Cloning and Analysis of a Surface Antigen Gene from Procyclic *Trypanosoma congolense*

A thesis submitted for the Degree of
Doctor of Philosophy at the
University of Glasgow

by

Rosemary Ann Lindsay Bayne

Institute of Genetics
Church Street
Glasgow

Aug. 1992

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

To David & Stephen

The research reported in this thesis is my own original work except where otherwise stated and has not been submitted for any other degree.

CONTENTS

ABBREVIATIONS

Chemicals

Ac	- acetate
Amp	- ampicillin
APS	- ammonium persulphate
ATP	- adenosine triphosphate
BSA	- bovine serum albumin
cDNA	- complementary DNA
CHAPS	- 3-[3-cholamidopropyl-dimethylammonio]-1-propanesulphonate
CIP	- calf intestinal phosphatase
Cm	- chloramphenicol
CNBr	- cyanogen bromide
DEPC	- diethyl pyrocarbonate
dH_2O	- Milli-Q deionized water
DNA	- deoxyribonucleic acid
DNase	- deoxyribonuclease
dNTP	- 2'-deoxy nucleotide
DTT	- dithiothreitol
ECL	- enzyme-linked chemoluminescence
EDTA	- ethylene diamine tetraacetic acid (disodium salt)
EtBr	- ethidium bromide
FCS	- foetal calf serum
FITC	- fluorescein isothiocyanate conjugate
gRNA	- guide RNA
HCl	- hydrochloric acid
HRP	- horse radish peroxidase
IPTG	- isopropyl thiogalactoside
Kan	- kanamycin
kDNA	- kinetoplast DNA
LMP	- low melting point (agarose)
mDNA	- mitochondrial DNA
MEM	- minimal essential medium
MOPS	- morpholinopropanesulphonic acid
mRNA	- messenger RNA
nt	- nucleotide

oligo	- oligonucleotide
$p(dN)_6$	- random hexanucleotides
PBS	- phosphate buffered saline
PEG	- polyethylene glycol
$poly(A)^+$	- polyadenylated
PSG	- phosphate buffered saline glucose
RNA	- ribonucleic acid
rRNA	- ribosomal RNA
RNase	- ribonuclease
RT	- reverse transcriptase
SDS	- sodium dodecyl sulphate
snRNA	- small nuclear RNA
SSC	- standard saline citrate
TEMED	- N,N,N',N' tetramethyl ethylenediamine
Tet	- tetracycline
tris	- tris (hydroxymethyl) amino methane
X-gal	- 5-bromo-4-chloro-3-indolyl-beta-D-galactoside

Measurements

A	- angstrom
bp	- base pair(s)
°C	- degrees centigrade
Ci	- Curie
cm	- centimetre
cpm	- counts per minute
g	- centrifugal force equal to gravitational acceleration
g	- gramme
h	- hour
kb	- kilobase (pairs)
kDa	- kilodaltons
kg	- kilogramme
kV	- kilovolt
l	- litre
min.	- minutes
M	- molar
mA	- milliamp

Mb - megabase pairs
mg - milligramme
ml - millilitre
mM - millimolar
mmol - millimoles
ng - nanogramme
OD_{260} - optical density at 260nm
PCV - packed cell volume
pI - isoelectric point
pmoles - picomoles
p.s.i. - pounds per square inch
sec. - seconds
temp. - temperature
u - units
uCi - microCurie
uF - microfarads
ug - microgramme
ul - microlitre
uM - micromolar
um - micron
V - volts
vol. - volume
v/v - volume/volume
W - watts
w/v - weight/volume
w/w - weight/weight
% - percent (w/v unless otherwise indicated)

Miscellaneous

A6 - ATPase subunit 6
ARP - alanine rich protein
BSF - bloodstream form
CAT - chloramphenicol acetyl transferase
CIP - calf intestinal phosphatase
COII - cytochrome oxidase subunit II
COIII - cytochrome oxidase subunit III

conc.	- concentrated
CYb	- apocytochrome b
EATRO	- East African Trypanosomiasis Research Organisation
ELISA	- enzyme-linked immunosorbent assay
ESAG	- expression site associated gene
Exo III	- exonuclease III
GPI	- glycosyl phosphatidyl inositol
GPI-PLC	- GPI-specific phospholipase C
GST	- glutathione-S-transferase
GUGM	- Glasgow University Genetics Monoclonal
IFA	- indirect immunofluorescence assay
M13 RF	- M13 replicative form
mab	- monoclonal antibody
medRNA	- mini exon derived RNA
MMLV RT	- Moloney murine leukaemia virus reverse transcriptase
m.o.i.	- multiplicity of infection
MURF1	- mitochondrial unidentified reading frame 1
M.Wt.	- molecular weight
ND1	- NADH Dehydrogenase subunit 1
ND4	- NADH Dehydrogenase subunit 4
ND5	- NADH Dehydrogenase subunit 5
ND7	- NADH Dehydrogenase subunit 7
ORF	- open reading frame
PAG	- procyclin associated gene
PAGE	- polyacrylamide gel electrophoresis
PARP	- procyclic acidic repetitive protein
PCR	- polymerase chain reaction
PFG	- pulsed field gel electrophoresis
p.f.u.	- plaque-forming units
pol I/II/III	- RNA Polymerase I/II/III
RF	- (M13) replicative form
RFLP	- restriction fragment length polymorphism
"Rxn-C"	- reaction mix minus dCTP
SL	- spliced leader
TBS	- tris buffered saline
TIM	- triose phosphate isomerase
TREU	- Trypanosomiasis Research Edinburgh University

TUTase - terminal uridinyl transferase
UTR - untranslated region
UV - ultraviolet light
VSG - variant surface glycoprotein
YNat - Yale Nannomonas antigen type

ACKNOWLEDGEMENTS

First of all I would like to thank my supervisor, Dave Barry, for his help and encouragement and giving me the opportunity to finish my studies when all seemed lost. Then there is everyone who has passed through the Barry and WUMP labs over the past seven years, and especially Jeremy, for their help, advice and friendship.

I would also like to thank Lawrence Tetley for the EM work, Carole Ross for providing cultures, Isobel Roditi and Terry Pearson for communicating unpublished work, the media ladies for providing materials and everyone in Genetics for making my time in Glasgow an enjoyable one.

Of course, none of this would have been possible without David who provided love, encouragement and financial assistance throughout or my mum and David's mum for looking after Stephen so that I could write this thesis - to them I will be eternally grateful. And then there's Stephen who slowed down the writing up of this work but made it all worthwhile.

This work was supported initially by a studentship from the Medical Research Council and subsequently by a project grant from the Overseas Development Administration.

SUMMARY

The presence of the variant surface glycoprotein coat, and its modulation by the process of antigenic variation have made it unlikely that a vaccine could be developed against mammalian stages of African trypanosomes. However, except for the metacyclic stage which is preadapted to life in the mammalian host, insect stages of these parasites do not possess the variant surface glycoprotein coat and display a common set of antigens on their surface dominated, in *T.brucei*, by a surface coat composed of the glycoprotein procyclin. If animals could be vaccinated against these common antigens then tsetse flies feeding on such vaccinated animals would ingest antibody with their bloodmeal which might inhibit development of trypanosome infections in the fly and therefore block transmission. Of a set of monoclonal antibodies which had been raised against living procyclic *Trypanosoma congolense*, one group appeared to be directed against molecules on the trypanosome surface and were strongly agglutinating and trypanocidal. Each monoclonal antibody in that group detected the same diffuse bands on western blots of SDS-PAGE of trypanosome lysates, indicating that they were against the same immunodominant antigen. The antigen has several unusual properties in common with procyclin and the work presented here was directed towards identifying the gene for this antigen.

Heterologous probing of a genomic library with the procyclin cDNA from *T.brucei* indicated that *T.congolense* has sequences which share a degree of homology to the repetitive elements in the procyclin gene but these sequences are not expressed and there is no true homologue of this gene in *T.congolense*.

The monoclonal antibodies did not detect the antigen gene from a cDNA expression library, probably because at least one of them appears to be directed against a carbohydrate epitope. However, differential screening of a cDNA library with first strand cDNA from procyclic and bloodstream stages detected several cDNAs, one of which contained an open reading frame with a high degree of homology to two cyanogen bromide peptide sequences derived from a Kilifi-type *T.congolense* surface antigen isolated by Beecroft *et al* (manuscript in preparation). Apart from a size difference defined by migration on SDS-PAGE, this antigen has identical properties to that detected by the set of monoclonal antibodies. The cDNA has an open reading frame coding for a protein of 256 amino acids which is rich in alanine and acidic residues. There is no *N*-glycosylation signal and no obvious

signal peptide but the amino terminus is hydrophobic and there is a potential signal at the carboxy terminus for the addition of a glycosyl phosphatidylinositol tail which could anchor the protein in the membrane.

This protein, which has been called ARP (for Alanine Rich Protein), has been expressed as a fusion protein in *E.coli* and used to raise antisera in rats and mice. These antisera label procyclic and epimastigote stages of *T.congolense* in indirect immunofluorescence and label the surface of procyclic cells in immunogold electron microscopy. In western blots of procyclic trypanosome lysates enriched for membranes they identify diffuse bands similar to those detected by the monoclonal antibodies.

Several other cDNA clones were isolated in the differential screen but none of these, except perhaps cDNAP1, proved to be genuinely stage-specific. cDNAP1 appears to be transcribed from telomeres but in which direction is not evident because it does not possess a poly(A) tail, although probing of northern blots suggests that some of the homologous transcripts are polyadenylated. cDNAs P7 and P8 appear to be single copy sequences in the trypanosome genome but no function can be ascribed to them as they do not contain open reading frames. cDNAPE encodes the ribosomal protein L29, is highly homologous to the the sequence of L29 from other species and may encode a cycloheximide resistant form of this protein in the stock of trypanosomes from which it was isolated. In conclusion, the apparent lack of stage-specific control at the level of transcription in trypanosomes has pushed the differential screening process to its limits and yielded sequences which are largely artifactual.

The ARP cDNA detects several bands on Southern blots of total *T.congolense* genomic DNA which represent multiple copies of the ARP gene. Two of these copies are linked in tandem within one clone isolated from a genomic library. Northern blot analysis reveals that, despite the cDNA having been isolated by differential screening, transcription of the ARP genes is not reproducibly stage-specific. Since ARP does not appear to be present in bloodstream stages, post-transcriptional controls must play a part in stage-specific expression. Further analysis of this gene and its protein and comparison with the properties of procyclin in *T.brucei* should facilitate elucidation of the function of these proteins and provide insight into the interaction of the parasite with its tsetse fly vector.

CHAPTER 1

INTRODUCTION

1.1 Trypanosome Species

The African trypanosome is the causative agent of Sleeping Sickness in man and Nagana in his domestic animals. While the human disease does not account for many deaths each year, the presence of the parasite throughout the tsetse belt of Africa prevents the farming of livestock in many of the most fertile regions of the continent. Control of this disease is therefore of great economic importance and in recent years much effort has been put into understanding the biology of the parasite with the hope of eventually developing therapeutic agents.

The Salivarian trypanosomes, so-called because they are transmitted via the saliva of their insect vectors, are members of the order Kinetoplastida and can be divided into three subgenera: *Trypanozoon*, *Nannomonas* and *Duttonella* (Hoare, 1972). The most widely studied of these is *Trypanosoma brucei* which is a member of *Trypanozoon* and is split into three subspecies - *T.b.rhodesiense* which causes acute sleeping sickness in humans and other animals, *T.b.gambiense* which causes the chronic form of the disease and *T.b.brucei* which is a cattle parasite and appears not to be infective to man. *Trypanosoma congolense*, which is a member of *Nannomonas*, is the main cattle trypanosome and *T.vivax*, another parasite of domestic livestock, is a member of the *Duttonella*. All of the above mentioned parasites are transmitted between their mammalian hosts by *Glossina*, the tsetse fly, and are therefore mainly restricted to the regions where this insect is present. However, some Salivarian trypanosomes have been spread further afield, including *T.equiperdum* which is a venereally spread parasite of horses and *T.evansi* which is transmitted by purely mechanical means. *T.vivax* also exists in South America where again it is transmitted mechanically.

1.2 The Life Cycle

The life cycles of the salivarian trypanosomes are complex, with growth stages and non-growing forms which are probably preadapted to life in their next host (reviewed by Vickerman, 1985). The life cycle of *T.brucei* is shown in Fig.1.1 and that of the other trypanosomes is quite similar although epimastigote and metacyclic stages of Nannomonads develop in the mouthparts of the fly rather than the salivary glands and *T.vivax* does not enter the fly's gut at all but develops entirely in the mouthparts. *T.brucei* displays pleomorphism in the mammalian bloodstream, with a long slender form which rapidly divides and intermediate and

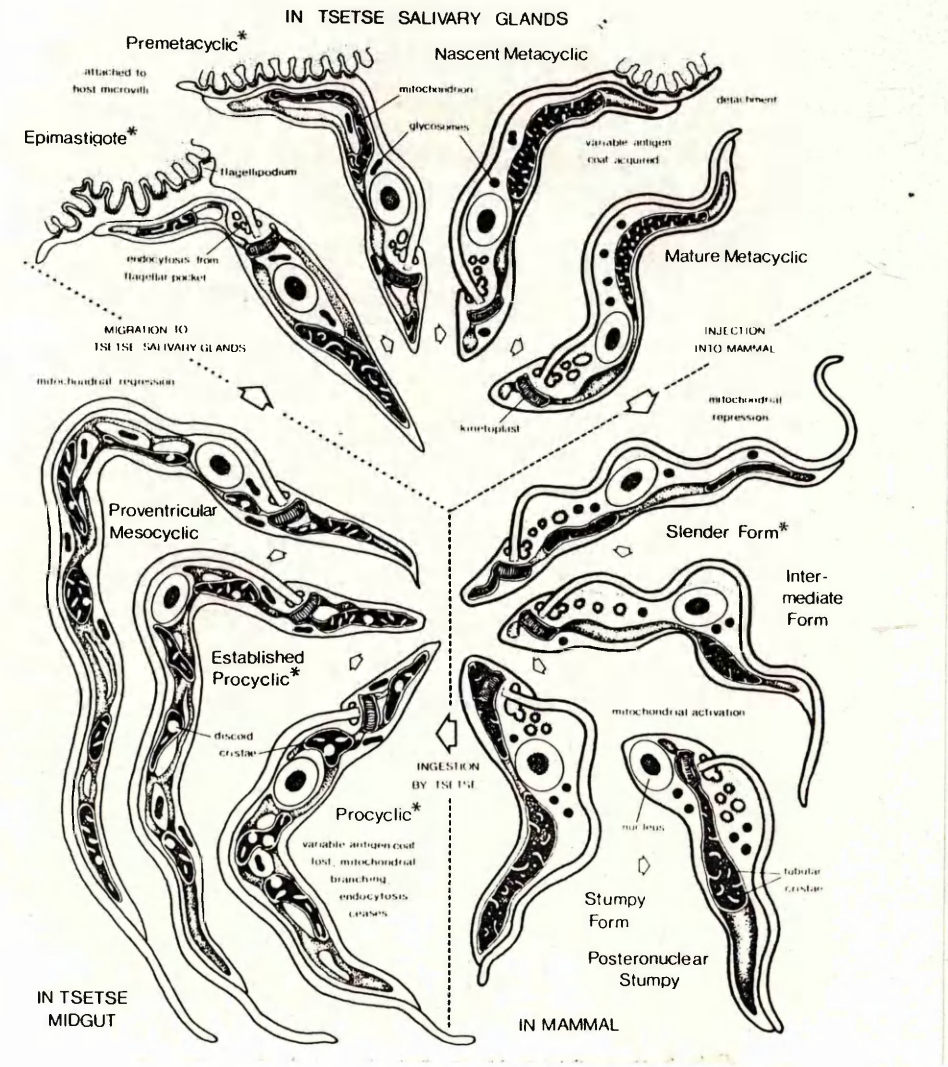


Fig. 1.1 The Lifecycle of *Trypanosoma brucei brucei* (from Vickerman, 1985)

non-dividing short stumpy forms. The short stumpy trypomastigotes are believed to be responsible for infection of the tsetse fly (Wijers & Willet, 1960; Vickerman & Barry, 1982). When the parasites enter the fly midgut they transform to the procyclic form which is indistinguishable morphologically, biochemically and immunologically from parasites grown *in vitro* in culture at 26°C. The infection in the fly develops by the multiplication of the procyclic form which migrates to the endoperitrophic space within the lumen of the peritrophic membrane which is concentric with the gut wall. At this stage, the parasites are thought to be sensitive to the presence of trypanocidal lectins produced by the tsetse fly (Maudlin & Welburn, 1987; Maudlin & Welburn, 1988). One characterized lectin is produced in response to the bloodmeal (Welburn *et al*, 1989) and can be sequestered by high levels of glucosamine administered with the bloodmeal (Welburn & Maudlin, 1989). Glucosamine is a possible product of the chitinase activity produced by endosymbiotic rickettsia-like organisms (RLOs) and their presence has been associated with the trypanosome susceptibility of tsetse flies (Baker *et al*, 1990; Maudlin *et al*, 1990; Welburn & Maudlin, 1991). This same lectin is however later required for maturation of established midgut infections (Welburn & Maudlin, 1989). Maturation is dependent on a signal from the fly and not on the number of cell divisions undergone by the trypanosome. A second lectin, present in the haemolymph, is essential to complete the maturation process (Welburn & Maudlin, 1990). From the endoperitrophic space, the parasites penetrate the proventriculus and then migrate to the salivary glands (in the case of *T.brucei*) or the mouthparts (in the case of *T.congolense*). Here the parasites differentiate into the epimastigote form which, depending on the trypanosome species, attaches to the salivary gland or mouthpart walls via flagellipodia. The epimastigote undergoes multiple divisions before metacyclogenesis to form the free, non-dividing metacyclic stage which is preadapted to life in the mammalian bloodstream and is responsible for infection of the new host.

1.3 The Variant Surface Glycoprotein

The African trypanosome evades its host's immune response by intermittently changing its surface coat in a process known as antigenic variation (Vickerman, 1969; reviewed by Barry, 1989). The coat is composed of a single moiety - the Variant Surface Glycoprotein (VSG; Cross, 1975) - and antigenic variation occurs by switching on and off different copies of the VSG genes, of which there are of the

order of 1000 different copies in the trypanosome genome (Van der Ploeg *et al*, 1982; reviewed by Pays, 1989). The VSG coat is an electron-dense layer of some 10^7 molecules (Turner *et al*, 1984) which covers the entire parasite surface. The role of the VSG in the mammalian bloodstream is twofold: to prevent lysis of the parasite by non-specific attack via the alternative complement pathway (Ferrante & Allison, 1983a) or by opsonization and macrophage uptake (Mosser & Roberts, 1982; Ferrante & Allison, 1983b); and to protect the other invariant surface molecules from specific antibody attack by acting as a replaceable surface.

The VSG is composed of two domains - a conserved C-terminal domain which is anchored to the membrane and a variable N-terminal domain which is exposed on the surface (Cross & Johnson, 1976). Most VSGs are distinct at the primary structure level but x-ray crystallographic data from two VSGs suggests that even those VSGs which do not share much homology at the primary sequence level have very similar 3-dimensional structures (Metcalf *et al*, 1986). This suggests a constraint on their structure, required to hold the coat together and prevent access of antibodies. There are families of VSG genes in the genome which are very similar in primary amino acid sequence but they are nonetheless immunologically distinct (Olafson *et al*, 1984; Pays *et al*, 1983; Hoeijmakers *et al*, 1980). The VSG sits in the membrane as a homo-dimer (Borst, 1986) and is attached via a glycosyl phosphatidyl inositol (GPI) tail which is added posttranslationally, after the carboxy-terminus of the nascent protein is cleaved off (Ferguson & Cross, 1984; Ferguson *et al*, 1985). A GPI-specific phospholipase C activity (GPI-PLC) has been found in bloodstream trypanosomes (Cardoso de Almeida & Turner, 1983; Ferguson *et al*, 1985; Hereld *et al*, 1986; Fox *et al*, 1986) but its function is unclear. It has been proposed that this enzyme might be involved in turnover of the VSG (Ferguson & Williams, 1988), either during growth or at a switch, but its location on the cytoplasmic side of intracellular membranes but not the plasma membrane (Bulow *et al*, 1989) makes it difficult to imagine how it could interact with the VSG (Carrington *et al*, 1991).

1.4 Procyclin

The VSG coat present on bloodstream forms of the parasite is rapidly lost on ingestion by the tsetse fly (Vickerman, 1969) and is not regained until development of the metacyclic stage which is preadapted to life in the mammalian bloodstream.

Loss of the VSG exposes a set of at least 25 conserved antigens on the procyclic surface (Seed, 1964; Gardiner *et al*, 1983), many of which are also present on bloodstream stages but remain masked by the VSG. By differential screening of cDNA libraries, Roditi *et al* (1987) and Mowatt & Clayton (1987) independently identified a gene whose protein product appeared to be expressed specifically in insect stages of *T.brucei*. The gene encodes a protein of predicted molecular weight about 14 kDa which has a cleavable signal peptide, a single *N*-glycosylation site and which is located on the procyclic surface, is abundant and immunodominant. Approximately 6 X 10^6 molecules of this protein, called procyclin or PARP (Procyclic Acidic Repetitive Protein), are present on the surface of the parasite (Clayton & Mowatt, 1989) and *in vitro* they start to appear on the surface as soon as the VSG starts to disappear (Roditi *et al*, 1989a; Ziegelbauer *et al*, 1990), disappearing again when the metacyclic VSGs start to be expressed. Despite its predicted size, procyclin migrates as a diffuse band of 45-60 kDa on SDS-PAGE gels and does not stain with any conventional protein stain except Stains-All (Clayton & Mowatt, 1989). The protein is highly acidic and has a mature carboxy-terminal domain composed of a (GluPro) dipeptide repeat which varies in size between different copies of the gene. One gene copy contains a pentapeptide repeat (GlyProGluGluThr) in place of the dipeptide repeat (Mowatt *et al*, 1989) but it is not known whether this gene product is functional. The variation in length of the repeat region probably in part accounts for the heterogeneous size of the protein in gels but glycosylation and the high proline content are also predicted to play a part. The function of procyclin is unclear but it has a highly unusual shape - a predicted rod-like structure protruding from the membrane with a globular domain at its end - and the whole molecule may be able to flip-flop from side to side (Roditi *et al*, 1989a; Roditi *et al*, 1989b). The rod-like structure is formed by the series of (GluPro) dipeptide repeats and, like the VSG, the protein is anchored to the membrane via a GPI tail (Clayton & Mowatt, 1989) although this tail is not able to be cleaved by the bloodstream-specific GPI-PLC and probably contains acylinositol and no myristic acid (Field *et al*, 1991a,b). The different forms of the glycolipid anchor may have functional implications for the differential control of VSG and procyclin membrane insertion in the lifecycle (Ferguson *et al*, 1991).

1.5 The Trypanosome Genome

The trypanosome genome is considered to be diploid with respect to housekeeping genes (Tait, 1980; Gibson *et al*, 1985), although each VSG gene is thought to be haploid (Borst *et al*, 1984; Gibson *et al*, 1985). The diploid genome size is about 70Mb (Borst *et al*, 1982) and composed of several chromosomes which do not condense fully at any stage in the cell cycle (Van der Ploeg *et al*, 1984a). Visualization of these chromosomes has only been possible through the technique of Pulsed Field Gel Electrophoresis (PFG) (Schwartz & Cantor, 1984) and it has been found that they are composed of 4 different size classes (Van der Ploeg *et al*, 1984a). The smallest size class is the minichromosomes which range in size from 50-150kb and are thought to exist solely to increase the telomeric VSG repertoire as they otherwise contain only repetitive DNA (Borst, 1986) and are not present in trypanosome species which do not display antigenic variation, or in some stocks of the more primitive *T.vivax* (Borst *et al*, 1984; Van der Ploeg *et al*, 1984b). The other size classes have been further resolved by improved separation conditions (Van der Ploeg *et al*, 1989) and in one stock of *T.brucei* are composed of five intermediate sized chromosomes of 200-430kb, nine chromosomes in the size range 680kb to 3Mb and four chromosomes of 3 to 5.7Mb that have only recently been resolved from the slot of the gel. This accounts for at least 80% of the trypanosome genome. Some material still remains in the slot and may either represent further, even larger, chromosome-length DNA molecules which are still too large to be resolved or a non-specific trapping by the kinetoplast DNA network which cannot enter the gel matrix. Housekeeping genes have been shown to be present on all of the larger chromosome size classes (Van der Ploeg *et al*, 1984a; Gibson & Borst, 1986) and their chromosomal location has been used to show the existence of homologous chromosomes despite size differences (Gottesdiener *et al*, 1990).

Genetic exchange in trypanosomes has been shown to occur in the tsetse fly (Tait, 1980; Jenni *et al*, 1986; Paindavoine *et al*, 1986). Iso-enzyme and RFLP markers have been used to identify novel combinations in metacyclic clones derived from tsetse probes (Sternberg *et al*, 1988; Tait *et al*, 1989). This process does not appear to be obligatory and exchange is able to occur between all combinations of stocks suggesting the lack of a mating type of the sort found in yeast (Turner *et al*, 1990).

A large number of genes in trypanosomes are present in multiple copies (Osinga *et al*, 1985; Ben Amar *et al*, 1988; Clayton, 1985; Michels *et al*, 1986;

Mowatt & Clayton, 1987; Tschudi & Ullu, 1988). These multiple copies are found in one or more tandem arrays with interstitial non-coding spacer regions (Thomashow *et al*, 1983; Ben Amar *et al*, 1988) and the whole array is often transcribed from a single promoter at the 5'-end of the cluster (Imboden *et al*, 1987; Muhich & Boothroyd, 1988; Tschudi & Ullu, 1988). Originally, transcription of the *hsp*70 heat shock genes in *T.brucei* was thought to be monocistronic (Glass *et al*, 1986) but more recent experiments by Lee & Van der Ploeg (1990b) indicate that they too are transcribed polycistronically. Huang & Van der Ploeg (1991a) have shown that the rarity of polycistronic *hsp*70 pre-mRNA in nascent RNA populations may be due to rapid processing by cleavage for polyadenylation. Single copy genes do exist (Swinkels *et al*, 1986) but these are usually genes which are only expressed at a low level, eg. Triose Phosphate Isomerase (TIM).

1.6 Transcription

All mRNAs in trypanosomes have a 39 nucleotide leader sequence at their 5'-end (the spliced leader (SL) or mini-exon; Sather & Agabian, 1985) which is not coordinately transcribed with the rest of the gene (Boothroyd & Cross, 1982; De Lange *et al*, 1984; Parsons *et al*, 1984; Guyaux *et al*, 1985). Instead this leader sequence is transcribed in a longer form 140bases long, called the SLRNA or medRNA (Campbell *et al*, 1984; Milhausen *et al*, 1984; Kooter *et al*, 1984), from tandem arrays of a 1.35kb repeat which exists in one or more clusters (De Lange *et al*, 1983; Michiels *et al*, 1983; Nelson *et al*, 1984). The 39 5' nucleotides are then attached to the nascent mRNA by a *trans*-splicing event which produces a y-branched side-product (Sutton & Boothroyd, 1986; Murphy *et al*, 1986), reminiscent of the lariat structure produced during cis-splicing in other systems. The process of *trans*-splicing appears to be very similar to *cis*-splicing in other respects as well - U2, U4 and U6 snRNAs are required for *trans*-splicing (Mottram *et al*, 1989; Tschudi & Ullu, 1990) and the SLRNA may play an additional catalytic role functionally equivalent to that played by U1 in *cis* splicing (reviewed by Perry & Agabian, 1991). In this respect it is interesting that *cis* splicing does not occur in trypanosomes, there being no introns. The two processes do not however appear to be mutually exclusive as both *cis*- and *trans*-splicing occur in the nematode *Caenorhabditis elegans* (Krause & Hirsch, 1987).

The SL provides a 5'-cap structure for the mRNAs (Laird *et al*, 1985) which they would otherwise lack if they are transcribed in long multi-cistronic transcription units. The mature 3'-ends of these mRNAs are produced by polyadenylation. The function of the SL may just be to provide this cap structure and a cutting mechanism for the polycistronic primary transcipts but a further role may be played in the translation of the mRNAs. An oligonucleotide complementary to the SL sequence can totally inhibit *in vitro* translation of all trypanosome mRNAs (Walder *et al*, 1986; Cornelissen *et al*, 1986). Such a role could conceivably be related to the fact that trypanosome ribosomes have an unusual structure, containing more than the usual number of rRNAs (Cordingley & Turner, 1980).

In most eukaryotes, there are three classes of RNA polymerase. RNA polymerase I (pol I) transcribes rRNAs, RNA polymerase II (pol II) transcribes mRNAs and RNA polymerase III (pol III) transcribes tRNAs and other small RNAs (Chambon, 1975). These can be distinguished by their varying sensitivities to the drug α-amanitin. RNA polymerase I is insensitive to 500ug.ml^{-1}, RNA polymerase III has intermediate sensitivity and RNA polymerase II is sensitive to levels as low as 5ug.ml^{-1}. In trypanosomes most protein coding genes are inhibited by 5ug.ml^{-1} of α-amanitin (Kooter & Borst, 1984) and are therefore presumably transcribed by a conventional polymerase II. Ribosomal and 5S RNA transcription also appear to have expected sensitivities to the drug (Dorfman *et al*, 1985) but VSG and procyclin transcription is insensitive to 1mg.ml^{-1} α-amanitin (Kooter & Borst, 1984; Koenig *et al*, 1989; Rudenko *et al*, 1989 & 1990). The question which has still not been fully answered is whether these protein coding genes are transcribed by RNA polymerase I as suggested by Shea *et al* (1987), by a separate polymerase or by RNA polymerase II containing an insensitive cofactor. Fractionation of RNA polymerases has only identified two (Earnshaw *et al*, 1987) or three (Grondal *et al*, 1989) different activities but genetic analysis of the genes for the large subunits of RNA polymerases (Evers *et al*, 1989; Smith *et al*, 1989) has identified two copies of the polymerase II large subunit, as opposed to the normal one, in addition to those for RNA polymerases I and III. Furthermore, one of these pol II genes, which differs from the other by a few polymorphisms, is only present in those trypanosome species which exhibit antigenic variation. The differences in this copy of the gene do not lie in the region associated with α-amanitin resistance variants in other organisms but the possibility remains that a cofactor which binds to the polymerase confers on it α-amanitin resistance. A modified polymerase II is therefore the most favoured model at present but it is still possible that polymerase I performs the

function as Zomerdijk *et al* (1991) and Rudenko *et al* (1991) have shown that a trypanosome rRNA promoter can promote transcription of marker mRNAs in transient and stable transformation assays, unlike the pol I promoters in other systems. This is probably because discontinuous transcription and *trans* splicing in trypanosomes removes the need for the polymerase transcribing the protein-coding genes to provide a 5' cap. By *in situ* hybridization analysis, Rudenko *et al* (1991) showed that this transcription was localized in the nucleolus of interphase procyclic *T.brucei*. The PARP genes also appear to be located and transcribed at the nucleolus (H.M .Chung & L.H.T. Van der Ploeg, unpublished; Rudenko *et al*, 1991). Transcription of VSG and PARP genes by pol I would not however explain the effect of Mn^{2+} and RNA polymerase inhibitors on VSG expression in nuclear run-on experiments which do not resemble the effects on rRNA (pol I) gene expression (Grondal *et al*, 1989), but is in agreement with the resistance of VSG and rRNA transcription to sarkosyl, while other protein-coding genes are sensitive (Rudenko *et al*, 1992). Why the trypanosome has chosen a different polymerase for VSG and procyclin transcription from other protein-coding genes is not clear but as both of these genes are expressed at a high level (10% and 3% of total mRNA respectively) it may be a means of ensuring that transcription levels are maintained.

1.7 Control of Gene Expression

Little is known about the control of gene expression in trypanosomes as, until very recently, transformation systems were not available. Recent successes (Laban & Wirth, 1989; Bellofatto & Cross, 1989; Kapler *et al*, 1990; Cruz & Beverley, 1990; ten Asbroek *et al*, 1990; Lee & Van der Ploeg, 1990a) in this area however suggest that some questions might now be investigated. No conventional promoters have been found for protein coding genes in trypanosomes and, without transfection, it has not been possible to identify what these might be. The observation that all protein coding genes are transcribed discontinuously has meant that the start of transcription cannot be found by analysis of steady-state transcripts. Nuclear run-on analysis in isolated nuclei has allowed the identification of transcription start sites for several highly expressed transcription units (Kooter *et al*, 1987; Pays *et al*, 1989a; Clayton *et al*, 1990) and this has been backed up by UV irradiation data in the case of VSG and procyclin transcription units (Johnson *et al*, 1987; Pays *et al*, 1990; Clayton *et al*, 1990). While the VSG gene is 45-60kb distant from its promoter, procyclin genes lie immediately downstream. Transcription of

these genes appears to be multi-cistronic, with several related genes being transcribed together (Kooter *et al*, 1987; Koenig-Martin *et al*, 1992). Such a process would seem to make sense when the different genes in a transcription unit need to be coordinately expressed, as in the alpha- and beta-tubulin gene cluster, but this is not always the case and in *Leishmania enriettii* the alpha and beta tubulin genes are in separate clusters (Landfear *et al*, 1983). In bloodstream VSG gene expression sites there are a number of other genes (Expression Site Associated Genes; ESAGs) coordinately transcribed with the VSG gene (Kooter *et al*, 1987; Pays *et al*, 1989a) but their steady-state RNA levels are several hundred-fold lower than the VSG gene itself (Cully *et al*, 1985). This suggests that posttranscriptional processes have an important role to play in control of gene expression in trypanosomes. This has become even more evident with the discovery that the promoter regions of at least some stage-regulated transcripts are active in the stages of the parasite where the genes they control are not expressed. VSG gene expression sites have been found to be active in procyclics (Pays *et al*, 1989b) although transcription does appear to terminate upstream of the VSG gene and most of the ESAGs. Likewise, nuclear run-on analysis of the promoter of the procyclin genes suggests that it is active to some extent in bloodstream forms (Pays *et al*, 1990). Transient transformation assays in the procyclic form of the parasite support this observation as both procyclin (Clayton *et al*, 1990; Rudenko *et al*, 1990) and VSG (Zomerdijk *et al*, 1990; Jeffries *et al*, 1991)) promoter regions appear to function in that stage.

How then is expression of these genes regulated in a stage-specific manner? By inserting other regions of DNA from these genes around the marker gene for Chloramphenicol Acetyl Transferase (CAT) in chimeric constructs with VSG and procyclin promoters it has been possible to modulate the level of CAT activity in transient expression assays of transfected trypanosomes in a stage-specific fashion (Jefferies *et al*, 1991). Endogenous bacterial CAT activity has produced artifacts in bloodstream form transfections, making these results suspect but the procyclic transfection results still stand. Two elements appear to be important - the splice acceptor region from the homologous gene and sequences in the 3'-untranslated region. So far the regions analysed have been quite large but further investigation should allow them to be reduced to the minimal size. Whether such elements will have structural importance at the primary sequence level or at a higher order of structure remains unclear as there are not enough stage-specific genes cloned to compare and one would expect that those important in procyclic and in bloodstream forms would be different by definition.

1.8 The Mitochondrion and RNA Editing

Trypanosomatids display many unique features, including *trans*-splicing, polycistronic transcription, and the packaging of the glycolytic machinery in microbody-like organelles called glycosomes (Opperdoes & Borst, 1977), but perhaps the most bizarre is the editing of mRNAs in the mitochondrion. The mitochondrion is a single organelle in trypanosomes and several of its functions are encoded in the kinetoplast DNA (kDNA). The kinetoplast DNA is a unique network of catenated circular DNA molecules which comprise two size classes (reviewed by Englund *et al*, 1982). There are 20-50 identical copies of the larger maxicircles which are 20kb in length in *T.brucei* and the equivalent to the mitochondrial DNA (mtDNA) of other eukaryotes (Simpson, 1987). These are interlinked by about 10,000 minicircles of around 1kb which are very heterogeneous in sequence in *T.brucei* (Chen & Donelson, 1980; Steinert & Van Assel, 1980; Borst *et al*, 1987) but homogeneous in *T.evansi* and *T.equiperdum* which interestingly do not have a functional mitochondrion (Fairlamb *et al*, 1978; Frasch *et al*, 1980). The function of these minicircles has only been elucidated with the discovery of RNA editing. This is the process whereby the sequence of primary transcripts of mitochondrial genes is altered, by the addition or deletion of uridine residues, to create functional mRNAs (reviewed by Simpson, 1987). There are no genes for edited transcripts in either the mitochondrion or the nucleus (Benne *et al*, 1986; van der Spek *et al*, 1988; Shaw *et al*, 1988). Along with the maxicircles (Bhat *et al*, 1990; Blum *et al*, 1990), the minicircles (Pollard *et al*, 1990; Sturm & Simpson, 1990a) encode short transcripts (60-80 bases long) called guide RNAs (gRNAs) which are complementary to regions of edited transcripts if G.U wobble base pairing is allowed and have tails of polyuridine. It is thought they guide the addition or deletion of uridine residues in the editing process. Each minicircle has the capacity to encode 3 gRNAs (Pollard *et al*, 1990) and there are at least 400 different minicircle sequence classes in *T.brucei* (Stuart, 1979), indicating that there is more than enough capacity within the kDNA to direct the editing of all edited transcripts. Multiple gRNAs may direct the editing of a single domain (Sturm & Simpson, 1990b; Decker & Sollner-Webb, 1990), reflecting inefficiency in the editing process and the potential for a proofreading system.

Not all transcripts in the mitochondrion are edited (*eg*. ND1, ND4, ND5, MURF1; Shaw *et al*, 1988; Feagin *et al*, 1988a; Campbell *et al*, 1989) and those that are, are edited to varying degrees. In several cases the AUG initiation codon is only

present in fully edited transcripts (*eg.* CYb, COIII, ND7, A6; van der Spek *et al*, 1988; Feagin *et al*, 1988b; Bhat *et al*, 1990) and the editing of the COIII gene in *T.brucei* is so extensive (Feagin *et al*, 1988a; Shaw *et al*, 1988) that the the gene was not distinguishable before editing was discovered. Editing produces functional open reading frames that encode proteins that are highly homologous within the order kinetoplastida (Stuart, 1991) and also equivalent to mitochondrial proteins in other eukaryotes which do not have editing.

The details of the editing mechanism are not known but it appears to occur posttranscriptionally (Stuart *et al*, 1988) in a 3'-5' direction (Abraham *et al*, 1988; Koslowsky *et al*, 1990) via a series of domains (Decker & Sollner-Webb, 1990). Two models were originally proposed for the mechanism of editing which use gRNA base pairing to determine mismatch recognition (Blum *et al*, 1990) or protection from random addition and deletion of uridines (Decker & Sollner-Webb, 1990). These models require several enzymatic activities including an RNA endonuclease, a terminal uridilyl transferase (TUTase), RNA ligase and a 3'-exonuclease. TUTase and RNA ligase activities have been found in kinetiplastid mitochondria (Bakalara *et al*, 1989). An alternative model has been proposed by Cech (1991) which involves transesterifications, akin to those involved in splicing, initiated by the free 3' hydroxyl group on the poly(U) tails of the gRNAs. Chimeric RNA molecules comprising gRNAs covalently linked to mRNAs have been observed *in vivo* by Blum *et al* (1991) and Harris & Hajduk (1992) have observed specific chimera formation between a CYb gRNA and the pre-edited CYb mRNA *in vitro* via the 3' poly(U) tail of the gRNA. This model removes the need for endonuclease and RNA ligase activities and is therefore more attractive. The role for TUTase might be to generate the poly(U) tail on the gRNAs (Cech, 1991). There are other aspects of the editing process which display similarities to splicing, the gRNA could be considered the functional eqivalent of the internal intron RNA guide in type I and type II introns and U1 in nuclear mRNA splicing. For this reason it has been suggested (Simpson & Shaw, 1989; Stuart, 1991) that editing takes place in a multi-enzyme complex or EDITOSOME, the equivalent of spliceosomes in splicing, although such structures have not yet been published.

The oxidative pathway in the mitochondrion is only fully functional in insect stages of *T.brucei* (Simpson, 1987). There are no spectrophotometrically detectable cytochromes in bloodstream forms and the citric acid cycle is incomplete. Bloodstream forms therefore derive their energy from glycolysis and possibly from

protein degradation. All maxicircle genes appear to be transcribed (Stuart, 1991) and some are edited in all stages of the lifecycle (Feagin & Stuart, 1988) but the CYb and COII mRNAs are only fully edited in the procyclic stage (Feagin *et al*, 1987; Feagin & Stuart, 1988) and the 3' domain of ND7 (Koslowsky *et al*, 1990) is only edited completely in bloodstream forms. There must be a transcript-specific mechanism to regulate editing throughout the lifecycle at the level of either gRNA production or use. Some transcripts are edited in the 3' untranslated region (3' UTR) and poly(A) tail (Campbell *et al*, 1989). This would clearly not affect the translation product *per se* but may alter the mRNA stability and therefore its abundance. Thus editing appears to play a regulatory role in mitochondrial biogenesis by modifying the functional capabilities of mRNAs.

1.9 Transmission Blocking Vaccines

Because of the variability in the surface coat it is unlikely that a vaccine could be designed against the mammalian stages of the trypanosome's life cycle. However, only a subset of the whole antigenic repertoire is expressed in the metacyclic stage (Hajduk *et al*, 1981; Crowe *et al*, 1983; Esser & Schoenbechler, 1985). This subset of the repertoire is highly predictable (Barry *et al*, 1983) and it was originally thought that a vaccine against the VSGs expressed at this stage might prevent development of the infection. However, the predictability is not absolute and the metacyclic repertoire displays antigenic drift over time (Barry *et al*, 1983), therefore making such an approach unlikely to succeed.

The other insect stages of the lifecycle, which do not possess the VSG coat, share common antigens on their surface, many of which are glycoproteins (Gardiner *et al*, 1983). With the exception of procyclin, little is known about these molecules. The presence of exposed carbohydrate groups has been shown by lectin analyses (Jackson *et al*, 1978) - in *T.congolense* the procyclic stage is reactive with concanavalin A and wheat germ agglutinin (Mutharia & Pearson, 1987). Some enzyme activities associated with the *T.brucei* procyclic surface have been reported (Barry, 1977) and two proteins, one associated with the membrane although not necessarily on the surface of *T.congolense* (Parish *et al*, 1985) and another found in several trypanosome species (Rovis *et al*, 1984) have been biochemically identified. Surface iodination of procyclic cells with ^{125}I followed by SDS-PAGE of lysates and autoradiography identified a minimum of 25 bands in *T.brucei* which probably

include proteins located on or associated with the cell surface (Gardiner *et al*, 1983) and reproducibly reveals 12 major components in *T.congolense* (F.A. Lainson & J.D. Barry, unpublished). Fourteen of the iodinated bands in *T.brucei* were detected by rabbit antisera raised against procyclic pellicular fractions (Gardiner *et al*, 1983) and if a vaccine could be designed against such common epitopes then tsetse flies feeding on such vaccinated animals would ingest antibody with their blood meal every 2-3 days and this might block the infection and prevent transmission (Barry, 1977; Murray *et al*, 1980). Indeed, feeding tsetse flies on animals immunized with the procyclic culture form had a marked effect on transmission of *T.brucei* and totally abolished *T.congolense* development (Maudlin *et al*, 1984; Murray *et al*, 1985). The difference in the level of effect between these two species is probably due to the site of development of later lifecycle stages in the fly. Epimastigote *T.congolense* develops in the mouthparts and is therefore readily exposed to antibody present in the bloodmeal. Epimastigote *T.brucei* on the other hand is sequestered in the salivary glands and trypanosomes of this species might only be exposed to trypanocidal antibodies when the parasites migrate from the fly's gut, through the oesophagus, before they reach the salivary glands. A transmission blocking vaccine for *T.congolense* is therefore more likely to be effective (Barry, 1986). However, Nantulya & Moloo (1988) had total abolition of *T.brucei* development with a *T.brucei* mab which was probably specific to procyclin.

In order to identify individual antigens which might form the basis of such a transmission blocking vaccine, a collection of monoclonal antibodies (mabs) was raised against living procyclic culture form *Trypanosoma congolense* (F.A.Lainson & J.D.Barry, unpublished) and several were shown to be agglutinating and trypanocidal towards the procyclic culture form. When these mabs were fed to tsetse flies with their bloodmeal, 40% fewer flies developed mature infections (I.Maudlin, unpublished). It would not be practical to attempt to immunize reservoir hosts but computer models (N. Putt, unpublished) suggest that such an approach in a cattle ranch situation might be successful in reducing transmission because flies feed only on the animals present in the ranch.

1.10 A Procyclic Surface Antigen

The group of mabs which were agglutinating gave strong fluorescence on the surface of procyclic and epimastigote trypanosomes in indirect immunofluorescence

assays (IFA) and all detected the same diffuse bands in western blots of procyclic lysates (F.A. Lainson & E. Kilbride, unpublished). This material detected in western blots does not stain with coomassie blue or silver stain and is acidic as it stains blue with Stains All. It is present in detergent lysates and can be isolated by organic extraction methods, indicating that it is closely associated with membranes. Partial purification has been achieved by cation exchange chromatography and excision from SDS-PAGE gel fractionations. Amino acid analysis indicates that it does have a protein component which is rich in alanine residues (Beecroft *et al*, 1992).

The aim of this thesis was to identify the gene(s) encoding this antigen(s), which from IFA studies is species and stage specific, and therefore obtain information about the protein(s) which may prove useful in both generating transmission blocking vaccines and in analysing the interaction of the parasite with its tsetse fly vector.

CHAPTER 2

MATERIALS & METHODS

2.1 List of Materials

Name	Source
General chemicals & organic solvents	B.D.H., Hopkins & Williams, Koch-Light Laboratories, May & Baker
Media	Davis, Oxoid
Agar	Davis, Difco
Biochemicals	Sigma
Antibiotics	Sigma
X-gal	Boehringer
IPTG	Boehringer, BRL
RNase A	Sigma
Proteinase K	Sigma
Foetal Calf Serum	Gibco BRL, NBL
Glutamine Solution	Gibco BRL, NBL
Eagle's MEM	Gibco BRL
5% CO_2	B.O.C.
Diethyl aminoethyl (DEAE) Cellulose DE-52	Whatman
Gigapack Packaging Extract	Stratagene Incorp.
Sequenase	United States Biochemicals
Freund's Adjuvant	Difco
Ponceau S	Sigma

Glutathione Beads	Pharmacia
Horse Serum	Gibco BRL
HRP-anti rat	Sigma
HRP-anti mouse	Promega
FITC-anti rat	Sigma
FITC-anti mouse	Sigma
ECL System	Amersham
Hybond N	Amersham
Nytran membrane	Schleicher & Schuell
Nitrocellulose	Schleicher & Schuell
Geneclean	Stratatech.
RNAGuard	Pharmacia
Enzymes	BRL, Promega, NBL, NEB, Boehringer, Pharmacia
Preweighed acrylamide	Biorad
oligo(dT) cellulose	BRL
oligo(dT) primer	BRL
$p(dN)_6$	Pharmacia
^{32}P-dNTPs	New England Nuclear
^{35}S-dATP	New England Nuclear
Formamide	Rose Chemicals
Agarose	IBI, BRL

X-ray Film	Amersham, Kodak, Fuji
DNA Markers	BRL
RNA Markers	BRL
SDS-PAGE Markers	Sigma

2.2 Trypanosomes

The TREU (Trypanosomiasis Research Edinburgh University) 1627 (originating from Gambia) and TREU 1457 (Nigeria) (Ross *et al*, 1985) stocks of *Trypanosoma congolense* were obtained as cloned procyclic populations from Dr C.A. Ross, Centre for Tropical Veterinary Medicine, Edinburgh. Stock M15 1/148 (Young & Godfrey, 1983) was obtained as a bloodstream stabilate from Dr I. Maudlin, Tsetse Research Laboratory, Bristol and transformed to procyclic cells by Dr C.A. Ross. The YNat 1.1 stock has been described by Rosen *et al* (1981).

2.3 Animals

Bloodstream trypanosomes were grown in 30g female BKTO mice or 250g female Wistar rats. Antisera were raised in 30g female Balb/c mice and 250g female Wistar rats. All animals were purchased from Bantin & Kingman Ltd. and maintained at the Institute of Virology, Glasgow University.

2.4 Bacterial Strains

All bacterial strains were derivatives of *Escherichia coli* K-12

Strain	Genotype	Ref./Source
JM101	*sup*E, *thi*, (*lac-pro*AB), {F', *tra*D36, *pro*AB, *lac*I^{q}Z ΔM15} Restriction: (r_k^+, m_k^-), mcrA$^+$	Yanisch-Perron *et al* (1985).

XL1-Blue	*rec*A1, *end*A1, *gyr*A96, *thi*, *hsd*R17(r_k^-, m_k^+), *sup*E44, *rel*A1, λ^-, (*lac*), {F', *pro*AB, *lac*I^qZΔM15, Tn10(tet^r)}	Bullock *et al* (1987).
PLK-F'	*rec*A, *lac*$^-$, *mcr*A(-), *mcr*B(-) *hsd*R^-, *hsd*M^+, *gal*$^-$, *sup*E, {F', *pro*AB, *lac*I^qZΔM15. Tn10(tet^r)}	Bullock *et al* (1987).
DL491	*hsd*R, *mcr*A, *rec*D, *sbc*C, *sbc*201, Tn10::*pho*R (derivative of NM621)	Whittaker *et al* (1988)
L87	*sup*E, *sup*F, *hsd*R, *trp*C, *met*D, *ton*A.	Amersham
NM514	*hsd*R, *arg*H, *gal*E, *gal*X, *Str*A *lyc*B7	Murray (1983).
LE392	*hsd*R514, (r_k^-,m_k^+), *sup*E44, *sup*F58, *lac*Y1 or (*lac*IZY)6, *gal*K2, *gal*T22, *met*B1, *trp*R55, λ(-).	Maniatis *et al* (1982)
Q359	*sup*E, *hsd*R, $\Phi 80^r$, (P2)	Karn *et al* (1980)
SURE™	e14$^-$ (*mcr*A), (*mcr*CB-*hsd*SMR-*mrr*)171, *end*A1, *sup*E44, *thi*-1, *gyr*A96, *rel*A1, *lac*, *rec*B, *rec*J, *sbc*C, *umu*C:Tn5 (kan^r), *uvr*C, {F' *pro*AB, *lac*I^qZΔM15, Tn10, (tet^r)}.	Greener (1990)

DS941	*rec*F143, *pro*A7, *str*31, *thr*1, *leu*6, *tsx*33, *mt*L1, *his*44, *arg*E3, *lac*Y1, *gal*K2, *ara*14, λ⁻, lacI^q, *lac*ZΔM15, *lac*Y$^+$, *sup*E44, *xyl*15.	D.Sherratt

2.5 Bacteriophage

Lambda ZAP II	Lambda insertion vector from which the inserted DNA can be excised as a pBluescript SK(-) phagemid. (Short *et al*, 1988).
Lambda gt10	Lambda immunity insertion vector for the cloning of small DNA fragments. (Huynh *et al*, 1985).
Lambda EMBL4	Lambda replacement vector used for the construction of genomic DNA libraries. (Frischauf *et al*, 1983).
M13mp18/19	Single stranded virion, used in producing template for DNA sequencing following cloning of exogenous DNA into the replicative form (RF) genome. (Messing, 1983).
R408	Stable interference resistant helper phage for excision of phagemid DNA from lambda ZAP II and production of
single-	stranded phagemid DNA for sequencing. (Russel *et al*, 1986).
VCSM13	Kan^r interference resistant helper phage derived from the M13 K07 mutant. Used as for R408. (Stratagene)

2.6 Plasmid Vectors

pBluescript II	phagemid cloning and sequencing vector (Short *et al*, 1988).
pUC19	pBR322-derived cloning vector (Yanisch-Perron *et al*, 1985).
pUC13	pBR322-derived cloning vector with Cm^r (Messing, 1983)

pGEX-2T — plasmid expression vector which produces IPTG-inducible fusions to the glutathione-S-transferase gene of *Schistosoma japonicum* (Smith & Johnson, 1988).

2.7 Probes

pPRO2001 — *T.brucei* procyclin cDNA clone (Roditi *et al*, 1987).

pTbαβT-1 — *T.brucei* genomic clone containing an alpha- and beta-tubulin repeat unit (Thomashow *et al*, 1983).

pTcTIM9 — cDNA clone of the *T.brucei* triose phosphate isomerase gene (Swinkels *et al*, 1986).

2.8 Trypanosome Culture Media

Heat Inactivated Foetal Calf Serum: The serum was heat inactivated by incubating at 56°C for 30 mins. and batch-tested for trypanosome growth.

Procyclic Growth Medium: Eagle's Minimal Essential Medium without L-glutamine (Gibco B.R.L.), supplemented with 20% (v/v) Heat Inactivated Foetal Calf Serum, 4mM L-glutamine and 0.1% D-glucose (optional).

109-C: 300ml Minimal Essential Medium with Earle's Salts without L-glutamine; 6ml 50X MEM Amino Acid Solution; 6ml 100X MEM Non-Essential Amino Acid Solution; 18ml 60mM HEPES; 0.09g L-glutamine; 0.09g L-proline; 0.3g glucose; 6mg adenosine. pH to 7.4 with 4N NaOH, filter sterilize, and add 10% (v/v) Heat Inactivated Foetal Calf Serum + 0.1% (v/v) of a $2mg.ml^{-1}$ solution of Haemin.

10X PBS: 84.3g Na_2HPO_4, 4.9g $NaH_2PO_4.2H_2O$, 26.6g NaCl, adjust pH to 8.0 and make up to 1l with dH_2O.

PSG: 1% D-glucose added to 10X PBS diluted 3:47 (v/v) with dH_2O and pH adjusted to 8.05.

2.9 *E.coli* Culture Media

L-Broth (LB): 10g tryptone, 5g yeast extract, 5g NaCl, 20mg thiamine, made up to 1l with dH_2O and adjusted to pH 7.0 with NaOH. Solid L-Agar was as LB, but with the addition of 12g.l^{-1} of No.3 Oxoid agar. For bacteriophage work, a supplement of 10mM $MgSO_4$ + 0.2% maltose (final concentration) was added to the LB.

2X YT Medium: 16g Bactotryptone, 10g yeast extract, 5g NaCl, made up to 1l with dH_2O.

Superbroth: 35g bactotryptone, 20g yeast extract, 5g NaCl, 0.02M $NaPO_4$, pH7.5 made up to 1l in dH_2O.

BBL Agarose Overlay: 10g trypticase peptone (BBL 11921), 5g NaCl, made up to 1l with dH_2O and adjusted to pH 7.2 with NaOH. Add 2.5g $MgSO_4.6H_2O$ before solidification with 6.5g agarose (Type 1. low EEO A6013).

Minimal Medium Plus Glucose (MM+G): 15g Taiyo agar made up to 400ml with dH_2O and supplemented with 50ml 10X M9 Salts, 7ml 100mM $MgSO_4$, 2ml 50mM $CaCl_2$, 5ml 20% glucose and 200ul 1% thiamine prior to use.

10X M9 Salts: 60g Na_2HPO_4, 30g KH_2PO_4, 10g NH_4Cl, 5g NaCl, made up to 1l with dH_2O.

M13 Soft Overlay: 6g Taiyo agar made up to 1l with dH_2O.

S.O.C. Transformation Recovery Medium: 2g Bactotryptone, 0.5g yeast extract, 1ml 1M NaCl, 0.25ml 1M KCl, added to 97ml dH_2O and autoclaved. Supplement with 1ml of 1M $MgCl_2$/ $MgSO_4$ + 1ml of 2M glucose. Filter sterilize.

Davis & Mingioli (D&M) Salts (4X): 28g K_2HPO_4, 8g KH_2PO_4, 4g $(NH_4)_2SO_4$, 19g Nacitrate, 0.4g $MgSO_4.7H_2O$, made up to 1l in dH_2O.

SM Buffer: Used for bacteriophage storage and dilution:- 5.8g NaCl, 2g $MgSO_4.7H_2O$, 50mM Tris.HCl, pH7.5, 0.01% gelatin, made up to 1l with dH_2O.

TM Buffer: 50mM Tris.HCl, pH7.5, 10mM $MgSO_4$.

2.10 Antibiotics and Media Supplements

Supplement	Stock	Final Conc.
Ampicillin	100mg.ml^{-1} in dH_2O	100ug.ml^{-1}
Tetracyclin	10mg.ml^{-1} in 1/1000 conc.HCl	10ug.ml^{-1}
Kanamycin	25mg.ml^{-1} in dH_2O	25ug.ml^{-1}
Chloramphenicol	25mg.ml^{-1} in ethanol	25ug.ml^{-1}
X-gal	40mg.ml^{-1} in dimethyl formamide	80ug.ml^{-1}
IPTG	0.5M in dH_2O	20mM

2.11 Buffer Solutions

2.11.1 Electrophoresis

a) DNA

10X TBE: 109g Tris, 55g boric acid, 9.3g $Na_2EDTA.2H_2O$ made up to 1l with dH_2O; pH 8.3.

50X TAE: 242.5g Tris, 18g NaAcetate, 18g $Na_2EDTA.2H_2O$, pH adjusted to 8.2 with glacial acetic acid and made up to 1l with dH_2O.

5X FSB: 10% Ficoll, 0.025% bromophenol blue, 0.025% orange G in 5X TBE.

5% Sequencing Gel Mix: 16.7ml 30% acrylamide:bis acrylamide (29:1 w/w) stock solution, 42g Urea, 10ml of 10X TBE, made up to 100ml with dH_2O and filtered through Whatman No.1. Polymerized by the addition of 800ul fresh 10% ammonium persulphate and 40ul TEMED.

Alkaline Gel Preparation Buffer: 50mM NaCl, 1mM EDTA

Alkaline Gel Running Buffer: 30mM NaOH, 1mM EDTA

5X Alkaline Loading Buffer: 150mM NaoH, 5mM EDTA, 5% Ficoll, 0.025% Bromocresol Green.

b) RNA

10X MOPS: 42g morpholinopropane sulphonic acid, 6.8g NaAcetate, 3.7g Na_2EDTA (pH8.0) made up to 1l with dH_2O.

"Northern Blue Juice": 50% glycerol, 10mM NaPhosphate, pH7.0, 1% bromophenol blue.

c) Protein

SDS-PAGE Resolving Gel: acrylamide:bis acrylamide (30:0.8 w/w) to desired percentage, 0.375M Tris.HCl (pH8.8), 0.1% SDS, 0.05% fresh ammonium persulphate, 0.05% (v/v) TEMED.

SDS-PAGE Stacking Gel: 4% acrylamide:bis acrylamide (30:0.8 w/w), 0.125M Tris.HCl (pH6.8), 0.1% SDS, 0.05% fresh ammonium persulphate, 0.05% (v/v) TEMED.

Running Buffer: 25mM Tris, 192mM glycine, 0.1% SDS

2X SDS Loading Buffer: 0.125M Tris.HCl (pH6.8), 10% (v/v) glycerol, 2% SDS, 5% (v/v) 2-mercaptoethanol, 0.05% bromophenol blue.

Coomassie Stain: $3mg.ml^{-1}$ Coomassie Brilliant Blue in 35% (v/v) methanol/ 10% (v/v) acetic acid.

Electroblotting Buffer: 25mM Tris, 192mM glycine, 20% methanol

2.11.2 DNA Manipulation

10X Restriction Buffers: React Buffers from B.R.L.

5X Ligation Buffer: 330mM Tris.HCl, pH7.5, 5mM Spermidine, 50mM $MgCl_2$, 75mM DTT, 0.2% gelatin.

2X *Bal*-31 Buffer: 1.2M NaCl, 24mM $MgCl_2$, 24mM $CaCl_2$, 40mM Tris.HCl, pH 8.0, 2mM EDTA.

10X *Taq* Polymerase Buffer (Promega): 0.5M KCl, 0.1M Tris.HCl (pH 9.0), 15mM $MgCl_2$, 1.0% v/v Triton X-100.

2X Denaturation Buffer: 0.4M NaOH, 0.4mM EDTA

5X Sequenase Reaction Buffer: 200mM Tris.HCl (pH7.5), 100mM $MgCl_2$, 250mM NaCl.

2.11.3 Hybridization Solutions

20X SSC: 3M NaCl, 300mM Na_3citrate, pH to 7.0.

50X Denhardt's Solution: 5g Ficoll, 5g polyvinylpyrrolidone, 5g BSA made up to 500ml with dH_2O.

Prehybridization Solution: 6X SSC, 5X Denhardt's Solution, 0.5% SDS, 200ug.ml^{-1} denatured herring sperm DNA.

Northern Prehybridization Solution: 50% (v/v) formamide, 5X SSC, 5X Denhardt's Solution, 0.1% SDS, 200ug.ml^{-1} denatured herring sperm DNA, 50ugml^{-1} each denatured poly(A) and poly(C).

Denaturing Solution: 1.5M NaCl, 0.5M NaOH.

Neutralizing Solution: 1.5M NaCl, 0.5M Tris.HCl, pH7.2, 1mM Na_2EDTA.

2.11.4 Radiolabelled Probe Preparation Solutions

Solution theta: 1.25M Tris.HCl, pH8.0, 0.125M $MgCl_2$.

Solution A (-C): 1ml Solution theta, 18ul 2-mercaptoethanol (14.3M), 5ul each 100mM dATP, dGTP and dTTP.

Solution B: 2M HEPES, pH to 6.6 with 4M NaOH.

Solution C: 50 OD units hexadeoxyribonucleotides (Pharmacia) in 556ul TE.

"Rxn-C": 20ul Solution A, 50ul Solution B, 30ul Solution C. Store at -20°C.

10X Kinase Buffer: 0.5M Tris.Hcl (pH 7.6), 0.1M $MgCl_2$, 50mM DTT, 1mM spermidine, 1mM EDTA.

5X MMLV RT Buffer: 250mM Tris.HCl, pH8.3, 375mM KCl, 50mM DTT, 15mM $MgCl_2$.

2.11.5 DNA Extraction, Purification and General Purpose Solutions

Phenol: All phenol used in the purification of DNA was redistilled and contained 0.1% 8-hydroxyquinoline. Phenol was buffered against 1M Tris.HCl, pH8.0 and then stored under TE buffer.

Phenol/Chloroform: phenol:chloroform:isoamyl alcohol 25:24:1 (v/v) mixture.

NET Buffer: 100mM NaCl, 100mM EDTA, 10mM Tris.HCl, pH8.0 in dH_2O.

DNase-free RNase A: 10mg.ml^{-1} stock in dH_2O boiled for 10 mins. to destroy any DNase contamination.

TE Buffer: 10mM Tris.HCl, 1mM EDTA, pH8.0.

TE (10:0.1) Buffer: 10mM Tris.HCl, 0.1mM EDTA, pH8.0.

STE Buffer: 150mM NaCl, 10mM Tris.HCl (pH7.5), 1mM EDTA

Birnboim Doly Solution I (BDI): 50mM glucose, 25mM Tris.HCl, pH8.0, 10mM EDTA.

Birnboim Doly Solution II (BDII): 0.2M NaOH, 1% SDS, prepared immediately prior to use.

Birnboim Doly Solution III (BDIII): 5M KAcetate: mix equal volumes 3M KAcetate + acetic acid, pH should be 8.0.

Geneclean NEW Buffer: 0.2M Tris.HCl, pH 7.5, 1M NaCl, 20mM EDTA in 25ml; dilute with 225ml dH_2O + 250ml ethanol before use and store at -20°C.

2.11.6 RNA Extraction

LiCl/Urea: 3M LiCl, 6M Urea in dH_2O.

RNA Mix: 10mM Tris.HCl, pH 8.0, 1mM EDTA, 0.02% SDS in dH_2O.

1X Binding Buffer: 0.5M NaCl, 10mM Tris.HCl, pH 7.5, 10mM EDTA, +/- 0.02% SDS in dH_2O.

Wash Buffer: 0.1M NaCl, 10mM Tris.HCl, pH 7.5, 10mM EDTA in dH_2O.

RNA Gel Sample Buffer (SB): 25ul formamide, 5ul 10X MOPS, 8ul 30% (v/v) formaldehyde solution added to RNA sample and volume made up to 50ul with dH_2O.

2.11.7 Protein Purification and Immunological Screening

MTPBS Triton: 150mM NaCl, 16mM Na_2HPO_4, 4mM NaH_2PO_4 (pH7.3), 1% (v/v) Triton X-100.

TBS: 10mM Tris.HCl, pH 8.0, 0.15M NaCl.

TBST: 10mM Tris.HCl, pH 8.0, 0.15M NaCl, 0.05% (v/v) Tween 20.

HST: 10mM Tris.HCl, pH7.4, 1M NaCl, 0.5% (v/v) Tween 20.

HRP Substrate Solution: 18mg 4-chloro-1-napthol in 6ml methanol. Add 94ml TBS + 25ul 30% (v/v) H_2O_2, mix well and use immediately.

2.12 General Techniques

Sterilization of Glassware and Media

Glassware and media were sterilized by autoclaving at 120°C, 15 p.s.i. for 15 mins., supplements and buffers at 105°C, 15 p.s.i. for 10 mins. and heat-sensitive solutions were filter sterilized through 0.22um filters.

Preparation of RNase-free Equipment and Media

Wherever possible, sterile disposable plasticware which was individually wrapped was used (eg. pipette tips, pipettes, tubes) as this can be considered essentially RNase-free. Any glassware to be used was baked at 200°C for 2 hours.

Solutions were prepared from bottles of chemicals kept solely for this purpose (no spatulas were used) and were made up in RNase-free dH_2O which had been treated with 0.05% DEPC before autoclaving.

Preparation of Phenol

Solid phenol was melted and redistilled and 8-hydroxyquinoline added to 0.1%, and then stored at -20°C. When required, the redistilled phenol was thawed and sufficient deionized water (dH_2O) added to produce two separate phases. The phenol could be stored at 4°C in this way ready for equilibration. The excess water was removed and an equal volume of 1M Tris.Cl, pH8.0 was added. The two phases were emulsified by vortexing and then separated again by brief centrifugation. This increased the pH of the phenol to around 7.0. The excess Tris was removed and the phenol equilibrated twice with an equal volume of TE. For phenol/chloroform, equilibrated phenol, chloroform and isoamyl alcohol were mixed in the ratio 25:24:1.

Organic Extractions

Phenol or phenol/chloroform was used to clean up DNA and RNA preparations or to remove enzymes between reactions. An equal volume of phenol or phenol/chloroform was added to the DNA or RNA solution and mixed by gentle inversion or vortexing depending upon the molecular weight of the nucleic acid concerned. The resulting emulsion was then separated into phenolic (lower) and aqueous (upper) layers by centrifugation, the aqueous layer being transferred to a fresh tube. This extraction could be repeated several times until no protein precipitate was visible at the interface. Residual phenol withdrawn with the aqueous layer could be removed by adding an equal volume of chloroform, vortexing, spinning and again withdrawing the upper aqueous layer. All organic extractions were followed by ethanol precipitation to remove residual traces of the organic solvent.

Precipitation of DNA and RNA

DNA was precipitated using either ethanol or isopropanol. In each case the DNA solution was made to 0.3M sodium acetate. For ethanol precipitation, 2 volumes of ethanol were added and the precipitation carried out on ice for 30 mins.. Isopropanol precipitation was performed by adding 0.6 volumes isopropanol to the DNA solution and incubating at room temperature for at least 10 mins.. Ethanol precipitation of RNA was as for DNA except that 2.5 volumes of ethanol were used and precipitation was either on ice for 30 mins or at -20°C

overnight.

Precipitated DNA (or RNA) was pelleted by centrifugation at 12,000*g* in a microcentrifuge for 5-10 mins. or in a Beckman J2-21 JA20 rotor at 17,400*g* for 20 mins. at 4°C. After decanting the supernatant fluid, traces of salt were removed by a 70% ethanol wash. The pellet was then dried by heating at 65°C in a dry heating block (DNA) or by lyophilization (RNA) before resuspending in the appropriate buffer.

2.13 Trypanosome Growth Conditions

Growth of Procyclic *T.congolense*

Procyclic forms of *T.congolense* can be cultured for long periods in a semi-defined medium (Brun, 1982) composed of Eagle's Minimal Essential Medium containing 20% (v/v) Heat Inactivated Foetal Calf Serum and 4mM glutamine. This can be supplemented with 0.1% glucose to aid cell growth if required. The cultures are maintained in upright tissue culture flasks, which have been gassed with 5% CO_2 (v/v) in air, and divided thrice weekly. The yield of trypanosomes is approximately $5x10^6$ cells.ml^{-1}.

Growth of Bloodstream Form *T.congolense*

Most stocks of *T.congolense* are not virulent in laboratory mice (Lumsden *et al*, 1973) and thus it is not usually possible to obtain sufficient numbers of bloodstream forms of this species. Fortunately, the M15 1/148 stock used in this study is virulent in mice (Young & Godfrey, 1983) and so these forms can be obtained by injecting a stabilate into mice and allowing the parasitaemia to develop, a process which normally takes about 7-9 days depending on the initial inoculum. The trypanosomes can then be isolated by cardiac puncture. A large proportion of *T.congolense* however are not recovered by this method, being retained in the peripheral bloodstream (Rosen *et al*, 1979). This may be due to their adherence to capillary walls (Banks, 1978). In order to obtain sufficient parasites it is therefore necessary to overcome this problem as far as possible. Several methods were utilised for this purpose and their merits are discussed in Chapter 4.

a) It has been observed (Rosen *et al*, 1979) that warming animals to 37°C prior to bleeding releases many more parasites into the general bloodstream so that they may be collected by cardiac puncture. The mice were therefore kept at 37°C for 30 mins. after the parasitaemia had reached the desirable level before bleeding and isolating the bloodstream forms. Parasites isolated by this method were used for the initial bloodstream RNA preparations used to screen the cDNA library and the initial northern blots.

b) The second method used to obtain bloodstream form RNA for confirming the northern data was to leave the parasitaemia to rise as high as possible before death of the host, prior to bleeding.

c) As neither of the above methods proved entirely satisfactory and produced conflicting results, rats were inoculated at a low level and the infection was allowed to pass through several waves of parasitaemia before collecting the blood prior to the parasitaemia reaching a high peak.

Transformation of *T.congolense* Bloodstream to Procyclic Forms.

Bloodstream forms of stock M15 1/148 were grown up in mice and then blood was prepared by exsanguination. The infected blood was mixed in a tissue culture flask with an equal volume of 109-C culture medium and incubated at 28°C for 2 days. After this time it was observed that the trypanosomes had altered in appearance to that of procyclic cells and that the cultures had started to divide. The blood cells were allowed to settle and the blood gradually diluted out with culture medium over the next few days.

The cultures grew rapidly at first to produce large numbers of procyclic trypanosomes but then for some unknown reason appeared to die. This happened in each of several attempts to establish a stable culture, despite modifying the serum concentration of the medium and the rate of diluting out the blood, but the parasites always appeared to die after approximately two weeks in culture. It is a common experience that *T.congolense* is more difficult than *T.brucei* to culture in this way (J.D. Barry, pers.comm.).

Isolation of Bloodstream Form Trypanosomes from Blood

Any VSG coated trypanosomes can be separated from blood cells, or other uncoated forms, by DEAE cellulose chromatography according to the method of Lanham & Godfrey (1970). The DEAE cellulose (Whatman DE52) is prepared by suspending in PBS and adjusting the pH to 8.05 with 5% phosphoric acid, allowing it to settle, and replacing with fresh PBS until the pH of the suspension is the same as the PBS buffer.

A 10ml column is required for every 1ml of blood and is prepared in a disposable syringe plugged with glass wool. The column is then reequilibrated by washing with at least 2 column volumes of PSG, pH8.05 before loading the blood onto the column. The blood is allowed to enter the column bed and then it is washed with PSG, pH8.05 and the eluate collected when it starts to turn a pearly colour due to the large number of trypanosomes present or when phase contrast microscopic examination reveals their presence. Eluate is collected until it becomes clear again. The number of trypanosomes isolated can be counted by use of a Neubauer haemocytometer and then pelleted by centrifugation before purification of nucleic acid.

2.14 Bacterial and Phage Growth

Growth of Bacteria

E.coli cells were grown aerobically by inoculating the culture medium, supplemented with the relevant antibiotic where appropriate, with a single colony and shaking at 37°C.

Plating Out of Bacteriophage

LB supplemented with 10mM $MgSO_4$ and 0.2% maltose was inoculated with a single colony of the desired *E.coli* plating strain and grown overnight at 37°C with vigorous shaking. The cells were then harvested by centrifugation at 1000*g* for 5 mins. and resuspended in ice-cold 10mM $MgSO_4$ at 1/50 the original volume. Cells prepared in this way could be used for several weeks although, for optimum plating efficiencies, fresh cells were prepared. The phage particles were

mixed with an appropriate volume of plating cells, depending on the size of plate to be used, and incubated at 37°C for 15 mins. to allow the phage to adsorb. This suspension was then added to molten BBL-agarose overlay cooled to 48°C, mixed gently and poured immediately onto a prewarmed LB plate. After allowing the overlay to harden, the plates were inverted and incubated at 37°C overnight.

Preparation of a High Titre Lambda Stock

50ul of a purified plaque was added to 100ul of plating cells and allowed to preadsorb at 37°C for 15 mins. This was then allowed to grow overnight in 10ml L-broth containing 0.2% maltose and 10mM $MgSO_4$ at 37°C. The next day the cells were pelleted and 5ml of the phage supernatant phase was preadsorbed to 5ml plating cells. This culture was grown at 37°C in 100ml L-broth with maltose and magnesium for 5 hrs and then shaken with 1ml chloroform for 15 mins. The cell debris was removed by spinning at 3,000*g* for 10 mins at 4°C and the high titre phage supernatant phase transferred to a fresh tube and stored at 4°C.

Preparation of Helper Phage

An overnight culture of *E.coli* XL1-Blue was diluted 1/10 in 2XYT and grown at 37°C for 1 hour. Helper phage were added to the flask at a multiplicity of infection (m.o.i.) of about 20:1 and the culture shaken for a further 30 mins at 37°C before addition of 25ug.ml^{-1} Kanamycin for phage VCSM13. (For phage R408 the Kanamycin was omitted.) The culture was then shaken at 37°C for a further 6 hours and the cells killed by heating at 65°C for 15 mins. Cellular debris was then removed by spinning the culture at 12,000*g* for 10 mins and transferring the supernatant phase containing the phage to a fresh tube. The phage were then titrated by plating dilutions on plates of XL1-Blue and incubating overnight. The helper phage were stored at 4°C and could be used for at least 1 month before the titre dropped.

Excision of pBluescript DNA from Lambda ZAP II

This was carried out according to the manufacturer's instructions using the helper phage R408.

2.15 Preparation of Nucleic Acids

2.15.1 RNA

Isolation of RNA from Trypanosomes

Bloodstream or procyclic form typanosomes were pelleted by centrifugation at 3000*g* for 5 mins and resuspended in at least 10 packed cell volumes (PCVs) of 3M LiCl/6M urea to release nucleic acid (Auffray & Rougeon, 1980; Bernards *et al*, 1981). The lysate was then sonicated or syringe passaged 20 times to shear the DNA and incubated on ice overnight to allow the RNA to precipitate while the DNA remained in solution. RNA was harvested by centrifugation at 15,000*g* for 1 hour at 0°C and resuspended in up to 4ml per 10^9 trypanosomes of RNA Mix which had been pretreated with 50ug.ml^{-1} Proteinase K at room temp. for 5 mins. This was incubated at 37°C for 10 mins to digest any protein present and then extracted twice with phenol/chloroform, back extracting the phenol with fresh RNA Mix each time. The RNA was then ethanol precipitated at -20°C overnight before resuspending in RNase-free dH_2O.

Purification of Poly(A)$^+$ RNA

Total RNA was diluted to a concentration less than 100ug.ml^{-1} and heated at 65°C for 10 mins then chilled on ice. An equal volume of 2X Binding Buffer was added and the RNA loaded on a Biorad Econocolumn containing 0.3g oligo(dT) cellulose equilibrated in Binding Buffer. The flow-through was collected, heated at 65°C for 10 mins, chilled on ice and reloaded to ensure that all poly(A)$^+$ RNA bound to the column. The column was washed with 15 column volumes Binding Buffer, followed by 10 column volumes Binding Buffer without SDS and then 10 column volumes Wash Buffer. Poly(A)$^+$ RNA was then eluted with RNase-free dH_2O. Fractions containing the poly(A)$^+$ RNA were identified and quantified by ultraviolet spectrophotometry, pooled, and ethanol precipitated at -20°C overnight before resuspension in RNase-free dH_2O.

2.15.2 Trypanosome DNA

Isolation of High Molecular Weight Trypanosome DNA

1-2 X 10^9 procyclic trypanosomes in a 50ml Falcon tube were washed twice in NET buffer and then pelleted at 3000*g* in a bench centrifuge. Assuming a packed cell volume (PCV) of $10^{10}.ml^{-1}$, the cells were resuspended in at least 10 PCVs NET buffer but usually not less than 5ml. The trypanosomes were then lysed to release their DNA by adding n-lauroyl sarcosine to a final concentration of 3% and gently inverting the tube several times until the suspension cleared. Proteinase K was added to a final concentration of $50ug.ml^{-1}$ and the lysate incubated at 37°C for 30 minutes to digest the protein present. The lysate was then extracted twice with phenol/chloroform, mixing very gently by rolling the tube to minimize shearing. After ethanol precipitation at room temperature the DNA was spooled out, rinsed briefly in 70% ethanol and resuspended in TE(10:1). Any contaminating RNA was removed by adding $100ug.ml^{-1}$ RNase A and incubating at 37°C for 30 minutes. The RNase was removed by 30 minutes digestion with $50ug.ml^{-1}$ Proteinase K at 37°C followed by phenol/chloroform extraction. The DNA was then reprecipitated with ethanol, spooled out and resuspended in TE.

Preparation of Kinetoplast DNA from Trypanosomes

The kinetoplast DNA (kDNA) is composed of a tight network of catenated circles which allows its purification from nuclear DNA by differential centrifugation.

A total genomic DNA preparation from M15 1/148 procyclic cells was diluted to 3ml in TE and then sheared by passing it through a 14G syringe needle which reduces the size of the nuclear DNA but does not affect the tight network of the kinetoplast DNA. This was split into two 1.5ml microcentrifuge tubes and spun at 27000*g* at 4°C for 1 hour in a Beckman JA20 rotor using rubber adaptors. The pelleted material containing the kDNA was resuspended in 0.5ml TE by rocking it overnight and then respun to ensure purification from any contaminating nuclear DNA. After gentle resuspension in 400ul of TE, the two aliquots were pooled and extracted twice with phenol/chloroform and then once with chloroform. The kDNA was then precipitated with 2.5vols. ethanol on ice, rinsed in 70% ethanol and resuspended in 20ul TE.

2.15.3 Plasmid DNA

Large-Scale Plasmid Isolation

A modification of the alkaline-lysis method of Birnboim & Doly (1979) was used for large scale plasmid DNA preparations. A 200ml overnight culture of plasmid-bearing cells was pelleted by centrifugation at 3,000*g* for 5 mins.. The pellet was resuspended in 5ml BDI solution and incubated on ice for 5 mins.. 10ml of freshly prepared BDII solution was added, mixed by gentle inversion, and incubated on ice for 5 mins.. 7.5ml BDIII solution was then added, again mixed gently and the lysate left on ice for 10 mins.. The white floccular precipitate was spun down by centrifugation at 31,000*g* for 20 mins. at 4°C and the supernatant liquid poured into a fresh tube through glass wool. The plasmid DNA was precipitated at room temperature by the addition of 0.6 volumes isopropanol, left for 10 mins. and then centrifuged at 12,000*g* for 15 mins. at room temperature. The resulting pellet was carefully dried and resuspended in 8ml TE.

The plasmid DNA was further purified by CsCl/EtBr equilibrium density gradient centrifugation: 9g CsCl was dissolved in 8ml plasmid-containing solution and 1ml of 10mg.ml^{-1} EtBr was added. After brief centrifugation to remove any insoluble precipitate formed, the suspension was transferred to a 12ml polypropylene ultracentrifuge tube and spun at 49,000 rpm (220,000*g*) at 20°C for 16 hours in a Beckman Ti70 rotor. The plasmid band was removed by hypodermic syringe and extensively extracted with n-butanol to remove the ethidium. The CsCl was eliminated by diluting the plasmid solution 3-fold in TE and precipitating the DNA with 2 volumes of ethanol.

Preparation of M13 RF DNA was essentially the same except that an overnight culture of *E.coli* JM101 was diluted 1:20 in 400ml 2XYT and shaken at 37°C for about 2 hours until the OD_{650} was about 0.45. A 2.5ml overnight culture of a purified M13 plaque in JM101 was added and the flask shaken for a further 4 hours at 37°C. After spinning down the cells as above, they were washed to remove any contaminating M13 phage by resuspending in 100ml D&M salts and respinning. The RF form DNA was then purified exactly as for plasmid DNA.

Small-Scale Plasmid Isolation

1.5ml of an overnight culture of cells containing the desired plasmid was pelleted in a microcentrifuge for 1 min. and the cell pellet resuspended in 100ul BDI. After incubation on ice for 5 mins, 200ul BDII was added, the tube contents mixed gently and left on ice for another 5 mins. 150ul of BDIII was added and, after 10 mins on ice the tube was spun for 5 mins at 12,000*g*. 400ul of the plasmid-containing supernatant phase was withdrawn to a fresh tube and extracted with an equal volume of phenol/chloroform before the addition of 1ml ethanol. After precipitation on ice for 10 mins, the plasmid DNA was recovered by centrifugation at 12,000*g* for 10 mins., washed in 70% ethanol, dried and resuspended in 30ul TE containing 10ug.ml^{-1} RNase A.

Preparation of Plasmid DNA for Sequencing

The purified small scale plasmid DNA preparation was incubated at 37°C for 15 mins to digest contaminating RNA and extracted once with phenol/chloroform. After ethanol precipitation it was resuspended in 20ul TE(10:0.1) and 20ul of 2X Denaturing Buffer was added. This was incubated at room temp. for 5 mins to denature the DNA and was then rapidly neutralized by the addition of 4ul of 2.5M NH_4Acetate, pH4.5. The DNA was recovered by ethanol precipitation and resuspended in 7ul TE(10:0.1), ready for sequencing.

Where CsCl-purified plasmid DNA was available, 2ug of the plasmid could be directly denatured and taken through the subsequent steps as indicated above.

An alternative to this method was that of Hsiao (1991), where RNase was not added to the small scale DNA preparation but instead, the DNA was precipitated for a second time, resuspended in 25ul TE, and 5ul was mixed directly with 1ul of primer and 1ul 1N NaOH. After incubation at 37°C for 10 mins to hydrolyse any RNA present and denature the template simultaneously, 1ul of 1N HCl was added, followed by 2ul of 5X Sequenase reaction buffer and incubated for a further 5 mins at 37°C to ensure annealing of the primer. Sequencing reactions were then performed as normal.

2.15.4 Phage DNA

Large-Scale Purification of Bacteriophage Lambda DNA

An overnight culture of plating cells was diluted 1/50 into 500ml prewarmed L-broth supplemented with 0.2% maltose and 10mM $MgSO_4$ and shaken at 37°C until the $O.D._{600}$ was about 0.3. High titre phage stock was added at a multiplicity of infection of around 0.2 and the culture shaken until it lysed. 10ml chloroform was added and the culture shaken for a further 30 mins. It was then cooled to room temp. and DNase I and RNase A each added to a final concentration of $5ug.ml^{-1}$ and left to digest for 30 mins. NaCl was then added to a concentration of 1M and the lysate chilled on ice for 1 hr. After spinning at 4,000*g* for 15 mins, 10% (w/v) polyethylene glycol (PEG) was added to the supernatant phase ,mixed gently to dissolve, and then left to precipitate overnight at 4°C. The phage precipitate was recovered by centrifugation at 12,000*g* for 5 mins at 4°C, concentrated by respinning, and then gently resuspended in 4ml TM. After chloroform extraction, the phage were loaded on a glycerol step gradient made up of 3ml 40% glycerol in TM with a 4ml 5% glycerol cushion and pelleted by spinning at 112,000*g* for 60 mins in a Beckman Ti70 ultracentrifuge rotor.

Small-Scale Preparation of Bacteriophage Lambda DNA

100ul of the appropriate plating cells were incubated with 50ul of a picked plaque at 37°C for 15 mins and then added to 10ml of L-broth supplemented with 10mM $MgSO_4$ and 0.2% maltose in a 50ml Falcon tube. The culture was shaken at 37°C overnight and centrifuged at 5,000*g* next morning to clear the lysate. 5ml of this cleared lysate were added to 5ml fresh cell overnight culture and incubated at 37°C for 15 mins before adding 100ml of prewarmed L-broth containing 0.2% maltose and 10mM $MgSO_4$ and shaking at 37°C until lysis again occurred. Chloroform was then added to 5% and the flask shaken for a further 15 mins. DNase 1 and RNase A were each added to a final concentration of $10ug.ml^{-1}$ and, after 30 mins digestion at 37°C, NaCl was added to a concentration of 1M. Cellular debris was removed by centrifugation at 3,000*g* for 10 mins at 4°C and PEG was added to the phage supernatant phase to a concentration of 10% (w/v). This was mixed gently until it had all dissolved. The phage were allowed to precipitate at 4°C overnight and recovered by centrifugation at 12,000*g* for 10 mins. and all supernatant fluid was removed from the tube walls. The phage pellet

was resuspended in 500ul SM and extracted once with chloroform to remove any remaining PEG. DNA was isolated by several extractions with phenol/chloroform followed by ethanol precipitation and resuspension in TE.

Preparation of Single-Stranded Template DNA

a) pBluescript

An overnight culture of XL1-blue cells containing the plasmid to be sequenced was diluted ten-fold into 30ml Superbroth and shaken at 37°C for 1 hour until the culture reached an OD_{600} of 0.3. VCSM13 helper phage were added at a multiplicity of infection of 20:1 and the culture shaken at 37°C for a further 6 hours. The cells were then removed by centrifugation at 12,000*g* for 10 mins at 4°C and the phage supernatant phase transferred to a fresh tube which could be stored in the refrigerator overnight. The phage were precipitated by the addition of 6ml 20% PEG/2.5M NaCl, incubating at room temp. for 10 mins, and then centrifugation at 12,000*g* for 10 mins. The resulting pellet was resuspended in 1ml TE(10:0.1), transferred to a microcentrifuge tube and reprecipitated by adding 1/5vol. 20% PEG/2.5M NaCl. The phage were recovered by spinning for 10 mins in a microcentrifuge, removing all of the supernatant fluid, and resuspending the pellet in 110ul TE(10:0.1). DNA was recovered by extracting the phage twice with 50ul phenol followed by 100ul chloroform and ethanol precipitation. After washing the pellets with 70% ethanol, the single-stranded template DNA was resuspended in 15ul TE(10:0.1) and a small aliquot (1-2ul) was run on a 0.8% agarose gel to check recovery.

b) M13

An overnight culture of *E.coli* JM101 was diluted 1/100 in 20ml 2XYT and a single plaque of the M13 clone to be sequenced was picked into it. The culture was incubated at 37°C for 5-6 hours and then transferred to a 40ml centrifuge tube for spinning at 12,000*g* for 10 mins in a Beckman JA20 rotor to remove cellular material. One fifth volume of 20% PEG/2.5M NaCl was added and the remainder of the purification performed as for pBluescript template above, finally resuspending the DNA in 50ul TE(10:0.1).

2.16 Cloning Techniques

Restriction Endonuclease Digestion of DNA

Plasmid or phage restriction digests were usually performed in a total volume of 20ul containing 0.5-2ug DNA, 2ul of the appropriate 10X restriction buffer and not more than 10% (v/v) restriction endonuclease. The reactions were incubated at the appropriate temperature, as indicated by the suppliers, for 1-2 hours, after which time digestion was usually complete. Multiple digests were performed where possible in a restriction buffer suitable for all enzymes but where this was not the case, the DNA was incubated with the enzyme using the lower salt concentration first, the reaction destroyed by heat inactivation or phenol extraction, and then the salt concentration increased before addition of the second enzyme. Digestion was arrested by either the addition of 5X FSB or, if serial treatments were required, phenol extraction and ethanol precipitation.

Genomic DNA digests were performed in varying volumes with a DNA concentration of 1ug/20ul and incubated for at least 5 hours in the appropriate restriction buffer.

Dephosphorylation of Vector for Cloning

Ligation can only occur if at least one of the 5'-ends involved in the ligation reaction is phosphorylated. Thus, by removing the 5' phosphate groups from the cohesive ends of the restricted cloning vector, religation of the vector arms is inhibited and so intermolecular ligation betwen the vector and the insert to be cloned is promoted. The enzyme used to perform this dephosphorylation is Calf Intestinal Phosphatase (CIP).

After digestion to produce the desired cloning site, 0.5u of CIP were added and the reaction was incubated for a further 20 mins at 37°C. EDTA was added to 15mM to stop the reaction and the tube was incubated at 70°C to destroy the phosphatase. The DNA was then extracted twice with phenol/chloroform, ethanol precipitated and resuspended in TE (10:1) at a concentration appropriate for ligation.

Ligation

For ligation of insert fragments into plasmid vectors, the reaction was carried out in a final volume of 20ul. When lambda vectors were used, the lambda DNA concentration was kept to a minimum of 500ug.ml^{-1} to allow formation of concatemers, required for packaging into phage heads, and the reaction was usually carried out in a 5ul volume.

Vector and insert DNA were mixed in a 0.5ml microcentrifuge tube with 1/5 vol. 5X Ligation Buffer, 1mM ATP (final concentration) and 1u (5u in the case of lambda ligations) T4 DNA Ligase. Ligations were then incubated at 4°C overnight.

Preparation of Competent *E.coli* Cells

A single colony was picked and grown overnight in L-broth at 37°C with vigorous shaking. This culture was then diluted 1/25 into 100ml L-broth and shaken vigorously until it reached an OD_{650} of 0.45. The cells were chilled on ice for 10 mins, transferred to precooled 40ml centrifuge tubes and then harvested by centrifugation at 1000*g* for 5 mins.. The pellet was gently resuspended in half the initial culture volume of ice-cold 50mM $CaCl_2$ and incubated on ice for 15 mins.. The cells were respun as above, the pellet resuspended in 1/15 the original culture volume of ice-cold 50mM $CaCl_2$, and left on ice for at least 1 hour or overnight to increase competence.

Transformation With Plasmid DNA

The transforming DNA, in a volume not exceeding 10ul and at a concentration below 50ng, was added to 200ul of competent cells and mixed gently by pipetting. The cells were left on ice for 30 mins. before undergoing heat shock at 42°C for 2 mins.. At this stage, 750ul of L-broth or S.O.C. transformation recovery medium was added and the cells incubated at 37°C for 30-60 mins. to allow expression of the antibiotic resistance gene before plating 1-200 ul of the transformation on LB plates containing the appropriate antibiotic plus X-gal and IPTG. The plates were then incubated at 37°C overnight.

Transformation With M13 DNA

This was performed basically as for plasmid DNA, except that, after heat shock, the cells were returned to ice for 30 mins., added to 100ul of JM101 plating cells plus 3ml M13 overlay containing X-gal and IPTG, and plated on MM+G plates at 37°C overnight.

Preparation of Cells for Electroporation

An overnight culture of the cells to be transformed was diluted 1/25 in 100ml L-broth and shaken at 37°C for 3 hours. The culture was then chilled on ice for at least 30 mins and gently pelleted by centrifugation. The cells were washed in 80ml of cold dH_2O and then 20ml cold dH_2O before resuspension in 2ml cold 10% glycerol in water. The cells were pelleted again, the tubes allowed to drain, and then resuspended finally in the remaining glycerol solution in the tube. The cells could then be used immediately or stored at -70°C until required.

Transformation of Bacteria by Electroporation

Ligation reactions were precipitated with ethanol and resuspended in dH_2O before electroporation. Frozen cells were allowed to thaw gently on ice and 1ul of the ligation to be transformed was added to 40ul of cell suspension. This mixture was then transferred to a cold electroporation cuvette with a 0.1cm gap and pulsed in a Biorad Gene Pulser at 25uF, 1.6kV, 400Ω with a time constant in the range of 6-10msec.. 1ml of 2X YT was added immediately to the cuvette to resuspend the cells and these were then transferred to a fresh tube and allowed to recover by incubating at 37°C for 1 hour. Dilutions were plated on LB plates supplemented with the appropriate antibiotic, plus X-gal and IPTG if required, and grown overnight at 37°C.

In vitro Packaging of Lambda DNA

This was carried out according to the manufacturer's instructions in either Gigapack Gold or Gigapack Plus.

Bal-31 Digestion of Genomic DNA

18ug of genomic DNA were mixed with 90ul 2X *Bal*-31 Reaction Buffer and made up to 180ul in dH_2O. The reaction mix was prewarmed to 30°C for 5 mins and then 30ul was withdrawn to a tube containing 3ul 0.25M EDTA on ice (T_0). 5u of *Bal*-31 was added to the remainder of the reaction mix which was incubated at 30°C for 5 (T_5), 10 (T_{10}), 20 (T_{20}), 40 (T_{40}) and 60 (T_{60}) mins, withdrawing 30ul into EDTA at each timepoint. The timepoints were then phenol/chloroform extracted and ethanol precipitated before resuspension in 20ul TE.

Preparation of Nested Deletions

A set of nested deletions of a particular sequence in pBluescript was prepared using the Stratagene Exo III-Mung Bean Nuclease kit exactly as indicated in the manufacturer's protocol using appropriate restriction sites.

Synthesis of Oligonucleotides

Oligonucleotides designed as probes, PCR primers or sequencing primers were synthesized using an Applied Biosystems PCR Mate synthesizer. The sequence of the oligonucleotide to be synthesized was typed into the machine and the synthesis performed on disposable columns conjugated with the 3'-most base of the sequence in a 3' to 5' direction.

After synthesis was complete, the oligonucleotide was removed from the column by incubating the column matrix in 1ml of concentrated ammonium hydroxide for 2 hours at room temp. and then spinning the matrix out by centrifugation for 5sec. in a microcentrifuge. The supernatant phase was transferred to a fresh screwcap vial, made up to 3ml with more of the concentrated NH_4OH and then incubated at 55°C overnight to remove protecting groups used during the synthesis. After cooling to room temp. the oligonucleotide was precipitated with 1/10 vol. 5M NH_4Acetate and 2.5 vols. ethanol at -70°C for at least 30 mins., recovered by spinning at 10000*g* in a JA20 rotor for 20 mins, washed in 70% ethanol and resuspended in 1ml TE. The concentration was measured by taking OD_{260} readings of 1/100 dilutions of the oligonucleotides in TE, assuming 1 OD_{260}= 33ug.ml^{-1}.

Sequencing Reactions

Template DNA (single- or double-stranded) was annealed with the appropriate primer and sequencing reactions were carried out using the Sequenase kit according to the manufacturer's instructions.

Computer Analysis of Sequence Data

Sequence data were analysed using the GCG Sequence Analysis Software Package (Devereux, Haeberli & Smithies, 1984) and release 27.0 of the GenEMBL Databank.

PCR Amplification of the 5'-Ends of cDNAP4

1ug of procyclic mRNA was reverse transcribed using oligonucleotides complementary to the sense strand of cDNAP4. The mRNA was mixed with 1ul RNA Guard, 4ul 5X MMLV-RT Buffer, 2ul 5mM dNTPs, 1pmole of the oligonucleotide and 200u of MMLV Reverse Transcriptase in a final volume of 20ul. The reaction was incubated at 37°C for 1 hour and then diluted to 500ul in TE. 5ul of this cDNA was then combined with 30 pmoles each of the primer used to prime the cDNA synthesis and an oligonucleotide (oligo 464) which contains the 3'-most 20 nucleotides of the trypanosome SL sequence attached to a linker-adapter sequence which had previously been designed for use in constructing a PCR-amplified cDNA library. This was added to 5ul 10X *Taq* Polymerase Buffer (Promega), 250uM dNTPs and 2.5u *Taq* DNA polymerase and made up to 50ul in deionized water. Control reactions, omitting the cDNA or each primer in turn, were also prepared and each reaction was covered with a drop of light mineral oil. The reactions were then placed in a thermal cycler and 25 cycles of amplification performed. The denaturation in each cycle was at 94°C for 1 min and was followed by an annealing for 2 mins at the temperature indicated in the experiment. The primers were then extended at 72°C for 2 mins with a slow ramp rate to prevent premature melting of the freshly annealed primers. The first 3 cycles were performed at a lower annealing temperature as at this stage only the 20 3' residues of oligo 464 would hybridize but after a couple of rounds of amplification the whole primer sequence should hybridize and therefore have a higher melting temperature.

2.17 Analytical Techniques

2.17.1 Preparation and Running of Electrophoretic Gels

a) DNA Gels

Agarose of the required percentage (w/v) was suspended in either tris-acetate (TAE) or tris-borate (TBE) buffer at 1X concentration and heated in a microwave oven until the agarose had dissolved. The solution was allowed to cool to below 50°C before being poured into the desired gel former. Samples were mixed with 1/5 vol. 5X FSB and gels run in the appropriate 1X buffer at 0.5V/cm.

b) Alkaline Gels

Again the required percentage of agarose was weighed out but this time it was suspended in Alkaline Gel Preparation Buffer, allowed to cool to below 50°C after heating, cast and set. The gel was then made alkaline by immersing in 1X Alkaline Running Buffer for 30 mins before loading the samples (in Alkaline Loading Buffer) and running slowly at 30V overnight.

c) RNA (Northern) Gels

2.5g agarose was suspended in 183ml dH_2O and heated to dissolve. The molten solution was then cooled to about 50°C before adding 25ml 10X MOPS and 42ml formaldehyde solution. 6ul RNA samples were mixed with 19ul SB, heated to 65°C for 10 mins., chilled on ice and mixed with 5ul "Northern Blue Juice". The gel was run in 1X MOPS at 200V.

d) Sequencing Gels

42g of urea (final concentration = 6M) was dissolved in 10ml 10X TBE plus 16.7ml of a (29:1 w/w) 30% acrylamide/bis-acrylamide solution by gentle heating. The volume was made up to 100ml with dH_2O (5% gel) and filtered through Whatman No.1 filter paper. 800ul of a fresh solution of 10% ammonium persulphate (APS) and 40ul of TEMED were added and the gel cast between 2 glass plates separated by 0.4mm spacers with a sharks tooth comb inserted in an inverted position to leave room for loading. The gel was allowed to polymerize for at least 2 hours or overnight with the plates clamped and covered with Saran wrap. Before loading samples, the gel was prerun in 1X TBE at 80W for 45 mins to

bring the temperature to about 50°C. Unpolymerized acrylamide and diffused urea were rinsed out of the gel with a syringe and the sharks tooth comb was placed teeth-down into the gel so that the teeth were touching the surface. After loading the samples the gel was run at 65W for the time reqired (1.5 hours for a short run, 4 hours to read further and 6 hours to read further still). The gel was then soaked in 10% methanol/10% acetic acid for 20 mins. to remove the urea before drying down onto Whatman 3MM paper using a vacuum gel drier.

e) Polyacrylamide Gels

The desired percentage of acrylamide gel was prepared from a (29:1 w/w) 30% acrylamide stock in TBE buffer, APS and TEMED were added to facilitate polymerization and the gel was cast between two glass plates in a vertical gel apparatus. The gel was run in 1X TBE at 200V.

f) SDS-PAGE Gels

Resolving gels were cast between two glass plates in a vertical gel apparatus and covered in 0.1% SDS to allow polymerization. A 4% stacking gel was cast above the resolving gel, after rinsing, and the comb inserted. Once the gel had polmerized, the comb was removed and the wells washed out with 1X Running Buffer. Samples were mixed with an equal volume of 2X SDS Loading Buffer and then boiled for 2 mins before loading and the wells filled up with running buffer. Minigels were run at 100V, larger gels at 200V, until the marker dye had almost reached the bottom of the gel.

2.17.2 Staining of Gels

DNA was visualized in agarose gels by staining the gel in 0.5ug.ml^{-1} EtBr, either during electrophoresis or afterwards for 20 mins, depending on whether the gel was only qualitative or to be used for accurate size measurement. The gel was then destained in dH_2O for at least 10 mins before visualization on a UV-transilluminator.

RNA markers from northern gels were visualized by rinsing the gel in dH_2O for 1 hour to remove the formaldehyde before staining in 1ug.ml^{-1} EtBr or acridine orange for 20 mins and destaining in dH_2O overnight.

Protein SDS-PAGE gels were stained in 3mg.ml^{-1} Coomassie Blue (dissolved in destaining solution) with gentle agitation for 1 hour and then destained in several changes of destain (7.5% methanol/7.5% acetic acid) overnight. The gel could then be dried in a vacuum gel drier at 80°C for long term storage. If the gel was to be blotted, staining was omitted but the protein bands could be visualised after blotting by incubating the blot in Ponceau S for 30secs. and then destaining in TBST for 2 mins. The position of the size markers could then be marked in pencil and the staining washed off by further incubation in fresh TBST.

2.17.3 Blotting of Nucleic Acids onto Nylon Membrane

For DNA gels, it was necessary to pretreat the gel before blotting. After staining and photography, the gel was incubated in 0.25M hydrochloric acid for 15 mins to depurinate the DNA and thus enable transfer of high molecular weight molecules. This step was not necessary if only small molecules of up to a few kb in size were to be transferred. The DNA was denatured by immersing the gel in Denaturing Solution for 45 mins with one change and neutralized in Neutralizing Solution, again for 45 mins with a change of buffer. The gel was then ready to be capillary blotted onto Hybond N or Nytran membrane in 20X SSC overnight. After blotting, the DNA was fixed onto the membrane by either baking it at 80°C for 2 hours (non-covalent) or exposing it to UV light on a transilluminator for 3-5 mins (covalent).

RNA gels required no pretreatment and were blotted directly onto Hybond N in 20X SSC overnight. RNA was fixed onto the membrane by UV-crosslinking.

For plaque lifts, a disc or sheet of nylon membrane was placed onto the prechilled plate and the membrane and plate marked to enable orientation of the filter later on. The membrane was left on the plate for 1 min and then transferred, plaque-side-up, to a sheet of filter paper soaked in Denaturing Solution. After 7 mins the membrane was placed on a sheet of filter paper soaked in Neutralizing Solution and left for 3 mins before being transferred to a fresh neutralizing sheet. Finally the membrane was washed briefly in 2X SSC before air drying and then baking at 80°C for 2 hours.

For colony lifts, the membrane was placed on a fresh plate and the colonies streaked out on the membrane in an array. After overnight incubation at 37°C the

membrane was removed and treated as for plaque lifts.

2.18 Hybridization Analysis

2.18.1 Isolation of DNA from Agarose

DNA fragments cut from LMP agarose gels could be used directly in probe preparation or ligations if the agarose was first melted by incubation at 70°C for 10 mins..

Two methods were used routinely to purify DNA fragments cut from standard agarose gels. In both cases it was necessary that the gel had been run in TAE buffer.

a) Geneclean

The gel fragment was treated exactly according to the manufacturer's instructions.

b) Spinning through Glass Wool

A 500ul microfuge tube was punctured at the bottom with a syringe needle and a small amount of siliconized glass wool was placed as a tight plug inside. This tube was then placed in a 1.5ml microfuge tube and the gel fragment put on top of the glass wool plug. The tubes were spun in a microcentrifuge at 6000*g* for 10 mins and the DNA collected in the displaced liquid while the dried out gel fragment remained in the plug.

2.18.2 Preparation of Hybridization Probes

a) Random Hexanucleotide Priming

10 to 50ng of the DNA to be radio-actively labelled (plasmid, gel fragment in LMP agarose or purified gel fragment) in a volume of 17ul was heated in a boiling water bath for 7 mins and cooled slightly before adding 5ul Rxn-C, 2ul [α-^{32}P]dCTP (20uCi; 800 Ci.mmol^{-1}) and 1ul Klenow fragment of DNA Polymerase (1u). The reaction was incubated at room temp. for 3-16 hours and then heated at

70°C for 10 mins to stop the reaction before separating from unincorporated counts through Sephadex G50. The purified probe was then denatured by heating in a boiling water bath for 5 mins and chilling rapidly on ice before adding to the hybridization mix.

b) End-Labelling of Oligonucleotide Probes

10pmoles of the oligonucleotide was mixed with 2.5ul 10X Kinase Buffer, 5ul [γ-^{32}P]dATP (50uCi, 6000Ci.mmol^{-1}), 10u T4 Polynucleotide Kinase and made up to 25ul in dH_2O. The reaction was incubated at 37°C for 1 hour and the enzyme then inactivated by heating to 70°C for 10 mins before separating the probe from unincorporated counts through Sephadex G25.

c) First-Strand cDNA Probes

50uCi each of [α-^{32}P]dATP and [α-^{32}P]dCTP (each 3000Ci.mmol^{-1}) were dried down to a volume of 2ul in a speedivac dessicator and added to 5ul 5X MMLV-RT Buffer, 1.25ul each 5mM dTTP and dGTP and 2.5ul oligo(dT) primer (100ug.ml^{-1}). The RNA to be used as template for the synthesis (20ug total RNA or 1ug mRNA) was freeze-dried to a volume of 4ul and added to the reaction mix with 400u MMLV Reverse Transcriptase, making the volume up to 25ul with RNase-free water. The reaction was incubated at 37°C for 1.5 hours and then 1ul each of 10mM dCTP and dATP chase was added before continuing the reaction for a further hour. An equal volume of 0.6N NaOH/20mM EDTA was added and the cDNA incubated at 65°C for 30 mins to hydrolyse the RNA before separating the probe through Sephadex G50. This method produced probes of 3-4 X 10^7cpm from 20ug total RNA.

2.18.3 Separation of Probes from Unincorporated dNTPs through Sephadex

A pasteur pipette was plugged with a small quantity of glass wool and filled to the neck with a grade of Sephadex (equilibrated in TE) appropriate to the size of probe to be purified. The probe labelling reaction was made up to 100ul in TE and allowed to enter the column. The column was then washed in TE and 250ul fractions collected in microcentrifuge tubes. The fractions containing the first peak of radiolabelled material, corresponding to the probe fraction, were assessed, either by Cerenkov counting in a scintillation counter or by a hand-held monitor, and pooled.

2.18.4 Hybridization

DNA and plaque blots were hybridized in aqueous buffers while RNA blots were hybridized in formamide buffers to preferentially select for the more stable RNA-DNA hybrids. Filters were prehybridized in prehybridization solution (5ml per 100 cm^2 of membrane) for at least 1 hour at the appropriate hybridization temperature for the probe to be used. This solution was then discarded and fresh prehybridization solution containing the denatured probe was added (~2ml per 100 cm^2 of membrane). The filters were hybridized at the appropriate temperature overnight either in polythene bags in a shaking water bath or in a rotary hybridization oven.

Filters were washed by briefly rinsing in 2XSSC at room temperature followed by sequential washes in 2XSSC at 65°C for 2X 15 mins, 2XSSC/0.1% SDS at 65°C for 30 mins and then a stringency wash for 1 hour. For high stringency for homologous probes this was 0.1XSSC/0.1% SDS at 65°C. When heterologous probes were used the washing temperature for all stages was reduced as described in the text and for oligonucleotide probes the salt concentration was increased to 5XSSC initially as well as dropping washing temperature.

2.18.5 Stripping of Nucleic Acid Blots for Rehybridization

Probes were removed from blots by boiling the filter in 0.1% SDS for 5-10 mins and cooling slowly to room temperature. The efficiency of stripping could be assessed by autoradiography. The filters were then ready to prehybridize before incubation with the next probe.

2.18.6 Autoradiography

^{32}P-labelled blots and gels were exposed to X-ray film either at room temp. for highly radio-active material or at -70°C with intensifying screens for most applications. Sequencing gels, which use ^{35}S were exposed directly onto Amersham Hyperfilm β-max at room temp. as this film contains a high density emulsion which improves resolution of the low energy β–particles from ^{35}S. Standard films were processed in a Kodak X-OMAT developing machine but, due to the extended fixing times required for the Amersham β-max film, this was developed by hand under safelight conditions. The film was developed for 5 mins.

in concentrated D19 developer and development was then stopped by placing the film in water for 2 mins.. Fixing was carried out for 10 mins. in a 1:3 dilution of Amfix and the film then extensively washed before drying in a drying oven.

2.19 Protein Analysis

Expression of Fusion Proteins in pGEX

An overnight culture of XL1-blue cells containing the pGEX construct to be expressed was diluted 1:10 in 100ml of L-broth supplemented with 100ug.ml^{-1} ampicillin and grown at 37°C shaking for 1 hour. IPTG was then added to a final concentration of 0.1mM to induce expression of the fusion protein and the culture grown for a further 3 hours. The cells were centrifuged at 12,000*g* for 5 mins at 4°C in a Beckman JA20 rotor and resuspended in 2ml of MTPBS containing 1% Triton X-100. This suspension was then lysed by sonicating on ice for 4 times 10sec. and spun in microcentrifuge tubes at 12,000*g* for 15 mins at 4°C. The supernatant phase was transferred to fresh tubes and the pellet resuspended in 1ml MTPBS/Triton. Samples of pellet and supernatant fractions could then be analyzed in SDS-PAGE to identify the fraction containing the fusion protein.

Selection of Fusion Proteins with Glutathione Beads

Soluble GST-fusion proteins can be selected by binding to Glutathione Sepharose 4B (Pharmacia; Smith & Johnson, 1988) and then eluting with excess glutathione. The supernatant fraction of the expression culture was mixed with 25ul of beads and incubated at room temp. for 2 mins. The beads were then pelleted by centrifugation and washed 5-10 times in MTPBS/Triton. This selection procedure was then repeated with fresh beads. At this stage the fusion protein can be eluted from the beads by incubating with excess glutathione but resuspending the beads in MTPBS, adding an equal volume of 2X SDS Loading Buffer and boiling for 2 mins is sufficient to release the fusion protein for analysis of recovery on SDS-PAGE.

Raising Antisera Against Fusion Proteins

50ug of the fusion protein/nitrocellulose suspension was made up to 100ul (for mice) or 250ul (for rats) in MTPBS/Triton in a 1.5ml microcentrifuge tube and an equal volume of Complete Freund's Adjuvant was added. The mixture was then vortexed vigorously until a thick emulsion was formed which did not separate into phases.

A small amount of blood was removed from the animals before immunization to prepare preimmune serum. The antigen emulsion was then injected intraperitoneally. The animals were left for 14 days to raise a primary immune response which was tested by immunofluorescence on acetone-fixed procyclic trypanosomes and they were then boosted with the same dose of antigen, this time emulsified with Incomplete Freund's Adjuvant. Again the animals were left for 14 days before taking more serum and reboosting if desired.

Preparing Serum from Whole Blood

Whole blood was removed by syringe and allowed to clot at room temp. for at least 1 hour. The unclotted material was then transferred to microcentrifuge tubes and spun at 10000*g* for 2 mins to remove the remaining blood cells. The serum supernatant was then transferred to fresh tubes and stored at -20°C.

Immunofluorescence

5-10ul of a suspension of trypanosomes was spotted onto a glass microscope slide, spread around a little and then pipetted off again to leave a thin smear of trypanosomes. Duplicate spots were made for each antiserum to be tested and then allowed to air dry before fixing for 5 mins in acetone. Each spot was drawn round with a Texel paint pen (Mark-Tex Corp., New Jersey) to create a small well and then incubated with 10ul of 1/10 or 1/100 dilutions of antiserum in PBS, in a humid environment at room temp. for 30 mins. The wells were then washed extensively in PBS before adding 10ul of a 1/50 dilution of rabbit anti-mouse (or rat) IgG-FITC-conjugate containing 20ug.ml^{-1} EtBr to allow visualisation of the nucleus and kinetoplast of unstained trypanosomes, and incubating again for 30 mins. After washing extensively in PBS, excess liquid was wiped away from the wells and 10ul of a 50:50 solution of PSG:glycerol was added before applying a coverslip. The smears could then be examined for fluorescence using Ploem

illumination with an FITC filter cube on a Leitz Orthoplan large field microscope and photographed.

Western Blots

After running proteins out on SDS-PAGE the gels were electroblotted onto two sheets of nitrocellulose in Electroblotting Buffer at 0.25amps for 2 hours, after which time the filters were ready to be probed.

Probing of Western Blots

The blot was first incubated in Ponceau S solution for 30sec. and then rinsed in TBST to stain the proteins present and identify the position of the molecular weight markers. The position of the markers was marked in pencil and the blot was then destained by washing several times in TBST. The filter was blocked in a 5% Marvel milk solution containing 10% (v/v) horse serum in TBST for at least 1 hour at room temp., or overnight at 4°C, and then incubated with the antiserum, at the appropriate dilution, in 2ml of blocking solution in a rolling Falcon tube, for 2 hours at room temp.. The blot was washed 3 times in TBST for 5-10 mins before adding the appropriate dilution of HRP-conjugated second antibody in blocking solution and incubating for 30 mins. The blot was washed again 3 times in TBST and then briefly rinsed in water before adding 10ml of a 50:50 mix of Amersham ECL Developing Solutions 1 & 2, which had been warmed to room temp., and shaking back and forth for 1 min. The blot was placed in a plastic bag and exposed to X-ray film for the necessary time to allow development of the chemoluminescent signal. The film was then processed as normal.

2.20 Library Construction

2.20.1 Procyclic cDNA Library

Construction of a Procyclic cDNA Library in Lambda ZAP II

The cDNA was synthesized using the Stratagene Uni-Zap cDNA Synthesis

Kit exactly according to the manufacturer's instructions, using 5ug of *T.congolense* procyclic poly(A)$^+$ RNA. First- and second-strand reaction controls were analyzed by alkaline agarose gel electrophoresis and autoradiography. The cDNA was size-fractionated through a column of Sepharose CL-4B prepared in a long 1ml plastic pipette and equilibrated in STE buffer. Eighteen 2-drop fractions were collected in microcentrifuge tubes and analysed by Cerenkov counts in a scintillation counter. Two peaks of radioactivity should have been observed, representing radiolabel incorporated into cDNA and unincorporated dNTPs but in fact there were 3 peaks. The first 2 peaks may have represented, respectively, full-length cDNA molecules and partially synthesised products. Both of these peaks of activity were pooled separately but in practice only the first peak was used to make the library.

One half of the size-fractionated cDNA was ligated to 1ug of lambda uni-Zap arms at 12°C overnight and each half of this was packaged in the Gigapack Gold Packaging kit according to the manufacturer's instructions.

Screening the cDNA Library with the Monoclonal Antibodies

1000pfu of the cDNA library was plated out on *E.coli* XL1-Blue on each of 10 LB plates and incubated at 42°C for 3.5 hours. Meanwhile, 2 nitrocellulose discs per plate were soaked in 10mM IPTG for 10 mins and then air dried. One nitrocellulose disc was placed on each plate and incubation was continued at 37°C for a further 2 hours, leaving the lid open for the last 20 mins. The filter was keyed to the plate with ink and then removed. The second filter was then placed on the plate and again incubated at 37°C for 3 hours, leaving the lid open for the last 20 mins. Once the filters had been removed from the plate they were washed in TBST for 30 mins and then stored in sealed bags until ready to probe with the mabs.

The filters were blocked in 3% Marvel milk in TBST for 60 mins and then incubated with the mab, diluted 1:100 in HST, for 1 hour. Excess antibody was washed off by rinsing the filters for 2 times 5 mins in TBST, 5 mins in HST and then 2 times 5 mins in TBST. HRPconjugated anti-mouse IgG was diluted 1:50 in HST and the filters incubated in this solution for 60 mins before washing again as above and then completing a final rinse in TBS. 18mg 4-chloro-1-napthol was dissolved in 6ml methanol and added to 94ml TBS + 25ul 30% hydrogen peroxide and the filters were agitated gently in this substrate for 30 mins before stopping the reaction by rinsing in several changes of water.

2.20.2 Genomic Library

Construction of a Size-Selected *Sau* 3AI Partial Genomic Library

a) Titration of *Sau* 3AI Partial Digests

10ug of *T.congolense* M15 1/148 genomic DNA was mixed with 10ul of 10X React 4 buffer and made up to 100ul in water. 20ul of this was transferred into tube 1 and 10ul into each of tubes 2 to 9 on ice. 4u of *Sau* 3AI were added to tube 1, mixed and then 10ul of the reaction were transferred to tube 2. The serial dilutions were repeated to tube 8, leaving tube 9 as an undigested control. Each tube was incubated at 37°C for exactly 60 mins and the reactions then stopped by adding EDTA to 20mM on ice. 2ul of 5X FSB were added to each tube and the reactions analysed on a 0.3% agarose gel.

The gel was stained with ethidium bromide and photographed. The tube producing the most fluorescence in the 18-22kb size range was identified and the concentration of *Sau* 3AI in this tube calculated. In order to obtain the maximum number of molecules in this size range the enzyme concentration should be half that calculated (Seed et al, 1982)).

b) Large-Scale *Sau* 3AI Partial Digestion

100ug of the same M15 1/148 genomic DNA were digested using the concentration of *Sau* 3AI calculated from the titration experiment and an additional 100ug were cut with 2-fold higher and lower concentrations of the enzyme. The reaction mixes were prewarmed at 37°C before addition of the enzyme and incubated at 37°C for exactly 60 mins as before. The reactions were stopped by the addition of EDTA to 20mM and 1ug of the digested DNA from each analyzed on a 0.4% agarose gel.

The partial digests were pooled, extracted twice with phenol/chloroform and ethanol precipitated at room temp.. The DNA was then spooled out and resuspended in 100ul TE.

c) Size-Fractionation of Partially Digested DNA

A linear 1.25M to 5M NaCl gradient in TE was prepared in a SW28

ultracentrifuge tube using a gradient former and the partially digested DNA loaded on top. The gradient was spun at 150,000*g* at 4°C overnight in a Beckman SW28.1 rotor before puncturing the bottom of the tube with a 21G needle and collecting 16X 0.5ml fractions.

20ul of each fraction was analysed on a 0.4% agarose gel and good size fractionation was observed. The fractions containing the DNA in the correct size range (18-22kb) were pooled and dialysed for several hours against 5l of TE to reduce the salt concentration. The DNA was then ethanol precipitated without the addition of salt and resuspended at a concentration of 250ug.ml^{-1}.

d) Preparation of λEMBL4 Arms

λEMBL4 DNA, at a concentration of 150ug.ml^{-1}, was first incubated in Annealing Buffer at 42°C for 1 hour and then incubated with T4 DNA Ligase at 37°C for 1 hour to ensure annealing of the *cos* ends which is required for packaging into phage heads. After phenol/chloroform extraction and ethanol precipitation, the EMBL4 DNA was then double digested with *Bam* HI and *Sal* I. The *Bam* HI creates *Sau* 3AI-compatible ends for ligating to the genomic DNA, while the *Sal* I destroys the "stuffer" fragment of this lambda replacement vector and thus increases the likelihood of creating recombinant phage molecules. The digest was phenol/chloroform extracted, ethanol precipitated and resuspended in TE at a concentration of 500ug.ml^{-1}.

CHAPTER 3

PROCYCLIN-HOMOLOGOUS SEQUENCES

3.1 Introduction

Several monoclonal antibodies (GUGM2.1, GUGM2.2, etc.), from a collection which was raised against procyclic *Trypanosoma congolense* (F.A. Lainson, unpublished), strongly label living procyclic trypanosomes in immunofluorescence, have agglutinating activity and detect a diffuse band of 29-36kDa in Western blots of trypanosome lysates subjected to SDS-PAGE. The protein(s) identified by this set of mabs has several properties which resemble those of the major procyclic surface antigen on *Trypanosoma brucei* - procyclin or PARP (F.A. Lainson, unpublished, Kilbride, 1992). These include higher solubility in CHAPS than in NP40 - *i.e.*,cell lysates which are enriched for membrane proteins, an inability to stain with any conventional protein stain except Stains-All, a low isoelectric point indicating an acidic protein, migration as a diffuse band in SDS-PAGE, abundance and immunodominance (Clayton & Mowatt, 1989). It was therefore not unreasonable to predict that the protein detected by the set of mabs was the *Trypanosoma congolense* equivalent of procyclin. The immunodominant epitope of procyclin in *T.brucei* is the (ProGlu) repeat region (Richardson *et al*, 1988) and this appears to be species-specific as antisera raised against this epitope do not cross-react with other species (Richardson *et al*, 1986). The mabs against *T.congolense* did not detect any epitope on the one stock of *T.brucei* tested but did react with the procyclic stage of all the *T.congolense* stocks tested and therefore are probably also species-specific. In their original paper on the molecular cloning of procyclin, Roditi *et al* (1987) probed Southern blots of genomic DNA from several other species of trypanosomatid, including *T.congolense*, with the procyclin cDNA, pPRO2001, and detected several bands at reduced stringency. This result suggested that there are homologues of the procyclin genes in other species.

The first experiment aimed at isolating the *T.congolense* gene was therefore to ascertain whether procyclin sequences were present in the available trypanosome stocks.

3.2 Are Procyclin-Like Sequences Present in *Trypanosoma congolense*?

Genomic DNA from *T.congolense* TREU 1627 procyclic and epimastigote stages was digested with *Eco* RI and *Hin*d III and electrophoresed on a 0.8%

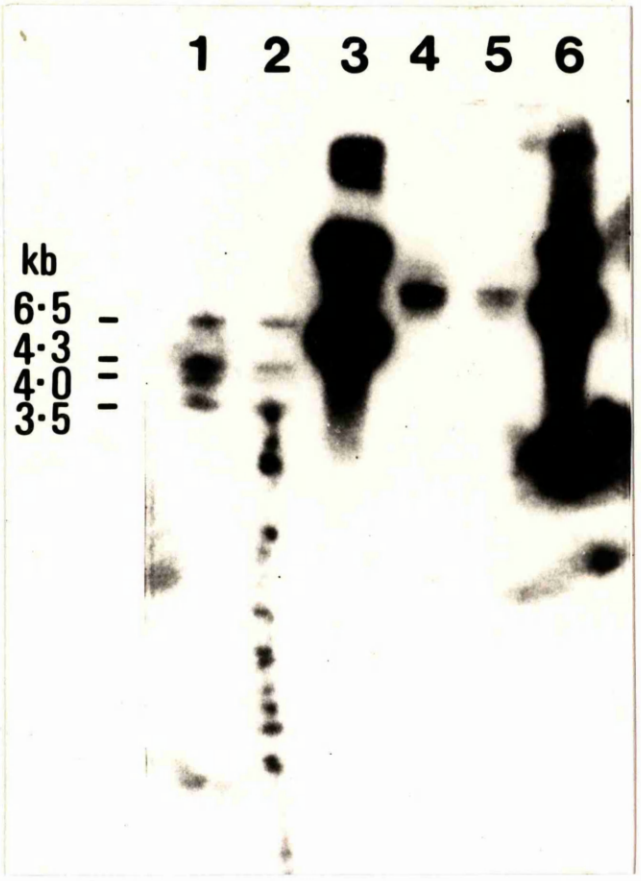


Fig. 3.1 Southern Blot Showing *T.congolense* has Sequences Homologous to *T.brucei* Procyclin.

Genomic DNA was digested with *Eco* RI (lanes 1-3) and *Hind* III (lanes 4-6), separated on a 0.8% agarose gel and blotted onto Nytran membrane. The blot was hybridized with ^{32}P- labelled pPRO2001 insert prepared by random hexanucleotide priming, washed with 0.1XSSC, 0.1% SDS at 60°C for 1 hour and exposed overnight for autoradiography. Lanes 1,2,4 & 5: *T.congolense* TREU 1627. Lanes 3 & 6: *T.b.rhodesiense* EATRO 2340.

agarose gel with *Eco* RI and *Hind* III digested *T.brucei* DNA run alongside as a positive control. The gel was then Southern blotted onto Nytran membrane and hybridized at 60°C with the radiolabelled 740bp Eco RI insert from pPRO2001, comprising the full-length *T.brucei* cDNA (kindly provided by Dr.I.Roditi). After washing to a stringency of 0.1XSSC at 60°C the blot was exposed for autoradiography. While the intensity of hybridization was much greater for *T.brucei* DNA than for *T.congolense* DNA (Fig. 3.1) there were several bands clearly visible in all tracks, indicating that procyclin-like sequences are present in the *T.congolense* stocks.

The sizes of the *Eco* RI fragments which hybridized to pPRO2001 were all appropriate for cloning into λgt10 and therefore an *Eco* RI genomic library was constructed in that vector in order to attempt to clone one or more of these sequences.

3.3 Construction of a Genomic Library in λgt10

5ug of *T.congolense* TREU 1627 DNA were digested with *Eco* RI and ligated at 4°C overnight to 1ug of λgt10 arms in a final volume of 5ul. 4ul of this ligation was then packaged using the Gigapack Plus packaging kit according to the manufacturer's instructions and the resulting library titrated on *E.coli* L87 and NM514. Insertion of fragments into the *Eco* RI site of λgt10 disrupts the *c*I gene, preventing recombinant phage from being able to lysogenise their host, and thus producing a clear plaque morphology on a permissive host such as L87. *E.coli* NM514 is a high frequency lysogenisation strain which prevents productive infections of lysogenic phage so that non-recombinant λgt10 will not produce plaques on this host. Recombinant phage however cannot lysogenise and therefore do grow on NM514 (Murray, 1983). Growth on this host thus selects for recombinant phages.

The titre of the library was 2.5 X 10^6 pfu, of which 93% were recombinant.

3.4 Isolation of a Clone With Homology to pPRO2001

As *Eco* R1 has a six base-pair recognition sequence it will cut DNA on average every 4096bp. Assuming a genome size of 70Mb (Borst *et al*, 1982) 1-2 X

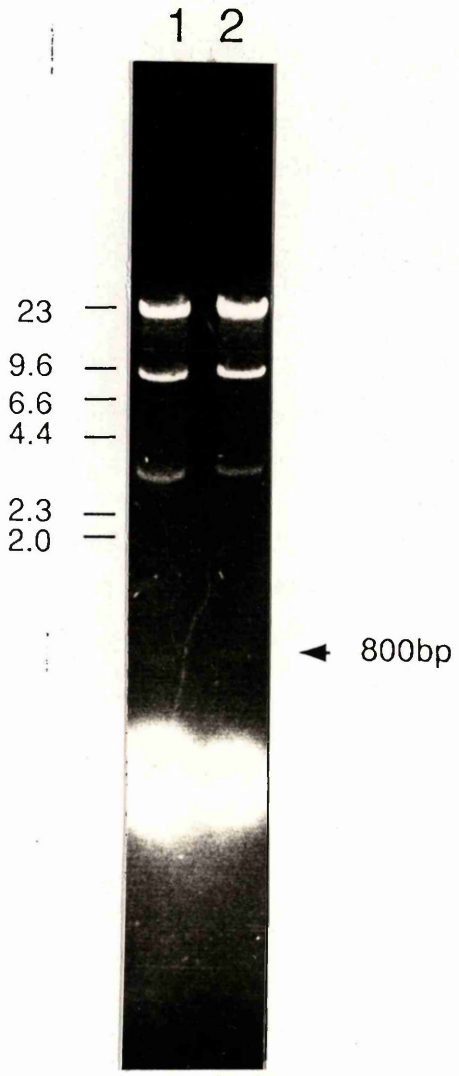


Fig. 3.2 λTcpro has Two *Eco* RI Inserts.

A small-scale DNA preparation from a fresh plaque of λTcpro was digested with *Eco* RI (lanes 1 & 2) and separated on a 0.6% agarose gel. Markers - λ*Hin*d III digest. Position of 800bp fragment which is barely visible in the photograph is indicated.

10^4 pfu should represent approximately one genome equivalent, which contains several sequences that hybridize with the procyclin probe. Approximately 30,000 pfu of the library was plated out on *E.coli* NM514 on a 24 X 24 cm petri dish and after overnight incubation, duplicate lifts were made onto Nytran membranes. In order to try and increase the hybridization signal, these membranes were probed with the pPRO2001 insert fragment at 55°C overnight and then washed at a low stringency (2XSSC at 55°C). After autoradiography, 3 plaques produced a positive signal. The region around these primary positives was cored from the plate and picked into SM buffer containing chloroform. After incubation to allow the phage to diffuse out, they were plated out to a density of several hundred plaques per plate for secondary screening with the same probe. Only one of the primary positives produced a signal on the secondary screen and all subsequent analysis was carried out on this phage, λTcPro. Tertiary screening of this phage did not produce a positive signal and small-scale phage DNA preparations produced only an *Eco* RI fragment of 800bp which did not agree with any of the band sizes detected in the genomic Southern blot. This band did not hybridize to pPRO2001. Since 2 of the primary positive plaques had also failed to hybridize on further screening it was suggested that the sequence(s) of interest might be unstable in NM514. *E.coli* DL491 is a *rec*D host and therefore some recombinogenic sequences are more stable when grown in this strain. A fresh plaque of λTcPro was picked from the secondary screen and grown up for small-scale DNA preparations. An *Eco* RI digest of this DNA produced 2 insert fragments of approximately 4kb and 800bp (Fig. 3.2), the former of which hybridized with the pPRO2001 probe. Only one *Eco* RI fragment should have been in the insert and it is not clear whether the presence of two fragments arose due to partial digestion when the library was constructed or to random insertions at ligation. The size of the 4kb fragment agreed with that of one of the bands detected in the genomic Southern blot (see Fig. 3.1).

3.5 Subcloning of λTcPro

Since λTcPro contained a double insert and it is simpler to resolve restriction maps from plasmids, the 4kb fragment was subcloned into pUC19. An *Eco* RI digest of λTcPro was ligated to pUC19 DNA which had been digested with *Eco* RI and dephosphorylated. The ligation reaction was used to transform competent *E.coli* DS941 to ampicillin resistance and recombinant colonies were selected by the blue-white screen on plates containing X-gal and IPTG. Digests of small-scale plasmid

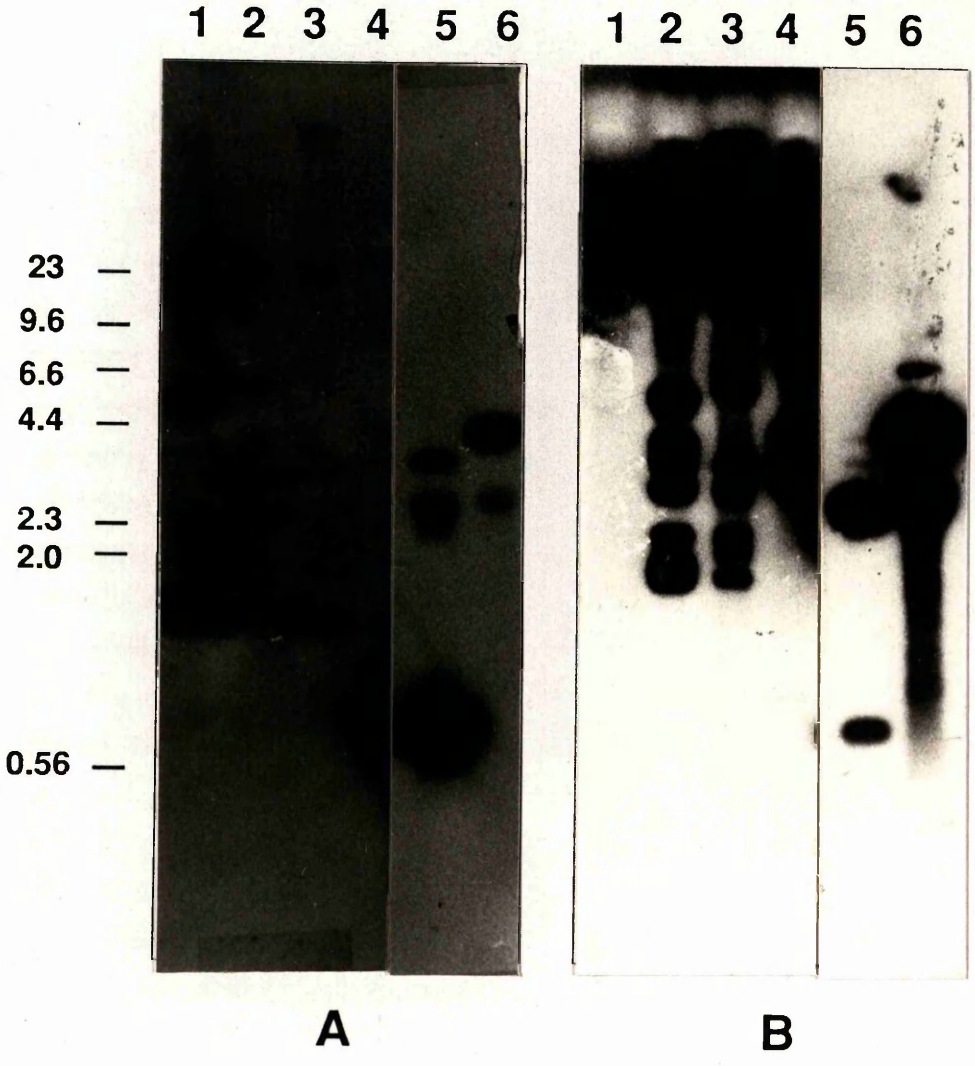


Fig. 3.3 Southern Blot Comparison of pPRO2001 and pTcpro Sequences.

DNA was digested with *Eco* RI and size-fractionated on a 0.8% agarose gel. After blotting, it was hybridized to radiolabelled pPRO2001 insert (A) or the 4kb *Eco* RI fragment of pTcpro (B) at 55°C. The blots were washed in 2XSSC/ 0.1% SDS at 55°C and exposed for autoradiography. Lane 1 - *T.b.brucei*, lane 2 - *T.congolense* TREU 1627, lane 3 - *T.congolense* TREU 1457, lane 4 - *T.congolense* "1/148", lane 5 - pPRO2001, lane 6 - pTcpro. Markers are a *Hin*d III digest of λ DNA.

DNA preparations identified two types of white colony - one containing the 800bp fragment from λTcPro and the other containing the desired 4kb fragment. The latter was designated pTcPro.

3.6 Analysis of pTcPro

In order to check further that pTcPro does indeed contain a homologue of the procyclin sequence, pPRO2001 DNA, pTcPro DNA and genomic DNA from three *T.congolense* stocks and one *T.brucei* stock were digested with *Eco* RI and divided into 2 equal samples which were loaded in duplicate sets of tracks on a 0.8% agarose gel. After blotting onto Nytran, one filter was probed with the pPRO2001 insert and the other with the 4kb fragment from pTcPro. After washing to a stringency of 2XSSC at 55°C and autoradiography, the patterns of hybridization were compared (Fig.3.3). The *T.brucei* probe detected *T.brucei* sequences strongly and *T.congolense* sequences much more weakly, as observed before (section 3.2), while the *T.congolense* probe gave the reciprocal result. The actual pattern of bands detected was identical between the two filters, suggesting that pTcPro does indeed contain sequences similar to procyclin.

3.7 Restriction Mapping of pTcPro

The 4kb *Eco* RI insert in pTcPro was restriction mapped by performing a series of single and double digests of the plasmid which were electrophoresed on agarose gels. By Southern blotting of these gels and hybridization with the pPRO2001 insert fragment, the region homologous to procyclin could then be identified (Fig. 3.4a). This proved to be a 1.0kb *Eco* RI-*Bgl* I fragment at the very 5'-end of the clone which, because the *Bgl* I site is unsuitable for cloning, was further subcloned as a slightly larger 1.15kb *Eco* RI-*Sph* I fragment in pUC19. This subclone allowed finer mapping of this region with blunt-end cutting enzymes which would facilitate subcloning in M13mp18 and 19 of fragments of a size amenable to sequencing.

The only enzymes which cut were *Rsa* I and *Hae* III, each of which produced a fragment of ~800-850bp. An *Rsa* I/*Hae* III double digest produced a band smaller than with either enzyme alone, suggesting that, either they cut once at opposite ends of the clone or that there was more than one site for each enzyme and

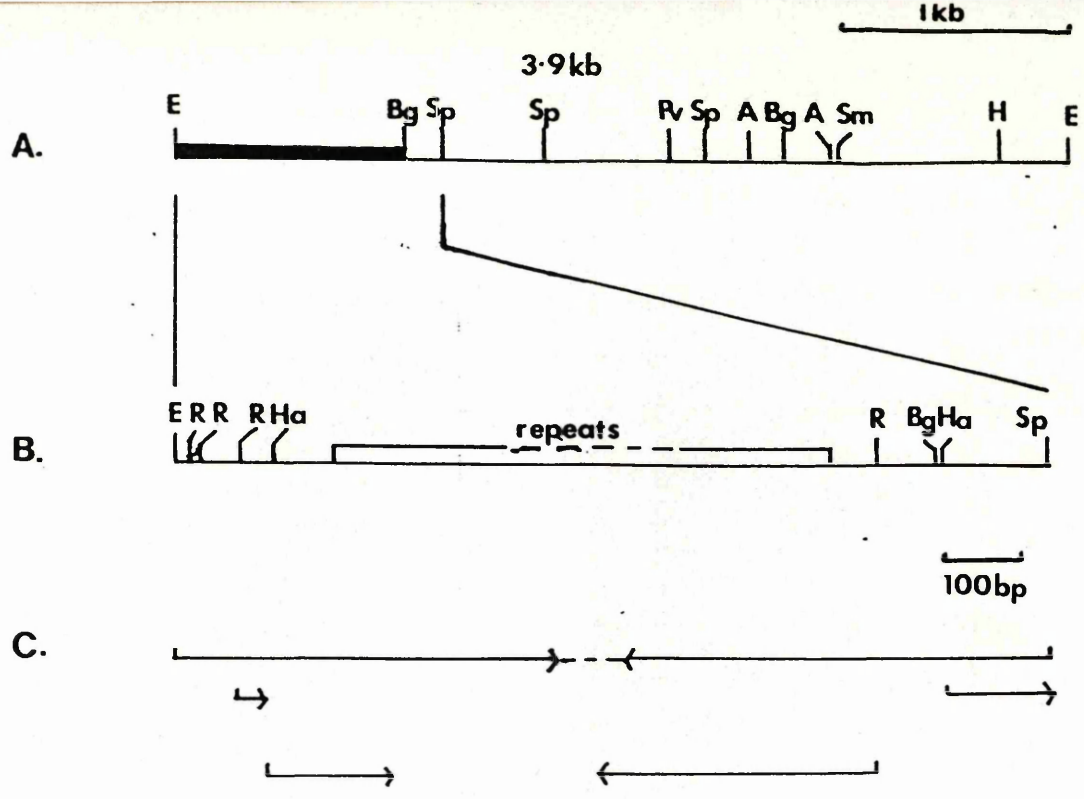


Fig. 3.4

Fig. 3.4 Restriction Map of pTcpro

(A) Shaded area indicates region hybridizing to pPRO2001. E (*Eco* RI), H (*Hin*d III), A (*Acc* I), Sm (*Sma* I), Sp (*Sph* I), Pv (*Pvu* II), Bg (*Bgl* I).

(B) Finer map of *Eco* RI/*Sph* I fragment. R (*Rsa* I), Ha (*Hae* III). Boxed region indicates the region containing the 18bp repeat and the dotted line indicates the uncertainty of a continuous repeat region or 2 blocks of repeats.

(C) Shows extent and direction of regions of subclone sequenced

Fig. 3.5 Nucleotide and Predicted Amino Acid Sequence of the pTcpro *Eco* RI-*Sph* I Fragment.

Sequence from *Sph* I end of clone. (GACCCTGACCCCGACCTT) repeats are boxed and in-frame stop codons indicated by a ∗. The underlined sequences indicate the possible point of overlap of the 2 ends of the clone allowing for reading errors towards the top of the gel and the different number of repeats in different M13 clones. The procyclin cDNA homology is shown above in lower case letters.

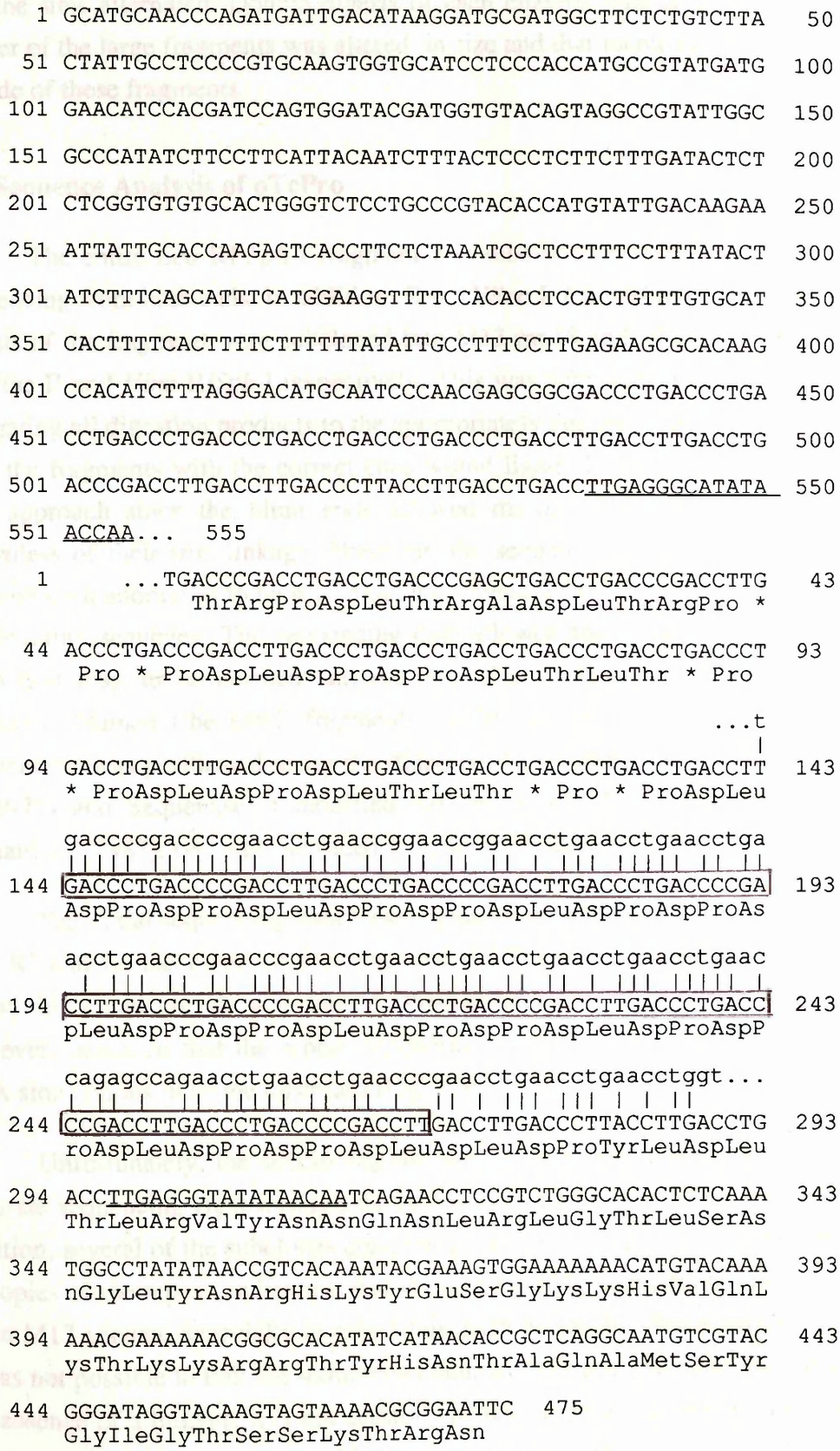

1 GCATGCAACCCAGATGATTGACATAAGGATGCGATGGCTTCTCTGTCTTA 50
51 CTATTGCCTCCCCGTGCAAGTGGTGCATCCTCCCACCATGCCGTATGATG 100
101 GAACATCCACGATCCAGTGGATACGATGGTGTACAGTAGGCCGTATTGGC 150
151 GCCCATATCTTCCTTCATTACAATCTTTACTCCCTCTTCTTTGATACTCT 200
201 CTCGGTGTGTGCACTGGGTCTCCTGCCCGTACACCATGTATTGACAAGAA 250
251 ATTATTGCACCAAGAGTCACCTTCTCTAAATCGCTCCTTTCCTTTATACT 300
301 ATCTTTCAGCATTTCATGGAAGGTTTTCCACACCTCCACTGTTTGTGCAT 350
351 CACTTTTCACTTTTCTTTTTTATATTGCCTTTCCTTGAGAAGCGCACAAG 400
401 CCACATCTTTAGGGACATGCAATCCCAACGAGCGGCGACCCTGACCCTGA 450
451 CCTGACCCTGACCCTGACCTGACCCTGACCCTGACCTTGACCTTGACCTG 500
501 ACCCGACCTTGACCTTGACCCTTACCTTGACCTGACCTTGAGGGCATATA 550
551 ACCAA... 555
1 ...TGACCCGACCTGACCTGACCCGAGCTGACCTGACCCGACCTTG 43
ThrArgProAspLeuThrArgAlaAspLeuThrArgPro *
44 ACCCTGACCCGACCTTGACCCTGACCCTGACCTGACCCTGACCTGACCCT 93
Pro * ProAspLeuAspProAspProAspLeuThrLeuThr * Pro
...t
94 GACCTGACCTTGACCCTGACCTGACCCTGACCTGACCCTGACCTGACCTT 143
* ProAspLeuAspProAspLeuThrLeuThr * Pro * ProAspLeu
gaccccgaccccgaacctgaaccggaaccggaacctgaacctgaacctga
144 GACCCTGACCCCGACCTTGACCCTGACCCCGACCTTGACCCTGACCCCGA 193
AspProAspProAspLeuAspProAspProAspLeuAspProAspProAs
acctgaacccgaacccgaacctgaacctgaacctgaacctgaacctgaac
194 CCTTGACCCTGACCCCGACCTTGACCCTGACCCCGACCTTGACCCTGACC 243
pLeuAspProAspProAspLeuAspProAspProAspLeuAspProAspP
cagagccagaacctgaacctgaacccgaacctgaacctgaacctggt...
244 CCGACCTTGACCCTGACCCCGACCTTGACCTTGACCCTTACCTTGACCTG 293
roAspLeuAspProAspProAspLeuAspLeuAspProTyrLeuAspLeu
294 ACCTTGAGGGTATATAACAATCAGAACCTCCGTCTGGGCACACTCTCAAA 343
ThrLeuArgValTyrAsnAsnGlnAsnLeuArgLeuGlyThrLeuSerAs
344 TGGCCTATATAACCGTCACAAATACGAAAGTGGAAAAAAACATGTACAAA 393
nGlyLeuTyrAsnArgHisLysTyrGluSerGlyLysLysHisValGlnL
394 AAACGAAAAAACGGCGCACATATCATAACACCGCTCAGGCAATGTCGTAC 443
ysThrLysLysArgArgThrTyrHisAsnThrAlaGlnAlaMetSerTyr
444 GGGATAGGTACAAGTAGTAAAACGCGGAATTC 475
GlyIleGlyThrSerSerLysThrArgAsn

Fig. 3.5

that the sites alternated. Double digests of each enzyme with *Bgl* I identified that neither of the large fragments was altered in size and that therefore the *Bgl* I site lay outside of these fragments.

3.8 Sequence Analysis of pTcPro

The entire *Eco* RI-*Sph* I fragment was subcloned into M13 mp18 and 19 for sequencing from both ends. In addition, *Eco* RI/*Rsa* I, *Rsa* I/*Hae* III and *Hae* III/*Sph* I digests of the fragment were subcloned into M13 mp18 and 19 cut with *Eco* RI/*Hin*c II, *Hin*c II and *Hin*c II/*Sph* I respectively. This was achieved by the shotgun method of ligating all digestion products to the appropriately cut vector under the premise that only the fragments with the correct ends would ligate. In hindsight, this was not the best approach since the blunt ends allowed multiple inserts to ligate together, regardless of their true linkage. However, the sequence data from the whole insert allowed such anomalies to be sorted out and several of the subclones did prove useful in obtaining sequence. The sequencing data allowed the correct *Rsa* I and *Hae* III restriction map to be defined and the map (Fig. 3.4b) agrees with the digestion products obtained (the small fragments would have run off the end of the gel). Sequencing was performed using the dideoxy chain terminating method (Sanger et al, 1977) and Sequenase, a modified version of T7 DNA polymerase (Tabor & Richardson, 1987), on single stranded template with universal M13 primers.

The initial sequencing data looked promising, with an 18bp repeat towards the *Eco* RI end of the clone of the sequence $(CCTGACCCCGACCTTGAC)_n$. In one frame this could encode a repeat of $(ProAspProAspLeuAsp)_n$. Further sequencing, however, indicated that the repeat degenerated further up- and downstream, putting TGA stop codons in frame upstream (Fig. 3.5).

Unfortunately, the subcloning did not produce any subclones which allowed accurate sequencing right through the repeat region and beyond for any distance. In addition, several of the subclones containing the repeat region had a different number of copies of the repeat, indicating that the sequence had recombined, as it is prone to do in M13 vectors, especially in a *rec*$^+$ host such as JM101. These factors meant that it was not possible to link the sequences obtained from each end with any certainty. In the absence of a linkage, it is not certain whether the repeat region is one continuous unit or whether there is non-repetitive sequence intervening.

None of the sequence obtained shares any significant homology with *T.brucei*

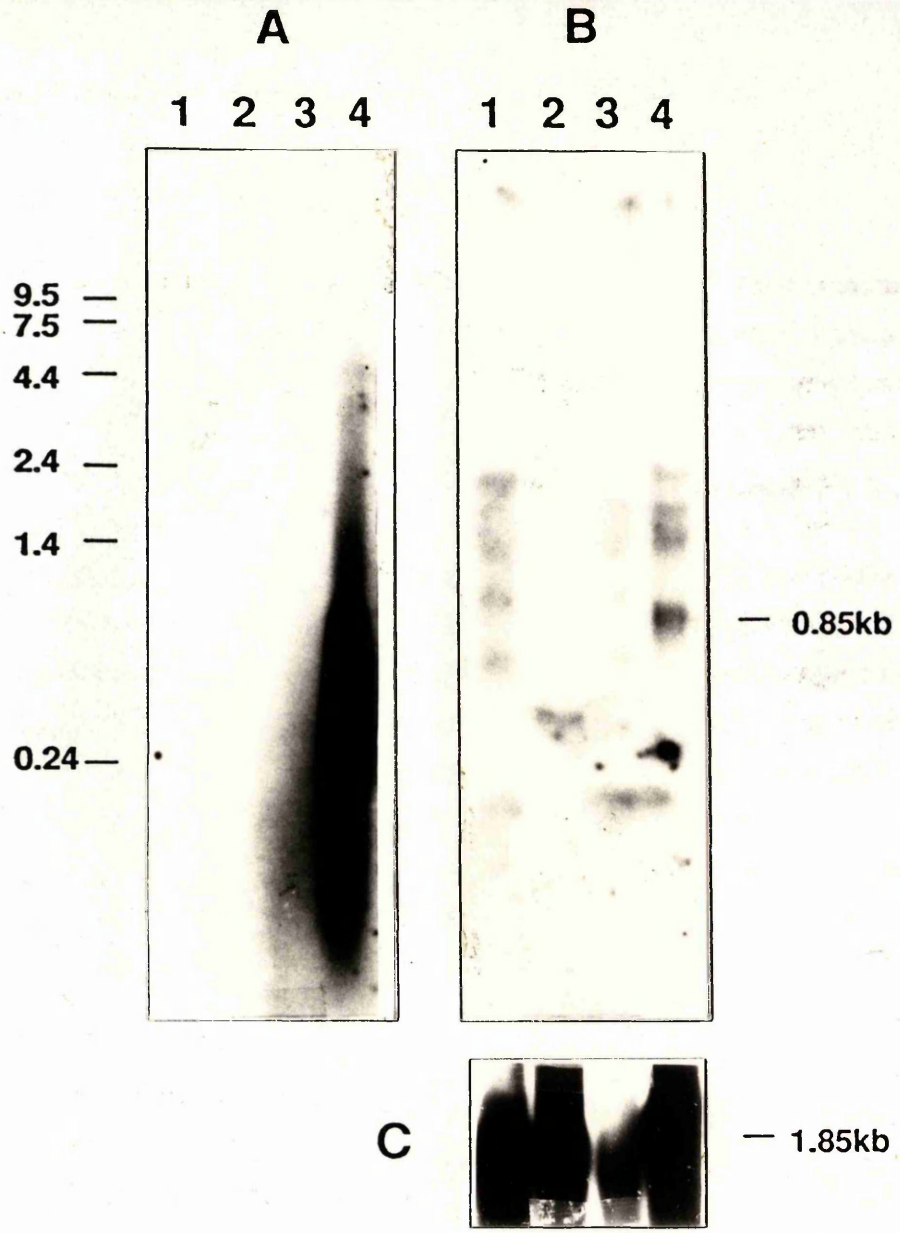


Fig. 3.6 Northern Blot Analysis of *T.congolense* Procyclic RNA.

For each blot, 20ug *T.congolense* total RNA (lane 1), 3ug poly(A)$^+$ RNA (lane 2), 20ug poly(A)$^-$ RNA (lane 3) and 20ug *T.b.rhodesiense* total RNA (lane 4), all from procyclic cells, were size-fractionated on formaldehyde gels and blotted onto Hybond N membrane. Blots were hybridized at 37°C in 50% formamide and washed to 2XSSC, 0.1% SDS at 55°C before autoradiography for 7 days (blots a and b) or 14 days (blot c) at -70°C with intensifying screens. Markers are the BRL RNA ladder.

(a) Blot probed with pPRO2001 insert. The *T.b.rhodesiense* RNA in this gel was observed to be somewhat degraded.

(b) Blot probed with pTcpro *Eco* RI-*Sph* I fragment.

(c) Blot (b) was stripped and reprobed with the 750bp *Hin*d III-*Eco* RI fragment of pTbαβT-1, containing the beta-tubulin gene from *T.brucei*, as a control for the presence of RNA in all tracks. Size of intact β-tubulin transcript is indicated.

procyclin, except in the repetitive region which is 66% homologous overall and about 93% homologous towards the *Eco* RI end of the repeat. If there is a non-repetitive region between two blocks of repeats it is still possible that the homology lies within this unsequenced region but the length of the repetitive region in one of the subclones suggests that there is not enough DNA left to encode a complete copy of the procyclin gene. A T_M analysis of the weak repeat homology suggests that the hybrid formed with the probe would be approaching its T_M at the temperature and salt concentrations used for hybridization and washing (the 14/15bp match at one end which is 73% GC has a predicted T_M of about 58°C in 0.39M Na^+), indicating why only a small amount of the probe hybridized and gave such a weak signal.

3.9 Northern Analysis of pTcPro

Total, poly(A)$^+$, and poly(A)$^-$ *T.congolense* procyclic RNA and total *T.brucei* procyclic RNA were size-fractionated on formaldehyde gels and blotted onto Hybond N membrane. The blots were then hybridized to the insert fragment of pPRO2001 or the *Eco* RI-*Sph* I fragment of pTcPro in 50% formamide at 37°C and washed to a stringency of 2XSSC at 55°C. With pPRO2001, the *T.brucei* RNA contained a hybridizing fragment of about 850bp which is the predicted size for the procyclin message, but did not detect anything in *T.congolense* RNA even on long exposure (Fig. 3.6a). The pTcPro fragment only detected rRNA bands in the *T.congolense* RNA which were also detected in *T.brucei* but in that case the predicted 850bp procyclin mRNA was also visible (Fig. 3.6b). To check that intact RNA was present in all tracks, the blot was stripped of its probe and reprobed with the 750bp *Hin*d III-*Eco* RI fragment from pTbαβT-1 (Thomashow et al, 1983) containing the *T.brucei* beta-tubulin gene. This detected a transcript of the expected size for beta-tubulin (1.85kb) in all of the *T.congolense* RNA tracks (less in poly(A)$^-$ as expected) as well as *T.brucei* (Fig. 3.6c), indicating that intact RNA was present and that therefore the procyclin-like sequence is not expressed detectably in these trypanosomes.

3.10 Discussion

Southern blot analysis indicated that sequences similar to procyclin were present in *Trypanosoma congolense* (Roditi *et al*, 1987 and this study) and one

such sequence has been cloned. The clone insert detected, in DNA from *T.brucei* and *T.congolense*, a set of bands of the same sizes as those detected by the procyclin cDNA. The insert also detected a band on Northern blots of *T.brucei* RNA corresponding in size to procyclin mRNA but neither the *T.congolense* nor the *T.brucei* probe was able to detect any transcript in *T.congolense* procyclics, implying that no copy of the "gene" is expressed.

Sequence analysis of the region of pTcPro which hybridizes to procyclin could detect no significant open reading frame except within a repetitive sequence that could encode a repeat of the peptide (AspProAspProAspLeu). This is not dissimilar from the dipeptide repeat (GluPro) in procyclin in terms of conservative amino acid changes, but the open reading frame did not continue upstream. It is not clear whether there are one or two blocks of the repeats, although the short piece of sequence obtained downstream of the repeats from the *Sph* I end of the clone suggests that the sequence from the two strands can be linked and that therefore there is only one block. Even if there is a non-repetitive region between the two regions of repeat sequenced it would not contain enough DNA to encode a complete copy of a procyclin-like molecule. This lack of protein coding capability is in agreement with the observed absence of any RNA expressed from this region in *T.congolense*.

The presence of different numbers of copies of the repeat in different subclones suggested that the repeats had recombined within the M13 vector. The original lambda clone was also unstable in a *rec*$^+$ host and a reduction in the number of repeats would explain why the 4kb band disappeared while the 800bp fragment remained constant. There was a smear present in the gel in the region of 4kb and this might have been due to the deletion of different numbers of copies of the repeat in individual phage.

The region of similarity to procyclin appears only to reside in the repeat region of pTcpro, implying that it is not a homologue of procyclin. However, only one of the four *T.congolense Eco* RI fragments which hybridized to pPRO2001 has so far been cloned. Could any of the other fragments contain the true procyclin gene? The Northern blot data indicate that this is unlikely given that no message was detected with pPRO2001 as probe. Any true copy of the gene should encode an mRNA with sufficient similarity to procyclin to be detected on blots at the reduced stringency conditions used, given that the gene should be expressed at a relatively high level and the coding part of the sequence is the most likely part to be

conserved. In addition, all four *Eco* RI fragments hybridized to pPRO2001 at similar intensity. A fragment that contained homology over a broader part of the sequence than the repeats alone would be expected to produce a more intense signal than the pTcpro fragment in which the homology is approaching its T_M at the temperature and salt concentration used. Each fragment must contain at least some repeats as they had in common hybridization to both probes. The similar high intensity of the signals with the pTcpro probe indicate that they are more similar to each other than to pPRO2001. Hybridization to pPRO2001 would therefore appear to have been purely fortuitous due to the similar nature of the repeat sequences in *T.congolense* to those of procyclin.

One observation with respect to the repeats is that the sequence is absolutely conserved between one repeat and the next. Except for differences within the repeat unit itself, there is no equivalent of "third base wobble". It is likely that the repeats evolved from a single unit but with time they might have diverged through the process of mutation. Gene conversion events from other copies of the repeats could however remove these mutations and maintain the sequence, especially if it is under functional constraints. The repeats do degenerate at the edges where presumably the constraints are weaker. What might the repeats be? They seem to be specific to *T.congolense* as they are not present in *T.brucei* outside the partially homologous coding sequence in the procyclin genes. The only bands that hybridize to pTcpro in genomic Southern blots of *T.brucei* DNA are those which contain copies of the procyclin gene and they are only detected weakly at reduced stringency. The repeats must therefore have arisen after the divergence of the two species or else have been lost from *T.brucei*.

The results indicate that there is no real homologue of procyclin in *T.congolense*. However, the similarity in properties between procyclin and the *T.congolense* protein detected by the GUGM collection of monoclonal antibodies suggests that there may be a different protein in this species which plays an analogous role, though it is not related at the level of primary structure.

CHAPTER 4

IDENTIFICATION AND ANALYSIS OF PROCYCLIC cDNAS

4.1 Introduction

It was hoped that the GUGM mabs might identify the gene for the *T.congolense* surface antigen from a cDNA expression library. This would however depend on the cDNA sequence being inserted in the correct orientation and reading frame for expression and the mabs recognising epitopes which do not rely on correct folding of the protein or on secondary modification. An alternative approach was therefore also undertaken.

The procyclin gene was first identified, by both Roditi *et al* (1987) and Mowatt & Clayton (1987), by differential screening of cDNA libraries with first strand cDNA made from bloodstream and procyclic form trypanosomes. This gene was found to be procyclic-specific and was the only such gene to be isolated by this method. From immunoelectron microscopy studies, procyclin is very abundant on the cell surface (Richardson *et al*, 1988) and the *T.congolense* surface antigen also appears to be present at a high level (E.Kilbride & L. Tetley, unpublished). It seemed therefore reasonable to assume that the gene encoding this antigen might be expressed at a similarly high level and thus might also be found by differential cDNA screening.

A differential screen of this sort is likely to detect several messages that are expressed stage-specifically although they would have to be expressed at a relatively high level in order to be detected without performing subtractive hybridizations. A differential screen of procyclic cDNA might therefore have been expected to detect other stage-regulated genes which may provide some insight into the control mechanisms used by the parasite.

What sort of genes might be expressed specifically in the procyclic stage? A major difference between bloodstream and procyclic forms is their energy requirements. Bloodstream forms live in an environment which is rich in glucose and they can obtain all of their energy needs from glycolysis (Donelson & Rice-Ficht, 1985) without requiring conventional mitochondrial activity. Procyclics on the other hand do not live in such an energy-rich environment and require an active mitochondrion. The mitochondrion exists as a single organelle in the trypanosome, encodes several of the genes for its own biogenesis in the kinetoplast DNA (Simpson, 1987), and its respiratory machinery is not fully functional in bloodstream stages but is activated in procyclics. Some of the mitochondrially-encoded genes are expressed at all stages of the lifecycle (Feagin & Stuart, 1988;

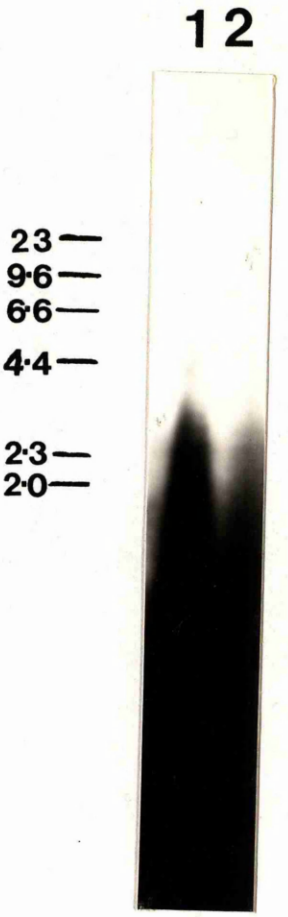


Fig. 4.1 Alkaline Agarose Gel Analysis of cDNA Synthesis Products.

First (lane 1) and second (lane 2) strand synthesis control reactions were run out on a 1.0% alkaline agarose gel which was dried at 60°C under vacuum before autoradiography at room temp. overnight. Markers were a *Hin*d III digest of λ DNA.

Stuart, 1991) and others appear to have their expression controlled posttranscriptionally by the process of RNA editing (eg. CYb, COII and ND7; Feagin *et al*, 1987; Feagin & Stuart, 1988; Koslowsky *et al*, 1990). It is however possible that some genes are controlled at the level of transcription and, in addition, there are the nuclear-encoded components of the mitochondrial machinery, some or all of which may be expressed in a stage-specific manner.

One problem with most stocks of *T.congolense* is that they do not grow well in rodents (Lumsden *et al*, 1973) and so obtaining sufficient material is very difficult. However, one line is available which undergoes complete development in the fly in 7 days and is also highly virulent in mice. This stock (M15 1/148; Young & Godfrey, 1983) was therefore obtained, along with procyclics transformed from these bloodstream forms. This allowed production of high parasitaemias in mice although there was still a problem in that the aetiology of *T.congolense* infections, whereby the majority of the parasites concentrate in the peripheral circulation, means that many of the parasites are not recovered by cardiac puncture (Rosen *et al*, 1979). The cause of this phenomenon is not clear (Hawking, 1975) but it has been proposed that this lack of recovery may be due to the parasites adhering to capillary walls. Banks (1978 and 1979) has observed that the parasites do adhere both *in vivo* to capillaries and *in vitro* to red blood cells. Two methods of partially overcoming this problem were utilised - warming the mice at 37°C before collecting the blood as suggested in Rosen *et al* (1979) and leaving the parasitaemia as late as possible before purification. The merits of each of these methods will be discussed later in this chapter. While this approach did not provide a large amount of material, sufficient RNA was obtained to make first strand bloodstream cDNA as a probe and for a few tracks on northern blots.

4.2 Construction of a Procyclic cDNA Library

5ug of *T.congolense* 1/148 poly(A)$^+$ RNA was used to construct a unidirectional cDNA library in lambda ZAP II using the Stratagene Uni-Zap cDNA Synthesis kit. Synthesis of the first and second strands of cDNA was monitored by adding radiolabelled [α-^{32}P]dCTP to the reaction mixes and analysing samples of the products on a 1% alkaline agarose gel which was then dried under vacuum at 60°C and exposed overnight for autoradiography (Fig. 4.1). The first strand synthesis produced material of a comparable size to the mRNA but the second

strand was shorter, although of good enough quality to proceed.

After ensuring that the cDNA had flush ends by the addition of the Klenow fragment of DNA Polymerase I, adding *Eco* RI adapters and digesting with *Xho* I, the cDNA was size-fractionated through a column of Sepharose CL-4B prepared in a long-bore 1ml plastic pipette and equilibrated in STE. Eighteen 2-drop fractions were collected and radioactivity measured in a scintillation counter. Instead of the expected 2 peaks of radioactivity, there appeared to be 3. The most plausible explanation for this is that the first peak represented full-length cDNA, the second peak short, partially-synthesised cDNA and the third peak unincorporated dNTPs. The first and second peaks were therefore pooled separately, leaving the tail of the second peak as it was likely to contain the *Xho* I-adapter fragment. Only the first peak was used in subsequent ligations, the second peak being retained for further ligations if that should have proved necessary.

Half of the first peak pool was ligated to 1ug of lambda uni-ZAP II arms according to the manufacturer's instructions and packaged in halves using the Gigapack Gold Packaging kit. The library was titrated on *E.coli* PLK-F' and found to contain 5.55 X 10^5 pfu. This was from half of the total cDNA and so a potential 1.1 X 10^6 pfu could be obtained which should contain sequences expressed even at a low level. The second aliquot was not immediately packaged but retained in the event of requiring an unamplified library at a later date, as neat packaging reactions are unstable.

The 5.55 X 10^5 pfu were amplified on *E.coli* PLK-F' to produce an amplified titre of 7.4 X 10^9 pfu.ml^{-1}.

4.3 Checking the Integrity of the Library

4.3.1 Insert Sizes

Six randomly chosen plaques were cored from the amplification titration plate and the pBluescript SK(-) phagemid containing the cDNA insert excised from each as indicated in the manufacturer's protocols. Small-scale plasmid DNA preparations were made from each, digested with *Eco* RI and *Xho* I to release the cDNA insert, and electrophoresed on a 0.8% agarose gel. The phagemid inserts varied in size from about 600bp to >4kb.

4.3.2 Sequence Representation

Approximately 400 pfu of the amplified library was plated out on *E.coli* XL1-blue and duplicate plaque lifts were probed with a random primed *Hind* III - *Eco* RI fragment from pTbαβT-1, covering the *T.brucei* beta-tubulin gene, at 55°C. After washing to a stringency of 2XSSC at 55°C, 2 plaques were detected indicating that beta-tubulin clones represented about 0.5% of the total library which is reasonably representative (alpha- and beta-tubulin represents about 10% of total cell protein and about 1% of mRNA; Thomashow *et al*, 1983; Seebeck *et al*, 1983).

4.3.3 Full-Length cDNAs

Plaque lifts were also probed with an end-labelled oligonucleotide containing the universal spliced leader (SL) sequence to give an idea of the number of full-length clones. Washing to 5XSSC at 37°C, this probe hybridized strongly with 10 plaques and also a few more weakly which either possessed only part of the SL or were due to spurious cross-reaction. Thus about 3% of the clone inserts appeared to be full-length. This is not particularly high but it is important to bear in mind that the SL is only a short sequence and many more clones may contain only a few base-pairs of it which would not hybridize to the probe, or may indeed terminate immediately 3' of the SL.

These results indicated that the library was of reasonable quality and ready for screening.

4.4 Screening of the Library With GUGM2.1

1000pfu of the library were plated out on each of 10 plates and induced to express β-galactosidase fusion proteins on IPTG-soaked nitrocellulose filters. The filters were then probed with the mab GUGM2.1 and binding detected by HRP-conjugated antimouse IgG using the 4-chloro-1-napthol-based developer.

No plaques reacted positively on any of the filters, suggesting that the GUGM2.1 epitope was not expressed although there was no good positive control available. This result however was not surprising as further studies with GUGM2.2 suggested that this mab may be specific to a carbohydrate epitope (E.Kilbride, 1992)

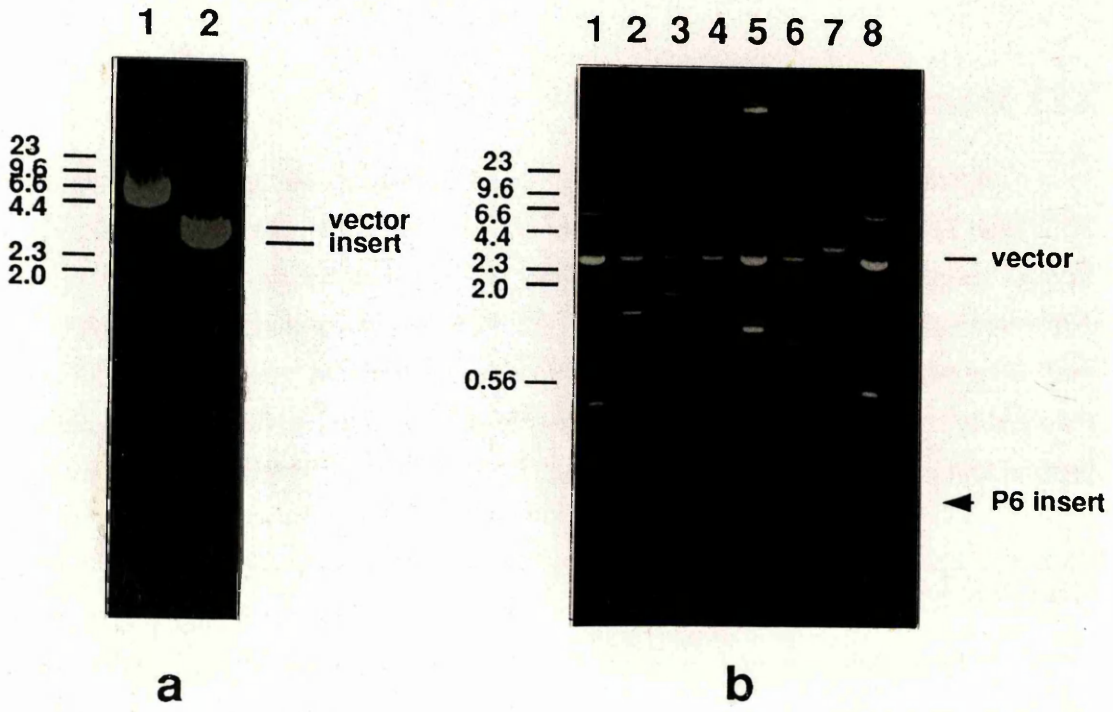


Fig. 4.3 Digestion of Procyclic-Specific cDNAs to Determine Insert Sizes.

Small-scale plasmid DNA preparations of the excised procyclic-specific cDNA clones were digested with *Eco* RI + *Xho* I and electrophoresed on 0.8% agarose gels to determine the size of the cDNA inserts. Positions of vector band and cDNAP6 insert are marked.

(a) Lane 1 - uncut cDNAP1, lane 2 - cDNAP1 cut with *Eco* RI and *Xho* I,

(b) Lane 1 - cDNAP2, lane 2 - cDNAP4, lane 3 - cDNAP5, lane 4 - cDNAP6, lane 5 - cDNAP7, lane 6 - cDNAP8, lane 7 - cDNAP10, lane 8 - cDNAPE.

and would therefore not detect the protein in a screen of this sort.

4.5 Differential Screening

Approximately 10^4 pfu of the library were plated out on *E.coli* XL1-Blue on a 24 X 24 cm LB plate and 4 duplicate lifts made onto Hybond N filters. One pair of filters was probed with 3.3 X 10^7 cpm of radiolabelled first strand cDNA prepared from total procyclic RNA and the other pair with 2.4 X 10^7 cpm of the same type of probe prepared from total bloodstream form RNA. It was necessary to derive these probes from total RNA as the number of bloodstream form cells obtained did not yield enough material for poly(A)$^+$ selection. As oligo(dT) was used to prime the cDNA synthesis however, only those RNAs possessing poly(A) tracts should have produced probe and the results indicated that this was indeed the case. Hybridization was for 24 hours in aqueous buffer at 65°C and the filters were washed to high stringency in 0.1XSSC at 65°C before autoradiography for 6 hours at -70°C with intensifying screens.

Many plaques hybridized with each probe. Some signals hybridized at different relative levels, but this was largely quantitative rather than qualitative. However, some plaques clearly reacted with only one of the probes and indeed there were more than 28 procyclic-specific signals in this primary screen. Some of these signals were more intense than others and the areas around 10 of them were plaque-purified for further analysis.

The areas around these primary procyclic-specific plaques were plated out at about 50-100 pfu per plate and duplicate plaque lifts on Hybond N were probed with the same hybridization solution used for the primary screen, to which had been added previously unused probe containing about 1/3 of the counts used initially. After this secondary screening, cDNAs P1,4,6,8 & E lit up strongly with the procyclic probe and not with the bloodstream probe while the signals for cDNAs P2,5,7 & 10 were much less intense but still appeared to be procyclic specific. A single plaque from each of these plates was purified and the pBluescript SK(-) phagemid containing the cDNA insert was excised from each.

Small-scale plasmid DNA preparations were made of each of the excised cDNAs and digested with *Eco* RI + *Xho* I to check for inserts (Fig. 4.3). All of the cDNAs except P10 produced insert fragments of different sizes with the double

```
                  MetProThrArgPheLysLysThrArgHisGlnArgGly
  1   ATTGCTTCGACATGCCGACTCGCTTCAAGAAGACACGCCATCAGCGTGGC    50

      SerThrPheCysGlyTyrGlyArgValGlyLysHisArgLysHisProSe
 51   TCCACGTTCTGTGGCTACGGCCGCGTCGGCAAACACCGCAAGCACCCTTC   100

      rGlyArgGlyAsnAlaGlyGlyGluHisHisHisArgIleAsnPheArgL
101   GGGCCGCGGTAATGCTGGCGGTGAGCACCATCACCGTATTAACTTCAGGA   150

      ysTyrHisProGlyTyrPheGlyLysCysGlyMetAsnHisTyrHisLys
151   AGTACCATCCCGGATACTTTGGGAAGTGCGGCATGAACCACTACCACAAG   200

      LysLysAsnAlaThrTrpLysProThrIleAsnLeuAspAsnLeuThrLy
201   AAGAAGAACGCGACGTGGAAGCCAACGATCAACCTGGACAACCTCACCAA   250

      sLeuIleThrArgAspGluAlaAlaMetAlaLysLysGlyGluValLeuP
251   GCTGATTACGAGGGACGAGGCCGCGATGGCAAAGAAGGGGGAGGTCCTTC   300

      roValIleAspLeuLeuAlaAsnGlyTyrAlaAsnValLeuGlyAsnGly
301   CTGTTATTGACCTGTTGGCCAATGGGTACGCGAACGTGCTTGGCAACGGC   350

      HisLeuGlnAlaProCysIleValLysAlaArgTrpValSerLysLeuAl
351   CATCTTCAGGCTCCTTGCATTGTGAAGGCTCGGTGGGTGAGTAAGCTCGC   400

      aAspLysLysIleArgLysAlaGlyGlyAlaValValLeuGlnAla *
401   TGATAAGAAGATCCGCAAGGCTGGTGGTGCGGTGGTGCTGCAGGCGTAGT   450

451   GCATGCGTGCGGTTTTCATAGCAATCATTTTTCATTTCTGTGCTTGTGTC   500

501   TGACTGAAAAAAAAAAAAAAAAAAA  526
```

Fig. 4.4 Sequence of cDNAPE.

cDNAPE was sequenced on both strands using the pBluescript polylinker primers on plasmid and single stranded templates. Additional sequence towards the 3'-end of each strand was obtained using internal primers (oligonucleotides 490 and 491) designed from sequence already obtained. (Spliced leader sequence at 5'-end of clone is underlined.)

digest and P10 produced a linear fragment larger than pBluescript alone, indicating that it too contained an insert. As this clone did not digest with *Eco* RI alone (data not shown), it appeared it had lost the *Eco* RI site at the 5'-end of the cDNA.

4.6 5' Sequence Analysis

As an initial screen to determine how many different sequences had been isolated from the differential screen, sequence analysis of the 5'-ends of the cDNA inserts was performed through the *Eco* RI site using the pBluescript T3 primer. cDNA P10 had a poly(A) tail at the 5'-end suggesting that the insert had been inserted the wrong way round and explaining why the *Eco* RI site had been lost. CDNAs P6 and P8 had exactly the same sequence at the 5'-end but, on a longer run of the gel, the sequences diverged further downstream, around about the point where cDNAP6 had a stretch of A residues which probably represented a poly(A) tail as the sequence then reentered the pBluescript polylinker. The cDNAP8 sequence continued further downstream before terminating with a poly(A) tail at a different site. cDNAP1 contained 29 copies of the sequence (CCCTAA) at its very 5'-end, which in one reading frame could encode a repeat of the dipeptide (ProAsn) and which is not dissimilar from the procyclin dipeptide repeat (ProGlu). It was therefore of some immediate interest although it is also exactly the sequence of *T.brucei* telomere repeats (Blackburn & Challoner, 1984). None of the other cDNAs had any sequence motifs of any obvious significance at this stage and none of the sequences, except cDNAPE, had any strong homologies to known sequences in the EMBL GenBank and NBRF databases, an observation perhaps not too surprising if the cDNAs were indeed procyclic-specific. cDNAPE (Figs. 4.4 and 4.7) had strong homology to the *Saccharomyces cerevisiae* (Kaufer *et al*, 1983), *Neurospora crassa* (Kreader & Heckman, 1987), mouse (Belhumeur *et al*, 1987), rat and *Tetrahymena* (Yao & Yao. 1991) ribosomal protein L29 (L27'). As this gene would be expected to be expressed at all stages of the lifecycle and therefore not to be isolated by a stage-specific screen, it was subjected to further analysis of stage-specificity.

4.7 Northern Analysis of the cDNA Clones

In order to determine which of the cDNAs identified by the differential screen represented true stage-specific transcripts, each was used to probe northern

cDNA	Intensity of Plaque Hyb^{n} (1)	Insert Size	Size of RNA Detected	Northern Probed [&Probing Order]	Stage-Specific?	Genomic Copy Number	Other Information
cDNAP1	++	2.4kb	7.5, 5.3, 3.5, 1.85, 1.23, 0.59kb	a[1] (Fig.4.20)	yes?(7) (A^+ & A^-)	multiple	telomeric
cDNAP2	+	450bp	1.2kb	a[2] (Fig. 4.5A)	no	1	
cDNAP4	+++	1.3kb	1.6kb	a[4] (Fig. 4.10a1)	+/-(3)	>5	ARP(6)
cDNAP5	+	3.4kb	nd(2)	b[2] (Fig. 4.5B)	not expressed	2	
cDNAP6	++	320bp	4.4kb	-	+/-(3)	1 or 2(5)	=cDNAP8
cDNAP7	+	1.1kb	1.3kb	a[3] (Fig. 4.10a2)	+/-(3)	1	
cDNAP8	++	890bp	4.4kb	b[1] (Fig. 4.9a)	+/-(3)	1 or 2(5)	=cDNAP6
cDNAP10	++	<500bp	-(4)	-	-(4)	-	
cDNAPE	++	530bp	0.9kb	b[3] (Fig. 4.6)	no	1	ribosomal protein L29

Table 4.1 Summary of Results from the cDNAs Isolated in the Differential Screen

(1) +++ - strong signal, ++ - quite strong signal, + - weak signal
(2) nd - none detected
(3) stage-specific in one northern, not others
(4) not analysed
(5) depends on stock
(6) see chapter 5
(7) only analysed with bloodstream RNA isolated from prewarmed mice

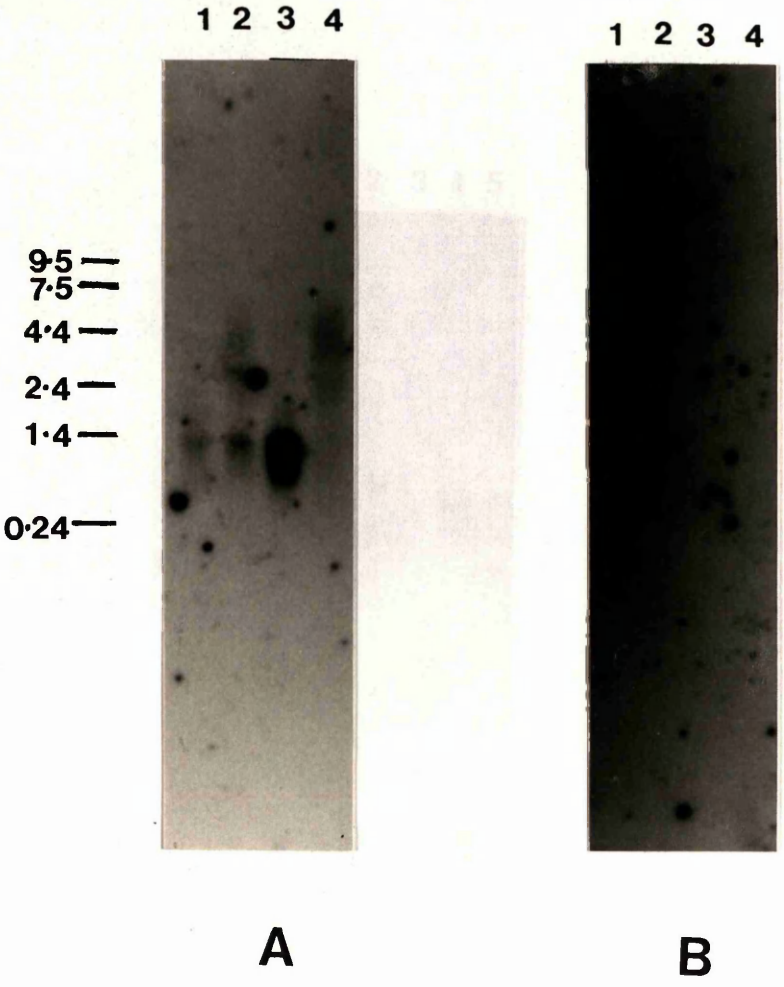


Fig. 4.5 Northern Blot Analysis of cDNAs P2 and P5.

20ug total bloodstream RNA (lane 1(A), lane 4(B)), 20ug total procyclic RNA (lane 2(A), lane 1(B)), 3ug poly(A)$^+$ procyclic RNA (lane 3(A), lane 2(B)) and 20ug poly(A)$^-$ procyclic RNA (lane 4(A), lane 3(B)) were electrophoresed on 1.0% formaldehyde gels and blotted onto Hybond N membrane. The blots was then probed with radiolabelled insert from cDNAP2 (A) and cDNAP5 (B) at 42°C in formamide buffer and washed to 0.1XSSC/0.1% SDS at 65°C before autoradiography for 5 days at -70°C with intensifying screens. Markers are the BRL RNA ladder. These were the 2nd probes on each blot (see Table 4.1).

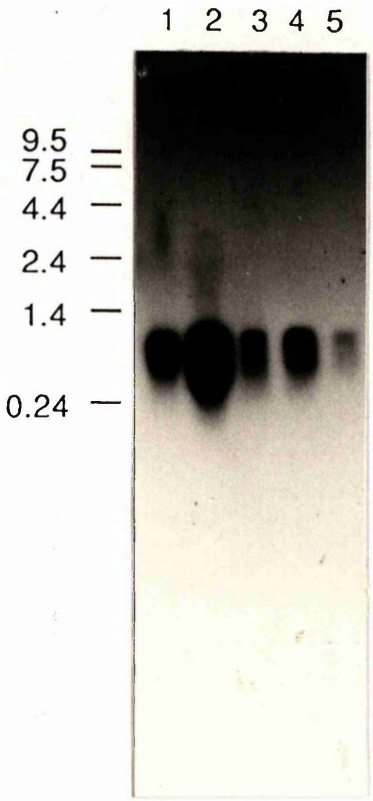


Fig. 4.6 Northern Blot Analysis of cDNAPE.

20ug *T.congolense* total (lane 1), 3ug poly(A)$^+$ (lane 2) and 20ug poly(A)$^-$ (lane 3) procyclic RNA and 20ug total bloodstream RNA (lane 4) and 20ug *T.brucei* total procyclic RNA (lane 5) were electrophoresed on a 1.0% formaldehyde gel and blotted onto Hybond N membrane. The blot was probed with radiolabelled insert from cDNAPE at 42°C in formamide buffer and washed in 0.1XSSC/0.1% SDS at 65°C before autoradiography overnight at -70°C with intensifying screens. Markers are the BRL RNA ladder. This was the 3rd probe on this blot (see Table 4.1).

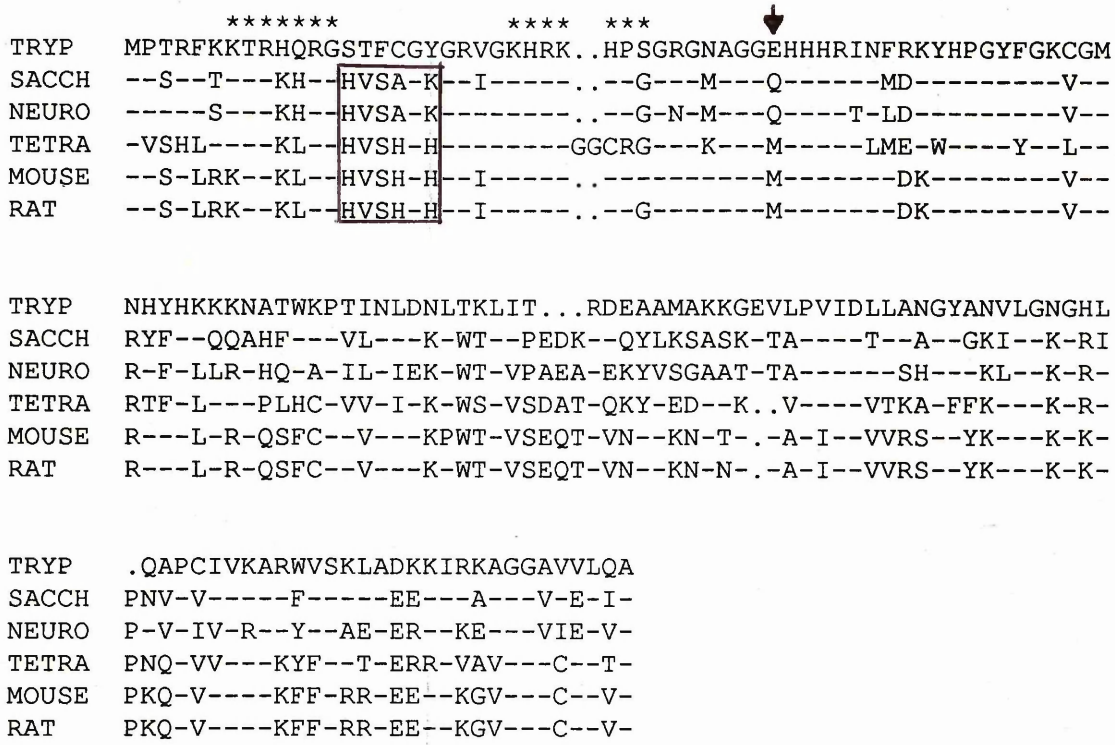


Fig. 4.7 Comparison of cDNAPE Sequence with the L29 Ribosomal Protein Gene from Other Species.

The predicted amino acid sequence from cDNAPE is shown aligned with the sequence of ribosomal protein L29 from other species in the database. TRYP - *Trypanosoma congolense*, SACCH - *Saccharomyces cerevisiae*, NEURO - *Neurospora crassa*, TETRA - *Tetrahymena*. The boxed region is the sequence conserved in other species which is not conserved in the *T.congolense* gene. Asterisks mark the nuclear localisation signals identified in yeast. An arrow marks the Glutamic acid at residue 38 which may indicate cycloheximide resistance (see discussion).

blots of bloodstream and procyclic RNA. Table 4.1 summarises the results. cDNA P2 (Fig. 4.5a) detected bands in both bloodstream and procyclic RNA and therefore does not encode a stage-specific transcript. Neither fragment of cDNAP5 could detect any signal in any RNA track even on prolonged exposure (Fig. 4.5b) suggesting that it may represent a transcript with a very low expression level. However, plasmid sequencing of the 3'-end of this clone did not reveal a poly(A) tail and the clone may not represent an mRNA at all. In either case it is not clear how the clone was selected from the differential screen but it was discarded.

4.7.1 CDNAPE encodes a *T.congolense* Ribosomal Protein Gene

Total, poly(A)$^+$ and poly(A)$^-$ procyclic stage RNA, total bloodstream stage RNA and total *T.brucei* procyclic stage RNA were run on a formaldehyde gel and blotted onto Hybond N. This blot was then hybridized with the insert fragment from cDNAPE at 42°C in 50% formamide and the filter washed in 0.1XSSC at 65°C (Fig. 4.6). The probe detected a transcript of about 550bp in all tracks, including bloodstream and *T.brucei* RNA, implying that it was not a stage-specific transcript but expressed in all stages. The size of the transcript detected is similar to that of L29 in other species (Kaufer *et al*, 1983; Kreader & Heckman, 1987; Belhumeur *et al*, 1987; Yao & Yao, 1991) and the sequence is highly homologous (55-58% identical and 70-75% similar at the amino acid level) to other L29 (L27') genes (Fig. 4.7). There is one short block of homology in other L29 proteins (HVSA/H at amino acid positions 14 to 17) that does not appear to be conserved in the *T.congolense* protein and the significance of this is not known. There are important structural differences between the ribosomes and rRNAs of trypanosomes and other eukaryotes (Cordingley & Turner, 1980) and this may have some relevance in this respect. The complete sequence appears to have been cloned as the first 4bp of the insert are the last 4bp of the SL sequence and the start of the open reading frame in other species is also present in the clone. The nuclear localization signals identified in the yeast L29 protein (Underwood & Fried, 1990) also appear to be largely conserved, at least at the essential arginine residues.

4.7.2 cDNAP1 Is Stage-Specific

A northern blot of total, poly(A)$^+$ and poly(A)$^-$ procyclic RNA plus total bloodstream RNA from *T.congolense* was probed with the insert from cDNAP1 and washed to high stringency in 0.1XSSC at 65°C. After 3 days exposure (Fig. 4.8a), 3

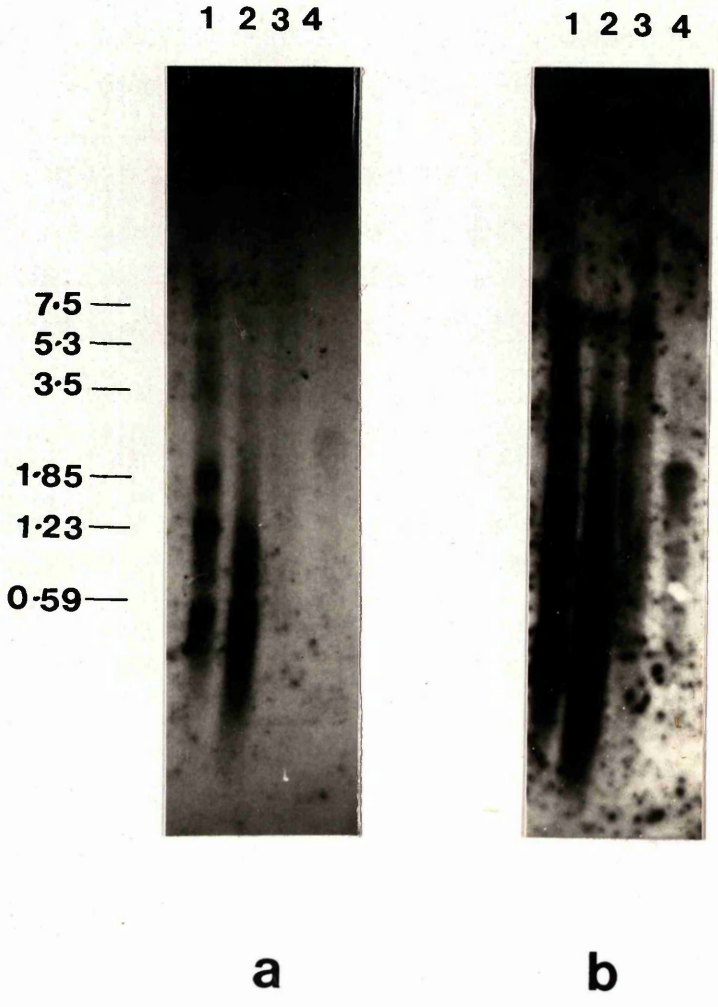


Fig. 4.8 Northern Blot Analysis of cDNAP1. 20ug total (lane 1), 3ug poly(A)$^+$ (lane 2) and 20ug poly(A)$^-$ (lane 3) procyclic RNA and 20ug total bloodstream RNA (lane 4) from *T.congolense* were electrophoresed on a 1.0% formaldehyde gel and blotted onto Hybond N membrane. The blot was probed with radiolabelled insert from cDNAP1 at 42°C in formamide buffer and washed in 0.1XSSC/0.1% SDS at 65°C before autoradiography for 3 days (a) and for 14 days (b) at -70°C with intensifying screens.

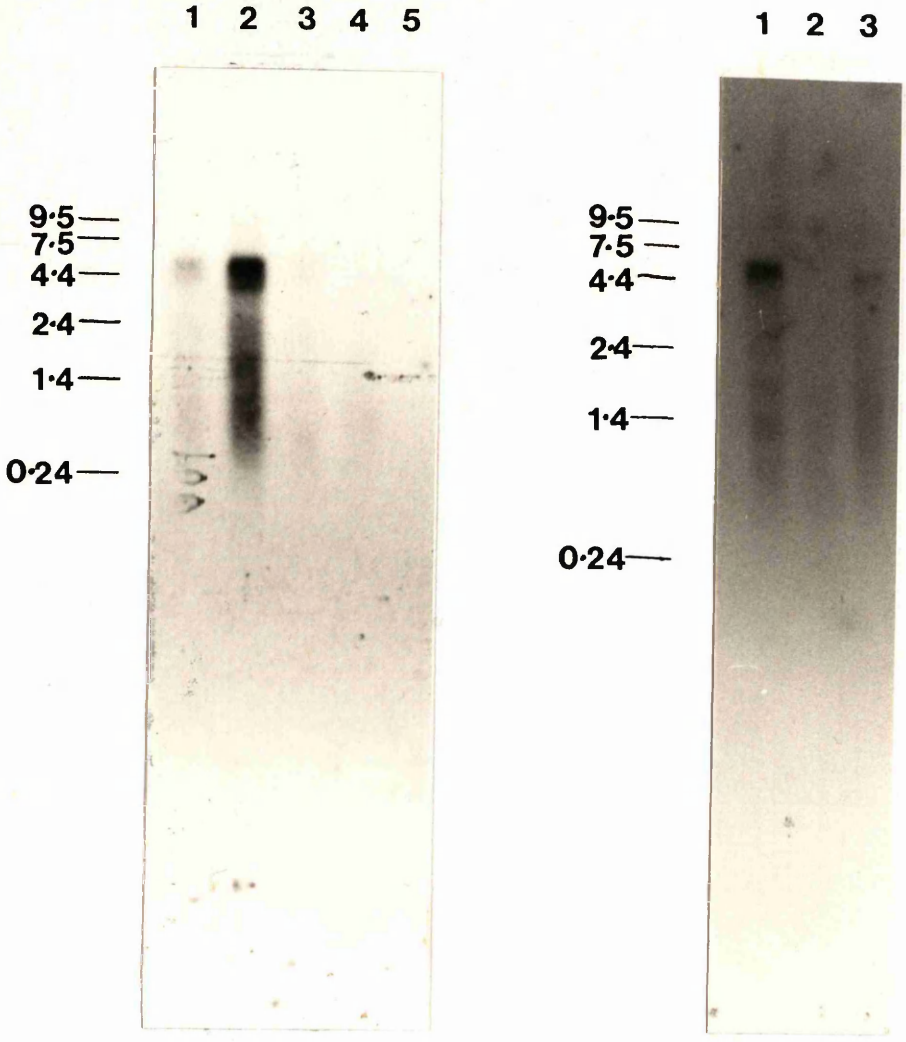
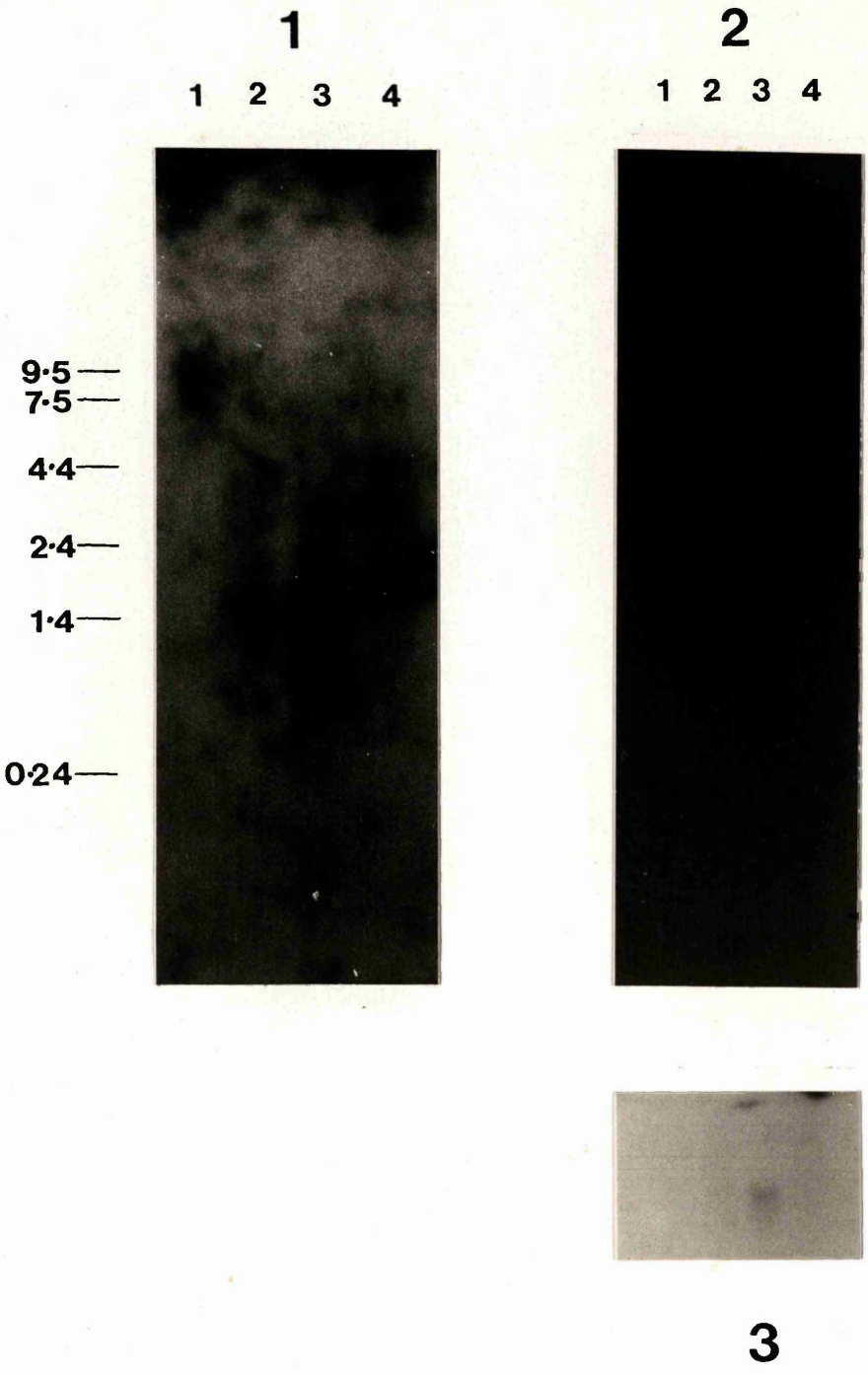


Fig. 4.9 Northern Blot Analysis of cDNAP8/6.

Procyclic and bloodstream *T.congolense* RNA and procyclic *T.brucei* RNA were electrophoresed on 1.0% formaldehyde gels and blotted onto Hybond N membranes. The blots were probed with radiolabelled cDNAP8 insert at 42°C in formamide buffer and washed in 0.1XSSC/0.1% SDS at 65°C before autoradiography for 5 days (a) or 6 days (b) at -70°C with intensifying screens.

(a) bloodstream RNA isolated after warming mice. (Lane 1 - 20ug total procyclic RNA, lane 2 - 3ug poly(A)$^+$ procyclic RNA, lane 3 - 20ug poly(A)$^-$ procyclic RNA, lane 4 - 20ug total bloodstream RNA, lane 5 - 20ug total *T.brucei* procyclic RNA.). This was the 1st probe on this blot (see Table 4.1).

(b) bloodstream RNA isolated late in infection. (Lane 1 - 20ug total bloodstream RNA, lane 2 - 20ug total procyclic RNA, lane 3 - 3ug poly(A)$^+$ procyclic RNA.



a

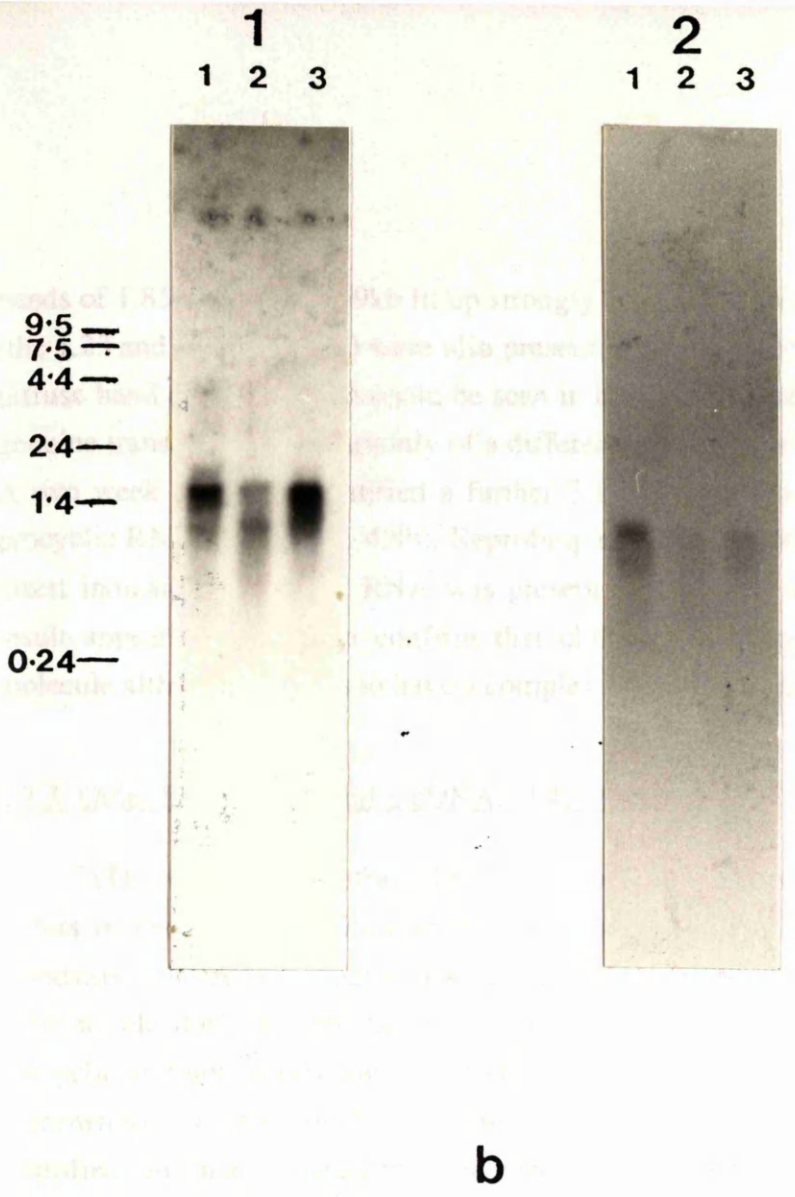


Fig. 4.10 Northern Blot Analysis of cDNAs P4 and P7.

RNA from procyclic and bloodstream stages of *T.congolense* was electrophoresed on 1.0% formaldehyde gels and blotted onto Hybond N membrane. The blots were hybridized with radiolabelled insert fragments from cDNAs P4 (1) and P7 (2) at 42°C in formamide buffer and washed in 0.1XSSC/0.1% SDS at 65°C before autoradiography overnight (P4) or for 5 days (P7a) or 6 days (P7b) at -70°C with intensifying screens. (a) Bloodstream RNA isolated after warming host. Lane 1 - 20ug total bloodstream RNA, lane 2 - 20ug total procyclic RNA, lane 3 - 3ug poly(A)$^+$ procyclic RNA, lane 4 - 20ug poly(A)$^-$ procyclic RNA. (b) Bloodstream RNA isolated late in infection. Lane 1 - 20ug total bloodstream RNA, lane 2 - 20ug total procyclic RNA, lane 3 - 3ug poly(A)$^+$ procyclic RNA. Blot (a), which had already been probed twice previously (see Table 4.1), was probed with P7 and then P4 in turn and then stripped and reprobed with the beta-tubulin probe (3), washing in 2XSSC/0.1% SDS at 55°C before autoradiography for 8 days at -70°C with intensifying screens, to control for the presence of RNA in all tracks and showed that by this stage much of the RNA had been stripped from the blot.

bands of 1.85, 1.23 and 0.59kb lit up strongly in total procyclic RNA, two of which (the 1.23 and 0.59kb bands) were also present in poly(A)$^+$ RNA and not poly(A)$^-$. A diffuse band of about 2.4kb could be seen in bloodstream form RNA but if it was a genuine transcript it was certainly of a different size to those in the procyclic tracks. A two week exposure identified a further 3 high molecular weight bands in all 3 procyclic RNA tracks (Fig. 4.8b). Reprobing of the blot with the *T.brucei* β-tubulin insert indicated that intact RNA was present in all tracks (data not shown). This result appears therefore to confirm that cDNAP1 is indeed a procyclic-specific molecule although it seems to have a complex expression pattern.

4.7.3 Other Procyclic-Specific cDNAs (P4, P6, P7 & P8)

cDNAP8 (and therefore cDNAP6 as well as it has the same 5' sequence, but differs in the position of its poly(A) tail) appeared to be procyclic specific when bloodstream RNA was isolated from infected mice which had been warmed at 37°C prior to bleeding. It detected a polyadenylated RNA of approximately 4kb in procyclic and not bloodstream RNA (Fig. 4.9a). The bloodstream RNA in this blot is known to be of good quality as subsequent reprobing with cDNAPE (see Fig. 4.6) identified an intact transcript. The size of the band detected by cDNAP8 is significantly longer than either of the cDNA clone inserts, with cDNAP6 only about 320bp and cDNAP8 less than 900bp in length. However, when this northern analysis was repeated with bloodstream RNA isolated late in infection without prewarming the animals, the transcript no longer appeared to be stage-specific (Fig. 4.9b).

cDNAs P4 and P7 also appeared to be procyclic specific in the first northern blots (Fig. 4.10a), encoding mRNAs of about 1.6kb and 1.3kb respectively. However in these cases the blot had been stripped and reprobed several times previously and it is not clear whether any intact bloodstream RNA remained when these probes were applied. Final screening of the blot with the β-tubulin probe (Fig. 4.10c) only detected a very weak signal in procyclic RNA (the cDNAPE probe was not yet available at that time). Plasmid sequencing of the 3'-ends of both of these cDNAs with the M13 -40 primer indicated that they did possess poly(A) tails. Again however, repetition of these northerns with RNA from unwarmed animals suggested that the transcripts were not stage-specific after all (Fig. 4.10b).

Since the two methods of increasing the yield of bloodstream form cells

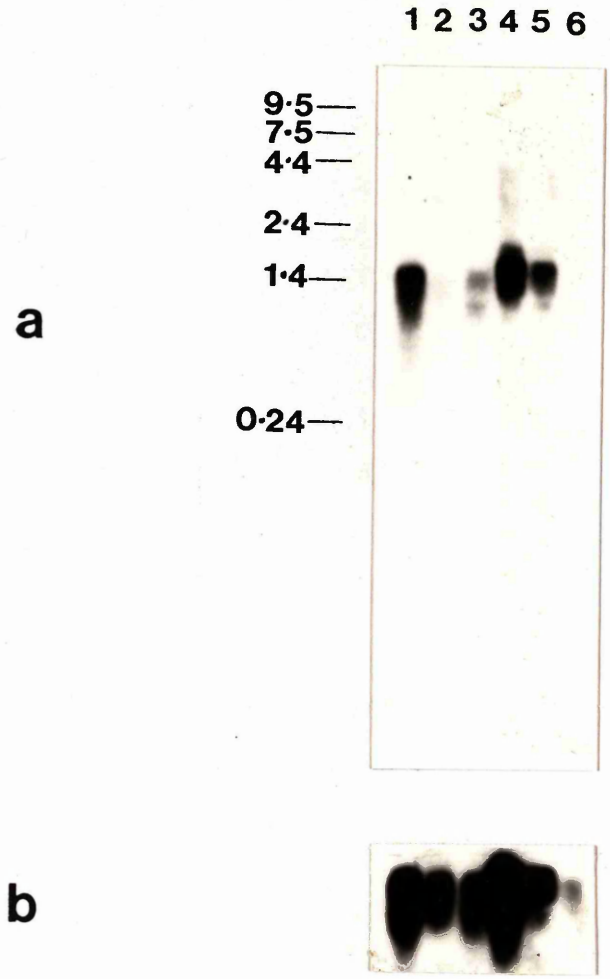


Fig. 4.11 Confirmation of Northern Blot Data.

Bloodstream form RNA was isolated from rodent-adapted *T.congolense* Ynat1.1 and electrophoresed alongside procyclic RNA from *T.congolense* M15 1/148 on a 1.0% formaldehyde gel then blotted onto Hybond N membrane. The blot was probed with radiolabelled insert from cDNAP4 (A) or cDNAPE as a control (B) in formamide buffer at 42°C and washed in 0.1XSSC/0.1% SDS at 65°C before autoradiography for 4 days (P4) or 2 days (PE) at -70°C with intensifying screens.

lane 1 - 3ug poly(A)$^+$ procyclic RNA, lane 2 - 20ug poly(A)$^-$ procyclic RNA, lane 3 - 20ug total procyclic RNA, lane 4 - 3ug poly(A)$^+$ bloodstream RNA, lane 5 - 20ug total bloodstream RNA, lane 6 - 20ug poly(A)$^-$ bloodstream RNA.

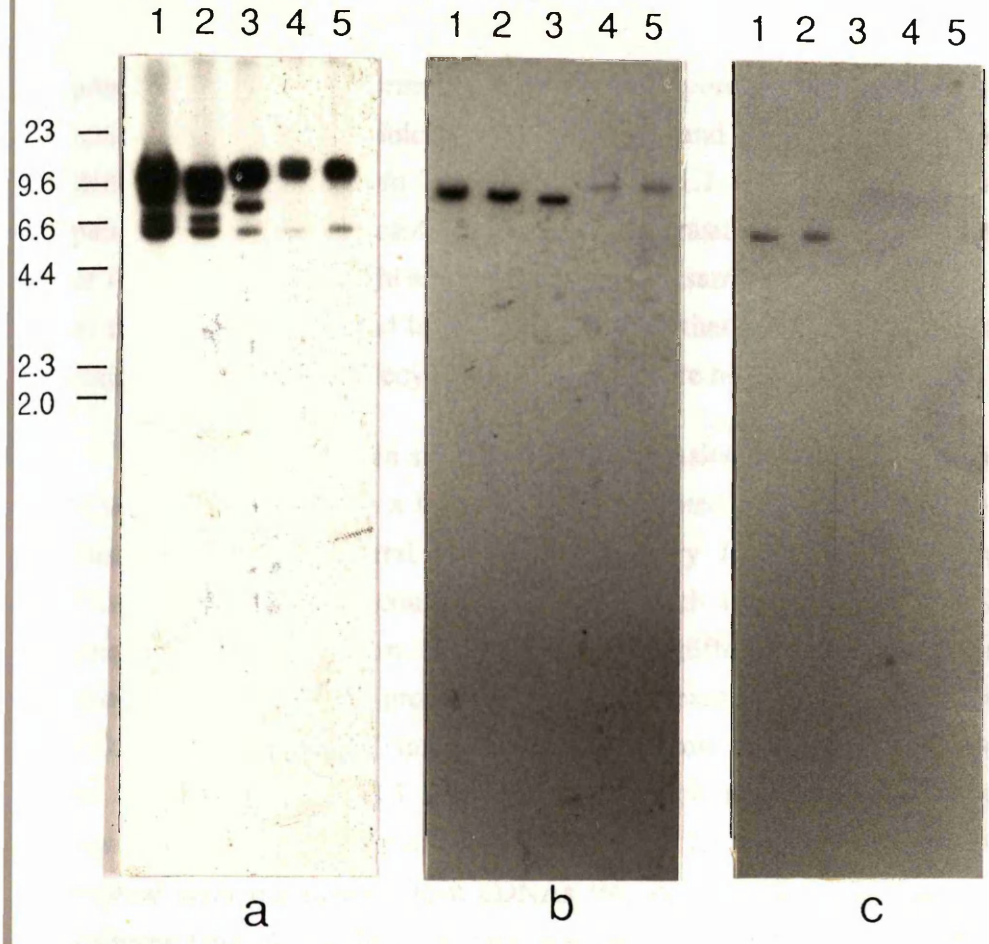


Fig. 4.12 Genomic Analysis of Several *T.congolense* Stocks.

Genomic DNA from *T.congolense* YNat1.1 (lane 1), M15 1/148 (lane 2), TREU 1627 (lane 3) and TREU 1457 (lane 4,5) was digested to completion with *Eco* RI and the products run out on a 0.8% agarose gel. The gel was blotted onto Hybond N membrane and hybridized with radiolabelled insert fragments of cDNAs P4 (a), P7 (b) and P8 (c) at 42°C in formamide buffer before washing in 0.1XSSC/0.1% SDS at 65°C and autoradiography. Size markers are a *Hin*d III digest of λ DNA.

produced conflicting northern results, a third approach was used which should have less effect on the physiology of the parasite and thus yield a more correct result. Bloodstream forms from *T.congolense* YNat1.1 were grown in rats over several parasitaemias and collected before the next parasitaemia had peaked. Northern blots of RNA purified from these parasites gave the same result with cDNAP4 (Fig. 4.11) as the parasites collected late in infection, i.e. that the cDNA was expressed in both stages of the parasite lifecycle and was therefore not stage-specific.

In order to obtain sufficiently high parasitaemias of these bloodstream forms it was necessary to use a line which was adapted to rodents such that the parasites remained in the general bloodstream. Barry (unpublished) had observed that *T.congolense* YNat1.1 could be adapted in such a way and so this line was used to prepare the bloodstream RNA. The use of different trypanosome stocks for the comparison of RNA in procyclic and bloodstream forms is not ideal but Rosen *et al* (1981) indicated that YNat 1.1 was derived from the Lister 1/148 stock which may be the same line as M15 1/148. Genomic zooblots of *Eco* RI-digested DNA from several *T.congolense* stocks including YNat1.1 and M15 1/148 were therefore probed with the inserts from cDNAs P4, P7 and P8 to ascertain their restriction patterns (Fig. 4.12). In each case, the banding patterns of both YNat1.1 and M15 1/148 were the same suggesting that they are indeed the same line of trypanosomes and that therefore the northern data are reliable.

4.8 cDNAP1

As cDNAP1 appeared to be stage-specific and had the potential to encode a protein with a dipeptide repeat not dissimilar from that found in procyclin, it was analysed further.

The 5' sequencing had indicated that cDNAP1 did not possess a SL sequence at its 5'-end and, since the potential open reading frame did not have an AUG start codon, it was decided to try and isolate a full-length version of the cDNA. Another 10^4 pfu of the amplified cDNA library were plated out and plaque lifts probed with the cDNAP1 insert. After washing to high stringency, 19 positives were identified which hybridized very strongly with the probe. This represented about 0.2% of the library, suggesting that cDNAP1 is highly expressed, in agreement with the northern data. The library was then reprobed with an oligonucleotide containing the SL sequence to identify clones which were full-length. Many plaques reacted with this

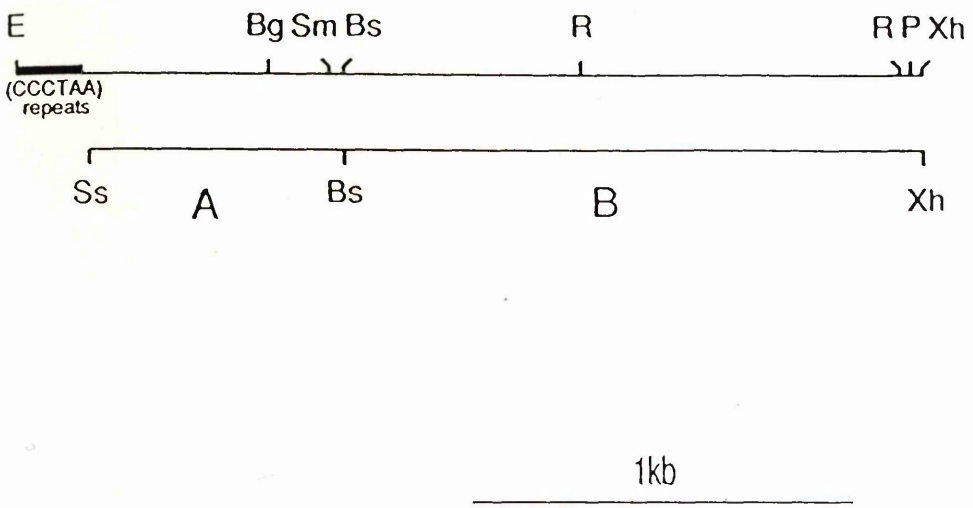


Fig. 4.13 cDNAP1 Restriction Map.

Black boxed region indicates CCCTAA repeats. Lower line indicates map of 1st timepoint sequencing deletion (T_1 deletion) used to prepare probes A and B used in the *Bal*-31 experiment (Fig. 4.19). E (*Eco* RI), Bg (*Bgl* II), Sm (*Sma* I), Ss (*Ssp* I), Bs (*Bst* XI), R (*Rsa* I), P (*Pst* I), Xh (*Xho* I).

```
                AsnProAsnProAsnProAsnProAsnProAsnProAsnProAsnPro
    1   CTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCT   50

                AsnProAsnProAsnProAsnProAsnProAsnProAsnProAsnProAs
   51   AACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAA  100

                nProAsnProAsnProAsnProAsnProAsnProAsnProAsnProAsnP
  101   CCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACC  150

                roAsnProAsnProAsnProAsnProAsnProValGlyArgMetGlyArg
  151   CTAACCCTAACCCTAACCCTAACCCTAACCCGGTTGGAAGGATGGGGAGG  200

                AspGluGlyTrpGlyHisArgLeuSerMetArgTrpLysAlaGlyProAr
  201   GATGAAGGGTGGGGGCATAGGCTGTCAATGAGGTGGAAGGCGGGCCCAAG  250

                gSerGluArgAlaAspLysArgValIleSerValTrpGluGlyAspProG
  251   GAGTGAACGGGCGGATAAGAGGGTTATTTCAGTTTGGGAAGGGGACCCGG  300

                luCysArgArgGlyGluArgIleValGluLeuValSerPheSerGluGlu
  301   AGTGTCGGAGGGGTGAAAGGATCGTGGAGCTGGTTTCCTTTTCAGAGGAA  350

                TrpGluMetGlyTrpLeuSerValValTyrGlyGlyGlyValTrpValTr
  351   TGGGAAATGGGATGGCTCTCAGTCGTGTATGGGGGTGGTGTTTGGGTTTG  400

                pSerValAlaGluLeuTrpLysSerLeuGlyGlyLeuPheArgIleAlaH
  401   GAGTGTGGCTGAGTTATGGAAGAGTCTTGGAGGTCTCTTTCGGATCGCCC  450

                isPheValPheGlyTrpTrpPheTrpSerAlaGlnAspLeuValGluVal
  451   ATTTCGTCTTTGGATGGTGGTTTTGGTCGGCTCAAGACCTGGTTGAAGTG  500

                ValSerValSerSerSerSerAsnGlyProLeuThrThrLeuTrpLeuSe
  501   GTCTCGGTTTCTTCCTCCTCCAATGGTCCACTGACCACTCTATGGCTTTC  550

                rThrThrGlyGluIleLeuCysSerProPhePheIleCysArgSerSerA
  551   CACGACGGGGGAGATTCTCTGCTCGCCTTTCTTCATTTGCCGTAGCTCAC  600

                rgIleArgSerLeuAlaLeuArgGlySerProArgTyrLeuLysAspEnd
  601   GGATACGCTCTCTTGCCCTGCGTGGGTCGCCCAGATACTTAAAGGACTGA  650

  651   CTACTCAATCTGCGTTTTGCCGGGTTCCGAGAATAGAGGTCTAATTGGGT  700

  701   AAGGTGCGCCCAATATTTGTTCCAGATCTTTTGTCTAAAGATGGCTTGAG  750

  751   AGGTGCCGCCAATGCCACCCGGCCACAGCCTTCACCAGGACCACAGGAGT  800

  801   TTCATCCCTGAAGGTTTTCATGGCCGTCAGTCGGGCCCCTGCGAGTCTCC  850

  851   TCGGAACAACTCATTTTGCCCCGGGAGAATGAAGTTGGTCTCGCTTTGAT  900

  901   CGTGCAGAGATAGAGAAGAGCGCCTCGACTGGATTTCCCTAAGCGATGCG  950

  951   AATCGAAAAGTTTGTAAGAGAGCTCCCAAATTTCGCGAGTTCCACGCTGA 1000

 1001   TGATAATCGCTTGCCAATTCTGTGAGGCGTCGCCGAAAGGCCCACCATTC 1050
                *********
 1051   TGGCCAAAAAAAAAATGAGGGACGACAGAAGCCTCCTTGCCGCCCTTCCAG 1100

 1101   CTTTCTTTTCCCACAGCTCCAGCAGCGCACAGGAAGGCACCAGAGAACGT 1150

 1151   CGTGCTGATTTTTATTGTGATTTTCTTCGGCCTCTTTCGATTTCCTAATT 1200

 1201   GCCATCTCAACCGCGTTCTCCTTCCTTCACATTTTCGTCGCCAAAGTGTC 1250

 1251   CCCACAACCTCTCACCCTCACCAATAAGAATGGACGAAATTAAATTTCTC 1300

 1301   AAGGTGTCTTTCCTCAGGTTTGAATTTTGTATTTTTTTTATGACGTTGTC 1350
```

```
1351  TTATGTGTTTTGTAAGAAATTATTATTCAGTTATAGATGCGTTTTGAGGG  1400
1401  GCTTTGCTGAGCGGGATAGGGAAGAGAATGACGTGATCTGCCAGACTCAC  1450
1451  ATTGCTTTTAAATACGATCATTCCAGCGTCTCTTCCTTTTTCGATCTCAT  1500
1501  GAGTCATTCGCGCCGCATAGGTACGGCTCTCGAACTCTCCTTCGACTTCC  1550
1551  CTTCCCACCTCGCATCATTCTCCATCTTTATTTACACCACGCAGGTTGCT  1600
1601  CGTCGGCTGCGTGGGAGGGCCTTGTGCACTCTGAATTGCTGCCTCTGCTG  1650
1651  GCGCCTCGTTCCGCGTTGATTCCTCAGTTCCTATGACGAAATATTCCACA  1700
1701  GTCTCCCCTCTTTGACCATCACATTTATTGGAATCCTCCAACTGTTTTTA  1750
1751  TGCAATGCTGACCATAGTGCCGTCTCTCGTTTCTTCATTACCACAGTGAC  1800
1801  CGGTTGCGAGGGCTGATTTGCCTGCGTGATTTTATTCGACACCTCGATAA  1850
1851  ACTCCCGCTCTCATTCCTTCTTTTCTGACATTGTATTCTTTTTGTTCTAT  1900
1901  AACTTCATGGAGAGGAGAAAAAGGTGAGTAAATCAATTCACAGTGATATC  1950
1951  ATATTTTTTTCGTGCACACAAAAATTTCCTGATATATTTTTTGTGGTATT  2000
2001  ATTTTTGATAATCACAAATAAATAAAGAGGAGGTTGGTAAAGAGTTTTTT  2050
2051  CCCCATCAAAAAGATTGGAAATATTCGTGTCGTTGAGGTTGTCCTTCCCA  2100
2101  GCCCCACTTTCGCGATCGTACGCGACTCACGCTTTGCTTTTGAAAAGGAG  2150
2151  AAGGTCTGCTCGATTTATGAGAATATTTTCCTCGACATTTTTCGCTTCCC  2200
2201  GCCTCTTGAGTGTAATAGCTTACAAATTGAGGTGTGCCAGGAGATTTCTT  2250
2251  TATGCTTTTTTGAGAGTCAGACGCTGTAATCACTCGAGATGCTTTCGTCT  2300
2301  TCCCCTCTGTATGGTCTGGACTTGGAGTTTCGTCTCTACTCCCATCATAC  2350
2351  TCTTTAACCTCACATCGATTCGTACTTTTCCCACTGCAGTCCGAAAATCT  2400
2401  TTTCGATCCTCATTACCTC  2419
```

Fig. 4.14 cDNAP1 Sequence.

cDNAP1 was sequenced on one strand by producing a set of nested deletions with Exonuclease III from the *Eco* RI end and using sequencing primers present in the pBluescript polylinker. The 3'-end was sequenced on the other strand from a subclone in pBluescript SK(+) and from plasmid sequencing of the original SK(-) clone. Putative reading frame containing (ProAsn) repeats is shown and the stretch of 9 "A" residues is marked with asterisks.

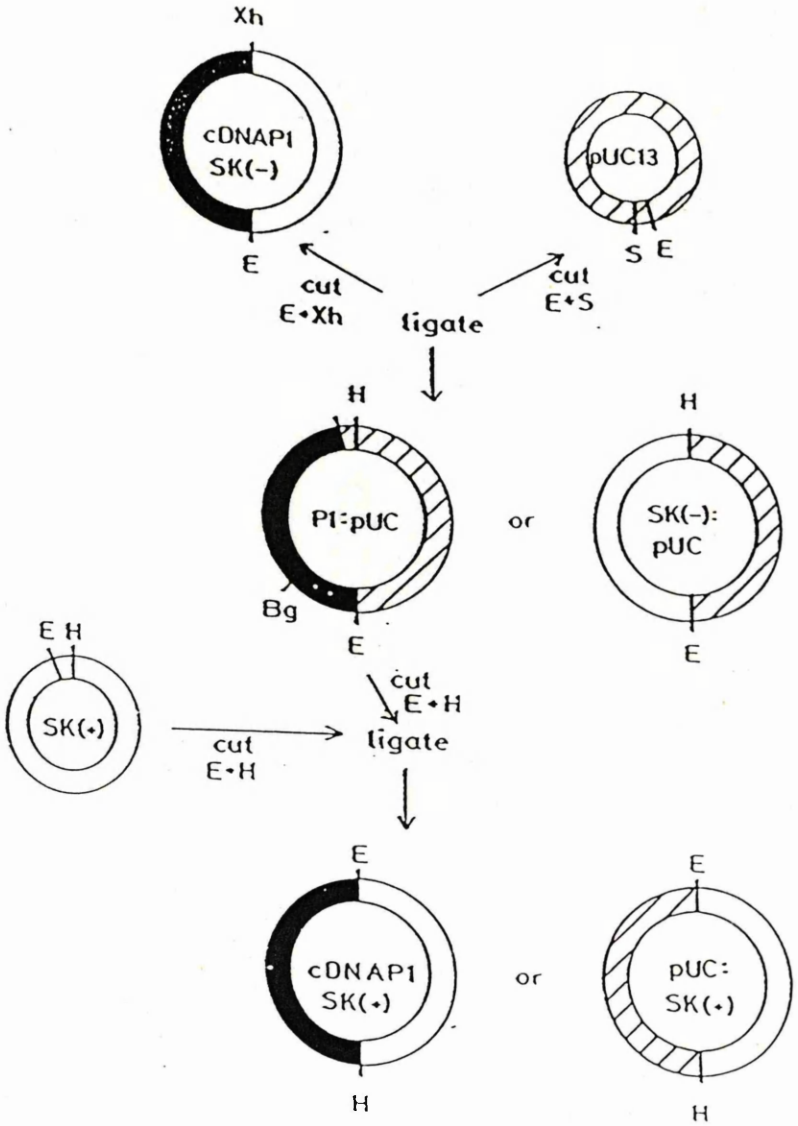


Fig. 4.15 Strategy for Cloning cDNAP1 in the Reverse Orientation.

cDNAP1 was cut with *Eco* RI and *Xho* I and ligated to pUC13 cut with *Eco* RI and *Sal* I. Resulting white colonies containing the cDNAP1 insert in pUC13 were digested with *Eco* RI and *Hind* III and ligated to pBluescript SK(+) cut with the same enzymes. Transformation yielded white colonies of the cDNAP1 insert in pBluescript SK(+) as well as a ligation of pUC13 to pBluescript. E (*Eco* RI), Xh (*Xho* I), S (*Sal* I), Bg (*Bgl* II), H (*Hind* III).

probe but none matched the plaques hybridizing to cDNAP1.

4.8.1 Mapping and Sequencing of cDNAP1

A restriction map of cDNAP1 was derived by digesting the plasmid with restriction endonucleases which, to simplify analysis, cut only once or twice in the vector. By performing single and double digests, an unambiguous map was obtained (Fig. 4.13) which had remarkably few restriction sites for a stretch of DNA 2.4kb long.

The sequence of the entire cDNA insert (Fig. 4.14) was obtained on one strand by creating a set of nested deletions from the 5'-end using Exonuclease III and Mung Bean Nuclease. The plasmid was digested to completion with *Xba* I and the 5'-overhang filled-in with thio dNTP derivatives to protect the end from attack by Exo III as no suitable enzyme site creating a 3'-overhang was available. This material was then digested to completion with *Eco* RI to create a 5'-overhang as an entry point for Exo III digestion. A set of twelve nested deletions was created, each approximately 200bp apart, and sequenced by producing single-stranded templates and using the T3 primer in the pBluescript polylinker. Unfortunately, the largest deletion (T_{12}) went right through the insert into the polylinker and so no sequence of the 3'-end was available from this strand. In order to obtain the 3' sequence it was therefore necessary to subclone the insert in the reverse orientation. Several attempts were made at ligating the excised insert directly into pBluescript SK(+) but each time the insert appeared to have gone back into residual SK(-) that was present. A more complicated cloning strategy was therefore adopted (Fig. 4.15).

The vector pUC13 is a chloramphenicol resistant plasmid with several unique cloning sites within its polylinker (Messing, 1983). An *Eco* RI site is available but there is no *Xho* I site. However, *Sal* I ends are compatible with *Xho* I ends although the combined site cannot be recut with either enzyme. A *Hin*d III site upstream of the *Sal* I site allows the insert to be re-excised. The *Eco* RI/*Xho* I insert from cDNAP1 was therefore ligated to *Eco* RI/*Sal* I cut pUC13, transformed into XL1-Blue cells, and chloramphenicol resistant colonies containing the desired construct were identified by the presence of a 0.7kb *Bgl* II/*Eco* RI fragment in digests of small-scale plasmid DNA preparations (Fig. 4.16a). The insert was then re-excised as a *Hin*d III/*Eco* RI fragment and the digest ligated to *Hin*d III/*Eco* RI cut pBluescript SK(+). Two types of ampicillin resistant transformant were obtained

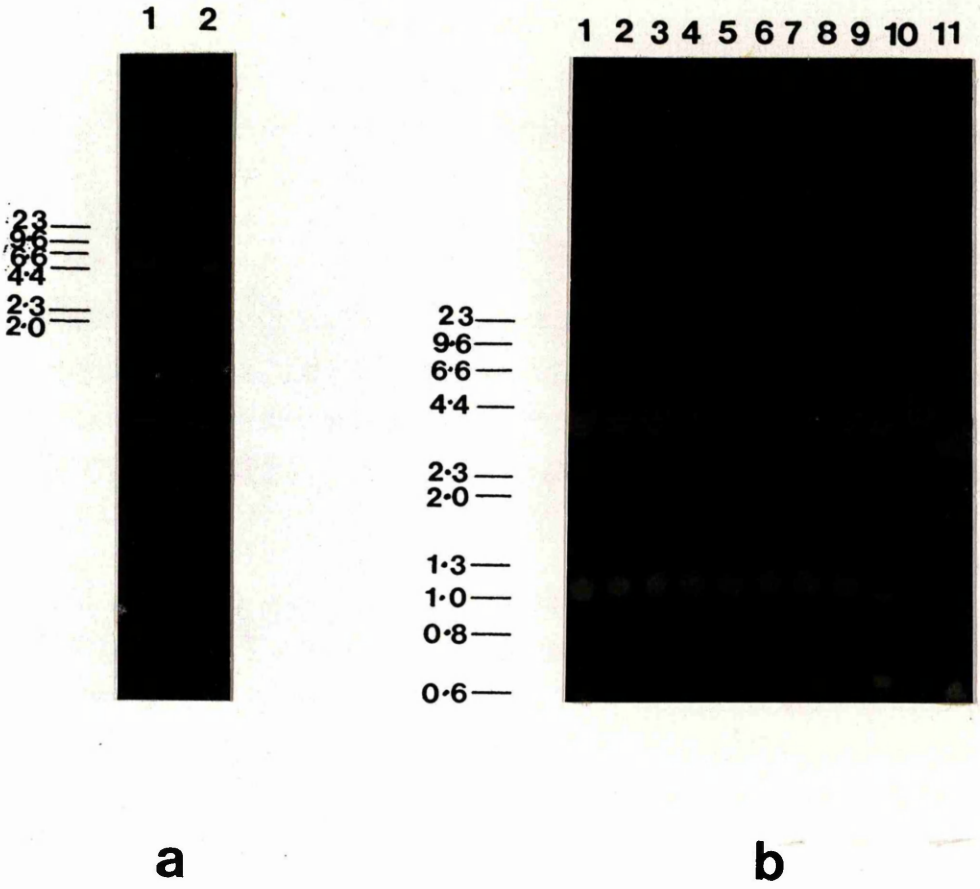


Fig. 4.16 Analysis of cDNAP1 Reverse Cloning Products.

(a) Identification of cDNAP1 clones in pUC13. Small-scale plasmid DNA preparations were digested with *Eco* RI + *Bgl* II and the products analysed on a 1.0% agarose gel stained with EtBr. (Lanes 1 & 2 - clones 1 & 2)

(b) Identification of cDNAP1 clones in pBluescript SK(+). Small-scale plasmid DNA preparations were digested with *Ssp* I and the products analyzed on a 1.0% agarose gel stained with EtBr. (Lanes 1-8 - clones 1-8, lane 9 - cDNAP1 SK(-), lane 10 - pBluescript SK(+), lane 11 - pBluescript SK(-).) Markers were λ DNA cut with *Hind* III and ΦX174 DNA cut with *Hae* III.

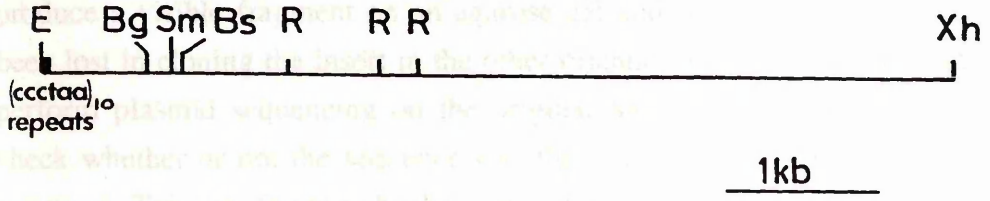


Fig.4.17 Map of cDNAP1b.

cDNAP1b, a homologue of cDNAP1 from the cDNA library, which had an insert significantly longer than cDNAP1 itself, was restriction mapped. (Compare with the map for cDNAP1 (Fig. 4.13).) E (*Eco* RI), Bg (*Bgl* II), Sm (*Sma* I), Bs (*Bst* XI), R (*Rsa* I), Xh (*Xho* I).

at a high level in the cloning ligation: blue colonies containing a ligation of pUC13:pBluescript which creates two complete copies of the β-galactosidase gene and white colonies containing the cDNAP1 insert in pBluescript SK(+). That these white colonies did indeed contain the desired construct was tested by performing an *Ssp* I digest which can differentiate between SK(+) and SK(-) (Fig. 4.16b). Template DNA preparations were then prepared from this reverse construct and sequenced using the T7 primer. The cDNA sequence did start at an *Xho* I site but there was no evidence of a poly(A) tail. It was possible that the poly(A) tail had been lost if there was a second *Xho* I site just 5' of the 3'-end which would not produce a visible fragment on an agarose gel and the small fragment would have been lost in cloning the insert in the other orientation. It was therefore necessary to perform plasmid sequencing on the original SK(-) clone using the T7 primer to check whether or not the sequence was the same as that obtained from the SK(+) construct. This was found to be the case and so this cDNA does not appear to have a poly(A) tail.

4.8.2 cDNAP1 Homologues

Despite none of the cDNAs homologous to cDNAP1 having a SL sequence at the 5'-end, one clone, cDNAP1b, had an insert significantly larger than cDNAP1 itself. This clone was therefore mapped with enzymes known to cut cDNAP1 and surprisingly produced a 5' map essentially identical to that of cDNAP1 while the 3' map was different (Fig. 4.17), having a 3.5kb extension. There was no *Pst* I site in this clone indicating that it may originate from a different gene copy. Sequencing of the 5'-end identified 10 copies of the (CCCTAA) repeat and a downstream sequence very similar to that of cDNAP1 with a few base substitutions (data not shown). Plasmid sequencing of the 3'-end of several of the homologues indicated that none possessed a poly(A) tail.

4.8.3 The Genomic Organisation of cDNAP1

In order to analyse the genomic organization of the cDNAP1 sequence, a Southern blot of *Eco* RI and *Hind* III digests of DNA from different stocks of *T.congolense* and from a control *T.b.rhodesiense* stock was probed with the 1.7kb *Bgl* II/*Xho* I 3' fragment of cDNAP1, with hybridization and washing being performed at 65°C to a stringency of 0.1XSSC (Fig. 4.18). This probe was used because it did not contain the (CCCTAA) repeats which were expected to hybridize

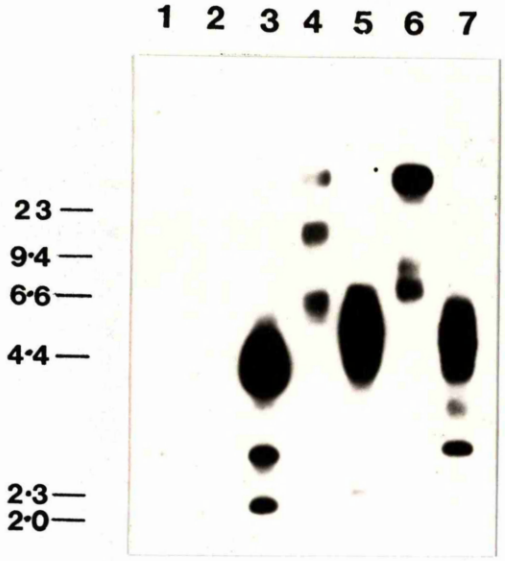
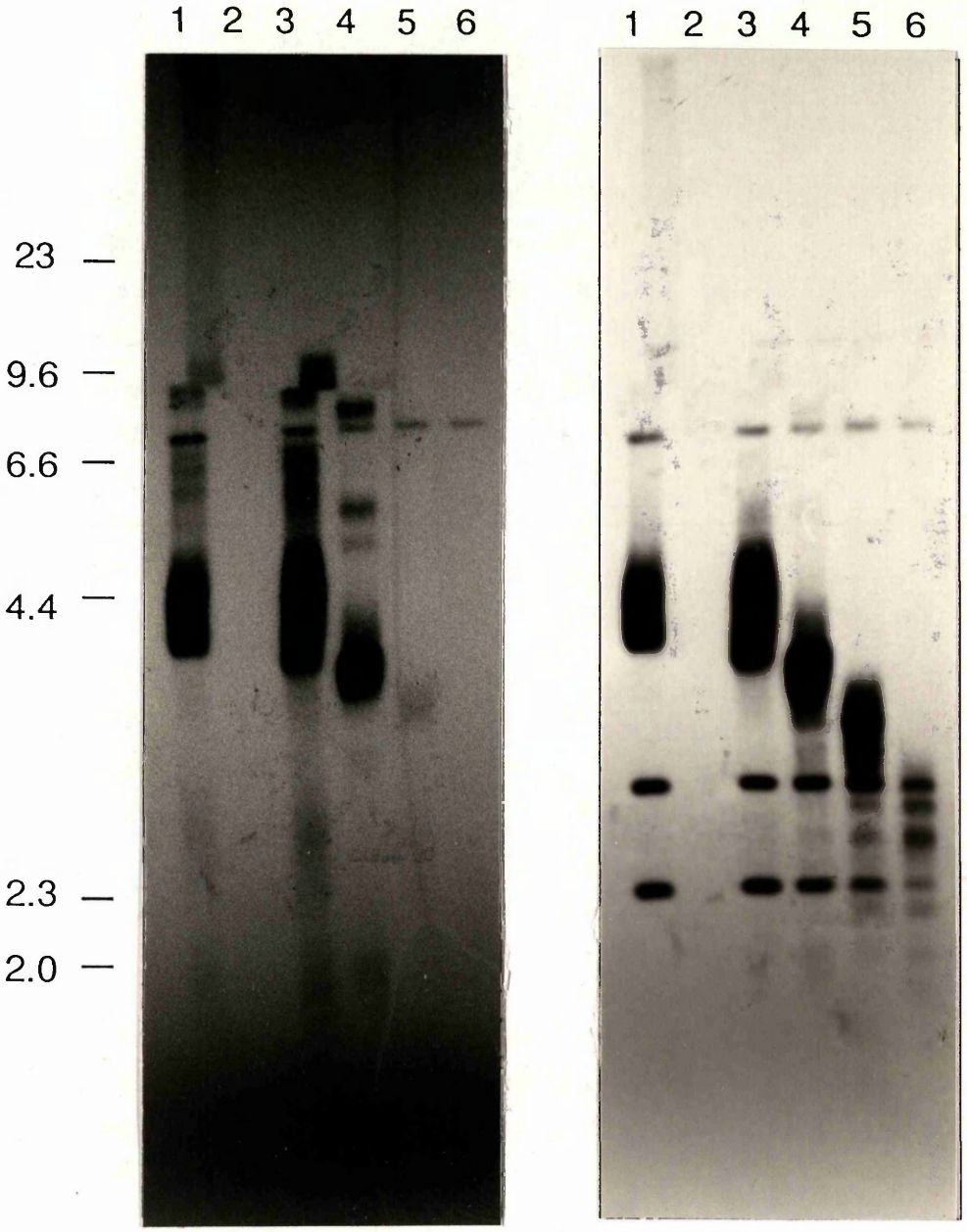


Fig. 4.18 Southern Blot Analysis of cDNAP1.

Genomic DNA from *T.b.rhodesiense* (lanes 1,2), *T.congolense* "1/148" (lanes 3,4), TREU 1457 (lanes 5,6) and TREU 1627 (lane 7) was digested with *Eco* RI (lanes 1,3,5,7) and *Hin*d III (lanes 2,4,6), electrophoresed on a 0.8% agarose gel and blotted onto Hybond N membrane. The blot was probed with the radiolabelled *Bgl* II-*Xho* I fragment of cDNAP1 (see Fig. 4.13) in aqueous buffer at 65°C and washed in 0.1XSSC/0.1% SDS at 65°C before autoradiography. Markers - *Hin*d III digest of λ DNA.

to telomeric sequences which share the same motif (Blackburn & Challoner, 1984) and might therefore confuse the results. While no signal was detectable in *T.b.rhodesiense* DNA, even on longer exposure, several signals were detected in both digests of *T.congolense* DNA from all 3 stocks, indicating that there were several copies of this sequence in the genome. The hybridization signals were not however tight bands and, since (CCCTAA) is the telomeric repeat unit in trypanosomes, this suggested that the sequence may reside at telomeres which vary in size with growth of the parasite (Bernards *et al*, 1983; Pays *et al*, 1983b; Van der Ploeg *et al*, 1984c) and therefore do not produce homogeneously sized fragments. A telomeric location could also explain the absence of any third base wobble in the repeat region as other constraints would be operative. In order to determine whether this sequence does indeed reside at telomeres, treatment of the DNA with the exonuclease *Bal*-31 was performed. As telomeres represent the ends of DNA molecules they are sensitive to exonucleases and therefore are the first signals to disappear when intact DNA is treated with a processive exonuclease such as *Bal*-31. Expression-linked copies of VSG genes were originally shown to reside at telomeres by an experiment of this sort (Laurent *et al*, 1983).

18ug of genomic DNA were digested with *Bal*-31 for increasing times and then *Eco* RI digests of the resulting products were fractionated on a 0.8% agarose gel and blotted onto Hybond N. It was noted that the DNA from the 5 min. *Bal*-31 digest had been lost at some stage during the subsequent procedures. The blot was probed with the *Bst* XI/*Ssp* I fragment of the first cDNAP1 sequencing deletion (T_1) (Fig. 4.13, probe A) which contains 440bp lying between the polylinker, immediately upstream of the 5'-end of the deletion, and a string of A residues (see Fig. 4.14). It was believed these may represent a poly(A) tail if the clone contained a double insert. After washing to 0.1XSSC at 65°C and autoradiography, it was observed that the hybridizing smear started to reduce in size after 10 mins of *Bal*-31 digestion and eventually disappeared (Fig.4.19a). This was indicative of the sequence being telomeric although one band of around 7kb was not degraded over the timecourse of the experiment and therefore may not be located at a telomere. This band gave a much weaker signal than the smear, indicating that it might be single-copy or else only partially related. The *Bal*-31 experiment was controlled by reprobing the blot with pTcTIM9, a cDNA clone of the *T.brucei* Triose Phosphate Isomerase gene, which is not telomeric and, as expected, the hybridizing fragment did not alter in size over the timecourse used (Fig. 4.19c).


1 2 3 4 5 6
23
9.6
6.6
4.4
2.3
2.0
a
1 2 3 4 5 6
b

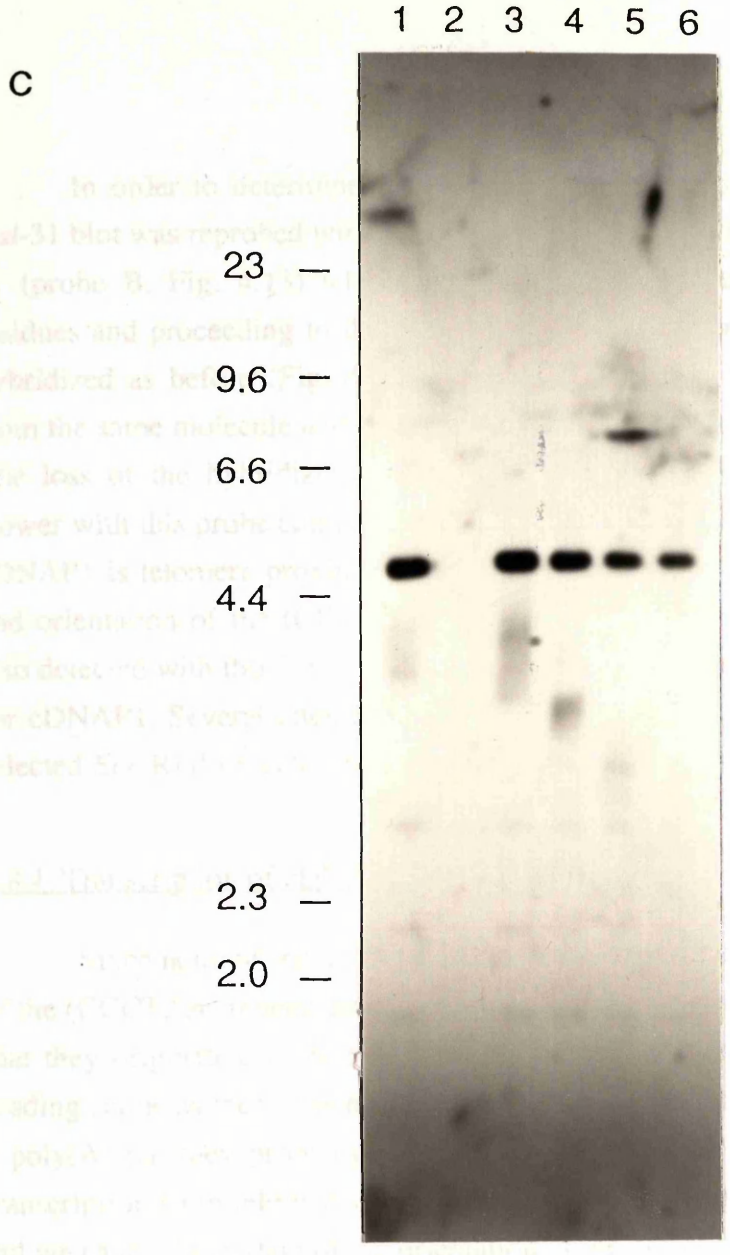


Fig. 4.19 Analysis of the *Bal*-31 Sensitivity of cDNAP1 Sequences.

T.congolense genomic DNA was digested for increasing times with the exonuclease *Bal*-31. After phenol/chloroform extraction and ethanol precipitation, the DNA was digested to completion with *Eco* RI and the products electrophoresed on a 0.8% agarose gel. The gel was blotted onto Hybond N membrane and probed with the radiolabelled *Bst* XI-*Ssp* I fragment of cDNAP1 (probe A in Fig. 4.13)(a), *Bst* XI-*Xho* I fragment of cDNAP1 (probe B in Fig. 4.13) (b) or pTcTIM9 (a *T.brucei* Triose Phosphate Isomerase gene cDNA clone) (c) at 65°C. The filters were washed in 0.1XSSC/0.1% SDS (2XSSC/0.1% SDS for pTcTIM9) at 65°C for 1 hour before autoradiography.

Markers - *Hin*d III digest of λ DNA. Lane 1 - no *Bal*-31 digestion; lanes 2-6 - 5, 10, 20, 40 & 60 minute digestions respectively with *Bal*-31.

In order to determine whether cDNAP1 does contain a double insert, the *Bal*-31 blot was reprobed with the *Bst* XI/*Xho* I fragment of cDNAP1 deletion clone T_1 (probe B, Fig. 4.13) which represents the 1.5kb starting at the string of A residues and proceeding to the 3'-end of the clone. Essentially the same fragments hybridized as before (Fig. 4.19b), indicating that both cDNAP1 probes originate from the same molecule and that the stretch of A's does not represent a poly(A) tail. The loss of the hybridizing signal with extent of *Bal*-31 digestion was slightly slower with this probe compared to the upstream probe, indicating that the 5'-end of cDNAP1 is telomere proximal and the 3'-end distal, as expected from the position and orientation of the (CCCTAA) repeats. The same 7kb non-telomeric band was also detected with this 3' probe suggesting that it may represent a real copy of a gene for cDNAP1. Several attempts were made to clone this fragment by ligating size-selected *Eco* RI digests to pBluescript but they were not successful.

4.8.4 Transcription of cDNAP1

Since none of the cDNAP1-like sequences appeared to have any sequence 5' of the (CCCTAA) repeats and appeared to be derived from telomeres, this suggested that they originated in the telomeric repeats and therefore that there was no open reading frame as there was no AUG start codon. With none of the clones possessing a poly(A) tail they probably do not represent true mRNAs but rather non-specific transcription from telomeres (Rudenko & Van der Ploeg, 1989). Without a poly(A) tail we cannot be certain of the orientation of the cDNA in pBluescript and therefore of the direction of transcription.

The question now arose as to whether cDNAP1 sequences are transcribed at all. It was thought possible that DNA may have contaminated the RNA preparation from which the library was made and, as this was the same RNA used in the northern blot (Fig. 4.8), this might also explain why there appeared to be transcripts. A new northern blot was prepared which had both newly prepared RNA and the original RNA used to construct the library. In addition, the original RNA was treated separately with RNase-free DNase 1 and DNase-free RNase A to assess whether either of these treatments affected the hybridization signal. The blot was probed with the 5' *Bst* X1/*Ssp* 1 fragment of the cDNAP1 T_1 deletion and washed to high stringency (Fig. 4.20). The DNase-treated track indicated that there was some DNA contamination of the original total RNA but the signal reduction was only in high molecular weight material and not in the region of the hybridizing bands. The

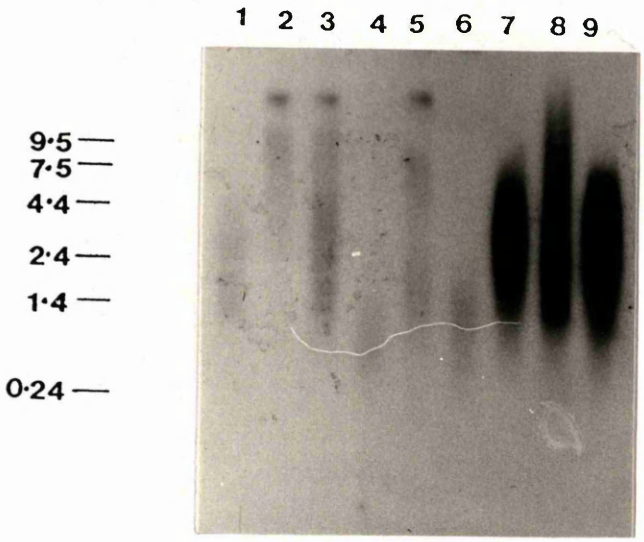


Fig. 4.20 Northern Blot Analysis to Check RNA Origin of cDNAP1.

The original *T.congolense* procyclic RNA preparation used to construct the cDNA library and a freshly isolated RNA preparation were electrophoresed on a 1.0% formaldehyde gel and blotted onto Hybond N membrane. The blot was hybridized with the radiolabelled *Bst* XI-*Ssp* I fragment of cDNAP1 at 42°C in formamide buffer and washed in 0.1XSSC/0.1% SDS at 65°C before autoradiography for 9 days at -70°C with intensifying screens.

Lane 1 - 20ug original RNA treated with 5u RNase-free DNase I for 30 mins. at 37°C, lane 2 - 20ug original RNA treated with 10ug.ml^{-1} DNase-free RNase for 15 mins. at 37°C, lane 3 - 20ug original total procyclic RNA, lane 4 - 3ug original poly(A)$^+$ procyclic RNA, lane 5 - 20ug original poly(A)$^-$ procyclic RNA, lane 6 - 20ug total bloodstream RNA, lane 7 - 20ug new total procyclic RNA, lane 8 - 3ug new poly(A)$^+$ procyclic RNA, lane 9 - 20ug new poly(A)$^-$ procyclic RNA. Markers are the BRL RNA ladder. This was the first probe on this blot (see Table 4.1)

RNase-treated track on the other hand showed only background signal suggesting that the majority of the hybridization was indeed due to RNA. Unfortunately the original mRNA preparation appeared to be degraded but in the fresh RNA tracks there was obvious signal in total RNA, poly(A)$^+$ and poly(A)$^-$, with higher molecular weight material only in the total and poly(A)$^-$ tracks. The bloodstream RNA track only produced a weak, low molecular weight signal with no clear bands. It is unlikely that this was due to degradation as subsequent reprobing of this blot with cDNAP2 indicated that this track did contain intact transcripts (Fig. 4.5a). If the transcripts of cDNAP1 are due to non-specific transcription from telomeres it is surprising that such transcription should be stage specific.

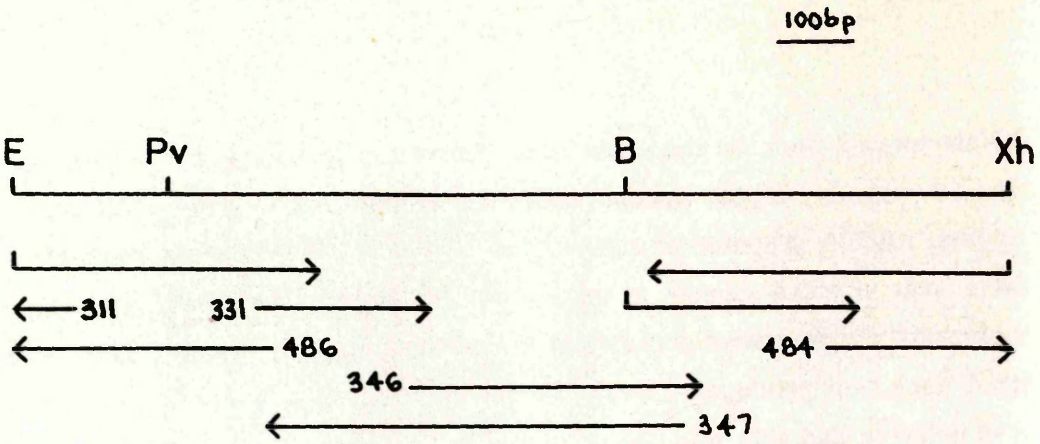
4.9 Sequence Analysis of cDNAs P4, P7 and P8/6

The other cDNAs which initially appeared to represent procyclic specific transcripts were sequenced in their entirety on both strands, to enable identification of significant open reading frames. This was achieved by a combination of the following methods. By cloning the insert in the opposite orientation into pBluescript KS(-) each strand of the cDNA could be sequenced using the pBluescript polylinker primers or specific oligonucleotide primers designed for the purpose of sequencing that region, using sequence information already obtained. Further sequence was obtained by subcloning the cDNAs into the appropriate pBluescript vector and performing plasmid or single-stranded sequencing with the polylinker primers. Regions of sequence obtained by each method are indicated in Figs. 4.21a, 4.22a and 4.23a.

4.9.1 cDNAP4

cDNAP4 (Fig. 4.21b) has a potential open reading frame of at least 256 amino acids. Whether the first AUG codon in this reading frame is the initiation codon was not clear at this stage, in the absence of upstream sequence, as the cDNA is not full-length. However, both the AUGs at nucleotides 41 and 53 possess the consensus sequence for initiation codons in flagellate protozoa (ANNAUGNC; Yamauchi, 1991) and initiation at either would produce a protein with no obvious signal peptide. The amino terminus of this predicted protein is reasonably hydrophobic but a signal peptide cleavage point cannot be predicted according to the method of Von Heijne (1988). However, later results (see chapter 5) which indicate that this protein is expressed on the surface of the trypanosome suggest that a signal

a



b

```
                                                    *
                                                MetThrThrT
  1 CTCGCACTCCTACTCCAAGCCAGCAAGAAGCGTGAATACCATGACGACAA  50
      *
    hrMetSerArgValLeuHisLeuMetThrValThrLeuLeuCysAlaArg
 51 CCATGTCCCGTGTCCTACACCTAATGACGGTCACACTGCTGTGCGCACGC 100

    ValGlyMetGlyGlnAlaSerAspAspAspAspCysGlyGlyGlnSerIl
101 GTGGGAATGGGCCAGGCTAGCGATGACGACGATTGTGGCGGGCAGAGCAT 150

    eProGlnLysValGluGluValGlnThrMetCysAspValAlaArgGlnL
151 TCCCCAAAAGGTGGAGGAGGTGCAAACGATGTGCGACGTTGCGCGGCAGC 200

    euArgAlaLeuGluThrAlaSerGlnSerAlaValAlaAlaValValSer
201 TGAGGGCCCTGGAGACCGCTTCCCAGTCCGCGGTGGCTGCCGTGGTTTCT 250

    SerAlaArgGluAlaSerGluAlaLysGluArgAlaGluLysAlaValGl
251 TCTGCCCGGGAGGCGTCGGAGGCGAAAGAGCGTGCGGAGAAAGCTGTGGA 300

    uArgAlaLysSerLysLysArgGlyValAspAlaAlaThrGluAlaAlaA
301 GCGCGCCAAATCGAAGAAGCGTGGTGTGGACGCGGCGACGGAAGCGGCTG 350

    laArgAlaAlaAlaAlaAlaGlnArgAlaGluThrValValSerAspAla
351 CAAGGGCTGCGGCTGCGGCCCAGCGCGCGGAGACGGTGGTGAGCGATGCG 400

    ArgLysHisAlaAlaAspLeuThrAlaAlaSerLysAspAlaIleGluTh
401 AGGAAGCACGCGGCAGACCTGACGGCGGCGTCGAAGGATGCTATCGAGAC 450

    rThrAspGluSerLeuArgLeuLeuAlaThrCysGluLysAlaAspGluP
451 GACCGACGAGTCGCTGCGCCTACTGGCCACATGCGAGAAAGCGGACGAGC 500

    roIleArgThrAlaAlaLysLysCysThrGlyAlaAlaAlaGluValThr
501 CCATCCGCACTGCTGCAAAAAAGTGCACGGGTGCCGCCGCCGAAGTCACG 550

    SerLysSerLeuGluSerAlaPheAspAlaLeuAlaGluLeuLeuProAs
551 TCCAAGTCCCTTGAGTCAGCGTTCGACGCTCTCGCGGAACTGCTACCGGA 600

    pGlyAlaAspAspIleArgGluHisGlyAlaValPheValLysGlyLeuL
601 TGGTGCGGACGACATCCGCGAGCACGGTGCCGTGTTCGTGAAGGGGCTGA 650

    ysSerLeuGluAspAspValArgThrAlaGlyGluAlaLysSerGluAla
651 AGTCTCTGGAGGATGACGTGCGCACGGCTGGAGAGGCAAAGAGCGAGGCG 700
```

```
      GluLysAlaGluGlyAspAlaAsnAspAlaAlaAspGlyAlaArgAlaVa
 701  GAGAAGGCTGAGGGCGATGCGAACGACGCGGCAGATGGTGCCCGTGCCGT   750

      lLeuThrGlyValCysValLeuLeuLeuLeuAlaAlaLeuHisPheSerA
 751  GCTGACGGGCGTGTGCGTGCTGCTGCTTCTGGCTGCACTGCACTTTTCTG   800

      laGlyLeu
 801  CGGGGCTGTGAGGCGGGATCCCCCAGCTCACCTGCCCGCGAGACATTATG   850

 851  CGTATTTTGAGTGACGAACTTTTATTCGCCCAACCTGGCATAATACCTTA   900

 901  TACAGTGAGCGTTATTTTTATTTCAACTTTCCCCTTCTTTTGTTATACTT   950

 951  TTTTATTCTTTGGCACGGTCATCATTCCTCTCCGTGTGACATTTCCTCTT  1000

1001  TTTTACATTGTGTGTTATTATGATTGTGTCACTGTTATCGCATGTGGTGT  1050

1051  ATTTATTTACTTCGCCTCTCCGGTCAGTACTGGGTTTCCGCTTTGGGATG  1100

1101  CGCCACGATTTCCCGCTGTACGAACCTTCCTTCCCACTGCTTTTACCCCC  1150

1151  TGCCACGCCGTAACAGGTGCGCACCGCTGTAGATTTCTGTGCACCTTTGT  1200

1201  TTCGGATCACGCTACTGGCGCCGTGGCTGATGGACATAAATAATCGGTTG  1250

1251  TGTTACTTTATTTTCTCTTGTTCATTTTGTTGGAAAAAAAAAAAAAAAAA  1300

1301  AAAAAAAAAAA  1311
```

c

```
PARP Aβ   635  CGCAGCCCTGTAGACTTC        652
               :::::::::::::: :::
cDNAP4   1170  CGCAGCCCTGTAGATTTCTGTG  1191
               : ::::::::::::::::::
PARP Aα   668  CACAGCCCTGTAGATTTCTGTG   689
```

Fig. 4.21 Map and Sequence of cDNAP4.

(a) Restriction map and sequencing strategy. E (*Eco* RI), Pv (*Pvu* II), B (*Bam* HI), Xh (*Xho* I). Numbers indicate oligonucleotide primer used.

(b) Nucleotide and predicted amino acid sequence. Asterisks mark possible start codons. Predicted cleavage point for GPI tail addition is marked with an arrow. Block of homology to 3' UTR of *T.brucei* procyclin message is boxed.

(c) Homology to procyclin 3' UTR

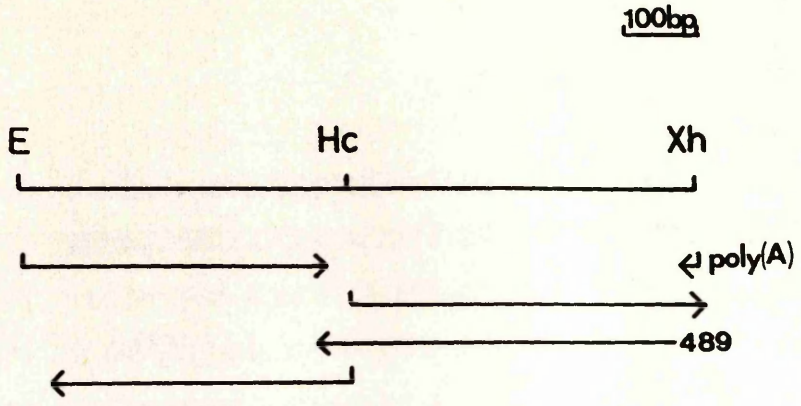
peptide might be present. The putative protein is rich in alanine residues (about 21.5%), contains no *N*-linked glycosylation signals, but has several serine residues which may indicate a potential for *O*-linked glycosylation although individual *O*-linked sites cannot be predicted (Wilson *et al*, 1991), and has a highly hydrophobic C-terminus which may be the signal for the addition of a glycosyl phosphatidyl inositol tail (Ferguson & Williams, 1988), as is found on several trypanosome surface molecules including procyclin and the VSG (Clayton & Mowatt, 1989; Ferguson *et al*, 1985). The predicted cleavage point for addition of such a tail is indicated. The protein, after processing at the carboxyl end but no other modification, would have a pI of 4.82 and would be 21% alanine.

cDNAP4 has a long potential 3' untranslated region (3' UTR) of 491bp and, since the initial northern result suggested this cDNA might be procyclic specific, it was interesting to compare the sequence of this region to that from procyclin. It has been shown (Jefferies *et al*, 1991) that sequences in the 3' untranslated region of the procyclin gene have a significant effect on the expression of Chloramphenicol Acetyl Transferase (CAT) when the CAT gene is cloned downstream of the procyclin and VSG promoters in expression vectors and introduced into procyclic cells of *T.brucei* by electroporation. Only 3 short regions of the procyclin 3' UTR are conserved between different copies of the gene in *T.brucei* (Mowatt & Clayton, 1988) and it is believed that one or more of these sequences may be important for message stability. A computer comparison of the procyclin and cDNAP4 3' UTRs using the programme BESTFIT identified a region of 22bp which had a 21/22 match for one procyclin cDNA and 17/18 match for another copy (Fig. 4.21c). This sequence contains the central 16bp conserved motif in procyclin 3' UTRs (Koenig *et al*, 1989) although it lies slightly further upstream of the poly(A) addition site (97bp compared to 76bp) and so its conservation in another gene from another trypanosome species is probably significant. In the light of the subsequent data which suggested that cDNAP4 is not expressed stage-specifically, this conserved region is unlikely to be the sole motif involved in the posttranscriptional control of procyclin expression but may play an important role in message stability or translation.

4.9.2 cDNAs P6 & P8

The sequences of cDNAs P6 and P8 (Fig. 4.22b) start at precisely the same nucleotide and yet, as the sequences diverge downstream, they cannot have arisen

a



b

```
  1  GGAGAAAAGGAGGTTTGTATTGAGAGGATGTGAAACTTGAATAACTACAT    50
 51  TTTGTTTGTTCGTCTGTTTTATGATGAGCAACGACTTGTATTCGAAAGAA   100
101  AACAAAAAAAAGAGGTATGTTCCAAATTGTATTGATAATTTTTTTCGTTT   150
151  TTTTATTTAAGGTGAGACTAACTCATTCTGAGAAGGAAAAGAGTAGTGTG   200
201  AAGGAAAATGATCAATAATATGAGAGGAAAAAAGGTTGGAAAAAGAAGAA   250
251  AGAATAGGGGATGAGTATCAACAAAAGCAAAAAATAGAAGCCAAATACCC   300
                aatgagaaaca(a)n
301  TTTAACGGTAGAAAGTAGCACAGAGAGAGGGGGGGGAATGTCCACGTTTG   350
351  GAGGAATGTCATGGGGCACGAGGGGCTCTGATGAGCAAAATATAGGAATG   400
401  ATGGGACCTAAGAAAGGAATTACAAAATGTTAACTTCAACAAGACATGTG   450
451  GAAAGAAGAAAAGGTGTGGGATTGAACTGATCGTTGCTCCTGTAACATGA   500
501  TAGGACATTCCGTGCTTACATGTACCGCATGTATGAAAAGGGGTTTGAAA   550
551  AATAATGCACAGAGTTCTTAAGAAAGAAATGAGAAGGAAGCTGGAGAGTA   600
601  TGATATGGATTTTCAAAGGGGGGGATTAGATGGAACGGAGGGCAAAGGGA   650
651  GGGGAACGACTCGTAATCGTGTGTAATGTTCACCTGTTGGTGTATACATT   700
701  TGGGGTGTGTTGGAGATGTAAAGAACACAAAATCCGAATTTCAGTGGGTT   750
751  TTCCACTCGTTGTGCGTCCCCGTCATGAAAACGTTTTTTTCTTATCATAC   800
801  TTTCCATTCTTTCTGACTTTTATCATTTCACCGCATTCGACTATAGTTAC   850
851  ATCTGTGTTTATTTTGTCGCTTCATTATACGCTTGTGAAAAAAAAA  896
```

Fig. 4.22 Map and Sequence of cDNAP8

(a) Restriction map and sequencing strategy. E (*Eco* RI), Hc (*Hinc* II), Xh (*Xho* I).

(b) Nucleotide sequence. Sequence of cDNAP6 at point of divergence from cDNAP8 is shown below in lower case.

a

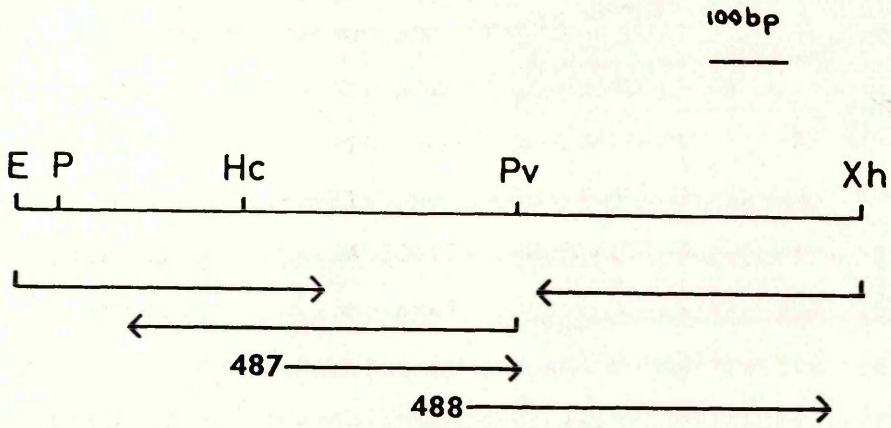


Fig. 4.23 Map and Sequence of cDNAP7.

(a) Restriction map and sequencing strategy. E (*Eco* RI), P (*Pst* I), Hc (*Hinc* II), Pv (*Pvu* II), Xh (*Xho* I). Numbers indicate oligonucleotide primers used.

(b) Nucleotide sequence.

b

```
   1  GACAACAAGATACGCCGCCGACATGCCGTGCAGCGTGACTCCAGAGCAGG   50
  51  AGCTGCAGGTCATCCAAACCATATGAGGCTTCGCATGCTGGGAAGTGAAG  100
 101  AATCAGGAGAGCGGCTTCGCATAAAAGTCGCCAGACGCTGCTTGAGTCGG  150
 151  TGGATGATGAAGCTGCTGTCGGAGCCGTGGAAGCGCTGTTGCGCCGCTAC  200
 201  GGGAAGGAAGTGAAGAAACTCGATGGCTCGTACGAAAGGGAGCGTGAGCT  250
 251  GAAGCGGCTCAGGACCGAGCGAGGGGCATTGCAGGGCTCACGGTTTGTTG  300
 301  ACGACGAGGCGGAGAGCGGGGAGGACGATGAGAGTGATAATGGCGAAGAA  350
 351  AATGAGGAGATTGACGGGAGGAGTAGCCACAACACACAGACGCTACAATG  400
 401  ATCGATGGTGAAAGGGGAAATGTGAAGGAGTTGTTACGGCTGCATTGGTT  450
 451  TTACACTTCAATGAGTCTTTGCTGCTTTTCACTCGTGTTCCCTTGACTTT  500
 501  TGATTGTGATGTGTGTTGTTGTTGTTGTTGTTCAAATTTCGCTAACGA   550
 551  TTTCTATCACCTTTTTGTGCTTGGGGGAGGGGGTCTGTGCCTCTCCGTGT 600
 601  TACTGGGGCGTTTGTTTTGGTGCTTGTGAGAAATCAGCAGCTGTGTGCTA 650
 651  TGTGCTTGTATGTGGAGGCATGTGTCTCTGATCCCCACACCTCGGGCCGG 700
 701  ACCAAGATGGAATGCTGAAGCACTAATGAAAAAATGTTATTGCTTCAAGA 750
 751  AAGGAGGAAATAAATAAGAGAGACCTCAGGAGGGCATGTTTTTCGTTTAT 800
 801  ATGCACATGATGTGGAGGCTGGTCACTTTGAATCACGTCTCCGCGTCCCG 850
 851  TGTGCTATTTACCCTTCTGTTTCAGGAGGGGGTGTTTGCTCAGCATTTAC 900
 901  GTTTGCCTCACGTGTCCCTTGTCATGGGCTATTCGACGCGCTGCCGACAG 950
 951  CTGCGGTGCGTCACGGACCTCACTGAGAGATTATTTGCCGGTGCATTTCC 1000
1001  GAGCTTCAGTCGTTGTTTGCTTACTTCCCTCTCTTTTAAAGCTGTTTATT 1050
1051  CAACAGGTGGGTAACGTTAGACCACTGGTACGAAAAAAAAAAAAAAAAAA 1100
```

from the same cDNA molecule. Neither cDNA contains any of the SL sequence and so neither is full-length, as was expected from a comparison of the sizes of the cDNA inserts (<300 and 896bp) and the mRNA (4kb) they detected in northern blot analysis. There is no open reading frame within the region cloned and the sequence is consistent with it being 3' untranslated region as it contains large runs of homopolymer.

4.9.3 cDNAP7

The sequence of cDNAP7 (Fig. 4.23b) is unremarkable except that it contains no significant open reading frame. The length of the clone insert (1.1kb) and the band it detects in northern analysis (1.3kb) suggests that most of the sequence has been cloned and so there is little potential for coding sequence upstream.

None of these cDNAs had any SL sequence at the 5'-end indicating that they were not full-length. In order to try and obtain full-length cDNA clones, 50,000 pfu of the cDNA library were plated out and quadruplicate lifts made onto Hybond N. One pair of filters was probed with the cDNA of interest while the other pair was probed with an end-labelled oligonucleotide containing the SL sequence. Hybridizing plaques were lined up to see whether any reacted with both the cDNA probe and the SL probe but no such coincident plaques were detected. Several plaques which hybridized with each cDNA probe were picked and excised anyway and small-scale plasmid preparations tested to see whether any of the inserts were bigger than those in the original cDNA clones. With cDNAP8, there were several clones which did appear to be bigger but on further analysis it was the vector fragment which appeared to hybridize.

4.10 Genomic Organisation of cDNAs P4, P7 and P8/6

Southern blots of *Eco* RI digests of *T.congolense* and *T.b.rhodesiense* genomic DNA were probed with the cDNA P4, P7 and P8/6 inserts in order to identify the organisation of the genomic sequences from which they had arisen. The blots were washed to 2XSSC at 65°C (Fig. 4.24). Weak hybridization to *T.brucei* DNA was observed in each case but to no greater extent than that found when *T.congolense* genomic DNA was probed with the *T.brucei* procyclin cDNA insert

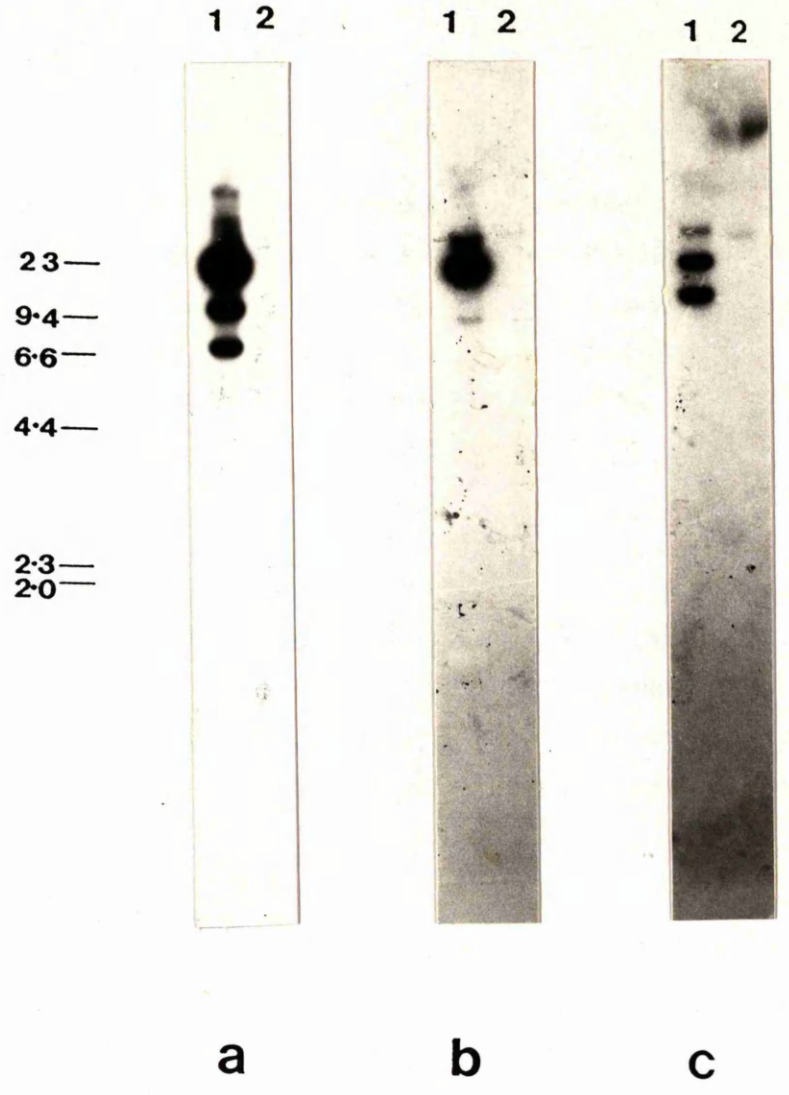


Fig. 4.24 Southern Blot Analysis of cDNAs P4, P7 and P8/6.

T.congolense "1/148" (lane 1) and *T.b.rhodesiense* EATRO 2340 (lane 2) genomic DNAs were digested with *Eco* RI, electrophoresed on a 0.8% agarose gel and blotted onto Hybond N membrane. The blots were probed with radiolabelled inserts of cDNAs P4 (a), P7 (b) and P8 (c) at 65°C in aqueous buffer and washed in 2XSSC/0.1% SDS at the same temp. before autoradiography. Markers were λ DNA digested with *Hin*d III,

and so the significance of this hybridization is not clear. cDNAs P8 & P6 identified two bands in *T.congolense* DNA which presumably represent the 2 copies of the gene encoding the 2 different cDNAs. Other *T.congolense* stocks (see section 4.13 & Fig. 4.12.) appear to have only one copy of this sequence. This could represent homozygosity of one locus with "1/148" being heterozygous. cDNAP4 identified 5 bands in *T.congolense* DNA and, as there is no internal *Eco* RI site in this clone, these presumably represent multiple copies of this gene. cDNAP7 appears to be single copy.

4.11 Construction of a *Sau* 3A1 Partial Genomic Library

In order to further analyse the genomic environment of the cDNAs, a genomic library was constructed.

4.11.1 Partial Digestion of the DNA

A series of tubes, each containing 1ug of *T.congolense* M15 1/148 DNA, were titrated for partial digestion by doubling dilutions of *Sau* 3A1 at 37°C for 1 hour. The resulting digests were fractionated on a 0.3% agarose gel (Fig. 4.25) and the enzyme concentration producing the largest amount of fluorescence in the range 18-22kb identified. This concentration was then used in a scaled-up reaction. Concentrations of twice and half as much enzyme were also used to ensure that DNA of the correct size was achieved. 100ug of DNA was digested with each concentration of enzyme after being prewarmed to 37°C before enzyme addition and each digest was allowed to proceed for exactly 1 hour. 1ug of each digest was electrophoresed on a 0.4% agarose gel and it was found that, while the highest concentration of *Sau* 3A1 used led to overdigestion, the other two digests scaled up accurately (Fig. 4.26). This material was extracted with phenol/chloroform and the DNA spooled out after ethanol precipitation at room temperature.

4.11.2 Size-Fractionation of the Genomic DNA

The partially digested genomic DNA was size-fractionated on a linear 1.25-5M NaCl gradient and 20ul samples from 0.5ml fractions run on a 0.4% agarose gel (Fig. 4.27). This indicated that there was good fractionation, with the first 3 fractions containing DNA that was too large to ligate into lambda. Fractions 4-8 contained DNA in the size range 18-22kb while fractions 9 onwards contained too much DNA of <10kb which could produce multiple inserts if ligated to λEMBL4. Fractions 4-8 were therefore pooled and dialysed against 5 litres of TE to reduce the

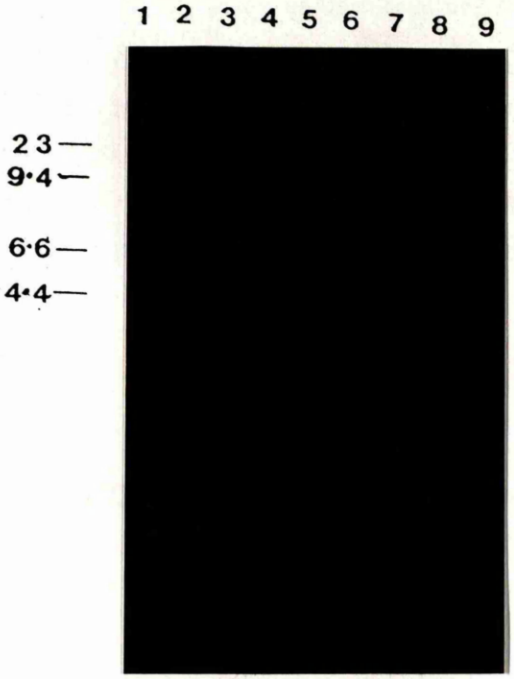


Fig. 4.25 Titration of *Sau* 3A1 Partial Digestion.

T.congolense M15 1/148 genomic DNA was digested with doubling dilutions of *Sau* 3A1 for 60 mins. at 37°C and electrophoresed on a 0.3% agarose gel before staining with EtBr. Lanes 1-8 - increasing dilutions of *Sau* 3A1, lane 9 - undigested control. Lane 7 appears to contain the most fluorescence in the region of 18-22 kb and this represents an enzyme concentration of 0.3125u.ug^{-1}. Markers - *Hin*d III digest of λ DNA.

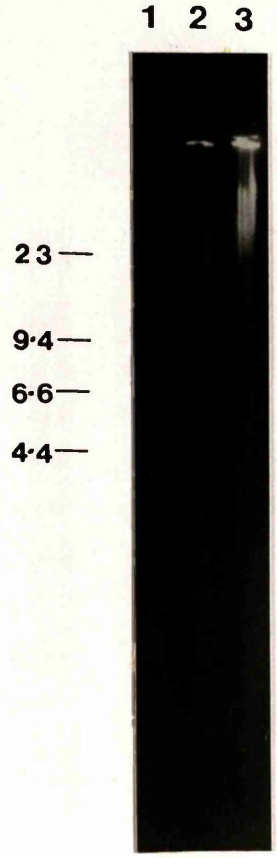


Fig. 4.26 Analysis of Large-Scale *Sau* 3A1 Partial Digestion.

100ug of *T.congolense* M15 1/148 genomic DNA was digested with 8 units (lane 1), 4 units (lane 2) and 2 units (lane 3) of *Sau* 3A1 for 60 mins. at 37°C and 1ug of each digest analysed on a 0.4% agarose gel after stopping the reactions with 20mM EDTA. The gel was then stained with EtBr. Markers - λ DNA cut with *Hin*d III.

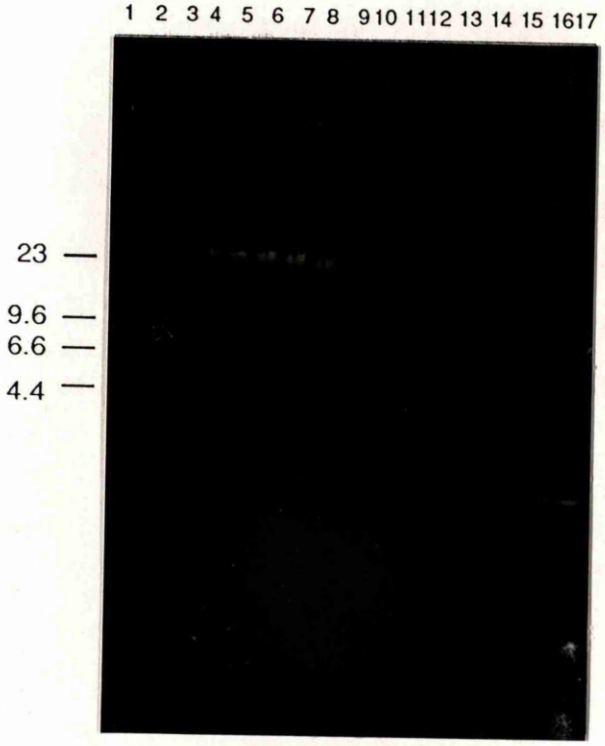


Fig. 4.27 Analysis of NaCl Gradient Fractionation.

The pooled *Sau* 3A1 partial digests were size-fractionated on a linear 1.25-5M NaCl gradient and 0.5ml fractions collected. 20ul of each fraction was analysed on a 0.4% agarose gel which was then stained with EtBr. Fractions 4-8 appeared to contain DNA in the correct size range of 18-22 kb and so were pooled for ligation into λEMBL4. Markers - λ DNA cut with *Hin*d III.

salt concentration before ethanol precipitation and resuspension in TE.

4.11.3 Preparation of the λEMBL4 Arms

The DNA ends created by digestion with *Sau* 3A1 are compatible with *Bam* HI ends and so the digested genomic DNA was inserted into the *Bam* HI site of λEMBL4. This is a replacement vector which has *Sal* I sites between the *Bam* HI sites and so a *Bam* HI/*Sal* I double digest is used to prevent the internal "stuffer" fragment from being able to ligate back in, and therefore increase the percentage of recombinant clones in the library (Frischauf *et al*, 1983). It was first necessary to anneal the *cos* ends of the vector and ligate them together to allow concatemer formation which is necessary for subsequent packaging of the DNA. The vector DNA was then digested with *Bam* H1 and *Sal* 1, phenol/chloroform extracted and ethanol precipitated after running a sample on a gel to check for digestion.

4.11.4 Ligation to the Vector Arms and Packaging

500ng of λEMBL4 arms were either self-ligated or ligated to 250ng of the size-selected genomic DNA digest in a final volume of 5ul. After an overnight incubation at 4°C, half of this ligation was packaged using the Gigapack Gold Packaging Kit.

4.11.5 Titration and Amplification of the Library

The library was titrated by plating aliquots on *E.coli* LE392, SURE™, and Q359. LE392 is a standard host for phage lambda and Q359 is isogenic with LE392 except that it is lysogenic for phage P2 and will therefore only permit growth of lambda phage with a *Spi*⁻ phenotype (Karn *et al*, 1980). As non-recombinant λEMBL4 is *Spi*⁺, it will not grow in this host. However, recombinant phage containing inserts have lost the stuffer fragment containing the *red* and *gam* genes, are *Spi*⁻, and so will plate on P2 lysogens. In this way it is possible to determine the percentage of recombinant phage in the library and indeed to select only these phage when screening. *E.coli* SURE™ is highly recombination deficient, allowing growth of phage containing inserts which might otherwise recombine, such as inverted repeats (Greener, 1990). It is however *rec*A⁺ which is essential for *Spi*⁻ phage replication.

The titre of the library was 8.9 X 10^5 pfu of which at least 85% were recombinant. One genome equivalent of DNA should be represented in approximately 10^4 pfu and therefore the unamplified library contained more than 75 genome equivalents. As the titre on *E.coli* SURE™ was at least as high as on *E.coli* LE392 it was decided to use the former as the preferred host after selecting out non-recombinant phages on *E.coli* Q359 in order to retain as many truly representative sequences as possible.

Approximately 20 genome equivalents (2 X 10^5 pfu) of the library were amplified in *E.coli* Q359, producing an amplified titre of 1.2 X 10^9 pfu.ml^{-1}. The remainder was kept unamplified in the event of sequences not being present at a representative level in the amplified library due to poor growth.

4.12 Screening of the Genomic Library with the Procyclic cDNAs

Approximately 5 X 10^4 pfu of the unamplified library were plated on *E.coli* Q359 and duplicate lifts of the plaques onto Hybond N filters were probed with the inserts from each of the cDNA clones, P4, P7 & P8, as well as with cDNAP1. After washing to 0.1XSSC at 65°C, around 10-12 plaques hybridized to cDNAP8, as would be expected of a gene present at 1-2 copies per haploid genome. Similar numbers of plaques were also detected by cDNAs P4 and P7 while cDNAP1 detected many plaques indicating that, despite selection against telomeres by this method of generating clones (fragments from chromosome ends would have a *Sau* 3A1 site at one flank and a non-ligatable telomere at the other), several telomeric copies were likely to have been cloned in the library. None of the cDNAP1-hybridizing clones taken through further screening contained a 7kb *Eco* RI fragment which hybridized to the probe and so were discarded and no further analysis of cDNAP1 was attempted.

Five positive plaque areas detected by each of the other cDNAs (P4, P7 & P8) were picked for further screening after which several genomic clones of each sequence were obtained.

4.13 Analysis of the Genomic Sequences

Before detailed analysis and mapping could be performed on the genomic

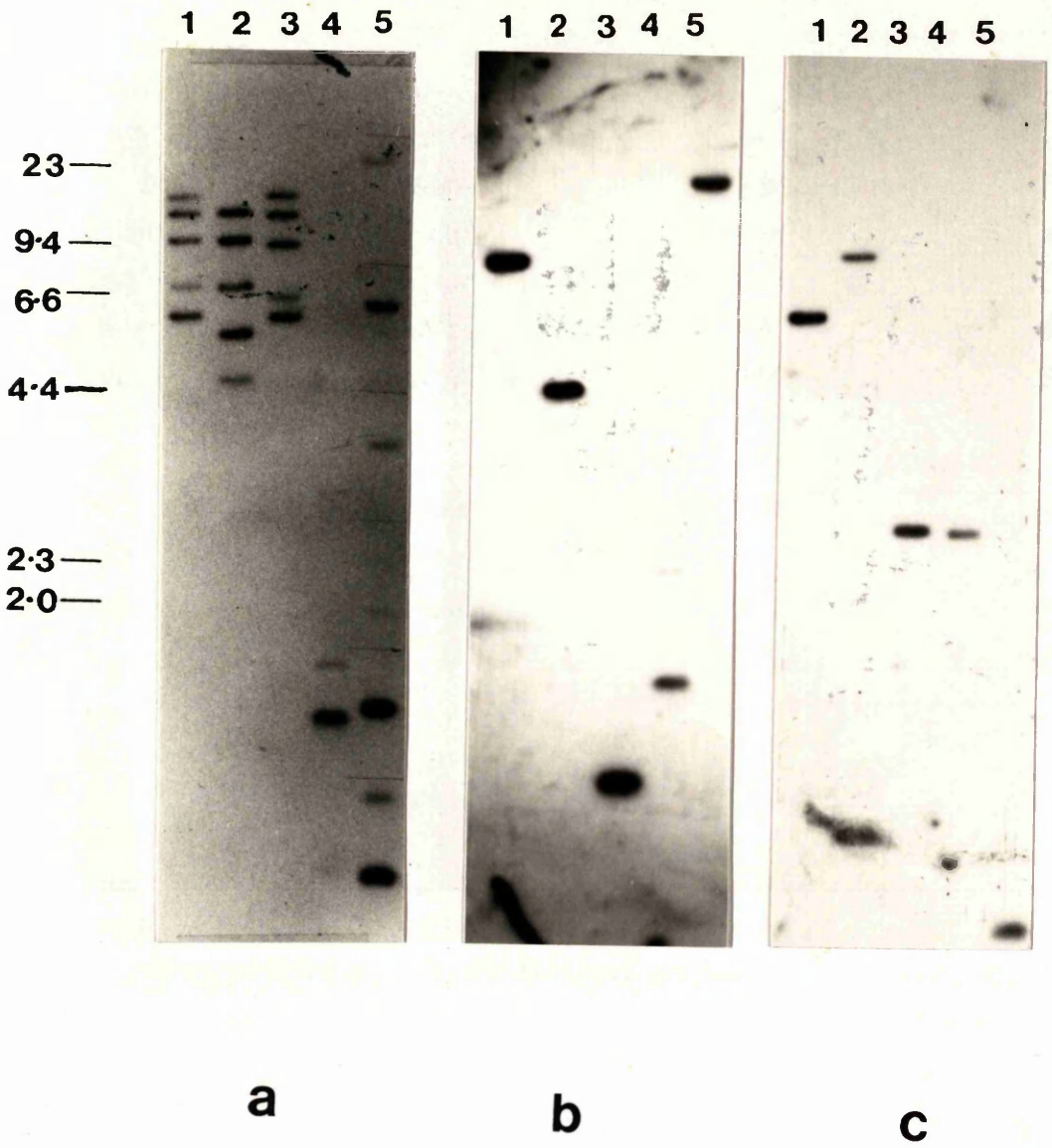
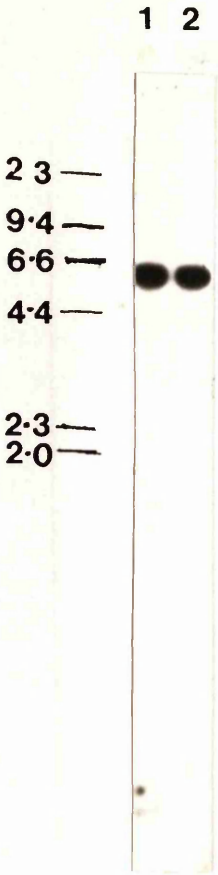


Fig. 4.28 Genomic Mapping of cDNAs P4, P7 and P8/6.

(a - c) *T.congolense* M15 1/148 genomic DNA was digested with *Eco* RI (lane 1), *Hin*d III (lane 2), *Pst* I (lane 3), *Pvu* II (lane 4) and *Bam* HI (lane 5), electrophoresed on a 0.8% agarose gel and blotted onto Hybond N membrane. The blots were probed with radiolabelled inserts from cDNAs P4 (a), P7 (b) and P8 (c) at 65°C in aqueous buffer and washed in 0.1XSSC/0.1% SDS at 65°C before autoradiography. Markers - λ DNA cut with *Hin*d III.



d

(d) Genomic DNA from *T.congolense* M15 1/148 procyclic cells transformed from the original bloodstream stock (lane 1) and the genomic DNA used to construct the λEMBL4 library (lane 2) were digested with *Eco* RI, electrophoresed on a 0.8% agarose gel and blotted onto hybond N membrane. The blot was then hybridized to radiolabelled cDNAP8 insert at 65°C, washed to a stringency of 0.1XSSC/0.1% SDS at 65°C and exposed for autoradiography. Markers - λ DNA cut with *Hin*d III.

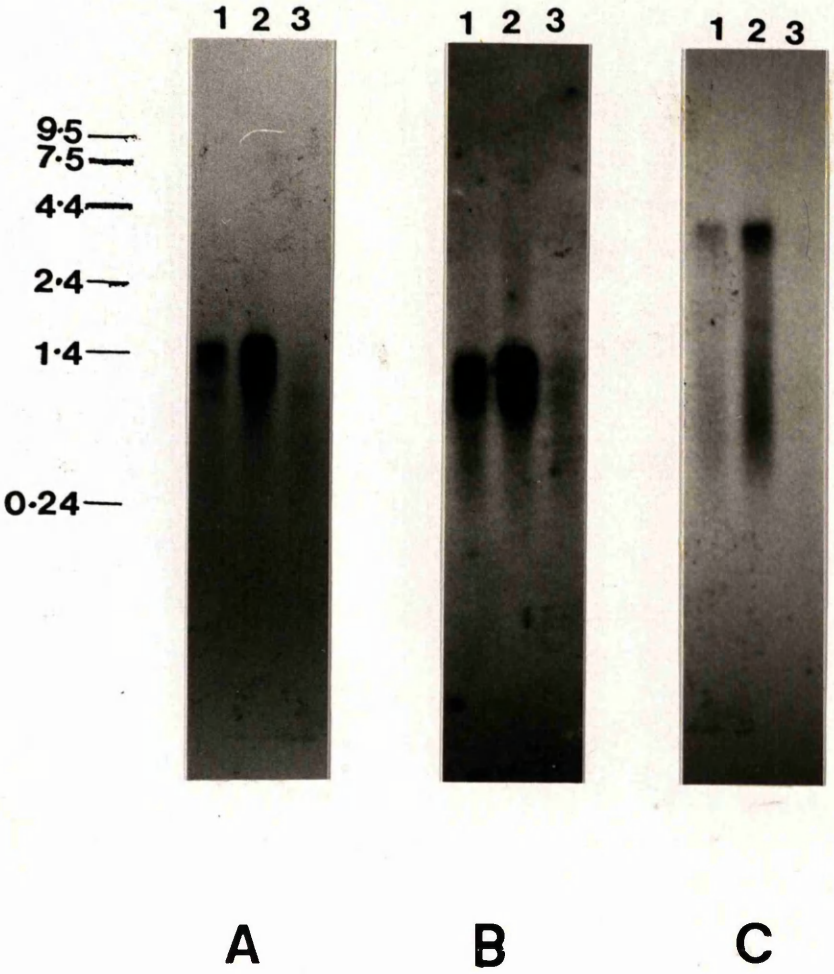


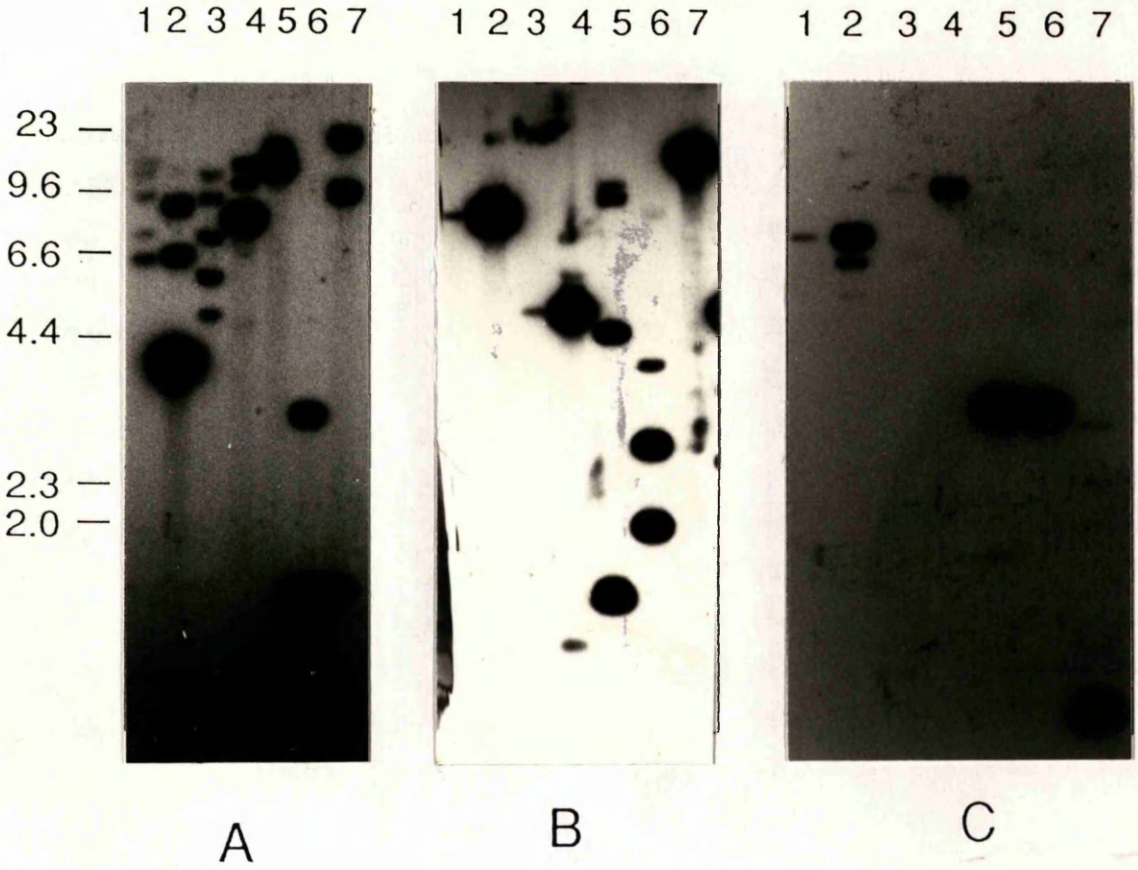
Fig. 4.29 Northern Blot Analysis to Check Expression of cDNAs P4, P7 & P8/6 in *T.congolense* M15 1/148 Procyclics.

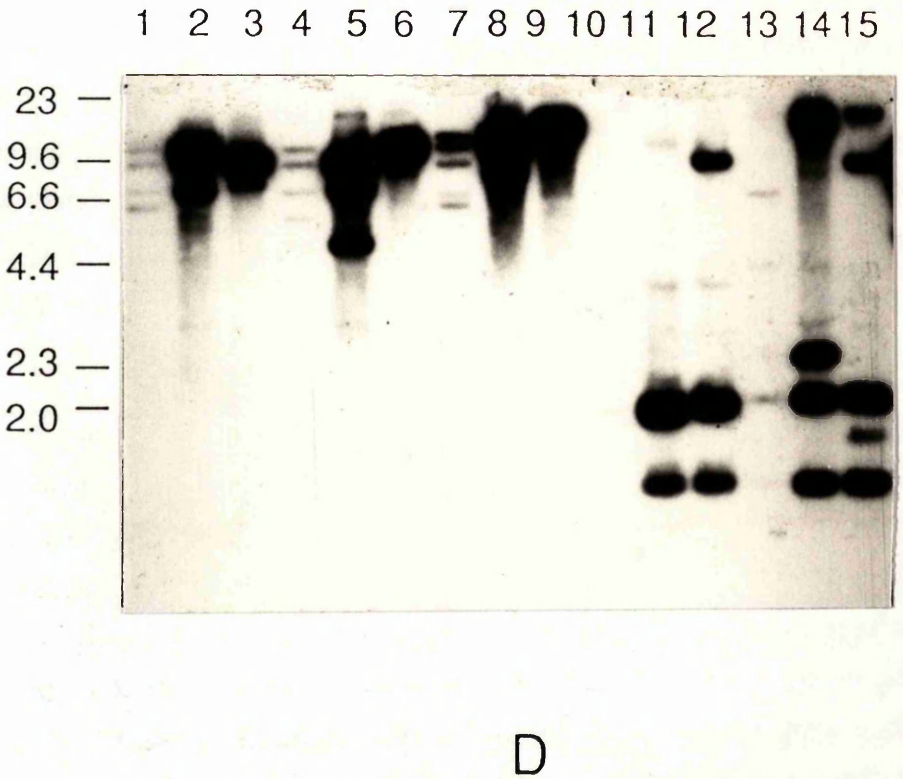
20ug total (lane 1), 3ug poly(A)$^+$ (lane 2) and 20ug poly(A)$^-$ (lane 3) RNA from *T.congolense* M15 1/148 procyclic cells was electrophoresed on a 1.0% formaldehyde gel and blotted onto Hybond N membrane. The blots were hybridized with radiolabelled insert fragments from cDNAs P4 (A), P7 (B) and P8 (C) at 42°C in formamide buffer and washed in 0.1XSSC/0.1% SDS at 65°C for 1 hour before autoradiography for 3 days (P4) or 6 days (P7 & P8) at -70°C with intensifying screens. Markers - BRL RNA ladder.

clones it was necessary to determine genomic maps more completely for each cDNA using Southern blotting. The DNA used to make the library was digested with the enzymes *Eco* RI, *Hin*d III, *Pst* I, *Pvu* II and *Bam* HI, run out on a 0.8% agarose gel and blotted onto Hybond N. The blot was then probed with each of the cDNAs separately and washed to high stringency (Fig. 4.28a-c). After autoradiography it was clear that cDNAP8 detected only one band in the *Eco* RI digest which was of a size different from that of the two bands detected previously. Reprobing of the original blot with the same probe still detected the original two bands indicating that the DNA used to make the genomic library must be from a different stock ("B") of trypanosomes from that used earlier ("A"). This problem arose because the procyclic cultures died for some unknown reason and could not be revived from stabilate, despite several attempts at recovery. Fresh cultures of what was believed to be the same trypanosome line were obtained from the Centre for Tropical Veterinary Medicine, Edinburgh and used to make the library.

It was at this stage necessary to determine which stock was the real M15 1/148, equivalent to the original stock ("C") of bloodstream forms obtained from Bristol. Because of poor yield of bloodstream forms, bloodstream trypanosomes of stock "C" were put into culture at 28°C and allowed to transform to procyclics. Although these cultures did not become established and lasted no more than about 2 weeks, enough DNA was isolated to compare to the stock "B" procyclic DNA. Both DNAs were digested with *Eco* RI and a Southern blot of the digests was probed with cDNAP8 (Fig. 4.28d). This showed that the stock "B" procyclic DNA gave a hybridization pattern identical to that from the bloodstream-derived procyclics from stock "C" and that therefore it was the procyclic stock "A" which had been misnamed.

As the cDNA library had been constructed and screened with procyclic RNA made from stock "A" but bloodstream RNA made from stock "C", it was possible that the reason for the apparent stage-specific expression of the cDNAs was in fact a stock-specific effect. It was clear from the Southern data that homologues for the genes are present in M15 1/148 but it was necessary to check by northern analysis that these copies are expressed. Total, poly(A)$^+$ and poly(A)$^-$ RNA from population "B" procyclics were run out on a formaldehyde gel, blotted onto Hybond N and probed with cDNAs P4, P7 and P8 at 42°C in 50% formamide buffer. After washing to 0.1 XSSC at 65°C and autoradiography it was found that these sequences are


1 2 3 4 5 6 7
1 2 3 4 5 6 7
1 2 3 4 5 6 7
23
9.6
6.6
4.4
2.3
2.0
A
B
C


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
23
9.6
6.6
4.4
2.3
2.0
D

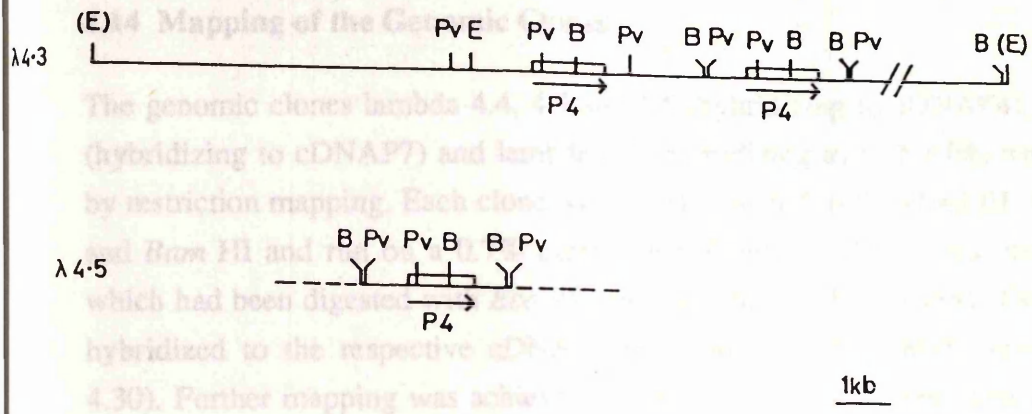


Fig. 4.30 Mapping Digests of Genomic Lambda Clones.

(A-C) Total genomic DNA from *T.congolense* M15 1/148 was digested with *Eco* RI (lane 1) and *Hin*d III (lane 3) and run out on 0.7% agarose gels alongside *Eco* RI (lane 2), *Hin*d III (lane 4), *Pst* I (lane5), *Pvu* II (lane 6) and *Bam* HI (lane 7) digests of the genomic lambda clones lambda 4.4 (A), lambda 7.5 (B) and lambda 8.2 (C). The gels were blotted onto Hybond N membrane and probed with the relevant radiolabelled cDNA insert at 65°C before washing in 0.1XSSC/0.1% SDS at 65°C for 1 hour and autoradiography. Markers - λ DNA cut with *Hin*d III.

(D) Total genomic DNA from *T,congolense* M15 1/148 was digested with *Eco* RI (lane 1), *Hin*d III (lane 4), *Pst* I (lane 7), *Pvu* II (lane 10) and *Bam* HI (lane 13) and run out on a 0.7% agarose gel beside the same digests of genomic clones lambda 4.3 (lanes 2,5,8,11 and 14) and 4.5 (lanes 3,6,9,12 and 15).The gel was probed with radiolabelled cDNAP4 insert at 65°C and washed in 0.1XSSC/0.1% SDS at 65°C for 1 hour before autoradiography. Markers - λ DNA cut with *Hin*d III.

(E) Preliminary restriction maps of genomic clones lambda 4.3 and 4.5 indicating positions of homology to cDNAP4. (E) - *Eco* RI sites in vector, Pv (Pvu II), B (Bam HI).

indeed expressed in M15 1/148 procyclics to produce transcripts of a similar size to those found in population "A" (Fig. 4.29).

4.14 Mapping of the Genomic Clones

The genomic clones lambda 4.4, 4.3 and 4.5 (hybridizing to cDNAP4), lambda 7.5 (hybridizing to cDNAP7) and lambda 8.2 (hybridizing to cDNAP8) were analysed by restriction mapping. Each clone was digested with *Eco* RI, *Hin*d III, *Pst* I, *Pvu* II and *Bam* HI and run on a 0.7% agarose gel alongside M15 1/148 genomic DNA which had been digested with *Eco* RI and *Hin*d III, and then blotted. The blots were hybridized to the respective cDNA probes and washed to high stringency (Fig. 4.30). Further mapping was achieved by performing double restriction digests with combinations of the above enzymes.

Eco RI and *Hin*d III fragments from the genomic clones lambda 4.4, 7.5 and 8.2, and which hybridized to the cDNA probes, were chosen for subcloning. Initially, the lambda clones were digested with the appropriate enzyme and, after phenol extraction and ethanol precipitation, the entire digest was ligated to pBluescript which had been cut and dephosphorylated as appropriate. After transformation of the ligations into XL1-blue, white colonies were picked onto arrays and colony lifts screened for hybridization to the cDNA probes. In this way, the 5.9kb *Hin*d III fragment from lambda 7.5 was subcloned. Although *Hin*d III inserts from lambda 4.4 were also detected in this way the colonies would not grow up and so had to be discarded. Other subclones (plambda4.4HIII, plambda4.4Eco and plambda7.5Eco) were obtained by purifying the desired bands from LMP agarose gels and ligating directly to dephosphorylated pBluescript.

Neither of these methods proved successful for subcloning the *Eco* RI fragment from lambda 8.2, despite several attempts. Transformants from a ligation of this fragment to pBluescript did on occasion hybridize to cDNAP8 but small-scale plasmid DNA preparations of these colonies did not contain the correct size of insert and indeed the plasmids did not produce the correct size of vector bands either, suggesting that they might be rearranged.

Since the *Eco* RI fragment appeared to be unclonable, an attempt was made to subclone other digestion fragments from lambda 8.2 but again colonies which hybridized to the cDNAP8 probe appeared to be rearranged.

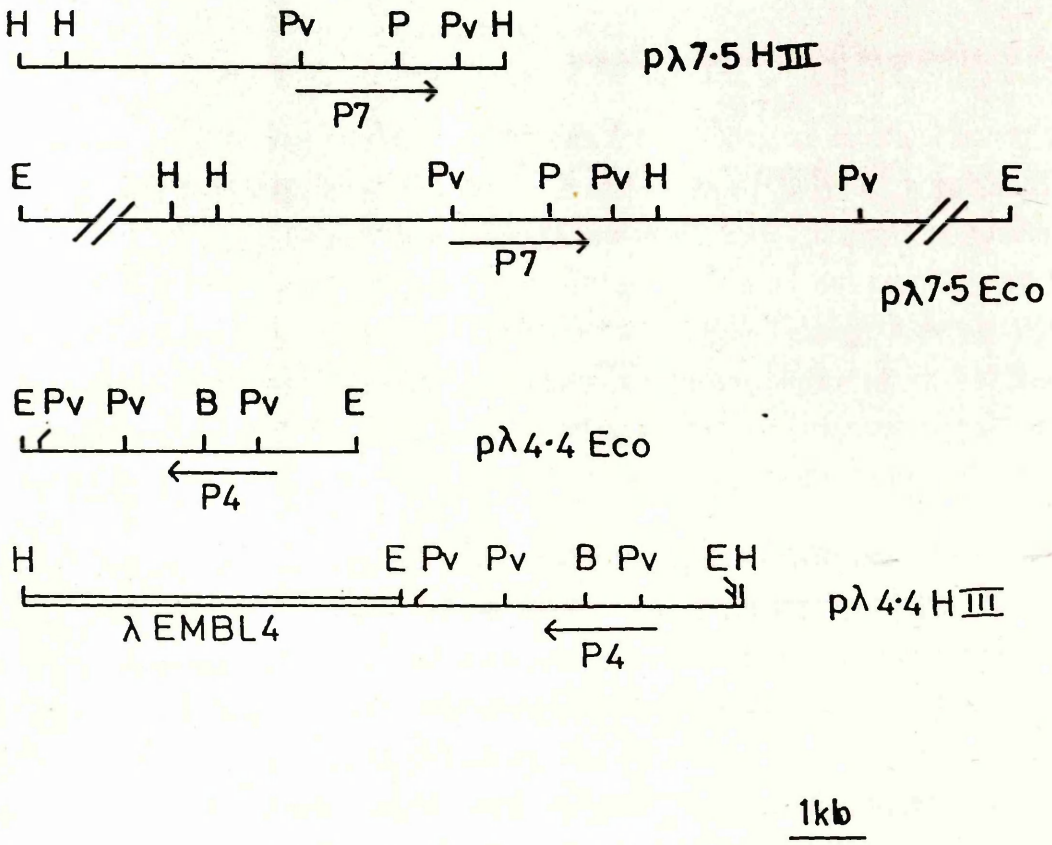


Fig. 4.31 Genomic Subclone Maps.

The maps of the genomic subclones pλ4.4Eco, pλ4.4Hind, pλ7.5Eco and pλ7.5Hind are shown with the region homologous to the cDNA probe indicated. H (*Hin*d III), Pv (*Pvu* II), P (*Pst* I), E (*Eco* RI), B (*Bam* HI).

Why should a large lambda clone of the cDNAP8 gene behave normally and yet subclones of this fragment get rapidly rearranged? The lambda clones were grown on *E.coli* SURE™ which, as indicated earlier, is highly deficient in its ability to allow recombination so that even inverted repeat sequences are stable in this host. The subclones, on the other hand, were grown in *E.coli* XL1-blue which, although *rec*A$^-$, is otherwise recombination competent. It was therefore decided to transform the subcloning ligations into *E.coli* SURE™. Unfortunately, because of its lack of recombination and repair functions, this host grows poorly and transformation efficiencies were found to be extremely low. In order to try and improve on this efficiency, an attempt was made to electroporate the DNA into the cells but again far lower efficiencies ($<10^6.ug^{-1}$) were achieved than expected and the subclones were not obtained. Whether this was due to the poor transformation efficiencies or unclonability is not clear.

4.15 Genomic Subclone Maps

Single and double digests of the genomic subclones were performed to derive basic restriction maps and the blotted gels were then probed with the cDNA inserts to ascertain the location of the genes on the maps (Fig. 4.31).

4.16 Sequencing 5' of cDNAs P4 and P7

In the case of each subclone of lambda 4.4 and lambda 7.5, genomic sequence from both up- and downstream of that cloned in the cDNA was present. Oligonucleotide primers (oligos 311 and 302) were therefore designed which were complementary to a region near the 5' ends of cDNAs P4 and P7 respectively, to allow generation of sequence from upstream of the cDNAs via plasmid sequencing of the genomic subclones.

The sequence obtained from plambda4.4Eco (Fig. 4.32) indicated that the region of overlap with the cDNA was identical except for a single base difference at nucleotide 23 in the cDNA sequence. (The two sequences were derived from different trypanosome stocks so the occasional base substitution is not surprising.) There are no AUG codons upstream of the cDNA sequence within the open reading frame, indicating that the AUG at position 41 in the cDNA is the initiation codon.

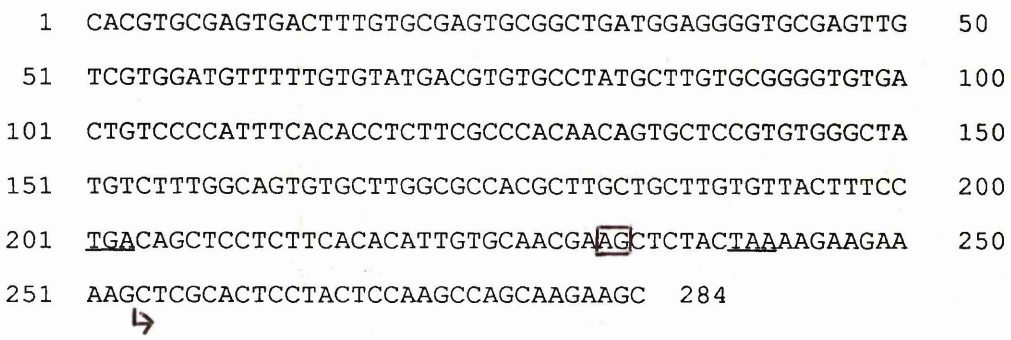


Fig. 4.32 Sequence of pλ4.4Eco Upstream of cDNAP4.

The sequence of pλ4.4Eco immediately upstream of the region homologous to cDNAP4 was obtained using the sequencing primer oligo 311, which is complementary to cDNAP4 from nucleotides 53 to 70, on plasmid template. The start point of cDNAP4 is marked with an arrow and in-frame stop codons upstream are underlined, indicating that the AUG at nucleotide 41 in the cDNA is the initiation codon. The AG 3'-splice site believed to be used in *trans* splicing is boxed (see section 5.2)

```
  1  CGGAATAGCCTTCGTCTTAGCTTGAGTGTGATGTTGGACGGGGGTTTGCG   50
 51  CCCACAATTGTGTGCTCTTTGCTTTGGGATACGCGCCTACCGTTGGTCAA  100
101  ATGGCGGTTGTTTGTCCAGTGGGAGATCGGTGTAAAAATTTAATGTGCCT  150
151  TTTGAAGCGTTTACGGTCAACTTTACGTAATTTACTTATTTTATAACTGG  200
201  ATGTCTTTTCTGCTTCCTGTTTTATTGGCTCCATTTCCGCTTATCATGTG  250
251  AGTGATGAGAGCGAAATGAGTTGCAGGTTTCCGAACCTAAACTGCGGTTG  300
301  TACTGCAGTTACCCCATTTTTCCTATCTACGGCGGCTGCCGTGTATAAAA  350
351  GAAAAGATAGACTGGGGTGATAAGAACGCTAGACAACAAGATACGCCGCC  400
401  GACATGCCGTGC  412
```

Fig. 4.33 Sequence of pλ7.5 Upstream of cDNAP7.

The sequence of pλ7.5Eco immediately upstream of the region homologous to cDNAP7 was obtained using the sequencing primer oligo 302, which is complementary to cDNAP7 from nucleotides 41 to 58 , on plasmid template. The start point of cDNAP7 is marked with an arrow.

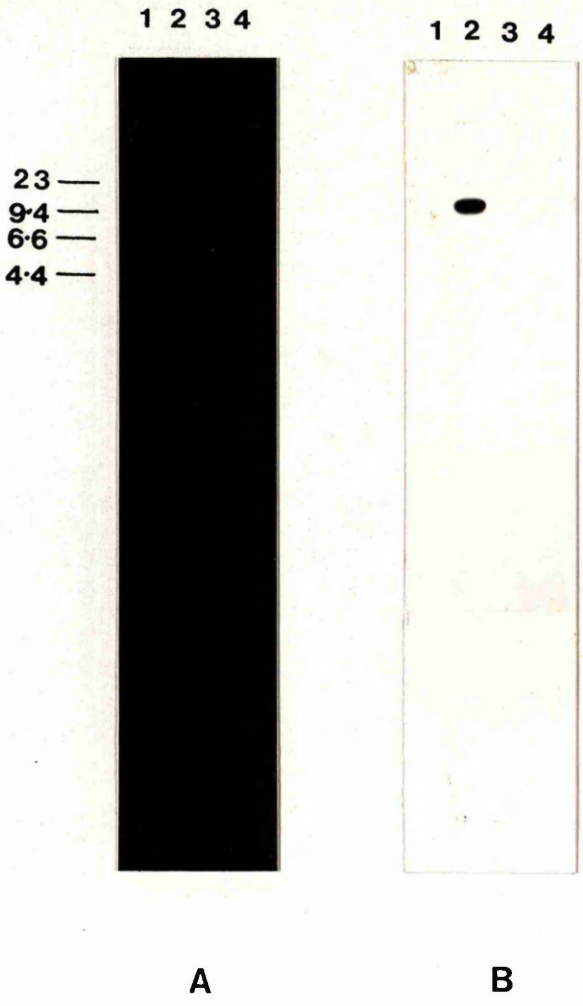


Fig. 4.34 Southern Blot Analysis to Check for the Possible Mitochondrial Origin of the cDNAP7 Gene.

The purified kinetoplast DNA network from *T.congolense* M15 1/148 was digested with *Eco* RI and electrophoresed on a 0.8% agarose gel alongside undigested kDNA and undigested and *Eco* RI digested total genomic DNA controls (A). The gel was blotted onto Hybond N membrane and hybridized with radiolabelled cDNAP7 insert at 65°C before washing in 0.1XSSC/0.1% SDS at 65°C for 1 hour and autoradiography (B).

Lane 1 - uncut total genomic DNA, lane 2 - *Eco* RI cut total genomic DNA, lane 3 - *Eco* RI cut kDNA, lane 4 - uncut kDNA.

Note the minicircle satellite band at 1kb in the total genomic DNA.

The sequence upstream of the cloned insert in cDNAP7 (Fig. 4.33) still provides no open reading frame, yet from size comparison it must contain the region that represents the 5'-end of the mRNA.

4.17 Are Any of the cDNAs Mitochondrially Encoded?

The mitochondrion of kinetoplastids is developmentally regulated both structurally and functionally, with the organelle being wholly functional only in insect stages of the lifecycle (Simpson, 1987). If these cDNAs were stage-regulated, the possibility arose that they may encode mitochondrial functions. cDNAP7 in particular was of interest because, despite the cDNA representing the majority of the length of the mRNA, and also having genomic sequence upstream, there was no obvious open reading frame. Several transcripts encoded in the mitochondrial DNA maxicircle do not have the same sequence as their DNA but are edited, by the addition and deletion of uridines, to produce a functional mRNA with an open reading frame (reviewed by Stuart, 1991). Partially-edited versions of these transcripts can be found (Bhat *et al*, 1990; Sturm & Simpson, 1990b; Decker & Sollner-Webb, 1990) and it was believed possible that cDNAP7 might represent one such partially-edited transcript if its genomic sequence was present in the kDNA maxicircle rather than the nucleus.

By searching for cDNAP7 sequences separately in kDNA and genomic DNA, this possibility could be investigated. The kinetoplast DNA network was prepared from one of the M15 1/148 total genomic DNA preparations and digested with *Eco* RI which digests the maxicircles into several bands and frees the minicircles which migrate as a single band on gels and can also be observed as a satellite of DNA in total genomic DNA preparations. This digest and uncut kDNA were run alongside uncut and *Eco* RI-cut total genomic DNA on a 0.8% agarose gel (Fig. 4.34a). A Southern blot of this gel was then probed with cDNAP7 and washed to high stringency (Fig. 4.34b). Neither the five maxicircle *Eco* RI fragments nor the minicircles in the kDNA tracks hybridized while the expected band was detected in total genomic DNA, excluding the possibility that cDNAP7 is mitochondrially encoded and therefore edited. The other cDNAs were not tested against kDNA as the genomic *Eco* RI fragments which they detect do not correspond to the sizes of the bands produced from *Eco* RI digests of kDNA.

4.18 Discussion

A unidirectional cDNA library was constructed from *T.congolense* procyclic poly(A)$^+$ RNA in lambda ZAP II and titrated on *E.coli* PLK-F'. It was necessary to use this host with phage from the primary library as the method used to construct the library generates hemi-methylated molecules which would be subject to restriction in certain hosts such as XL1-Blue (Raleigh & Wilson, 1986). After passing through PLK-F' (eg. during amplification), the library is then stable in XL1-Blue. The titration suggested that the library was of a good size as transcripts present at a level of 0.0002% should still be represented at least once.

Randomly picked plaques yielded inserts of various sizes, including 0.6 - 4 kb and, as the average length of a transcript in trypanosomes is around 2kb, it seemed that some full-length cDNA molecules should be present. The sizes of some of the inserts however were small. As only the first peak of radiolabel from the size-fractionation step was used to construct the library, this indicates that the assumption that the second peak was composed of only partially synthesised products was correct. Clearly small transcripts such as that for the L29 ribosomal protein were still represented in the library. Despite the indication that large cDNA inserts were present in the library, only about 3% of the plaques hybridized to the SL sequence and therefore contained full-length cDNAs. However, in the case of the L29 ribosomal protein cDNA, the insert was essentially full-length as the last 4bp of the SL sequence were present at the 5'end of the sequence but such a short part of the SL would not have hybridized under the hybridization and washing conditions used. Even in the case of cDNAP4, where the SL sequence is not present at all, results indicate that the mRNA is only of the order of around 55bp longer at the 5'-end than the cDNA (see chapter 5) and the AUG initiation codon is present in the insert.

The initial hope that the library could be expressed and the GUGM monoclonal antibodies used to probe for the surface antigen gene did not prove fruitful. Subsequent western blot analysis of the purified antigen, which had been treated with periodate to remove carbohydrate moieties prior to probing with the mabs, did not produce any signal (E.Kilbride, 1992) while the untreated controls gave the expected signal and antisera that were against protein epitopes did react with the periodate-treated blot. This result suggests that GUGM 2.2 is directed against carbohydrate or carbohydrate/protein epitopes and thus would clearly not be expected to react on blots of expressed cDNA fusion proteins where appropriate

glycosylation does not occur.

In the differential screen, there appeared to be many differences detected by the procyclic and bloodstream probes, but these were largely due to quantitative differences in level of hybridization rather than qualitative, stage-specific signals. Interestingly, a few of the bloodstream cDNA-hybridizing plaques did not appear to hybridize to the procyclic probe. Whilst it is to be expected that there are bloodstream-specific transcripts (the VSG for example), these should not be present in a procyclic cDNA library. The most likely explanation is that the apparently bloodstream-specific transcripts in the library are just expressed at a much lower level in the procyclic form, to the extent where the hybridization signal is not detected. This is also likely to be the case for some of the procyclic-specific signals, especially as the bloodstream cDNA probe was not labelled to quite such a high specific activity, but several of these procyclic clones did give intense signals with the procyclic probe and were therefore more likely to be genuine stage-specific cDNAs. The subsequent northern analyses indicated that this was not the case as all but one of the cDNA clones taken through secondary screening were expressed in both stages of the life cycle.

While single digests indicated that cDNAP10 contained an insert, it was not possible to re-excise it with *Eco* RI and *Xho* I as the *Eco* RI site appeared to be missing. 5'-end sequence analysis of this clone indicated that a poly(A) tail was located at this end, *ie.* in the wrong orientation, and that the *Xho* I end had ligated directly to the *Eco* RI site in the vector. The only explanation for this is that the two flanks had formed flush ends, either by chew-back of the overhanging end or by filling in of the recessed end, and had then been blunt-end ligated. There also appeared to be a poly(A) tail at the other end of the clone, inserted correctly. Such aberrant ligation products have also been found in other cDNA libraries in lambda ZAP II generated in our laboratory and also in a custom-made drosophila cDNA library obtained from the manufacturer. Clones with no poly(A) tail at either end, as found in the case of cDNAs P5 and P1, also appeared to be quite common. These are not likely to have been due to DNA contamination of the library, as the northern blot of RNA treated with DNase and RNase indicated, at least in the case of cDNAP1, that signal was due to RNase-sensitive material and not DNase-sensitive material. This tends to indicate that these clones are artifactual products of either the method of generating the cDNA or of the vector.

Two methods were utilised to obtain sufficient bloodstream forms for RNA

purification. The first, described by Rosen *et al* (1979) involved warming the mice at 37°C before bleeding. This method produced RNA which displayed no reactivity with several of the cDNAs in northern analysis, suggesting that these cDNAs were procyclic specific. The second required that terminal parasitaemias be used to maximize trypanosome numbers, but the animals were not warmed prior to bleeding. This second method produced RNA which gave contradictory results, suggesting that the cDNAs were not stage-specific and that the transcripts were also expressed in steady-state bloodstream RNA. Why should these two methods produce such contrasting results? The main concern with the first method is that the warming of the mice may be stressful and that a stress response might be invoked. This might affect the trypanosomes as well since the parasite depends on host fluid for so much and alteration in host physiology (*eg.* fat mobilization, *etc.*) may result in a major alteration of the parasite's environment. Clearly however, a classic stress response, in which general transcription is repressed, is not involved as many transcipts are still present. While the second method involves no unnatural procedures, the possibility arises that a large number of the parasites have differentiated to the *T.congolense* equivalent of the short stumpy form (Nantulya *et al*, 1978; these are actually longer than the dividing forms in this species) by this late stage in the infection. As the short stumpy stage is pre-adapted to life in the tsetse fly (Vickerman & Barry, 1982), it is possible that they may have already activated some gene functions required by the procyclic stage. Indeed, the short stumpy form of *T.brucei* already has an incomplete mitochondrial respiratory machinery. Neither of these isolation methods is therefore ideal and in order to obtain a true result, the trypanosomes were grown in rats over several waves of parasitaemia to increase numbers available in the general bloodstream. They were then harvested by cardiac puncture before the parasitaemia had peaked. RNA derived from bloodstream forms isolated in this way indicated that cDNAP4 transcripts at least were not stage-specific. A possible conclusion of this work is that the clones were identified in the differential screen because the RNA preparation used to make the bloodstream cDNA probe was derived from trypanosomes isolated after warming the mice. The absence of any genuine procyclic-specific clones in the library is in agreement with the findings of Roditi *et al* (1987) and Mowatt & Clayton (1987) who found no other stage-specific transcripts in their procyclic libraries from *T.brucei* by differential screening except the procyclin message. If there is no procyclin in *T.congolense* then there would be no highly expressed stage-specific transcripts, but this suggests that molecules on the surface of *T.congolense* procyclics are not controlled in a stage-specific manner at the level of transcription.

In the absence of any readily detectable stable mRNA in *T.congolense* which is specific to the procyclic stage, the detection system used will have been pushed to its limits so that it is hardly surprising that junk and artifactual transcripts should be found.

While a ribosomal protein cDNA clone was of little relevance to the main aims of the project, it was interesting for two reasons: firstly this sequence is expressed at a relatively high level and produces a strong hybridization signal with both *T.congolense* and *T.brucei* RNA. The only other probe available in the laboratory for a highly expressed housekeeping gene was pTgBαβT-1, a genomic clone containing the alpha- and beta-tubulin repeat from *T.brucei* (Thomashow *et al*, 1983). Unfortunately the sequence of this repeat does not appear to be well conserved in *T.congolense* as the probe does not produce a very strong hybridization signal on DNA or RNA of this species even at the reduced stringency conditions of 2XSSC at 55°C. The ribosomal protein cDNA was therefore much more appropriate as a control probe in northern blots. Secondly, alteration in ribosomal protein L29 is one basis of cycloheximide resistance in yeast (Kaufer *et al*, 1983). Resistance to this drug can be brought about by the conversion of a glutamine residue at position 38 to glutamic acid. A recent study (Yao & Yao, 1991) showed that site directed mutagenesis of methionine to glutamic acid at the equivalent position in L29 in cycloheximide-sensitive *Tetrahymena* also produced resistance to the drug. cDNAPE encodes a glutamic acid at this position which suggests that the protein it encodes would confer resistance to cycloheximide. In general, *T.brucei* at least is sensitive to this drug (Ehlers *et al*, 1987 and S.V. Graham, pers.comm.) but whether the *T.congolense* stock used to construct the library was resistant is not known and we have been unable to revive liquid nitrogen stabilates of this stock of unknown origin. If a cycloheximide resistant form of this protein could be introduced into a suitable trypanosome line by the newly available technique of gene replacement (Lee & Van der Ploeg, 1990a; ten Asbroek *et al,* 1990; Cruz & Beverley, 1990; Eid & Sollner-Webb, 1991) then it may provide an ideal marker for genetic studies.

The *Bal*-31 experiment clearly indicates that the cDNAP1 sequence is derived from telomeres and this would certainly explain why the clone is lacking many restriction sites as trypanosome telomeres tend to be barren for these (Hoeijmakers *et al*, 1980). What is more surprising is that the cloned 2.4kb fragment does not appear to be repetitive distal to the (CCCTAA) repeats (which start from the telomere inwards towards the centromere), because subtelomeric regions of both

trypanosomes and other organisms tend to be composed of degenerate repeats of the telomeric repeat itself and other longer sequences (Blackburn & Challoner, 1984; Van der Ploeg *et al*, 1984c). Another unexpected result is that there does not appear to be a homologue of the non-$(hexamer)_n$ region of cDNAP1 in *T.brucei*, despite the hexamer itself being identical between the two species. It is possible that only a subset of the *T.congolense* telomeres contain the cDNAP1 sequence and that these are different from the other telomeres and those present in *T.brucei* as there is no sign of *T.brucei* VSG characteristics in the cDNAP1 sequence. A telomeric motif in *Plasmodium berghei*, which exists as a highly repetitive element in the genome, also appears to be species-specific (Pace *et al*, 1987). Another possible explanation for these results is that the (CCCTAA) repeats present in cDNAP1 are not located at the end of the telomere, but rather are present further inside. The telomeric clone isolated from *T.brucei* by Blackburn & Challoner (1984) also contained the hexamer repeat between other DNA sequences, one of which was the 3'-end of a VSG gene. These sequences would still be sensitive to *Bal*-31 but may be outwith the telomere itself. Thus the possibility of telomeric copies of cDNAP1 encoding "real" genes still remains and, if so, would be the first example of a gene other than a VSG gene to be located at a telomere in African trypanosomes. The cDNAP1 sequence may have an important function and yet it must be specific to, at most, the subgenus *Nannomonas*.

The stage-specific expression of cDNAP1 is rather strange if it is due to non-specific transcription from telomeres. Transcription from telomeres has been observed in other trypanosomatids by Rudenko & Van der Ploeg (1989) who have suggested that it results from readthrough downstream of telomerically located genes, but it is not stage-specific, occurring in both bloodstream and procyclic stages. Telomere repeat-containing RNA has also been observed in higher eukaryotes, eg. humans (Cheng *et al*, 1991)). In trypanosomatids, this transcription appears to be predominantly unidirectional - from a chromosome internal position towards the chromosome end - except in *T.lewisi* where 40% of the transcription appears to proceed in the other direction (Rudenko & Van der Ploeg, 1989). As there is no poly(A) tail in any of the cDNAP1 clones, the direction of transcription is not clear. If the transcription is indeed non-specific, diffuse bands would be expected in northern blots. There is a background smear in the northern blots which may indicate a heterogeneous size population but there are discrete bands as well. The apparent stage-specificity may be explained by the fact that all of the bloodstream RNA used in the analysis of cDNAP1 was isolated by the method that

indicated other cDNAs were stage-specific. This may not be a genuine result.

One explanation for the absence of poly(A) tails in the cDNAP1 clones is that they represent pre-mRNAs which are still to be processed. The cDNA clone inserts are much longer than the bands on northerns, indicating that this may be so. However, the incidence of such clones in the library rather exceeds that expected for precursor RNA that should be short-lived and at least the majority of the transcripts would be expected to have a poly(A) tail. Transcription of telomeres in *T.brucei* appeared also to be distributed between poly(A)$^+$ and poly(A)$^-$ RNA (Rudenko & Van der Ploeg, 1989) and this may be a general feature of telomere transcription.

The presence of a genomic copy of the cDNAP1 sequence which is not apparently telomeric is intriguing. It is possible that it does represent the "real" gene, but it is at a very low copy number. However, several of the cDNA clones have different restriction maps and appear therefore to have different origins. An accurate elucidation of cDNAP1 sequences and their expression will require much more analysis, but this was outside the scope of this study. In addition, the presence of aberrant sequences in lambda ZAPII cDNA libraries may mean that cDNAP1 clones are artifactual products of the library construction, although the data obtained suggest this to be unlikely.

Are cDNAs P6 and P8 the products of differential polyadenylation of transcripts from the same gene copy or transcripts from different gene copies? The sequence of each is identical until 10bp upstream of the poly(A) tail in the shorter cDNAP6, where they diverge, suggesting that there are two different copies of the gene with different polyadenylation sites. The presence of two *Eco* RI bands in genomic DNA from the same trypanosome stock, which hybridize to either of the cDNA probes, agrees with this hypothesis as there are no internal *Eco* RI sites in the probes. Whether the two copies are alleles or from different loci is not evident from this blot alone but the presence of only one hybridizing fragment in several other *T.congolense* stocks (Fig. 4.12) suggests that they are homozygous for the gene and that P6 and P8 do represent different alleles in the stock of unknown origin. Which allele is present in the apparently homozygous stocks has not been investigated.

The question arises as to why no longer cDNA clones for P8/6 could be isolated from the library when the transcript is clearly much longer than the clones (4kb compared to 896bp) and the library contains much longer sequences for other genes. In light of the results from the genomic subcloning, where P8 homologous

sequences could not be subcloned in pBluescript from lambda, this may have been due to rearrangement and so the length of cDNAP8 may be the maximum size clonable and this might also explain why cDNAP6 had an identical 5'-end when it is clearly not derived from the same original cDNA molecule. How the cloning problem might be overcome, in order to obtain sequence of this unclonable upstream region and to ascertain what might be happening, is not yet clear. Either direct sequencing of the lambda clone or PCR amplification from mRNA using oligonucleotides, one complementary to the 5'-most sequence of cDNAP8 and the other containing the SL sequence, followed by sequencing without cloning may prove useful.

The sequence and length of cDNAP8 is consistent with its being derived from 3' untranslated region and so the function of this gene is not clear. Upstream sequence will have to be obtained before any conclusions can be drawn but it is interesting in that it is apparently single copy, an uncommon occurrence in trypanosomes.

cDNAP7 encodes a polyadenylated transcript of around 1.3kb which is expressed from a single copy gene. The kDNA blot indicates that the gene must be present in the nucleus, not the mitochondrion. To date, editing has only been observed for mitochondrially-encoded transcripts, not those encoded in the nucleus. There is still the possibility that nuclear editing does exist but it is unlikely. However, there does not appear to be any significant open reading frame in the cDNAP7 sequence and the sequence does not detect any obvious homologies in the database. The "gene" is conserved between different trypanosome stocks suggesting that it is functional but what that function might be is not evident.

cDNAP4 encodes a polyadenylated transcript of around 1.6kb which does not appear to be stage specific, except that it is at a lower level in bloodstream forms which have been isolated from animals which had been warmed to 37°C prior to bleeding. The Southern blot data indicate that it is a middle repetitive gene, with at least 5 copies in the genome as there are no *Eco* RI, *Hin*d III or *Pst* I sites in the cDNA but 5 bands are detected in genomic DNA in each case. The restriction map of lambda4.3 suggests that there may be more than 5 copies as there is more than one copy of cDNAP4 in this clone. Tandem linkage of multicopy genes is very common in African trypanosomes.

The sequence of cDNAP4 looks unremarkable except that it contains a

potential open reading frame of 256 amino acids. Upstream sequencing of plambda4.4Eco using a primer (oligo 311) complementary to the sequence at the 5'-end of the cDNA suggests that the first AUG codon in the cDNA is the initiation codon. There are no other AUGs upstream that lie downstream of a series of stop codons and the sequence around this AUG has a good match to the flagellate protozoa initiation consensus (Yamauchi, 1991). The putative protein contains no obvious signal peptide, although its amino terminal sequence is relatively hydrophobic, and there are no N-linked glycosylation signals but there is a good potential glypiation signal and putative C-terminal extension (Ferguson & Williams, 1988). The predicted molecular weight of the unprocessed protein is 26.5kDa and it has a predicted pI of 4.82. Neither the protein nor the DNA sequence detected any significant homologies to sequences already in the EMBL, GenBank or NBRF databases but this gene proved to be of significant interest and its analysis is described in more detail in the next chapter.

Posttranscriptional control appears to play an important part in gene regulation in trypanosomes. Many promoters appear to be constitutive (Pays *et al*, 1989b; Pays *et al*, 1990) and control of their genes must occur at a later stage than transcription initiation. The study of promoter function has only recently been possible with the development of a DNA transformation system in *T.brucei* (Laban & Wirth, 1989; Bellofato & Cross, 1989), but transient expression assays on the procyclin and VSG promoter regions by Jefferies *et al* (1991) have indicated that elements in the 3' UTRs of these genes have a significant effect on the level of Chloramphenicol Acetyl Transferase (CAT) expressed from a reporter gene inserted downstream of these promoters in transient expression vectors. Inhibition of protein synthesis leads to superinduction of procyclin mRNA levels (Dorn *et al*, 1991) suggesting that mRNA stability is controlled by a specific nuclease. Only a few short regions of homology exist between different copies of the procyclin 3'UTRs (Mowatt & Clayton, 1988) and it is believed that they may have a role to play in transcript stability or translation. The presence of one of these conserved blocks of homology in the 3' UTR of cDNAP4 is therefore of considerable interest. If transcripts from this gene had also been procyclic specific, as was originally thought, then this would have provided a strong indication that this sequence motif is involved in the control of gene expression from procyclic-specific genes via their transcription or by altering the stability/ instability of their transcripts as occurs in the genes for ompA (Melefors & von Gabain, 1988) and polynucleotide phosphorylase (Portier *et al*, 1978) in *E.coli* and several genes in higher eukaryotes

(Mullner & Kuhn, 1988; Levine *et al*, 1987; Wilson & Treisman, 1988; Brewer & Ross, 1988). This hypothesis could easily be tested by placing this motif downstream of the procyclin promoter/CAT construct in transient expression assays. However, as the cDNAP4 gene does not appear to be stage-specific, the sequence may play a more mundane role in gene expression. None of the other cDNAs sequenced in this study contain this motif but it will be interesting to see if any other trypanosome transcripts, yet to be analysed, do possess it. Another, unrelated, 16bp consensus sequence is found in the 3' UTR of VSG genes in *T.brucei* (Matthyssens *et al*, 1981; Boothroyd *et al*, 1981; Lenardo *et al*, 1984). It has been suggested (Koenig *et al*, 1989) that it too might play a role in gene expression in bloodstream forms. The presence of this consensus in *T.congolense* VSG genes would lend support to this hypothesis but the two *T.congolense* VSG gene cDNA sequences published to date have no 3' UTR (Strickler *et al*, 1987).

CHAPTER 5

ALANINE RICH PROTEIN - A SURFACE ANTIGEN ON PROCYCLIC *TRYPANOSOMA CONGOLENSE*

5.1 Introduction

In an independent study, as a follow-up to their studies on the procyclin protein of *T.brucei* (Richardson *et al*, 1986; Richardson *et al*, 1988), Beecroft *et al* (manuscript in preparation) had also raised monoclonal antibodies to the *T.congolense* procyclic culture form. As in our own study, these antibodies were predominantly against protein(s) on the trypanosome surface, which in this case migrated as broad bands of around 50-60kDa and 22-28kDa. This protein(s) appears on the surface of procyclic form *T.congolense* with similar kinetics to *T.brucei* procyclin during transformation from bloodstream forms (T.Pearson, pers. comm.). Apart from the size differences, which may in part be due to the percentage of acrylamide used in the gels, this protein(s) has properties similar to those which we had observed. Amino acid microanalyses performed on each band indicated that they were both rich in alanine and acidic residues, as was the case with the protein identified by our laboratory. Another possible explanation for the difference in size of the two proteins is that the Kilifi-type stock of *T.congolense* used by Beecroft *et al* (manuscript in preparation) may actually be a different species of trypanosome from the savannah-type *T.congolense* used in our study, as defined by various criteria including isoenzyme electrophoretic variation, RFLPs detected by repetitive DNA probes and molecular karyotypes, (Knowles *et al*, 1988; Majiwa *et al*, 1985 and 1986a,b; Masake *et al*, 1988). Species differences may well lead to size differences in the protein.

Microsequencing of purified antigen was unsuccessful due to blockage at the amino-terminus but Beecroft *et al* (manuscript in preparation) were able to obtain some peptide sequence after cyanogen bromide cleavage of the upper of the observed bands. The sequence of 2 cleavage peptides was kindly made available for comparison with the predicted amino acid sequences encoded by cDNA clones isolated in our laboratory. Both peptide sequences were found to have a high degree of homology to regions of the cDNAP4 open reading frame (20/23 amino acids in the case of peptide 2 and 18/22 amino acids in the case of peptide 5; Fig. 5.1), with most differences representing conservative amino acid changes or else being located towards the end of the peptides where the peptide sequence was becoming ambiguous. As both peptides were found to be present separately in the one coding sequence it seemed reasonable to assume that cDNAP4 represented a gene for the antigen identified by Beecroft *et al* and, by virtue of the similarity to results from our laboratory, the antigen detected by the GUGM mabs.

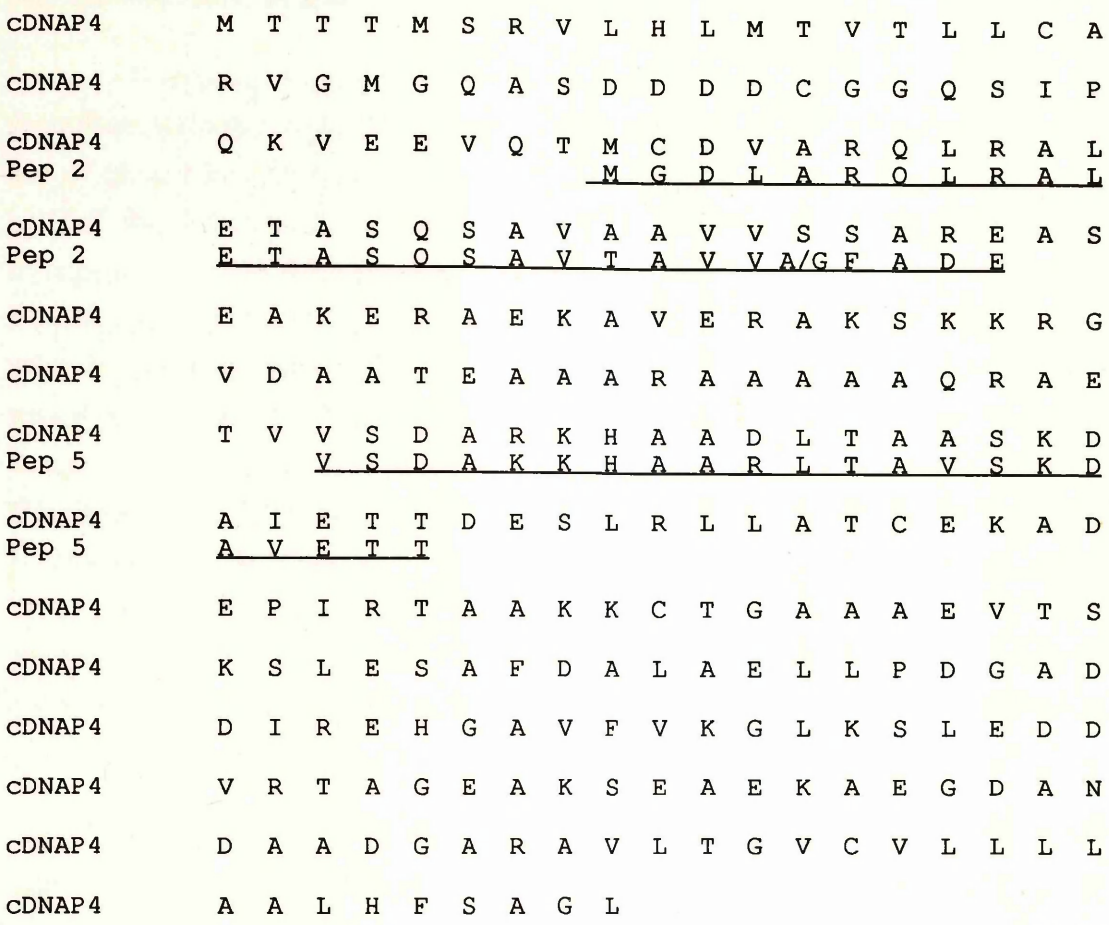


Fig. 5.1 Homology Between the cDNAP4 Open Reading Frame and CNBr Peptides 2 and 5 from a *T.congolense* Procyclic Surface Antigen.

The two CNBr peptide sequences obtained from Dr.T.Pearson (underlined) are shown aligned with the cDNAP4 open reading frame.

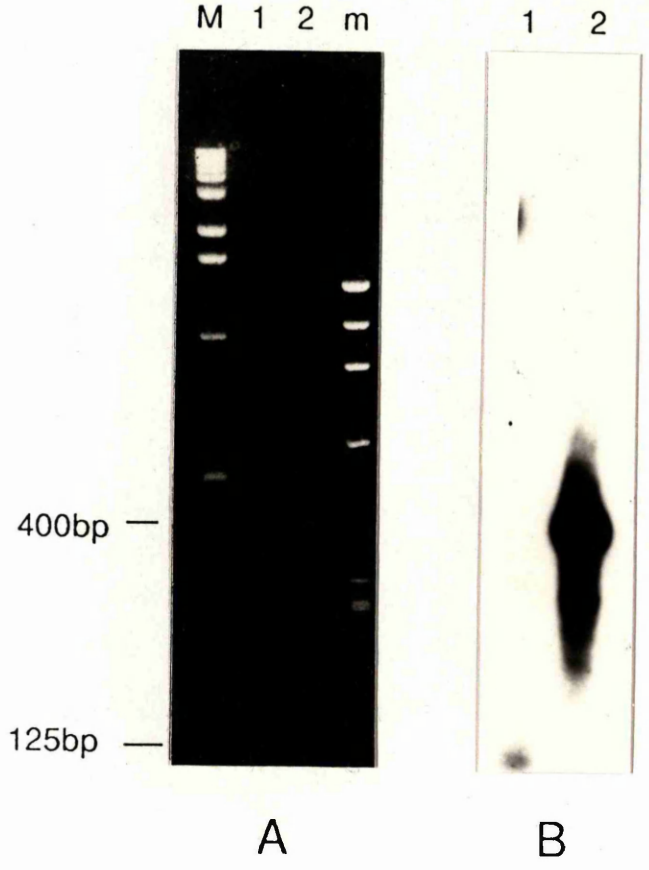


Fig. 5.2 Analysis of the Location of the 5'-End of the cDNAP4 Message Using the Polymerase Chain Reaction.

First strand cDNA was synthesized from total procyclic mRNA using the cDNAP4-specific oligonucleotides 311 and 486. After dilution, a fraction of this cDNA was amplified by the polymerase chain reaction (PCR) using an oligonucleotide (464) containing 20 nucleotides of the trypanosome SL and the cDNAP4-specific primer used for the cDNA synthesis. The products were run on a 1.5% agarose gel, which was stained with EtBr (A), blotted and probed with radiolabelled pλ4.4Eco, washing to a stringency of 0.1X SSC/0.1% SDS at 65°C before autoradiography (B). M - BRL 1kb ladder, m - ΦX174 cut with *Hae* III. Lane 1 - oligo 311-primed PCR product, lane 2 - oligo 486-primed PCR product.

5.2 Identification of the 5' End of The P4 mRNA

All trypanosome mRNAs have a universal spliced leader (SL) sequence at their 5'-ends (Sather & Agabian, 1985). As the absence of any SL sequence at the 5'-end of cloned P4 cDNAs showed a full-length cDNA had not been isolated, the 5'-end of the P4 mRNA was investigated. An oligonucleotide (oligo 311) that hybridises 53bp downstream from the start of cDNAP4 and had been previously used to sequence 5' of the cDNA sequence in a genomic clone of P4 (pλ4.4Eco) was used to prime reverse transcription from procyclic mRNA. After dilution, the resulting first-strand cDNA was amplified by the polymerase chain reaction (PCR) using oligo 311 and an oligonucleotide containing the 3' 20 nucleotides of the trypanosome SL (oligo 464, which has a linker adapter sequence upstream that had previously been designed for generating a PCR amplified cDNA library) as primers. Three rounds of amplification were carried out at the annealing temperature of 47°C followed by 22 rounds at the increased temperature of 53°C. Control reactions containing no cDNA template or only one of the oligonucleotide primers were also carried out.

After amplification, the products were electrophoresed on agarose (Fig. 5.2a, lane 1) and polyacrylamide gels and a single band of approximately 125bp was identified which only appeared when the cDNA template and both oligonucleotide primers were present in the reaction (data not shown). As it did not prove possible either to sequence this PCR product directly or to clone it into pBluescript for sequencing, an alternative approach to verify that it did indeed represent amplification of the 5'-end of the P4 mRNA was to use an oligonucleotide different from oligo 311 which would hybridize to cDNAP4 further downstream and therefore produce a different size of extension/amplification product. The size of the products should agree with the location of the priming oligonucleotide if they represent genuine extensions of P4. Therefore another primer (oligo 486) which had been used to generate sequence from cDNAP4 and which hybridizes 333bp downstream of the cDNA start was used in place of oligo 311 to prime first strand cDNA synthesis and PCR amplification. Again the products were electrophoresed on polyacrylamide and agarose gels (Fig. 5.2a, lane 2) and, as with the first experiment, a single product was obtained only when all components were present in the amplification reaction. This time the product was about 400bp in size. Both of these experiments predict addition of the SL about 20bp upstream of the start of cDNAP4 (each product contains 35 + 17/18bp of primer sequences) and indeed

there is an AG 3'-splice consensus sequence in this region (nts -22 + -21) in the genomic clone pλ4.4Eco (see Fig. 4.32) and the consensus extends to the UUUCPy-(4-40)-AG consensus for 3' splice sites in trypanosomes observed by Laird (1989). There are also short polypyrimidine tracts in the region -40 to -5 with respect to this 3' splice site. These have been shown to be important for the efficiency of *trans* splicing (Huang & Van der Ploeg, 1991b). While this would predict a mRNA slightly smaller than expected from northern blot data it is consistent with the location of the already deduced AUG start codon which must lie within 30-70 nucleotides of the splice acceptor site (Agabian, 1990).

In order to verify that these PCR products were genuine extension products from the P4 mRNA, the agarose gel was blotted onto Hybond N and probed with the genomic subclone pλ4.4Eco, which contains sequences upstream of the cDNAP4 start. The blot was washed to a stringency of 0.1XSSC at 65°C and autoradiography indicated that the 125bp and 400bp bands did hybridize to this probe (Fig. 5.2b). Although the probe contains the oligonucleotide sequences, products containing these sequences alone would not hybridize at the stringency conditions used and so it could be concluded that the bands did represent extension products of the P4 mRNA.

5.3 Ligation of the cDNAP4 Coding Sequence into the pGEX-2T Expression Vector

In order to investigate whether cDNAP4 does represent the gene for the antigen detected by the GUGM mabs, it was necessary to express the cloned gene in order to produce material for raising antisera whose specific reactivity could be compared with those of the mabs. An 850bp *Bam* HI fragment, running from the *Bam* HI site just upstream of the cDNAP4 insert in the pBluescript polylinker to the *Bam* HI site just downstream of the TGA stop codon of cDNAP4, was cut from an agarose gel and purified by centrifuging the gel fragment through glass wool. It was then ligated to dephosphorylated, *Bam* HI cut pGEX-2T, a translational fusion expression vector containing the *Schistosoma japonicum* glutathione-S-transferase (GST) gene (Smith & Johnson, 1988), and transformed into *E.coli* XL1-blue. As there is no blue-white selection with pGEX vectors, 12 colonies were picked and those containing an insert of the correct size were identified by performing *Bam* HI digests on small-scale plasmid preparations (Fig. 5.3a). A single-site ligation would

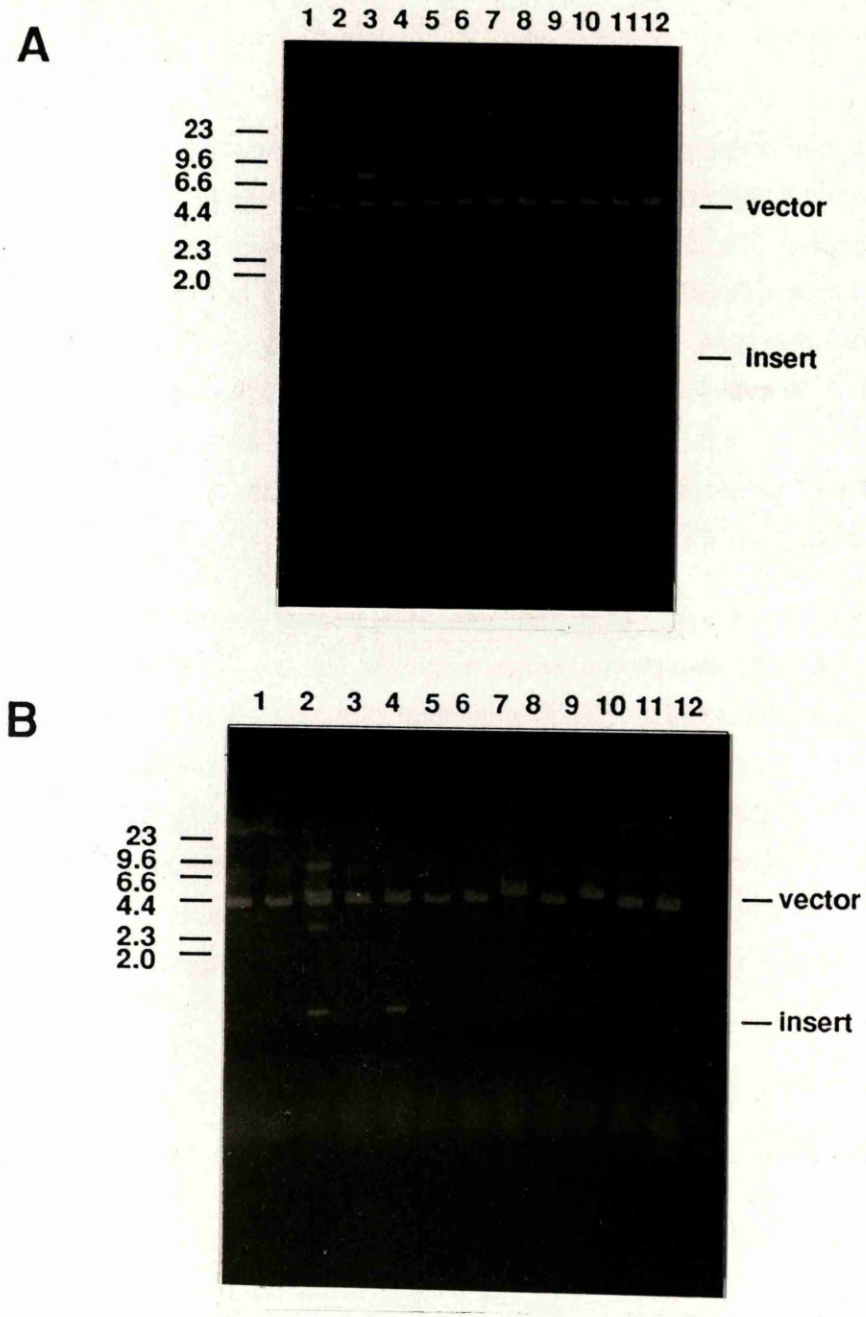
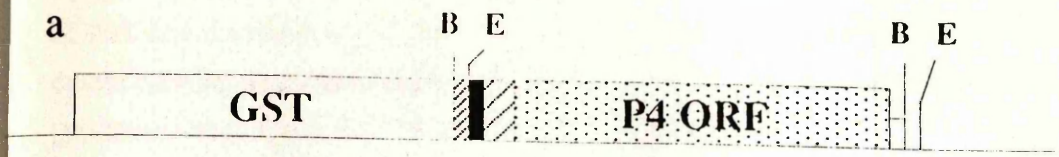


Fig. 5.3 Insertion and Orientation of the cDNAP4 *Bam* HI fragment in pGEX-2T.

Small-scale plasmid DNA preparations of XL1-Blue transformants from a pGEX 2T : cDNAP4 *Bam* HI fragment ligation were digested with *Bam* HI (A) and *Eco* RI (B) and the products analysed on 0.8% agarose gels stained with EtBr. Markers are a *Hin*d III digest of λ DNA. Vector and insert bands are indicated.

a



pBluescript polylinker
Eco RI adapter
P4 5' UTR
GST stop and 3' UTR

b

pGEX-2T pBlue polylinker *Eco* RI adapter cDNAP4

CTG GTT CCG CGT GGA TCC CCC GGG CTG CAG GAA TTC GGC ACG AGC TCG CAC

P4 ORF

TCC TAC TCC AAG CCA GCA AGA AGC GTG AAT ACC ATG

Fig. 5.4 Map of pP4gex.

a) Construction of pP4gex. Large boxed region indicates protein coding sequence.

b) sequence of pP4gex around the *Bam*HI insertion site.

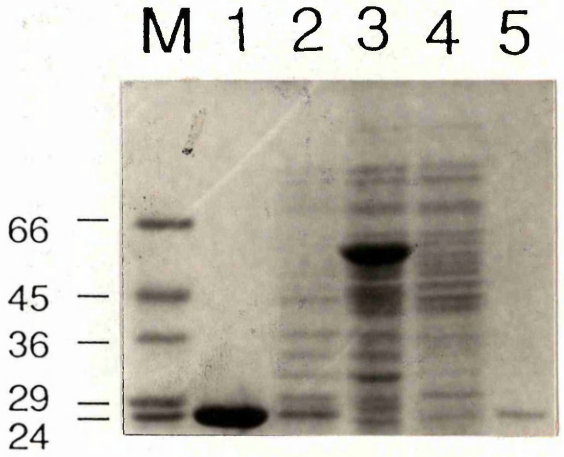


Fig. 5.5 Expression of pP4gex.

The pP4gex construct in XL1-Blue was expressed by induction with IPTG and samples of the pellet (lane 3) and supernatant (lane 4) fractions analysed on a 10% SDS-PAGE gel, stained with Coomassie blue. The remaining supernatant fraction was selected with glutathione beads and a sample of this was also analysed (lane 5). Non-recombinant pGEX 2T expression products were run alongside as a control (lane 1 - supernatant, lane 2 - pellet). M - Sigma Low M.Wt. size markers.

allow the insert to ligate in both orientations so the *Eco* RI sites at the 5'-end of the cDNA and downstream of the *Bam* HI site in pGEX 2T were used to ascertain the orientation of the insertion in each clone. Insertion in the wrong orientation should produce a very short *Eco* RI fragment which would run off the end of the gel while ligation in the correct orientation should produce a fragment of about 850bp. In this way (Fig. 5.3b), clones 3 and 5 were identified as having P4 inserted in the correct orientation and plasmid sequence analysis of the small-scale plasmid preparations, using oligo 311 as primer, helped to ensure that the coding sequence was in-frame with the GST sequence of pGEX (Fig. 5.4b). Clone 5, which has been named pP4gex (Fig. 5.4a), was used in all subsequent analyses.

5.4 Expression of the pP4gex Fusion Protein

A 100ml culture of *E.coli* XL1-blue containing the pP4gex construct was induced to express the GST-P4 fusion protein and the pellet and supernatant fractions were analysed on a 10% SDS-PAGE gel with expressed non-recombinant pGEX-2T as a control. Coomassie staining (Fig. 5.5) indicated that, while GST itself is a soluble protein, present in the supernatant fraction, the pP4gex fusion protein is highly insoluble, being located entirely in the pellet fraction (as indicated by the absence of any fusion protein even in the enriched, glutathione-selected, supernatant fraction (lane 5)). However, the pP4gex fusion protein appears to have been highly expressed, at a level similar to that of GST alone, and is the predicted size (26kDa of GST + 26.5kDa of P4 ORF + 22 amino acids from the *Eco* R1 adapter and polylinkers).

5.5 Raising Antisera Against the pP4gex Fusion Protein

As the pP4gex fusion protein is insoluble, it was not possible to purify it away from other proteins present in the bacterial lysate using glutathione-selection columns as had been hoped. The alternative was therefore to run the pellet fraction of the expression culture on a large SDS-PAGE gel, electroblot this onto nitrocellulose, stain the blot with Ponceau S to identify the position of the fusion protein and excise that strip. After destaining in TBST, the nitrocellulose strip was cut into very fine pieces and macerated in 1ml of PBS by sonicating ten times for 10sec. (Harlow & Lane, 1988).

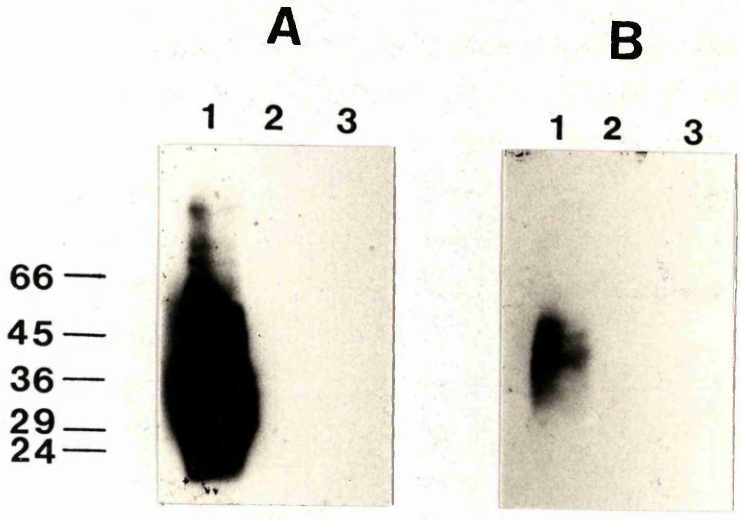


Fig. 5.6 Probing the pP4gex Fusion Protein with GUGM 2.2.

Non-recombinant pGEX 2T (lane 1) and pP4gex (lane 2) expression products were run on a 10% SDS-PAGE gel alongside a CHAPS lysate of 5 X 10^6 procyclic *T.congolense* TREU 1627 (lane 3). After blotting onto two sheets of nitrocellulose and blocking, the products were probed with a 1/500 dilution of GUGM 2.2 ascites fluid followed by a 1/2000 dilution of HRP- anti-mouse IgG (Promega). The filters were developed using the Amersham ECL system and exposed for 1 sec.. Markers were Sigma Low M.Wt. size markers.

(A) Nitrocellulose filter closest to gel
(B) Second filter

Antisera against the pP4gex fusion protein were raised in mice and rats by intraperitoneally injecting an estimated 25ug of the fusion protein as nitrocellulose suspension, emulsified with Complete Freund's Adjuvant. Three mice were inoculated and normal mouse serum was taken from another two mice in the same litter; it was not possible, for ethical reasons, to obtain adequate volumes of pre-immune serum from the injected mice. The rats were controlled in a better manner, by taking pre-immune serum from each rat and inoculating two rats with the pP4gex fusion protein and one with the supernatant fraction from a pGEX 2T expression culture containing about the same amount of GST protein as the fusion protein suspension. After two weeks, the animals were boosted with another 25ug of fusion protein in Incomplete Freund's Adjuvant and left for a further two weeks before collecting serum samples. Because of the low titre of antibody in the rat sera (assessed by titration on western blots), these animals were boosted again at this stage and more serum collected after another 12 days.

5.6 Testing The Monoclonal Antibodies Against the pP4gex Fusion Protein

1ul of the pP4gex pellet was run on a 10% SDS-PAGE gel alongside 1ul of non-recombinant pGEX 2T supernatant control and 2ul (5 X 10^6 trypanosomes) of a CHAPS lysate of *T.congolense* 1627 procyclic cells and electroblotted onto 2 layers of nitrocellulose filters. The blots were stained with Ponceau S to identify the position of the molecular weight markers and then destained in TBST before blocking the filters and probing with a 1:500 dilution of GUGM2.2 mouse ascites fluid. After addition of the HRP-conjugated anti-mouse second antibody, diluted 1:2000, the filters were developed using the Amersham ECL system (Fig. 5.6). The filter closer to the gel contained a diffuse band of hybridization in the trypanosome lysate track at around 29-50kDa. The second filter had a similar pattern although the signal was weaker. Nothing was detected in the pP4gex or pGEX 2T tracks suggesting that they did not possess the epitope detected by GUGM2.2. This would not be surprising if the epitope is carbohydrate as the fusion protein is not glycosylated.

5.7 Immunofluorescence With The Anti-pP4gex Antisera

Acetone-fixed, dried smears of *T.congolense* TREU 1627 procyclic cells

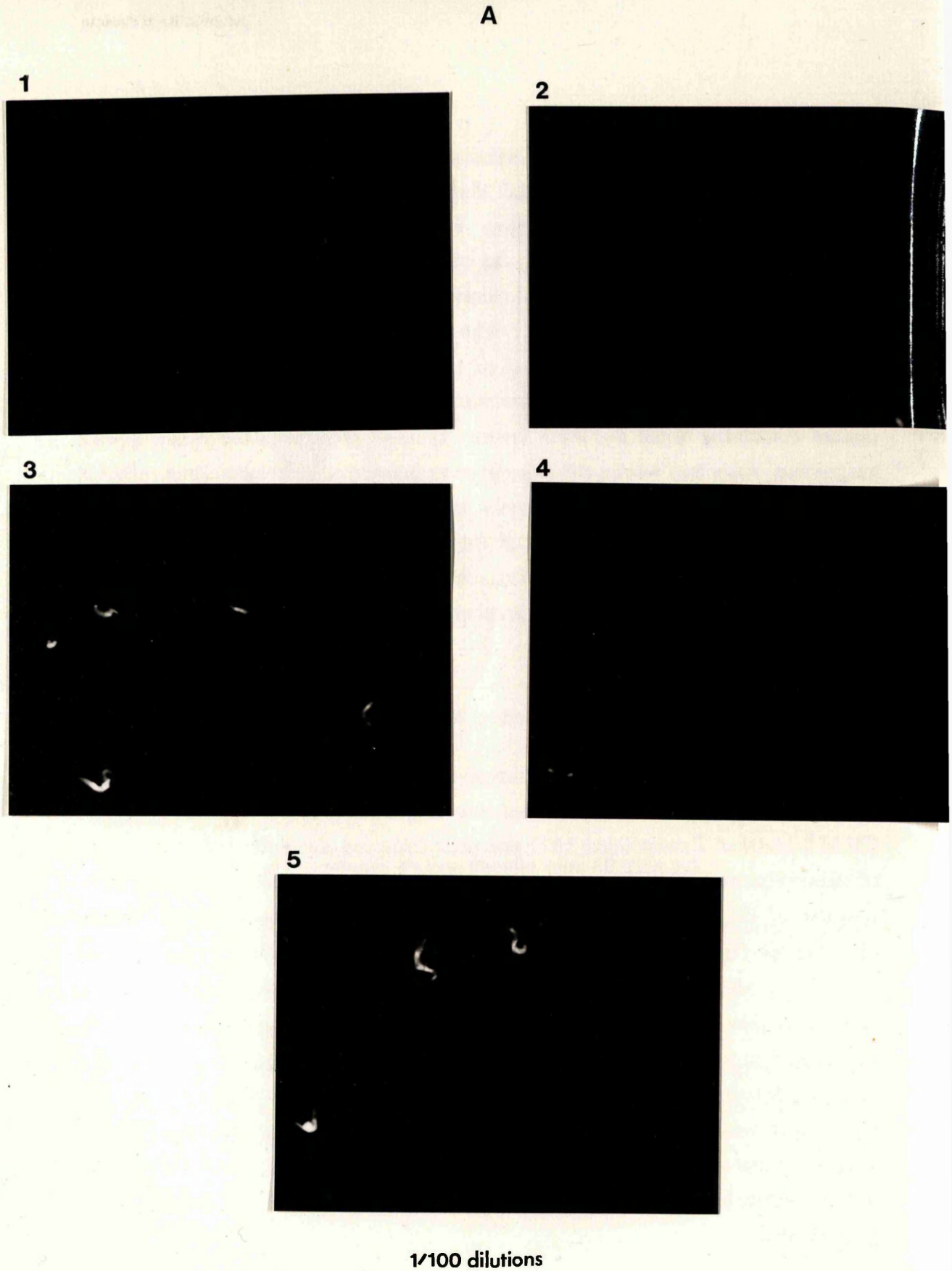
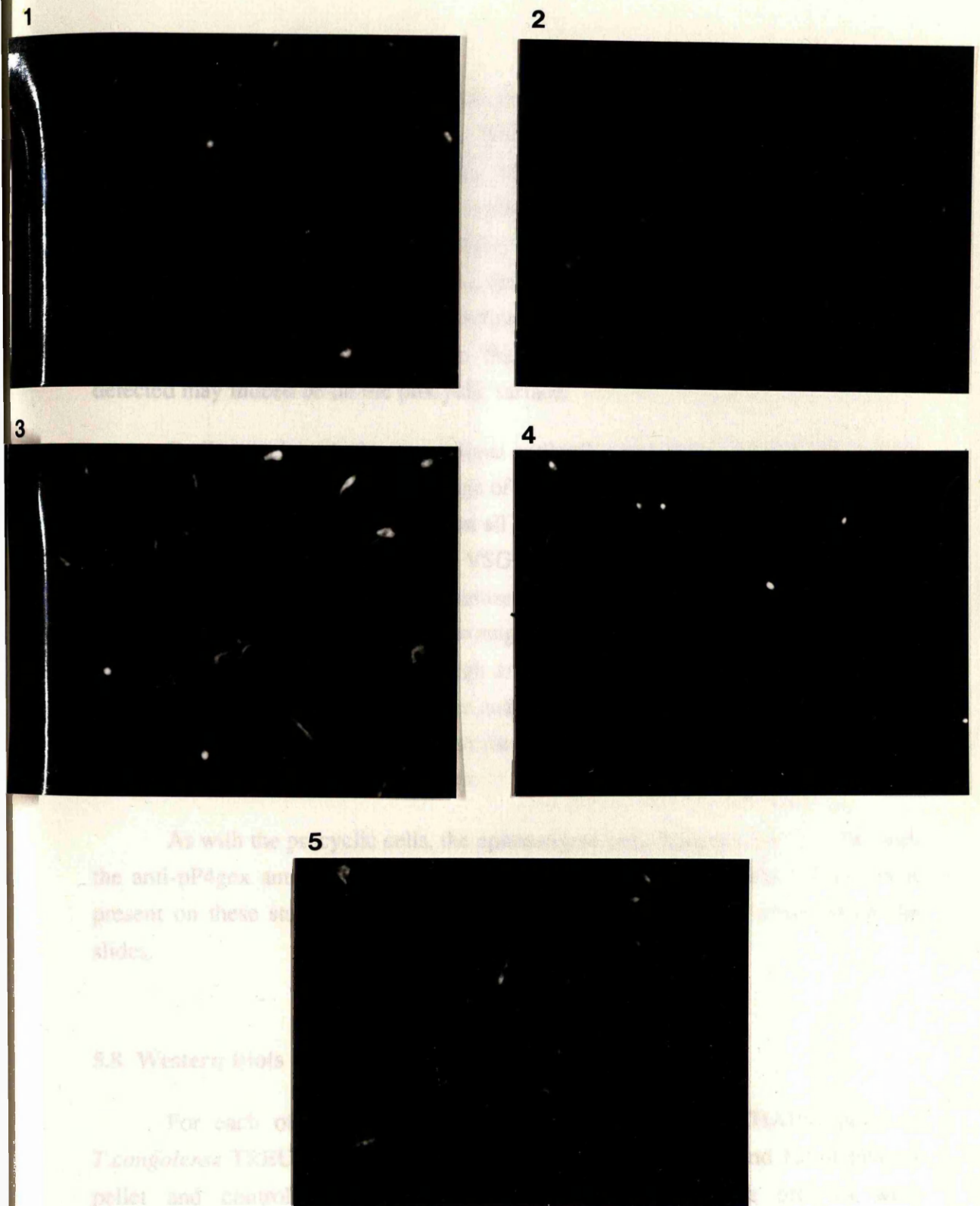


Fig. 5.7 Immunofluorescence with the Polyclonal Antisera.

Procyclic *T.congolense* TREU 1627 (A) and epimastigote *T.congolense* TREU 1457 (B) smears were acetone-fixed on slides and probed with the polyclonal antisera:

B



1 - normal mouse serum, 2 - pre-immune rat serum, 3 - mouse anti-pP4gex, 4 - rat anti-pGEX, 5 - rat anti-pP4gex. After washing, they were labelled with FITC-conjugated anti-mouse or anti-rat antibodies and examined by UV-microscopy. Magnification is 1250X.

were incubated with 1:10 and 1:100 dilutions of the mouse and rat antisera in PBS and, after washing, labelled with FITC-conjugated anti-mouse or anti-rat IgG second antibodies (diluted 1:50 in PBS). The smears were examined by fluorescence microscopy (Fig. 5.7a). While the control sera in each case did not label the trypanosomes to any extent, the antisera raised against the pP4gex fusion protein labelled the entire trypanosome, with a few brighter patches evident on some cells. Acetone-fixing does not require the protein to have a surface disposition to label but the eggshell-like appearance of the fluorescence suggested that the epitope(s) detected may indeed be on the procyclic surface.

Earlier work with the monoclonal antibodies had suggested that the antigen was not restricted to the procyclic stage of the *T.congolense* lifecycle (Lainson *et al*, unpublished) but rather was present on all insect stages, except the metacyclic stage, which has initiated synthesis of the VSG coat. In order to determine whether the protein detected by the anti-pP4gex antisera was also present on other insect stages of the parasite, a flask of epimastigote stage *T.congolense* TREU 1457 was obtained from Dr.C.Ross at CTVM, Edinburgh and the culture supernatant phase used to make smears for immunofluorescence analysis. These cultures were also producing metacyclic cells and it was hoped that these would also be present in the smears to test for their labelling with the antisera.

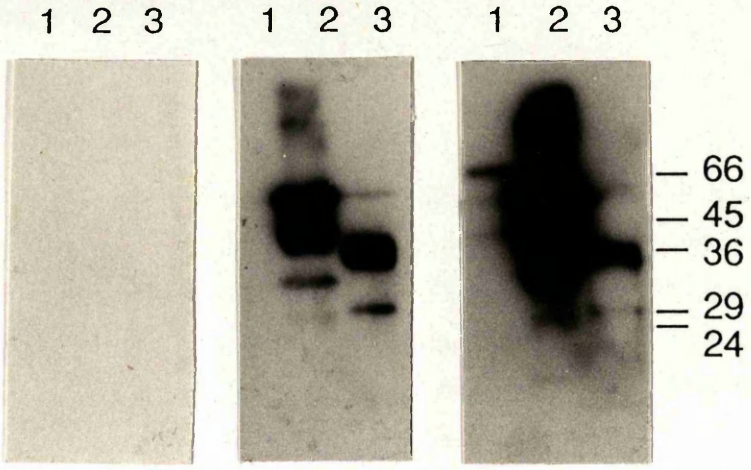
As with the procyclic cells, the epimastigote cells fluoresced (Fig. 5.7b) with the anti-pP4gex antisera but not the control sera, indicating that the P4 protein is present on these stages. Unfortunately, no metacyclic cells were observed on the slides.

5.8 Western Blots

For each of the test and control antisera, 2ul of a CHAPS lysate of *T.congolense* TREU 1627 procyclic cells ($5X10^6$ trypanosomes), and 1ul of pP4gex pellet and control pGEX-2T supernatant (about 1ug fusion protein) were electrophoresed on an SDS-PAGE minigel and electroblotted onto two sheets of nitrocellulose. Each pair of filters was probed with the indicated dilutions of antiserum and second antibody and developed using the Amersham ECL system.

With the mouse antisera (Fig. 5.8a), nothing was detected with normal mouse serum as expected, while the pP4gex fusion protein was detected at the

A


1 2 3
1 2 3
1 2 3
66
45
36
29
24
normal
mouse 1
mouse 2

B

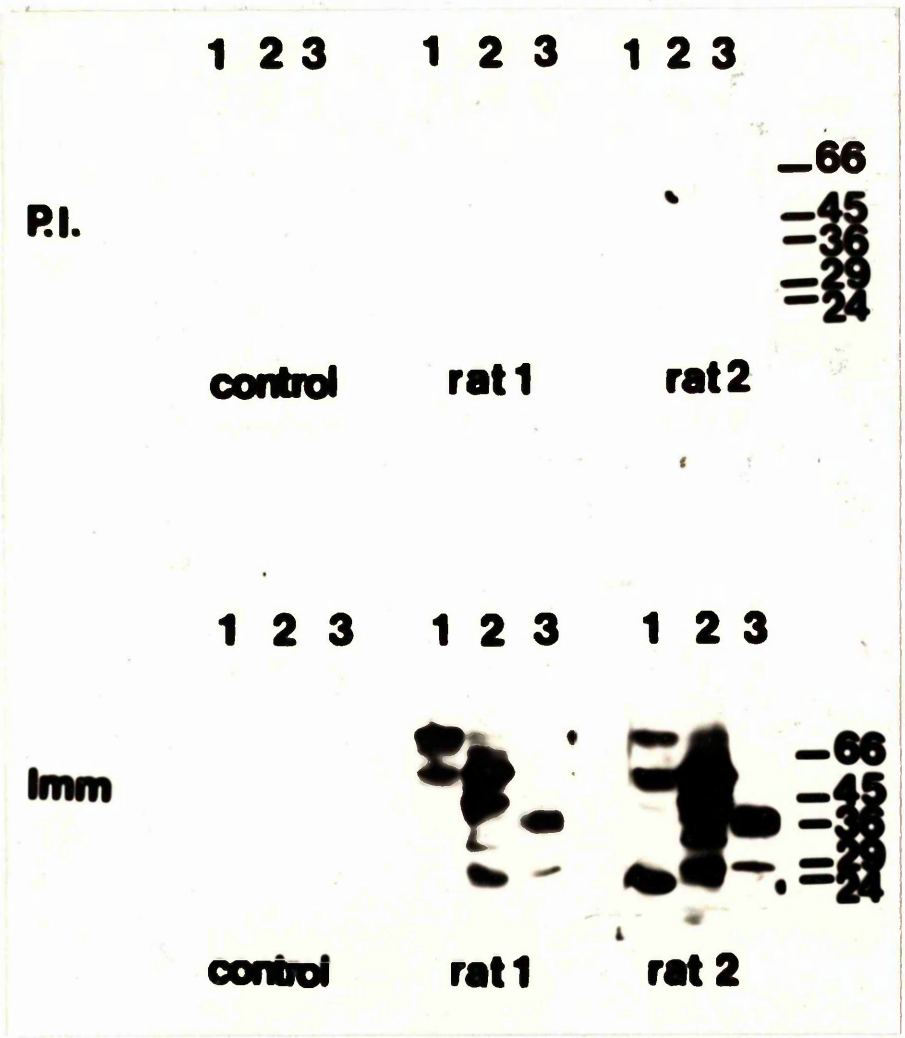

1 2 3
1 2 3
1 2 3
P.I.
66
45
36
29
24
control
rat 1
rat 2
1 2 3
1 2 3
1 2 3
Imm
66
45
36
29
24
control
rat 1
rat 2

Fig. 5.8 Western Blots Probed with the Polyclonal Antisera.

pGEX 2T (lane 1), pP4gex (lane 2) expression products and *T.congolense* TREU 1627 procyclic CHAPS lysate (lane 3) were run on a 10% SDS-PAGE gel, blotted onto nitrocellulose and probed with 1/500 dilutions of the mouse (A) and 1/100 dilutions of the rat (B) polyclonal antisera as indicated. After washing, they were incubated with HRP- anti-mouse IgG (1/2000 dilution) or HRP- anti-rat IgG (1/1000 dilution) and developed using the Amersham ECL system, exposing for 1 sec. (A) or 10 sec. (B). P.I. - pre-immune sera, Imm. - immune sera. Markers were Sigma Low M.Wt. size markers.

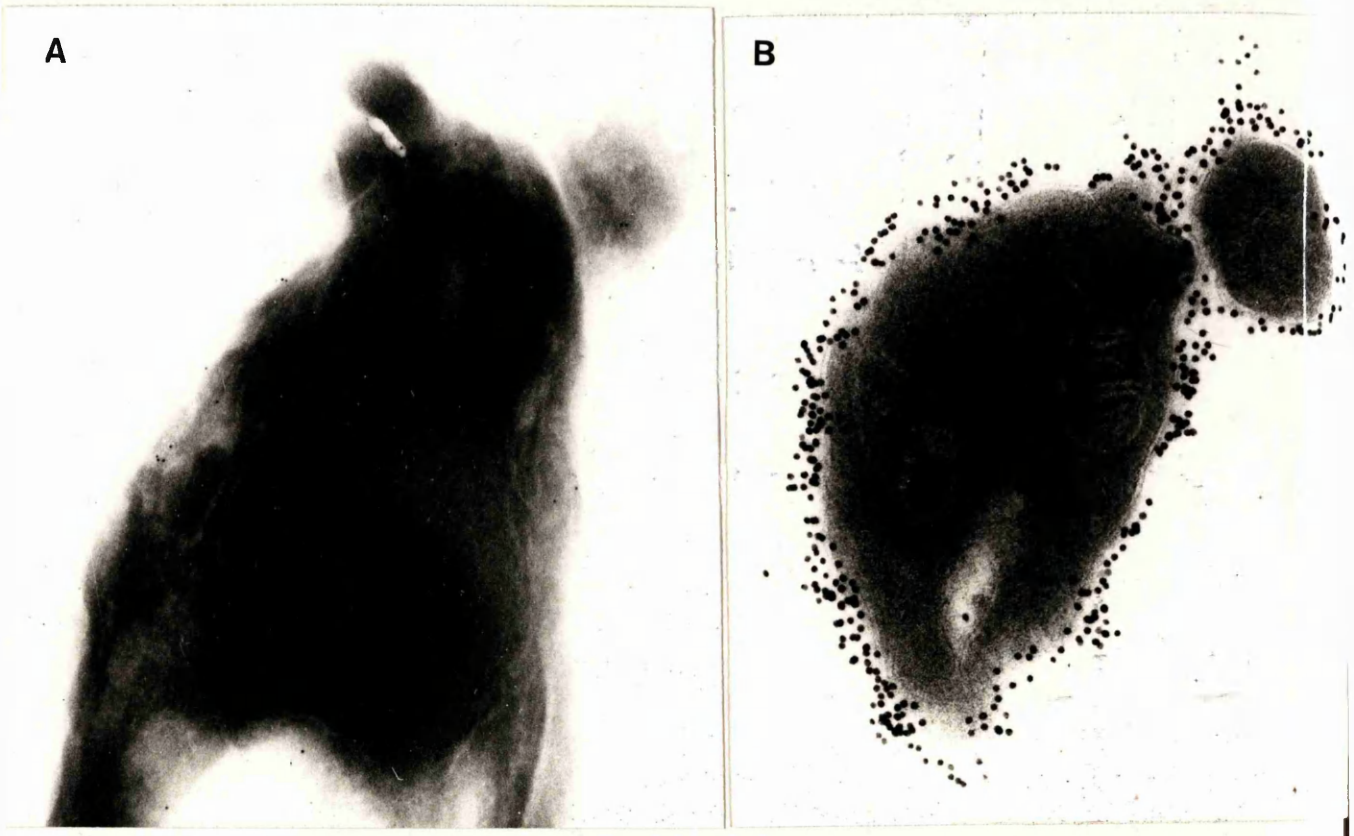


Fig. 5.9 Electron Micrographs of Immuno-gold Labelling of *T.congolense* Procyclic Cells with Rat Anti-pP4gex Antisera.

Live procyclic *T.congolense* TREU 1627 cells were labelled with a 1/10 dilution of the rat pre-immune (a) or anti-pP4gex (b) antisera in PBS containing 1% BSA for 1 hour at 37^{o}C before gently fixing in formaldehyde and adding a 1/10 dilution of goat anti-rat IgG conjugated to 10nm gold particles, under the same conditions for a further hour at 37^{o}C. The cells were then embedded in methyl cellulose/uranyl acetate according to Tokuyasu (1986) and cryosectioned and examined under the transmission electron microscope. Magnification is 38,000X (a) and 96,000X (b).

expected apparent molecular weight (55kDa) with each of the anti-pP4gex antisera. The pGEX-2T GST protein itself did not appear to be labelled with the mouse sera, despite the fact that this protein is a part of the fusion construct, indicating that the P4 segment of the fusion may be immunodominant. Each of the anti-pP4gex antisera also detected a diffuse band of around 36-40kD and a band of about 29kDa in the trypanosome lysates.

The rat antisera gave a very similar result (Fig. 5.8b) although the preimmune sera did detect a few bands in the bacterial expression lysates on longer exposures, but these were not the fusion protein itself and nothing was detected in the trypanosome lysates with preimmune sera. In the case of the rat sera however, while the pGEX-2T GST protein was detected only weakly on a longer exposure by the control anti-GST serum, suggesting it was a poor immunogen, with one of the anti-pP4gex sera (rat 2) it labelled quite strongly although not with the other (rat 1). The pGEX-2T control antiserum however detected nothing in the trypanosome lysate track indicating that it is probably the P4 epitopes which are present in trypanosome lysates and not the pGEX epitopes.

5.9 Immunogold-Labelling and Electron Microscopy

This was performed for us very kindly by Dr.L.Tetley in the EM unit of the Department of Zoology, Glasgow University.

T.congolense TREU 1627 procyclic cells were gently fixed in formaldehyde and incubated with the mouse anti-pP4gex antiserum, with the normal mouse serum as a negative control. After labelling with a Protein A-colloidal gold conjugate and embedding in resin, ultra thin sections were made and examined in the transmission electron microscope. Unfortunately, the labelling was not very efficient but it was clear that the pP4gex epitope was located on the surface of the trypanosomes (data not shown). This experiment has now been repeated with fresh conjugate and the rat antisera, but in order to obtain good labelling (in the absence of a good positive control) it was necessary to use the mildest form of treatment, *i.e.* to label the live cells with the antibody before formaldehyde fixation and then perform cryosectioning (Fig. 5.9). The labelling is clearly specific to the immune serum, is on the parasite's surface and is quite dense, although not to the extent that is found with procyclin.

```
CDNAP4    1  CTCGCACTCCTACTCCAAGCCAGCAAGAAGCGTGAATACCATGACGACAA   50
λ4.4         ----------------G-------------------------------

CDNAP4  360  CGGCTGCGGCCCAGCGCGCGGAGACGGTGGTGAGCGATGCGAGGAAGCAC  409
λ4.4         ----------A-----------------------------------------

CDNAP4  410  GCGGCAGACCTGACGGCGGCGTCGAAGGATGCTATCGAGACGACCGACGA  459
λ4.4         ------------------------------------------A-----

CDNAP4  460  GTCGCTGCGCCTACTGGCCACATGCGAGAAAGCGGACGAGCCCATCCGCA  509
λ4.4         ------------------------------------------------

CDNAP4  510  CTGCTGCAAAA  520
λ4.4         -----------

CDNAP4  558  CCCTTGAGTCAGCGTTCGACGCTCTCGCGGAACTGCTACCGGATGGTGCG  607
λ4.4         ------------------------------------------------

CDNAP4  608  GACGACATCCGCGAGCACGGTGCCGTGTTCGTGAAGGGGCTGAAGTCTCT  657
λ4.4         -----G------------------------------------------

CDNAP4  658  GGAGGATGACGTGCGCACGGCTGGAGAGGCAAAGAGCGAGGCGGAGAAGG  707
λ4.4         ------------------------------------------------
λ4.3/5                    ------------------------------------

CDNAP4  708  CTGAGGGCGATGCGAACGACGCGGCAGATGGTGCCCGTGCCGTGCTGACG  757
λ4.4         ------------------------------------T----------C
λ4.3/5       ------------------------------------T----------

CDNAP4  758  GGCGTGTGCGTGCTGCTGCTTCTGGCTGCACTGCACTTTTCTGCGGGGCT  807
λ4.4         ------------------------------------------------T---
λ4.3/5       ------------------------------------------------T---

CDNAP4  808  GTGAGGCGGGATCCCCCAGCTCACCTGCCCGCGAGACATTATGCGTATTT  857
λ4.4         ------------------
λ4.3/5       ----------------------T---------------C----------

CDNAP4  858  TGAGTGACGAACTTTTATTCGCCCAACCTGGCATAATACCTTATACAGTG  907
λ4.3/5       -------GC--------------------A--------------------

CDNAP4  908  AGCGTTATTTTTATTTCAACTTTCCCCTTC  937
λ4.3/5       -------------------------T---
```

Fig. 5.10 Sequence Comparison of Different Genomic Copies of cDNAP4.

The partial sequence data are aligned with sequence identities marked (-) and sequence differences indicated.

5.10 Comparison of Different cDNAP4 Gene Copies

The question arises as to why the cDNAP4 protein migrates as more than one species in SDS-PAGE. As there are several different copies of the gene it is possible that one or more copies produce each of the different products, although the northern blot data (Fig. 4.10) indicate that the mRNA is a single-sized species. In order to test this hypothesis, partial sequence analysis was performed on subclones generated from each of the genomic clones of cDNAP4 (Fig. 5.10). In each case, the sequence was almost identical to that of the cDNA, with only dispersed single base differences that were either not in the coding region or, at most, led to a conservative amino acid substitution. While the entire coding sequence was not investigated in each case, the restriction maps were identical over this region. The sequences of the subclones, and in some cases the local restriction maps, did however diverge up- and downstream outside the cDNA sequence itself, indicating that different gene copies were being studied. This analysis is not complete as not every gene copy has been cloned but the high degree of sequence conservation between gene copies from two different trypanosome stocks (the cDNA and genomic clones are derived from different trypanosome populations) suggests that it is unlikely that different gene copies encode the different versions of the antigen. It would have been unlikely not to have cloned one copy which differed from the others.

5.11 Protein Structure Analysis

As procyclin is specific to *T.brucei* and the P4 protein (henceforth referred to as ARP [Alanine Rich Protein]) appears from the mab studies to be specific to *T.congolense*, and as both are present at high levels on the surface of insect stages of the parasite, it is possible that the two proteins share an analogous function. This could be structural or functional. If there is a structural role then it might be expected that the two proteins would have similar higher order structures. In order to determine whether or not this is the case, the two protein sequences were subjected to computer-aided secondary structure analysis using the program "PEPTIDESTRUCTURE" and the results expressed in tabular and graphical form using "PLOTSTRUCTURE" (Fig. 5.11).

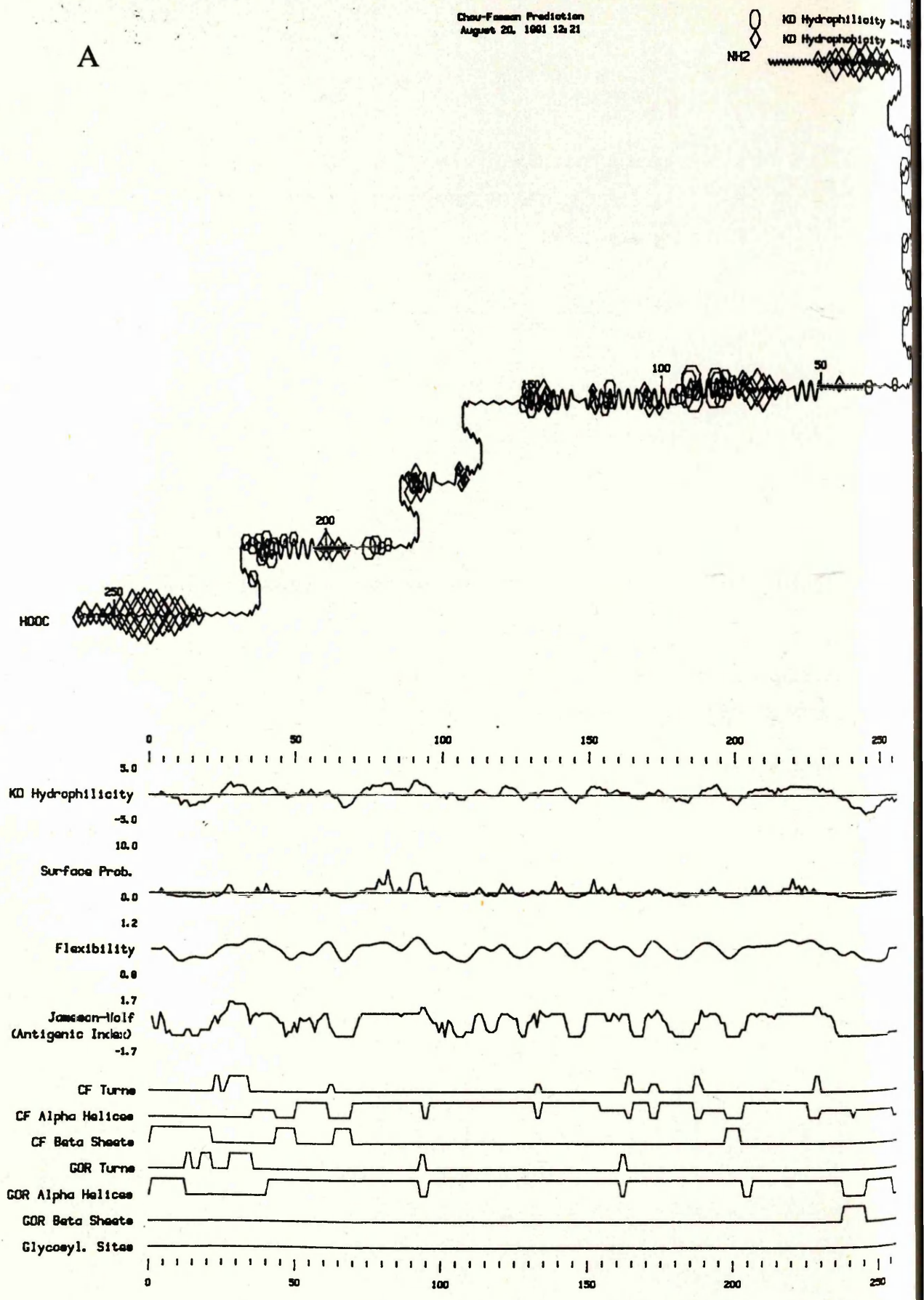
The mature procyclin protein is predicted to have an extended rod-like

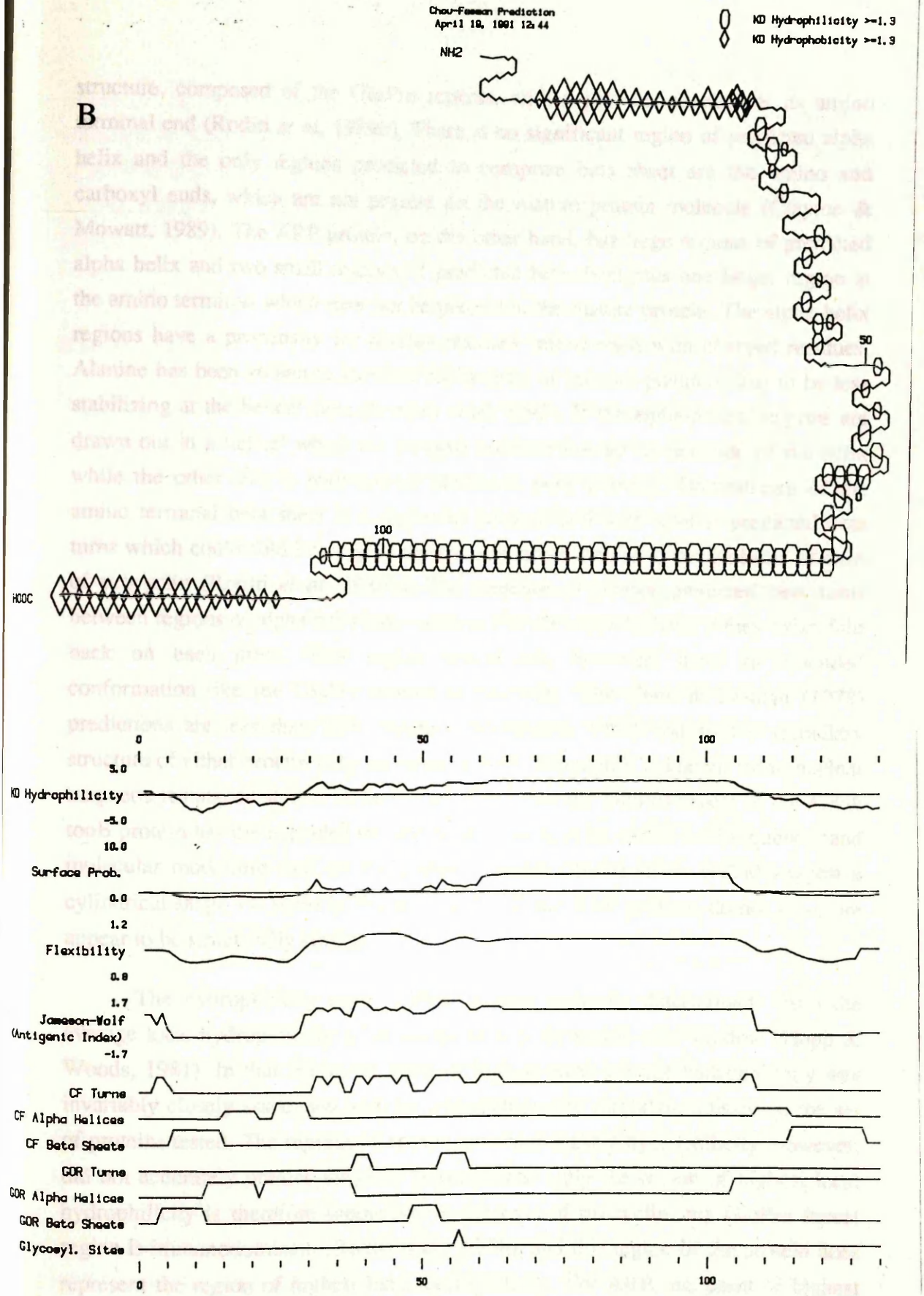
Fig. 5.11 cDNAP4 Protein Structure Analysis and Comparison to Procyclin.

The cDNAP4 and procyclin open reading frames were subjected to computer secondary stucture analysis using the programme "PEPTIDESTRUCTURE" and the results interpreted by the programme "PLOTSTRUCTURE".

(A) cDNAP4

(B) procyclin.


A
Chou-Fasman Prediction
August 20, 1991 12:21
KD Hydrophilicity >=1.5
KD Hydrophobicity >=1.5
NH2
HOOC
50
100
200
250
0
50
100
150
200
250
5.0
KD Hydrophilicity
-5.0
10.0
Surface Prob.
0.0
1.2
Flexibility
0.0
1.7
Jameson-Wolf
(Antigenic Index)
-1.7
CF Turns
CF Alpha Helices
CF Beta Sheets
GOR Turns
GOR Alpha Helices
GOR Beta Sheets
Glycosyl. Sites
0
50
100
150
200
250


Chou-Fasman Prediction
April 18, 1991 12:44
KD Hydrophilicity >=1.3
KD Hydrophobicity >=1.3
B
NH2
50
100
HOOC
0
50
100
5.0
KD Hydrophilicity
-5.0
10.0
Surface Prob.
0.0
1.2
Flexibility
0.8
1.7
Jameson-Wolf
(Antigenic Index)
-1.7
CF Turns
CF Alpha Helices
CF Beta Sheets
GOR Turns
GOR Alpha Helices
GOR Beta Sheets
Glycosyl. Sites
0
50
100

structure, composed of the GluPro repeats, with a globular domain at its amino terminal end (Roditi *et al*, 1989b). There is no significant region of predicted alpha helix and the only regions predicted to comprise beta sheet are the amino and carboxyl ends, which are not present on the mature protein molecule (Clayton & Mowatt, 1989). The ARP protein, on the other hand, has large regions of predicted alpha helix and two small regions of predicted beta sheet plus one larger region at the amino terminus which may not be present in the mature protein. The alpha-helix regions have a propensity for alanine residues interspersed with charged residues. Alanine has been shown to stabilize alpha-helix at internal positions but to be less stabilizing at the helical caps (Serrano *et al*, 1992). If the alpha-helical regions are drawn out in a helical wheel the charged residues line up on one side of the helix while the other side is hydrophobic (D.Barry, pers. comm.). Downstream of the amino terminal beta sheet is a region of random coil with several predicted beta turns which could fold into a globular domain rather similar to the globular domain of procyclin (Roditi *et al*, 1989b). The presence of further predicted beta turns between regions of alpha helix may indicate that the alpha helices, if they exist, fold back on each other. This region would not, however, have an extended conformation like the GluPro repeats in procyclin. The Chou & Fasman (1978) predictions are less than 55% accurate (Nishikawa, 1983) and so the secondary structure of either protein is by no means certain, although it is known from nuclear magnetic resonance studies (Evans *et al*, 1986) that the GluPro repeat in the *E.coli* tonB protein has the extended rod structure intrinsic in its amino acid sequence and molecular modelling data on the procyclin repeat (Roditi *et al*, 1989a) suggest a cylindrical shape 14-18nm in length. The PARP and ARP proteins do not therefore appear to be structurally similar.

The hydrophilicity trace predicts protein antigenic determinants from the average local hydrophilicity of amino acids in a six amino acid window (Hopp & Woods, 1981). In that study, the point of highest local average hydrophilicity was invariably closely associated with an immunologically important epitope in the set of proteins tested. The regions of second and third highest hydrophilicity, however, did not accurately predict antigenic determinants. Only the region of highest local hydrophilicity is therefore predictive. In the case of procyclin, the GluPro repeat region is immunodominant (Roditi *et al*, 1989b), and this region of the protein does represent the region of highest local hydrophilicity. For ARP, the point of highest local hydrophilicity is around amino acid 82. This region lies at the end of an alpha helical domain, is predicted to have a high surface probability on the criteria of

Emini *et al*, (1985), and scores highly in the Jameson-Wolf antigenic index (Jameson & Wolf, 1988). The hydrophilicity trace also indicates that the N-terminus of ARP is relatively hydrophobic indicating that it may represent a signal peptide despite lacking the classical features described by von Heijne (1983, 1985 and 1986).

5.12 Discussion

A cDNA clone (cDNAP4) has been isolated which contains an open reading frame that shares a high degree of sequence homology (75 and 83% respectively) with two peptides from a Kilifi-type *T.congolense* surface antigen (ARP) that has been isolated by Beecroft *et al* (manuscript in preparation). Translation of this open reading frame produces a protein of 256 amino acids which has no *N*-glycosylation signals. It is not possible to predict individual *O*-glycosylation sites (Wilson *et al*, 1991). Although there is no obvious signal peptide at the amino terminus, this region is hydrophobic over a length (22 amino acids) typical for a signal peptide and there is a potential signal for the addition of a glycosyl phosphatidyl inositol tail at the carboxyl end. The protein is rich in alanine residues (21.5%) and acidic.

The high degree of sequence conservation with the Beecroft *et al* antigen indicates that this cDNA probably encodes the homologous protein in the *T.congolense* stock used to construct the library. The large number of similarities between the Beecroft *et al* antigen and the antigen detected by a set of monoclonal antibodies raised against *T.congolense* procyclic culture forms also indicate that these two proteins are probably homologous. By inference therefore, cDNAP4 may represent the gene for the antigen that reacts with the monoclonals. This hypothesis could only be tested indirectly because at least one of the monoclonal antibodies appears to be reactive against a carbohydrate epitope and therefore will not react with fusion proteins induced from clones containing the cDNAP4 coding sequence. Polyclonal antisera raised in mice and rats against the cDNAP4 fusion protein from pP4gex detected a diffuse band of 36-40kDa and a sharper band of around 29kDa in western blots of a *T.congolense* procyclic CHAPS lysate which is enriched for surface proteins. The monoclonal antibodies detect bands in similar positions in such lysates (E.Kilbride, 1992 and this study). In addition, antigen enriched by anion exchange chromatography from CHAPS lysates or purified by organic extraction and hydrophobic interaction chromatography is reactive in western blots

with both the monoclonal and polyclonal antibodies (E.Kilbride, 1992). In IFA or immuno-electron microscopy however, the anti-pP4gex antisera did not label trypanosomes as strongly as did the mabs. One explanation for this might be that the polyclonal antisera are against protein epitopes while at least one of the mabs is against a carbohydrate epitope. While the protein epitopes will be readily accessible in western blots of trypanosome lysates, they may be partially masked on the trypanosome surface by carbohydrate moieties.

Electron microscopy of immuno-gold labelled procyclic cells confirmed that the antigen is located on the surface of procyclic *T.congolense*(L. Tetley & E. Kilbride, unpublished). The polyclonal antisera against the cDNAP4 fusion protein were also tested in such an assay and again the labelling appeared to be on the exposed surface. The protein encoded by cDNAP4 must therefore either be inserted in the membrane or peripherally attached to it by ionic bonds. As with several other major surface antigens in trypanosomes (the VSG and procyclin), this is probably achieved via the post-translational addition of a glycosyl phosphatidyl inositol tail since a potential glypiation signal is present at the carboxyl terminus of the ARP protein. How the protein reaches the membrane, however, is not clear as a signal peptide is not obvious. In other proteins linked to the membrane in this way, such signals are present and indeed, protein translocation appears to be a prerequisite for GPI tail addition (Caras, 1991). Either the hydrophobic amino terminus is a genuine signal peptide or else some novel form of modification must label the ARP protein for translocation to the plasma membrane.

Having the sequence of only one cDNA clone, it is possible that the membrane form of the protein is encoded by another copy of the gene which does have the signals required for passage to the membrane. Indeed, Southern blot analysis indicates that there are several copies of the cDNAP4 sequence present in the genome. However, partial sequence analysis could detect no significant differences between several of these copies isolated from a genomic library. As not every copy was sequenced, this possibility still remains, but it appears a less likely explanation. In addition, the single PCR product from each of the two oligonucleotide primers suggests a single 5' sequence.

Why might there be more than one ARP band in SDS-PAGE? Beecroft *et al* (manuscript in preparation) performed amino acid composition analysis on both of the diffusely staining regions which were detected by their mabs. Although they shared acidity and richness in alanine residues they were not identical. Clearly, the

fact that they share epitopes suggests that they must be related but they need not be encoded by the same gene. Since the partial sequence analysis of genomic copies of cDNAP4 indicates that it is unlikely that different copies encode different proteins, then the two protein bands must either be encoded by widely divergent gene sequences or else share the same gene, with one band representing a posttranslational modification of the other. One band may represent a cleavage product associated with insertion into the membrane as cleavage would alter the amino acid composition. Beecroft *et al* (manuscript in preparation) did observe large quantities of water-soluble antigen in their ELISA tests as well as material that was solubilized only after detergent extraction. Watersoluble and membrane-bound forms of procyclin also exist (Richardson *et al*, 1988; Clayton & Mowatt, 1989). The presence of large amounts of partially- or unmodified protein suggests that there is a high turnover of the antigen on the surface so that large quantities of the protein are continuously being synthesised. Alternatively, the unmodified, free form may have a separate function of its own.

What function might the ARP protein play on the procyclic surface? The northern blot results, representation of cDNAP4 in the cDNA library and strength of signal in western blots indicate that the gene is highly expressed and the protein present at relatively high levels, representing a major surface antigen. The behaviour of the protein and non-reactivity of the mabs with bacterially-expressed fusion protein suggest that ARP is heavily modified in some way, possibly by *O*-linked glycosylation. Despite the ARP sequence being highly conserved between highly divergent stocks of *T.congolense*, it appears to be subgenus specific - DNA probes and the mabs do not cross-react significantly with *T.brucei*. Beecroft *et al* (manuscript in preparation) found that their mabs reacted with procyclic cells from another member of the *Nannomonas* subgenus, *T.simiae*. By analogy, all members of the *Trypanozoon* subgenus are reactive with anti-procyclin antibodies. Since ARP is specific to *T.congolense* and procyclin specific to *T.brucei*, and since both proteins are major surface antigens, highly acidic and immunodominant relative to other procyclic antigens, could they perform a similar function? The secondary structure predictions for ARP and PARP do not immediately indicate any structural similarities, except for a possible globular domain near the amino terminus. However, it took very careful study of VSGs before their structural similarities became evident. In the VSG, very little conservation is obvious at the level of primary structure (only the positions of cysteine residues), there is slightly more similarity at the secondary level but it is only at the level of tertiary and quaternary

structure that the extensive structural conservation is obvious (Cohen *et al*, 1984; Freymann *et al*, 1984; Gurnett *et al*, 1986; Metcalf *et al*, 1986; Jahnig *et al*, 1987). It has been proposed (Roditi *et al*, 1989a; Beecroft *et al*, manuscript in preparation) that procyclin might be involved in tropism, promoting differentiation of epimastigote *T.brucei* in the tsetse salivary glands. Since *T.congolense* does not develop in the salivary glands it would not require this protein but may require another - ARP? - for its own differentiation in the fly mouthparts (Roditi & Pearson, 1990). This proposal however is purely speculative as there is no evidence to support the hypothesis. Why would procyclin be expressed so early after fly ingestion (immediately VSG expression is switched off) if it is not required until later on? A much more plausible explanation is that the two proteins serve a purely protective role.

The tsetse fly gut is probably a very harsh environment, containing many proteolytic activities which could prove harmful to the procyclic trypanosome. Without a protective coat, essential proteins on the trypanosome surface could be destroyed. An alternative hypothesis is therefore that procyclin and ARP may serve to protect these proteins (Richardson *et al*, 1988; Roditi *et al*, 1989a). In support of this, ARP appears to be insensitive to many proteolytic enzymes (E.Kilbride, 1992), with only chemical agents such as CNBr able to cleave it. Since the protein sequence indicates that there are potential cleavage sites for several of these enzymes, they are presumably rendered inaccessible either by protein folding or carbohydrate attachment. With a purely protective role, the actual structure of the protein would be of little importance, and so the two trypanosome species could have recruited different gene products for this function. Only the *Nannomonas* and *Trypanozoon* subgenera have to pass through the fly gut during their development. Protection from proteolytic and other trypanocidal activities is therefore likely to have been a relatively recent acquisition in insect stages of the parasite and if this occurred subsequent to the divergence of the two subgenera then it is not surprising that different proteins would be recruited for this purpose. Functional constraints would however require that certain criteria be met and this would explain why the two proteins have many similar properties despite their obvious differences. Alternatively, one species might have rejected its original protection in favour of a better one, more suited to its lifestyle. Perhaps the procyclin-like sequences observed in *T.congolense* are remnants or precursors of a procyclin gene.

Apart from sequence and structural differences, one major difference

between procyclin and ARP is that the ARP mRNA is constitutively expressed, being present in the bloodstream as well as procyclic stages, while the procyclin message is not present at high levels in steady state bloodstream form RNA. It has not been conclusively proved that ARP protein is not expressed in bloodstream forms but the monoclonal antibodies do not label these stages in ELISA or immunoblots (Beecroft *et al*, manuscript in preparation; Lainson *et al*, unpublished) and the protein seems to appear with similar kinetics to procyclin on transformation of bloodstream forms to procyclics (T.Pearson, pers. comm.). If indeed it is not present, then control of expression must be posttranscriptional, either at the level of translation or beyond.

CHAPTER 6

CONCLUDING REMARKS

This thesis has taken several approaches to identify the gene for the procyclic surface antigen detected by a set of monoclonal antibodies raised against procyclic *T.congolense*. Molecules on the surface of insect stages of the parasite are likely to be involved with its interaction with its tsetse fly vector and by studying such molecules it may be possible to look at the basis of the trypanosome's survival in the potentially harsh environment of the fly's midgut.

The ARP gene is highly expressed at the RNA level and the indications are that this is extended to the protein as well. As a major, dominant surface antigen it represents the alternative surface coat and in this context has a fascinating structure, apparently entirely different from that of procyclin in *T.brucei*. No homologue for the highly expressed procyclin was found in *T.congolense*. While procyclin was known to have subgenus-specific epitopes, this came as somewhat of a surprise as the two surface antigens share several biochemical and immunological properties and there are sequences similar to parts of the procyclin gene present in the *T.congolense* genome.

Although the gene for ARP was identified by differential screening of a cDNA library, it does not appear to be transcribed in a stage-specific manner. Immunofluorescence analysis with the monoclonal antibodies suggests that the protein is expressed at least at a much higher level in insect stages compared to bloodstream forms and that therefore expression of ARP must be subject to control at the level of translation or beyond. In *T.brucei* the procyclin gene is also transcibed in both procyclic and bloodstream stages but in bloodstream forms the mRNA appears to be very unstable and is only present at very low levels in steady state RNA. The control of the genes for these two major surface antigens therefore appears to be very different. Thus the observation of Roditi *et al* (1987) and Mowatt & Clayton (1987) that the procyclin cDNA was the only highly expressed stage-specific molecule in their libraries may explain why no stage-specific molecules (with the possible exception of cDNAP1) were identified in this study.

The differential screen identified several interesting sequences although none, except perhaps cDNAP1, is genuinely transcribed in a stage-specific manner. The discovery of a ribosomal protein gene which might confer cycloheximide resistance could prove very useful as a dominant selectable marker.

There are several copies of the ARP gene within the *T.congolense* genome and their sequences appear to be highly conserved both between loci and between

the two stocks used to construct the libraries. This high degree of sequence conservation is in contrast to the greater degree of divergence of the ARP proteins between the Kilifi-type stock studied by Beecroft *et al* (manuscript in preparation) and the stocks studied here and is probably yet another example of the likelihood that the Kilifi stocks represent a different species of trypanosome.

Analysis and comparison of procyclin and ARP may provide insight into the interaction of the parasite and its tsetse fly vector, an area which has to date concentrated on the side of the tsetse fly. The presence of a lectin which is antagonistic to the early development of trypanosome infections in the fly but at a later stage actually promotes development of the infection (Maudlin & Welburn, 1988) is intriguing. As both procyclin and ARP are potentially heavily glycosylated, they might interact with this lectin in some way. Recently, the GPI anchor of procyclin has been analysed (Ferguson *et al*, unpublished). It has a highly unusual glycosidic side-chain composed of nine galactose residues, nine *N*-acetyl glucosamine residues and five sialic acid residues and is probably branched (M. Ferguson, pers.comm.). The presence of sialic acid is highly unusual for trypanosomes but it is possible that it has been sequestered from the serum during growth *in vitro*. A transialidase activity has recently been identified in *Trypanosoma cruzi* (Parodi *et al*, 1992). Whatever the structure of the side-chain finally proves to be, it must be a very large molecule which would cover a large area of the trypanosome surface and may take up the space between the procyclin protein moieties. If the GPI tail attached to ARP also has such a large side chain then it may explain the inaccessibility of the protein to proteases.

Future Work

The results of this work immediately suggest several points for further investigation.

1/ In order to test the ARP antigen for its possible use in a transmission-blocking vaccine, the pP4gex fusion protein will be used to immunize rabbits. Blood from these animals can then be taken to feed flies which will be analysed for the development of *T.congolense* infections.

2/ Both procyclin and the VSGs are transcibed from multicistronic expression sites by a polymerase which is highly resistant to α-amanitin. Since ARP is also a major

surface antigen and may share its function with procyclin, it is not unreasonable to predict that it too might be transcribed by an α-amanitin-resistant polymerase. Nuclear run-on analysis on procyclic *T.congolense* in the presence and absence of α-amanitin, using as substrate the P4 genomic clones, should provide the answer to this question and also give some indication as to whether the promoter for ARP transcription lies just upstream of the gene, as is the case for procyclin. The presence of genes equivalent to PAGs (procyclin associated genes; Koenig-Martin *et al*, 1992) or ESAGs in the ARP transcription unit should also be investigated and the presence of downstream sequences in the P4 genomic clones should prove useful in this context. It is possible that the cDNAP1 sequence identified in this study lies downstream of ARP in the same transcription unit. Rudenko & Van der Ploeg (1989) suggested that transcription of telomeric sequences in bloodstream *T.brucei* may be due to continuation of transcription downstream of the VSG genes as the transcription was resistant to α-amanitin. However, telomere transcription was also resistant to this drug in procyclic cells and the location of such sequences within ARP or PARP transcription units could explain this observation. There is no indication that PARP genes are located near telomeres as they are not sensitive to *Bal*-31 (Roditi *et al*, 1987) but the PARP gene lies at the start of its transcription unit and not the end like the VSG so that the digestion would have to be much more extensive before any effect was observed.

3/ Now that stable transformation of trypanosomatids by homologous recombination seems feasible (Cruz & Beverley, 1990; ten Asbroek *et al*, 1990; Eid & Sollner-Webb, 1991), it might be possible to investigate the function of procyclin and ARP. If all copies of the ARP or procyclin genes could be disrupted then the absolute requirement of these proteins could be tested. A more complex experiment would be to analyse effects on tropism after transfecting the *T.brucei* gene into *T.congolense* or *vice versa*, having deleted or disrupted the native genes. The presence of multiple copies of each of these genes would however make such experiments very difficult to perform.

4/ Roditi *et al* (unpublished) have analysed the role of the conserved motif in the 3' UTR of procyclin. The sequence was mutated and cloned downstream of the chloramphenicol acetyl transferase (CAT) gene under the control of the procyclin promoter. The construct was then transfected into procyclic *T.brucei* and its effect on CAT expression analysed by comparison to the wildtype sequence. Initial results have indicated that there is a reduction in CAT expression in procyclics when this

motif is mutated but the nature of the mutation may have some influence on this result (I. Roditi, pers.comm.). In contrast, S.Graham (unpublished) in our laboratory has deleted the conserved motif from the 3' UTR of procyclin in transfection constructs which have the CAT gene under the control of the procyclin promoter and analysed the effect compared to the wildtype sequence in transfected *T.brucei* procyclics. Additional experiments removing a larger segment of the 3' UTR, placing an oligonucleotide comprising the conserved 16bp downstream or replacing the procyclin 3' sequence with that of ARP have also been tested in such an assay. The results of these experiments indicate that the motif has no effect on CAT expression. Further experiments will be necessary to resolve these contradicting results and stable transformation in the relevant gene locus might be more useful.

5/ A cycloheximide resistance gene would be extremely useful as a dominant selectable marker for genetic studies which have until now depended on iso-enzyme and RFLP (restriction fragment length polymorphism) markers. Since sexual recombination in trypanosomes is not obligatory and appears to be a rare event (Sternberg *et al*, 1988; Turner *et al*, 1990), the ability to select for novel combinations of genetic markers has obvious advantages. The sequence of cDNAPE suggests that it may encode a cycloheximide resistant L29 ribosomal protein but unfortunately the stock of trypanosomes from which it was cloned is no longer available to test. Since the work for this thesis was completed, PCR amplification of first strand cDNA from the M15 1/148 stock has been performed (S.Loker, unpublished) using an oligonucleotide internal to cDNAPE and the SL oligonucleotide as primers. The resulting product, which was the expected size for cDNAPE, was then cloned in pBluescript. Sequence analysis of this clone indicates that it is essentially identical to cDNAPE (S.Loker, pers.comm.) and that it too encodes a glutamic acid at position 37 in the amino acid sequence. Experiments are now underway to test the sensitivity of M15 1/148 procyclic cells to cycloheximide and the cDNA will also be tested for its ability to confer cycloheximide resistance when transformed into yeast.

REFERENCES

Abraham J.M., Feagin J.E. and Stuart K. (1988) Characterization of cytochrome c oxidase III transcripts that are edited only in the 3' region. Cell, 55, 267-272.

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

Auffrey C. and Rougeon F. (1980) Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumour RNA. Eur. J. Biochem., 107, 303-314.

Bakalara N., Simpson A.M. and Simpson L. (1989) The Leishmania kinetoplast-mitochondrion contains terminal uridylyltransferase and RNA ligase activities. J. Biol. Chem., 264, 18679-18686.

Baker R.D., Maudlin I., Milligan P.M.J., Molyneux D.H. and Welburn S.C. (1990) The possible role of Rickettsia-like organisms in trypanosomiasis epidemiology. Parasitol., 100, 209-217.

Banks K.L. (1978) Binding of *Trypanosoma congolense* to the walls of small blood vessels. J. Protozool., 25, 241-245.

Banks K.L. (1979) *In vitro* binding of *Trypanosoma congolense* to erythrocytes. J.Protozool., 26, 103-108.

Barry J.D. (1977) Studies on surface antigens of *Trypanosoma brucei.* PhD. Thesis, University of Glasgow.

Barry J.D. (1986) Surface antigens of African trypanosomes in the tsetse fly. Parasitol. Today, 2, 143-145.

Barry J.D. (1989) African trypanosomes: an elusive target. In New Strategies in Parasitology, McAdam K.P.W.J. ed., Churchill Livinston, London.

Barry J.D., Crowe J.S. and Vickerman K. (1983) Instability of the *Trypanosoma brucei rhodesiense* metacyclic variable antigen repertoire. Nature, 306, 699-701.

Beecroft R.P., Roditi, I and Pearson, T.W. (1992/3) Identification and characterisation of a procyclin-like molecule from procyclic culture forms of *Trypanosoma congolense.* Manuscript in preparation.

Belhumeur P., Paterno G.D., Boileau G., Claverie J.-M. and Skup D. (1987). Isolation and characterization of a murine cDNA clone highly homologous to the yeast L29 ribosomal protein gene. Nucl. Acids Res., 15, 1019-1029.

Bellofatto V. and Cross G.A.M. (1989) Expression of a bacterial gene in a trypanosomatid protozoan. Science, 244, 1167-1169.

Ben-Amar M., Pays A., Tebabi P., Dero B., Seebeck T., Steinert M. and Pays E. (1988) Structure and transcription of the actin gene of *Trypanosoma brucei.* Mol.

Benne R., Van Den Borg J., Brakenhoff J., Sloof P., Van Boom J. and Tromp M. (1986) The major transcript of the frameshifted coxII gene from trypanosome mitochondria contain four nucleotides that are not encoded in the DNA. Cell, 46, 819-826.

Bernards A., Van der Ploeg L.H.T., Liu A.Y.C., Borst P., Boothroyd J.C.. Coleman S. and Cross G.A.M. (1981) Activation of trypanosome variant surface glycoprotein genes involves a duplication-transposition leading to an altered 3' end. Cell, 27, 497-505.

Bernards A., Michels P.A.M., Lincke C.R. and Borst P. (1983) Growth of chromosome ends in multiplying trypanosomes. Nature, 303, 592-597.

Bhat G.J., Jayarama G., Koslowsky D.J., Feagin J.E., Smiley B.L. and Stuart K. (1990) An extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6. Cell, 61, 885-894.

Birnboim H.C. and Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res., 7, 1515-1523.

Blackburn E.H. and Challoner P.B. (1984) Identification of a telomeric rDNA sequence in *Trypanosoma brucei*. Cell, 36, 447-457.

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

Blum B., Sturm N.R., Simpson A.M. and Simpson L. (1991) Chimeric gRNA Messenger RNA Molecules with Oligo(U) Tails Covalently Linked at Sites of RNA Editing Suggest That U-Addition Occurs by Transesterification. Cell, 65, 543-550.

Boothroyd J.C., Paynter C.A., Cross G.A.M., Bernards A. and Borst P. (1981) Variant surface glycoproteins of *Trypanosoma brucei* are synthesised with cleavable hydrophobic sequences at the carboxy and amino terminal. Nucl. Acids Res., 9, 4743-4745.

Boothroyd J.C. and Cross G.A.M. (1982) Transcripts coding for variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5' end. Gene, 20, 281-289.

Borst P. (1986) Discontinuous transcription and antigenic variation in trypanosomes. Ann. Rev. Biochem., 55, 701-732.

Borst P., Van der Ploeg M., Van Hoek J.F.M., Tas J. and James J. (1982) On the DNA content and ploidy of trypanosomes. Mol. Biochem. Parasitol., 6, 13-23.

Borst P., Bernards A., Van der Ploeg L.H.T., Michels P.A.M., Liu A.Y.C., De Lange T. and Sloof P. (1984) Gene rearrangements controlling the expression of surface antigen genes in trypanosomes. In Molecular Biology and Host Parasite Interactions, Agabian,N. and Eisen H., eds., pp205-217.

Borst P., Fase-Fowler F. and Gibson W.C. (1987) Kinetoplast DNA of *Trypanosoma evansi*. Mol. Biochem. Parasitol., 23, 31-38.

Brewer G. and Ross J. (1988) Poly(A) shortening and degradation of the 3' A + U-rich sequences of human c-myc mRNA in a cell-free system. Mol. Cell.Biol., 8, 1697-1708.

Brun R. (1982) Cultivation of procyclic trypomastigotes of *Trypanosoma congolense* in a semi-defined medium with direct adaptation of bloodstream forms. Z. Parasitenkd., 67, 129-135.

Bullock W.O., Fernandez J.M. and Short J.M. (1987) XL1-Blue: a high efficiency plasmid transforming *rec*A *Escherichia coli* strain with beta-galactosidase selection. Biotechniques, 5, 376-378.

Bulow R., Griffiths G., Webster P., York-Dieter S., Opperdoes, F. and Overath P. (1989) Intracellular localization of the glycosylphosphatidylinositol-specific phospholipase C of *Trypanosoma brucei*. J. Cell Sci., 93, 233-240.

Campbell D.A., Thornton D.A. and Boothroyd J.C. (1984) Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. Nature, 311, 350-355.

Campbell D., Spithill T., Samaras T., Simpson A. and Simpson L. (1989) Apparent uridine addition in the poly A tail of *Leishmania major* ND1 transcript. Mol. Biochem. Parasitol., 36, 197-199.

Caras I.W. (1991) An Internally Positioned Signal Can Direct Attachment of a Glycophospholipid Membrane Anchor. J. Cell Biol., 113, 77-85.

Cardoso de Almeida M.L. and Turner M.J. (1983) The membrane form of the variant surface glycoprotein of *Trypanosoma brucei*. Nature, 302, 349-352.

Carrington M., Walters D. and Webb H. (1991) The Biology of the Glycosylphosphatidylinositol-Specific Phospholipase-C of *Trypanosoma brucei*. Cell Biol. Int. Rep., 15, 1101-1114.

Cech T.R. (1991) RNA editing - world's smallest introns. Cell, 64, 667-669.

Chambon P. (1975) Eukaryotic RNA Polymerases. Ann. Rev. Biochem., 44, 613-638.

Chen K. and Donelson J. (1980) The sequence of two kinetoplast DNA minicircles of *Trypanosoma brucei*. Proc. Natl. Acad. Sci. USA, 77, 2445-2449.

Cheng J.-F., Smitch C.L. and Cantor C.R. (1991) Structural and transcriptional analysis of a human subtelomeric repeat. Nucl. Acids Res., 19, 149-154.

Chou P.Y. and Fasman G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol., 47, 45-148.

Clayton C.E. (1985) Structure and regulated expression of genes encoding fructose bisphosphate aldolase in *Trypanosoma brucei*. EMBO J., 4, 2997-3003.

Clayton C.E. and Mowatt M.R. (1989) The procyclic acidic repetitive proteins of *Trypanosoma brucei*. Purification and post-translational modification. J. Biol. Chem., 264, 15088-15093.

Clayton C.E., Fueri J.P., Itzhaki J.E., Bellofatto V., Sherman D.R., Wisdom G.S., Vijayasarathy S. and Mowatt M.R. (1990) Transcription of the procyclic acidic repetitive protein genes of *Trypanosoma brucei*. Mol. Cell. Biol., 10, 3036-3047.

Cohen C., Reinhardt B., Parry D.A., Roelants G.E., Hirsch W. and Kanawe B. (1984) Alpha-helical coiled-coil structures of *Trypanosoma brucei* variable surface glycoproteins. Nature, 311, 169-171.

Cordingley J.S. and Turner M.J. (1980) 6.5S RNA: preliminary characterization of small RNAs in *Trypanosoma brucei*. Mol. Biochem. Parasitol., 1, 91-96.

Cornelissen A.W.C.A., Verspieren M.P., Toulme J.J., Swinkels B.W. and Borst P. (1986) The common 5' terminal sequence on trypanosome mRNAs: a target for anti-messenger oligodeoxynucleotides. Nucl. Acids Res., 14, 5605-5614.

Cross G.A.M. (1975) Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. Parasitol., 71, 393-417.

Cross G.A.M. and Johnson J.G. (1976) Structure and organisation of the variant-specific surface antigens of *Trypanosoma brucei*. In Biochemistry of Parasites and Host Parasite Relationships, Van den Bossche, H. ed., Elsevier, Amsterdam, pp413-420.

Crowe J.S., Barry J.D., Luckins A.G., Ross C.A. and Vickerman K. (1983) All metacyclic variable antigen types of *Trypanosoma congolense* identified using monoclonal antibodies. Nature, 306, 389-391.

Cruz A. and Beverley S.M. (1990) Gene replacement in parasitic protozoa. Nature, 348, 171-173.

Cully D.F., Ip H.S. and Cross G.A.M. (1985) Coordinate transcription of variant surface glycoprotein genes and an expression site associated gene family in *Trypanosoma brucei*. Cell, 42, 173-182.

Decker C.J. and Sollner-Webb B. (1990) RNA editing involves indiscriminate U changes throughout precisely defined editing domains. Cell, 61, 1001-1011.

De Lange T., Liu T.A.Y.C., Van der Ploeg L.H.T., Borst P., Tromp M.C., and Van Boom J.H. (1983) Tandem repetition of the 5' mini-exon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription? Cell, 34, 891-900.

De Lange T., Michels P.A.M., Veerman H.J.G., Cornelissen A.W.C.A. and Borst P. (1984) Many trypanosome messenger RNAs share a common 5' terminal sequence. Nucl. Acids Res., 12, 3777-3790.

Devereux J., Haeberli P. and Smithies O. (1984) A comprehensive set of sequence analysis programmes for the VAX. Nucl. Acids Res., 12, 387-395.

Donelson J.E. and Rice-Ficht A.C. (1985) Molecular biology of trypanosome antigenic variation. Microbiol. Rev., 49, 107-125.

Dorfman D.M., Lenardo M.J., Reddy L.V., Van der Ploeg L.H.T. and Donelson J.E. (1985) The 5.8S ribosomal RNA gene of *Trypanosoma brucei*: structural and transcriptional studies. Nucl. Acids Res., 13, 3533-3549.

Dorn P.L., Aman R.A. and Boothroyd J.C. (1991) Inhibition of protein synthesis results in superinduction of procyclin (PARP) mRNA levels. Mol. Biochem. Parasitol., 44, 133-139.

Earnshaw D.L., Beebee T.J.C. and Gutteridge W.E. (1987) Demonstration of RNA polymerase multiplicity in *Trypanosoma brucei*. Biochem. J., 241, 649-655.

Ehlers B., Czichos J. and Overath P. (1987) Repression and reactivation of the variant surface glycoprotein gene in *Trypanosoma brucei*. FEBS Lett., 225, 53-58.

Eid J. and Sollner-Webb B. (1991) Stable integrative transformation of *Trypanosoma brucei* that occurs exclusively by homologous recombination. Proc. Natl. Acad. Sci. USA, 88, 2118-2121.

Emini E.A., Hughes J.V., Perlow D.S. and Boger J. (1985) Induction of Hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J. Virol., 55, 836-839.

Englund P.T., Hajduk S.L. and Marin J.C. (1982) The molecular biology of trypanosomes. Ann. Rev. Biochem., 51, 695-726.

Esser K.M. and Schoenbechler M.J. (1985) Expression of two variant surface glycoproteins on individual African trypanosomes during switching. Science, 229, 190-193.

Evans J.S., Levine B.A., Trayer I.P., Dorman C.J. and Higgins C.F. (1986) Sequence-imposed structural constraints in the *Ton*B protein of *E.coli*. FEBS Lett., 208, 211-216.

Evers R., Hammer A., Kock W., Jess W., Borst P., Memet S. and Cornelissen A.W.C.A. (1989). *Trypanosoma brucei* contains two RNA polymerase II largest subunit genes with an altered C-terminal domain. Cell, 56, 585-597.

Fairlamb A.H., Weislogel P.O., Hoeijmakers J.H.J. and Borst P. (1978) Isolation and characterization of kinetoplast DNA from bloodstream form of *Trypanosoma brucei*. J. Cell Biol., 76, 293-309.

Feagin J.E., Jasmer D.P. and Stuart K. (1987) Developmentally regulated addition of nucleotides within apocytochrome b transcripts in *Trypanosoma brucei*. Cell, 49, 337-345.

Feagin J.E. and Stuart K. (1988) Developmental aspects of uridine addition within mitochondrial transcripts of *Trypanosoma brucei*. Mol. Cell. Biol., 8, 1259-1265.

Feagin J.E., Abraham J.M. and Stuart K. (1988a) Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. Cell, 53, 413-422.

Feagin J.E., Shaw J.M., Simpson L. and Stuart K. (1988b) Creation of AUG initiation codons by addition of uridines within cytochrome b transcripts of kinetoplastids. Proc. Natl. acad. Sci. USA, 85, 539-543.

Ferguson M.A.J. and Cross G.A.M. (1984) Myristylation of the membrane form of a *Trypanosoma brucei* variant surface glycoprotein. J. Biol. Chem., 259, 3011-3015.

Ferguson M.A.J., Low M.G. and Cross G.A.M. (1985) Glycosyn 1-sn-1,2-dimyristylphosphatidylinositol is covalently linked to *Trypanosoma brucei* surface glycoprotein. J. Biol. Chem., 260, 14545-14555.

Ferguson M.A.J. and Williams A.F. (1988) Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. Ann. Rev. Biochem., 57, 285-320.

Ferguson , M.A.J., Masterson W.J., Homans S.W. and Mcconville M.J. (1991) Evolutionary Aspects of GPI Metabolism in Kinetoplastid Parasites. Cell Biol. Int. Rep., 15, 991-1005.

Ferrante A. and Allison A.C. (1983a) Alternative pathway activation of complement by African trypanosomes lacking a glycoprotein coat. Parasite Immunol., 5, 491-498.

Ferrante A. and Allison A.C. (1983b) Natural agglutinins to African trypanosomes. Parasite Immunol., 5, 539-546.

Field M.C., Menon A.K., and Cross G.A.M. (1991a) Developmental Variation of Glycosylphosphatidylinositol Membrane Anchors in *Trypanosoma brucei* - Identification of a Candidate Biosynthetic Precursor of the Glycosylphosphatidylinositol Anchor of the Major Procyclic Stage Surface Glycoprotein. J. Biol. Chem., 266, 8392-8400.

Field M.C., Menon A.K., and Cross G.A.M. (1991b) A Glycosyl-phosphatidylinositol Protein Anchor from Procyclic Stage *Trypanosoma brucei* - Lipid Structure and Biosynthesis. EMBO J., 10, 2731-2739.

Fox J.A., Duszenko M., Ferguson M.A.J., Low M.G. and Cross G.A.M. (1986) Purification and characterisation of a novel glycan- phosphatidylinositol-specific phospholipase C from *Trypanosoma brucei*. J. Biol. Chem., 261, 15767-15771.

Frasch A.C.C., Bernards A., Van der Ploeg L.H.T., Borst P, Hoeikmakers J.H.J., Van den Burg J. and Cross G.A.M. (1980) The genes for the variable surface glycoproteins of *Trypanosoma brucei*. In The Biochemistry of Parasites and Host-Parasite Relationship: The Host-Invader Interplay. Van den Bossche H., ed., pp. 235-239, Elsevier/North-Holland, Amsterdam.

Freymann D.M., Metcalf P., Turner M.J. and Wiley D. (1984) The 6A resolution X-ray structure of a variable surface glycoprotein from *Trypanosoma brucei*. Nature, 311, 167-169.

Frischauf A.M., Lehrach H., Poustka A. and Murray N. (1983) Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol., 170, 827-842.

Gardiner P.R., Finerty J.R. and Dwyer D.M. (1983) Iodination and identification of surface membrane antigens in procyclic *Trypanosoma rhodesiense*. J. Immunol., 131, 454-457.

Gibson W.C., Osinga K.A., Michels P.A.M. and Borst P. (1985) Trypanosomes of subgenus *Trypanozoon* are diploid for housekeeping genes. Mol. Biochem. Parasitol., 16, 231-242.

Gibson W.C. and Borst P. (1986) Size-fractionation of the small chromosomes of *Trypanozoon* and *Nannomonas* trypanosomes by pulsed field gradient gel electrophoresis. Mol. Biochem. Parasitol., 18, 127-140.

Glass D.J., Polvere R.I., and Van der Ploeg L.H.T. (1986) Conserved sequences and transcription of the *hsp*70 gene family in *Trypanosoma brucei*. Mol. Cell. Biol., 6, 4657-4666.

Gottesdiener K., Garcia-Anoveros J., Lee M.G.-S. and Van der Ploeg L.H.T. (1990) Chromosome organization of the protozoan *Trypanosoma brucei*. Mol. Cell. Biol., 10, 6079-6083.

Greener A. (1990) *E.coli* SURE™ Strain: clone "unclonable" DNA. Strategies, 3, 5-6.

Grondal E.J.M., Evers R., Kosubek K. and Cornelissen A.W.C.A. (1989) Characterization of the RNA polymerases of *Trypanosoma brucei*: Trypanosomal mRNAs are composed of transcripts derived from both RNA polymerases II and III. EMBO J., 8, 3383-3389.

Gurnett A.M., Ward J., Raper J. and Turner M.J. (1986) Purification and characterization of membrane-form variant surface glycoproteins of *Trypanosoma brucei*. Mol. Biochem. Parasitol., 20, 1-13.

Guyaux M., Cornelissen A.W.C.A., Pays E., Steinert M. and Borst P. (1985) *Trypanosoma brucei*: a surface antigen mRNA is discontinuously transcribed from two distinct chromosomes. EMBO J., 4, 995-998.

Hajduk S.L., Cameron C.R., Barry J.D. and Vickerman K. (1981) Antigenic variation in cyclically transmitted *Trypanosoma brucei*. Variable antigen type composition of metacyclic trypanosome populations from the salivary glands of *Glossina morsitans*. Parasitol., 83, 595-607.

Harlow E. and Lane D. (1988) Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory. U.S.A.

Harris M.E. and Hajduk S.L. (1992) Kinetoplastid RNA editing: *in vitro* formation of cytochrome b gRNA-mRNA chimeras from synthetic substrate RNAs. Cell, 68, 1091-1099.

Hawking F. (1975) Circadian and other rhythms of parasites. Adv. Parasitol., 13, 123-182.

Hereld D., Krakow J.L., Bangs J.D., Hart G.W. and Englund P.T. (1986) A phospholipase C from *Trypanosoma brucei* which selectively cleaves the glycolipid of the variant surface glycoprotein. J. Biol. Chem., 261, 13813-13819.

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

Hoeijmakers J.H.J., Frasch A.C.C., Bernards A., Borst P. and Cross G.A.M. (1980) Novel expression linked copies of the genes for variant surface antigens in trypanosomes. Nature, 284, 78-80.

Hopp T.P. and Woods K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. acad. Sci. USA, 78, 3824-3828.

Hsiao K.-C. (1991) A fast and simple procedure for sequencing double stranded DNA with Sequenase. Nucl. Acids Res., 19, 2787.

Huang J., and Van der Ploeg L.H.T. (1991a) Maturation of Polycistronic Pre-mRNA in *Trypanosoma brucei*: Analysis of *trans* Splicing and Poly(A) Addition at Nascent RNA Transcripts from the *hsp*70 Locus. Mol. Cell. Biol., 11, 3180-3190.

Huang J., and Van der Ploeg L.H.T. (1991b) Requirement of a Polypyrimidine Tract for *trans* Splicing in Trypanosomes - Discriminating the PARP Promoter from the Immediately Adjacent 3' Splice Acceptor Site. EMBO J., 10, 3877-3885.

Huynh T.V., Young R.A. and Davis R.W. (1985) Constructing and screening cDNA libraries in lambda gt10 and lambda gt11. In DNA cloning: A practical approach., D.M. Glover, ed., vol.1 pp49-78, IRL Press, Oxford.

Imboden M.A., Laird P.W., Affolter M. and Seebeck T. (1987) Transcription of the intergenic regions of the tubulin gene cluster of *Trypanosoma brucei*: evidence for a polycistronic transcription unit in a eukaryote. Nucl. Acids Res., 15, 7357-7368.

Jackson P.R., Honigberg B.M. and Holt S.C. (1978) Lectin analysis of *Trypanosoma congolense* bloodstream trypomastigote and culture procyclic surface saccharides by agglutination and electron microscopic techniques. J. Protozool, 25, 471-477.

Jahnig F., Bulow R., Baltz T. and Overath P. (1987) Secondary structure of the variant surface glycoproteins of trypanosomes. FEBS Lett., 221, 37-42.

Jameson B.A. and Wolf H. (1988) The antigenic index: a novel algorithm for predicting antigenic determinants. Comput. Appl. Biosc., 4, 181-186.

Jefferies D., Tebabi P. and Pays E. (1991) Transient Activity Assays of the *Trypanosoma brucei* Variant Surface Glycoprotein Gene Promoter - Control of Gene Expression at the Posttranscriptional Level. Mol. Cell. Biol., 11, 338-343.

Jenni L., Marti S., Schweizer J., Betschart B., Le Page R.W.F., Wells J.M., Tait A., Paindavoine P., Pays E. and Steinert M. (1986) Hybrid formation between African trypanosomes during cyclical transmission. Nature, 322, 173-175.

Johnson P.J., Kooter J.M. and Borst P. (1987) Inactivation of transcription by UV irradiation of *Trypanosoma brucei* provides evidence for a multicistronic transcription unit that includes a variant surface glycoprotein gene. Cell, 51, 273-281.

Kapler G.M., Coburn C.M. and Beverley S.M. (1990) Stable transfection of the human parasite *Leishmania major* delineates a 30-kilobase region sufficient for extrachromosomal replication and expression. Mol. Cell. Biol., 10, 1084-1094.

Karn J., Matthes H.W.D., Gait M.J. and Brenner S. (1980) A new selective phage cloning vector, lambda2001, with sites for *Xba* I, *Bam* HI, *Hin*d III, *Eco* RI, *Sst* I and *Xho* I. Gene, 32, 217-224.

Kaufer N.F., Fried H.M., Schwindinger W.F., Jasin M. and Warner J.R. (1983) Cycloheximide resistance in yeast: the gene and its protein. Nucl. Acids Res., 11, 3123-3135.

Kilbride E. (1992) A surface antigen from *Trypanosoma congolense.* MSc. Thesis, University of Glasgow.

Knowles G., Betschart B., Kukla B.A., Scott J.R. and Majiwa P.A.O. (1988) Genetically discrete populations of *Trypanosoma congolense* from livestock on the Kenyan coast. Parasitol., 96, 461-474.

Koenig E., Delius H., Carrington M., Williams R.O. and Roditi I. (1989) Duplication and transcription of procyclin genes in *Trypanosoma brucei.* Nucl. Acids Res., 17, 8727-8739.

Koenig-Martin E., Yamage M. and Roditi I. (1992) *Trypanosoma brucei*: a procyclin-associated gene encodes a polypeptide related to ESAG 6 and 7 proteins. Mol. Biochem. Parasitol., in press.

Kooter J.M. and Borst P. (1984) Alpha-amanitin insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. Nucl. Acids Res., 12, 9457-9472.

Kooter J.M., De Lange T., and Borst P. (1984) Discontinuous synthesis of mRNAs in trypanosomes. EMBO J., 3, 2387-2392.

Kooter J.M., van der Spek H.J., Wagter R., d'Oliveira C.E., van der Hoeven F., Johnson P.J. and Borst P. (1987) The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in *Trypanosoma brucei.* Cell, 51, 261-272.

Koslowsky D.J., Bhat G.J., Perrollez A.L., Feagin J.E. and Stuart K. (1990) The MURF 3 Gene of *Trypanosoma brucei* contains multiple domains of extensive editing and is homologous to a subunit of NADH dehydrogenase. Cell, 62, 901-911.

Krause M. and Hirsch D. (1987) A *trans*-spliced leader sequence on actin mRNA in *C.elegans.* Cell, 49, 753-761.

Kreader C.A. and Heckman J.E. (1987) Isolation and characterization of a *Neurospora crassa* ribosomal protein gene homologous to CYH2 of yeast. Nucl. Acids Res., 15, 9027-9042.

Laban A. and Wirth D.F. (1989) Transfection of *Leishmania enriettii* and expression of chloramphenicol acetyltransferase gene. Proc. Natl. Acad. Sci. USA, 86, 9119-9123.

Landfear S.M., McMahon-Pratt D. and Wirth D.F. (1983) Tandem arrangement of tubulin genes in the protozoan parasite *Leishmania enriettii*. Mol. Cell. Biol., 3, 1070-1076.

Lanham S.M. and Godfrey D.G. (1970) Isolation of salivarian trypanosomes from man and other animals on DEAE cellulose. Exp. Parasitol., 28, 521-532.

Laird P.W. (1989) *Trans* splicing in trypanosomes - archaism or adaptation? Trends Genet., 5, 204-208.

Laird P.W., Kooter J.M., Loosbroek N. and Borst P. (1985) Mature mRNAs of *Trypanosoma brucei* possess a 5' cap acquired by discontinuous RNA synthesis. Nucl. Acids Res., 13, 4253-4266.

Laurent M., Pays E., Van Meirvenne N., Matthyssens G., Williams R.O. and Steinert M. (1983) DNA rearrangements linked to expression of a predominant surface antigen gene of trypanosomes. Nature, 302, 263-266.

Lee M.G.-S. and Van der Ploeg L.H.T. (1990a) Homologous recombination and stable transfection in the parasitic protozoan *Trypanosoma brucei*. Science, 250, 1583-1587.

Lee M.G.-S., and Van der Ploeg L.H.T. (1990b) Transcription of the heat shock 70 locus in *Trypanosoma brucei*. Mol. Biochem. Parasitol., 41, 221-232.

Lenardo M.J., Rice-Ficht A.C., Kelly G., Esser K.M. and Donelson J.E. (1984) Characterization of the genes specifying two metacyclic variable antigen types in *Trypanosoma brucei rhodesiense*. Proc. Natl. Acad. Sci. USA, 81, 6642-6646.

Levine B.J., Chodchoy N., Marzluff W.F. and Skoultchi A.I. (1987) Coupling of replication type histone mRNA levels to DNA synthesis requires the stem-loop sequence at the 3'-end of the mRNA. Proc. Natl. Acad. Sci. USA, 84, 6189-6193.

Lumsden W.H.R., Herbert W.J. and McNeillage G.J.C. (1973) Techniques with Trypanosomes. Churchill Livingstone, Edinburgh and London.

Majiwa P.A.O., Masake R.A., Nantulya V.M., Hamers R. and Matthyssens G. (1985) *Trypanosoma* (*Nannomonas*) *congolense*: identification of two karyotypic groups. EMBO J., 4, 3307-3313.

Majiwa P.A.O., Young J.R., Hamers R. and Matthyssens G. (1986a) Minichromosomal variable surface glycoprotein genes and molecular karyotypes of *Trypanosoma* (*Nannomonas*) *congolense*. Gene, 1510, 183-192.

Majiwa P.A.O., Hamers R., Van Meirvenne N. and Matthyssens G. (1986b) Evidence for genetic diversity in *Trypanosoma (Nannomonas) congolense.* Parasitol., 93, 291-304.

Maniatis T., Fritsch E.F. and Sambrook J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory. U.S.A.

Masake R.A., Nyambati V.M., Nantulya V.M., Majiwa P.A.O., Moloo S.K. and Musoke A.J. (1988) The chromosome profiles of *Trypanosoma congolense* isolates from Kilifi, Kenya and their relationship to serodeme identity. Mol. Biochem. Parasitol., 30, 105-112.

Matthyssens G., Michiels F., Hamers R., Pays E. and Steinert M. (1981) Two variant surface glycoproteins of *Trypanosoma brucei* have a conserved C-terminus. Nature, 293, 230-233.

Maudlin I., Turner M.J., Dukes P. and Miller N. (1984) Maintenance of *Glossina morsitans morsitans* on antiserum to procyclic trypanosomes reduced infection rates with homologous and heterologous *Trypanosoma congolense* stocks. Acta Trop., 41, 253-257.

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

Maudlin I. and Welburn S.C. (1988) The role of lectins and trypanosome genotype in the maturation of midgut infections in *Glossina morsitans*. Trop. Med. Parasit., 39, 56-58.

Maudlin I., Welburn S.C. and Mehlitz, (1990) The relationship between rickettsia-like-organisms and trypanosome infections in natural populations of tsetse in Liberia. Trop. Med. Parasitol., 41, 265-267.

Melefors O. and von Gabain A. (1988) Site-specific endonucleolytic cleavages and the regulation of stability of *E.coli omp*A mRNA. Cell, 52, 893-901.

Messing J. (1983) New M13 vectors for cloning. Methods Enzymol., 101, 20-78.

Metcalf P., Blum M., Freymann D., Turner M.J. and Wiley D. (1986) Two variant surface antigens of *Trypanosoma brucei* of different sequence classes have similar 6 A resolution X-ray structure. Nature, 325, 84-86.

Michels P.A.M., Poliszczak A., Osinga K.A., Misset O., Van Beeumen J., Wierenga R.K., Borst P. and Opperdoes F.R. (1986) Two tandemly linked identical genes code for the glycosomal glyceraldehyde-phosphate dehydrogenase in *Trypanosoma brucei.* EMBO J., 5, 1049-1056.

Michiels F., Matthyssens G., Darville M., Cravador A., Steinert M. and Hamers R. (1983) Gene activation and re-expression of a *Trypanosoma brucei* variant surface glycoprotein. EMBO J., 2, 1185-1192.

Milhausen M.R.G., Nelson S., Sather M., Selkirk N. and Agabian N. (1984) Identification of a small RNA containing the trypanosome spliced leader: a donor of the shared 5' sequences of trypanosomatid mRNAs? Cell, 38, 721-729.

Mosser D.M. and Roberts J.F. (1982) *Trypanosoma brucei*: recognition *in vitro* of two developmental forms by murine macrophages. Exp. Parasitol., 54, 310-316.

Mottram J., Perry K.L., Lizardi P.M., Luhrmann R., Agabian N. and Nelson R.G. (1989) Isolation and sequence of four small nuclear U RNA genes of *Trypanosoma brucei* subsp. *brucei:* Identification of the U2, U4 and U6 RNA analogs. Mol. Cell. Biol., 9, 1212-1223.

Mowatt M.R. and Clayton C.E. (1987) Developmental regulation of a novel repetitive protein of *Trypanosoma brucei*. Mol. Cell. Biol., 7, 2838-2844.

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

Mowatt M.R., Wisdom G.S. and Clayton C.E. (1989) Variation of tandem repeats in the developmentally regulated procyclic acidic repetitive proteins of *Trypanosoma brucei*. Mol. Cell. Biol., 9, 1332-1335.

Muhich M.L. and Boothroyd J. (1988) Polycistronic transcripts in trypanosomes and their accumulation during heat shock: evidence for a precursor role in mRNA synthesis. Mol. Cell. Biol., 8, 3837-3846.

Mullner E.W. and Kuhn L.C. (1988) A stem-loop in the 3' untranslated region mediates iron-dependent regulation of transferrin receptor mRNA stability in the cytoplasm. Cell, 53, 815-825.

Murphy W.J., Watkins K.P. and Agabian N. (1986) Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans* splicing. Cell, 47, 517-525.

Murray M., Barry J.D., Morrison W.I., Williams R.O., Hirumi H. and Rovis L. (1980) A review of the prospects for vaccination in African trypanosomiasis. World Animal Rev., 33, 14-18.

Murray M., Hirumi H. and Moloo S.K. (1985) Suppression of *Trypanosoma congolense*, *T. vivax* and *T. brucei* infection rates in tsetse flies maintained on goats immunized with uncoated forms of trypanosomes grown *in vitro*. Parasitol., 91, 53-66.

Murray N.E. (1983) Phage lambda and molecular cloning. In Lambda II, Hendrix R.W., Roberts J.W., Stahl F.W. and Weisberg R.A. eds. Cold Spring Harbor Laboratory. U.S.A.

Mutharia L.M. and Pearson T.W. (1987) Surface carbohydrates of procyclic forms of African trypanosomes studied using fluorescence activated cell sorter analysis and agglutination with lectins. Mol. Biochem. Parasitol., 23, 165-174.

Nantulya V.M., Doyle J.J. and Jenni L. (1978) Studies on *Trypanosoma congolense*: 1. On the morphological appearance of the parasite in the mouse. Acta Trop., 35, 329-337.

Nantulya V.M. and Moloo S.K. (1988) Suppression of cyclical development of *Trypanosoma brucei brucei* in *Glossina morsitans centralis* by an anti-procyclics monoclonal antibody. Acta Trop., 45, 137-144.

Nelson R.G., Parsons M., Selkirk M., Newport G., Barr P.J. and Agabian N. (1984) Sequences homologous to the variant antigen mRNA spliced leader are present in Trypanosomatidae which do not undergo antigenic variation. Nature, 308, 665-667.

Nishikawa K. (1983) Assessment of secondary structure prediction of proteins: comparison of Chou-Fasman method with others. Biochim. Biophys. Acta, 748, 285-299.

Olafson R.W., Clarke M.W., Keilland S.O., Barbet A.F. and McGuire T.D. (1984) Amino-terminal sequence homologies amongst variant surfce glycoproteins of the African trypanosomes. Mol. Biochem. Parasitol., 2, 287-298.

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

Osinga K.A., Swinkels B.W., Gibson W.G., Borst P., Veeneman G.H., Van Boom J.H., Michels P.A.M. and Opperdoes F.R. (1985) Topogenesis of microbody enzymes: a sequence comparison of the genes for the glycosomal (microbody) and cytosolic phosphoglycerate kinases of *Trypanosoma brucei*. EMBO J., 4, 3811-3817.

Pace T., Ponzi M., Dore E. and Frontali C. (1987) Telomeric motifs are present in a highly repetitive element in the *Plasmodium berghei* genome. Mol. Biochem. Parasitol., 24, 193-202.

Paindavoine P., Zampetti-Bosseler F., Pays E.,Schweizer J., Guyaux M., Jenni L. and Steinert M. (1986) Trypanosome hybrids generated in tsetse flies by nuclear fusion. EMBO J., 5, 36313636.

Parish N.M., Morrison W.I. and Pearson T.W. (1985) Identification of an antigen specific to *Trypanosoma congolense* using monoclonal antibodies. J. Immunol., 134, 593-597.

Parodi A.J., Pollevick G.D., Mautner M., Buschiazzo A., Sanchez D.O. and Frasch A.C.C. (1992) Identification of the gene(s) coding for the *trans*-sialidase of *Trypanosoma cruzi*. EMBO J., 11, 1705-1710.

Parsons M.R.G., Nelson K.P., Watkins N. and Agabian N. (1984) Trypanosome mRNAs share a common 5' spliced leader sequence. Cell, 38, 309-316.

Pays E. (1989) Pseudogenes, chimaeric genes and the timing of antigen variation in African trypanosomes. Trends Genet., 5, 389391.

Pays E., Delauw M.F., Van Assel S., Laurent M., Vervoort T., Van Meirvenne N. and Steinert M. (1983a) Modifications of a *Trypanosoma brucei* antigen gene repertoire by different DNA recombinational mechanisms. Cell, 35, 721-731.

Pays E., Laurent M., Delinte K., Van Meirvenne N. and Steinert M. (1983b) Differential size variations between transcriptionally active and inactive telomeres of *Trypanosoma brucei*. Nucl. Acids Res., 11, 8137-8147.

Pays E., Tebabi P., Pays A., Coquelet H., Revelard P., Salmon D. and Steinert M. (1989a) The genes and transcripts of an antigen gene expression site from *Trypanosoma brucei*. Cell, 57, 835-845.

Pays E., Coquelet H., Pays A., Tebabi P. and Steinert M. (1989b) *Trypanosoma brucei*: posttranscriptional control of the variable surface glycoprotein gene expression site. Mol. Cell. Biol., 9, 4018-4021.

Pays E., Coquelet H., Tebabi P., Pays A., Jefferies D., Steinert M., Koenig E., Williams R.O. and Roditi I. (1990) *Trypanosoma brucei*: constitutive activity of the VSG and procyclin promoters. EMBO J., 9, 3145-3151.

Perry K. and Agabian N. (1991) mRNA processing in the Trypanosomatidae. Experientia, 47, 118-128.

Pollard V.W., Rohrer S.P., Michelotti E.F., Hancock K. and Hajduk S.L. (1990) Organization of minicircle genes for guide RNAs in *Trypanosoma brucei*. Cell, 63, 783-790.

Portier C., Dondon L., Grunberg-Manago M. and Regnier R. (1978) The first step in the functional inactivation of the *Escherichia coli* polynucleotide phosphorylase messenger is a ribonuclease III processing at the 5' end. EMBO J., 6, 2165-2170.

Raleigh E.A. and Wilson G. (1986) *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. Proc. Natl. Acad. Sci. USA, 83, 9070-9074.

Richardson J.P., Jenni L., Beecroft R.P. and Pearson T.W. (1986) Procyclic tsetse fly midgut forms and culture forms of African trypanosomes share stage- and species-specific surface antigens identified by monoclonal antibodies. J. Immunol., 136, 2259-2267.

Richardson J.P., Beecroft R.P., Tolson D.L., Liu M.K. and Pearson T.W. (1988) Procyclin: an unusual immunodominant glycoprotein surface antigen from the procyclic stage of African trypanosomes. Mol. Biochem. Parasitol., 31, 203-216.

Roditi I., Carrington M. and Turner M.J. (1987) Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*. Nature, 325, 272-274.

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

Roditi I., Dobbelaere D., Williams R.O., Masterson W.,Beecroft R., Richardson J.P. and Pearson T.W. (1989b) Expression of *Trypanosoma brucei* procyclin as a fusion protein in *E.coli*. Mol. Biochem. Parasitol., 34, 35-44.

Roditi I. and Pearson T.W. (1990) The procyclin coat of African trypanosomes (or the not-so-naked trypanosome). Parasitol. Today, 6, 79-82.

Rosen N.L., Onodera M., Patton C., Lipman M. and Richards F.F. (1979) *Trypanosoma congolense*: isolation and purification. Exp. Parasitol., 47, 378-383.

Rosen N.L., Onodera M., Hotez P.J., Bogucki M.S., Elce B., Patton C., Konigsberg W.H., Cross G.A.M. and Richards F.F. (1981) *Trypanosoma congolense*: surface glycoproteins of two early bloodstream variants. Exp. Parasitol., 52, 210-218.

Ross C.A., Gray M.A., Taylor A.M. and Luckins A.G. (1985) *In vitro* cultivation of *Trypanosoma congolense*: establishment of infective mammalian forms in continuous culture after isolation from the blood of infected mice. Acta Trop., 42, 113-122.

Rovis L., Musoke A.J. and Moloo S.K. (1984) Failure of trypanosomal membrane antigens to induce protection against tsetse-transmitted *Trypanosoma vivax* or *T.brucei* in goats and rabbits. Acta Trop., 41, 227-236.

Rudenko G. and Van der Ploeg L.H.T. (1989) Transcription of telomere repeats in protozoa. EMBO J., 8, 2633-2638.

Rudenko G., Bishop D., Gottesdiener K. and Van der Ploeg L.H.T. (1989) Alpha-amanitin resistant transcription of protein coding genes in insect and bloodstream form *Trypanosoma brucei*. EMBO J., 8, 4259-4263.

Rudenko G., Le Blanq S., Smith J., Lee M. G.-S., Rattray A. and Van der Ploeg L.H.T. (1990) Procyclic acidic repetitive protein (PARP) genes located in an unusually small alpha-amanitin resistant transcription unit: PARP promoter activity assayed by transient DNA transfection of *Trypanosoma brucei*. Mol. Cell. Biol., 10, 3492-3504.

Rudenko G., Chung H.M.M. , Pham V.P., and Van der Ploeg L.H.T. (1991) RNA Polymerase I Can Mediate Expression of CAT and *neo* Protein-Coding Genes in *Trypanosoma brucei*. EMBO J., 10, 3387-3397.

Rudenko G., Lee M.G.-S. , and Van der Ploeg L.H.T. (1992) The PARP and VSG Genes of *Trypanosoma brucei* Do Not Resemble RNA Polymerase-II Transcription Units in Sensitivity to Sarkosyl in Nuclear Run-on Assays. Nucl. Acids Res., 20, 303-306.

Russel M., Kidd S. and Kelley M.R. (1986) An improved filamentous helper phage for generating single-stranded plasmid DNA. Gene, 45, 333-338.

Sanger F., Nicklen S. and Coulson A.R. (1977) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA, 74, 5463-5467.

Sather S. and Agabian N. (1985) A 5'-spliced leader is added in *trans* to both alpha and beta tubulin transcripts in *Trypanosoma brucei*. Proc. Natl. Acad. Sci. USA, 82, 5695-5699.

Schwartz D.C. and Cantor C.R. (1984) Separation of yeast chromosome sized DNAs by pulsed field gradient gel electrophoresis. Cell, 37, 67-75.

Seebeck T., Whittaker P., Imboden M., Hardman N. and Braun R. (1983) Tubulin genes of *Trypanosoma brucei*: a tightly clustered family of alternating genes. Proc. Natl. Acad. Sci. USA, 80, 4634-4638.

Seed J.R. (1964) Antigenic similarity among culture forms of the "*brucei*" group of trypanosomes. Parasitol., 54, 593-596.

Seed B., Parker R.C. and Davidson N. (1982) Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digestion. Gene, 19, 201-209.

Serrano L., Neira J.-L., Sancho J. and Fersht A.R. (1992) Effect of alanine versus glycine in alpha-helices on protein stability. Nature, 356, 453-455.

Shaw J.M., Feagin J.E., Stuart K. and Simpson L. (1988) Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. Cell, 53, 401-411.

Shea C., Lee M.G.-S. and Van der Ploeg L.H.T. (1987) VSG gene 118 is transcribed from a cotransposed pol-I like promoter. Cell, 50, 603-612.

Short J.M., Fernandez J.M., Sorge J.A. and Huse W.D. (1988) Lambda Zap: a bacteriophage lambda expression vector with *in vivo* excision properties. Nucl. Acids Res., 16, 7583-7600.

Simpson L. (1987) The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication, and evolution. Ann. Rev. Microbiol., 41, 363-382.

Simpson L. and Shaw J. (1989) RNA editing and the mitochondrial cryptogenes of kinetoplastid protozoa. Cell, 57, 355-366.

Smith D.B. and Johnson K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene, 67, 31-40.

Smith J.L., Levin J.R., Ingles C.J. and Agabian N. (1989) In trypanosomes the homolog of the largest subunit of RNA polymerase II is encoded by two genes and has a highly unusual C-terminal domain structure. Cell, 56, 815-827.

Steinert M. and Van Assel S. (1980) Sequence heterogeneity in kinetoplast DNA: reassociation kinetics. Plasmid, 3, 7-17.

Sternberg J., Tait A., Haley S., Wells J.M., Le Page R.W.F., Schweizer J. and Jenni L. (1988) Gene exchange in African trypanosomes: characterisation of a new hybrid genotype. Mol. Biochem. Parasitol., 27, 191-200.

Strickler J.E., Binder D.A.L., Italien J.J., Shimamoto G.T., Wait S.W., Dalheim L.J., Novotny J., Radding J.A., Konigsberg W.H., Armstrong M.Y.K., Richards F.F. and Lalor T.M. (1987) *Trypanosoma congolense* : structure and molecular organization of the surface glycoproteins of two early bloodstream variants. Biochem., 26, 796-805.

Stuart K. (1979) Kinetoplast DNA of *Trypanosoma brucei*: Physical map of the maxicircle. Plasmid, 2, 520-528.

Stuart K. (1991) RNA editing in mitochondrial messenger RNA of trypanosomes. Trends Biochem. Sci., 16, 68-72.

Stuart K., Feagin J.E. and Abraham J.M. (1988) RNA editing: the creation of nucleotide sequences in mRNA - a minireview. Gene, 82, 155-160.

Sturm N.R. and Simpson L. (1990a) Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. Cell, 61, 879-884.

Sturm N.R. and Simpson L. (1990b) Partially edited mRNAs for cytochrome b and subunit III of cytochrome oxidase from *Leishmania tarentolae* mitochondria: RNA editing intermediates. Cell, 61, 871-878.

Sutton R.E. and Boothroyd J.C. (1986) Evidence for *trans* splicing in trypanosomes. Cell, 47, 527-535.

Swinkels B.W., Gibson W.C., Osinga K.A., Kramer R., Veeneman G.H., Van Boom J.H. and Borst P. (1986) Characterization of the gene for the microbody (glycosomal) triosephosphate isomerase of *Trypanosoma brucei*. EMBO J., 5, 1291-1298.

Tabor S. and Richardson C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc Natl. Acad. Sci. USA, 84, 4767-4771.

Tait A. (1980) Evidence for diploidy and mating in trypanosomes. Nature, 287, 536-538.

Tait A., Turner C.M.R., Le Page R.W.F. and Wells J.M. (1989) Genetic evidence that metacyclic forms of *Trypanosoma brucei* are diploid. Mol. Biochem. Parasitol., 37, 247-256.

ten Asbroek A.L.M.A., Ouellette M. and Borst P. (1990) Targeted insertion of the neomycin phosphotransferase gene into the tubulin gene cluster of *Trypanosoma brucei*. Nature, 348, 174-175.

Thomashow L.S., Milhausen M., Rutter W.J. and Agabian N. (1983) Tubulin genes are tandemly linked and clustered in the genome of *Trypanosoma brucei*. Cell, 32, 35-43.

Tokuyasu (1986) Application of cryoultramicrotomy to immunochemistry. J. Microscopy, 143, 139-149.

Tschudi C.E. and Ullu (1988) Polygene transcripts are precursors to calmodulin mRNAs in trypanosomes. EMBO J., 7, 455-463.

Tschudi C. and Ullu E. (1990) Destruction of U2, U4 or U6 small nuclear RNA blocks *trans* splicing in trypanosome cells. Cell, 61, 459-466.

Turner C.M.R., Sternberg J., Buchanan N., Smith E., Hide G. and Tait A. (1990) Evidence That the Mechanism of Gene Exchange in *Trypanosoma brucei* Involves Meiosis and Syngamy. Parasitol., 101, 377-386.

Turner M.J., Cardoso de Almeida M.L., Gurnett A.M., Raper J. and Ward J. (1984) Biosynthesis, attachment and release of variant surface glycoproteins of the African trypanosome. Curr. Topics Microbiol. Immunol., 117, 23-55.

Underwood M.R. and Fried H.M. (1990) Characterization of nuclear localizing sequences derived from yeast ribosomal protein L29. EMBO J., 9, 91-99.

Van der Ploeg L.H.T., Valerio D., De Lange T., Bernards A., Borst P. and Grosveld F.G. (1982) An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. Nucl. Acids Res., 10, 5905-5923.

Van der Ploeg L.H.T., Cornelissen A.W.C.A., Barry J.D. and Borst P. (1984a) Chromosomes of kinetoplastida. EMBO J., 3, 3109-3115.

Van der Ploeg L.H.T., Schwartz D.C., Cantor C.R. and Borst P. (1984b) Antigenic variation in *Trypanosoma brucei* analyzed by electrophoretic separation of chromosome sized DNA molecules. Cell, 37, 77-84.

Van der Ploeg L.H.T., Liu A.Y.C. and Borst P. (1984c) The structure of the growing telomeres of trypanosomes. Cell, 36, 459-468.

Van der Ploeg L.H.T., Smith C.L., Polvere R.I. and Gottesdiener K.M. (1989) Improved separation of chromosome-sized DNA from *Trypanosoma brucei*, stock 427-60. Nucl. acids Res., 17, 32173227.

van der Spek H., van den Burg J., Croiset A., van den Broek M., Sloof P. and Benne R. (1988) Transcripts from the frameshifted MURF3 gene from *Crithidia fasciculata* are edited by U insertion at multiple sites. EMBO J., 7, 2509-2514.

Vickerman K. (1969) On the surface coat and flagellar adhesions in trypanosomes. J.Cell Sci., 5, 169-193.

Vickerman K. (1985) Developmental cycles and biology of pathogenic trypanosomes. Brit. Med. Bull., 41, 105-114.

Vickerman K. and Barry J.D. (1982) African Trypanosomiasis. In Immunology of Parasitic Infections (2nd ed.), Cohen S. and Warren K.S. eds. Oxford: Blackwell Sci. Publ.

von Heijne G. (1983) Patterns of amino acids near signal sequence cleavage sites. Eur. J. Biochem., 133, 17-21.

von Heijne G. (1985) Signal sequences: the limits of variation. J. Mol. Biol., 184, 99-105.

von Heijne G. (1986) Towards a comparative anatomy of N-terminal topogenic protein sequences. J. Mol. Biol., 189, 239-242.

von Heijne G. (1988) Transcending the impenetrable: how proteins come to terms with membranes. Biochim. Biophys. Acta, 947, 307-333.

Walder J.A., Eder P.S., Engman D.M., Brentano S.T., Walder R.Y., Knutzon D.S., Dorfman D.M. and Donelson J.E. (1986) The 35-nucleotide spliced leader sequence is common to all trypanosome messenger RNAs. Science, 233, 569-571.

Welburn S.C., Maudlin I. and Ellis (1989) Rate of trypanosome killing by lectins in midguts of different species and strains of *Glossina*. Med. Vet. Entomol., 3, 77-82.

Welburn S.C. and Maudlin I. (1989) Lectin signalling of maturation of *T.congolense* infections in tsetse. Med. Vet. Entomol., 3, 141-145.

Welburn S.C. and Maudlin I. (1990) Haemolymph lectin and the maturation of trypanosome infections in tsetse. Med. Vet. Entomol., 4, 43-48.

Welburn S.C. and Maudlin I. (1991) Rickettsia-like organisms, puparial temperature and susceptibility to trypanosome infection in *Glossina morsitans*. Parasitol., 102, 201-206.

Whittaker P.A., Campbell A.J.B., Southern E.M. and Murray N.E. (1988) Enhanced recovery and restriction mapping of DNA fragments cloned in a new lambda vector. Nucl. Acids Res., 16, 6725-6736.

Wijers D.J. and Willet K.C. (1960) Factors that may influence the infection rate of *Glossina palpalis* with *Trypanosoma gambiense*: the number and morphology of the trypanosomes present in the blood of the host at the time of the infective feed. Ann. Trop. Med., 54, 341-349.

Wilson T. and Treisman R. (1988) Removal of poly(A) and consequent degradation of *c-fos* mRNA facilitated by 3' AU-rich sequences. Nature, 336, 396-399.

Wilson I.B.H., Gavel Y. and von Heijne G. (1991) Amino acid distributions around *O*-linked glycosylation sites. Biochem. J., 275, 529-534.

Yamauchi K. (1991) The sequence flanking translational initiation site in protozoa. Nucl. Acids Res., 19, 2715-2720.

Yanisch-Perron C., Vieira J. and Messing J. (1985) Improved M13 cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. Gene, 33, 103-119.

Yao M.-C. and Yao C.-H. (1991) Transformation of *Tetrahymena* to cycloheximide resistance with a ribosomal proetin gene through sequence replacement. Proc. Natl. Acad. Sci. USA, 88, 9493-9497.

Young C.J. and Godfrey D.G. (1983) Enzyme polymorphism and the distribution of *Trypanosoma congolense* isolates. Ann. Trop. Med. Parasit., 77, 467-481.

Ziegelbauer K., Quinten M., Schwartz H., Pearson T.W. and Overath P. (1990) Synchronous differentiation of *Trypanosoma brucei* from bloodstream to procyclic forms *in vitro*. Eur. J. Biochem., 192, 373-378.

Zomerdijk J.C.B.M., Ouellette M., ten Asbroek A.L.M.A., Kieft R., Bommer A.M.M., Clayton C.E. and Borst P. (1990) The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. EMBO J., 9, 2791-2801.

Zomerdijk , J.C.B.M., Kieft R., Shiels P.G. and Borst P. (1991) Alpha-Amanitin-Resistant Transcription Units in Trypanosomes - A Comparison of Promoter Sequences for a VSG Gene Expression Site and for the Ribosomal RNA Genes.

APPENDIX

(a) Sequencing and PCR primers

302 - CTGCAGCTCCTGCTCTGG

311 - GGTGTAGGACACGGGACA

331 - GAAGAAGCGTGGTGTGGA

346 - AAAGTGCACGGGTGCCGC

347 - ATGCCAGGTTGGGCGAAT

484 - TCGCCTCTCCGGTCAGT

485 - CACGAACACGGCACCGT

486 - TGCAGCCGCTTCCGTCG

487 - TGAGGAGATTGACGGGA

488 - TGCCTCTCCGTGTTACT

490 - CATTGTGAAGGCTCGGT

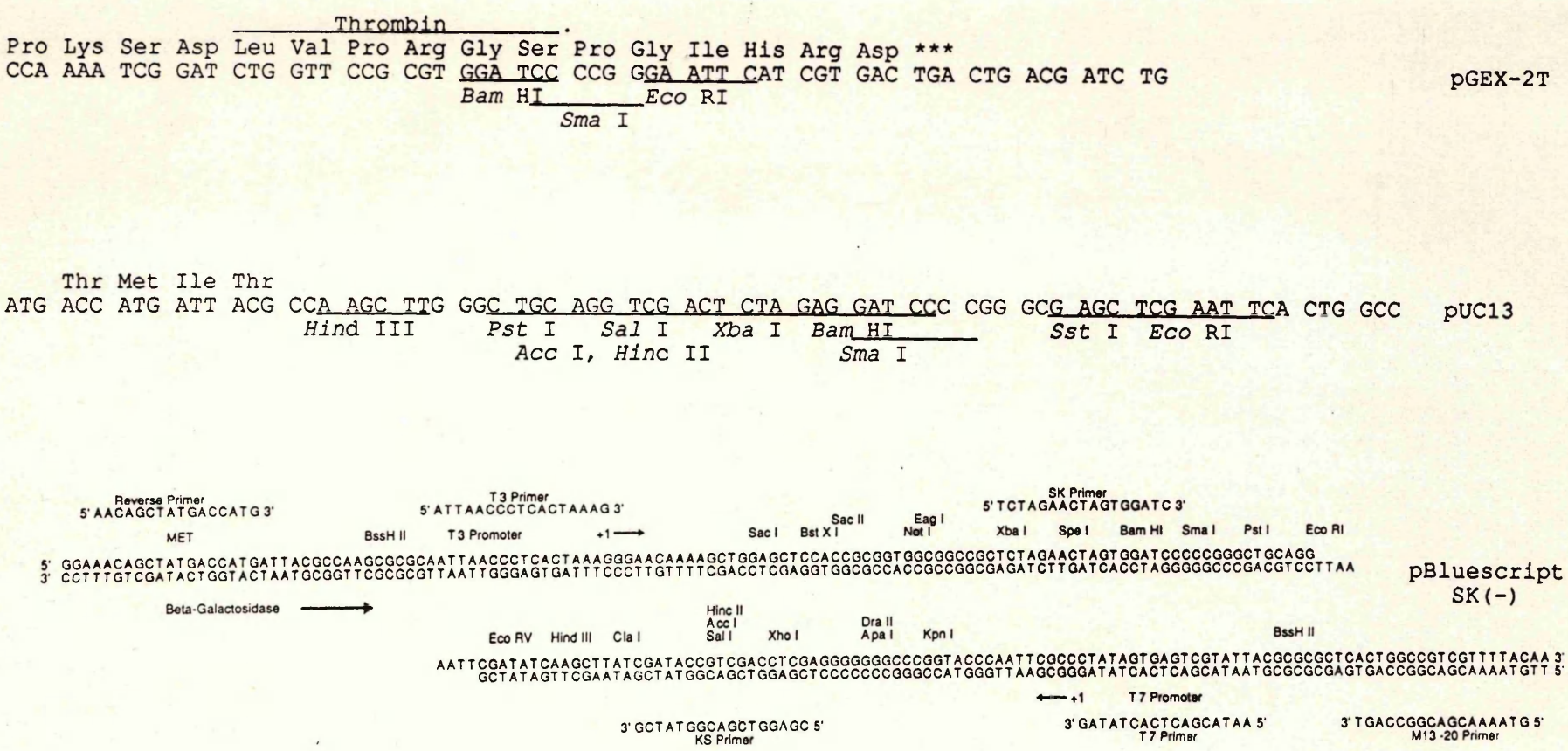
491 - TACCCATTGGCCAACAG

464 - GGGCGGCCGCACAGTTTCTGTACTATATTG

(b) Probes

SL - AACGCTATTATTAGAACAGTTTCTGTACTATATTG

Appendix I Oligonucleotide Sequences



Appendix II Polylinker Sequences of Plasmid Vectors pGEX-2T, pUC13 and pBluescript SK(-).