* I is Leucine (L). *Rap* GTPases normally have a leucine residue at this position and the action of *rap* GTPases is dependent on the translocation of the protein from a membrane to the cytosol. The presence of leucine in the same position in *tbrlp* I suggests that its function is also dependent on the translocation of the protein from a membrane to the cytosol. To gain a greater understanding of the evolution of *tbrlp* I and its function, *tbrlp* I was expressed as a fusion protein and, a conjugated peptide was synthesised in order to raise *tbrlp* I specific anti peptide serum. These studies are described in Chapter 6.

Chapter 6

Development of tools for functional analyses of *tbrlp* I

CHAPTER 6

DEVELOPMENT OF TOOLS FOR THE FUNCTIONAL ANALYSES OF *tbrlp* I

6.0 Introduction

A gene, *tbrlp* I, with the characteristics of two members of the *ras* subfamily (*ras* and *rap*), was cloned and sequenced from *T. brucei*. Results from the evolutionary analysis of *tbrlp* I, suggest that it is an ancestral *ras/rap* gene. Studies of the functions of *ras* and *rap* subgroup members in higher eukaryotes, have shown that the proteins are involved in cellular growth and differentiation. *Ras* and *rap* GTPases are known to have an antagonistic relationship, i.e *ras* GTPase activity is terminated when *rap* is activated and vice versa. An example of this relationship is the ability of a mammalian *rap* GTPase to suppress the phenotypic transformation of cultured cells induced by K-*ras* (Kitayama *et al*, 1989). The sequence of *tbrlp* I shows that it is definitely a *ras* subfamily member which suggests that it could be involved in cellular growth regulation. However, due to its mixed features, it is not clear whether its function is predominantly *ras*-like, *rap*-like or a combination of both.

The aim of this section of the project was to develop tools which can be used for functional analysis of *tbrlp* I. Two approaches were taken: the first approach was the expression of *tbrlp* I as a fusion protein and the second approach was, the production of *tbrlp* I - specific anti -serum. These approaches would provide reagents which could be used for subcellular localisation and functional analysis of the native protein.

6.1 Methods available for the production of fusion proteins

A number of systems were considered for the expression of *tbrlp* I as a fusion protein. It was decided to try three of the most commonly used methods, namely, the Glutathione -S-transferase (GST) gene fusion system, the maltose binding protein fusion protein system, and the Qiaexpressionist (histidine tag) system. These systems are based on the integration of a cloned gene into an expression vector and the subsequent expression of the gene as a fusion protein. Each system provides a unique domain which forms part of the fusion protein. In the GST system, the gene to be expressed is fused with *Schistosoma japonicum* GST while in the case of

the maltose binding protein system the gene is fused to the maltose binding protein of *E. coli*, which is encoded by the *malE* gene. The Qiaexpressionist system allows the expression of proteins that are fused to six histidine residues provided in the vector construct. Each system has a specific affinity method for the purification of the expressed proteins and the fusion protein can be cleaved at a factor Xa or thrombin cleavage site using the appropriate protease, thus allowing the preparation of pure recombinant protein. In order to maximise the probability of expressing a soluble fusion protein, both the full length *tbrlp* I gene and a truncated portion lacking the C-terminus were inserted into all three expression vector systems.

6.11 *Development of a PCR method for rapid checking of the orientation of inserts*

To facilitate the analysis of transformants, a more cost effective and less time consuming method than DNA sequencing was developed for the screening of colonies by direct PCR. The colonies were lysed and the DNA denatured in a microwave for two minutes (Kilger and Schmid, 1994) followed by resuspension in an adequate amount of 11.1x PCR mix (refer Materials and Methods) and the gene specific primers required for the amplification of the insert. Using the vector forward primer and a gene specific reverse primer, positive amplification of a specific fragment indicated the presence of the correct insert in the correct orientation (see Figure 6.0). As a negative control, vector forward and gene specific forward primers were used (i.e both facing the 5' $\rightarrow$ 3' direction) so that the presence of a PCR product would indicate that the insert was in the wrong orientation. Using this method, many colonies could be screened simultaneously and results were normally available at least three hours from the beginning of the procedure. Table 6.0 provides a list of all the gene specific primers used in the expression analysis of *tbrlp* I as a fusion protein, some of which were also used in determining the orientation of the insert. Figure 6.0 provides a diagram showing the results expected from the use of gene and vector specific primers in determining the orientation of *tbrlp* I within the vector.

Table 6.0

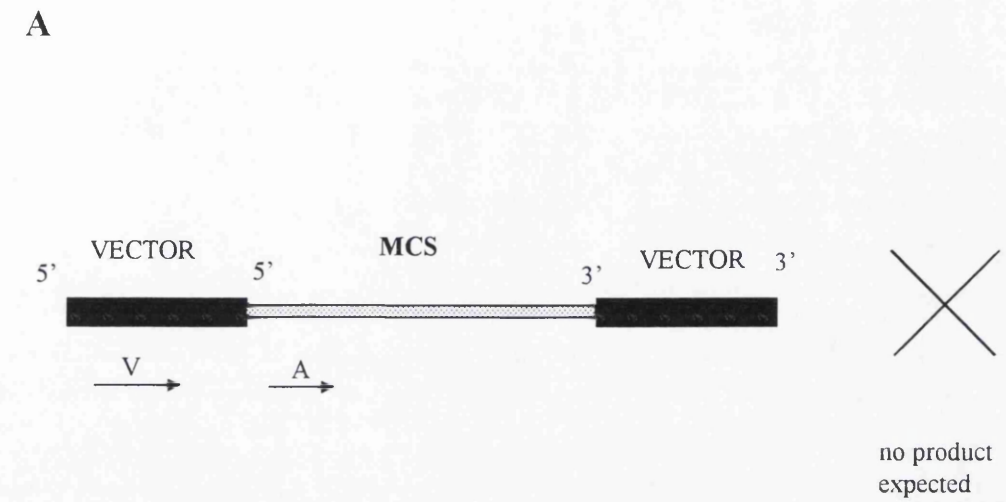
List of gene specific primers used in the expression of *tbrlp* I as a fusion protein.

Gene specific primers	Sequence
5' tbrlp	ATG AGG AAT ATT AAC CTC GTC GTT TTG GGT
5' tbbam	CGC GGA TCC AGG AAT ATT AAC CTC GTC
5' Ecotb	CCG GAA TTC ATG AGG AAT ATT AAC CTC GTC
3' utbrlp	TCA GAG CAT AGT GCA TCC CGA TTT CTT TTT CCT A
3' tbrlp	TTG CTT GCC TCT TCA CTA CTC ACC GCA CGG
3' tbsph	ACA TGC ATG CTC AGA GCA TAG TGC ATC CCG
3'Xbatb	CTA GTC TAG ATC AGA GCA TAG TGC ATC CCG

Vector specific primers used:

1) 5' pGEX - GGG CTG GCA AGC CAC GTT TGG TG
2) M13/pUC sequencing primer - CGC CAG GGT TTT CCC AGT CAC GAC
3) malE (5' pMAL-c2 primer) - GGT CGT CAG CAT GTC GAT GAA GCC
4) T7 (3' end of bluescript) - CGG GAT ATC ACT CAG CAT AAT G
5) T3 (5' end of bluescript) - AAT TAA CCC TCA CTA AAG GG

FIGURE 6.0



An illustration of the results expected from a PCR method which was developed to determine the orientation of the insert in the multiple cloning site (MCS) of the vector. Primers used in determining the orientation of the insert were v, vector specific i.e, either pGEX, pMAL or pQE specific, A (any of the 5' gene specific primers, refer Table 6.0) and B (any of the gene specific primers, refer Table 6.0). Two reactions were performed for each colony that was picked. If the insert was in the wrong orientation, the reactions carried out with the situation illustrated in B would not produce a fragment of the expected band size.

6.2 The expression of *tbrlp I* as a Glutathione-S-Transferase fusion protein

Two types of inserts were used to make GST fusion proteins containing *tbrlp* I. The first was a truncated version of *tbrlp* I (referred to as t-*tbrlp* I), containing the amino terminal, and is truncated at a point which is 283 bp 5' of the first stop codon. The truncated form therefore lacks half of the carboxyl terminal extension region. The second insert was a full length copy of *tbrlp* I. Figure 6.1 shows an illustration of the PCR products (i.e the truncated and full length forms of *tbrlp*) and the gene specific primers used in their amplification. The PCR products were non-directionally ligated into the *Eco* RI site of pGEX-5X-3 (as described in the manufacturers manual, pGEX-5X-1 and 2 were also used to ensure that the insert was cloned into the right vector). Transformation with t-tbrlp I was successful and 8 colonies were picked. DNA was extracted from overnight cultures, digested for 1 hour with *Eco* RI and the DNA separated on a 1.5 % agarose gel. Figure 6.2 shows that 6 out of 8 clones (lanes 2-7) possessed inserts of the expected size (i.e 398 bp). Wizard minipreps of the 6 clones were prepared and each one was sequenced to ensure that they were in the right frame. However, the results showed that all 6 inserts were in the wrong orientation.

FIGURE 6.1

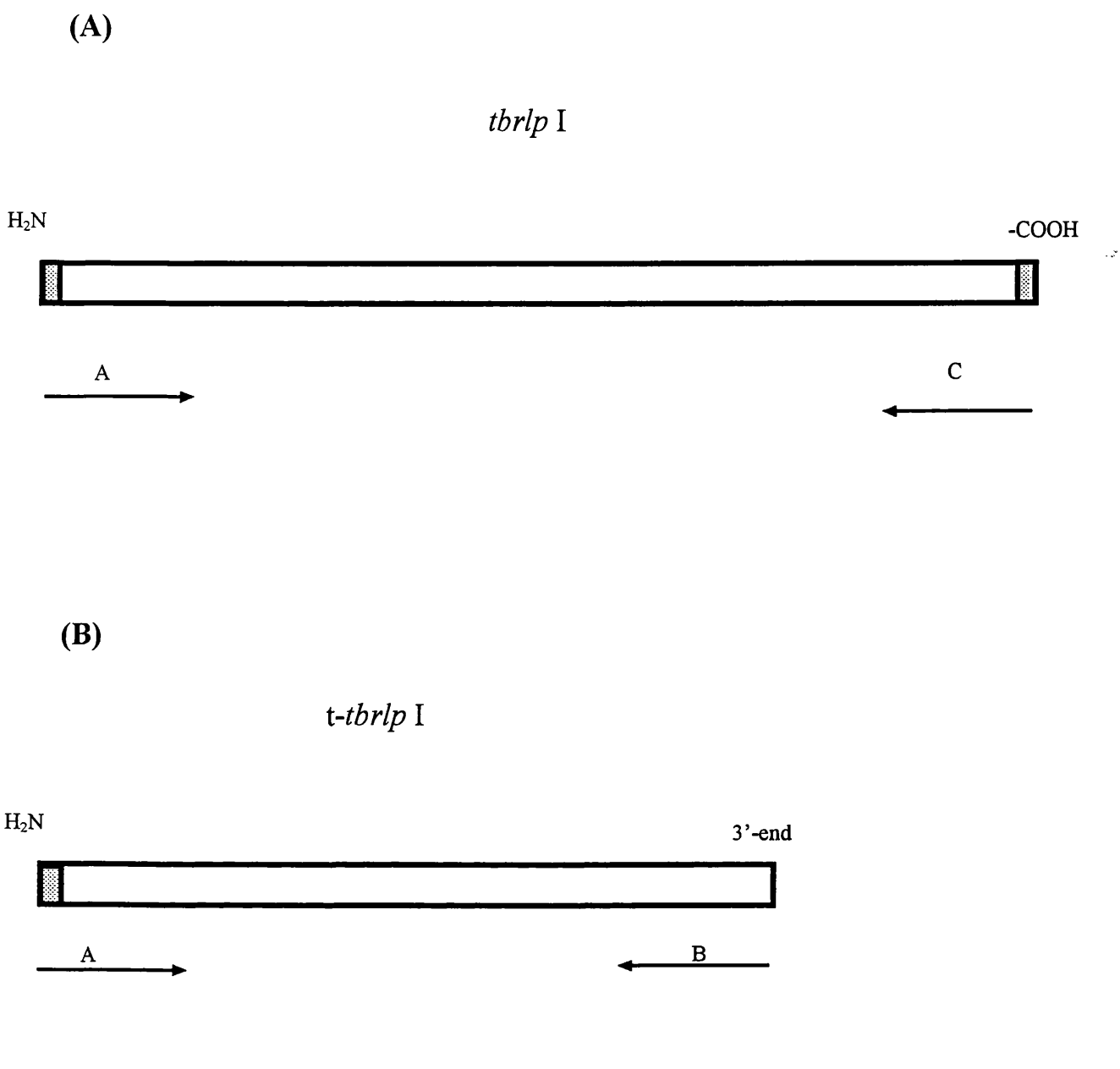


Figure 6.1 is an illustration of the gene specific primers (A, B & C) used in the amplification of *tbrlp* I, a 681 bp PCR product and t-*tbrlp* I, a 398 bp product. A -5' tbrlp I; B - 3' tbrlp I; C - 3'utbrlp I (refer section 6.11, for primer sequences). NH_2 - amino terminus; -COOH - carboxyl terminus; 3'-end - i.e the 3' end of *tbrlp* I minus the carboxyl terminus. The PCR product illustrated in (A) is *tbrlp* I and that in (B) is t-*tbrlp* I (i.e truncated *tbrlp* I).

IGURE 6.2

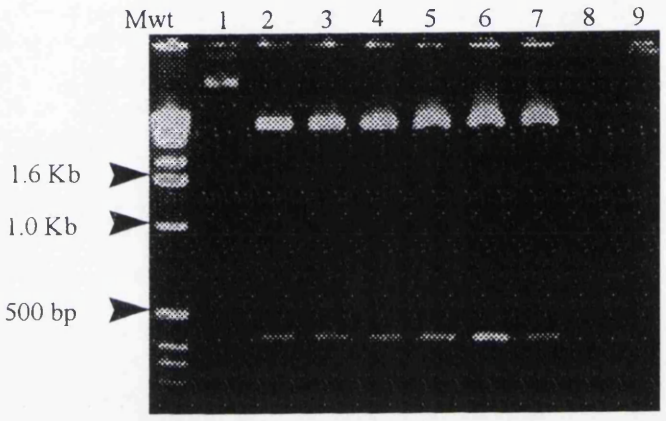


Figure 6.2 shows a 1.5 % ethidium bromide stained agarose gel of 8 *Eco*R I digested clones, 6 of which had the expected insert size, i.e 398 bp (lanes 2-7). Mwt - 1 Kb ladder.

The procedure was repeated, using non-directional cloning with the full length *tbrlp* I insert. Figure 6.3 shows a photograph of *Eco* RI digests of DNA (separated on 1.2 % agarose gels) from twelve transformants produced from the ligation of *tbrlp* I into pGEX-5X-2. Ten of the transformants (lanes 1 & 2, 4 & 5, and 7-12) had inserts of the expected band size (i.e 681 bp). However, sequencing showed that all 10 transformants had inserts in the wrong orientation.

To investigate why the cloning process was only generating clones in the wrong orientation, the ligated mixture of DNA was tested to ensure that it contained inserts ligated into the vector in the correct orientation. Aliquots of the ligation mix were amplified using the polymerase chain reaction method with primer combinations from vector and gene specific sequences which allowed the distinction between inserts in either orientation. Figure 6.4 is a schematic diagram illustrating the PCR strategy used to determine the orientation of inserts in the ligation mix and the expected band sizes. Briefly, the strategy entailed using 3 different primer combinations: (1) A vector forward primer (V) and a gene specific forward primer (A) which would yield no product if the insert is in the correct orientation. If on the other hand the insert is in the wrong orientation, a 700 bp product would be expected. (2) Vector forward primer (V) and a gene specific reverse primer (B). The expected size of the product with this combination is 681 bp if the insert is in the correct orientation; and (3) A vector forward primer and a gene specific reverse primer (C). The expected product size from this combination is 431 bp, if the insert is in the correct orientation.

FIGURE 6.3

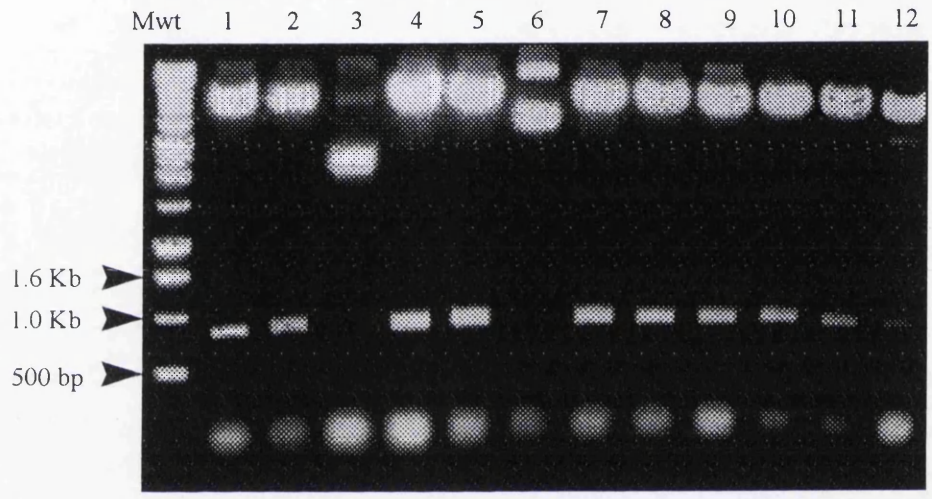
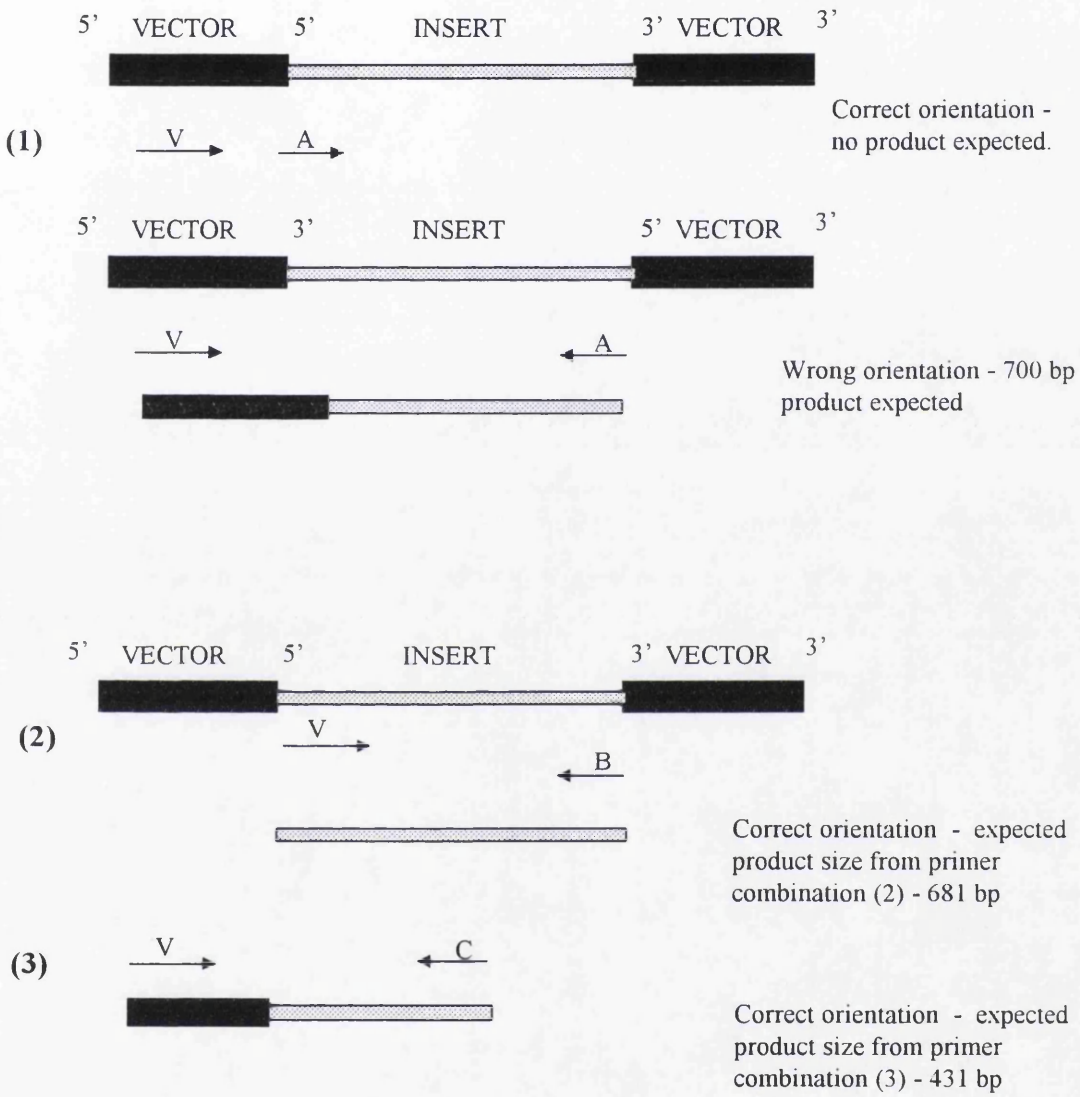


Fig 6.3 is an ethidium bromide stained gel showing the results of *Eco* RI digests of12 transformants. 10 of the transformant, in lanes 1 & 2, 4 & 5 and 7-12 contained the expected insert size i.e, approximately 700 bp . Mwt - 1 Kb ladder.

FIGURE 6.4



Three primer combinations were used in the PCR amplifications (1, 2 & 3). No product was expected from the first amplification with primer combination (1) if the insert was in the correct orientation. A band of 700 bp was expected if the insert was in the wrong orientation. Primer combination (1) was 5'pGEX, a vector forward primer (V) and 5'tbrlp, a gene specific forward primer (A) (see Figure 6.5, lane 1). Product sizes of 681 bp and 431 bp were expected from the amplification reactions with primer combinations 2 and 3 respectively, if in the correct orientation. (2) 5'tbrlp (A), a gene specific forward primer and 3' utbrlp (B), a gene specific reverse primer (see Figure 6.5 lane 4). (3) 5'pGEX (vector forward primer (V)) and 3'tbrlp (C), a gene specific reverse primer (see Figure 6.5, lane 2). See section 6.11 for primer sequences.

The PCR products from each of the three pairs of primers were separated on 1 % agarose gels. Figure 6.5 is a photograph showing the results from the three PCR amplifications of 1μl of the ligation mix (template). A positive control product (lane 4), was amplified from TREU 927 genomic DNA and primer combination (3), alongside the three test reactions. The expected band size from the positive control (*tbrlp* I) was 681 bp. The results from this experiment showed that there was a ligation product in the wrong orientation in lane 1. The size of the product in lane 2 corresponds to that of an insert in the correct orientation. Lane 3 also shows the presence of the insert in the ligation mix. Thus, the results suggest that, the cloning had selectively removed the representation of the sequence ligated into the vector in the correct orientation.

To investigate whether inserts could be introduced in the correct orientation during cloning, a primer was designed with a restriction site at the 5' end. The primers, 5'Bamtb and 3'utbrlp (refer section 6.11 for sequences) were used in the amplification of the insert. The amplification reaction was performed with *Pfu* polymerase and therefore the product from the amplification was blunt ended. A product with a 5' *Bam* HI and a 3'*Sma* I site was recovered after digesting with *Bam* HI. However, despite the presence of an insert in the ligation mix, I was unable to isolate any clones, which theoretically should have been in the right frame and orientation, due to the failure of the colonies to grow after transformation.

The results from this approach, suggest that a recombinant protein incorporating *tbrlp* I is either toxic to the host cell or inhibits the normal growth of the bacterial host. Only transformants that were out of frame or in the wrong orientation could grow. The simplest explanation of this is that leaky expression in the vector resulted in the premature expression of the fusion protein and a resulting toxic effect on the host prevented growth of all clones containing the insert in the right orientation and frame. To overcome these difficulties, a second approach was taken using a different expression vector.

FIGURE 6.5

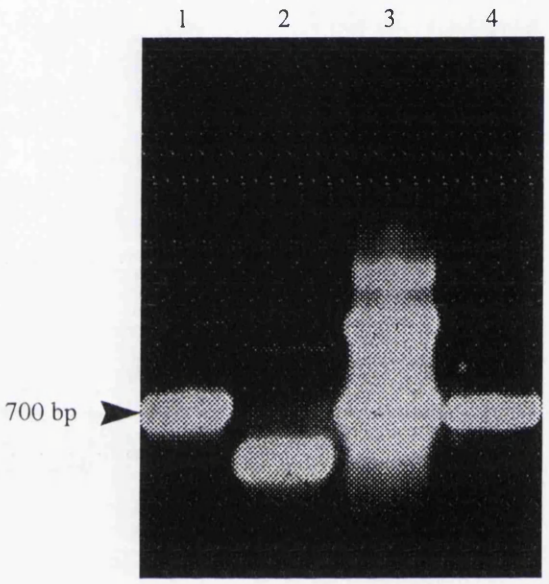


Figure 6.5 shows PCR products separated on a 1 % agarose gel and visualized by ethidium bromide staining. The products were derived from the following primer combinations; in lane 1, 5'pGEX + 5'tbrlp; in lane 2, 5'pGEX + 3'tbrlp and in lane 3, 5'tbrlp + 3'utbrlp (template used for reactions in lanes 1-3 - ligation mix). 5'tbrlp + 3'utbrlp were used in the amplification of the product in lane 4 from *T. brucei* 927 genomic DNA.

6.3 The expression of *tbrlp* I as a maltose binding protein (MBP) fusion protein

The MBP fusion protein system has two expression vectors, pMAL-c2 and pMAL-p2. Both vectors allow the insertion of cloned genes and their expression as fusion proteins linked to MBP which is a product of the *malE* gene of *E. coli*. The polylinker site of the vector is situated downstream of the *malE* gene and upstream of the lacZα gene. The insertion of a cloned gene into the multiple cloning site should therefore result in the disruption of the β-galactosidase α-fragment activity, normally present in the vector and thus facilitate the selection of positive recombinants plated on Xgal/agar. The expressed protein is linked to the MBP via the amino terminal of the protein. Each vector contains a recognition site for Factor Xa between the MBP and the inserted gene. The MBP can be cleaved at this site leaving the recombinant protein for functional analyses unhindered by the presence of the MBP. pMAL-c2 was chosen for the expression of *tbrlp* I because it is generally found to be more efficient at producing fusion proteins than pMAL-p2.

Gene specific primers were designed with 5' (*Eco* RI) and 3' (*Xba* I) restriction sites (sites which are present in the polylinker of pMAL-c2). The primers were used in the amplification of *tbrlp* I from genomic DNA of strain TREU 927 and ligated into pMAL-c2. Transformations were carried out according to the manufacturers instructions. Twelve transformants were picked, but only one (pmtb 2.6) was sequenced from both ends, (using the malE primer for the 5' end of the sequence and M13 for the 3' end) and shown to contain the sequence corresponding to *tbrlp* I and to be in frame with MBP. The insert was subcloned into Bluescript and sequenced from both ends, to ensure that the sequence corresponded to that of pmtb 2.6 and *tbrlp* I. Glycerol stocks of both clones were prepared and stored at -20° C.

Small scale preparations were made of the pmtb 2.6 fusion protein (see Materials and Methods) and five different fractions were isolated: uninduced, induced, insoluble matter, crude extract and the crude extract bound to the amylose resin (used in the purification of the fusion protein). These were all separated on a 12 % SDS-polyacrylamide gel and stained with Coomassie blue in order to identify the expressed polypeptide(s). The predicted size of *tbrlp* I is 25 kDa while MBP has a molecular size of 42 kDa, thus the expected size of the fusion protein is 67 kDa. The results, presented in Figure 6.6, show a polypeptide of 69 kDa in lanes 5 and 6 suggesting that the fusion protein was being expressed. Molecular size estimations, using the GCG programs, are normally based on peptide sequences and processes such as isoprenylation and glycosylation are not taken into consideration, therefore, the predicted size of the fusion protein may be either slightly higher or lower than predicted. In the case of *tbrlp* I, the size of the fusion protein indicated that the predicted molecular size of the product of *tbrlp* I may have been

underestimated. Two polypeptides with molecular weights of 50 and 52 kDa respectively were also observed on the Coomassie stained gel. The presence of these polypeptides was initially attributed to proteinase degradation and a number of measures were taken to eliminate the degradation.

In order to carry out GTP-binding activity studies, a large scale preparation of the fusion protein was made and protein concentrations were determined using the Biorad Bradford Assay Kit (using BSA as a standard). Extracts of the induced fractions were collected from the column and the protein concentrations determined. The fractions with the highest protein concentrations were separated on 12 % SDS polyacrylamide gels and stained with Coomassie blue in order to determine the molecular weight of the expressed proteins. Due to some smaller bands that were observed on the gel from the small scale expression of pmtb 2.6, a cocktail of protease inhibitors were used to decrease what was suspected to be proteinase degradation. The column buffer contained the following protease inhibitors: 5 mM EDTA, 5 mM Benzamidine, 1 μM Aprotinin and 1 μM Leupeptin.

A Coomassie stained SDS PAGE separation, of fractions of the expressed protein (data not shown), showed that there was no significant decrease in the number of smaller polypeptides. These could, therefore, be due to initiation at sites internal to the start codon and not protein degradation or alternatively be due to the lack of inhibition of *E. coli* proteinases. This would account, to some extent, for the lack of change in the number of smaller polypeptides observed in the presence of a cocktail of protease inhibitors. However, the presence of the band with a molecular weight which was close to that of the expected band size led me to initiate the next step of the project, that is, to determine whether pmtb 2.6 binds GTP.

FIGURE 6.6

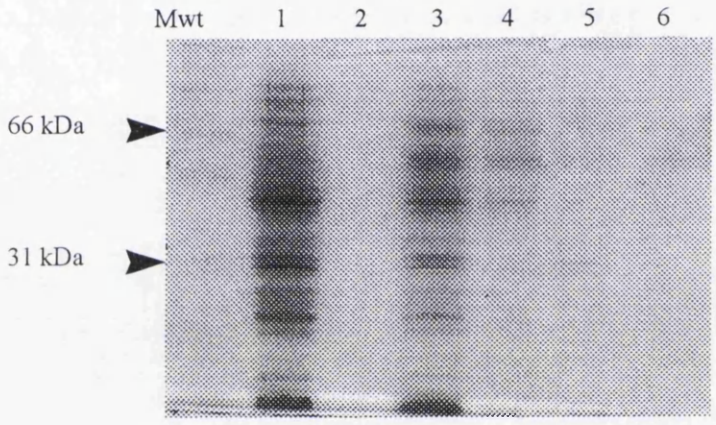


Figure 6.6 shows a coomassie blue stained SDS polyacrylamide gel with the following separated samples: lane 1, uninduced pmtb 2.6; lane 3, induced pmtb 2.6; lane 4, insoluble matter; lane 5, crude extract and lane 6 crude extract bound to the amylose resin. The 69 kDa band can be seen in lanes 5 and 6. Protein markers from NEB were used in the determination of molecular sizes of the visualized polypeptides.

6.4 GTP - binding activity of pmtbs 2.1 and 2.6

All members of the *ras* superfamily have the ability to bind and hydrolyse GTP. It was, therefore, essential to establish whether pmtb 2.6 could bind GTP. Two identical gels loaded with the following samples were separated by SDS polyacrylamide gel electrophoresis and assayed for GTP binding using the method of Coulter and Hide (1995): pmtb 2.1 (from the set of positive colonies picked earlier); the putative GTP binding protein pmtb 2.6; MBP (which served as a negative control as it is unable to bind GTP), and the spun detergent extract of bloodstream trypanosomes (positive control for GTP binding). Biorad SDS-PAGE molecular weight markers were used for the determination of molecular weight sizes. One of the gels was Coomassie stained in order to establish the presence of all protein fractions on the gel. Samples from the second gel were transferred onto a polyvinylidene difluoride (PVDF) membrane, blocked for 1 hour at room temperature (RT) and subsequently incubated for 2 hours at RT in [α -^{32}P]GTP. After the incubation period, the filter was washed and put to film. Results from the assay (data not shown) were negative for pmtb 2.1, 2.6 and MBP, while being positive for the extracts from bloodstream trypanosomes (positive control). The results suggest that the fusion proteins from pmtb 2.1 and pmtb 2.6 do not bind GTP.

The lack of GTP binding by the fusion proteins could have been due to a number of reasons such as (1) MBP is a large protein of 42 kDa and the fusion protein may have refolded in such a way that access to the GTP - binding domain by [α -^{32}P] GTP may have been blocked by the MBP.

(2) The inability of the fusion protein to bind GTP may be due to the expression of a eukaryotic recombinant protein in a bacterial expression system. Biological activity of members of the *ras* GTPase superfamily is dependent on post translational modification of the proteins. The bacterial host used in the expression of *tbrlp* I, i.e *E. coli*, does not produce prenyl transferases which are necessary post translational modification. Expression of the gene within a eukaryotic host which does produce prenyl transferases might therefore lead to a biologically active fusion protein capable of binding GTP. (3) While the assay of Coulter and Hide (1995) measures binding of GTP to some polypeptides in trypanosome extracts, there is no evidence to suggest that it measures the binding of GTP to the *tbrlp* gene product. Thus, although the positive control demonstrated GTP binding to trypanosome extracts, the assay may not be a suitable assay for detecting the binding of GTP to the *tbrlp* product. It may, therefore, be necessary to further develop the assay to detect GTP binding in a greater range of *T. brucei* GTP binding proteins.

6.41 *Development of conditions for GTP binding in pmtb 2.6*

In order to enhance correct folding of the fusion protein, the experiment described in section 6.4 was repeated using additional steps known to assist protein folding in other systems. Two gels were run and the proteins transferred onto PVDF membranes cut to size and treated as follows: membrane 1 was treated with 6 M urea and membrane 2 with 6 M guanidine-HCl for 30 minutes at room temperature. The initial incubation period ensures the complete denaturation of the proteins. The concentrations of the refolding reagents were decreased by a twofold dilution every 10 minutes, until a final concentration of 0.94 M was achieved. On achieving final concentrations of 0.94 M for both urea and guanidine-HCl, the membranes were rinsed in blocking buffer. This step ensured that all traces of the urea (for membrane 1) and guanidine-HCl (for membrane 2) were removed. The membranes were then blocked overnight at 4° C and subsequently assayed for GTP-binding (as previously described).

An increase in the intensity of the radioactive bands from the control samples, i.e bloodstream and procyclic spun detergent extracts treated in urea, suggested that urea is a better folding reagent than guanidine-HCl. Therefore, in order to enhance proper folding, it would be better to routinely refold the proteins in urea. However, despite the fact that the urea treatment resulted in more radioactivity bound to the positive controls (i.e spun detergent extracts of bloodstream and procyclic trypanosomes), the fusion proteins still showed no binding activity. The results therefore suggest that, the lack of GTP - binding activity of the fusion protein may not be entirely due to poor refolding.

6.42 *Factor Xa cleavage of pmtb 2.6*

As a Factor Xa cleavage site occurs between the MBP and the inserted recombinant protein, cleavage using Factor Xa would release the recombinant protein from the MBP. In this form the GTP binding domain of pmtb 2.6, could not be stearically hindered by the MBP and therefore might be able to bind GTP.

Two expressed purified samples, pmtb2.6, *cyp*-6 and a negative control (i.e not affected by Factor Xa cleavage) MBP, were subjected to Factor Xa treatment over a period of 36 hours with samples taken at 2, 4, 8, 24 and 36 hours. *cyp*-6 is a *C. elegans* cyclophilin B isoform, which has been expressed as a soluble MBP fusion protein (Page *et al*, 1996), and can be cleaved to a limited extent by Factor Xa. This control was added so as to ensure that the reaction had actually occurred as Factor Xa cleavage is unreliable in some circumstances.

Aliquots of the samples after the 36 hour incubation period with Factor Xa, were run on 12 % SDS polyacrylamide gels and Coomassie stained to detect the polypeptides. Figure 6.7 shows a photograph of a 12 % Coomassie stained gel, showing the cleaved and uncleaved, positive control; cleaved and uncleaved pmtb 2.6 and "cleaved" MBP. Even though the sample in lane 4, i.e cleaved pmtb 2.6, appeared fainter than the uncleaved sample, no cleavage product was detected at 25 kDa, which is the predicted size of the recombinant protein without the MBP domain, nor was there the expected 25 kDa cleavage product from *cyp*-6. Samples from the Factor Xa cleavage reaction, were taken out at the onset of the reaction, 2, 8, 24 and 36 hrs respectively. The samples, plus MBP and factor Xa which served as negative controls, were separated on 12 % SDS polyacrylamide gels, transferred onto PVDF membrane, treated with decreasing concentrations of urea and subsequently assayed for GTP binding activity. Positive control spun detergent extracts of bloodstream and procyclic forms of *T. brucei* showed GTP binding while neither MBP nor any of the fractions of the "cleaved" fusion protein (i.e samples 2 and 3) showed GTP binding. Factor Xa also did not bind GTP (data not shown). The lack of cleaved pmtb 2.6 and therefore GTP-binding ability, could be due to the inefficiency of the cleavage reaction by Factor Xa. As there was no evidence of a cleaved product from pmtb 2.6 (or the positive control, *cyp*-6), the results from the GTP binding assay were not relevant as in the absence of total cleavage, the MBP molecule could still block access of the GTP binding domain by [α -^{32}P] GTP.

The inability to remove the MBP domain from the fusion protein, suggested the use of another vector for the expression of *tbrlp* I. As the Qiaexpressionist system has only 6 residues of histidine attached to the fusion protein the GTP -binding domain is less likely to be obscured by the size of the histidine tag.

FIGURE 6.7

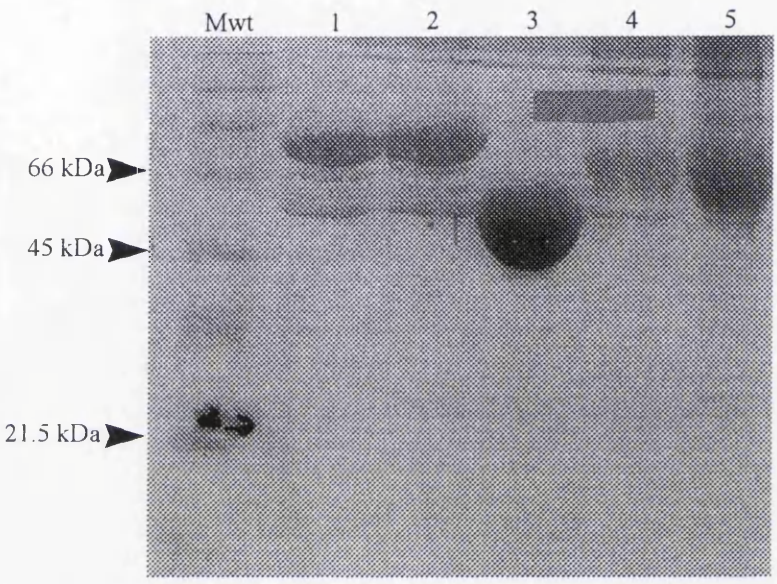


Figure 6.7 shows a Coomassie stained gel of the results from the Factor Xa cleavage reaction. In lane 1, cleaved *cyp*-6 (expected size of cleaved recombinant protein - 25 kDa); lane 2, uncleaved *cyp*-6; lane 3, MBP (not cleaved by Factor Xa); lane 4, cleaved pmtb2.6 and in lane 5, uncleaved pmtb 2.6.

6.5 The expression of *tbrlp* I as a pQE fusion protein

The Qiaexpressionist system was the third approach used in the production of a fusion protein from *tbrlp* I. Primers were designed with 5' *Bam* HI and 3' *Sph* I restriction sites (required for cloning into the vector) and used in the amplification of *tbrlp* I. The ends of the PCR product were digested with *Bam* HI and *Sph* I and cloned into pQE 30 (one of the qiaexpressionist vectors). Two attempts were made, to produce His-tagged fusion protein using the pQE 30 expression vector. Small scale preparations were made (according to the manufacturers manual), but in both cases did not become translucent on lysing. This is found to be characteristic of insoluble pQE fusion proteins.

The Qiaexpressionist system is known to be extremely sensitive to hydrophobic regions, particularly transmembrane sequences and signal peptides. In future a more appropriate approach might be to use the truncated version of *tbrlp* I in the production of the fusion protein. The truncated version of *tbrlp* I lacks the carboxyl terminus and is less hydrophobic than the full length gene and may be less likely to form an insoluble protein. A second approach to producing a soluble protein would have been to solubilise the proteins (from the full length gene) in either 6 M guanidine hydrochloride or 8 M urea. The refolding reagents have no effect on the affinity of the histidine tag for the affinity matrix and therefore purification of the proteins can still be achieved. The process of finding the right conditions for the expression of a truncated form of *tbrlp* I and/or the refolding of the solubilised full length fusion protein was not undertaken due to time constraints and thus precluded any further attempts to obtain an active fusion protein of *tbrlp* I.

6.6 Generation of antibodies to *tbrlp* I

One of the strategies that would be of use for further studies of *tbrlp*, would be the production of anti-sera against *tbrlp*. The use of anti-serum in immunolocalization, immunoprecipitation and functional studies is one of the most commonly used strategies in the study of proteins. The failure to produce a *tbrlp* fusion protein prevented this from being used as a source of *tbrlp* for generating antibodies. As an alternative approach a peptide was synthesised from which *tbrlp* specific anti-serum could be raised and used in further studies of the *ras*-like protein.

A synthetic KLH-peptide was produced for the purpose of raising *tbrlp* specific anti-peptide serum (hereafter referred to as anti-tbrlp pep). The peptide used in the production of the

anti-serum was a hydrophilic 12-mer and designed from the amino acid sequence of the carboxy-terminal extension region (i.e the unique region) of *tbrlp* I (refer Chapter 4, section 4.7, Figure 4.13). The peptide sequence of anti-tbrlp pep is as follows: SEVRDEVHQLEL. Two rabbits were used to produce the anti-peptide serum but as testing revealed that only one of them produced any anti-peptide antibodies, only this one will be considered in the following sections.

6.61 *tbrlp* I-*specific anti-peptide serum*

Pre-immune serum taken from the rabbit was stored at -20° C until required for blocking studies. Immune serum was obtained after immunisation with tbrlp-pep, followed by four boosts at 3-4 week intervals. The reactivity of the resulting anti-serum was tested by Western blot analysis of bloodstream and procyclic detergent extracts. Initially the optimum conditions required for the testing of the anti-serum were determined. A range of dilutions of the anti-serum were tested: 1:200, 1:500 and 1:1000. Two dilutions of the second antibody were used, 1:400 (manufacturers recommendation) and 1:800. Results from these tests showed that 1:200 of the first antibody and 1:800 of the second antibody produced results with a minimum amount of background to noise ratio (data not shown). These were therefore adopted as the optimum dilutions required for determining the specificity of the reactions.

Figure 6.8 shows the results obtained from Western blot analysis of both the supernatant and the pellet fractions of spun detergent extracts (SDEs) of bloodstream and procyclic trypanosomes, using both preimmune and immune serum. Three bands with the following molecular weights, 23.3 kDa, 26.7 kDa and 30.3 kDa, were detected by the anti-serum and were not detected by preimmune serum. The expected size of the protein product of *tbrlp* I is 25 kDa, which falls within the range of the peptides recognised by anti-tbrlp pep. In addition to the triplet of polypeptides at 23 - 30 kDa, a series of higher molecular weight polypeptides were detected, which were not observed using the pre-immune serum, showing that the immune serum is not totally specific to the lower molecular weight polypeptides. Anti-tbrlp pep appears to be specific for polypeptides found within the pellet fractions of both bloodstream and procyclic detergent extracts and would suggest that *tbrlp* I is membrane localised. The presence of the polypeptides in the pellet fractions of both bloodstream and procyclic detergent extracts, confirms that the putative product of *tbrlp* I is not stage specific, as was shown by the results from the Northern blot (Chapter 4, section 4.9, Figure 4.20).

FIGURE 6.8

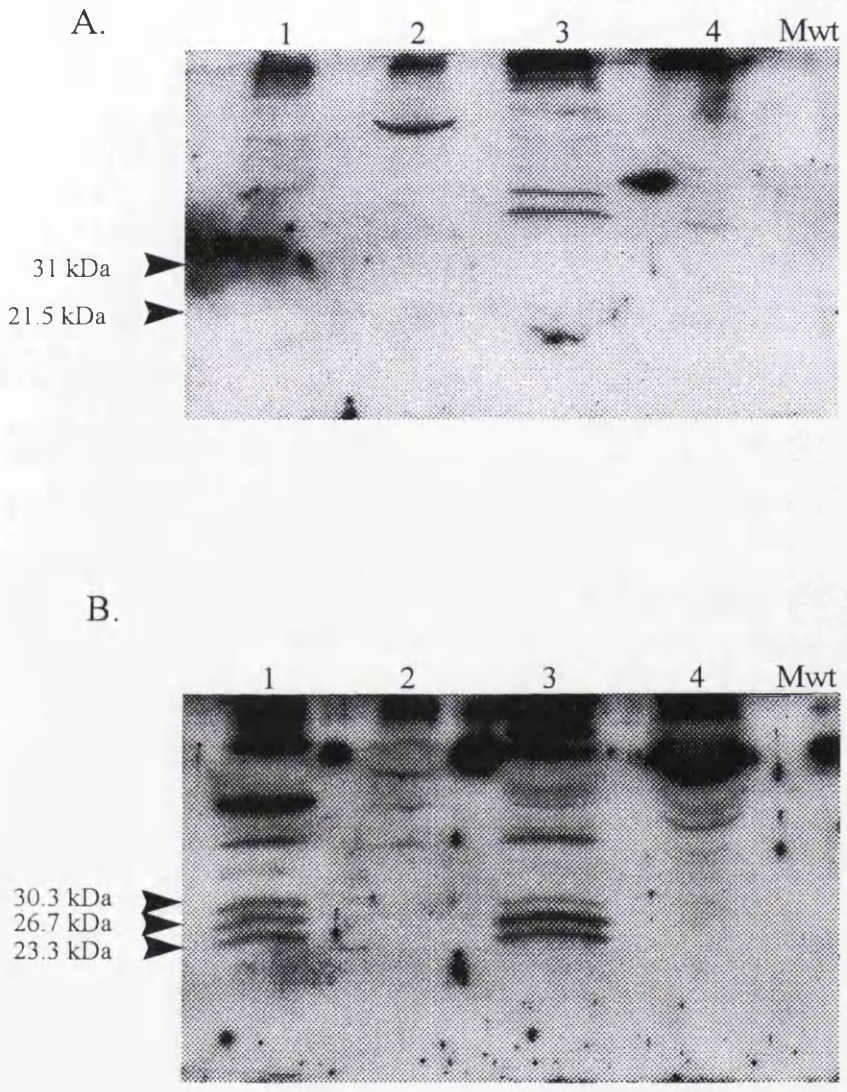


Figure 6.8 shows the results obtained from Western blot analysis of the supernatant and pellet fractions of bloodstream and procyclic trypanosomes. Membrane A was incubated in preimmune serum and membrane B was a duplicate which was incubated in *tbrlp* specific anti-serum. In lane 1 of each of the membranes, bloodstream pellet; lane 2, bloodstream supernatant; lane 3, procyclic pellet; lane 4, procyclic supernatant. Mwt - protein molecular marker.

6.62 *Peptide blocking studies*

Peptide blocking studies were carried out on SDEs of both bloodstream and procyclic trypanosomes, to determine whether the three polypeptides of 23.3, 26.7 and 30.3 kDa, are specifically recognised by anti-tbrlp pep. The peptide is identical to that used in the production of the anti-serum, but unconjugated (hereafter referred to as tbrlp-pep). A 1 mg ml^{-1} stock solution of tbrlp I-pep was prepared, stored at -20° C and used when required. Preliminary experiments were carried out to determine the most suitable concentration required for the blocking of the polypeptides observed in both SDEs of bloodstream and procyclic trypanosomes.

Figure 6.9 shows results of the peptide blocking studies, which clearly show that a 12 μg ml^{-1} solution of tbrlp -pep, completely blocks the recognition of the three polypeptides (molecular weight masses of 23.3, 26.7, and 30.3 kDa respectively). This confirms that anti-tbrlp pep recognises polypeptides of approximately the expected molecular size of *tbrlp* in trypanosome extracts. The higher molecular weight polypeptides, were however, not blocked by the peptide, which indicates that they are not *tbrlp* specific polypeptides. The recognition of the higher molecular weight polypeptides might therefore be eliminated by affinity purifying the anti serum. What remains to be determined is whether the three bands are due to the presence of the two different copies of *tbrlp* (i.e *tbrlp* I and II); pre-/post-translational modification products or the cross-reactivity of anti-tbrlp pep with other related polypeptides

FIGURE 6.9

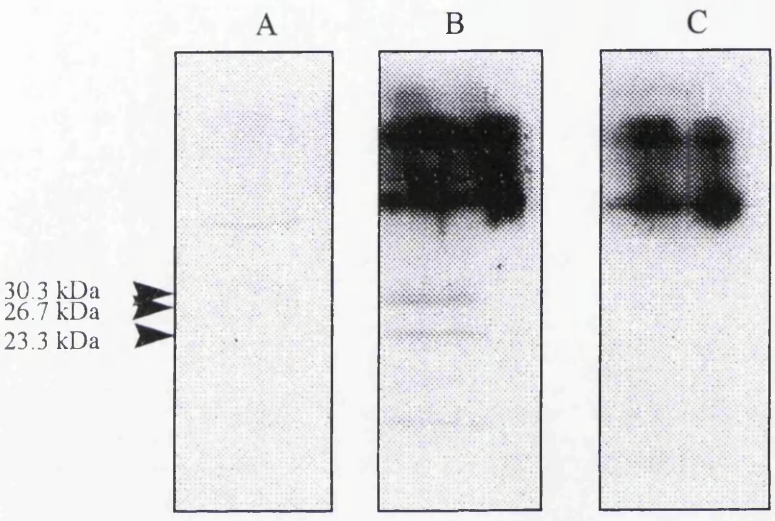


Figure 6.9 shows the results from the peptide blocking studies on SDS-PAGE separated total bloodstream SDE. Panel A was incubated in a 1:200 dilution of preimmune serum; B, 1:200 dilution of *tbrlp* specific anti serum and panel C in a 1:200 dilution of *tbrlp* specific anti serum preincubated in 12 μg ml $^{-1}$ of tbrlp-pep for 1 hour at 4°C.

6.7 Discussion

To generate tools for functional studies on *tbrlp*, experiments were undertaken to construct an active recombinant protein and to generate antibodies against *tbrlp* I. A number of attempts were made to produce a fusion protein. The GST system was unsuccessful in expressing a viable fusion protein and the only bacterial colonies that were produced were the transformants which had inserts in the wrong frame or inserts in the right frame but the wrong orientation. This would suggest that the expressed proteins could have an adverse effect on the *E. coli* host.

A fusion protein of the expected size was produced by the maltose binding protein system. The fusion protein, however, did not have the ability to bind GTP. This may have been due to the GTP-binding domain of pmtb 2.6 being obscured by the large maltose binding protein domain. To circumvent this, experiments were undertaken to cleave the recombinant protein using the Factor Xa cleavage reaction. This reaction is known to be inefficient in some cases, and it was therefore not surprising that attempts, at cleaving the maltose binding protein from the recombinant, were unsuccessful. Therefore, the lack of GTP-binding activity of *tbrlp* I could not be unequivocally attributed to an MBP blocked GTP-binding domain. An alternative explanation could be the lack of proper renaturation or refolding of the proteins. Using two different reagents capable of catalysing protein refolding in other systems, namely urea and guanidine-HCl, proteins transferred onto PVDF membranes were refolded and assayed for GTP-binding activity. Results from the assay of the positive controls (i.e SDEs of bloodstream and procyclic trypanosomes) suggest that the refolding of proteins in this system is more efficient in urea than guanidine-HCl.

The only conclusions one can draw from these results are that the lack of GTP-binding activity may have been due to the inability of [α -^{32}P] GTP to gain uninhibited access to the GTP-binding domain of the fusion protein or that, the lack of GTP-binding activity may be due to the lack of machinery necessary for the production of a fully active fusion protein, i.e the lack of prenyl transferases due to the expression of a eukaryotic protein in a bacterial expression system. This explanation is particularly appealing, as the sequence of *tbrlp* suggests that it undergoes isoprenylation and therefore might need to be post translationally modified in order for it to become active, i.e bind GTP.

Due to the inability to cleave the maltose binding protein from the fusion protein, another method was employed to determine whether or not GTP-binding activity could be established in the absence of the large MBP. The Qiaexpressionist system was used in the production of a fusion protein from *tbrlp* I. This system was employed as the expressed protein has only 6 histidine residues attached to it, therefore, if the lack of GTP-binding activity was due to an MBP

blocked GTP-binding domain, the pQE fusion protein might show some activity, as it is less likely to block the *tbrlp* GTP binding domain. Two attempts were made at the expression of *tbrlp* I as a pQE 30 fusion protein. In both attempts the protein was found to be insoluble. Due to the limited amount of time available, these experiments were not carried further. To do so would have entailed finding the conditions necessary for the production of a soluble fusion protein. However, a number of approaches are available for the solublilisation of the fusion proteins, such as the use of decreasing molarities of either 6 M guanidine-HCl or 8 M urea. Conditions required to enhance proper folding of the proteins will have to be determined. This procedure will not in anyway affect the affinity of the His tag for the Ni-NTA resin which is used in the purification of pQE fusion proteins. Another approach would be the use of the truncated form of *tbrlp* I in the production of the fusion protein.

An alternative approach would be the use of a baculoviral expression system, as this would result in the production of the protein in an eukaryotic expression system, thus allowing the post translational modification of proteins. This is required not only for subcellular localisation, but also for the full biological activity of the fusion protein (Willumsen *et al*, 1984; Porfiri *et al*, 1995). A fusion protein produced in this way could therefore be used to define some of the functions of *tbrlp* I.

Anti-peptide serum raised against a region within the unique C-terminal extension domain of *tbrlp* I, recognised three polypeptides in the pellet fraction of both bloodstream and procyclic spun detergent extracts. Recognition of all three polypeptides was competed out by a 12 $\mu g\ ml^{-1}$ solution of the unconjugated peptide. This suggests that the anti-serum recognises more than one polypeptide within the SDEs of bloodstream and procyclic trypanosomes and could be due to the expression of the second copy of *tbrlp* I. Preliminary sequence analysis of part of the putative second copy of *tbrlp* I i.e *tbrlp* II, shows a high degree of sequence identity which would predict cross reactivity. It is possible that *tbrlp* II is a different size and could be modified in a different way to *tbrlp* I, and, therefore, could represent one of the polypeptides detected by anti-tbrlp pep. However, further sequence data are required to determine this. An alternative explanation for the three polypeptides could be the occurrence of pre-/post-translational modification leading to products of different sizes, all of which contain the epitope recognised by the anti-serum but have different electrophoretic mobilities. Another possibility is the degradation of the product of *tbrlp* I, during sample preparation. The expected molecular size of the product of *tbrlp* I is 25 kDa, which is approximately the same size as one of the polypeptides detected by the GTP binding assay described in Chapter 3. However, due to the occurrence of several small GTP binding proteins of different subfamilies within the same molecular weight range, a number of experiments would have to be conducted in order to

establish that they are actually the same protein. The immunoprecipitation of the HPLC derived fractions with anti-tbrlp pep would be good approach to establishing whether the 24 kDa polypeptides are indeed *ras* subfamily GTP-binding proteins. Affinity purification of anti-tbrlp pep on a peptide column would produce a highly specific anti-serum which could be used in the determination of the subfamily to which the previously isolated 24 kDa polypeptides (see Chapter 3) belong and in the determination of the subcellular location of *tbrlp* I.

The next line of approach would be to determine whether or not some of the observed bands are due to pre-/post-translational modification products, before embarking on subcellular localisation studies. The strategy used by Field *et al*, 1996, highlighting the presence of an isoprenylation pathway in *T. brucei*, can be used (with a few modifications) to identify the products recognised by anti-tbrlp pep. Mevalonate is a key intermediate in the isoprenyl biosynthetic pathway from acetoacetyl CoA (Hancock, 1995). The inhibition of the synthesis of mevalonate is achieved by inhibiting the enzyme hydroxymethylglutaryl-coenzyme A reductase, using inhibitors like compactin and mevinolin (Field *et al*, 1996). Cells deprived of mevalonate tend to take up exogenous mevalonate for the synthesis of isoprenoids and squalene (Field *et al*, 1996). Briefly, bloodstream (or procyclic) trypanosomes deprived of mevalonate could be incubated in culture with exogenous radioactive mevalonate. After the incubation period, the *ras* protein (i.e *tbrlp*) could be immunoprecipitated and separated by SDS polyacrylamide gel electrophoresis. These results would provide information on the size of the isoprenylated proteins recognised by anti-tbrlp pep. A second approach would be to separate the hydrophilic (unprocessed) and hydrophobic (processed) fractions (Hancock, 1995) of bloodstream or procyclic trypanosomes using Triton X-114. The *ras* GTPase can then be immunoprecipitated from the hydrophilic (which represents the unprocessed i.e non isoprenylated proteins) and hydrophobic (isoprenylated proteins) fractions, separated on an SDS polyacrylamide gel and detected by Western blotting to determine the state of the proteins (i.e processed or unprocessed) and the sizes of bands recognised by the anti-serum. The results from such an analysis should provide information on the pre/post translational status of the proteins recognised by anti-tbrlp pep.

In order to use the anti-serum in the isolation of native *tbrlp* I, it would be necessary to demonstrate that anti-tbrlp pep recognises the expressed product of *tbrlp* I. Once this has been established, the antiserum could be used in the isolation of native *tbrlp* I (and/or *tbrlp* II, if it recognises both proteins), which could be used in studying the function of the proteins in *T. brucei*. Results from the Northern analysis and anti peptide binding studies have shown that *tbrlp* I is not stage-specific and the amino acid sequence suggests that it is a membrane bound protein. These results suggest that the function of *tbrlp* I may be essential for a process common

to both life cycle stages of the trypanosome such as growth and differentiation. If *tbrlp* I functions in the same manner as do *ras* GTPases of higher eukaryotes (i.e if it is involved in the regulation of cellular growth and differentiation) then, the development of tools for the analysis of the role of *tbrlp* I would result in the generation of data which may lead to a greater understanding of the mechanisms of trypanosomal growth regulation.

Chapter 7

Discussion

CHAPTER 7

DISCUSSION

7.0 General Discussion

The aims of this project were to investigate the presence of small GTP binding proteins and, in particular, members of the *ras* subfamily in *Trypanosoma brucei* and to develop the tools and reagents needed to study the role of *ras* subfamily GTPases in trypanosomal development. Studies in higher eukaryotes show that members of the *ras* subfamily are directly involved in the transduction of cellular signals, which result in the regulation of cell growth and differentiation (Hall, 1990; Bokoch, 1993; Bokoch, 1996). The isolation of a member of this subgroup from *T. brucei* would, therefore, be the first step towards the development of tools for the study of the role of *ras* subgroup GTPases in cellular signalling in *T. brucei*.

The project was divided into four concurrently run parts: (1) The development of a GTP binding assay suitable for screening large numbers of fractions generated by HPLC column purification methods; (2) The identification of *ras* like GTP binding protein from *T. brucei*. There were two approaches to this stage: a) the isolation of a *ras* like GTP-binding protein by means of HPLC purification of small GTP binding proteins from bloodstream trypanosomes; and b) the isolation of genes encoding *ras* like GTP binding proteins using molecular methods; (3) The evolutionary analysis of a *ras* like GTP binding protein, *tbrlp* I and (4) The development of tools needed for the study of the role of *tbrlp* I in the signalling pathways of *T. brucei*.

The presence of low molecular weight GTP binding proteins in *T. brucei* was confirmed by the *in situ* GTP binding assay of Coulter and Hide (1995) following polyacrylamide gel electrophoresis of trypanosome protein extracts. This assay showed the presence of GTP binding proteins within a range of 18 - 35 kDa and is consistent with sizes of members of the mammalian *ras* superfamily (Kahn, Der and Bokoch, 1992). Some of the bands were observed solely in the bloodstream or the procyclic extracts suggesting that the proteins are stage specific. A routine GTP binding assay was developed for the handling and concentration of large numbers of sample using conditions similar to the *in situ* GTP binding assay of Coulter and Hide (1995). This was shown to be suitable for identifying HPLC fractions positive for GTP binding. The first step used to purify *ras* like GTP binding proteins from *T. brucei* was HPLC anion-exchange chromatography. Under specific separation conditions a single 24 kDa polypeptide was identified

which bound GTP. As protein can be detected down to 10 ng by silver staining (Guillemett and Lewis, 1983), the fact that this band could not be visualised on a silver stained SDS polyacrylamide gel suggested that there was less than 10 ng of the protein on the gel. Thus, there was a need to scale up the amount of starting material to recover suitable amounts of this purified GTP binding protein. A survey of a number of studies, in which small GTP binding proteins have been purified to near homogeneity (e.g from human placenta and bovine brain extracts, Evans *et al*, 1986 and Waldo *et al*, 1987), showed that large amounts of starting material are required for the purification of a workable amount of small GTP binding proteins. Bovine brain is known to contain 5-10 fold more *ras* GTP binding protein than any other mammalian tissue (Yamashita *et al*, 1988) and an initial amount of 858 mg of crude membrane protein from bovine brain produced only 11 μg of small GTP binding proteins. This suggests that to achieve purity by the use of successive HPLC chromatography methods, the amount of *T. brucei* starting material needs to be scaled up. The initial protein content of 14 mg of the bloodstream spun detergent extract (equivalent to 1 ml of packed bloodstream trypanosomes) resulted in less than 10 ng of the 24 kDa purified small GTP binding protein. Using the values for human placenta for comparison 60 ml of packed bloodstream trypanosomes would be required to produce 11 μg of purified GTP binding protein. However, the actual amount of material required would be well in excess of 60 ml of packed trypanosomes as a purification protocol would also have to be developed. The HPLC purification approach was, therefore, discontinued due to the number of rats that would have to be infected to produce 60 ml of trypanosomes (approximately 200) and success in other areas.

A heterologous anti-human *ras* antibody was used to investigate the presence of members of the *ras* subgroup of proteins in *T. brucei*. Due to the cross reactivity of the antibody, it was difficult to interpret the results with any certainty. Future approaches using this method should be aimed at developing specific conditions under which this *ras* antibody reacts specifically and testing a further range of antibodies.

The next strategy was to isolate genes encoding *ras* subfamily GTPases, using degenerate primer PCR. This resulted in a 280 bp PCR product from *T. brucei* TREU 869 cDNA. The PCR product was used to obtain the sequence of a full length gene, *tbrlp*, which encoded a *ras* subfamily GTP binding protein.

tbrlp has an open reading frame of 681 bp (227 amino acids) and a predicted molecular size of 25 kDa, which is consistent with its classification as a small GTPase. In addition to the molecular size, *tbrlp* has all the conserved sequence motifs characteristic of a member of the *ras* GTPase subfamily. High stringency hybridisation, with the full length *tbrlp* as the probe, indicated that the gene is probably single copy, but a putative second copy was subsequently

identified using *tbrlp* specific 3' and 5' primers, and appeared to be upstream of, and tandemly linked to, *tbrlp*. The possible existence of *tbrlp* as a multi copy gene is not surprising as many housekeeping genes in trypanosomes are found to be multi copy (Hadjuk *et al*, 1992; Tschudi, 1995). Due to the existence of a putative second copy of *tbrlp*, the original gene was named *tbrlp* I and the second gene, *tbrlp* II. The full length sequence of *tbrlp* II has not yet been determined but preliminary sequence suggests that it will be virtually identical to *tbrlp* I. Northern analysis indicates that *tbrlp* I is expressed in both bloodstream and procyclic life cycle stages of the trypanosome and suggests that it is an essential gene which has a role in processes common to both life cycle stages of the trypanosome.

tbrlp is the first member of the *ras* subfamily to have been cloned from *T. brucei*. Approximately 10 small GTPases have been cloned and fully sequenced from *T. brucei*: *ran* (Field and Boothroyd, 1995) and at least nine *rab* GTPase genes (El Sayed *et al*, 1995; Field *et al*, 1995). An alignment of *tbrlp* I and these other *ras* like genes from *T. brucei* shows that the only sequence identity lies within the six conserved sequence motifs of the GTP binding domain. *tbrlp* I is, therefore, a previously undescribed *ras* like gene from *T. brucei*. *tbrlp* I has some novel features which make it a unique *ras* subfamily member. The GTP binding domain of *tbrlp* has 45 % and 41 % amino acid sequence identity to that of human *rap* 2 and *ras* h subgroup genes respectively. This feature makes it difficult to classify *tbrlp* as either a *rap* or a *ras* subgroup member. In mammalian cells functional studies of *ras* and *rap* indicate that the two proteins exist in an antagonistic relationship (Lerosey *et al*, 1991; Bokoch, 1993; Altschuler *et al*, 1995; Nassar *et al*, 1995) and together, they seem to be involved in the regulation of cellular growth and differentiation (Hall, 1990; Bokoch, 1996). The existence of this "mosaic" *ras*-like protein in *T. brucei*, an organism which diverged early in eukaryotic evolution, suggests that the protein could represent an ancestral *ras/rap* gene which existed before the divergence of the gene into *ras* and *rap*. Due to the mixed features of *tbrlp* I, phylogenetic analysis was carried out to determine its evolutionary significance.

Six different hypotheses were put forward to explain the possible evolutionary origins of *tbrlp* I. On testing, only one of the hypotheses could be used to explain the origins of *tbrlp* I. The results from the phylogenetic analysis of *tbrlp* I, using a number of tree building methods, showed that *tbrlp* I is an ancestral *ras/rap* gene. Conversely, this evolutionary analysis of *tbrlp* I suggests that the kinetoplastids diverged at a point prior to the duplication and divergence of a primeval GTP binding domain to form the *ras* and *rap* genes of higher eukaryotes but after the evolution of *rab* and *ran* GTP binding domains. This accounts for the presence of *ran* and *rab* GTPases in *T. brucei*. The eukaryotes higher up the evolutionary scale, must, therefore, have diverged after the duplication and divergence of the *ras/rap* ancestral subfamily gene. This

hypothesis agrees with the findings of Fernandes *et al*, 1993 and Maslov *et al*, 1995, with regards to the order of divergence of the eukaryotes on the SSU rRNA tree of life.

The *ras/rap* like similarities of *tbrlp* suggest that it has a role in cellular signalling. To investigate this role, a series of reagents and tools were required. These were: a *tbrlp* fusion protein with the ability to bind GTP and *tbrlp* specific anti-sera (from either the whole fusion protein or to a peptide conserved in the variable region of *tbrlp* i.e unique to *tbrlp*) which could be used in the determination of the subcellular localisation of *tbrlp*, immunoprecipitation and further functional analyses. Two approaches were taken. The first approach entailed the expression of a fusion protein using *tbrlp* I and three different expression systems were used in attempts to produce a suitable fusion protein. Only one of the expression systems, the MBP fusion system, produced a soluble fusion protein which lacked GTP binding ability. This may have been due to the GTP binding domain of the protein being obscured by the large size of the MBP. Factor Xa cleavage of the MBP did not succeed in resolving this question. On the other hand the lack of GTP binding could have been due to the fact that *tbrlp* I is a eukaryotic protein expressed in a bacterial system which does not contain the enzymes required for eukaryotic post translational modifications. To test this, *tbrlp* I could be expressed in a eukaryotic expression system such as a baculovirus system which contains all the machinery necessary for post translational modification e.g prenyl transferases which are not present in *E. coli*. Recent experiments have shown that a viable mevalonate pathway exists in trypanosomes (Field *et al*, 1996) which is necessary for the production of the conjugated isoprene groups involved in post translational protein modification. *tbrlp* has a carboxyl terminal isoprenylation signal which implies that, in order to express a viable and biologically functional protein, all the necessary cellular machinery must be made available. A fusion protein produced in such a way could be a useful tool for biochemical analyses. GTP binding assays could be carried out on the fusion protein, as the fully functional recombinant protein is expected to bind GTP. More specific functional analyses on *tbrlp* I can be carried out once the ability of the recombinant protein to bind GTP is shown. Such a protein could also be used in antibody production for immunoprecipitation and immunolocalisation studies. These studies could shed light on the mode of action and subcellular localisation of *tbrlp*.

The second approach to developing analytical tools was to raise *tbrlp* specific anti-peptide sera. A peptide was synthesised using the amino acid sequence derived from the carboxyl terminal extension region of *tbrlp* I. Using this anti-serum, anti-tbrlp pep, *tbrlp* I was detected in the detergent insoluble fractions of both bloodstream and procyclic trypanosomes. This result suggests that *tbrlp* I is membrane localised. However, due to the recognition of three different polypeptides, it appears as though the anti-serum in its current state does not solely recognise

tbrlp. Results from peptide blocking studies show that the three bands are effectively blocked when the antiserum is preincubated in a 12 μg ml^{-1} solution of the unconjugated peptide. These bands fall within the 17-30 kDa range, consistent with the expected size of *tbrlp* I. This suggests that, in addition to *tbrlp* I, the anti-serum recognises other proteins. Preliminary sequence analysis of *tbrlp* II indicate that the sequence is identical to that of *tbrlp* I. It is unlikely, therefore, that the other polypeptides are due to *tbrlp* II. One explanation for the multiple bands is that the anti-serum might be recognising pre/post translational modification products of *tbrlp* I which might have the same epitopes as the full length protein but have different electrophoretic mobilities.

7.1 Future investigations

In order to carry out further studies on the role of *tbrlp* I in trypanosomal development, the tools currently available for these studies i.e *tbrlp* specific anti-serum and a fusion protein, would have to be further developed in the following manner: The anti-serum in its present form cross reacts with a few high molecular weight peptides and so would need to be affinity purified in order to reduce the cross-reactivity with other epitopes. Commercially prepared affinity purification kits are currently available and can be used in the purification process. The identity of the three bands must be established i.e are they a result of pre-/post translational modification products of *tbrlp* I, *tbrlp* II or degradation products of *tbrlp* I, which may have been formed during sample preparation? The method used by Field *et al*, 1996, which showed the presence of an isoprenylation pathway in *T. brucei,* could be modified and used in the determination of pre/post translational modification products (see Discussion, Chapter 6). The use of peptide sequencing in the determination of the full length sequences of the three polypeptides could also be used to determine the pre/post translational modification status of the polypeptides recognised by anti-tbrlp pep.

The full length sequence of *tbrlp* II remains to be determined. Primers designed from the sequence can be used to amplify the full length gene from genomic DNA. The full length genes from *tbrlp* I and II could then be expressed as fusion proteins in the baculoviral system. A eukaryotic expression system should provide all the machinery necessary for post translational modification (i.e prenyl transferases) and, therefore, the expressed proteins should be biologically functional. The recombinant proteins could be used in antibody production as there is a possibility that antibodies to the recombinant proteins may be more specific than the anti-peptide serum.

Having developed the analytical tools and reagents, a number of questions have to be addressed in order to shed light on the function of *tbrlp*: (a) Does it bind and hydrolyse GTP? (b) Is *tbrlp* essential for the viability of both bloodstream and procyclic trypanosomes? (c) Will the transfection of trypanosomes with mutant *tbrlp* affect normal growth patterns of the parasite? (d) Where is it localised? Studies of *ras* and *rap* GTPases in higher eukaryotes have shown that the two proteins are localised in different cellular fractions. *ras* proteins are found on the inner surface of the plasma membrane (Lowy and Willumsen, 1991) while *rap* proteins are found on the endoplasmic reticulum and endosomal/lysosomal compartments (Zerial and Huber, 1995). As *tbrlp* I has features of both *ras* and *rap* it is important to determine its cellular location. (e) Does it bind to other molecules such as the trypanosomal equivalent of GTPase activating proteins, guanine nucleotide exchange factors or a *raf* protein kinase homologue? (f) Are there any significant changes in GTP hydrolysis when trypanosomes are exposed to extracellular signals known to be directly involved in trypanosomal growth regulation (e.g TNF-α, IFN-γ and EGF)? (g) Is the activation of *tbrlp* dependent on phosphorylation and if so is it a substrate for Protein kinase A, C or both? Some studies on the phosphorylation patterns of *ras* subfamily GTPases have reported the phosphorylation of three mammalian *ras* GTPases by both Protein kinases A and C or by protein kinase C alone (Altschuler and Lapetina, 1993). Results from studies on mammalian *rap* 1 GTPases, show that they are phosphorylated solely by Protein kinase A (Altschuler and Lapetina, 1993). The phosphorylation patterns of *tbrlp* should therefore give an indication as to whether the function of *tbrlp* is *ras* like or *rap* like.

As preliminary data suggest that *tbrlp* I and II are identical, it is likely that their functions are also identical. Gene knockouts would have to be undertaken in order to determine the function of *tbrlp* I and II. *tbrlp* is expressed in both bloodstream and procyclic trypanosomes indicating that it might be essential for a basic process common to both life cycle stages. A double knockout might therefore be lethal. The inducible gene expression system developed by Wirtz and Clayton (1995) could be used to determine the effects of single and double knockouts on the growth patterns of trypanosomes. Briefly, this system makes use of trypanosomes transfected with a tetracycline repressor of *E. coli*. The DNA of the gene to be studied (in this case, *tbrlp*) can be integrated into the vector, under the control of a PARP promoter containing a tet operator. Results from the double knockout of *tbrlp* I would indicate whether its presence in the *T. brucei* genome is essential for normal trypanosomal development. However, normal trypanosomal development in the *tbrlp* I knockouts would suggest that *tbrlp* II may be able to function in an homologous manner and therefore might compensate for the loss of *tbrlp* I. In order to determine whether both *tbrlp* I and II are essential for normal trypanosomal development it would be necessary to knockout *tbrlp* II in the *tbrlp* I knockouts. A lethal effect on the

trypanosomes would indicate that the two genes are essential for normal trypanosomal development. A tet operator-*tbrlp* II construct could then be introduced into the knockout strain, and used in the maintenance of the expression of *tbrlp* II, whilst the wild type allele is knocked out in order to observe the effect of "turning off" *tbrlp* II. The same process could be repeated with the introduction of a tet operator-*tbrlp* I construct into the knockout strain, so as to observe the effect *tbrlp* I has on trypanosomal development after the elimination of the wild type allele.

Different mutants of *tbrlp* I and II could be produced, using site directed mutagenesis (which would entail the substitution of individual amino acids, within the GTP binding domain, known to be essential for normal protein function) and used to determine the effect of particular mutations on the viability of the trypanosomes using the inducible gene expression system of Wirtz and Clayton (1993).

The binding of guanine nucleotides to mutant *tbrlp* can be determined and compared to the binding by wild type *tbrlp*. This can be measured *in vitro* by determining the amounts of GTP and GDP found in immunoprecipitated *tbrlp*, using both radioactive (based on the amount of GDP converted to [γ-^{32}P]GTP from [γ-^{32}P]ATP and nucleoside diphosphate kinase) and non radioactive methods (the amount of GTP present can be measured by determining the amount of ATP produced from ADP and nucleoside diphosphate using a luciferase assay, Gibbs, 1995). *In vitro* measurements of GTP and GDP of trypanosomes exposed to extracellular signals known to be directly involved in the regulation of trypanosomal growth, should shed light on the effects of these signals on GTP-binding proteins. Studies have shown that there is an increase in cAMP levels just before the transformation of long slender trypanosomes to short stumpy forms and from short stumpy forms to procyclics. The increase in levels are known not to be associated with the heterotrimeric G-proteins (Pays et al, 1997) but could be due to other GTP binding/hydrolysis reactions of small GTP binding proteins such as *tbrlp*.

The cellular activity of all *ras* and *rap* GTPases in mammalian cells is under the regulation of GTPase activating proteins (GAPs) and guanine nucleotide exchange factors (GEFs). It would be interesting to investigate the presence of the trypanosomal equivalents of GAPs, and GEFs and their effect on the activation of *tbrlp*. Studies in mammalian cells have shown that GEFs form highly stable complexes with *ras* proteins in nucleotide depleted media. Based on this observation, a number of binding assays are currently available for the detection of GEF-*ras* complexes in nucleotide depleted media using known GEFs e.g yeast Cdc 25. Yeast GEF is structurally related to mammalian GEF and has in the past been used in the immunoprecipitation of nucleotide depleted mammalian *ras* proteins. The possibility therefore exists that it could also be structurally related to the trypanosomal homologue of yeast GEF. Briefly, the procedure would entail the precipitation of Glutathione-S-transferase tagged GEF and

tbrlp complexes using glutathione-agarose beads. Unbound protein would be washed away and the complexes detected by immunoblotting with *tbrlp* specific anti-serum and visualising by enhanced chemiluminesence (Hart and Powers, 1995). A complete absence of bands should be observed in the lanes containing nucleotide bound forms of *tbrlp*. The same approach could be used in the detection of GAPs, with the *ras*-GAP complexes being in the samples containing nucleotide bound *tbrlp*. Both guanine exchange releasing factors and GTPase activating proteins bind *ras* and *rap* at the effector binding domain (which is identical) and have highly conserved structural features (Maekawa *et al*, 1993). The effector binding domain of *tbrlp* I has approximately 70 % amino acid identity to human *rap* 2 and 66 % identity to human H-*ras*. It therefore, might be possible to coprecipitate *tbrlp* and GEFs/GAPs, using known mammalian/yeast Glutathione-S-transferase tagged GEFs/GAPs. Once the ability of *tbrlp* to bind GEF/GAP has been established the trypanosomal homologues could be isolated by immunoprecipitation, using mammalian/yeast GEF/GAP heterologous antisera.

Raf is a serine/threonine kinase known to be involved in signalling pathways essential for growth and differentiation of mammalian cells (Zerial and Huber, 1995). It is found downstream of *ras* and physically interacts with a number of molecules including *ras*, *rap* and MAP kinase kinase. *Raf* kinases have three distinct domains, two of which are found at the amino terminal (CR 1 and CR 2) and are conserved between *raf* family members (Rapp *et al*, 1988), and one which is found at the carboxyl terminal end and represents the kinase domain i.e CR 3. CR 1 appears to be involved in interactions with *ras*, which are necessary for the regulation of cell growth and differentiation. The *ras-raf*-MAP kinase kinase pathway has been found to be functional in a number of cell types and in several organisms including *Caenorhabditis elegans* and *Drosophila melanogaster* (Dickson *et al*, 1992; Lu *et al*, 1993; Han *et al*, 1993). The presence of a *ras*-like gene in *T. brucei* suggests, by analogy to other organisms, that a *raf* kinase may also be present that interacts with the *tbrlp* product in a classical signalling pathway. A number of approaches could be considered to provide evidence for a *raf* kinase in *T. brucei*. These include attempting to clone a *raf* kinase from trypanosomes and the use of heterologous antisera to detect the presence of a *raf* kinase activity or polypeptide.

7.2 Possible roles of *tbrlp*

Ras GTPases of the *ras* subfamily are evolutionarily conserved proteins. They are found across a wide range of organisms from the slime moulds to humans (Lerosey et al, 1991). Studies in mammalian cells have shown that they are involved in the control of several physiological processes, such as growth and cellular differentiation (Hall, 1990). The presence of a *ras* like GTPase in *T. brucei*, indicates that there could be a *ras* linked signalling pathway which could be involved in the regulation of growth and differentiation of the trypanosome.

Currently little is known about cellular signalling pathways in *T. brucei* and in the light of this one can only speculate on the role of *tbrlp* I in these pathways. An adenylate cyclase receptor has been isolated from *T. brucei*, is encoded by ESAG 4 (Paindavoine, 1992) and is found to be activated by calcium. The activation of adenylate cyclase results in the initiation of the cAMP cascade, which is known to be linked to a number of physiological changes in other organisms, such as growth and differentiation in mammalian cells. Studies have shown that mammalian (Milligan, 1995; Neer, 1995) and *Phytomonas* (Farber et al, 1995) adenylyl cyclases are activated by heterotrimeric G proteins and this in turn results in the initiation of the cAMP cascade. Studies of the effect of heterotrimeric G proteins (with nonhydrolysable analogs of GTP) on the activation of adenylyl cyclase, show that the heterotrimeric G proteins are not involved in the stimulation of trypanosomal adenylate cyclase. Yeast adenylyl cyclases are found to be stimulated by monomeric G proteins. Based on these observation, the possibility exists that the trypanosomal adenylyl cyclases might also be stimulated by *ras* GTPases. In *T. brucei*, the cAMP cascade is known to be involved in the regulation of cell growth and differentiation (Mancini and Patton, 1981; Reed *et al*, 1985), and if *ras* GTPases (i.e *tbrlp*) are involved in the stimulation and inhibition of this cascade, then it would suggest another possible role for *tbrlp*. *tbrlp* I has a C-terminal cAMP dependent protein kinase (Protein kinase A) phosphorylation site, with 75 % amino acid identity to that of human *rap* 1b. The activation of human *rap* 1b is dependent on phosphorylation by cAMP (Altschuler and Lapetina, 1993; Altschuler et al, 1995). The activity of *tbrlp* might therefore be regulated by the cAMP cascade in *T. brucei* and thus might play a regulatory role in the proliferation and differentiation of the parasites.

Homologues of some of the components found in the mammalian MAP kinase pathway have been identified in *T. brucei*. Studies from higher eukaryotes have shown that the activation of *ras* is an important intermediate in the stimulation of the MAP kinase cascade. If *tbrlp* I and II are involved in either the cAMP or MAP kinase pathways then it would suggest that *tbrlp* is important for growth and differentiation in *T. brucei*. This might make it a suitable drug target.

Figure 7.0 is schematic diagram of a model showing the possible mode of activation of membrane bound *tbrlp*, based on the presence of the effector binding domain which has 70 % amino acid identity to that of human *rap* 2 and 66 % identity to human H-*ras*, and is involved in signal transduction based on the activation of *raf* (in mammalian cells), the presence of a GTP binding domain, a cAMP dependent phosphorylation site and the assumption that a *raf* homologue (tbraf) exists in *T. brucei* and is involved in trypanosomal cell signalling.

FIGURE 7.0

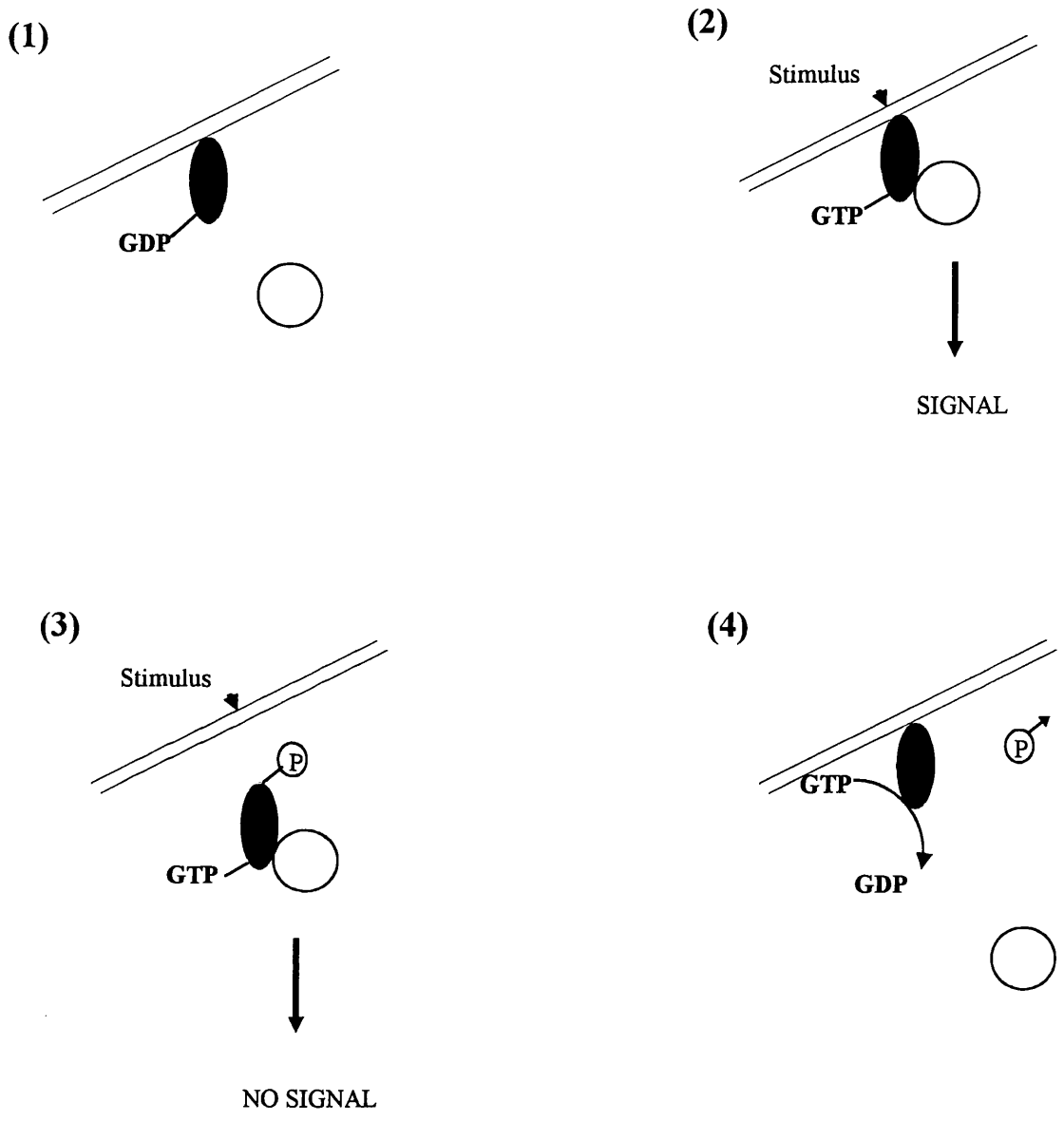


Figure 7.0 represents a schematic diagram of a proposed model for the activation of *tbrlp* (●), in the presence of a putative trypanosomal *raf* serine/threonine kinase homologue (tbraf -O). (1) In the absence of an appropriate stimulus, *tbrlp* is membrane bound and inactive, i.e bound to GDP. (2) On stimulation *tbrlp* is activated by the physical replacement of GDP by GTP. In this state it binds to and recruits *raf* to the membrane (as do activated *ras* proteins in mammalian cells) and activates effectors downstream of *raf*. This activation could cause the stimulation of a Protein kinase A pathway which results in the phosphorylation of *tbrlp*, at the C-terminal cAMP dependent phosphorylation site (3) and the dissociation of the *tbrlp*-tbraf complex from the membrane, thus breaking contact with the external stimulus and its receptor. The binding of *ras* to *raf* in mammalian cells is known to cause the inhibition of the guanine nucleotide exchange reaction (Sprang, 1995), thus in this state the hydrolysis of GTP is the favoured reaction. The change to the inactive state causes a conformational change which decreases the affinity of *tbrlp* for tbraf and thus reverts to its membrane bound state (4).

7.3 *tbrlp*, a putative anti-trypanosomal drug target

The main requirements of a good trypanosome drug target are (1) The target should be essential for the survival of the organism (2) It should have structural and functional differences or absent in the mammalian host. (3) It should be a low abundance molecule, such that only small doses of inhibitors are needed. (4) The availability of a detailed knowledge of structural features and effects of inhibitors. If it is essential to the survival of the trypanosome, *tbrlp* would be a potentially good drug target because, it is sufficiently different from the mammalian host *ras* subfamily GTPases, it is probably a low abundance protein and much is known about the structure, function and inhibitors of *ras* proteins in mammals. Furthermore, drugs could be designed to inhibit its action but not have a detrimental effect on the mammalian host.

The question as to whether a *tbrlp* I like gene exists in the other kinetoplastids remains to be determined. The results of the evolutionary analysis of *tbrlp* I suggest that a *tbrlp* I like gene might be present in all the kinetoplastids. Thus it could also be a potential drug target for the treatment of *Leishmaniasis* and South American trypanosomiasis. Pharmaceutical companies might therefore find the prospect of producing a single drug for the treatment of three different, but equally devastating diseases, more appealing.

References

References

Abaru, D. E. (1985). Sleeping sickness in Busoga, Uganda, 1976 - 1983. *Tropical Medicine and Parasitology* **36**: 72-76.

Abath, F. G. C., and Werkhauser, R. C. (1996). The tegument of Schistosoma mansoni: functional and immunological features. *Parasite Immunology* **18**: 15-20.

Agabian, N. (1990). Trans splicing of nuclear pre-mRNAs. *Cell* **61**: 1157-1160.

Allred, D. R. (1995). Immune evasion by *Babesia bovis* and *Plasmodium falciparum*, cliff-dwellers of the parasite world. *Parasitology Today* **11** (3): 100-104.

Altschuler, D., and Lapetina, E. G. (1993). Mutational analysis of the cAMP-dependent protein kinase-mediated phosphorylation site of *rap* 1b. *Journal of Biological Chemistry* **268** (10): 7527-7531.

Altschuler, D., Peterson, S. N., Ostrowski, M. C., and Lapetina, E. G. (1995). Cyclic AMP-dependent activation of *rap* 1b. *Journal of Biological Chemistry* **270** (18): 10373-10376.

Avise, J. C. (1994). Molecular markers, natural history and evolution. Published by Chapman and Hall.

Avruch, J., Zhang, X.-F., and Kyriakis, J. M. (1994). *Raf* meets *ras*: completing the framework of a signal transduction pathway. *Trends in Biochemical Sciences* **19**: 279-283.

Bacchi, C. J., Brun, R., Croft, S. L., Alicea, K., and Buhler, Y. (1996). In vivo trypanocidal activities of new S-adenosylmethionine decarboxylase inhibitors. *Antimicrobial agents and Chemotherapy* 1448-1453.

Bacchi, C. J., Garofalo, J., Mokenhaupt, D., McCann, P. P., Diekema, K. A., Pegg, A. E., Nathan, H. C., Mullaney, E. A., Chunosoff, L., Sjoerdsma, A., and Hunter, S. H. (1983). *In vivo* effect of α-DL-difloromethylornithine on the metabolism and morphology of *Trypanosoma brucei brucei*. *Molecular and Biochemical Parasitology* **7**: 209-225.

Bailey, J. W., and Smith, D. H. (1992). The use of the acridine orange QBC technique in the diagnosis of African Trypanosomiasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **86**: 630.

Balber, A. E. (1990). The pellicle and the membrane of the flagellum, flagellar adhesion zone and flagellar pocket: functionally discrete surface domains of the bloodstream form of African trypanosomes. *Critical Reviews in Immunology* **10**: 177-201.

Barbacid, M. (1987). *Ras* genes. *Annual Review of Biochemistry* **56**: 779-827.

Barcinski, M. A., and Costa-Moreira, M. E. (1994). Cellular response of protozoan parasites to host-derived cytokines. *Parasitology Today* **10** (9): 352-354.

Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yeild from lambda bacteriophage templates. *Proceedings of the National Academy of Sciences, USA* **91**: 2216-2220.

Barry, J. D. (1989). African Trypanosomiasis in vaccination strategies of tropical diseases 197-217. Edited by Liew, F.Y.

Barry, J. D., and Emery, D. L. (1984). Parasite development and host responses during the establishment of *Trypanosoma brucei* infection transmitted by the tsetse fly. *Parasitology* **88**: 67-84.

Bastin, P., Coppens, I., Saint-Remy, J. M., Baudhuin, P., Opperdoes, F. R., and Courtoy, P. J. (1994). Identification of a specific epitope on the extracellular domain of the LDL-receptor of *Trypanosoma brucei brucei*. *Molecular and Biochemical Parasitology*, **63**: 193-202.

Bastin, P., Stephan, A., Raper, J., Saint-Remy, J. M., Opperdoes, F. R., and Courtoy, P. J. (1996). A Mr 145000 low-density lipoprotein (LDL)-binding protein is conserved throughout the Kinetoplastida order. *Molecular and Biochemical Parasitology* **76**: 43-56.

Bellafatto, V., Fairlamb, A. H., Henderson, G. B., and Cross, G. A. M. (1987). Biochemical changes associated with a-difloromethylornithine uptake and resistance in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **25:** 227-238.

Beranger, F., Goud, B., Tavitian, A., and Degunzburg, J. (1991). Association of the *ras*-antagonistic *Rap*1/Krev-1 proteins with the golgi complex. *Proceedings of the National Academy of Sciences. USA* **80** (5): 1606-1610.

Bhattacharya, S., Chen, L., Broach, J. R., and Powers, S. (1995). *Ras* membrane targeting is essential for glucose signaling but not for viability in yeast. *Proceedings of the National Academy of Sciences, USA* **92**: 2984-2988.

Bitonti, A. J., Dumont, J. A., and McCann, P. P. (1986a). Characterisation of *Trypanosoma brucei brucei* S-adenosy-L-methionine decarboxylase and its inhibition by Berenil, pentamidine and methylglyoxal bis (guanylhydrazone). *Biochemical Journal* **237**: 518-521.

Bitonti, A. J., McCann, P. P., and Sjoerdsma, A. (1986). Necessity of antibody response in the treatment of African Trypanosomiasis with a-difloromethylornithine. *Biochemical Pharmacology* **35**: 331-334.

Block, C., Janknecht, R., Herrmann, C., Nassar, N., and Wittinghofer, A. (1996). Quantitative structure-activity analysis correlating *ras/raf* interaction *in vitro* to *raf* activation *in vivo*. *Nature Structural Biology* **3** (3): 244-251.

Boguski, M. S., and McCormick, F. (1993). Proteins regulating *ras* and its relatives. *Nature* **366**: 643-653.

Bokoch, G. M. (1993). Biology of the *rap* proteins, members of the *ras* superfamily of GTP-binding proteins. *Biochemical Journal* **289**: 17-24.

Bokoch, G. M. (1996). Interplay between *ras*-related and heterotrimeric GTP binding proteins: lifestyles of the big and little. *Faseb Journal* **10**: 1290-1295.

Bokoch, G. M., and Der, C. J. (1993). Emerging concepts in the *ras* superfamily of GTP-binding proteins. *FASEB Journal* **7**: 750-759.

Borst, P. (1991). Transferrin receptor, antigenic variation and the prospect of trypanosome vaccine. *Trends in Genetics* **7** (10): 307-309.

Borst, P., Fase-Fowler, F., Frasch, A.C. C., Hoeijmakers, J. H. J., and Weijers, P. J. (1980). Characterisation of DNA from *Trypanosoma brucei* and related trypanosomes by restriction endonuclease digestion. *Molecular and Biochemical Parasitology* **6**: 13-23.

Boshart, M., and Mottram, J. C. (1997). Protein phosphorylation and protein kinases in trypanosomatids. *Trypanosomiasis and Leishmaniasis; Biology and control*: 227-244. Edited by G. Hide, J.C. Mottram, G.H. Coombs and P.H. Holmes.

Bourne, H. R., Lustig, K. D., Wong, Y. H., and Conklin, B. R. (1992). Detection of coincident signals by G-proteins and adenylyl cyclase. *Cold Spring Harbour Symposia on Quantitative Biology* **57**: 145-148.

Bourne, H. R., Sanders, D. A., and McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism. *Nature* **349**: 117-127.

Brittingham, A., and Mosser, D.M. (1996). Exploitation of the complement system by Leishmania promastigotes. *Parasitology Today* **12** (11): 444-447.

Brown, J. R., and Doolittle, W. F. (1995). Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proceedings of the National Academy of Sciences, USA* **92**: 2441-2445.

Brun, R., and Schonenberger, M. (1979). Cultivation and *in vitro* cloning of procyclic culture forms of *T. brucei* in semi-defined medium. *Acta Tropica* **36**: 289-292.

Burgering, B. M. T., and Bos, J. L. (1995). Regulation of *ras*-mediated signalling: more than one way to skin a cat. *Trends in Biochemical Sciences* **20**: 18-22.

Buxbaum, L. U., Milne, K. G., Werbovetz, K. A., and Englund, P. T. (1996). Myristate exchange in the *Trypanosoma brucei* variant surface glycoprotein. *Proceedings of the National Academy of Sciences, USA* **93**: 1178-1183.

Buxbaum, L. U., Raper, J., Opperdoes, F. R., and Englund, P. T. (1994). Myristate Exchange: A second glycosyl phosphatidylinositol myristoylation reaction in African trypanosomes. *Journal of Biological Chemistry* **269** (48): 30212-30220.

Cano, E., and Mahadevan, L. C. (1995). Parallel signal processing among mammalian MAPKs. *Trends in Biochemical Sciences* **20**: 117-122.

Cappai, R. (1993). Cloning and charaterization of a golgi associated GTP-binding protein homolog from *Leishmania major*. *Molecular and Biochemical Parasitology* **62** (1): 73-82.

Cassel, D., Shoubi, S., Glusman, G., Cukierman, E., Rotman, M., and Zilberstein, D. (1991). *Leishmania donovani*: characterization of a 38-kDa membrane protein that cross-reacts with the mammalian G-protein transducin. *Experimental Parasitology* **72**: 411-417.

Certa, U. (1991). Malaria vaccine. *Experientia* **47**: 147-163.

Chan, M. M., and Fong, D. (1994). Plant microtubule inhibitors against trypanosomatids. *Parasitology Today* **10** (11): 448-451.

Chavrier, P., Gorvel, J.-P., Stelzer, E., Simons, K., Gruenberg, J., and Zerial, M. (1991). Hypervariable C-terminal domain of *rab* proteins acts as a targeting signal. *Nature* **353**: 769-772.

Chen, J., DeVivo, M., Dingus, J., Harry, A., Li, J., Sui, J., Carty, D. J., Vlank, J. L., Exton, J. H., Stoffel, R. H., Inglese, J., Lefkowitz, R. J., Logothetis, D. E., Hildebrandt, J. D., and Iyengar, R. (1995). A region of adenylyl cyclase 2 critical for regulation by G-protein $\beta\gamma$ subunits. *Science* **268**: 1166-1169.

Chomczynski, P., and Saachi, N. (1987). Single step method of RNA isolation by acid guanidium thiocyanate phenol chloroform extraction. *Analytical Biochemistry* **162** (1): 156-159.

Clayton, C., Hausler, T., and Blattner, J. (1995). Protein Trafficking in kinetoplastid protozoa. *Microbiological Reviews* 325-344.

Clayton, C., and Michels, P. (1996). Metabolic compartmentation in African Trypanosomes. *Parasitology Today* **12** (12): 465-471.

Cook, J., Rubinfeld, B., Albert, I., and McCormick, F. (1993). *Rap*V12 antagonizes *ras*-dependent activation of ERK1 and ERK2 by LPA and EGF in Ra 1 fibroblasts. *EMBO Journal* **12**: 3475-3485.

Coppens, I., Bastin, P., Courtoy, P. J., Baudhuin, P., and Opperdoes, F. R. (1991). A rapid method purifies a glycoprotein of Mr 145,000 as the LDL receptor of *Trypanosoma brucei brucei*. *Biochemical and Biophysical Research Communications* **178**: 185-191.

Coppens, L., Baudhuin, P., Opperdoes, F. R., and Courtoy, P. J. (1988). Receptor for the low-density lipoproteins on the hemoflagellate *Trypanosoma brucei*. Purification and involvement in growth of the parasite. *Proceedings of the National Academy of Sciences, USA* **85**: 6753-6757.

Coso, O. A., Diaz-Anel, A., Martinetto, H., Muschietti, J. P., Kazanietz, M., Fraidenraich, D., Torres, H. N., and Flawia, M. M. (1992). Characterization of a Gi-protein from *Trypanosoma cruzi* epimastigote membranes. *Biochemical Journal* **287**: 443-446.

Coulter, L. J., and Hide, G. (1995). *Trypanosoma brucei*: Characterisation of a life cycle stage-specific G-protein. *Experimental Parasitology* **80**: 308-318.

Cross, G. A. M. (1990). Cellular and genetic aspects of antigenic variation in trypanosomes. *Annual Reviews of Immunology* **8**: 83-110.

Cross, G. A. M., and Takle, G. B. (1993). The surface trans-sialidase family of *Trypanosoma cruzi*. *Annual Reviews of Microbiology* **47**: 385-411.

Daum, G., Eisenmann-Tappe, I., Fries, H.-W., Troppmair, J., and Rapp, U. R. (1994). The ins and outs of *raf* kinases. *Trends in Biochemical Sciences* **19**: 474-479.

Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. (1978). A model for evolutionary change in proteins. *Atlas of Proteins Sequence Structure* **5** (3): 345-352. Edited by Dayhoff, M. Washington D.C. National Biomedical Research Foundation.

De Rijk, P., Van der Peer, Y., Van den Broeck, I., and De Wachter, R. (1995). Evolution according to large ribosomal subunit RNA. *Journal of Molecular Evolution* **41**: 366-375.

Der, C.J., Finkel, T., and Cooper, G.M. (1986a). Biological and biochemical properties of human ras^{H} genes mutated at codon 61. *Cell* **44**: 167-176.

Dirie, M. F., Otte, M. J., Thatthi, R., and Gardiner, P. R. (1992). Comparitive studies of *Trypanosoma* (*Duttonella*) *vivax* isolates from Columbia. *Parasitology* **106**: 21-29.

Docampo, R., and Pignataro, O. P. (1991). The inositol phosphate/diacylglycerol signalling pathway in *Trypanosoma cruzi*. *Biochemical Journal* **275**: 407-411.

Doolittle, R. F., and Feng, D. F. (1990). A nearest neighbour procedure for relating progressively aligned amino acid sequences. *Methods in Enzymology* **183**: 659-669.

Drouin, G., Moniz de Sa, M., and Zuker, M. (1995). The *Giardia lamblia* actin gene and the phylogeny of eukaryotes. *Molecular Evolution* **41**: 841-849.

Durnez, P., Pernambuco, M. B., Oris, E., Arguelles, J.-C., Mergelsberg, H., and Thevelein, J. M. (1994). Activation of trehalase during growth induction by nitrogen sourced in the yeast *Saccharomyces cerevisiae* depends on the free catalytic subunits of cAMP-dependent protein kinase, but not on functional *ras* proteins. *Yeast* **10**: 1049-1064.

Efron, B., and Tibshirani, (1991). Statistical data analysis in the computer-age. *Science* **253** (5018): 390-395.

Efron, B., and Gong, G. (1983). A leisurely look at the bootstrap, jack-knife and cross validation. *American Statistician* **37**: 36-48.

Efron, B., Halloran, E., and Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences, USA* **93**: 7085-7090.

Eisenschlos, C., Flawia, M. M., Torruella, M., and Torres, H. N. (1986a). Interaction of *Trypanosoma cruzi* adenylate cyclase with liver regulatory factors. *Biochemical Journal* **236**: 185-191.

Eisenschlos, C., Paladini, A. A., Molina y Vedia, L., Torres, H. N., and Flawia, M. M. (1986b). Evidence for the existence of an Ns-type regulatory protein in Trypanosoma cruzi membranes. *Biochemical Journal* **237**: 913-917.

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

Engelberg, D., Zandi, E., Parker, C. S., and Karin, M. (1994). The yeast and mammalian *ras* pathways control transcription of heat shock genes independently of heat shock transcription factor. *Molecular and Cellular Biology* 4929-4937.

Evans, T., Brown, M. L., Fraser, E. D., and Northrup, J. K. (1986). Purification of the major GTP binding proteins from human placental membranes. *Journal of Biological Chemistry* **261** (15): 7052-7059.

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

Farber, M. D., Montagna, A. E., Paveto, C., Dollet, M., Sanchez-Moreno, M., Osuna, A., Torres, H. N., and Flawia, M. (1995). Adenylyl cyclase and G-proteins in *Phytomonas*. *Journal of Eukaryotic Microbiology* **42** (3): 257-260.

Farrar, M. A., Alberola-Ila, J., and Perimutter, R. M. (1996). Activation of the *Raf*-1 kinase cascade by coumermycin-induced dimerization. *Nature* **383**: 179-184.

Feig, L. A., and Schaffhausen, B. (1994). The hunt for *ras* targets. *Nature* **370**: 508-509.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39** 783-791.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package). Abstract.

Feng, D. F., and Doolittle, R. F. (1990). Progressive alignment and phylogenetic tree construction of protein sequences. *Methods in Enzymology* **183**: 375-387.

Fernandes, A. P., Nelson, K., and Beverley, S. M. (1993). Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa; perspectives on the age and origins of par*as*itism. *Proceedings of the National Academy of Sciences, USA* **90**: 11608-11612.

Field, M., Field, H., and Boothroyd, J. C. (1995). A homologue of the nuclear GTPase *ran*/TC4 from *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **69**: 131-134.

Field, M. C., and Boothroyd, J. C. (1995). *Trypanosoma brucei*: molecular cloning of homologues of small GTP-binding proteins involved in vesicle trafficking. *Experimental Parasitology* **81**: 313-320.

Fitch, W. M. (1995). Uses for evolutionary trees. *Philosophical Transactions of the Royal Society of London* **349**: 93-102.

Fraidenraich, D., Pena, C., Isola, E. L., Lammel, E. M., Coso, O., Anel, A. D., Pongor, S., Baralle, F., Torres, H. N., and Flawia, M. M. (1993). Stimulation of Trypanosoma cruzi adenylyl cyclase by an αD-

Hancock, J. F. (1995). Prenylation and palmitoylation analysis. *Methods in Enzymology* **255** (Part A): 237-239.

Hancock, J. F., Cadwallader, K., Magee, A. I., Newman, C., Giannakouros, T., Fawell, E., Armstrong, J., Paterson, H. F., and Marshall, C. J. (1991). Prenylation of *ras* and *ras*-related proteins. *The superfamily of ras-related genes* 15-22. Edited by D.A. Spandidos, Plenum press, New York.

Hancock, J. F., Magee, A. I., Childs, J. E., and Marshall, C. J. (1989). All *ras* proteins are polyisoprenylated but only some are palmitoylated. *Cell*, 57, 1167-1177.

Hart, M., and Powers, S. (1995). *Ras*-Cdc25 and *rho*-Dbl binding assays: Complex formation *in vitro*. *Methods in Enzymology* **255** (Part A): 129-135.

Henderson, G. B., and Fairlamb, A. H. (1987). Trypanothione metabolism: a chemother*ap*eutic target in trypanosomatids. *Parasitology Today* **3** (No.10): 312-315.

Hide, G. (1994). The elusive trypanosome. *Parasitology Today* **10** (3): 85-86.

Hide, G., Cattand, P., Le Ray, D., Barry, J. D., and Tait, A. (1990). The identification of *T. brucei* subspecies using repetitive DNA sequences. *Molecular and Biochemical Parasitology* **39**: 213-226.

Hide, G., Gray, A., Harrison, C. M., and Tait, A. (1989). Identification of an epidermal growth factor receptor homologue in trypanosomes. *Molecular and Biochemical Parasitology* **36**: 51-59.

Hide, G., and Tait, A. (1991). The molecular epidemiology of parasites. *Experientia* **47**: 128-142.

Hide, G., Tait, A., and Keith, K. (1990). Characterization of *Trypanosoma brucei* protein kinases and a growth factor receptor. *Biochemical Society Transactions* **18**: 733-735.

Hide, G., Tait, A., Maudlin, I., and Welburn, S. C. (1996). The origins, dynamics and generation of *Trypanosoma brucei rhodesiense* epidemics in East Africa. *Parasitology Today* **2** (2): 50-55.

Hoare, C. A. (1972). The trypanosomes of mammals. A zoological monograph. Oxford Blackwell Scientific Publications.

Hua, S., and Wang, C. C. (1994). Differential accumulation of a protein kinase homolog in *Trypanosoma brucei*. *Journal of Cellular Biochemistry* **54**: 20-31.

Hughes, D. A., Ashworth, A., and Marshall, C. J. (1993). Complementation of byr1 in fission yeast by mammalian MAP kinase kinase requires coexpression of *Raf* kinase. *Nature* **364**: 349-352.

Hughes, D. A., Yoshiyuki, I., and Yamamoto, M. (1991). Regulation of the *ras* pathway in the fission yeast Schizosaccharomyces pombe. *The superfamily of ras-related genes* 41-47. Edited by D.A. Spandidos, Plenum press, New York.

Hurwitz, N., Segal, M., Marbach, I., and Levitzki, A. (1995). Differential activation of yeast adenylyl cyclase by *ras*1 and *ras*2 depends on the conserved N terminus. *Proceedings of the National Academy of Sciences USA* **92**: 11009-11013.

Ives, H. E. (1991). GTP binding proteins and growth factor signal transduction. *Cellular Signalling* **3** (6): 491-499.

Iwabe, N., Kuma, K., and Miyata, T. (1996). Evolution of gene families and relationship with organismal evolution: *rap*id divergence of tissue-specific genes in the early evolution of chordates. *Molecular Biology and Evolution* **13** (3): 483-493.

Jernigan, J. A., and Pearson, R. D. (1993). Chemother*apy* of leishmaniasis, Chagas' disease and African Trypanosomiasis. *Current Opinion in Infectious Diseases* **6**: 794-802.

Kahn, R. A., Der, C. J., and Bokoch, G. M. (1992). The *ras* superfamily of GTP binding proteins: guidelines on nomenclature. *FASEB Journal* **6**: 2512-2513.

Karin, M. (1991). Signal transduction and gene control. *Current Opinion in Cell Biology* **3**: 467-473.

Kataoka, T., Powers, S., Cameron, S., Fasano, O., Goldfarb, M., Broach, J. and Wigler, M. (1985). Functional homology of mammalian and yeast *ras* genes. *Cell* **40** (1): 19-26.

Kawata, M., Matsui, Y., Kondo, J., Hishida, T., Teranishi, Y., and Takai, Y. (1988). A novel small molecular weight GTP-binding protein with the same putative effector domain as the *ras* proteins in bovine brain membranes. *Journal of Biological Chemistry* **263** (35): 18965-18971.

Keith, K., Hide, G., and Tait, A. (1990). Characterization of protein kinase C like activities in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* 43: 107-116.

Kim, S., Mizoguchi, A., Kikuchi, A., and Takai, Y. (1990). Tissue and subcellular distribution of the Smg-21/*rap*1/Krev-1 proteins which are partly distinct from those of c-*ras* p21s. *Molecular and Cellular Biology* **10** (6): 2645-2652.

Kitayama, H., Sugimoto, Y., Matsuzaki, T., Ikawa, Y.,and Noda, M. (1989). A *ras*-related gene with transformation suppressor activity. *Cell* **56**: 77-84.

Koener, T., de Raadt, P., and Maudlin, I. (1995). The Uganda sleeping sickness epidemic revisited: a case of mistaken identity. *Parasitology Today* **8**: 303-306.

Komba, D. B., Chimfwembe, E. C., and Nantulya, V. M. (1992). Multicentre evaluation of an antigen-detection ELISA for the diagnosis of *Trypanosoma brucei rhodesiense* sleeping sickness. *Bulletin of the World Health Organisation* **70** (1): 57-61.

Krengel, U., Schlichting, I., Scherer, A., Schumann, R., Frech, M., John, J., Kabsch, W., Pai, E. F., and Wittinghofer, A. (1990). Three-dimensional structures of H-*ras* p21 mutants molecular basis for their inability to function as signal switch molecules. *Cell* **62**: 539-548.

Kuhner, M. K., and Felsenstein, J. (1994). A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution* **11**(3): 459-468.

Kuzoe, F. A. S. (1993). Current situation of African trypanosomiasis. *Acta Tropica* **54**: 153-162.

Lanham, S. M., and Godfrey, D. G. (1970). Isolation of salivarian trypanosomes from man and other animals using DEAE-cellulose. *Experimental Parasitology* **28**: 521-534.

Lee, M. G. S., Bihain, B. E., Russell, D. G., Deckelbaum, R. J., and Van der Ploeg, L. H. T. (1990a). Characterization of a cDNA encoding a cystein-rich cell surface protein located in the flagellar pocket of the protozoan *Trypanosoma brucei*. *Molecular and Cellular Biology* **10**: 4506-4517.

Leevers, S. J., and Marshall, C. J. (1992). MAP kinase regulation - the oncogene connection. *Trends in Cell Biology* **2**: 283-286.

Lerosey, I., Pizon, V., Tavitian, A., and de Gunzburg, J. (1991). Phosphorylation of *rap* proteins by the cAMP-dependent protein kinase. *The superfamily of ras-related genes* 31-39. Edited by D.A. Spandidos, Plenum press, New York.

Ligtenberg, M. J., Bitter, W., Kieft, R., Steverding, D., Janssen, H., Calafat, J., and Borst, P. (1994). Reconstitution of a surface transferrin binding complex in insect form *Trypanosoma brucei*. *EMBO Journal* **13**: 2565-2573.

Lipman, D. J., and Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* 227: 1435-1441.

Lowy, D., and Willumsen, B. M. (1993). Function and regulation of *ras*. *Annual Review of Biochemistry* 62: 851-891.

Lu, X.-Y., Perkins, L. A., and Perrimon, H. (1993). The torso pathway in drosophila - A model system to study receptor tyrosine kinase signal transduction. *Development* SS: 47-56.

Lucas, R., Magez, S., Songa, B., Darji, A., Hamers, R., and De Baetselier, P. (1993). A role for TNF during African trypanosomiasis: involvement in parasite control, immunosuppression and pathology. *Research in Immunology* **144**: 370-376.

Majiwa, P. A. O., Maina, M., Waitumbi, J. N., Mihok, S., and Zweygarth, E. (1993). *Trypanosoma (Nannomonas) congolense*: molecular characterization of a new genotype from Tsavo, Kenya. *Parasitology* **106**: 151-162.

Mancini, P. E., and Patton, C. L. (1981). Cyclic 3': 5' adenosine monophosphate levels during the developmental cycle of *Trypanosoma brucei brucei* in the rat. *Molecular and Biochemical Parasitology* **3**: 19-31.

Marais, R., Light, Y., Paterson, H. F., and Marshall, C. J. (1995). *Ras* recruits *raf*-1 to the plasma membrane for activation by tyrosine phosphorylaton. *EMBO Journal* **14** (13): 3136-3145.

Marrack, P., and Kappler, J. (1994). Subversion of the immune system by pathogens. *Cell* **76**: 323-332.

Marshall, M. S. (1993). The effector interactions of p21*ras*. *Trends in Biochemical Sciences* **18**: 250-254.

Marshall, M. S., Davis, L. J., Keys, R. D., Mosser, S. D., Hill, W. S., Scolnick, E. M., and Gibbs, J. B. (1991). Identification of amino acid residues required for *ras* p21 target activation. *Molecular and Cellular Biology* 3997-4004.

Martin, B. R., Voorheis, H. P., and Kennedy, E. L. (1978). Adenylate cyclase in bloodstream forms of *Trypanosoma* (Trypanozoon) *brucei* spp. *Biochemical Journal* **175**: 207-212.

Marx, J. (1993). 2 major signal pathways linked. *Science* **262** (5136): 988-.

Maslov, D. A., Avila, H. A., Lake, J. A., and Simpson, L. (1994). Evolution of RNA editing in kinetoplastid protozoa. *Nature* **368:** 345-348.

Maslov, D. A., and Simpson, L. (1995). Evolution of pa*ras*itism in kinetoplastid protozoa. *Parasitology Today* **11** (1): 30-32.

Matsuura, A., and Anraku, Y. (1994). Genetic interaction between the *ras*-cAMP pathway and the Dis2s1/Glc7 protein phosphatase in Saccharomyces cerevisiae. *Molecular and General Genetics* 242: 257-262.

Matthews, K. R., and Gull, K. (1994). Cycles within cycles: The interplay between differentiation and cell division in *Trypanosoma brucei*. *Parasitology Today* **10** (12): 473-476.

Maudlin, I., and Welburn, S. C. (1987). Lectin mediated establishment of midgut infections of Trypanosoma congolense and *Trypanosoma brucei* in Glossina morsitans. *Tropical Medical Parasitology* **38**: 167-170.

Maudlin, I., and Welburn, S. C. (1994). Maturation of trypanosome infections in tsetse. *Experimental Parasitology* **79**: 202-205.

McClure, M. A., Vasi, T. K., and Fitch, W. M. (1994). Comparitive analysis of multiple protein sequence alignment methods. *Molecular Biology and Evolution* **11** (4): 571-592.

McConville, M. J., and Ferguson, M. A. J. (1993). The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes. *Biochemical Journal* **294:** 305-324.

McCormick, F. (1990). The world according to GAP. *Oncogene* **5**: 1281-1283.

McCormick, F. (1993). How *ras* receptors turn *ras* on. *Nature* **363**: 15-16.

McNamara, J. J., Mohammed, G., and Gibson, W. C. (1994). *Trypanosoma* (Nannomona) *godfreyi* sp. nov. from tsetse in The Gambia: biological and biochemical characterization. *Parasitology*, 109, 496-509.

Milligan, G., Kim, G. D., Mullaney, I., and Adie, E. J. (1995). Regulation of cellular G(S) alpha levels and basal adenylyl cyclase activity by expression of the beta-2-adreno-receptor in neuroblastoma cell lines. *Molecular and Cellular Biology* **149**: 213-216.

Minato, T., Wang, J., Akasaka, K., Okada, T., Suzuki, N., and Kataoka, T. (1994). Quantitative analysis of mutually competitive binding of human *raf*-1 and yeast adenylyl cyclase to *ras* proteins. *Journal of Biological Chemistry* **269** (33): 20845-20851.

Molyneux, D. H., Pentreath, V., and Doua, F. (1996). African Trypanosomiasis in man. *Manson's Tropical Diseases, 12th edition*: 1171-1196. Edited by G.C. Cook. Publishers - W. B. Saunders Company Limited.

Moodie, S. A., and Wolfman, A. (1994). The 3Rs of life: *Ras, Raf* and growth regulation. *Trends in Genetics* **10** (2): 44-48.

Moores, S. L., Schaber, M. D., Mosser, S. D., Rands, E., O'Hoara, M. B., Garsky, V. M., Marshall, M. S., Pompliano, D. L., and Gibbs, J. B. (1991). Sequence dependence of protein isoprenylation. *Journal of Biological Chemistry* 266 (22): 14603-14610.

Morishita, T., Mitsuzawa, H., Nakafuku, M., Nakamura, S., Hattori, S., and Anraku, Y. (1995). Requirement of *Saccharomyces cerevisiae ras* for completion of mitosis. *Science* **270**: 1213-1215.

Moro, A., Ruiz-Cabello, F., Fernandez-Cano, A., Stock, R. P., and Gonzalez, A. (1995). Secretion by *Trypanosoma cruzi* of a peptidyl-prolyl cis-trans isomerase involved in cell infection. *EMBO Journal* **14:** 2483-2490.

Nantulya, V. M. (1991). Molecular diagnosis of parasites. *Experientia* **47**: 142-145.

Nantulya, V. M., Doua, F., and Molisho, S. (1992a). Diagnosis of *Trypanosoma brucei* sleeping sickness using an antigen detection enzyme-linked immunosorbent assay. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **86**: 42-45.

Nantulya, V. M., Lindqvist, K. J., Stevenson, P., and Mwangi, E. K. (1992b). Application of monoclonal antibody based antigen detection enzyme-linkes immunosorbent assay (antigen ELISA) for field diagnosis of bovine trypanosomiasis at Nguruman, Kenya. *Annals of Tropical Medicine and Parasitology* **86** (No.3): 225-230.

Nassar, N., Horn, G., Herrmann, C., Scherer, A., McCormick, F., and Wittinghofer, A. (1995). *Ras* and *rap* opposition. *Nature Structural Biology* **2** (7): 536.

Neer, E. J. (1995). Heterotrimeric G proteins: Organizers of transmembrane signals. *Cell* **80**: 249-257.

Nei, M. (1987). Phylogenetic trees. *Molecular evolutionary genetics* 287-326. Columbia University press, New York.

Neuman-Silberberg, F. S., Bhattacharya, S., and Broach, J. R. (1995). Nutrient availability and the *ras*/cyclic AMP pathway both induce expression of ribosomal protein genes in *Saccharomyces cerevisiae* but by different mechanisms. *Molecular and Cellular Biology*, 3187-3196.

Nishida, E., and Gotoh, Y. (1993). The MAP kinase cascade is essential for diverse signal transduction pathways. *Trends in Biochemical Sciences* **18**: 128-131.

Ogueta, S. B., Solari, A., and Tellez-Inon, M. T. (1994). *Trypanosoma cruzi* epimastigote forms possess a Ca2+-calmodulin dependent protein kinase. *FEBS Letters* **337**: 293-297.

Olsson, T., Bakhiet, M., Hojeberg, B., Ljungdahl, A., Edlund, C., Andersson, G., Ekre, H. P., Fung Leung, W. P., Mak, T., Wigell, H., Fiszer, U., and Kristensson, K. (1993). CD8 is critically involved in lymphocyte activation by a *T. brucei brucei*-released molecule. *Cell*, 72, 715-727.

Opperdoes, F. R., and Borst, P. (1977). Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei*: the glycosome. *FEBS Letters* **80**: 360-364.

Oz, H. S., Huang, H., Wittner, M., Tanowitz, H. B., Bilezikian, J. P., and Morris, S. A. (1994). Evidence for guanosine triphosphate-binding proteins in *Trypanosoma cruzi*. *American Journal of Tropical Medicine and Hygiene* **50:** 620-631.

Page, A. P., MacNiven, K., and Hengartner, M. O. (1996). Cloning and biochemical characterization of the cyclophilin homologues from the free-living nematode *Caenorhabditis elegans*. *Biochemical Journal* **317**: 179-185.

Pai, E. F., Kabsch, W., Krengel, U., Holmes, K., John, J., and Wittinghofer, A. (1989). Structure of the guanine nucleotide binding domain of the Ha-*ras* oncogene product p21 in the triphosphate conformation. *Nature* **341**: 209-214.

Pai, E. F., Krengel, U., Petsko, G. A., Goody, R. S., Kabsch, W., and Wittinghofer, A. (1990). Refined crystal structure of the triphosphate conformation of H-*ras* p21 at a 1.35 A resolution: implications for the mechanism of GTP hydrolysis. *EMBO Journal* **9**: 2351-2359.

Paindavoine, P., Rolin, S., Van Assel, S., Geuskens, M., Jauniaux, J. C., Dinsart, C., Huet, G., and Pays, E. (1992). A gene from the VSG expression site encodes one of several transmembrane adenylate cyclases located on the flagellum of *Trypanosoma brucei*. *Molecular and Cellular Biology* **12**: 1218-1225.

Parsons, M., and Carter, V. (1993). Protein kinases in divergent eukaryotes: identification of protein kinase activities regulated during trypanosome development. *Proceedings of the National Academy of Sciences, USA* **90**: 2656-2660.

Parsons, M., Valentine, M., Deans, J., Schieven, G. L., and Ledbetter, J. A. (1991). Distinct patterns of tyrosine phosphorylation during the life cycle of *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **45**: 241-248.

Pays, E. (1988). Expression of variant-specific antigen genes in African Trypanosomiasis. *Biology of the Cell* **64**: 121-130.

Pays, E., Rolin, S., and Magez, S. (1997). Cell signalling in trypanosomatids. *Trypanosomiasis and Leishmaniasis: Biology and control* 119-225. Edited by G. Hide, J.C. Mottram, G.H. Coombs and P.H. Holmes.

Pearce, E. J., Basch, P. F., and Sher, A. (1986). Evidence that the reduced surface antigenicity of developing *Schistosoma mansoni* schistosomula is due to antigen shedding rather than host molecule acquisition. *Parasite Immunology* 8: 79-89.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences, USA* **85**: 2444-2448.

Pelech, S. L., and Sanghera, J., S. (1992). Mitogen-activated protein kinases: versatile transducers for cell signaling. *Trends in Biochemical Sciences* **17**: 233-238.

Pepin, J., and Milord, F. (1994). The treatment of Human African Trypanosomiasis. *Advances in Parasitology* **33**: 1-47.

Pitt, G., Milona, N., Borleis, J., Lin, K., Reed, R., and Devreotes, P. (1992). Structurally distinct and stage-specific adenylyl cyclase genes play different roles in Dictyostelium development. *Cell* **69**: 305-315.

Porfiri, E., Evans, T., Bollag, G., Clark, R., and Hancock, J. F. (1995). Purification of baculovirus-expressed recombinant *ras* and *rap* proteins. *Methods in Enzymology* **255**: 13-21.

Racagni, G., Garcia-de Lema, M., Domenech, C. E., and Machano-de-Domenech, E. E. (1992). Phospholipids in *Trypanosoma cruzi*: phosphoinositide composition and turnover. *Lipids* **27**: 275-278.

Rangel-Aldao, R., Tiana, F., Fernandez, V., Gomach, G., Abate, T., and Montereano, R. (1988). Cyclic AMP as an inducer of the cell differentiation of *Trypanosoma cruzi*. *Biochemistry International* **17**: 337-344.

Raper, J., Nussenzweig, V., and Tomlinson, S. (1996). The main lytic factor of *Trypanosoma brucei brucei* in normal human serum is not high density lipoprotein. *Journal of Experimental Medicine* **183**: 1023-1029.

Rapp, U. R., Heidecker, G., Huleihel, M., Cleveland, J. L., Choi, W. C., Pawson, T., Ihle, J. N., and Anderson, W. B. (1988). *Raf* family serine protein-kinases in mitogen signal transduction. *Cold Spring Harbour Symposia on Quantitative Biology* **53** (11): 173-184.

Reed, S. L., Fierer, A. S., Goddard, D. R., Colmerauer, M. E. M., and Davis, C. E. (1985). Effect of theophylline on differentiation in *Trypanosoma brucei*. *Infection and Immunity* **49**: 844-847.

Reiner, S. L. (1994). Parasites and T helper cell development; some insights. *Parasitology Today* **10** (12):.

Reinstein, J., Schlichting, I., and Wittinghofer, A. (1990). Structurally and catalytically important residues in the phosphate binding loop of adenylate kinase of *Escherichia coli*. *Biochemistry* **29**: 7451-7459.

Rifkin, M. R. (1978). Identification of the trypanocidal factor in normal human serum: high density lipoprotein. *Proceedings of the National Academy of Sciences, USA* 75: 3450-3454.

Roditi, I., and Pearson, A. (1990). The procyclin coat of African trypanosomes. *Parasitology Today* **6**: 79-82.

Roditi, I., Schwarz, H., Pearson, T. W., Beechcroft, R. P., Liu, M. K., Richardson, J. P., Buhring, H. J., Pleiss, J., Bulow, R., Williams, R. O., and Overath, P. (1989). Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of *Trypanosoma brucei*. *Journal of Cell Biology* **108**: 737-746.

Rolin, S., Halleux, S., Van Sande, J., Dumont, J., Pays, E., and Steinert, M. (1990). Stage-specific adenylate cyclase activity in *Trypanosoma brucei*. *Experimental Parasitology* **71**: 350-352.

Rolin, S., Hanocq-Quertier, Paturiaux-Hanocq, F., Nolan, D., Salmon, D., Webb, H., Carrington, M., Voorheis, P., and Pays, E. (1996). Simultaneous but independent activation of adenylate cyclase and

glycophosphatidylinositol-phospholipase c under stress conditions in *Trypanosoma brucei*. *Journal of Biological Chemistry* **271**.

Rolin, S., Paindavoine, P., Hanocq-Quertier, J., Hanocq, F., Le Ray, D., Overath, P., and Pays, E. (1993). Transient adenylate cyclase activation accompanies differentiation of *Trypanosoma brucei* from bloodstream to procyclic forms. *Molecular and Biochemical Parasitology* **61**: 115-126.

Ross, D. T., Raibaud, A., Florent, I. C., Sather, S., Gross, M. K., Storm, D. R., and Eisen, H. (1991). The trypansome VSG expression site encodes adenylate cyclase and leucine-rich putative regulatory gene. *EMBO Journal* **10**: 2047-2053.

Ruben, L., Haghighat, N., and Campbell, A. (1990). Cyclical differentiation of *Trypanosoma brucei* involved changes in the cellular complement of calmodulin-binding proteins. *Experimental Parasitology* **70**: 144-153.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406-425.

Salmon, D., Geuskens, M., Hanocq-Quertier, J., Nolan, D., Reuben, L., and Pays, E. (1994). A novel heterodimeric transferrin receptor encoded by a pair of VSG expression site-associated genes in *Trypanosoma brucei*. *Cell* **78**: 75-86.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning. *Cold Spring Harbour Laboratory Press*.

Sanchez, M. A., Zeoli, D., Klamo, E. M., Kavanaugh, M. P., and Landfear, S. M. (1995). A family of putative receptor-adenylate cyclases from *Leishmania donovani*. *Journal of Biological Chemistry* **270**: 17551-17558.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences (USA)* **74**: 5463-5467

Schechter, P. J., and Sjoerdsma, A. (1986). Difloromethylornithine in the treatment of African Trypanosomiasis. *Parasitology Today* **2** (8): 223-224.

Schell, D., Borowy, N. K., and Overath, P. (1991). Transferrin is a growth factor of the bloodstream form of *Trypanosoma brucei*. *Parasitology Research* **77**: 558-560.

Schell, D., Evers, R., Preis, D., Ziegelbauer, K., Kiefer, H., Lottspeich, F., Cornelissen, A. W. C. A., and Overath, P. (1990). A transferrin-binding protein of *Trypanosoma brucei* is encodes by one of the genes in the variant surface glycoprotein gene expression site. *EMBO Journal* **10**: 1061-1066.

Schenkman, S., and Eichinger, D. (1993). *Trypanosoma cruzi* trans-sialidase and cell invasion. *Parasitology Today* **9**: 218-225.

Schlegel, M. (1994). Molecular phylogeny of eukaryotes. *Trends in Ecology and Evolution* **9** (9): 330-335.

Schoeniger, M., and von Haesler, A. (1993). A simple method to improve reliability of tree reconstructions. *Molecular Biology and Evolution* **10**: 471-483.

Screaton, G. R., Bangham, C. R. M., and Bell, J. I. (1993). Direct sequencing of single primer PCR products: a rapid method to achieve short chromosomal walks. *Nucleic Acids Research* **21** (9): 2263-2264.

Simarro, P. P., and Asumu, P. N. (1996). Gambian trypanosomiasis and synergism between melarsoprol and eflornithine; first case report. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **90**: 315.

Sommer, J. M., and Wang, C. C. (1994). Targeting proteins to the glycosomes of African trypanosomes. *Annual Reviews of Microbiology* **48**: 105-138.

Sourdis, J., and Nei, M. (1988). Relative efficiencies of the maximum parsimony and distance matrix methods in obtaining the correct phylogenetic tree. *Molecular Biology and Evolution* **5** (3): 298-311.

Sprang, S. R. (1995). How *ras* works: structure of a *rap-raf* complex. *Structure* **3** (No.7): 641-643.

Stark, K. R., and James, A. A. (1996). Anticoagulants in vector arthropods. *Parasitology Today* **12** (No.11): 430-437.

Sternberg, J. M., and McGuigan, F. (1994). *Trypanosoma brucei*: mammalian epidermal growth factor promotes the growth of the African Trypanosome bloodstream form. *Experimental Parasitology* **78**: 422-424.

Sternberg, J., and Tait, A. (1990). Genetic exchange in African trypanosomes. *Trends in Genetics* **6** (10): 317.

Sternweis, P. C., and Robishaw, J. D. (1984). Isolation of two proteins with high affinity for guanine nucleotides from membranes of bovine brain. *Journal of Biological Chemistry* **259** (22): 13806-13813.

Steverding, D., Stierhof, Y. D., Chaudhri, M., Ligtenberg, M., Schell, D., Beck-Sickinger, A. G., and Overath, P. (1994). ESAG 6 and 7 products of *Trypanosoma brucei* form a transferrin binding complex. *European Journal of Cell Biology* **64**: 78-87.

Strimmer, K., and von Haesler, A. (1996). Quartet Puzzling - A quartet maximum-likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution* **13** (7): 964-969.

Sun, J., Kale, S. P., Childress, A. M., Pinswasdi, C., and Jazwinski, S. M. (1994). Divergent roles of *ras*1 and *ras*2 in yeast longevity. *Journal of Biological Chemistry* **269** (28): 18638-18645.

Swindle, J., and Tait, A. (1995). Trypanosomatid genetics. *Molecular Biology of Parasitic Protozoa* 6-34.

Sztein, M. B., and Kierszenbaum, F. (1993). Mechanisms of development of immunosuppression during Trypanosoma infections. *Parasitology Today* **9** (11): 424-428.

Tachado, S., and Schofield, L. (1994). Glycophosphatidylinositol toxin of *Trypanosoma brucei* regulates IL-1a and TNF-α expression in macrophages by protein tyrosine kinase mediated signal transduction. *Biochemical and Biophysical Research Communications* **205**: 984-991.

Tait, A. (1983). Sexual processes in the kinetoplastida. *Parasitology* **86**: 29-57.

Tait, A., Babiker, E. A., and LeRay, D. (1984). Enzyme variation in *Trypanosoma brucei* spp. 1. Evidence for the subspeciation of *Trypanosoma brucei gambiense*. *Parasitology* **89**: 311-326.

Tait, A., and Turner, C. M. R. (1990). Genetic exchange in *Trypanosoma brucei*. *Parasitology Today* **6** (3): 70-75.

Tang, W. J., and Gilman, A. G. (1992). Type specific regulation of adenylyl cyclase by G-protein beta-gamma (βγ) subunits. *Science* **254** (5037): 1500-1503.

Tateno, Y., Takezaki, N., and Nei, M. (1994). Relative efficiencies of the maximum-likelihood, neighbor-joining and maximum parsimony methods when substitution rate varies with site. *Molecular Biology and Evolution* **11** (2): 261-277.

Thevelein, J. M., Van Aelst, L., Durnez, P., and Hofmann, S. (1991). The signal transduction pathway upstream of CDC25 - *ras* - adenylate cyclase in the yeast Saccharomyces cerevisiae and its relationship to nutrient control of cell cycle progression. *The Superfamily of ras-related genes* 57-66. Edited by D.A Spandidos, Plenum press, New York.

Tobin, J. F., Laban, A., and Wirth, D. F. (1991). Homologous recombination in Leishmania enrietti. *Proceedings of the National Academy of Sciences, USA* **88**: 804-808.

Tokiwa, G., Tyers, M., Volpe, T., and Futcher, B. (1994). Inhibition of G1 cyclin activity by the *ras*/cAMP pathway in yeast. *Nature* **371**: 342-345.

Tomlinson, S., and *Rap*er, J. (1996). The lysis of *Trypanosoma brucei brucei* by human serum. *Nature Biotechnology* **14**: 717-721.

Torruella, M., Flawia, M., Eisenschlos, C., Molina y Vedia., Rubinstein, C. P., and Torres, H. N. (1986). *Trypanosoma cruzi* adenylate cyclase activity. Purification and characterization. *Biochemical Journal,* **234**: 145-150.

Truc, P., Aerts, D., McNamara, J. J., Claes, Y., Allingham, R., Le Ray, D., and Godfrey, D. G. (1992a). Direct isolation in vitro of *Trypanosoma brucei* from man and other animals and its potential value for the diagnosis of Gambian trypanosomiasis. *Trans. Roy. Soc. of Trop. Med & Hyg.* **86**: 627-629.

Truc, P., Aerts, D., McNamara, J. J., Claes, Y., Allingham, R., Le Ray, D., and Godfrey, D. G. (1992b). Isolation in vitro of *Trypanosoma brucei* from man and other animals and its potential value for the diagnosis of *gambiense* trypanosomiasis. *Tropical Royal Society of Medicine and Hygiene* **86**: 627-629.

Tschudi, C. (1995). Transcription and RNA Processing in *Trypanosoma brucei. Molecular Approaches to Parasitology* 255-268.

Valencia, A., Chardin, P., Wittinghofer, A., and Sander, S. (1991). The *ras* protein family: Evolutionary tree and role of conserved amino acids. *Biochemistry* **39** (19): 4637-4648.

Van de Peer, Y., Neefs, J., De Rijk, P., and De Wachter, R. (1993). Reconstructing evolution from eukaryotic small-subunit RNA sequences: calibration of the molecular clock. *Journal of Molecular Evolution* **37**: 221-232.

Van Den Abeele, J., Rolin, S., Claes, Y., Le Ray, D., Pays, E., and Coosemans, M. (1995). *Trypanosoma brucei*: stimulation of adenylate cyclase by proventriculus and oesophagus tissue of the tsetse fly, *Glossina morsitans morsitans. Experimental Parasitology* **81**: 618-620.

Van der Ploeg, L. H. T., Cornelissen, A. W. C. A., Barry, J. D., and Borst, P. (1984a). Chromosomes of Kinetoplastida. *EMBO Journal* **3**: 3109-3115.

Vanhamme, L., and Pays, E. (1995). Control of gene expression in trypanosomes. *Microbiological Reviews* 223-240.

Verzotti, E., Geymonat, M., Valetti, F., Lanzetti, L., and Giunta, C. (1994). In the budding yeast *Kluyveromyces marxianus*, adenylate cyclase is regulated by *ras* protein(s) *in vitro. Yeast* **10**: 993-1001.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. *Bri. Med. Bull.* **41** (2): 105-114.

Vickerman, K. (1994). The evolutionary expansion of the trypanosomatid flagellates. *International Journal for Parasitology* **8**: 1317-1331.

Vickerman, K., Tetley, L., Hendry, K. A. K., and Turner, C. M. R. (1988). *Biology of African trypanosomes in the tsetse fly.*

Voorheis, H. P., and Martin, B. R. (1982). Local anaesthetics including benzyl alcohol activate adenylate cyclase in *Trypanosoma brucei* by a calcium-dependent mechanism. *European Journal of Biochemistry* **123**: 371-376.

Waldo, G. L., Evans, T., Fraser, E.D., , Northup, J. K., Martin, M. W., and Harden, T. K. (1987). Identification and purification from bovine brain of a guaninie-nucleotide-binding protein distinct from Gs, Gi and Go. *Biochemical Journal* **246**: 431-439.

Wang, C. C. (1995). Molecular mechanisms and therapeutic approaches to the treatment of African Trypanosomiasis. *Annual Reviews of Pharmacological Toxicology* **35**: 93-127.

Warne, P. H., Viciana, P. R., and Downward, J. (1993). Direct interaction of *ras* and the amino-terminal region of *Raf*-1 in vitro. *Nature* **364**: 352-355.

Webster, P., and Russell, D. G. (1993). The flagellar pocket of trypanosomatids. *Parasitology Today* **9**: 201-206.

Weiden, M., Osheim, Y. N., Beyer, A. L., and Van der Ploeg, L. H. T. (1991). Chromosome structure: DNA nucleotide sequence element of a subset of the minichromosomes of the protozoan *Trypanosoma brucei*. *Molecular and Cellular Biology* **11** (8): 3823-3834.

Welburn, S. C., Dale, C., Ellis, D., Beechcroft, R., and Pearson, T. W. (1996). Apoptosis in procyclic *Trypanosoma brucei rhodesiense* in vitro. *Cell Death and Differentiation* **3**: 229-236.

Werbovetz, K. A., Bacchi, C. J., and Englund, P. T. (1996). Trypanocidal analogs of myristate and myristoyllysophosphatidylcholine. *Molecular and Biochemical Parasitology* **81**: 115-118.

WHO (1986). Epidemiology and control of African Trypanosomiasis. *World Health Organization Technical Report Series* (No.739).

WHO (1994). Tropical Disease Research Highlights. *Twelfth Programme Report UNDP/World Bank/Special Programme for Research & Training in Tropical Diseases (TDR)* 115-123.

Willumsen, B. M., Christensen, A., Hubbert, N. L., Papageorge, A. G., and Lowy, D. R. (1984a). The p21 *ras* C-terminus is required for transformation and membrane association. *Nature* **310**: 583-586.

Willumsen, B. M., Norris, K., Papageorge, A. G., Hubbert, N. L., and Lowy, D. R. (1984b). Harvey murine sarcoma virus p21 *ras* protein: biological and biochemical significance of the cysteine nearest the carboxy terminus. *EMBO Journal* **3** (11): 2581-2585.

Wirtz, E., and Clayton, C. (1995). Inducible gene expression in Trypanosomiasis mediated by prokaryotic repressor. *Science* **268** (5214): 1179-1183.

Wittinghofer, A. (1994). The structure of transducin $G_{\alpha t}$: More to view than just *ras*. *Cell* **76**: 201-204.

Yamashita, T., Yamamoto, K., Kikuchi, A., Kawata, M., Kondo, J., Hishida, T., Teranishi, Y., Shiku, H., and Takai, Y. (1988). Purification and characterization of c-Ki-*ras* p21 from bovine brain crude membranes. *Journal of Biological Chemistry* **263** (32): 17181-17188.

Yang, Z. (1993). Maximum-Likelihood estimation of phylogeny from DAN sequences when substitution rates differ over sites. *Molecular Biology and Evolution* **10** (6): 1396-1401.

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* **39**: 306-314.

Yang, Z., Goldman, N., and Friday, A. (1994). Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Molecular Biology and Evolution* **11** (2): 316-324.

Yin, Z., K.,, Smith, R. J., and Brown, A. J. P. (1996). Multiple signalling pathways trigger the exquisite sensitivity of yeast gluconeogenic messenger RNAs to glucose. *Molecular Microbiology* **20** (4): 751-764.

Zerial, M., and Huber, L. A. (1995). *Guidebook to the small GTPases*. Publishers: Sambrook and Tooze, Oxford University Press.

Zhang, K., Noda, M., Vass, W. C., Papageorge, A. G., and Lowy, D. R. (1990). Identification of small clusters of divergent amino acids that mediate the opposing effects of *ras* and Krev-1. Science **249**: 162-165.

Zhang, X.-F., Settleman, J., Kyriakis, J. M., Takeuchi-Suzuki, E., Elledge, S. J., Marshall, M. S., Bruder, J. T., Rapp, U. R., and Avruch, J. (1993). Normal and oncogenic p21*ras* proteins bind to the amino-terminal regulatory domain of c-*Raf*-1. *Nature* **364**: 308-313.

Zhikang, Y., Smith, R. J., and Brown, A. J. P. (1996). Multiple signalling pathways trigger the exquisite sensitivity of yeast gluconeogenic mRNAs to glucose. *Molecular Microbiology* **20** (4): 751-764.

GLASGOW UNIVERSITY LIBRARY