ANALYSIS OF METACYCLIC TELOMERES IN TRYPANOSOMA BRUCEI

by

Paul Gerard Shiels B.A. (Mod.).Dublin.
Institute of Genetics
University of Glasgow

Presented in submision for the
degree of Doctor of Philosophy.

March 1990.

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

DECLARATION.

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

Paul G. Shiels.

This thesis is dedicated to my Parents.

ACKNOWLEDGEMENTS

I would like to thank Professor D. Sheratt for giving me the opportunity to undertake my Ph.D in Glasgow and Dr. J.D. Barry for his supervision, encouragement and criticisms.

Thanks is due to all others in the Genetics department who helped make this research possible, and to Rab, Doris and Agnes for the laughs. Special thanks to Jackie, Ian, Tim, Andy, Chris and all in the G.U.C.S.C. for their support and encouragement.

Many thanks to Davy M^{c} Combe for help with photography and entertainment in taxi queues, and especially to Lesley for everything.

Finally I would like to thank the Bhoys for Saturdays and all the Gael in Glaschu for keeping me sane, go raibh maith'ad.

CONTENTS

Summary.

Trypanosoma brucei spp. are the agents responsible for African sleeping sickness in man and Nagana in cattle. The organisms have the ability to evade the host's immune system by antigenic variation of their surface coat. The surface coat of the infective forms is composed of a single molecular species, the variant surface glycoprotein (VSG).

Each specific VSG is encoded by a separate gene, expression of which occurs in a loose hierarchical order. T.brucei has the coding capacity for approximately 10^3 VSG genes, which are found either in chromosomal clusters or at telomeric loci; it is only from the latter that the gene may be expressed.

Telomeric expression sites (ESs) utilized during bloodstream infection are complex, typically between 40 - 60 Kb long, containing several non-VSG expression site associated genes (ESAGs) and preceded by a long barren region devoid of restriction sites. Transcription of telomeric ESs is insensitive to the toxin alpha-amanitin and transcription of a given ES appears to be mutually exclusive of other ESs in vivo. By virtue of their location, chromosomal internal VSG genes need to be transposed to a telomere for expression. The transposition event is duplicative and produces an expression linked copy of the gene. Telomeric VSG genes, however, can be expressed either by duplicative transposition or reciprocal recombination to an active ES, or by in situ activation.

The metacyclic stage of the trypanosome life cycle is the first to

express VSG genes. The metacyclic form utilizes a highly predictable subset of VSG genes (M-VSGs), comprising only 1-2% of the entire VSG gene repertoire, which appear to be expressed by a distinct and dominant mechanism to that employed during chronic bloodstream infection.

Direct molecular analysis of M-VSG gene expression is precluded by the paucity of metacyclic forms in the salivary exudate of the tsetse fly vector and the transient nature of this developmental stage which cannot be cultured in vitro. M-VSG gene expression, however, is still extant in the host bloodstream in the first few days following fly bite. Analysis during this period is compounded by the polyclonal origin of the M-VSG genes expressed by individual organisms and instability of VSG expression.

The work described in this thesis focuses on attempts to clone an M-VSG gene telomere and to gain insight into the predictability and stability of the M-VSG repertoire, by analysis of the telomere in a model trypanosome line which circumvents some of the problems associated with previous systems.

Cloning and analysis of the BC telomeres for the M-VSG genes for GUTat 7.1 and ILTat 1.22 revealed that each had a remarkably small barren region and shared no sequence homology with ESs used in chronic bloodstream infection, apart from the 70 bp repeat sequence constituting the barren region 5'of the VSG gene.

Transcriptional analysis of the ILTat 1.22 metacyclic ES, utilising the model line of trypanosomes expressing the gene in situ, revealed

that the ES is extremely short in comparison to bloodstream ESs, extending no more than 3.5 - 4 Kb 5' of the VSG gene.

One other region of the ILTat 1.22 BC and GUTat 7.1-2 BC telomeres appeared to be transcriptionally active. This comprised a genomic repetitive element which also was transcribed in procyclic trypanosomes.

The structural individuality of these telomeres was proposed as underlying the stability and physical distinction of the M- VSG repertoire. This hypothesis is supported by an epidemiological analysis of the ILTat 1.22 BC telomere over a 24 year period and spanning diverse epidemic foci of infection.

Within Kenyan epidemic foci this telomere is present unaltered in all the stocks investigated over the the period examined and suggests that spread of the disease in East Africa is principally by mechanical transmission; this has important consequences for tackling the disease at source.

Cloning and analysis of these telomeres now facilitates characterisation of metacyclic ES control elements and comparisons with other M-ES to be made.

ABBREVIATIONS.

ENZYME KEY.

B:	Bam HI.	Hc: Hin cII.	Sc: Sca I.
Bg:	Bgl II.	K: Kpn I.	T: Taq I.
C:	Cla I.	P: Pst I.	X: Xba I.
E:	Eco RI.	Pv: Pvu II.	
H:	Hin dIII.	S: Sph I.	

CHEMICALS.

ATP : Adenosine triphosphate.

BSA : Bovine serum albumin.

$CHCl_3$: Chloroform.

CsCl : Caesium chloride.

ddH_2O : Distilled deionized water.

DEPC : Diethylpyrocarbonate.

DNA : Deoxyribonucleic acid.

cDNA : Complementary DNA.

kDNA : Kinetoplast DNA.

DTT : Dithiothreitol.

EDTA : Ethylenediamine tetra-acetic acid (disodium salt).

HCL : Hydrochloric acid.

IPTG : Isopropylthiogalactoside.

KCL : Potassium chloride.

$MgCl_2$: Magnesium chloride.

$MgSO_4$: Magnesium sulphate hepta hydrate salt.

Na- : Sodium - salt.

NaOH : Sodium hydroxide.

RNA : Ribonucleic acid.

mRNA : Messesger RNA.

Poly A^+- : Poly adenylated -.

SDS : Sodium dodecyl sulphate.

Tris : Tris (hydroxymethyl) amino methane.

TE : Tris EDTA.

Measurements.

bp : base pairs.

°C : degrees centrigade

g : centrifugal force equal to gravitational acceleration.

gm : grammes.

Kb : kilobase pairs.

Kd : kilodaltons.

Kg : kilogrammes.

L : litre.

M : molar.

Mb : Megabase pairs.

mg : milligrammes.

ml : millilitres.

mM : millimolar.

nm : nanometre.

pH : acidity (neg.$\log_{10}$ molar $[H^+]$).

w/v : weight / volume.

ug : microgrammes.

ul : microlitres.

Techniques.

PFGE : Pulse field gradient electrophoresis.

OFAGE : Orthogonal field alternation gel electrophoresis.

Miscellaneous.

BC : Basic copy

BOTat : Bordeaux Trypanozoon antigen type.

Deme : Population.

EATRO : East Africa trypanosomiasis research organization.

ELC : Expression linked copy.

L-ELC : Lingering ELC.

EMBL : European Molecular Biology Organization.

Focus : Region of endemic infection.

GUTat : Glasgow University Trypanozoon antigen type.

ILTat : ILRAD Trypanozoon antigen type.

LMP : Low melting point.

McAb : monoclonal antibody.

M.O.I. : multiplicity of infection.

P.F.U. : Plaque forming units.

RFLP : Restriction fragment length polymorphism.

Serodeme : serologially defined population.

STIB : Swiss Tropical Institute Basel.

UTAR : Ugandan trypanosome antigenic repertoire.

UV : Ultra violet.

VAT : Variable antigen type.

B-VAT : bloodstream VAT.

I-VAT : ingested VAT.

M-VAT : metacyclic VAT.

VSG : Variant surface glycoprotein.

WRATat : Walter Reed Army Institute antigen type.

Zymodeme : Population defined by isoenzyme analysis.

Chapter 1.

ANTIGENIC VARIATION IN AFRICAN TRYPANOSOMES.

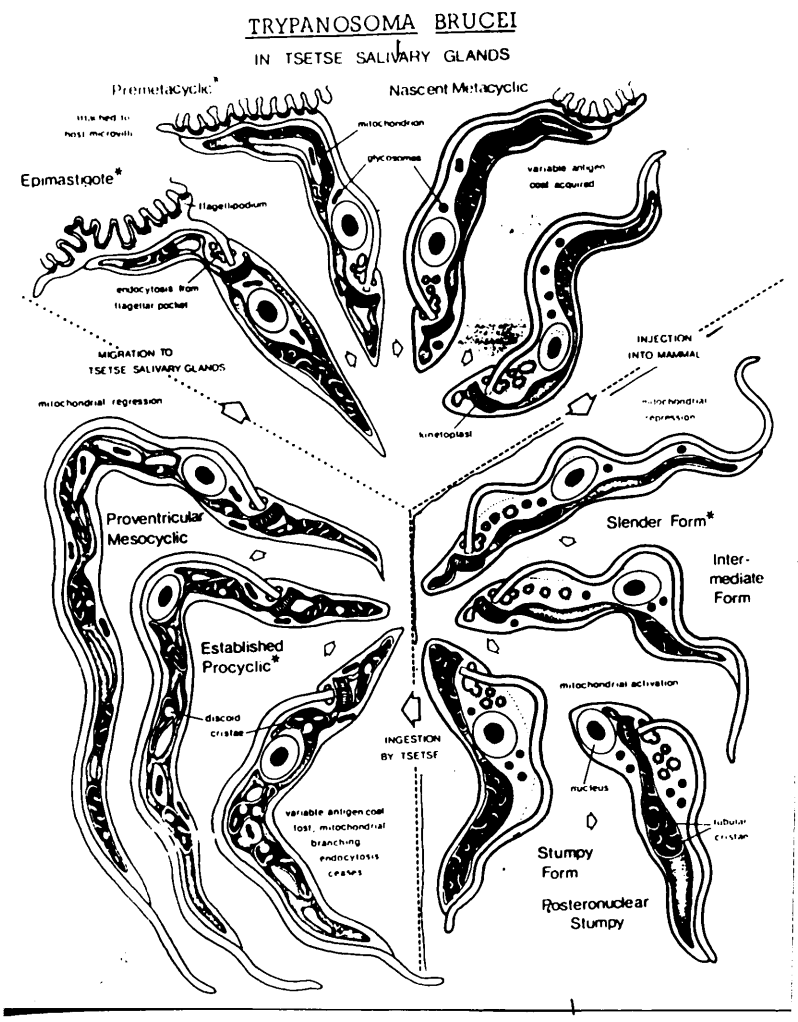
1.1 INTRODUCTION.

Trypanosomes are protozoan flagellates belonging to the order Kinetoplastida, characterized by the possession of a singular mitochondrion with a DNA-containing nucleoid, the kinetoplast, within which the DNA complement is arrayed in catenates of two distinct classes, minicircles and maxicircles.

Members of the genus Trypanosoma are parasites of man and his domestic animals, and as such represent an extensively studied group of organisms. This genus comprises two groups, the Stercoraria, which develop as vertebrate infective forms in the hindgut of their transmitting insect vector, and the Salivaria, which develop as infective forms in the mouthparts or salivary glands of the tsetse fly (Glossina sp.) and are transmitted via inoculation during the insects' bloodmeal (Hoare, 1972; Lumsden and Evans, 1976). Both groups contain human pathogens.

The Stercorarian Trypanosoma cruzi is the agent responsible for Chagas' disease in the New World, while the African salivarian trypanosomes belonging to the T. brucei species are infective to man and his domestic animals. T. brucei brucei is infective to most domestic livestock and as a consequence makes large tracts of Africa unsuitable for farming. This problem is heightened by the fact that T. brucei rhodesiense and T. brucei gambiense are the cause of sleeping sickness in man.

Figure 1 Life Cycle of T. brucei.
Vickerman (1969).



The life cycle of the salivarian trypanosome involves transmission between the vertebrate host and the tsetse fly vector and is typified by that for T. brucei sp. (Figure 1.1), although mechanical transmission, the transfer of infected blood between mammals without involving the developmental changes that occur in the tsetse fly, can also occur on the proboscis of tsetse flies or other haematophagous flies.

1.2 LIFE CYCLE:

Infection of the host is a result of the transmission of up to 10^4 trypanosomes during the tsetse feeding. There is then multiplication at the wound site, resulting in the formation of a chancre and subsequent infection of the lymph and blood systems.

Unlike the Stercorarian T. cruzi, which is an intracellular parasite, salivarian trypanosomes proliferate within the host's circulatory system as pleomorphic bloodstream forms ranging from a long slender form, which divides by binary fission, to the short stumpy form which does not divide and is thought to be pre-adapted to life in the tsetse fly (Vickerman and Barry 1982). This has not yet been formally proven, but the long slender form does show a differential death rate compared to the stumpy form in the posterior midgut of the fly, and this may be an indirect indication of the postulated preadaptation of the latter form to life in the fly (Turner et al., 1988). The basis of this preadaptation is unknown, but may be a consequence of the derepression of mitochondrial functioning beginning in these forms.

Within the bloodstream the trypanosomes evade the host immune system by antigenic variation of their surface coat (Vickerman, 1969; 1979). The whole surface of the infective forms is covered by a unique species of Variant Surface Glycoprotein (VSG) which protects the organism against non-specific immune defences in the host and specific immune response against membrane proteins. The VSG antigen does, however, stimulate a host humoral immune response which destroys the organism. To overcome this particular response, bloodstream populations are heterotypic for the VSG expressed (ie. a range of VSGs is being expressed by the population) and hence a small proportion of the population expressing a different variable antigen type (VAT) escape immune detection and go on to generate a new population. In this way, the infection is prolonged by successive waves of parasitaemia, as each new VAT is detected and the organisms expressing it destroyed.

The surface coat is not required in the tsetse fly and, when the trypanosomes are reingested by the fly, the surface coat is lost rapidly and the parasites transform into procyclic trypomastigotes, with concomitant changes in the ultrastructure of the mitochondrion. Mitochondrial activity recommences in the stumpy form and is completely regenerated in the fly to allow for the activation of the electron transport system (Brown et al., 1973). The loss of the VSG is completed within 48 hours of ingestion by the fly (Barry and Vickerman, 1979) and is the result of a cessation of VSG expression at the transcriptional level (Overath et al., 1983). Furthermore, this transformation can be mimicked in vitro (Brown et al., 1973, Overath et al., 1983). Procyclic trypanosomes are not mammalian infective and are readily lysed by immune serum due to the absence of the

VSG coat.

Development within the fly continues after migration to the salivary glands where attachment to the epithelium occurs and the epimastigote stage of the life cycle appears. How the trypanosome gets to the salivary gland is unknown. It is the attached epimastigote form that gives rise to the free metacyclic forms. The latter reacquires the VSG coat and thus acquires mammalian infectivity. The cycle is completed with injection of the trypanosomes into the host. Such a life cycle is dependent on the insect vector, though this is not necessarily reflected in the range of the disease in Africa. Though generally correlated with the tsetse belt, mechanical transmission of the disease by other blood sucking insects, without the trypanosome completing its lifecycle, allows expansion of the range of trypanosome outside the tsetse belt. Outside this range other vectors are also employed by different Trypanozoon species. T.evansi employs bloodsucking insects and vampire bats, while T.equiperdum is transmitted venereally by horses.

1.3 GENOME AND PLOIDY:

The trypanosome haploid genome is small, comprising about 4.4×10^7 Kb DNA (Borst et al., 1980) as determined by renaturation experiments. The single copy complement is around 68%, the remainder being composed of 20% middle repetitive and 12% highly repetitive DNA (Borst et al., 1980; 1982). The latter grouping mainly consists of satellite DNA which is heavily composed of tandem arrays of a 177 bp A+T rich repeat (Sloof et al., 1983).

The genome of the trypanosome is considered to be diploid. Fluorometric measurements of the nuclear DNA content give comparative agreement with those determined for nuclear complexity by DNA renaturation kinetics (Borst et al., 1980; 1982). This is further supported by both isoenzyme analysis (Tait, 1980; Tait et al., 1984,1985; Gibson et al., 1980) and RFLP analysis (Gibson et al., 1985) of a number of housekeeping genes.

Analysis of trypanosome karyotypes is hampered by the fact that visual inspection of the chromosomes cannot be achieved directly, as they do not condense prior to, or during, mitosis (Vickerman and Preston, 1970), possibly as the result of the organism not possessing histone H1 (Hecker and Gander,1985). This problem has been relieved to an extent by the development of new electrophoretic techniques for the separation of large DNA molecules, such as Pulse Field Gradient (PFG) and Orthogonal Field Alternation Gel Electrophoresis (OFAGE) (Schwartz and Cantor, 1984; Van der Ploeg et al., 1984a; Carle and Olsen, 1984).Use of these procedures in the analysis of the T. brucei genome has revealed four main chromosome classes:

(i) An unresolvable fraction sizing at greater than 2 Mbs, which fails to leave the gel slot;

(ii) At least 6 chromosomes of between 1-2 Mbs;

(iii) A fraction containing 5-7 intermediate sized chromosomes from 200-700 Kb;

(v) Up to 100 minichromosomes of 25-100 Kb (Van der Ploeg et al., 1984 a, b, c; 1985; Johnson and Borst, 1986).

There does not appear to be any rigid organization to the genome and the minichromosome complement may vary significantly between

strains and species. T. equiperdum and T. vivax, for example, have few minichromosomes compared to T.brucei. There may be, however, correlation between the extent of the VSG gene repertoire and the minichromosome number. Minichromosomes harbour both VSG genes and associated telomeric repeat sequences, and their presence in the genome may purely be to increase the potential antigen repertoire of the organism (Van der Ploeg et al., 1984d). No other type of gene has been found on minichromosomes. Such a pool of VSG genes would enhance the trypanosome's ability to reprogramme the order of VSG gene expression during chronic infections and thus give it a selective advantage in a pre-infected host (Borst et al., 1983) since minichromosomes possess telomeres and thus can participate in telomeric exchange. Southern analysis of PFG gels has been used to examine the distribution of genes in the various chromosome classes. Interestingly, all housekeeping genes analysed appear to be on chromosomes of 1-2 Mb or greater in size. Those examined to date include Triose phosphate isomerase (TIM), Phosphofructokinase (PFK), Glyceraldehyde phosphate dehydrogenase (GAPDH), Tubulin and rRNA genes (Gibson et al., 1985; Gibson and Borst, 1986; Van der Ploeg et al., 1984d). In contrast, VSG genes are found on all chromosome classes (Van der Ploeg et al., 1984d). It is likely that this distribution once again reflects the evolution of an increased pool of telomeres available for antigenic switching and VSG gene expression. The evidence for this is indirect, and stems from the observation that essential or housekeeping genes are on the largest chromosomes, which appear to be diploid in complement (Gibson and Borst, 1986). The intermediate chromosomes are possibly haploid, while the ploidy of the minichromosomes is indeterminate (Van der Ploeg et al., 1984d). The latter two classes

show variation in content both between trypanosome stocks and even during the sequential transmission of a given trypanosome clone; hence the importance of having housekeeping genes on the larger chromosomes.

Mitotic dysjunction of trypanosome chromosomes is still not understood. Only 10 kinetochores have been observed (M.Steinert, pers. comm.); far too few to deal with the 100 or more chromosomes observed using PFGE or OFAGE. One possibility for control of the segregation of the surplus chromosomes is that they are attached to specific elements of the nuclear matrix, possibly at centromeric and/or telomeric regions, as has been observed in yeast (Amati and Gasser, 1988).

The participation of minichromosomes in mitotic dysjunction and their subsequent variance in number may be a reflection of their size. In yeast small artificial chromosomes segregate with decreased fidelity below a threshold size (Murray and Szostak, 1983). The presence of extensive amounts of satellite sequence in the minichromosomes may be an attempt to facilitate reaching a size suitable for more stable dysjunction. One might also envisage a limited number of matrix sites for attachment for segregation. This factor in conjunction with a size limit factor may help explain the anomalous observations of Zampetti-Bosseler et al.(1986) and Paindavoine et al. (1986), in that the reported haploid metacyclic DNA content may reflect a non-stochastic loss of some minichromosomal DNA during given periods of the life cycle.

1.4 THE VARIANT SURFACE GLYCOPROTEIN:

The Variant Surface Glycoprotein (VSG) coats the trypanosome surface as a densely packed monolayer, 12-15 nm thick, which normally is composed of a homogeneous, antigenically distinct species of the molecule (Vickerman, 1969; Vickerman and Luckins, 1969). The coat protects the trypanosome by masking antigenically invariant components of the plasma membrane which do elicit specific antibodies, and by antigenic variation of this protective screen (Barry, 1979; Tetley et al.,1981).

The coat is composed of approximately $1.2x10^7$ VSG molecules (Turner et al., 1984) and, though typically it is of one specific variant, 'mosaic' coated forms have been observed in T. brucei (Esser and Schoenbechler, 1985) and T. equiperdum (Baltz et al., 1986).

The capacity of the trypanosome to generate antigenically distinct coats is large; over 100 serologically distinct VSGs have been identified in T. equiperdum (Capbern et al., 1977),and T. brucei has the coding capacity in its genome for up to 10^3 distinct VSG genes (Van der Ploeg et al., 1984b).

In T. brucei each mature VSG monomer has a molecular weight of approximately 55-65 kd. Size estimates have also been made for the T. congolense VSG, at 55-65 kd, and T. vivax at 44 kd, though the number of data for such estimates is very limited (Strickler et al., 1987; Gardiner et al., 1987).

Specific variants differ in amino acid sequence, isoelectric

point, peptide map and carbohydrate content (Cross, 1975). The T. brucei VSG is typically between 450-500 amino acids long and is composed of two domains, an N-terminal domain comprising approximately three-quarters of the amino acid sequence, and a C-terminal domain making up the remainder (Johnson and Cross, 1979). This latter domain is membrane orientated, and shows inter VSG conservation at the primary sequence level, indicative of both a structural constraint with respect to membrane interaction or processing and intramolecular interactions on the surface of the organism (Rice-Ficht et al., 1981; Majumder et al., 1981; Matthysens et al., 1981; Boothroyd and Cross, 1982).

The VSG tail, which anchors it in the plasma membrane, is a glycosyl-phosphatidylinositol moiety attached to the carboxy terminal of the peptide, whose existence is not predicted by sequence analysis of VSG cDNAs as these predict the molecule to be translated with the hydrophobic amino acid tail coded for within the VSG structural gene (Boothroyd et al., 1981; Rice-Ficht et al., 1981).

1.5 VSG GENES: Genomic Organization and Structure.

Each specific variable antigen type (VAT) expressed in the course of infection is the product of a single VSG gene. VSG gene sequences have been estimated at 1050 copies in the T. brucei nucleus as deduced by their presence in 9% of clones from a cosmid library (Van der Ploeg et al., 1982a). Such an estimation is of course, biased by the presence of pseudogenes. VSG genes are found on all chromosome classes (Van der Ploeg et al., 1984a) in two distinct environments, either chromosome internally or at telomeric loci. It is

only from a telomeric locus, however, that a VSG gene may be expressed (Williams et al., 1982; De Lange and Borst, 1982; Pays et al., 1981).

Chromosome internal genes constitute the bulk of the VSG gene repertoire and typically occur in tandem clusters, each gene flanked 5' by a characteristic imperfect direct repeat element 70-80 bp in length which are AT rich, referred to henceforth as 70bp repeats,. The size variation of individual repeat regions is due to imperfection in the number of individual TAA triplets that precede a more conserved 44bp region at the 3' end of the repeat unit, though it should be stated that the functional ends of the unit have not been determined (Van der Ploeg et al., 1982 a; Liu et al., 1983; Campbell et al., 1984). Telomeric VSG genes are also preceded 5' by 70bp repeats, usually an order of magnitude greater in number than for VSG genes located chromosome internally, and these constitute 'barren regions' (ie. devoid of restriction sites) (Pays et al., 1982; Van der Ploeg et al., 1982a). The telomeric VSG gene typically has 50 or more such repeats 5', while chromosome internal genes have as few as 4-5 repeats (Bernards et al., 1984 a, b).

The downstream environment of a telomeric VSG gene is composed of a simple hexameric repeat of consensus 5'-TTAGGG-3'. This repeat is typical of those found at the ends of other eukaryotic chromosomes, (Van der Ploeg et al., 1984c; Blackburn and Challoner, 1984) and fits the protozoan telomeric repeat consensus of 5'(n(T)nPuGGG)n-3'.

The VSG gene is flanked initially 3' by a 'T' rich stretch

followed by a 29bp repeat derived from the simple hexamer, then followed by several kilobases of the simple hexamer (GGGATT). The precise structure of the distal end of the telomere (the terminus) is unknown, but is likely to be similar to that of other lower eukaryotes which possess a 'G' strand overhang of approximately 12bp (Henderson & Blackburn, 1989). As with other eukaryotic telomeres, those of the trypanosome show length variation over increasing cell generations, offset by occasional deletion of blocks of hexameric repeats (Bernards et al., 1983). Runs of these repeats are interrupted by single stranded breaks and hence are good substrate for recombination events.

A novel feature of trypanosome telomeres is that they appear to grow faster when transcriptionally active, up to an average 10bp per generation (Pays et al., 1983). This may be a consequence of the telomere being under torsional stress when transcribed (Greaves and Borst, 1987) and in a more open chromatin conformation, or possibly the lack of modified bases on an actively expressed telomere (Bernards et al., 1984b) may allow enhanced access for the telomerase that would be responsible for addition of the repeats. Another possibility is that modification is linked directly in a slower rate of growth for the telomere.

In keeping with these ideas is the observation that loss of repeats from the telomere is also more readily occurring when it is actively transcribed (Pays et al., 1983).

1.6 ACTIVATION OF VSG GENES.

VSG genes can be expressed only when residing at a telomeric locus, (Pays et al., 1981a: De Lange and Borst, 1982; Williams et al., 1982). The majority of VSG genes are, however, chromosome internal, so expression requires genomic re-arrangements. Molecular analysis of antigenic switching has revealed a number of distinct mechanisms for achieving novel VSG expression, which are illustrated in Figure 1.2.

1.6.1 Duplicative transposition.

When chromosome internal VSG genes are transposed to a telomeric locus to be expressed, the VSG gene previously at that locus is gene converted. In order to conserve its complement of VSG genes, the donor gene, or basic copy (BC) is retained intact as part of the genome, while a duplicated copy, the expression linked copy (ELC) enters the telomeric expression site (ES) where not only is it transcribed, but also it can act as the substrate for further conversion events (Williams et al., 1979, Pays et al., 1981; Hoeijmakers et al., 1980).

That the ELC represents the transcriptionally active copy of the VSG gene is apparent by its sensitivity to DNase I (Pays et al., 1981b) and single stranded endonucleases (Greaves and Borst, 1987).

Transposition to a telomeric ES involves sequence 5' of the VSG gene, and for chromosome internal genes has a 5' boundary within the 70bp repeat region (Van der Ploeg et al., 1982b; Michels

et al., 1983, Campbell et al., 1984; Aline et al., 1985a; De Lange et al., 1986). Sometimes, however, the 3' terminus can be within the VSG coding sequence, and in such cases results in the formation of a chimeric VSG gene whose C-terminal is that of the previous gene at the ES (Michels et al., 1982; Roth et al., 1986). This has obvious implications for the evolution of the VAT repertoire by generating novel VSG coding sequences, through partial conversion events.

The 3' boundary of the duplication event is determined by homology blocks comprising the conserved 14-mer region of the VSG coding sequence and for telomeric BCs, the telomeric repeat sequences (Michels et al., 1983; De Lange, 1986). The mechanisms by which the gene conversion involved in the duplication event is achieved has not been determined formally, but a double strand break model (Szostak et al., 1983; Sun et al., 1989; Fincham and Oliver, 1989), seems likely as the basis of M-VSG gene analysis (this study: Matthews et al in prep.).

1.6.2 Telomeric VSG genes.

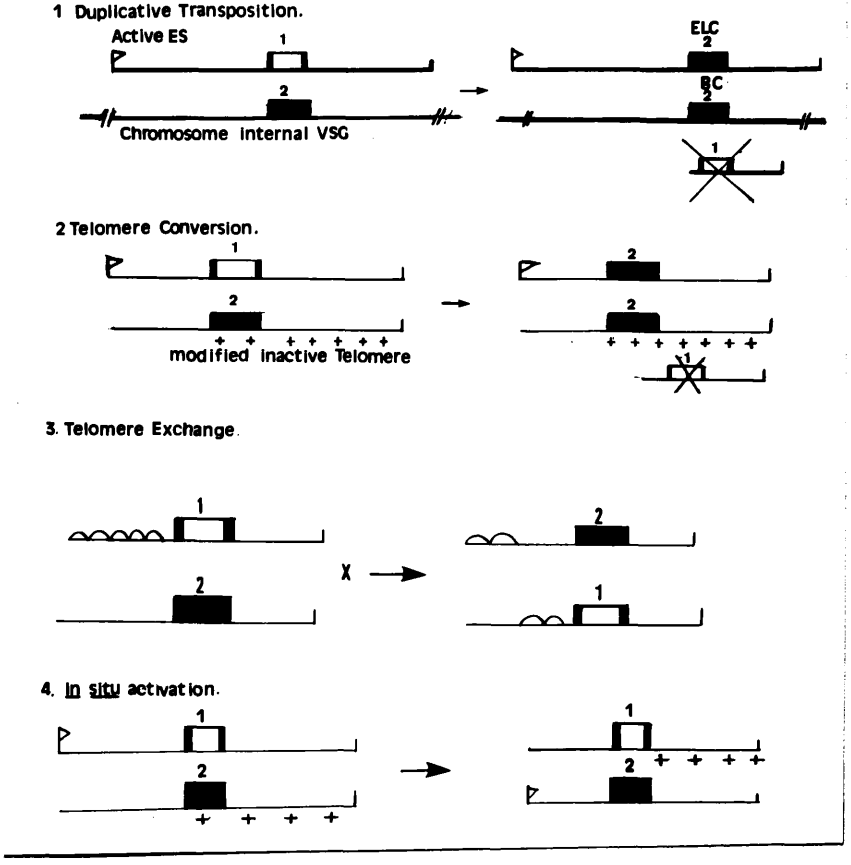
VSG genes already residing at telomeres have a number of options available for expression. They can be expressed in situ, if the antigenic switch entails a new ES becoming activated while the previously active site is switched off, or they can reciprocally recombine into an active ES by telomere exchange or by telomere conversion (see Figure 1.2). Telomeric conversion is analogous to the duplicative transposition of chromosome internal VSG genes. The 3' conversion boundary, due to the extensive range of telomeric repeats can extend to the end of the chromosome, though this has not been

demonstrated experimentally as the exact physical structure of the typanosmome telomere ends have yet to be elucidated (De Lange et al., 1983; Bernards et al., 1984; Laurent et al., 1983; Bernards et al., 1985; Kooter et al., 1988). Similarly, the 5' terminal may extend upstream of the 70bp repeat region utilising other regions of homology between expression sites (Cully et al., 1985; Kooter et al., 1987, 1988; Murphy et al., 1987; Alexandre et al., 1988; Timmers et al., 1987).

Upstream regions of homology also allow for reciprocal exchange between the telomere in the active ES and a transcriptionally inactive telomere. Such exchanges are readily mediated by homology between telomeres in the 70bp region or through other upstream sequences, such as expression site associated genes (ESAGs).

This form of antigenic switch does not alter the VSG copy number and is apparently a rare occurrence (Pays et al., 1985; Shea et al., 1986). This rarity is puzzling, given that there are extensive homologies between telomeres, and may reflect a novelty of active ES chromatin structure. Alternatively, these observations can be interpreted as implying that antigenic switching employs a distinctive biochemical mechanism driven at a rate greater than that for spontaneous recombination events, and that reciprocal exchanges involved in antigenic switching are occurring outwith this mechanism and at a background recombination frequency.

Figure 1.2 Modes of activation of VSG genes.



Another switching mechanism which does not alter the VSG gene number is in situ activation, which occurs when the active ES locus switches, thus initiating transcription of a new VSG gene at the newly activated ES (Van der Ploeg et al., 1984a; Myler et al., 1984; Pays et al., 1983; Laurent et al., 1983; Bernards et al., 1984a). This mode of antigenic switching indicates that trypanosomes contain more than one potential ES, as initially observed by Van der Ploeg et al (1984a) who noted that expression could occur from different chromosome class not only following an antigenic switch, but also for the expression of a given VSG gene in different trypanosome populations. Molecular analysis using ES derived probes has indicated that there may be between 15-25 ES available for VSG gene expression (Cully et al., 1985; Kooter et al., 1987; Murphy et al., 1987). This figure, however, is probably an overestimate given that the probes used often represent sequences duplicated and triplicated within an ES, as is discussed later.

A consequence of in situ expression is that the trypanosome has the potential to retain an ELC on an inactive telomere as a "lingering" ELC (L-ELC). (Young et al., 1983; Laurent et al., 1984b; Michels et al., 1984). L-ELCS have a high probability of expression early in infection following cyclical transmission through the tsetse fly (Laurent et al, 1984b).

1.7 TELOMERIC EXPRESSION SITES:.

1.7.1 ORGANIZATION.

A number of telomeric ES utilized by VSG genes expressed during bloodstream infection have now been subjected to detailed molecular analysis (Cully et al., 1985; Kooter et al., 1987; Alexandre et al., 1988; Pays et al., 1989a) in an attempt to characterize the elements necessary for VSG gene transcription and its control. These ES are complex polycistronic transcription units, containing at least seven non-VSG coding units, termed Expression Site Associated Genes (ESAGs) (Cully et al., 1985; Kooter et al., 1987; Pays et al., 1989a) which lie 5' to the 70bp repeat region, and are co-ordinately expressed with the VSG gene in an active ES. The relative order of the ESAGs within the individual ES appears to be conserved, though intergenic distances between respective ESAGs may vary, as may copy number. Each ESAG constitutes a member of an isogenic family comprising 14-25 members within the genome, (Cully et al., 1985, 1986). Their distribution appears limited to the larger trypanosome chromosomes; none has been observed on a mini-chromosome telomere (Alexandre et al., 1988; Cully et al., 1985; Gibbs and Cross, 1988; Kooter et al., 1987, 1988; Murphy et al., 1987).

The function of putative ESAG products remains unknown, though sequence analysis indicates that some may be membrane proteins (Cully et al., 1986; Kooter et al., 1987) judging by the presence of long hydrophobic stretches in the C-terminal region; one of these, ESAG 1, the only ESAG formally shown to have an

expressed polypeptide, does not appear to have a dimyristyl-phosphatidylinositol lipid anchor as found on the VSG (Cully et al., 1986).

Sequence analysis of ESAG 4 indicates that the predicted polypeptide shows sequence similarity to the active site region of yeast adenylate cyclase, a demonstrated transmembrane protein (Pays et al., 1989a).

The co-ordinate expression of ESAGs with the VSG gene has led to speculation that their products have a function in antigenic variation (Cully et al., 1985; Kooter et al., 1987). This view is difficult to reconcile with the detection of ESAG transcripts in procyclic trypomastigotes (Alexandre et al., 1988; Zomerdijk et al., 1990; S. Graham, pers. comm) and the absence of similar sequence elements in T. congolense (Murphy et al., 1987) and T. vivax (J.D. Barry unpub. obs.), both of which undergo antigenic variation. It is formally possible that T. congolense and T. vivax have diverged to such an extent from Trypanozoon trypanosomes, which all appear to possess ESAGS (Kooter, 1988), that the equivalent ESAG sequences are not readily detectable, even under low stringency Southern analysis. A further consideration is that the ESAG products may be compartmentalized within the cell, though there is no direct evidence for this. The presence of an endoplasmic reticulum signal peptide on only some ESAG products indicates that their products may have different cellular locations and thus different functions.

Individual ESAGs in different ESs show a degree of sequence conservation, 71% at the primary sequence level for ESAG1 and

67% amino acid sequence conservation outwith conserved amino acid substitutions (Cully et al., 1985, 1986). Direct analysis of ESAG products by in situ hybridization is precluded by their low concentrations within the cell, approximately 0.01% of the total cellular protein (Cully et al., 1986). An alternative strategy may be to utilize cellular fractionation with immuno-labelling techniques to locate specific ESAG products.

1.7.2 Transcription:- the RNA polymerase.

A feature of the ESs characterized to date is their large size, ranging between 40Kb and 60Kb (Kooter et al., 1987; Johnson et al., 1987; Alexandre et al., 1988; Gibbs and Cross, 1988; Pays et al., 1989a). All the genes within the ES appear to be under the control of a single promoter, and the whole polycistronic transcription unit is transcribed by an atypical RNA polymerase.

Eukaryotic RNA polymerases are classified as types I, II and III depending on their sensitivity to the toxin alpha-amanitin (Cochet-Meilac and Chambon, 1974). RNA polymerase I, which transcribes the rRNA genes, is generally insensitive to alpha-amanitin while RNA polymerase II transcribes protein coding genes and is highly sensitive to the toxin (rev. Sentenac, 1985). The VSG ES, however, is transcribed by a polymerase insensitive to alpha-amanitin, a novelty among eukaryotes, which indicates that the protein coding genes within the VSG ES are not transcribed by a typical eukaryotic RNA polymerase II enzyme.

The presence of sequences resembling the T. brucei ribosomal

DNA promoter (ie. for polymerase I) (White et al., 1986) within the boundaries of putative promoter regions for two VSG ESs (Shea et al., 1987; Alexandre et al., 1988) initially led to suggestions that the VSG ES was transcribed by an RNA polymerase I type enzyme. Detailed sequence analysis to the actual promoter region of a number of ESs (Pays et al., 1989a; Zomerdijk et al., 1990), refutes this suggestion as a consensus RNA polymerase I promoter is absent from the actual transcription start region.

Cloning of the large sub-unit genes for all the T. brucei RNA polymerases has identified two putative RNA polymerase II large subunit genes (Evers et al., 1989). Sequence analysis of these two RNA polymerase II large subunit genes reveals that the C- terminals of their predicted polypeptide differ from the analogous enzyme in other eukaryotes, where the C-terminal domain is conserved and has been ascribed a role in transcription factor and promoter recognition, and as a site for the phosphorylation which is necessary for transcribing activity (Cadena and Dahmus, 1987; Tower and Solner-Webb, 1988).

The trypanosome RNA polymerase II appears to lack the C-terminal heptapeptide tandem repeat ubiquitous among other eukaryotic type II polymerases (Evers et al., 1989). The presence of these repeats appears critical to the function of the enzyme (Nonet et al., 1987) though the reason for this remains unclear.

The possibility exists that the modified C-terminal domain of the trypanosome type II polymerase renders it insensitive to alpha-amanitin and that the VSG ES is transcribed by a type II

enzyme in accordance with the transcription of protein coding genes in all other eukaryotes. This might then also account for the novel promoter sequence employed by the VSG ES (Pays et al., 1989a; Zomerdijk et al., 1990).

Support for this comes from the observation that post-translational modifications of type II polymerase in mouse and Drosophila which confer increased alpha-amanitin insensitivity, map to the C-terminal of the large subunit (Greenleaf et al., 1979; Greenleaf, 1983; Bartolomei and Corden, 1987). One relevant point, however, is the presence of two distinct RNA polymerase II enzymes in T. cruzi, which does not undergo antigenic variation (Hodo and Hatcher, 1986). It is thus formally possible, in the absence of a functional analysis of trypanosome polymerases, that a modified RNA polymerase I is employed for VSG gene transcription. A number of distinct alpha-amanitin resistant polymerase activities have been observed in the course of the trypanosome developmental cycle (Tittawella, 1988) one of which could be envisaged as being specific for VSG transcription by means of subunit modification. Such an event occurs in the soil protist Acanthamoeba castellannii by means of a promoter binding factor and a modified RNA polymerase I (Paule et al., 1984; Bateman and Paule, 1986). Similar situations have been described for Maize rDNA transcription where an activated subform of RNA polymerase I is required (Tower and Solner-Webb, 1987). Purification of the relevant polymerase from trypanosomes and its detailed functional analysis will be required to settle this issue.

1.7.3 Transcription: trans-splicing and message capping.

All trypanosome mRNAs contain a common untranslated 39nt sequence at their 5' end, termed the "mini-exon" or "spliced leader" sequence (Van der Ploeg et al., 1982a; Boothroyd and Cross, 1982; Parsons et al., 1984; Freistadt et al., 1987, 1988), henceforth referred to as the mini-exon.

This mini-exon is derived from a 1.35Kb coding unit present at approximately 200 copies in the genome and found in tandem reiteration in two main clusters (De Lange et al., 1983; Michels et al., 1983; Nelson et al., 1983; Dorfman and Donelson, 1985) present on large chromosomes (Van der Ploeg et al., 1984b) and in a number of smaller clusters termed 'orphons' present elsewhere in the genome (Nelson et al., 1983). Homologous sequences have been found in related kinetoplastids (De Lange et al., 1984b; Nelson et al., 1984; Muhich et al., 1984) and analogous leader sequences have been detected in Caenorhabditis elegans (Krause and Hirsh, 1987) and Brugia malayi (Takacs et al.1988).

A 140nt mini-exon derived (med) RNA is obtained by RNA polymerase II transcription of the 1.35 Kb repeat in T. brucei (Kooter et al., 1984; Campbell et al., 1984; Milhausen et al., 1984) and this medRNA contains the 39nt mini exon at its 5' end. Addition of the mini exon to the 5' end of each mRNA appears to occur by a trans-splicing mechanism (rev. Borst, 1986) necessitated by the separate transcription of the mini-exon and the structural genes in other transcription units. Therefore, the mini exon coding unit need not be on the same chromosome as a given structural gene (Van der

Ploeg et al., 1984b; Guyaux et al., 1985) nor be transcribed by the same polymerase, as exemplified by the alpha-amanitin insensitive character of the VSG ES (Laird et al 1985; Kooter et al, 1984).

Trans-splicing of the VSG ES (Murphy et al., 1986; Sutton and Boothroyd, 1986; Laird et al., 1987) is thought to occur co-transcriptionally, judging by the inability to isolate large precursor transcripts from the VSG ES (Kooter et al., 1987) and from other polycistronic transcription units in T. bucei, such as that for the tubulin gene cluster (Milhausen et al., 1984; Imboden et al., 1985; Kimmel et al., 1985).

The trans-splicing reaction is considered analogous to cis-splicing reactions in higher eukaryotes, except that a branched intermediate, as opposed to a lariat structure, is formed on ligation and subsequent processing of the medRNA and structural gene transcript (Sutton and Boothroyd, 1986). The mechanism of trans-splicing has been reviewed extensively elsewhere (see Borst, 1986; Agabian et al., 1987).

It is the medRNA that provides each message from the ES with its cap structure (Laird et al., 1985; Lenardo et al., 1985). The trypanosome cap, however, is unusual in a number of aspects, containing N^7-Methylguanosine (Sutton and Boothroyd, 1988) and four modified nucleotides (Freistadt et al., 1987). All previously identified cap structures have either 0, 1, or 2-2'-0 methylated nucleotides (rev. Shatkin, 1976; Bannerjee, 1986). Such modifications, while apparently heterogeneous on the caps of mRNAs in other organisms (Bannerjee, 1980), appear homogenous in trypanosomes

(Freistadt et al., 1988).

A number of roles have been ascribed for cap structures including a primary function in influenza virus transcription (Bouloy et al., 1980), in vitro RNA splicing (Konarska et al., 1984; Krainer et al., 1984) and possibly 3' mRNA processing (Hart et al., 1985; Mowry and Steitz, 1987) in HeLa cells.

The novel structure of the trypanosome mRNA cap may be required for either trans-splicing of primary transcripts or translation of mRNAs. It will be interesting to ascertain whether primary transcripts are capped separately from that of the medRNA. An analysis of discontinuous transcription in C. elegans and a comparison with the trypanosome system is an obvious route to tackling these issues.

The observation that T. brucei mRNA can be translated in a heterologous system (Hoeijmakers et al., 1980), however, argues against a role for this novel cap structure being required to specifically bind trypanosome translation initiation factors.

Although the genes within the ES are co-ordinately transcribed at the same rate (Kooter et al., 1987; Johnson et al., 1987; Pays et al., 1989), the relative amounts of their stable mRNAs differ greatly. The VSG mRNA constitutes 5-10% of the total mRNA population (De Lange et al., 1984b) while ESAG 1 mRNA constitutes only less than 0.01% (Cully et al., 1985).

This is not an unusual feature of trypanosome transcription

units. For example, that containing the three genes for phosphoglycerate kinase (PGK) (Gibson et al., 1988) shows differential developmental mRNA regulation (Osinga et al., 1985; Gibson et al., 1988). This regulation occurs post- transcriptionally, though exactly how this is achieved is undetermined. Presumably differential stability or splicing of pre-mRNAs accounts for this.

1.7.4 Control of VSG expression:

VSG expression in vivo is generally mutually exclusive, with only one telomeric ES being active at a time. Observations of trypanosomes expressing more than one VSG on their surface coat in vitro (Baltz et al., 1986), indicate that the exclusive expression of VSG genes in vitro is a selective constraint, as mixed coated organisms stand an increased chance of immune detection and immune lysis, possibly mediated by the inability of heterospecific surface antigen interactions to provide a suitable barrier to serum components (Cardoso de Almeida et al., 1984), though this is a formally untested hypothesis. It is likely that double expressors arising during antigen switching, are a consequence of old and new VSG molecules mixing on their surface and are thus rapidly detected by host antibodies.

A number of models can be constructed to explain the apparent use of only one ES during bloodstream infection, though all suffer a series of drawbacks as detailed below:

1.7.4.1. Attachment Models.

1.7.4.1.1 Nucleolar compartmentalization model.

Shea et al., (1987) have proposed a model based on the exclusive attachment of an ES telomere to a site within the nucleolus. Attachment to such a site is proposed to access the ES to the necessary transcriptional machinery, principally RNA polymerase I, which these workers considered to be the polymerase used for VSG ES transcription, based on the presence of a sequence similar to a trypanosome rDNA promoter sequence within the region of a putative VSG transcription 'start' site (Shea et al., 1987). This model is attractive in a number of respects, namely (i) the nucleolus is a ready sink for RNA polymerase I, used for rDNA transcription, (ii) telomeres are known to form interactions with components of the nuclear membrane (rev. Lima-de-Faria, 1983), (iii) the active ES is under torsional stress, indicative of interaction with the nuclear membrane (Greaves and Borst, 1987), (iv) The active ES is unmodified and therefore recognizable, while all inactive ESs are modified (Bernards et al., 1984b; Pays et al., 1984). The model fails, however, to account for the existence of more than one ES being expressed at a time (Baltz et al., 1986) and does not allow for the possibility that the VSG ES is transcribed by a modified RNA polymerase II enzyme, given that the actual promoter for the ES contains no homology to the rDNA promoter (Pays et al., 1989a; Zomerdijk et al., 1990). The putative transcription starts containing a sequence similar to an rDNA promoter (Shea et al., 1987; Alexandre et al., 1988) appears to be artifactual.

1.7.4.2 Nuclear Attachment Model.

Mutually exclusive telomere expression is controlled by attachment to a site on the nuclear matrix (Bernards et al., 1984b). Only an unmodified telomere is considered as being competent to attach and hence become transcriptionally active. Attachment would once again account for the active ES being under torsional stress, as judged by its sensitivity to single stranded nucleases (Greaves and Borst, 1987), and stochastic modification of telomeric sequences could then be used to account for the observed patterns of antigenic variation (Bernards et al, 1984). Once again this 'attachment' model fails to account for 'double expressors' without invoking the presence of more than one attachment site, and it is unclear how modification and attachment are coupled, though their correlation with ES activity is strong. A further difficulty exists with respect to the correlation between modification and transcriptional inactivity of other ES telomeres, in that the promoter of the ES is 40-60Kb 5' of the VSG gene and appears to be unmodified (Zomerdijk et al., 1990). The telomere modification of inactive ESs ends approximately 1Kb 5' of the VSG gene, hence it is difficult to envisage it influencing promoter activity by long range chromatin interactions, especially in the light of the observation that the nucleosomal repeat of inactive and active transcription units is identical (Sloof et al., 1983; Greaves and Borst, 1987). It is still formally possible, however, that the promoter region of inactive ESs is modified in an as yet unknown manner. The major stumbling block with attachment models is that no experimental demonstration of attachment has been achieved (P. Borst, pers comm). An explanation for the torsional stress generated in the active ES is thus still awaited.

1.7.4.2 The Activator Model.

On transmission from the mammalian host to the tsetse fly vector VSG synthesis is rapidly shut off. In vitro observations indicate that this is concomitant with a specific decrease in the half-life of VSG mRNA from 4.5 to 1.2 hours and a 100 fold decrease in the steady state level of the message, which reaches undetectable levels within 24 hours (Ehlers et al., 1987). Inhibition of protein synthesis in bloodstream trypanosomes mimics the response observed in the procyclic form and indicates that VSG transcription is under the control of a short lived "activator", a positive regulatory protein or proteins (Ehlers et al., 1987). This agrees with the marked temperature dependence of VSG transcription which is rapidly lost with decreasing temperature concomitant with loss of RNA polymerase from its template (Alexandre et al., 1988). As to how an 'activator' or 'activators' select a given ES is unknown. It is tempting to speculate that such a factor(s) may influence polymerase specificity and characterisation.

A position effect relative to the genes coding for such a factor(s), where developmental regulation of the expression of these genes is controlled by virtue of their locus being adjacent to a cell-cycle gene or within its polycistronic transcription unit, may explain this. Preferential activation of a given ES could be due to subsequent use of an ES hierarchy determined by the mechanics of the switching process. This model is considered further in the Chapter 6.

If a number of such coding units exist they then could be used

to account for double expresser production and anamnestic re-expression (see general discussion).

1.7.4.3 The Repressor Model.

Pays and Steinert (1988) have proposed a model based on the existence of ESAGs in the VSG ES. It is proposed that one ESAG encodes a short lived negative control factor that prevents the activation of all non-transcribed ESs possibly by DNA modification. Activation in the absence of the repressor would then be stochastic, as the first activated ES modifies all the others. Depletion of the factor, for example, in the transformation to the procyclic form, would then allow demodification of all the other telomeres until VSG re-expression recommenced in procyclic cells. This model fails to account for all ESAGs being transcribed in the procyclic form, albeit as part of a non-VSG transcription unit (Alexandre et al., 1988; S.V. Graham and J.D. Barry unpub. obs.) and does not explain inter ES switching.

1.7.4.4 Mobile Element Models.

Use of a mobile enhancer/activator element for VSG ES activation could be used to account for many of the observations on antigenic switching. It might be envisaged that such an element would alter long range chromatin structure by insertion upstream of the VSG promoter and thus facilitate access of the transcriptional machinery. Target site availability might then determine the hierarchy of ES activation and the preferential use of only a few ESs (Liu et

al., 1985). Transposition of such an element would be required to be conservative to account for ES exclusivity, though duplicative transposition events can be invoked to explain double expressors. No physical change has, however, been observed in the vicinity of the VSG ES promoter when active and inactive (Zomerdijk et al., 1990), but this does not rule out the presence of such an element further upstream.

Analogies can be drawn between this model and mobile control element models in prokaryotes such as antigenic variation in Borrelia spp., Neisseria spp. and Salmonella typhimurium (rev. Seifert and So, 1988), where a rate of antigenic variation of 10^{-2}-10^{-4} per cell generation is compatible with that for trypanosomes (Turner and Barry, 1989). There has not yet been confirmation of any of these models.

1.8 Hierarchy of Sequential VSG expression: ES usage.

A necessary feature not generally covered in models attempting to account for VSG ES regulation, is the sequential order of VSG expression. Individual VSG genes appear to have a discrete probability for expression in the course of bloodstream infection, with telomeric BCs being preferentially expressed early in infection and chromosome internal BCs duplicated to a telomeric ES later in infection. (Laurent et al., 1983; Michels et al., 1984; Liu et al., 1985, Delauw et al., 1987). This sequence of expression is hierarchical and entails a non-stochastic use of individual ESs (Michels et al., 1984; Laurent et al., 1984; Lui et al., 1985). The reason for the preferential use of an individual ES is unclear. The resolution of this

problem will require detailed analysis of the environment and physical characteristics of a number of such sites. It is tempting to speculate that the position of individual ES loci governs the use of ESs, as it appears to do for the sequence of expression of VSGs. This feature is inherent in a mobile enhancer / activator model for ES activation where target site distribution for the putative mobile enhancer / activator governs the potential for activation of a given site.

VSG genes expressed late in chronic infection tend to be aberrant (Laurent et al., 1984) or composites of multiple partial conversion events (Roth et al., 1986) as the trypanosome's VAT repertoire becomes exhausted. This type of conversion is of obvious benefit in enhancing the evolution of the VAT repertoire. An additional outcome may be that the BCs for such VSG genes can be re-programmed within the sequence of antigen expression: should their ELC be retained as a L-ELC, since the L-ELC is preferentially expressed early in bloodstream infection (Laurent et al., 1984; Michels et al., 1984).

A position effect is not the only factor governing the place of an individual VSG within the sequence of expression. The hierarchy of expression also appears influenced by the composition of VSG genes' environment. Thus, for telomeric VSGs the extent of homology blocks between individual ESs, will influence the probability of conversion events, while both chromosome internal and telomeric BCs have 70bp common to their 5' flanks which might similarly influence the potential of a given VSG gene for expression. Indeed, the number of 70bp repeats 5' of a gene has been proposed to be directly correlated with the frequency of expression (Liu et al., 1983; Aline et

al., 1985a; Timmers et al., 1987).

Within the hierarchy of VSG gene expression is a separate sub-set, the Metacyclic-VSG genes (M-VSG) which show a number of unusual properties which are discussed below.

1.9 Metacyclic VSG genes.

Metacyclic form trypanosomes are the first to possess a VSG coat. VSG expression at this stage of the life-cycle is limited to a small, highly predictable subset comprising no more than 1-2% of the entire VAT repertoire (Crowe et al., 1983; Turner et al., 1988) and the genes for which appear to reside exclusively on the telomeres of the largest chromosomes (Cornelissen et al., 1985a, Lenardo et al., 1984; Delauw et al., 1987).

Expression of M-VSG genes is independent of that for B-VSG genes (Turner et al., 1985) and their activation within the salivary gland of the tsetse fly vector is stochastic (Tetley et al., 1987). The predictability of the M-VAT repertoire in a system geared for antigenic diversity, and its expression distinct from that of the B-VSG repertoire are puzzling. Direct molecular investigation of M-VSG expression is not readily achieved as the numbers of these organisms obtainable yield insufficient nucleic acids for analysis, a feature compounded by the antigenic instability during growth of populations from metacyclic cells. In vitro observations indicate that transition to bloodstream trypomastigotes occurs within 6 hours of entering the host (Brun et al., 1984). The proceeding chapters discuss in detail the difficulties in cloning and analysing M-VSG genes with the aims of;

(i) Defining the anatomy of a M-ES telomere.

(ii) Deliniating the M-VSG transcription unit if possible.

(iii) Comparing the M-ES with a typical B-ES.

(iv) Attempting to explain the stability, predictability and separate expression of the M-VAT repertoire.

MATERIALS AND METHODS:

2.1 Plasmids : Source and Description.

pTgB 221.1 : Bernards et al., 1984. B-ES clone.

pTg 221.8 : Kooter et al.,1987. "

pTg 221.9 : " "

pTg 221.12 : " "

pTcV 7.1-14 : Cornelissen et al., 1985a. cDNA clone for GUTat 7.1.

pMG 7.1-1 : " Genomic clone for GUTat 7.1.

pRibo : White et al., 1986. Plasmid clone containing the T.b.brucei ribosomal repeat unit.

pTbabT-1 : Thomashow et al., 1983. Plasmid clone containing the T.b.brucei alpha and beta tubulin genes.

pTIM : Swinkels et al., 1987. Genomic clone for the T.b.brucei TIM gene.

pUC 19 : Vierra and Messing, 1982. General purpose cloning vector.

M13 mp18/19 : Messing, 1983. Sequencing vector.

2.2 E.coli stocks.

Relevant Genotype and source.

NM538 : hsd R mcr A sup F. Kaiser and Murray, (1985).

NM539 : " " " (P2cox3). Kaiser and Muray (1985).

NM621 : hsd R mcr A mcr B rec D1009

thy$^+$ P1 transductant of SB 204. Whittaker et al.(1988).

DL491 : hsd R mcr A mcr B rec D sbc C sbc 201

Tn10::pho R derivative of NM621.Whittaker et al;. (1988).

JM101 : tsd$^+$ mrr$^+$ (lac -pro AB)/F' lac I^q lac Z /M15. Vieira and Messing, (1982).

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 lambda$^-$ lac I^q lac^z M15 lac y$^+$ sup E44 xyl 15. Source, D.Sheratt.

2.3 Buffers and Solutions :

TE :10 mM Tris, 1 mM EDTA

pH - as required by adjustment with HCl

NET :100 mM NaCl, 100 mM EDTA, 10 mM Tris, pH 8.0.

20xSSC :3M NaCl, 0.3M Na Citrate.

SM : 100 mM Tris - HCl, pH 7.5., 10 mM $MgS0_4$, 1 mM EDTA pH 8.0. 1 ml 1% gelatine L^{-1}.

10X Ligation Buffer : 0.675 M Tris - HCl, pH 7.5., 0.1 M $MgCl_2$, 0.15 DTT, 10 mM Spermidine, 10 mM ATP.

1X RT buffer : 0.05 M Tris - HCl, pH 8.3, 6 mM $MgCl_2$, 40 mM KCL, 1 mM DTT, 0.1 mg ml^{-1} BSA.

Denaturing Solution : 0.5 M NaOH, 1.5 M NaCl.

Neutralising solution : 3 M Na Acetate, pH 5.5., 1.5 M NaCl

L-Broth : 10g Bactotryptone (Difco), 5g Yeast extract (Difco), 10g NaCl, 1L ddH_20, final pH adjusted to 7.2.

L-agar as above plus 15g agar (Sigma, UK) per litre.

L-agarose as above but 1% agarose (Sigma type I) instead of 1.5% agar.

BBL-broth : 10g BBL trypticase (Baltimore Biological Labs),

5g NaCl, 1L ddH_20, pH 7.2.

BBL - agarose, as above plus 10g agarose (Sigma, type I).

NZCYM - broth : 10g NZ amine A, 5g NaCl, 5g Yeast extract, 1g Casamino acids, 1L ddH_20, pH 7.5.

- agarose, as above plus 15g agarose (Sigma, type I) and 2g $MgSO_4$.

LA/IPTG/X-GAL plates: L-AGAR base, 25ug ml^{-1} IPTG, 25ug ml^{-1} X-GAL, (+ desired antibiotic if required).

TBE: 0.089M Tris-borate, 0.089M Boric Acid, 0.002M EDTA

Blue Dextran tracking dye: 500ug.ml^{-1} sheared, salmon sperm DNA, 100ug ml^{-1} RNA, 1% Phenol red, 1% Blue dextran pH 8.0.

GTE : 50 mM Glucose, 25 mM Tris - HCl, pH 8.0, 10 mM EDTA.

RSB : 1 volume 0.1M, Tris, 9 volumes DMSO, 1 volume 1M LiCl.

METHODS.

2.4 Bloodstream Trypanosome Growth, Isolation and DNA Preparation.

Trypanosome clones stability expressing GUTat 7.1 belonging to the EATRO (East African Trypanosomiasis Research Organization) 2340 stock (Cornilissen et al., 1985a), or ILTat 1.22 derived from the EATRO 795 stock (Turner and Barry, 1989), were grown and isolated as described by Hajduk and Vickerman (1981) utilising CFLP mice and Wistar rats. DNA isolation was performed as detailed by Bernards et al (1981).

2.5 GROWTH, ISOLATION AND DNA PURIFICATION FOR PROCYCLIC TRYPOMASTIGOTES.

Approximately 10^7 bloodstream trypomastigotes isolated by exsanguination between days 7-9 post infection in a total volume of 0.1ml blood were diluted 100 fold in Hills medium (Brun et al.,1979) and incubated at 26°C for 48 hours, then transferred to $25cm^2$ plastic tissue culture flasks (Gibco, UK) and maintained at a concentration of $\sim 10^7 ml^{-1}$ until sufficient numbers were available for nucleic acid isolation.

2.6 RNA isolation: Bloodstream and Procyclic forms.

Trypanolysis was performed by resuspending purified trypanosomes in 10 volumes 3M Lithium chloride and 6M Urea, per volume of trypanosomes. High molecular DNA present was then sheared by sonication for 30 seconds, and the lytic mix incubated for a minimum of 5 hours at 0°C, to allow the RNA present to precipitate. RNA was recovered by centrifugation at 15,000g for 60 minutes at 0°C, using a Beckman JA 20:1 rotor, and resuspended in 1ml 10mM Tris, pH 7.5, 10mM EDTA and 0.1% SDS, pretreated with $50ugml^{-1}$ proteinase K for 5 minutes at room temperature. Resuspension of the pellet was achieved with very gentle pipetting, and incubated for 10 minutes at 37°C, followed by phenol: chloroform:iso-amyl alcohol extraction and ethanol precipitation, using 2.5 volumes of ice cold ethanol in the presence of 0.3M Sodium Acetate in the aqueous phase.

Precipitation was allowed to proceed for 30 minutes at 0°C

and the RNA harvested by centrifugation at 15,000g for 15 minutes at °C using a Beckman JA 20.1 rotor.

The RNA pellet was then stored at -20°C under 70% ethanol plus Sodium acetate to a final concentration of 0.3M.
It should be noted that all implements and vessels involved in the handling of RNA throughout the procedures described were DEPC treated to remove all unwanted RNase activity.

2.7 LAMBDA METHODOLOGY.

Construction of trypanosome genomic libraries in Lambda EMBL 4 was as described by Kaiser and Murray (1985). Screening, isolation, maintenance and DNA isolation for phage clones followed the methods of Benton and Davies (1977), with the modified protocol for phage showing poor growth detailed below.

2.7.1 PLATE LYSATE METHOD FOR POOR GROWTH PHAGE.

5 X 10^5 p.f.u. of phage were mixed at a M.O.I of one with E. coli cells following the method of Benton and Davies (1977), and incubated at 37°C for 30 minutes. 20mls of melted top agarose at 47°C was then added and mixed with the phage/E. coli, and poured onto a fresh 22cm^2 plate (Nunc, Gibco, UK) containing 200 mls hardened bottom BBL agarose, and allowed to set for 15 minutes at room temperature. The plate was then inverted and incubated at 37°C for 6-12 hours or until the plaques just touched and no patches of uninfected bacterial growth were visible.

The plate was then righted and swamped with 30mls SM and placed at 4°C for a minimum of two hours with gentle agitation. The SM was then harvested by pipetting and 5mls fresh SM used to wash over the plate, which was then allowed to drain by standing in a tilted position for 15 minutes and the SM collected as before. The harvested phage in SM were then treated as for liquid lysis preparations as described by Benton and Davies (1977).

2.8 PLASMID SUBCLONING.

2.8.1 Competent E. coli production.

A single colony of the desired E. coli strain was grown in 1ml LB at 37°C overnight, and used to infuse 100mls LB in a 250ml flask. The culture was then incubated at 37°C with shaking until just turbid, corresponding to an OD_{550nm} of 0.2-0.3. The culture was then cooled on ice for 30 minutes and the cells harvested in a pre-cooled Beckman JA.14 rotor at 2.5K, 4°C, 5 minutes.

The pellet was then dissolved in 6mls 100mM $CaCl_2$ at 4°C, and incubated for 30 minutes at 4°C. The suspension was then divided into 1ml aliquots and spun at full speed in an eppendorf centrifuge for 10 seconds at 4°C. The pellets were then gently redissolved in 100mM $CaCl_2$ (4°C) using a wide bore plastic pasteur pipette (Sterilin, UK) and incubated for 6-24 hours, 0°C. The cells were then ready for use or for freezing. Frozen stocks were made by addition of glycerol to 17.6%.

2.8.2 Transformation of competent E. coli.

Fifty microlitre aliquots of competent E. coli were incubated at 0°C for thirty minutes along with 10ng of the DNA to be transformed into the bacteria. The cells were then heat shocked by incubation at 42°C for 45 seconds and then placed on ice for a further 5 minutes. One millilitre of LB was then added and the cells incubated with shaking at 37°C for 60 minutes. 0.1ml aliquots from a serial dilution of the resultant culture were then plated out on;

(i) L agar/IPTG/X-GAL plates supplemented with any required antibiotic, and grown at 37°C overnight. Recombinant colonies were identified by a white coloration as opposed to blue for non-recombinants.

(ii) L agar plus desired antibiotic and growth with 37°C overnight. Recombinant colonies were identified by hybridization of colony lifts from individual plates with specific probes derived from the insert of the recombinant plasmid used for the transformation.

The status of the recombinant colony in both methods was further checked by mini-prep analysis and restriction digestion. Colonies so isolated were then used for plasmid isolation.

2.8.3 Construction of Recombinant Molecules.

The plasmid vector was prepared by digestion to completion with the requisite endonuclease to generate compatible ends for cloning the desired insert material and then phosphatased using calf intestinal

phosphatase following the manufacturers recommended conditions (Boehringer - F.R.G.). The phosphatased vector was then phenol extracted, chloroform extracted and ether extracted. The vector was then ethanol co-precipitated with the insert DNA at an insert: vector ratio of 3:1 and total DNA amount of ~0.5ug, then washed in 70% ethanol and air dried prior to resuspension in 8ul TE. 1ul 10X ligation buffer was then added and mixed gently with 1 unit T_4 DNA ligase.

(i) Ligation was then allowed to proceed for 12-16 hours at 15°C, and checked by agarose gel electrophoresis for the presence of ligation products.

(ii) In the situation of the insert DNA having been isolated in LMP agarose, the agarose was melted at 65°C for 10 minutes and then the desired volume for ligation removed and diluted X3 with ddH_20 at 37°C. This was then added to ethanol precipitated prepared vector along with 10X ligation buffer to a final concentration of 1X, and 1 unit T_4 DNA ligase. The total ligation volume was kept at a maximum of 20ul.

Ligation was then achieved by incubation at 4°C for 48 hours, to minimize the activity of any contaminating endonucleases present in the agarose from the insert DNA isolation. Success of the ligation was checked as previously.

2.9.1 Isolation of plasmid DNA.

Isolation of plasmid DNA was by the method of Birnboim and Doly (1977) with the modifications for small scale preparation described below.

2.9.1 Rapid plasmid miniprep.

A single bacterial colony was used to generate at 20 ml overnight culture by growth LB plus the relevant antibiotic. A 1.5ml aliquot was removed and spun in an eppendorf benchtop centrifuge at full power for 5 seconds to harvest cells which were then resuspended in 0.1ml GTE plus fresh lysozyme to 1mg ml^{-1}, by vortexing, then incubated on ice for 10 minutes. 0.2 mls 2M NaOH/0.2 % SDS was then added and mixed in by gentle inversion and incubated on ice for 5 minutes, followed by the addition of 150ul 5M Potassium acetate and further incubation on ice for 1 hour. The mix was then spun in an eppendorf centrifuge for 5 minutes at full power. The top 400ul of the supernatant was carefully pipetted off and split into two 200ul aliquots, which were then ethanol precipitated on ice for 30 minutes, centrifuged for 10 minutes (full power, eppendorf centrifuge) and the pellets redissolved in 0.1M Sodium acetate. The solution was then ethanol precipitated and spun as previously prior to recovery of the DNA by dissolution of the pellet in 30ul TE.

2.10 FRAGMENT ISOLATION.

Isolation of a designated DNA fragment for use as a radiolabelled probe, or for cloning purposes, was achieved by

complete digestion of the DNA with the restriction endonuclease(s) necessary to isolate the fragment, followed by horizontal agarose gel electrophoresis. The gel in all instances was made using LMP agarose (Sigma, UK) to a percentage in accordance with the need for suitable resolution of the desired fragment size. Visualization of the electrophoresised digestion products was performed using longwave UV transillumination on the gel, stained with 0.5 mg ml^{-1} ethidium bromide.

The fragment to be isolated was then cut from the gel, using a clean scalpel blade, in the minimum volume of agarose and transferred into an eppendorf tube for storage at -20°C.

2.11 PLAQUE LIFTS/COLONY LIFTS.

Plaque and colony lifts were performed using Hybond-N (Amersham) utilizing the manufacturers recommended protocol. In brief, duplicate plaque or colony lifts were taken for each plate, washed in 2X SSC, 5 minutes, denatured for 7 minutes on Whatmann 3MM filter paper, soaked in denaturing solution, neutralized X2, for 5 minutes on Whatmann 3MM paper soaked in neutralizing solution and washed in 2X SSC for 5 minutes. The filters were then air dried for 30 minutes and oven baked at 80°C for 1 hour prior to hybridization.

2.12 SOUTHERN BLOTTING.

DNA samples were subject to restriction endonuclease digestion under conditions recommended by the manufacturers

(Behtesda Research Laboratories) and using the manufacturers recommended buffer ("REACT" buffers - BRL). Digests were analyzed by horizontal agarose gel electrophoresis at $5Vcm^{-1}$ in Borate buffer. Visualization of the digestion products was achieved by staining the gel in $0.5mg\ ml^{-1}$ ethidium bromide in Borate buffer, for 30 minutes at room temperature. The gel was then destained in excess ddH_20 for 30 minutes at room temperature and photographed. Southern blotting of the DNA was performed essentially as described by Southern (1979), with modifications in accordance with the use of Nylon membrane as the blotting agent.

The gel was denatured for 30 minutes at room temperature in 0.5M NaOH, 1.5 NaCl and neutralized for 30 minutes in 1.5M NaCl 3M Potassium Acetate, pH 5.5). The gel was then placed, inverted, on a perspex blotting platform covered with Whatmann 3MM filter paper, whose ends formed wicks in a tank of 20X SSC which acted as the transfer solution. A cut to size piece of Hybond-N nylon membrane (Amersham PLC) was then placed on the gel, avoiding trapping air bubbles, and covered with two sheets of Whatman 3MM filter paper followed by two absorbent nappies (Boots, PLC), a perspex plate and a 1Kg weight. Transfer was allowed to proceed overnight after which the nylon filter was washed in 2X SSC, air dried for 30 minutes at room temperature and either oven baked at 80°C for 1 hour or longwave UV baked for 3 minutes.

2.13 NORTHERN BLOTTING.

RNA samples were freeze dried and resuspended in 25ul fresh RSB, vortexed to aid dissolution and heated to 70°C for 10 minutes,

then snap cooled in ice prior to analysis by horizontal agarose gel electrophoresis of 5V cm^{-1}. Electrophoresis was performed using 1% agarose gels in formaldehyde gel buffer. Visualisation of the electrophoresed RNA samples or markers was achieved by soaking in excess ddH_20 for 1 hour, staining in 10ug ml^{-1} Acridine orange for 10 minutes and destaining in excess ddH_20 overnight prior to visual analysis under longwave UV irradiation.

Blotting was performed as for the Southern blot, using 20X SSC to transfer the RNA, and Hybond-N as the blotting membrane, followed by longwave UV fixation for 3 minutes.

2.14 HYBRIDIZATION.

^{32}P labelled probes were prepared by Hexanucleotide random priming as described by Feinberg and Vogelstein (1983), with specific activities of more than 10^8 cpm/ug being obtained. Hybridization of Southern blots and Northern blots were performed according to manufacturers recommendations (Amersham, PLC).

Post hybridization washing was at the desired temperature and SSC concentration, once again in accordance with the manufacturers recommendations. Autoradiography was performed using X-OMAT S film (Kodak) with incubation at -70°C under two intensifying screens. The length of autoradiographic exposure varied in accordance with the experiment, and is indicated separately with each hybridization described subsequently.

2.15 'REVERSE NORTHERN': Reverse transcription of Poly[A^+] RNA.

Poly[A^+] RNA was prepared from total RNA by the method of Graham and Birnie (1986). 0.5ul of a 1mg ml^{-1} poly[A^+] RNA stock was then incubated in 5ul 5X RT buffer plus 5ul oligo dT (1mg/ml^{-1}) and 6.5 ul ddH_20, for 10 minutes at 65°C and rapidly cooled on ice. To this 1ul was added, from 10mM stocks of dATP, dGTP and dTTP plus 2.5ul ^{32}P-dCTP and 2.5ul Reverse transcriptase (Boehringer, FRG).

The reaction was incubated at 42°C for 1 hour and the labelled cDNA separated out over a 'Nick' column (Amersham, UK). The labelled cDNA was tracked using a blue dextran tracking dye and eluted in a minimum volume of TE prior to hybridization. Hybridization using the labelled cDNA as probe was as described for Southern analysis.

2.16 NUCLEAR 'RUN-ON'.

Trypanosome nuclei preparation and run-on reactions were performed as described by Kooter et al., (1987). Inhibition of RNA polymerase activity during run-on was through use of the inhibitor alpha-amanitin (Sigma, UK), as described by Kooter and Borst (1984).

2.17 SEQUENCING REACTIONS.

Fragments to be sequenced were subcloned into M13 mp18 and mp19, and sequencing reactions carried out utilizing a modification of the ^{35}S-dideoxy method of Sanger et al. (1977), with a commercial sequencing kit, 'Sequenase' (U.S. Biochem.). All

reactions and conditions were carried out in accordance with the manufacturers recommendations.

Sequencing products were analysed by vertical polyacrylamide gel electrophoresis, using 15% gels run at 65V, 55°C in borate buffer. The length of run was determined by the migration of the individual 'Sequenase' tracking dyes relative to the end of the gel, with a short run corresponding to the first dye front reaching two thirds down the gel, a medium run corresponding to the first dye just running off the gel and longer runs by the similar migration of the second dye front.

The visual analysis of sequencing reaction products was achieved by fixing the gel in 10% methanol/ 10% acetic acid for 20 minutes at room temperature, washing briefly by immersion in water, and blotting the gel onto 3MM Whatmann filter paper. The gel was then vacuum dried at 80°C for 1 hour prior to autoradiography, without intensifying screens, overnight at room temperature.

Chapter 3.

Cloning of the GUTat 7.1 and ILTAT 1.22 M-VSG Genes

Introduction :

Trypanosoma brucei rhodesiense, in common with other African salivarian trypanosomes, displays the ability to undergo antigenic variation of its surface coat (reviewed, Barry, 1986). The surface coat is constituted entirely by a single species of VSG which provides the parasite with a physical barrier to non-specific immune attack and endows the population with the ability to escape specific immune lysis, through the appearance of individual trypanosomes of novel VSG.

Acquisition of mammalian infectivity and the appearance of the VSG surface coat are concomitant and first occurs in the tsetse fly vector, with the development of the metacyclic form (Vickerman, 1969).

This form constitutes a transient developmental stage, being amitotic and free swimming within the salivary gland lumen. Upon entering the host the metacyclic trypomastigotes rapidly develop into bloodstream forms, probably within as little as 6 hours (Brun et al., 1984). Metacyclic trypanosome populations have a number of features distinctive from, and contrasting with, those of bloodstream trypanosome populations. While bloodstream stages have the capacity to express, one at a time, between 10^2-10^3 separate VSGs

(Capbern et al., 1977; Van der Ploeg et al., 1982b) only a small distinct subset of these is available for expression by the metacyclic stage.

In T. brucei this comprises no more than 27 VATs as defined by antibody analysis (Turner and Barry, 1989). In one incomplete study, a pool of 16 monoclonal antibodies (McAbs) has been used to identify most or all of the M-VATs in a stock of T. brucei (Esser and Schoenbechler, 1985), while in T. congolense the entire M-VSG repertoire of one stock numbers 12 (Crowe et al., 1983). Thus it is apparent that the M-VAT repertoire is defined by only 1-2% of the total repertoire.

Given the dynamism of the trypanosome genome, the M-VAT repertoire has one very unusual characteristic : it is conserved within a given serodeme (Barry et al., 1979, 1985; Hajduk and Vickerman, 1981). However, it is not totally conserved, displaying a low level of change with time (Barry et al., 1983). Though the metacyclic stage of the life cycle is transient, M-VSG gene expression persists within the host bloodstream for the first few days following fly bite, where M-VSG interswitching is prevalent and seemingly preferential (Barry et al., 1985; Turner and Barry, 1989).

A number of lines of evidence indicate that M-VSG gene expression is distinct, and separately controlled, from B-VSG gene expression. The principal evidence for this comes from experiments showing that the composition of the M-VAT repertoire is uninfluenced by the VAT composition of the trypanosome population initially ingested by the tsetse fly.

In transmissions where a clonal population is introduced to the fly, the predominant VAT in ingested trypanosome population, the I-VAT, following cessation of VSG synthesis in the fly gut, is re-expressed at a high level in early parasitaemias but is not expressed in the metacyclic population (Hajduk and Vickerman, 1981). This re-expression is referred to as the 'anamnestic' effect, whereby the trypanosome appears to retain a memory of the last expressed bloodstream VSG gene. This observation, coupled with the prevalence of inter M-VSG switches in early parasitaemias is indicative of separate control for M- and B-VSG expression, the former apparently being dominant at this period. Molecular evidence to explain the 'anamnestic' effect lends support to the presence of a dichotomy of control of M- and B-VSG gene expression.

In vitro transformation of bloodstream to procyclic stage, which is thought to mimic initial infection in the fly, results in the last expressed ELC being retained in the genome, though it is transcriptionally silent (Overath et al., 1983).

Furthermore, when an ELC is retained in the genome after a non-duplicative switching event in bloodstream trypanosomes, this ELC, known as the lingering ELC (L-ELC), is preferentially reactivated in subsequent switching events (Laurent et al., 1984b), presumably by virtue of already being in an expression site. There is a high potential for inter-expression site switching during bloodstream antigenic variation (Liu et al., 1985). Thus, if an ELC such as the I-VSG gene is already at a bloodstream VSG gene expression site it should be preferentially reactivated when the

bloodstream expression system recommences. Indeed, Delauw et al. (1985) have followed an I-VSG gene through cyclical transmission in the tsetse fly and have shown the I-VSG gene retained as a transcriptionally silent L-ELC in procyclic cells from the fly gut. The I-VSG gene was demonstrated subsequently to be transcriptionally reactivated without duplication in the first parasitaemia following cyclical transmission.

Further evidence comes from the observation that even when the I-VAT is a normal component of a given stock's M-VAT repertoire, that is, it is ingested by the fly while being expressed by the bloodstream mechanism, it does not significantly elevate the level of that given VAT within the metacyclic population, indicating that the M-VSG gene expression and 'anamnestic' re-expression are unlinked (Turner et al., 1986). This apparently occurs whether or not the M-VSG gene is being expressed via an ELC or in situ in the ingested population (K. Matthews and J.D. Barry, personal communication). Indeed, antibody analysis of the activation of M-VSG genes in the fly indicates it to be a random and polyclonal event, coinciding with the initiation of VSG synthesis; again supporting the hypothesis of a separate M-VSG gene subset (Tetley et al., 1987).

Direct molecular analysis of M-VSG gene expression, necessary to gain an insight into the control of antigenic variation, is hampered on a number of fronts.

(i) There is no reliable reverse genetics system for trypanosomes.

(ii) The metacyclic stage of the life cycle is transient and rapidly develops to the bloodstream stage following fly bite. (iii) Individual

tsetse flies only extrude a few thousand trypanosomes during a bloodmeal ($>3 \times 10^3$/proboscis - Harley et al., 1966) far too few to permit investigation at the DNA or RNA level. Two indirect strategies, however, have been used to investigate the nature of M-VSG gene expression. The first of these involves the use of large numbers of flies to infect large numbers of mice to produce early parasitaemias, from which trypanosomes are harvested typically after 5 days, with antibody mediated lysis then being used to enrich for a given VAT or VATs. Using this approach Lenardo et al. (1984,1986) have analysed two M-VSG genes, both of which were present as single copies in the genome, situated on the telomeres of the large chromosome class. Surprisingly, these genes were not abutted by a 5' barren region and, possibly as a direct consequence of this precluding ELC formation, appeared to be expressed in situ. The polyclonal origin of the trypanosomes expressing a given VAT used in such an analysis, however, does not rule out the possibility that there were multiple ELCs present, all of differing fragment sizes and hence at too low a level to detect by standard Southern analysis. Such a scenario also entails other sequences 5' to the VSG gene acting as a starting point for the conversion event during switching, as these workers have also observed for one of their M-VSG genes (Lenardo et al., 1986).

A second strategy employs cloning single copy M-VSG genes still being expressed in early bloodstream parasitaemias. This approach overcomes the problems associated with the polyclonal origins of trypanosomes used in the first system. Such clonal propagation and analysis of M-VAT expressing trypanosomes gives some data contrasting with those of the previous study. While the

M-VSG genes again appear on the telomeres of the largest chromosome class, they have 5' barren regions and appear to utilize an ELC expression mechanism in early parasitaemias (Delauw et al., 1987). Whether the latter mechanism is a true reflection of M-VSG expression or is a consequence of the indirect system of analysis used not reflecting accurately M-VSG expression, cannot be determined by these observations.

Neither system employed gives an indication of whether there is anything unique on the M-VSG gene telomeres to distinguish them from B-VSG gene expression sites, or how and why this subset of VSG genes is conserved and its expression predominant at this stage of the life cycle. In order to gain some insight into these problems I have attempted to clone BC M-VSG genes and some of their 5' environment, to allow for a comparison with previously characterized telomeres harbouring B-VSG genes. The strategy employed was to clone the BC VSG genes from a genomic library constructed in lambda EMBL 4. The use of the lambda vector allows the isolation of large tracts of DNA upstream of the VSG gene, thus increasing the likelihood of jumping over any 5' barren region. The large numbers of recombinants obtainable using such a vector would also increase the chances of isolating recombinants from a region with little availability of 3' or 5' cloning sites, due to the repetitive nature of the sequences flanking the VSG gene on either side. Isolation of a M-VSG gene and its 5' environment would then allow a more direct approach in comparison of these genes with those expressed in the bloodstream and thus provide some indication as to the maintenance, control and propagation of this specific VSG subset.

Results :

3.1 Construction of a genomic library in lambda EMBL 4.

A library of the EATRO 2340 trypanosome genome was constructed in lambda EMBL 4 following the methodology of Kaiser and Murray (1985), utilising insert fragments in the size range 18 - 22 Kb and E. coli host strains NM 538 and NM 539.

In order to establish how well the library represented the T.b.rhodesiense genome, prior to screening for sequences of interest, the primary library was screened with genomic DNA probes for the single copy gene Triose phosphate isomerase (TIM), and the multicopy rDNA coding unit. A primary library was chosen as opposed to an amplified library as the latter may not have been as representative of the typanosome genome, given that individual phage clones may amplify differentially. For TIM, 24,000 pfu were screened by plaque hybridization , while only 3000 were screened for the multicopy sequence. TIM should be present at ~ 2 copies per 20000 p.f.u., while the 110 copies of the ~18.5 Kb of rDNA in the genome should be present in ~30 of the 3000 screened , as determined by the equation $N = \ln(1-P) / \ln(1-f)$, (Clarke and Carbon, 1975), were P is the probability of finding a given unique sequence represented in N plaques, and f is the average library insert size expressed as a fraction of the haploid genome. The equation can be extended for multicopy sequences if one knows the total copy number of the sequence.

The screening for TIM yielded 2 positive signals (data not shown), indicating that at least for this single copy sequence, the

library was representative. The rDNA probe identified 33 positive clones (data not shown), again lending support to the view that the library was representative of the trypanosome genome. This is only a first approximation, as it was not verified that these clones were bona fide positive plaques and whether or not their numbers reflected differential phage viability.

3.2.1 Screening the EATRO 2340 T.b.rhodesiense genomic library for M-VSG genes.

The EATRO 2340 genomic library was screened extensively using cDNA probes for M-VSG genes which had previously been subjected to a brief physical mapping analysis (Cornelissen et al., 1985a). The four genes, GUTats 7.1, 7.2, 7.13 and 7.15, were found to reside on the telomeres of the largest chromosome class in the trypanosome genome, and appeared to be activated in bloodstream trypanosomes by an ELC mechanism (Cornelissen et al., 1985a). GUTats 7.2, 7.13 and 7.15 were single copy in the genome as the EATRO 2340 stock employed for the library construction, while GUTat 7.1 was double copy.

Successful cloning of any of these VSG genes was dependent on the availability of Sau 3A sites 3' and 5' to each gene, this was thought to be limited due to the presence of repetitive sequences on their 3' and 5' flanks. It was hoped that use of large insert DNA in making the library would resolve the difficulty of cloning upstream of the 70 bp repeat region 5' to the VSG gene, and thus finding suitable cloning sites. To help overcome this problem further, large numbers

of recombinant phage were exhaustively screened in both the amplified and unamplified libraries. Screening was performed on duplicate filters using specific ^{32}P radio-labelled cDNAs, with hybridization being carried out for 16 hours at 68°C. Post hybridization washing was performed to a stringency of 0.1 X SSC, 68°C, prior to autoradiography. While negative plaques on the filters acted as an inherent negative control for the screening, 5ng samples of each cDNA were individually spotted onto Nylon membrane and were included in the hybridization as a positive control : in each case the probe did hybridize to this control DNA. Post hybridization washing of the control strip was as for the plaque containing filters. Compiled results for the screenings are presented in Table 3.1.

Three positive clones were identified, picked and plaque purified. The individual purified phage were designated lambda MT 7.1-1, 7.1-2 and 7.1-3, MT being an abbreviation of metacyclic telomere. The extreme scarcity of M-VSG gene clones within the library, despite initial testing having shown the library to be representative of the trypanosome genome (section 3.1), is indicative of the difficulty of cloning telomeric genes by standard methods. The presence of a Bam HI site already in the GUTat 7.1 VSG gene should have facilitated easier cloning of this gene, providing a suitable 5' cloning site was available, as was the case. Obviously there must be other considerations to account for the inability to isolate other M-VSG gene clones, such as phage stability and cloning site availability.

3.2.2 Cloning of the ILTat 1.22 BC VSG gene.

In an attempt to remedy the problems encountered in the exhaustive screening of the EATRO 2340 library to isolate M-VSG genes, a second size selected library was constructed in lambda EMBL 4 and plated on the E. coli host NM 621, chosen in the hope of alleviating problems associated with phage stability that may have attributed to the cloning difficulties encountered with the EATRO 2340 library. The genotype of NM621, unlike NM539 is recD$^-$, mcrB$^-$, and thus more conducive to cloning repetitive and methylated sequences respectively, which can lead to phage instability (see discussion).

Library construction again followed the method of Kaiser and Murray (1985) with 18-22Kb fragments once again being chosen with which to make the library. NM621, however, is not a P2 lysogen, and hence the SPI genetic system for enhanced selection of recombinant phage on NM 539 could not be used. The loss of the SPI selection system was balanced by the potential benefits of the NM621 genotype, and the recovery of the recombinant phage enhanced by physical removal of the vector arm linker fragments during vector preparation (Kaiser and Murray, 1985).

A second feature incorporated into the library construction was the use of an EATRO 795 trypanosome stock as opposed to EATRO 2340. The EATRO 795 stock offered the advantage of a more amenable system for transcriptional analysis of any M-VSG genes cloned from this stock, as is described in Chapter 4.

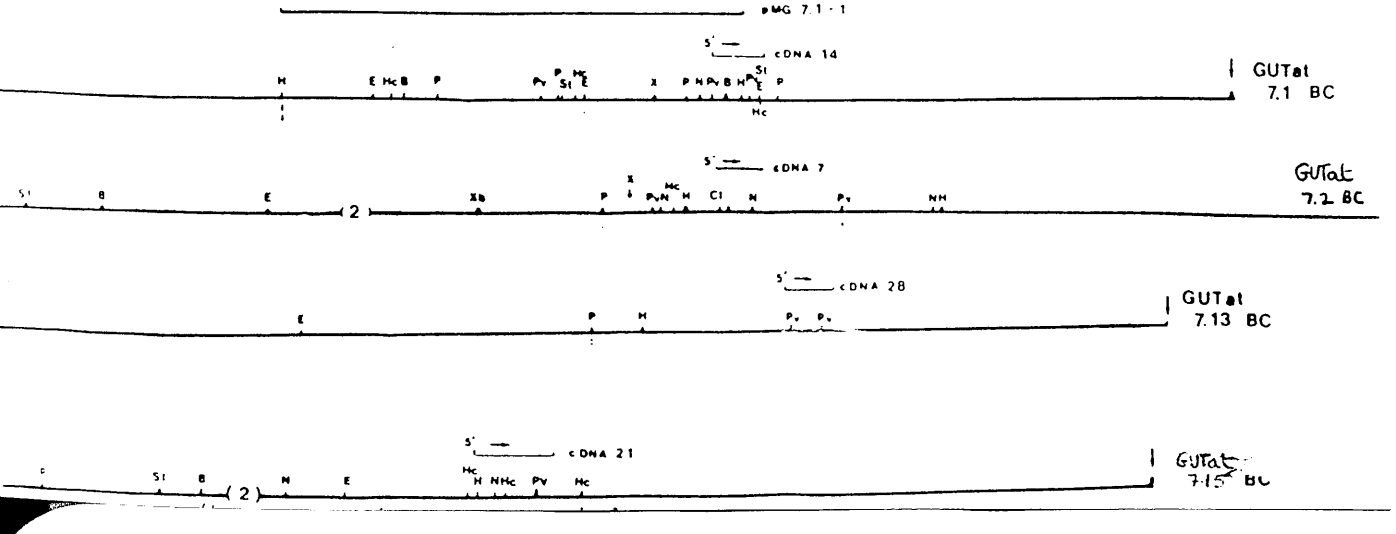
The maps opposite show the BC telomeres for the GUTat 7.1, 7.2, 7.13, 7.15 VSG genes (after Cornelissen et al, 1985a). The full length cDNAs used as probes to construct these maps are indicated above each respective telomere. Screening of the lambda library was performed using fragments derived from the 5' end of each cDNA as detailed in the text. The genomic clone for the GUTat 7.1 BC telomere, pMG 7.1-1 is delineated above the GUTat 7.1 BC telomere map. An updated version of this map is presented and discussed extensively in section 3.4.2.

Table 3.1.

Screening of the EATRO 2340 lambda library for M-VSG genes GUTats 7.1, 7.2, 7.13 and 7.15

	Total PFU Screened		Positive Recombinants	
GUTat	Unamplified	Amplified	Unamplified	Amplified
7.1	1.4×10^6	3×10^5	2	1
7.2	3×10^6	3×10^5	0	0
7.13	1.4×10^6	3×10^5	0	0
7.15	1.3×10^6	3×10^5	0	0

Probes : 7.1 pTcV 7.1-14 (Cornelissen et al., 1985a)
7.2 pTcV 7.2-7 "
7.13 pTcV 7.13-26 "
7.15 pTcV 7.15-21 "



To get a first approximation of how representative a library had been constructed, it was screened for the known single copy gene triose phosphate isomerase (TIM) and the multicopy sequence of the rDNA. Screening was performed as described in section 3.1.

The TIM probe yielded two positive signals from a total of 3×10^4 p.f.u., which is within reason given that this number of phage covered the genome just under three times. Screening 2060 plaques with rDNA probe yielded 25 positive signals. Given that there are 110 copies of the 18.5Kb rDNA repeat sequence in the genome, this figure is a reasonable first approximation of the library being representative of the EATRO 795 genome; screening for the M-VSG genes was thus commenced.

110,000 p.f.u. corresponding to 11 genome equivalents, were initially screened in duplicate, using the same for cDNA probes as detailed in Table 3.1. Hybridizations were performed at 68°C for 16 hrs, with post-hybridization washing to a stringency of 0.1 x SSC, 68°C.

Two positive clones were isolated, both corresponding to ILTat 1.22, the EATRO 795 equivalent of GUTat 7.1, which was single copy in this stock's genome. These clones, designated lambda MT 1.22 A and lambda MT 1.22 B, were picked and plaque purified.

Subsequent miniprep and restriction analysis revealed each to be an independent isolate of recombinants carrying the same insert (data not shown). One clone only, lambda MT 1.22 B was selected for further analysis. Failure to find clones of the other M-VSG genes

sought is examined in the discussion of this chapter.

3.3. Physical mapping of the M-VSG clones.

To enable analysis of the composition of the telomeres harbouring the GUTat 7.1 and ILTat 1.22 VSG gene copies, physical maps were constructed for each clone. The map of lambda MT 7.1-3, however, is taken as identical to lambda MT 7.1-1, as these appear to be independent isolates of the same clone on the basis of restriction analysis (data not shown).

The map for lambda MT 1.22B clone was constructed in most detail as the single copy nature of the ILTat 1.22 BC telomere in the EATRO 795 stock employed for later investigations, allowed for greater ease of analysis.

3.3.1.

Physical mapping of lambda MT 7.1-1 and lambda MT 7.1-2..

Mapping of the GUTat 7.1 clones was achieved by a combination of restriction digest pattern analysis and by partial digestions in conjunction with Southern analyses, using end labelled vector arms as probes. The latter strategy was employed to determine, for a given enzyme, the order of sites relative to each of the vector arms. Accurate fragment sizing and cross checking of site order was achieved by analysis of complete single or double digest patterns (see Figure 3.3).

3.3.1.1. Construction of the basic physical maps for lambda MT 7.1-1 and lambda MT 7.1-2.

Bam HI and Eco RI partial digest series of each lambda clone were analysed with uncut controls by 0.3% agarose gel electrophoresis and Southern blotted prior to hybridization with the respective short or long arm fragment of EMBL4 for 16 hrs at 68°C. Blots were washed post-hybridization to a stringency of 0.1X SSC, 68°C. Figure 3.3.1.1 and 3.3.1.2 present representative results of this analysis for lambda MT 7.1-1. and lambda MT 7.1-2.

In figure 3.3.1.1 the smallest fragment detectable by the probe is 17Kb, which corresponds to the vector 9.2 Kb short arm plus insert DNA, as the Bam HI cloning site at this arm has not been reconstituted. Thus the insert section of this fragment must be 7.8Kb. The probe then detects bands increasing in size from 23Kb

to 26.8Kb and finally an uncut phage band at 45.8Kb. Thus by subtraction of the sizes of the phage arms the insert must be 17.6Kb (see map Figure 3.3.1.3). The lambda MT 7.1-2 analysis is shown in figure 3.3.1.2. In this clone the Bam HI site at the vector short arm is reconstituted, hence the probe detects itself as the smallest band in the partial digest series followed by a 9.4Kb fragment. The latter is not well resolved but is most apparent from lane 6 and 7 (see ethidium stained gel, panel A). The remaining fragments size in increasing order at 13.2Kb , 19.2Kb , 26.8Kb and 45.8Kb , the last being the size of the uncut phage. The deduced map is shown in Figure 3.3.1.3.

3.3.2. Physical mapping of lambda MT 1.22.

The lambda MT 1.22B clone was mapped in detail by a combination of the analysis of restriction digest patterns and Southern blotting and hybridization, utilizing specific fragments from the cloned insert as probe. The data from the restriction analyses are summarized in Figure 3.3.2.1 and Table 3.3. Panel A shows the derived restriction map of the clone. Panel B comprises representative Southern analyses from which the map in Panel A and data in Table 3.3 have been derived. Hybridizations were performed at 68°C for 16 hours with post-hybridization washing to 0.1X SSC, 68°C. Restriction mapping gives an initial size estimate for the insert in lambda MT 1.22 B of 17.75Kb.

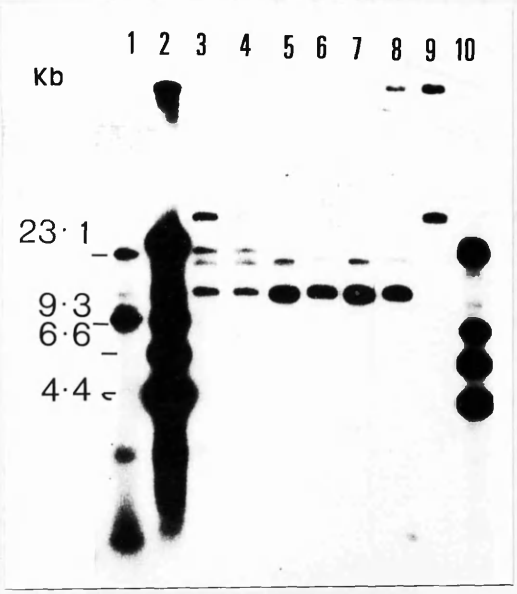
Probe A, the Pst I/Hin dIII fragment from the cDNA clone

pTcV 7.1-14, detects fragments all of which are predicted from the map of the plasmid clone pMG 7.1-1 (Cornelissen et al., 1985a), plus a small region stretching 50bp downstream of the region defined by the plasmid clone (ref: Figure 3.3.2.1, Panel B for autoradiograph). It furthermore defines the region of the clone containing the ILTat 1.22 VSG gene.

One anomaly exists with Probe B (see Panel C), in that it failed to detect an 8Kb Eco RI fragment that overlaps the 5' Bam HI site of Probe B (see map Panel A). This may reflect poor transfer of larger fragments in the gel during Southern blotting, with available probe being 'mopped up' by more abundant smaller fragments during hybridization as a consequence. It is difficult to judge the likelihood of such a scenario.

The lambda MT 1.22B map indicates that the single copy ILTat 1.22 BC telomere is equivalent to that for GUTat 7.1-2, on the basis of the Bam HI and Eco RI sites in the respective lambda clones.

Figure 3.3.1.1 Bam HI partial digest mapping of lambda MT7.1-1.



Lane 1.Lambda Kpn I.

2.Lambda Hin dIII.

3.Lambda MT 7.1-1 x 0.03 units Bam HI.

4. " " 0.06 " "

5. " " 0.12 " "

6. " " 0.25 " "

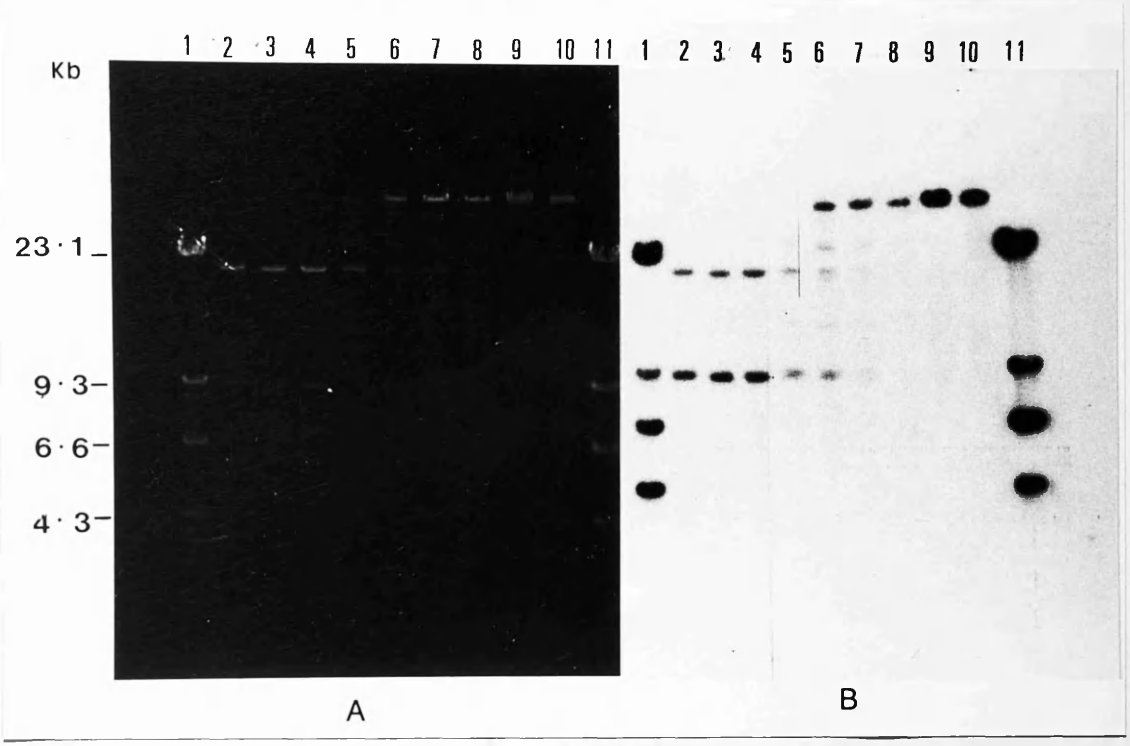
7. " " 0.5 " "

8. " " 1.0 " "

9. " " 0.0 " "

10. Lambda Hin dIII.

Figure 3.3.1.2. Bam HI partial mapping of the lambda MT7.1-2



Lane 1. Lambda Hin dIII.

2. Lambda MT 7.1-2 x 2.0 units Bam HI.

3. " " 1.0 " "

4. " " 0.5 " "

5. " " 0.25 " "

6. " " 0.12 " "

7. " " 0.06 " "

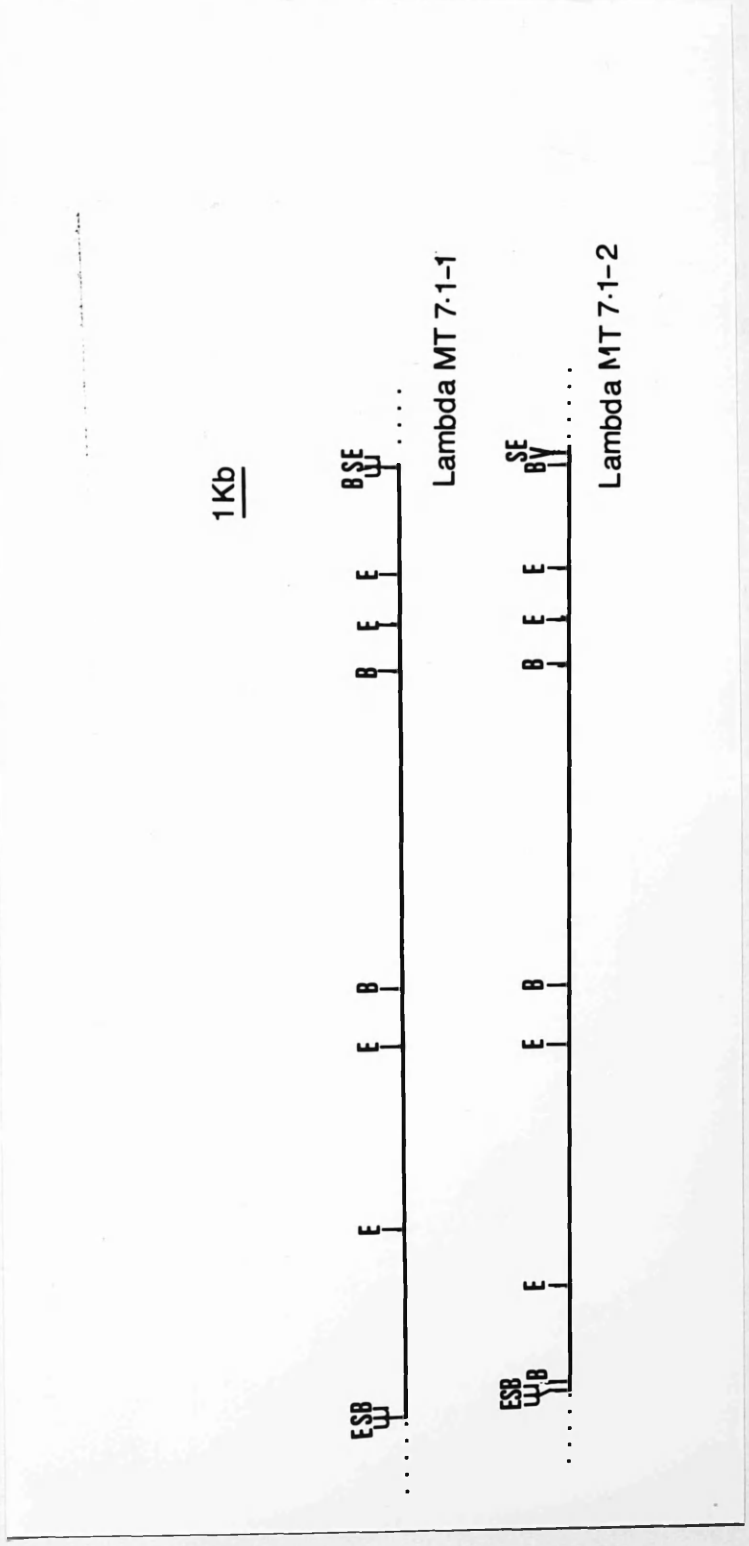
8. " " 0.03 " "

9. " " 0.015 " "

10. " " 0.007 " "

11. Lambda Hin dIII.

Figure 3.3.1.3 Restriction maps of lambda MT 7.1-1 and 7.1-2.

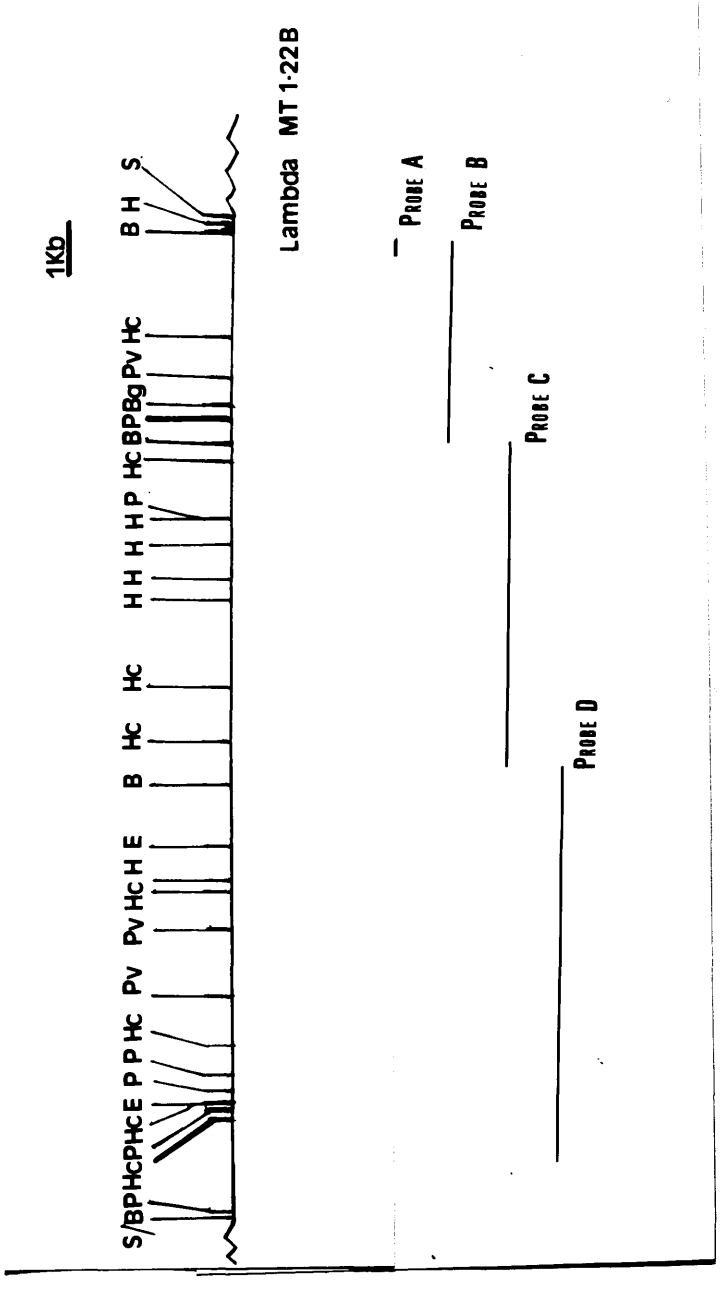


3.4. Genomic mapping of the GUTat 7.1 telomeres in EATRO 2340.

An interesting feature of the lambda MT 7.1 maps is that they differ from that of Cornelissen et al. (1985a) with respect to the Eco RI site immediately 5' to that carrying the VSG gene (Figure 3.3.1.3). Instead of the expected 2.5Kb Eco RI fragment present in the genomic map, in the lambda MT clone maps there is a 1Kb Eco RI fragment. In order to verify the fidelity of the lambda MT maps with respect to their genomic locations, it was decided to construct basic genomic maps for each telomere.

Figure 3.3.2.1

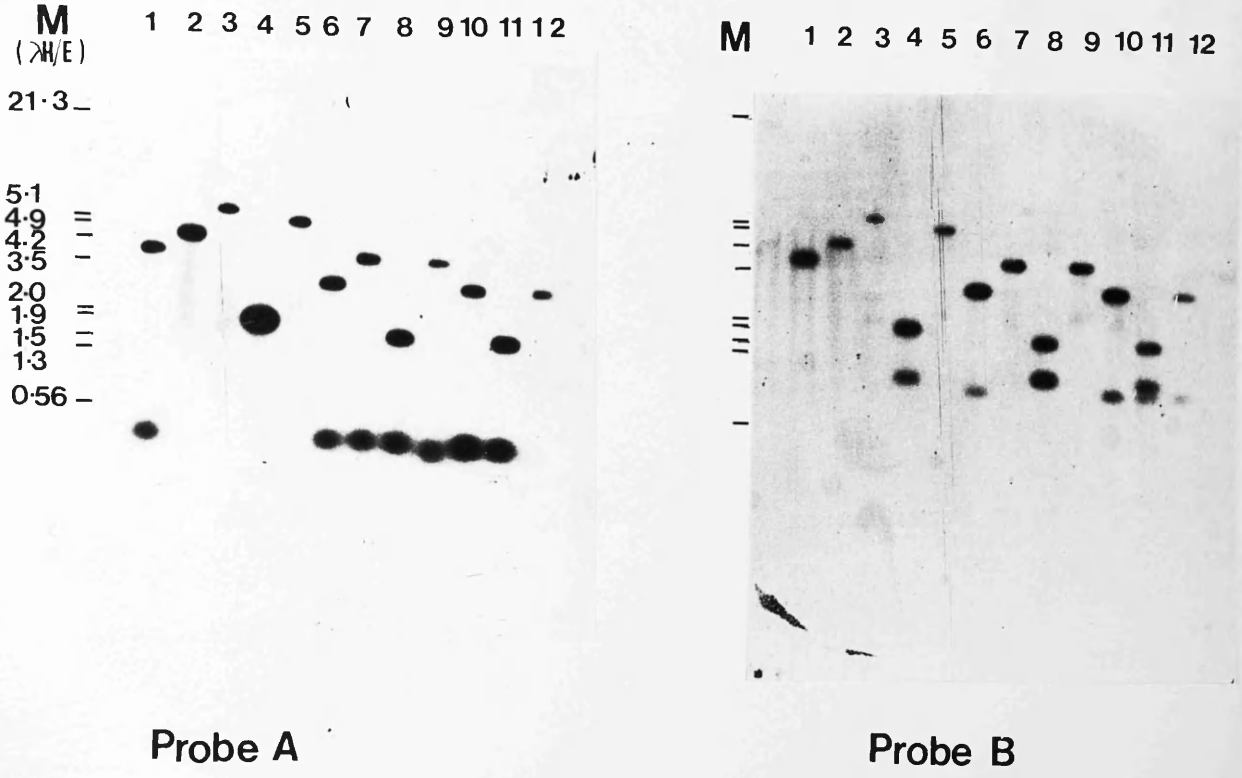
Panel A : Physical map of lambda MT 1.22 B.



Restriction map of the Lambda MT 1.22 B clone with probes used for construction of this physical map indicated below.

Figure 3.3.2.1

Panel B: Southern analysis of lambda MT 1.22 B.



The panels above show Southern analyses of the lambda MT 1.22 B clone used in the construction of the physical map of the clone shown in Panel A of the figure. Probes used in the Southern analyses are indicated beneath the aforementioned physical map. Posthybribization washing of the blots shown was to 0.1X SSC, 65 ^{0}C. The blot shown hybridized with probe B is a reprobing of that shown for the probe A analysis and the marker indications shown are identical. The respective digests used for the analyses are as indicated in order in Table 3.3, Lane 1 coresponding to a Bam HI digest.

Table 3.3 Restriction data from the mapping of lambda MT 1.22 B.

PROBE A	PROBE B	PROBE C	PROBE D
DIGEST		FRAGMENT SIZES (Kb)	
B	3.8, 0.2	3.8	6.05	7.6
Bg	4.5	4.5	>23	>23
C	6.2	6.2	6.2, >23	>23
E	2.1	2.1, 1.0 (8.0)	8.0	8.0,4.7,1.95
H	5.25,0.15,	5.25	5.25,4.7,0.6 0.4,0.4	4.7,2.2,0.5
Hc			3.8,2.7,2.2, 0.95	2.7,2.7,2.7, 1.0,0.5
P			9.8,1.8	9.8,1.8,O.4, 0.4
Pv			9.8	9.8,5.0,1.2
B/Bg	3.0, 0.2	3.0, 0.8	6.05	7.6
B/C	3.8, 0.2	3.8	4.5, 1.5	7.6
B/E	1.9, 0.2	1.9, 1.0 (8.0)	6.05	4.75,1.85, 1.05
B/H	3.8,0.15	3.8	3.3, 3.3, 0.6 0.4. 0.4	3.3, 2.2, 1.6, 0.5
B/Hc			3.8, 1.2, 0.95, 0.4	2.7, 1.7, 1.0, 0.5
B/P			4.7, 1.25	5.05, 1.8, 0.4, 0.4
B/Pv			6.05	3.8, 2.6,

			1.2
B/Bg/C 3.8, 0.2	3.0, 0.8	4.55, 1.5	7.6
B/Bg/E 1.9, 0.2	1.9, 1.0,0.8	6.05	4.75, 1.85, 1.05
B/Bg/H 3.8, 0.15	3.0, 0.8	3.3, 1.3, 0.6,0.4, 0.4	3.3, 2.2, 1.6, 0.5

The possibility of the lambda MT clones being scrambled was not to be unexpected, given that the 'A+T' richness of the 70bp repeat regions makes for highly recombinogenic secondary structures. Figure 3.4.1 shows a series of Southern blots of T. brucei EATRO 2340 genomic digests probed with specific fragments from the lambda MT 7.1-1 and lambda MT 7.1-2 clones. Hybridizations were all performed at 68°C for 16 hrs, post- hybridization washing was to 68°C, 0.1X SSC. Probe fragments are indicated on the maps shown in Figure 3.4.1 Panel A, while the compiled genomic maps resulting from this analysis are shown in Figure 3.4.2, with restriction data presented in Table 3.4. The genomic mapping from this analysis shows agreement with the lambda MT clone maps, indicating that, at the level of resolution employed, the two lambda clones represent faithful copies of their respective telomeres. One anomaly exists in the mapping using Probe 6 where a 5 Kb Hin dIII that overlaps the 3' end of the probe (see map Figure 3.4.2) is not detected. This may be a consequence of the region of overlap between the probe and the Hin dIII fragment being too small to give suitable hybridization. Relevant, in this respect, is the observation that the breakpoint between the two GUTat 7.1 BC telomeres appears to occur in this region. Two Hin dIII / Eco RI double digest fragments predicted from the map (Figure 3.4.2) are on faintly visible in the Southern analysis, these are arrowed in the figure.

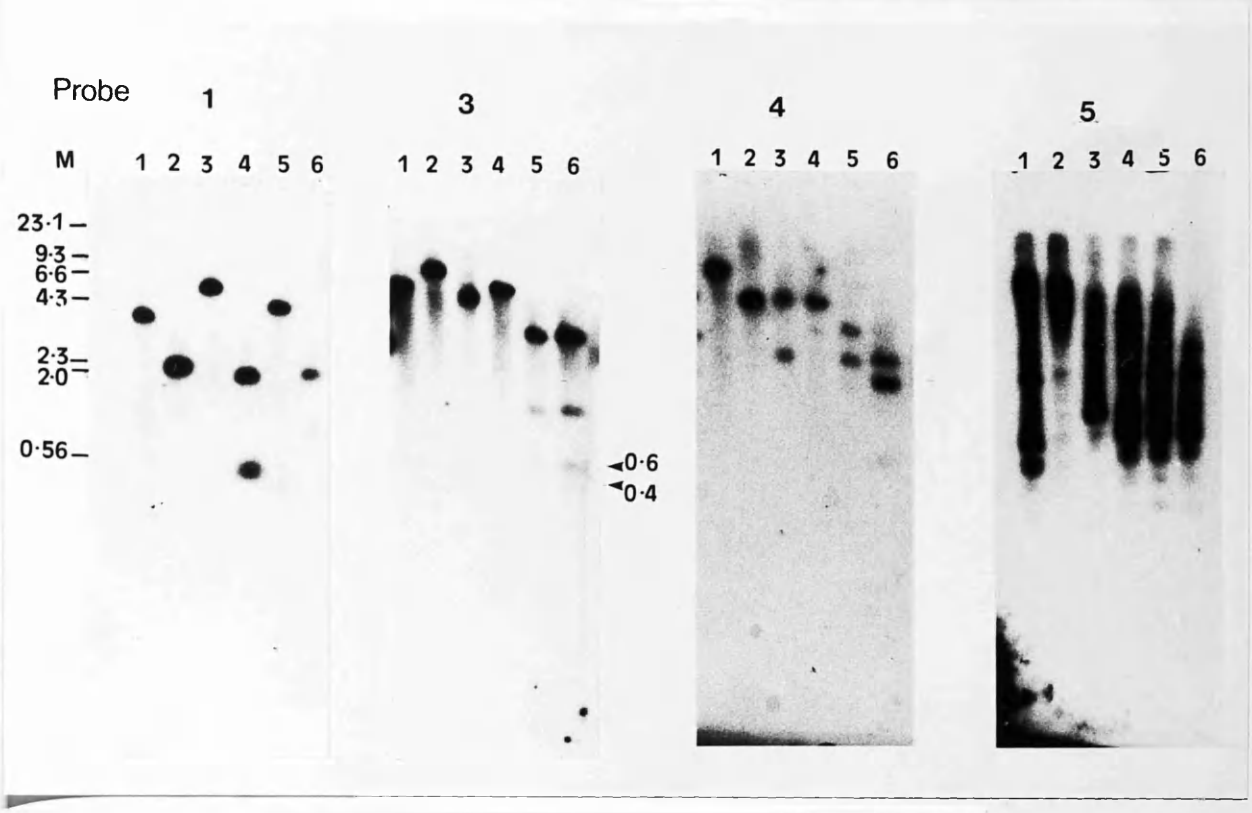
What is surprising from these Southern analyses is that nearly all the upstream areas of the GUTat 7.1 BC telomeres are not constituted by repetitive sequence elements. Most of the

Figure 3.4.1

Genomic Southern analysis of the GUTat 7.1 telomeres in EATRO 2340

Panel A : Maps for lambda MT 7.1-1 and 7.1-2 showing the probes used for Southern analyses.

B Panels B and C : Southern analyses of EATRO 2340 digests



C

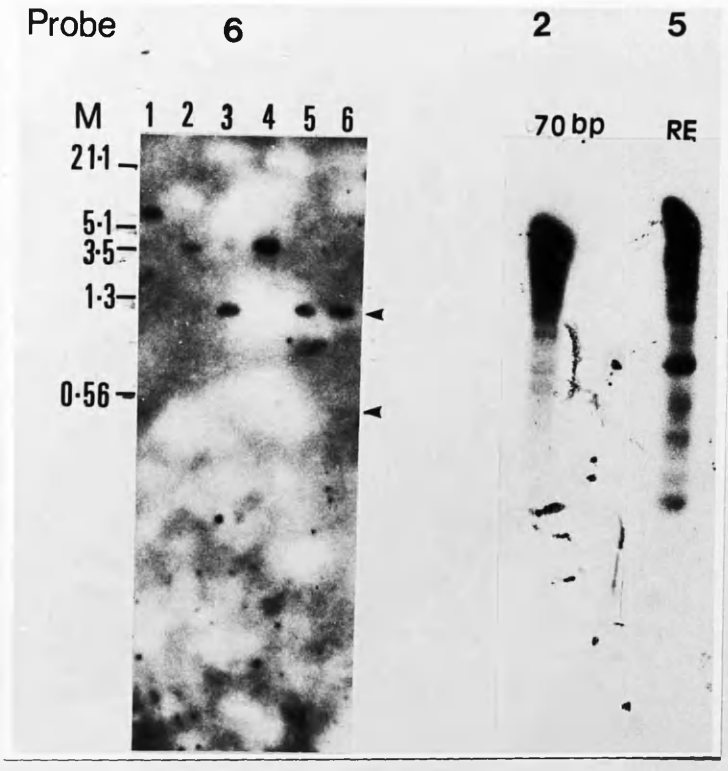
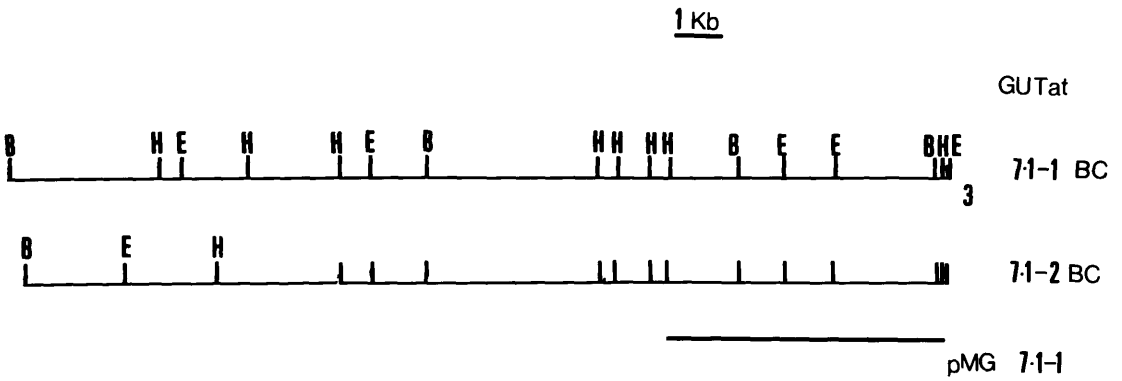


Figure 3.4.2 Genomic maps of the GUTat 7.1 BC telomeres.



The figure illustrates the genomic maps for the GUTat 7.1 BC telomeres in EATRO 2340. The extent of the plasmid genomic clone pMG 7.1-1, for the GUTat 7.1 gene is shown below the genomic maps. Restriction sites are only labelled on the 7.1-2 BC telomere where they difer from those on the 7.1-1 telomere.

Table 3.4 Restriction data from the mapping of the GUTat 7.1 BC telomere.

PROBE	1	3	4	6
DIGEST		FRAGMENT SIZE (Kb)		
B	3.8	6.05	7.6	8.0
E	2.2	8.0	4.75	3.5
H	5.25	5.25, 4.7 0.6,0.4,0.4	5.0, 2.2	1.7, 1.7
B/E	1.9, 0.3	6.05	4.75	3.5,3.5
B/H	3.8, 0.15	5.25,3.3, 0.6,0.4,0.4	5.0,2.2, 1.7	1.7,1.7, 1.7
B/E/H	1.9, 0.15	3.3,1.3,0.6, 0.4,0.4	1.7,1.7, 1.3, 0.6, 0.6	1.7,1.2 0.6

probes employed only detect single or a few bands on the genomic Southern blots. This is in stark contrast to what has been observed for other trypanosome telomeres, principally those employed as B-VSG gene expression sites. Two areas of the lambda MT 7.1-2 clone, one of which is common to lambda MT 7.1-1, detect genomic repetitive sequence. Probe 2 defines one such region and corresponds to a 1Kb Eco RI fragment, not 2.5Kb as predicted by Cornelissen et al. (1985a), which contains 70 bp repeat sequence (Panel C, 2) (see also section 3.3.1.3). The other corresponds to an unidentified genomic repeat sequence which lies on the lambda MT 7.1-2 clone within the region defined by Probe 5 (Panel C, 5).

3.5. Assessment of the Veracity of lambda MT 1.22 B.

Two immediate questions to be answered about the lambda MT 1.22 B clone were:

(i) whether or not it was a true representation of its genomic context?
(ii) were the same sequence elements present on the ILTat 1.22 BC telomere as those on the GUTat 7.1 BC telomere?

To gauge the status of the lambda MT 1.22B clone with respect to its genomic context, Bam HI digests of the clone were electrophoresed alongside similarly digested samples of EATRO 795 (ILTat 1.22) genomic DNA and hybridized with Probes 1, 3 and 5 respectively (see Figure 3.5.1) after Southern blotting. Hybridizations and subsequent posthybridization washings were performed at 68°C, and to a final posthybridization stringency of 0.1X SSC.

Figure 3.5.2 presents the results of this analysis.

The figure shows a composite of the results of the Southern analyses, with 12 hour autoradiographic exposures of the genomic tracks (lanes 2, 4 and 6 respectively), alongside 1 hour exposures of the positive control lambda MT 1.22 B lanes. The shorter exposure time employed in lanes 1, 3 and 5 was to allow for greater clarity of observation of the comparative signals from the genomic and clone digests. Longer exposures of the clone digest gave 'burn out' signals due to the quantity of homologous material present. For this reason also the lambda MT 1.22 B lanes were hybridized separately to prevent probe being 'mopped up', and thus reducing the signal from the genomic digest.

In each set of tracks a band of equivalent size is observed in the genomic and lambda clone digests. Probe 1 detects a 3.8 Kb Bam HI band in the genomic (lane 2) and lambda MT 1.22 B digest. The small 0.2 Kb Bam HI/Sau3A fragment also detectable by Probe 1 (see Figure 3.3.2.1 Panel B, section 1) is not present in this analysis due to the percentage of the agarose gel used in combination with the length of gel employed to give suitable resolution of the genomic digest. The band sizing at greater than 23 Kb in the genomic digest (lane 2) is the ILTat 1.22 ELC.

Probe 1 is also able to detect the 3' telomere fragments on each telomere, BC and ELC, but these are not visible in this analysis, possibly due to the combination of a short exposure in conjunction with a limitation in the quantity of probe employed, and the amount of homologous material present in the genomic digest.

Probe 3 detects a 6.05 Kb Bam HI fragment in both the lambda MT 1.22B digest (lane 3) and genomic digest (lane 4), while Probe 4 detects the expected 7.6 Kb Bam HI fragment in each (lanes 5 and 6 respectively). The results indicate that at the level of resolution employed for these analyses, the lambda MT 1.22B clone is a true representation of its genomic context.

As is apparent from Figure 3.5.2, each of the probes used only detected single bands in the genomic Southern blots, a feature discussed more fully in the discussion section of this chapter. Two regions did, however, detect genomic repeated sequences, as defined by Probes 2 and 5.

Probe 5 detected a highly reiterated genomic sequence, as did the equivalent region of lambda MT 7.1-2. This is examined further in the following chapter (see Chapter 4., section 4.8).

The second region on the clone detecting multicopy sequence corresponds to the 70 bp repeat sequence 5' of the VSG gene. This region is extremely brief and comprises a 410 bp Pst I/PvuII fragment. A detailed definition of this region is presented in Chapter 4, Figure 4.1.

Sequencing of this region reveals only a single complete 70 bp repeat unit present (K. Matthews and J.D. Barry, unpub. obs.).

Figure 3.5.1

Probes used to determine the status of lambda MT 1.22 B.

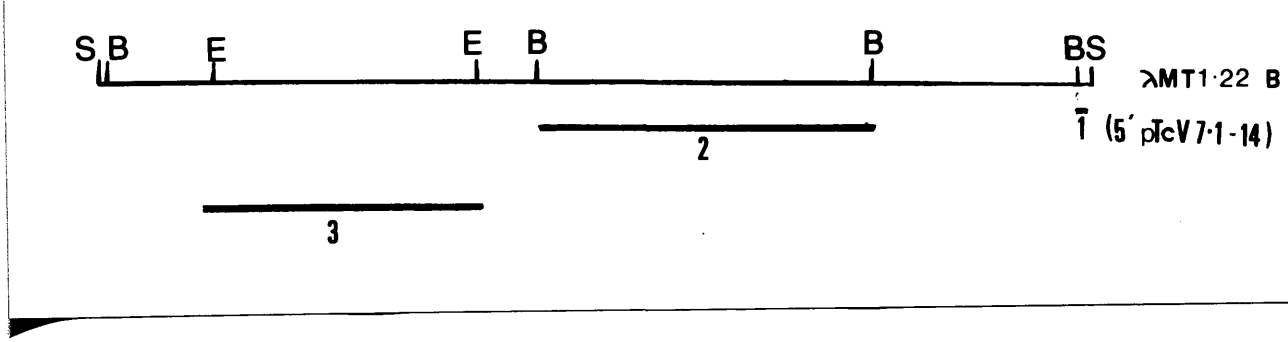
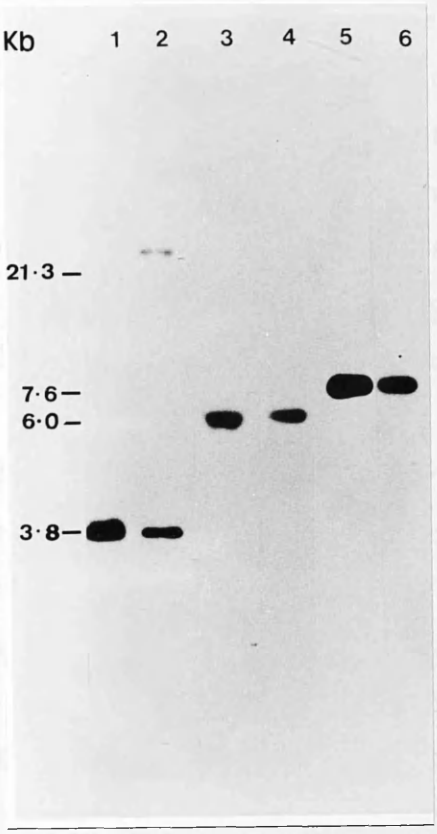


Figure 3.5.2

Assessment of the veracity of the lambda MT 1.22B clone.



Lane 1. Lambda MT 1.22 B x Bam HI -probe 1.

Lane 2. EATRO 795 x Bam HI - probe 1.

Lane 3. Lambda MT 1.22 B x Bam HI - probe 3.

Lane 4. EATRO 795 x Bam HI - probe 3.

Lane 5. Lambda MT 1.22 B x Bam HI -probe 4.

Lane 6. EATRO 795 x Bam HI - probe 4.

3.6. Genomic mapping of the ILTat 1.22 BC telomere.

The availability of the 'bona fide' lambda clone for the ILTat 1.22 BC telomere allowed for the genomic fine mapping of the area 5' of the VSG gene and beyond the region defined by the pMG 7.1-1 clone. This also presented the opportunity to cross compare the map with that for the GUTat 7.1 BC telomere, the equivalent telomere in the EATRO 2340 stock, and to assess if there were any differences in the EATRO 795 stock carrying the ILTat 1.22 BC telomere, outwith the region defined by lambda MT 1.22 B.

Mapping was once again performed by Southern analysis of genomic digests using probes derived from the lambda MT 1.22 B clone, plus the Pst I/Hin dIII fragment from pTcV 7.1-14. The probes are illustrated beneath a map of lambda MT 1.22 B in Figure 3.6.1. All hybridizations were performed at 68°C for 16 hours, followed by posthybridization washing to a stringency of 0.1X SSC, 68°C.

The compiled mapping data are presented in Figure 3.6. The probes used in the analysis are delineated in Figure 3.6.1 beneath the map of the BC telomere for ILTat 1.22. Individual fragments detected by each probe are tabulated in Table 3.6.

Probe 1, a 350 bp Pst I/Hin dIII fragment from pTcV 7.1-14 detects both the BC and ELC for ILTat 1.22, (Figure 3.6.2). In the Bam HI digest (lane 1) a 3.8 Kb (BC) and a > 23 Kb (ELC fragment are detected along with a 7.5 Kb (BC) and 8.5 Kb (ELC) telomeric fragments. The 3.8 Kb fragment provides a good internal

positive control for the VSG gene and is diagnostic of its presence in the EATRO 795 stock. The ELC being on a fragment of greater than 23 Kb is in keeping with it having an extensive barren region. This feature is further exemplified by the large ELC fragments in the BglII and ClaI lanes (lanes 2 and 3 respectively), both of which size at more than 23 Kb, while in the Hin dIII digest (lane 5) the ELC is observed on a 16 Kb fragment. The size of the latter is conducive to cloning in a lambda vector. A 2.2 Kb doublet is detected in the Eco RI digest (lane 4) indicative of the enzyme having a recognition site prior to the breakpoint for ELC formation on the BC telomere, hence the probe cannot distinguish between the BC and ELC telomeres. The faint bands at 7.5 and 8.5 Kb observable in lanes 1, 6, 7, 10 and 14 correspond to fragments extending 3' from sites 5' to the Hin dIII site in Probe 1, to the telomere terminus (see Figure 3.6.3).

One anomaly exists in lane 14, where these fragments should not be detected as the digest includes Hin dIII, and thus placing them outside the region defined by the probe. In this digest, however, Hin dIII appears not to have cut, hence the detection of an ELC fragment of 20 Kb defined 3' by a Bam HI site and 5' by a Cla I site. The percentage of the gel used for the mapping, 0.6%, proved too low, with the extended run to detect small fragments, of < 500 bp. The sizes and presence of any such fragments are inferred in the compilation of the map shown in Figure 3.6.3, and in Table 3.6. One Pst I site (marked * on the genomic map, Figure 3.6.3) consistently showed partial digestion during the mapping analysis. This is not an unusual feature of trypanosome telomeres and has been amply documented (Bernards et al., 1984; Lenardo et al., 1984). This

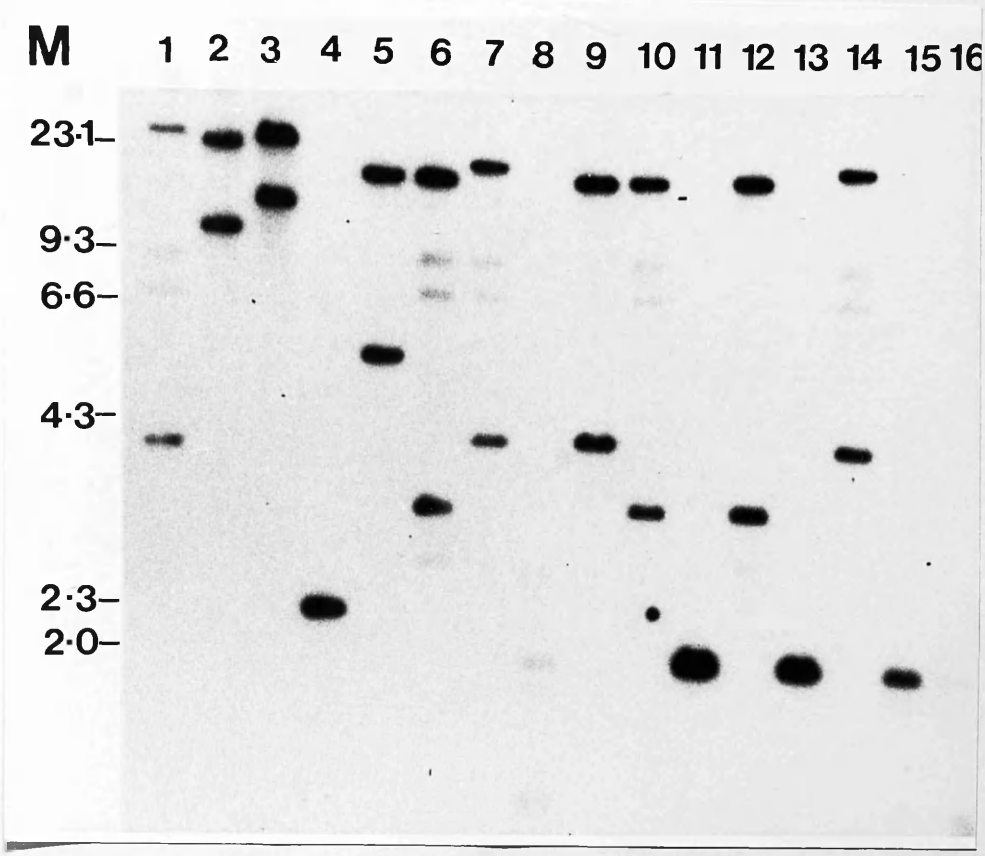
site, however, is not detected in the lambda clone or in pTcV 7.1-14, though sequence analysis has revealed its presence in pMG 7.1-1 (K. Matthews and J.D. Barry, unpub. obs.). The absence of this site in pTcV 7.1-14 and lambda MT 1.22 B may reflect a polymorphism at this site in these recombinant molecules.

3.7 Attempts to clone the ILTat 1.22 ELC.

Previous efforts to clone the ILTat 1.22 ELC in an EMBL 4 library plated on NM 539 E.coli host cells all proved fruitless. The reasons for this may be the presence of both direct and indirect repeat regions typical of expression site telomeres in conjunction with barren regions and unclonable regions of indeterminate nature (see discussion to this chapter).

Mapping of the ILTat 1.22 VSG gene (see section 3.6.) indicated that its ELC resided on a 16 Kb Hin dIII fragment in the particular stock examined (Figure 3.6.2), and thus may be suitable for cloning in a lambda vector. To overcome the problems associated with insert instability host strains, NM 621 was employed on which to screen the library, as it contained a suitable genotype for the propagation of repetitive sequence in the insert fragment and for coping with some sequence modifications. The potential problem of an 'unclonable' stretch on the ELC telomere was considered an indeterminate factor, but as this host strain had not previously been tested for ELC cloning from a size selected total digest library, it was considered worth attempting . To construct the library, a fragment range of 15-17 Kb was excised in agarose from a total Hin

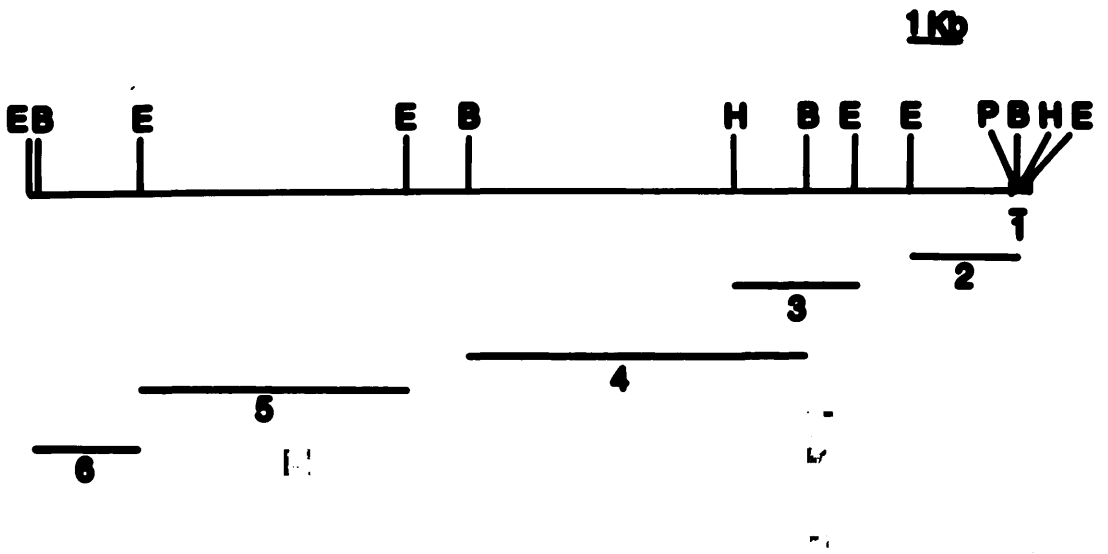
Figure 3.6.2 Probe 1 genomic mapping of IlTat 1.22 BC telomere.



Lane 1.B, 2.Bg, 3.C, 4.E, 5.H, 6.B/Bg, 7.B/C, 8.B/E, 9.B/H, 10. B/Bg/C, 11.B/Bg/E, 12.B/Bg/H, 13.B/C/E/, 14.B/C/H, 15.B/E/H, 16.B/Bg/C/E/H.

Markers (M) are all shown in Kb.

Figure 3.6.3 Map of the ILTat 1.22 telomeres.
Probes.



Map of the lambda MT 1.22 B clone with probes used for genomic mapping delineated below.

Table 3.5 Restriction fragment sizes from the mapping of the ILTat 1.22BC telomere.

PROBE	1		2
DIGEST	FRAGMENT SIZE (Kb)		
B	>23, 8.5, 7.5, 3.8	B	>23,8.5, 7.5, 3.8
Bg	>23, 10.5	Hc	2.2
C	>23, 13	P	1.4, 1.0
E	2.2	Pv	2.1, 0.55
H	16, 5.25	B/Hc	2.0, 0.2
B/Bg	16, 8.5, 7.5, 3.0	B/P	1.4, 0.65, 0.35
B/C	20, 8.5, 7.5, 3.8	B/Pv	2.1, 0.35, 0.2
B/E	1.9, 0.3	B/Hc/P	1.35, 0.65, 0.2
B/H	16, 3.8, 0.5	B/Hc/Pv	1.65, 0.35, 0.2
B/Bg/C	>23, 8.5 7.5, 3.0	B/P/Pv	1.4, 0.35, 0.3, 0.2
B/Bg/E	1.9, 0.3	Hc/P	1.35, 0.85
B/Bg/H	16, 3.0, 0.15	Hc/Pv	1.65, 0.5
B/C/E	1.9, 0.3	P/Pv	1.4, 0.55, 0.3
B/C/H	16, 8.5, 7.5, 3.8	Hc/P/Pv	1.35., 0.5, 0.3
B/E/H	1.9, 0.15	B/Hc/P/Pv	1.35,0.35,0.3,0.2

PROBE 3		3	
DIGEST		FRAGMENT SIZE (Kb)	
B	6.05,3.8	B	6.05,3.8
Bg	>23,10.5	Hc	3.8,2.2,2.2
C	>23,13	P	9.8,1.8,1.35
E	8.0	Pv	9.8
H	5.25	B/Hc	3.8,1.8,0.4
B/Bg	6.05,0.8	B/P	1.35,1.3,0.5
B/C	3.8,1.5	B/Pv	6.05,1.35
B/E	6.05,0.9	B/Hc/P	1.35,0.8,0.4
B/H	3.8,1.35	B/Hc/Pv	3.8,1.3.0.45,0.4

B/Bg/C	1.5,0.8,(3.0)
B/Bg/E	6.05,0.8,(3.0)
B/Bg/H	1.35,0.8
B/C/E	1.5,0.9
B/C/H	3.8,1.35
B/E/H	1.35,0.9
B/Bg/C/E/H	1.35,0.8,0.1.

PROBE 4 4

DIGEST	FRAGMENT SIZE (Kb)		
B	6.05	B	6.05
Bg	>23	Hc	3.8,2.7,2.2,0.95
C	13,13	P	9.8,1.8
E	8.0	Pv	9.8
H	5.25,4.7,0.6,0.4,0.4	B/Hc	3.8,0.95,0.9,,0.4
B/Bg	6.05	B/P	4.75,1.3
B/C	4.35,1.7	B/Pv	6.05
B/E	6.05	B/Hc/P	2.9,0.95,0.9,0.9,0.4
B/H	3.3,1.35,0.6,0.4,0.4	B/Hc/Pv	3.8,0.95,0.9,0.4
B/Bg/C	4.35,1.5	B/Hc/P/Pv	2.9,0.95,0.9,0.9,0.4
B/Bg/E	6.05		
B/Bg/H	3.3,1.35,0.6,0.4,0.4		
B/C/E	4.35,1.7		
B/C/H	3.3,1.35,0.6,0.4,0.3,0.1		
B/E/H	3.3,1.35,0.6,0.4,0.4		
B/Bg/C/E/H	3.3,1.35,0.6,0.4,0.4		

PROBE 5

DIGEST	FRAGMENT SIZE (Kb)
B	7.6
E	4.75
H	4.7,4.7,2.2,0.5
B/E	4.75
B/H	3.3,2.2,1.6,0.5
B/E/H	2.2,1.6,1.6,0.5
Hc	2.7,2.7,1.0,0.5
P	9.8, (5.0,0.4,0.4)
B/Hc	2.7,1.7,1.0,0.5
B/P	5.0,1.8,0.4
B/Hc/P	2.7,1.7,0.5,0.5,0.3.,0.2

dIII digest of genomic DNA from ILTat 1.22 expressing trypanosomes and ligated to lambda DASH vector arms.

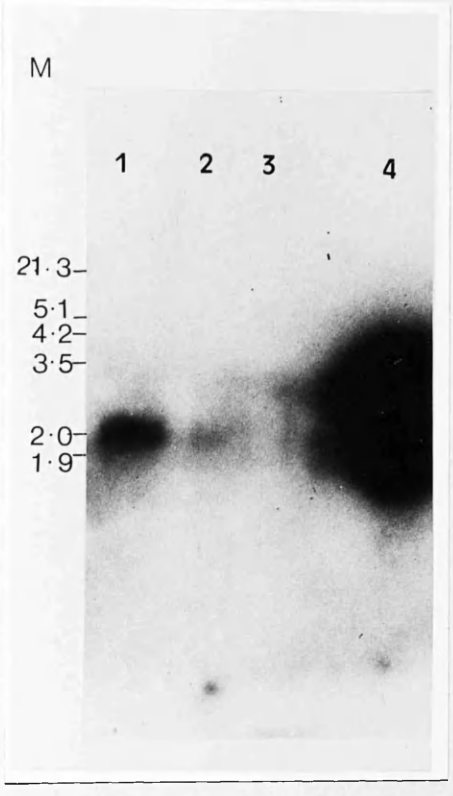
Screening of a total of 6 x 10^3 p.f.u. with the 350 bp Pst I /Hin dIII fragment of pTcV 7.1-14 at 68°C for 16 hours, followed by posthybridization washing to 0.1X SSC, 68°C, detected three positive recombinants, which were picked and plaque purified.

All these recombinants gave very small plaques and consequently low titres (10^4 p.f.u. per plaque - data not shown). To compensate for this, DNA preparation from each recombinant necessitated use of a plate lysate method to enhance DNA recovery.

Hin dIII digests of each recombinant clone are shown hybridized with the Pst I /Hin dIII fragment of pTcV 7.1-14 in Figure 3.7.1. Hybridizations were for 16 hours at 68°C, followed by posthybridization washing to a stringency of 0.1X SSC, 68°C. None of the clones appeared to have an insert corresponding to a single Hin dIII fragment of 16 Kb, and thus were scrambled. The VSG gene is readily detectable on 2 of the 3 clones (lanes 1 and 2). The third clone (lane 3) had too little DNA present to give a suitably strong signal. In the former two clones a 2.1 Kb band was detected, indicating that a rearrangement may have occurred in the vicinity of the 70 bp region.

The pMG 7.1-1 control (lane 4) yields two signals on Hin dIII /Eco RI digestion. The fragment at 2 Kb carrying the VSG gene is readily observed, as is a plasmid band (upper band, lane 4) which is detected due to the presence of a small vector

Figure 3.7.2 Southern analysis of ILTat 1.22 ELC clones.



Lane 1. Lambda ELC 1.22 A x H

2. Lambda ELC 1.22 B x H

3. Lambda ELC 1.22 C x H

4. pMG 7.1-1 x H/E

Marker (M) sizes are shown in Kb.

section on the probe.

No further work on these clones was pursued as a consequence of their being scrambled.

Discussion.

The results in this chapter have documented attempts to clone M-VSG genes from T.b.rhodesiense genomic libraries. Cloning for one of four M-VSG genes sought proved successful, with both GUTat 7.1 BC genes present in the EATRO 2340 stock, and the equivalent BC VSG gene, ILTat 1.22, present as a single copy in the EATRO 795 stock, being isolated.

Attempts to isolate GUTats 7.2, 7.13 and 7.15 BC genes in EATRO 2340 and ILTats 1.62,1.63 and 1.64 BC genes in EATRO 795 proved unsuccessful. Cloning of the GUTat 7.1 BC genes did, however, require exhaustive screening of what appeared to be a representative genomic library in both its amplified and unamplified forms. The apparent under-representation of the M- VSG genes in the library may be attributed to a number of factors, detailed below, which were taken into consideration when attempting to clone the ILTat 1.22 BC from the EATRO 795 stock and are discussed relative to this.

(i) Lack of Sau 3A sites 3' or 5' to the VSG gene, so that suitable fragments of the correct size for cloning are not produced.

(ii) Instability of the recombinant phage carrying M-VSG gene in the E. coli host used for plating out the library due to :

- Inverted repeats in the cloned sequence.
- The presence of direct repeats in the clone sequence.
- Modified bases acting as targets for host restriction systems.

(iii) Unclonable trypanosome telomeric regions of unknown nature.

The availability of Sau 3A sites either 3' or 5' to the M-VSG gene, while formally a consideration in the screening process, is not a major problem with the GUTat 7.1 or ILTat 1.22 genes, which already have a Bam HI site in the VSG gene coding region. Furthermore, the ability to isolate both BC copies of GUTat 7.1 from the EATRO 2340 stock, and the ILTat 1.22 BC from the EATRO 795 stock, on approximately 18 Kb fragments suggests that its under-representation in both libraries was not significantly affected by cloning site availability. It is unknown as to what extent this consideration affects the other M-VSG genes sought. It is known, however, that ILTat 1.61, a double copy gene in the EATRO 795 stock, and GUTat 7.15, its equivalent in EATRO 2340, possesses an internal Sau 3A site (Keith Matthews and J.D. Barry, unpub. obs.).

In retrospect one might allow for the consideration of cloning site availability on 18-22 Kb fragments by performing a series of Sau 3A partial digestions on genomic DNA, then cutting to completion with a restriction enzyme with a site internal in the VSG gene. Such digests could then be probed with cDNA fragments from regions of the VSG gene 5' to the site for the enzyme used for complete digestion. This would eliminate detection of large fragments running from the VSG gene to the telomere terminus, and allow observation of fragments running 5' from the VSG gene. Sau 3A sites internal in the VSG coding region would also be readily observed.

Two major sequence organizations have been shown to affect phage stability and hence representation within a genomic library;

these are direct and indirect (palindromic) repeats.

Sequences present as direct repeats, especially if the repeat unit is multicopy within a given lambda clone, can readily be lost or reduced in number by unequal crossing over between individual repeat units. This can result in the insert fragment being reduced to a size at which it is no longer able to participate in packaging and therefore support the formation of viable phage. The 70 bp repeats 5' to VSG genes provide such a substrate. This problem, however, can be overcome by use of a host strain carrying mutations in rec A, B and C.

The rec B^-C^- genotype is required to allow propagation of lambda EMBL 4 recombinant phage. Such phage are gam$^-$, by virtue of the gam gene being on the phage stuffer fragment (see Kaiser and Murray, 1985) and as a consequence, are susceptible to Exonuclease V, the Rec BC enzyme. This would otherwise prevent formation of phage concatamers suitable for packaging.

The use of a rec A^- mutant prevents homologous recombination by the main pathway in E. coli K12 derived hosts, thus it facilitates propagation of recombinant phage carrying directly repeated sequences.

The presence of inverted repetitions with a lambda clone may make it unstable or inviable in many wild type E. coli hosts (Leach and Stahl, 1983), the reasons for this are unclear, but they may provide a strong selection for DNA loss via 'looping out'.

Once again this problem may be circumvented using a rec BC^- sbc B^- host, as has been successfully demonstrated for a number of human genes containing long palindromes (reviewed, Wyman and Wertman, 1987). Such sbc B^- mutants were initially identified as rec B^- or rec C^- suppressors, and their presence as part of the host genotype improves rec BC^- cell growth properties (Barbour et al., 1970). As to how and why sbc B^- mutants stabilize palindromes is unknown.

Considerations on the nature of the host strain used to propagate the libraries may thus be critical in respect of the failure to isolate M-VSG genes other than GUTat 7.1 and ILTat 1.22.
NM 539, the host strain used for propagation of the EATRO 2340 library is rec A^- rec BC^- and thus should cope with direct repetitions, but lacks the sbc B^- mutation to alleviate any problems caused by palindromes.

NM 621, used for the EATRO 795 library, is rec A^- rec B^- rec C^- sbc B^- and thus should eliminate problems associated with sequence repetition. Isolation of the ILTat 1.22 BC clones, while requiring less exhaustive screening than for GUTat 7.1 clones, still indicated under-representation of this telomere in the EATRO 795 library. This fact, coupled with the inability to clone the other 3 ILTat M-VSG genes sought, strongly suggests that other factors contribute to the general cloning problem.

A possible contributary factor in this respect, influencing sequence representation within a genomic library, is the restriction of foreign DNA by E.coli endonucleases. Both NM 539 and NM 621

are Mcr$^-$ which allows for cloning of sequences modified by the presence of methyl cytosine (Raleigh et al., 1988), similarly each strain is Rgl$^-$, which confers insensitivity to restriction in sequences carrying 5-hydroxymethylcytosine (Raleigh and Wilson, 1986), and hsdR$^-$ which prevents restriction by the EcoK endonuclease. Neither strain, however, accounts for the possible action of a third methylation dependent site specific restriction locus Mrr, which targets N^6-methyladenine residues (Heitman and Model, 1987). No suitable Mrr$^-$ host strains are currently available for general cloning purposes. The presence of N^6-methyladenine residues in trypanosome expression sites has not been demonstrated, though a graduation of non-methylcytosine modification, running telomere to centromere, has been observed on inactive sites (Bernards et al., 1984b; Pays et al., 1984). It may thus be that the Mrr system is causing the cloning difficulties associated with these telomeric sequences. It is, however, difficult to reconcile this hypothesis with the ILTat 1.22 ELC clones, all of which were scrambled. These clones were derived from an active expression site where the telomere should not be modified. Furthermore the genotype of NM 621 should cope with the long stretches of repetitive sequences typical of the B-VSG expression site. Thus it appears that a further factor is obstructing successful cloning; at least of the ELC.

Such difficulties are not unique to M-VSG gene cloning. Unclonable regions of unknown nature have been reported for a number of B-VSG gene expression sites (Alexandre et al., 1988) and have required 'jump' cloning procedures.

Relevant in this respect is the presence of a stretch of

sequence on the ILTat 1.61 telomere unclonable in plasmid vectors (K. Matthews and J.D. Barry, unpub. obs.). This sequence can be cloned in lambda EMBL 4, albeit requiring extensive screening and sequence loss from the lambda clone, and indicates that its inability to be cloned in plasmid vectors could be due to inadvertant expression and subsequent lethality to the host cell, or possibly be due to it causing replicon instability. Analysis of this sequence from the lambda clone may answer this problem and provide insight into the presence of unclonable stretches on B-VSG expression sites, all of which have been reported for sites cloned in a series of plasmid vectors (Kooter et al., 1987; Alexandre et al., 1988; Gibbs et al., 1988).

The possibility remains of unknown modification on trypanosome telomeres and novel unclonable regions being responsible for the cloning difficulties observed. Use of suitable Mrr$^-$ hosts may in the future circumvent these difficulties.

Physical mapping of the GUTat 7.1 and ILTat 1.22 BC telomeres reveals a striking disparity with characterized B-VSG gene expression sites. Unlike most telomeric B-VSG genes which are preceded by over 40 Kb of 70 bp (Michels et al., 1983), the GUTat 7.1 and ILTat 1.22 BC genes are only preceded by a very small 70 bp repeat region which can be defined within a 1Kb EcoRI fragment on each lambda clone. This region is 'bona fide' in the lambda clones in that they appear to be true representatives of their genomic context, as judged by the agreement of restriction maps for the lambda clones and genomic maps for each telomeric BC gene. The genomic maps are updated, and in the instance of that for ILTat

1.22, more comprehensive than the original (Cornelissen et al., 1985a). The 70 bp region of the pMG 7.1-1 plasmid clone for GUTat 7.1 has been subsequently sequenced and shown to comprise no more than one complete 70 bp repeat unit within a 410 bp Pst/PvuII fragment (K. Matthews and J.D. Barry, unpub. obs.).

The brevity of the 70 bp repeat region is quite significant, as it has been proposed that activation of a given VSG gene by duplicative transposition may be correlated directly with the number of 70 bp repeat units it possesses on its 5' flank (Liu et al., 1983; Aline et al., 1985a; Timmers et al., 1987). By such a hypothesis one would expect infrequent activation of GUTat 7.1 or ILTat 1.22, yet it is known that these genes are activated at high frequency in bloodstream infections in a number of trypanosome stocks (the ETat 1.2 VSG gene - Mc Neilage et al., 1969; ILTat 1.22 - Miller and Turner, 1981; GUTat 7.1 - Barry et al., 1985). This argues strongly against any direct correlation between frequency of VSG gene activation and the number of 70 bp repeat units, at least for telomeric VSG genes.

The ILTat 1.22 ELC, as judged from the genomic map of this telomere (Figure 3.6.1), appears to have a typically large barren region of approximately 14 Kb.

The upstream maps of each BC telomere for GUTat 7.1 appear identical at the resolution used in mapping, for approximately 10 Kb upstream of the VSG gene. Beyond this the maps diverge and show no apparent cross homology, as judged by genomic Southern analysis (Figure 3.4). This implies that the duplication of this telomere is

from a breakpoint very close to the Bam HI site, 10 Kb 5' to the Bam HI site internal within the VSG gene coding region (Figure 3.4.3), and suggests that the duplication was most likely the result of a telomere conversion event. It is the GUTat 7.1-2 telomere that is equivalent to the ILTat 1.22 BC telomere as judged by their genomic maps.

Significantly, the majority of the upstream sequences on all three telomeres appear as single bands on genomic Southern blots, with only two regions showing as multicopy, and one of these comprises 70 bp repeat sequence. Both these regions are found on the GUTat 7.1-2 and its equivalent, ILTat 1.22 BC telomeres.

This then allows a number of questions to be posed concerning the anatomy of these M-VSG gene telomeres.

(i) Is the genomic repeat sequence present on the GUTat 7.1-2 and ILTat 1.22 BCs a novel or related ESAG to those on B-VSG expression sites?

(ii) Outwith the 70 bp repeat sequence do the M-VSG telomeres share any downstream homology with their B-VSG counterparts?

(iii) Given the lack of extensive multicopy sequence on the GUTat 7.1 and ILTat 1.22 BC telomeres, does this structural individuality reflect the disparity in M- and B-VSG gene expression systems?

These questions are pursued in the following chapter as the anatomy of each M-VSG telomere is dissected.

An unusual feature of this area of the telomere was the presence of a PstI site within the VSG coding region (see Figure

3.6.1) not detectable in the lambda or a cDNA clone containing the VSG gene. This site appears modified as it is only detected in genomic digests through the presence of partial digest products. Sequence analysis of the 5' end of the VSG gene from plasmid clone pMG 7.1-1 confirm its presence.

The detection of modified Pst I and Puv II sites on M-VSG gene telomeres is not unusual and appears coincident with telomere inactivity (Lenardo et al., 1984, 1986). The fact that genomic mapping in the instance of the ILTat 1.22 BC was performed using DNA from a 1.22 expressor clone, may explain this observation, though this was the only observable instance on the telomere and was in no way as extensive as reported for the other telomeres (Lenardo et al., 1984, 1986). It is difficult to envisage how this modification could be perpetuated in the lambda clone. It is possible the lambda MT 1.22 B clone represents a copy of a telomere with a small polymorphism at this site, as appears the situation with the cDNA clone pTcV 7.1-14.

Chapter 4.

Anatomy of the BC telomeres for GUTat 7.1 and ILTat 1.22.

Introduction

As detailed in the previous Chapter, indirect analysis of metacyclic forms is facilitated by the observation that early bloodstream forms appear to express M-VSG genes for the first few days within the host bloodstream, and that this expression is distinct from that utilized in mature bloodstream infections, as judged by phenotypic analysis (Turner et al., 1988).

Furthermore, this expression is concomitant with high rates of interswitching between M-VATs (Barry et al., 1985). Indirect methods of analyses, however, still suffer from a series of drawbacks. Firstly, metacyclic populations are polyclonal for the expression of individual VSG genes (Tetley et al., 1987). Secondly, fly transmitted trypanosome lines undergo antigenic variation at rates which are several orders of magnitude greater than those displayed by rodent adapted lines (Turner and Barry, 1989). Thus it is not normally possible to obtain trypanosomes which are both pleomorphic enough for transmission via the tsetse fly yet stable enough to be clonally propagated to levels sufficient for obtaining materials for molecular analysis. These problems have been circumvented, however, by the use of a novel line of trypanosomes with the necessary combination of fly transmissability and rodent virulence (Turner and Barry, 1989), which subsequently allows for the isolation of trypanosomes from salivary exudate or early bloodstream populations to be clonally

propagated to levels sufficient for DNA and RNA isolation and analysis.

The qualities of this line of trypanosome, belonging to the ILTat serodeme of the EATRO 795 stock, have been used to develop a model system to study M-VSG gene expression against a more representative background of antigen switching events. The pattern of antigenic variation exhibited by the model line is similar to that exhibited by conventional fly transmitted lines. The former, however, can be propagated to levels sufficient for molecular analysis without extensive switching between VATs, whereas with the latter by the time such levels are reached the trypanosomes have lost expression of the initially expressed VAT.

A further advantage of this model line is that the trypanosomes are not immunogenic during short term infections in mice being cloned prior to levels sufficient for immunogenicity being attained. Intervention by antibodies thus does not arise, hence expression of the initially expressed VAT is maintained. Furthermore, no differences in the pattern of antigenic variation are observed between infections in immunosuppressed and non-immunosuppressed mice (Turner and Barry, 1989). The features inherent in this model line thus present the necessary credentials for a detailed molecular analysis of a M-VSG gene expression.

Utilization of this model line indicates that antigenic switching between M-VSG genes occurs at a rate of 10^{-2} to 10^{-3} switches per cell generation (Turner and Barry, 1989) and implies that the switches

are biochemically 'driven' and not the result of background recombination events as suggested previously (Borst, 1986; Greaves and Borst, 1987; Pays and Steinert, 1988). This observed rate of switching is comparable to rates observed for antigenic variation in other systems, such as Borrelia hermsii and Neisseria spp. (Seifert and So, 1988), and is indicative of an optimum rate being crucial in maintaining a balance between the parasite growth rate and host detection and suppression of its numbers.

A second crucial observation from use of this system is that M-VSG gene expression occurs in situ, with ELC formation occurring rapidly, concomitant with the switch to B-VSG expression, (S. Graham and J.D. Barry, unpub. obs.). The observations of M-VSG gene activation by ELC formation as reported previously (Cornelissen et al., 1985a; Delauw et al., 1987) may have been a consequence of the indirect systems of analysis using rodent adapted lines, and that the mode of activation observed reflected a switch to B-VSG expression having already occurred.

Possession of a lambda clone for the ILTat 1.22 BC telomere from the EATRO 795 stock used for the model system allows for a more direct approach in the analysis of the M-VSG expression system.

Dissection of the expression site for the ILTat 1.22 M-VSG gene would allow for a number of pertinent questions to be answered concerning the anatomy of an M-VSG gene expression site telomere, some of which are listed below.

(i) Are there any unique features possessed by M-VSG telomeres?
(ii) Is there any crosshomology with characterized B-VSG gene expression sites?
(iii) What size is the M-VSG gene transcription unit?
(iv) Is transcription of the M-VSG gene by a similar type of polymerase as for B-VSG genes?
(v) What is the nature of the M-VSG gene promoter, if present?
(vi) Are any upstream sequence elements on the M-VSG gene telomere transcribed during bloodstream infection?
(vii) What is the nature of the genomic repeat sequence present on lambda MT 7.1-2 and lambda MT 1.22 B?

These questions are relevant to trying to assess the individual characteristics of M-VAT telomeres and in gaining insight into the nature of the distinctive M-VAT expression system.

RESULTS :

4.1. Southern analysis of lambda clones representative of the M-VSG gene BC telomeres.

The physical maps of lambda MT 7.1-1, lambda MT 7.1-2 and lambda MT 1.22 B, as detailed in the previous chapter, appeared to be accurate with respect to the equivalent genomic map for each respective telomere. These maps did not, however, completely agree with a plasmid clone of the BC telomere, pMG 7.1-1 (Cornelissen et al., 1985a), and thus it was decided to delineate the location of the VSG and 70 bp repeat sequence accurately using the lambda clones corresponding to each telomere.

To this end lambda MT 7.1-1, lambda MT 1.22 B were restriction enzyme digested and hybridized with the 5' Pst I-Hin dIII fragment of pTcV 7.1-14. These two were chosen as representative of the two individual BC telomeres carrying the VSG gene; GUTat 7.1-2 being equivalent to ILTat 1.22 in the EATRO 2340 genome, as is apparent from a comparison of the restriction maps (see Chapter 3, section 3.4 and section 3.6). Hybridizations were performed for 16 hours at 68°C, and posthybridization washing to a stringency of 0.1X SSC, 68°C. The result for lambda MT 7.1-1 is presented in figure 4.1, that for lambda MT 1.22 B has been presented previously in Chapter 3, figure 3.3.2.1., panel B. In each instance a 3.8 Kb Bam HI fragment belonging to the recombinant phage is detected in agreement with the equivalent genomic map (Cornelissen et al., 1985a) (see map, Chapter 3, Figures 3.3.1.3 and 3.3.2.1). For each clone, the hybridizing Eco RI fragments are

smaller than predicted from the genomic map, but this is explicable if the clones' 3' cloning site lies 5' of the genomic Eco RI site, as is the case for both the lambda MT clones. Thus, the 3' cloning site in lambda MT 7.1-1 appears to be the Bam HI site within the region covered by the cDNA, so that a Bam HI/Eco RI double digest yields the expected 1.9 Kb fragment detectable using the 5' fragment of the cDNA; which is in agreement with the genomic map. In lambda MT 1.22 B the 3' cloning site appears to be between the Bam HI and Eco RI genomic sites, and thus the detected Eco RI fragment is smaller than the predicted 2.2 Kb, as opposed to the 2.1 Kb, but again the double digest band sizes correctly at 1.9 Kb . The position of the 3' cloning site relative to the genomic map also explains the unpredicted size of the Pst I fragment in lambda MT 7.1-1 with respect to its genomic equivalent. The Eco RI maps of the lambda MT 7.1-1 and lambda MT 7.1-2 differ in only one aspect, the presence of a 3.5 Kb band in the former and a 4.7 Kb band in the latter (see maps Chapter 3, Figure 3.3.1.3 and 3.3.2.1). As detailed earlier in Chapter 3, these clones appear to represent both copies of the GUTat 7.1 telomere in the EATRO 2340 stock of T.b. rhodesiense. Mapping indicates that lambda MT 7.1-2 and lambda MT 1.22 B are copies of the same telomere from each respective EATRO stock.

4.2. Definition of the 70 bp repeat region 5' to GUTat 7.1 and ILTat 1.22 BC genes.

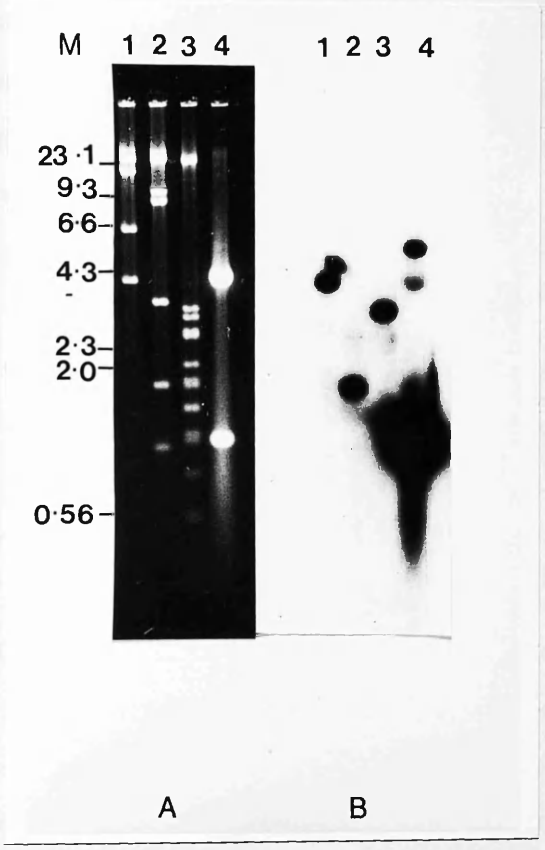
Figure 4.2 shows the result of probing lambda clones representative of GUTat 7.1 and ILTat 1.22 BC telomeres with a Pst I/TaqI fragment from pTgB 221.1 (Bernards et al., 1984), a

genomic clone from the 221 expression site consisting solely of 40-45 repeat units of 70 bp repeat sequence. Hybridizations were performed at 37°C, 3X SSC. The relaxed stringency was employed to detect any related sequence possibly on either of the clones used in the analysis.

Even at the relaxed stringency employed 70 bp repeat sequence is only detected on a single band in the individual digests of each respective clone. In the lambda MT 7.1-1 digests, representative of the M-VSG BC telomere in EATRO 2340, the 70 bp repeat sequence is defined within a 1 Kb Eco RI fragment (Panel A, lane 2). In the lambda MT 1.22 B digests, the 70 bp repeat sequence can be delineated within a smallest common fragment of 410 bp which corresponds to the overlap between a 1.35 Kb Pst I fragment and a 2.1 Kb PvuII fragment (see Chapter 3, Figure 3.3.2.1), both of which hybridize with the probe (lanes 3 and 4, Panel B). The latter observation confirms the sequence analysis of the same region from the pMG 7.1-1 plasmid clone (K. Matthews and J.D. Barry, unpub. obs.), which showed only one complete 70 bp repeat present.

Together these observations are quite unusual, as telomeric VSG genes previously characterized are all flanked by large stretches of 70 bp repeats (see discussion, Chapter 3). Whether this reflects a feature specific to M-VSG genes is open to conjecture, and is discussed later.

Figure 4.1 Definition of the VSG gene on the Lambda MT 7.1-1 clone.



Panel A : 0.6% agarose gel analysis of Lambda MT 7.1-1 digests.

Panel B : Southern analysis with pTcV 7.1-14 cDNA fragment.

Digests :

Lambda MT 7.1-1 x ;

Lane 1. B, 2.E, 3.H.

Lane 4. pTcV 7.1-14 x P.

Lambda x H markers (M) are shown in Kb.

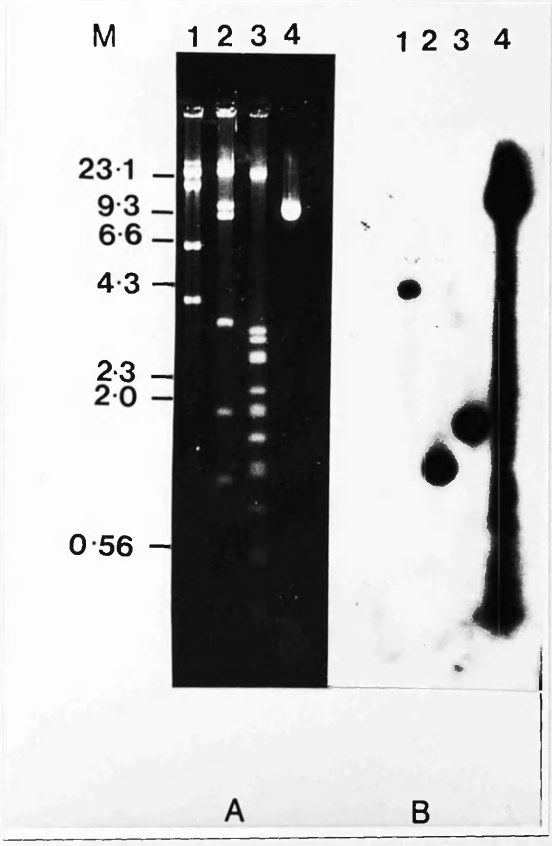
4.3. Comparative analyses between the M-VSG BC telomeres and a typical B-VSG gene expression site.

The lack of multicopy sequence in genomic Southern blots hybridized with probes derived from lambda MT 7.1-1, lambda MT 7.1-2 and lambda MT 1.22 B (Chapter 3, section 3.4, 3.5 and 3.6) is surprising, given the high density of representative sequences from multigene families present in B-VSG expression sites, which constitute the ESAGs that typify such sites.

The question arises as to whether any sequences related to the ESAGs of previously characterized B-VSG expression sites are present on the M-VSG BC telomeres? To this end specific ESAG probes from the 221 B-VSG expression site were hybridized with lambda MT 7.1-1 and lambda MT 7.1-2 under relaxed conditions (either 40°C or 45°, 16 hours hybridization) and washed to a posthybridization stringency of 3X SSC, 40°C. The low stringency employed was to account for ESAGs being a divergent multigene family and to realise the possibility that any such sequences on the GUTat 7.1 telomere may only show weak sequence similarity to those ESAGs previously characterized on other trypanosome telomeres, some of which have been reported to be only detectable at the low stringency used here (Cully et al., 1985). Figure 4.3.1 presents a composite of the results of these hybridizations with a 0.58 Kb, 0.96 Kb and 6.85 Kb, Pst I fragments, (from pTgB 221.8, a plasmid clone from the 221 expression site corresponding to ESAGs 1,2 and 3 respectively). Hybridization was also performed with a 2.6 Kb Bgl III/Sca I fragment from another 221 expression site clone, pTgB 221.12, containing the promoter for the B-VSG gene expression site (for details on pTg221.8, -.9 and -.12 see Kooter et al., 1987).

Figure 4.2.1

Definition of the 70 bp repeat region on Lambda MT 7.1-1.



Panel A : 0.6% agarose gel analysis of Lambda MT 7.1-1 digests.

Panel B : Southern analysis with a pTgB 221.2 x P/T fragment containg 70 bp repeat sequence, at a posthybridization stringency of 37 °C, 3X SSC.

Digests :

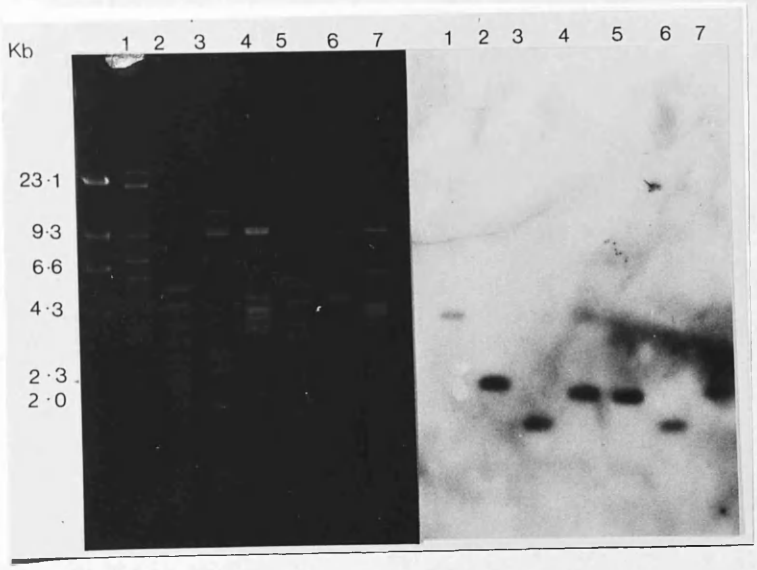
Lambda MT 7.1-1 x ;

Lane 1. B, 2.E, 3.H.

Lane 4. pTgB 221.2 x Pst I / Taq I.

Figure 4.2.2

Definition of the 70 bp repeat region on Lambda MT 1.22 B.



Panel A of the figure shows an ethidium bromide stained 0.6 % agarose gel analysis of lambda MT 1.22 B restriction digests. The corresponding Southern blot hybridized with the 70 bp probe is shown in Panel B. Posthybridization washing was to 3X SSC, 37^{0}C. Digests shown are;
Lane 1.Bam HI, 2. Hin cII, 3.Pst I, 4. Pvu II, 5. Bam HI/Hin cII , 6. Bam HI/Pst I, 7. Bam HI/ Pvu II.

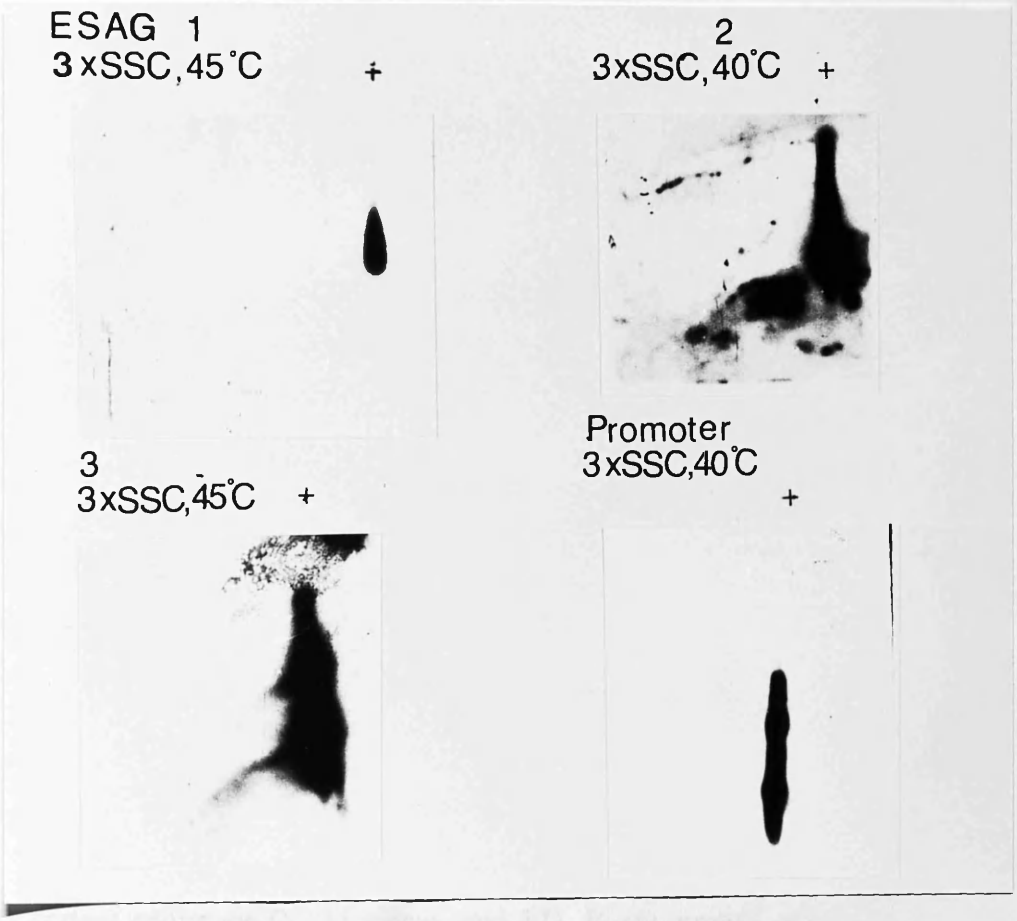
In all instances only the positive control lanes (marked + in Figure 4.3.1), corresponding to 221 expression site clones, give a detectable signal. No signals were detected in the lambda clone digests (situated to the left of the positive control on all the blots shown) even on prolonged exposure of the blots.

A drawback of this analysis is that it only uses specific ESAG probes, and not more extensive regions of the 221 expression site. Thus detection of intergenic regions that may be common to both sites may be precluded. This problem was tackled by hybridizing specific probes derived from lambda MT 1.22 B onto plasmid clones containing all of the 221 expression site elements. Hybridizations were performed at 5X SSC, 42°C, with posthybridization washing to 3X SSC, 42°C.

Hybridization to positive control digests were performed separately from those of the 221 expression site clones, to prevent 'mop up' of the probe by the control which might otherwise have occluded detection of positive signal.

The results of these hybridizations are shown in Figure 4.3.2, along with a map detailing the relevant probe used for each hybridization. The positive controls used for each hybridization shown in Figure 4.3.2 correspond to plasmid subclones derived from lambda MT 1.22 B, (the constructions of which are described in section 4.8). pMT 1.22-B2 (lane 2) corresponds to the subclone of the 6.05 Kb Bam HI fragment of lambda MT 1.22 B, while pMT 1.22-B3 (lane 1) corresponds to the 7.6 Kb Bam HI fragment of the lambda clone.

Figure 4.3.1 Screening of the Lambda MT 7.1-1 and 7.1-2 clones for ESAGs.



Probes used for this analysis are indicated above the respective blots, as are the posthybridizational stringencies. Positive control lanes (+) are indicated on the blots with lambda clone digests situated to the remaining lanes on the left hand side. The positive controls are;

ESAG 1. pTg 221.8 x P.

ESAG 2. pTg 221.8 x P.

ESAG 3. pTg 221.8 x P.

Promoter. pTg 221.12 x Bg/S.

When hybridized to Probe A (Panel A2) and washed to a stringency of 3X SSC, 42°C, only the positive control signal is observed, even on prolonged exposure (> 7 days). At the same stringency, probe B yields a strong positive control signal (Panel B1, lane 1) and no strong bands in all other lanes. On long exposures of > 1 week, a faint signal is just visible (marked * on Figure in Panel B1) corresponding to fragments from pTgB 221.8 and pTgB 221.9 which carry ESAG 3. These signals disappear at 3X SSC, 45°C, even on very long (> 2 weeks).

Exposures (Panel B2) indicating that they were background.

A further feature of this investigation is that neither probe cross hybridized indicating that they possess no common sequence.

The overall implication from these hybridizations is that a typical B-VSG expression site (that for 221) has no homology with the ILTat 1.22 M-VSG gene expression site. A similar analysis using pMG 7.1-1 produced the same result, with the exception of the 70 bp repeat sequence (K. Matthews and J.D. Barry, unpub. obs.).

While these analyses indicate a lack of gross structural and sequence homology between the two types of expression site, they do not exclude the presence of small regions common to both, perhaps as control elements in VSG expression. It is surprising, however, that no hybridization with the B-VSG expression site promoter was detected, given that this is highly conserved among such sites (Zomerdijk et al., 1990; Pays et al., 1989a), though it is still

Figure 4.3.2 Analysis of the lLTat 1.22 BC telomere for the presence of B-ES elements

Panel A1 : Map showing the Lambda MT 1.22 B fragments used as probes for the analysis.

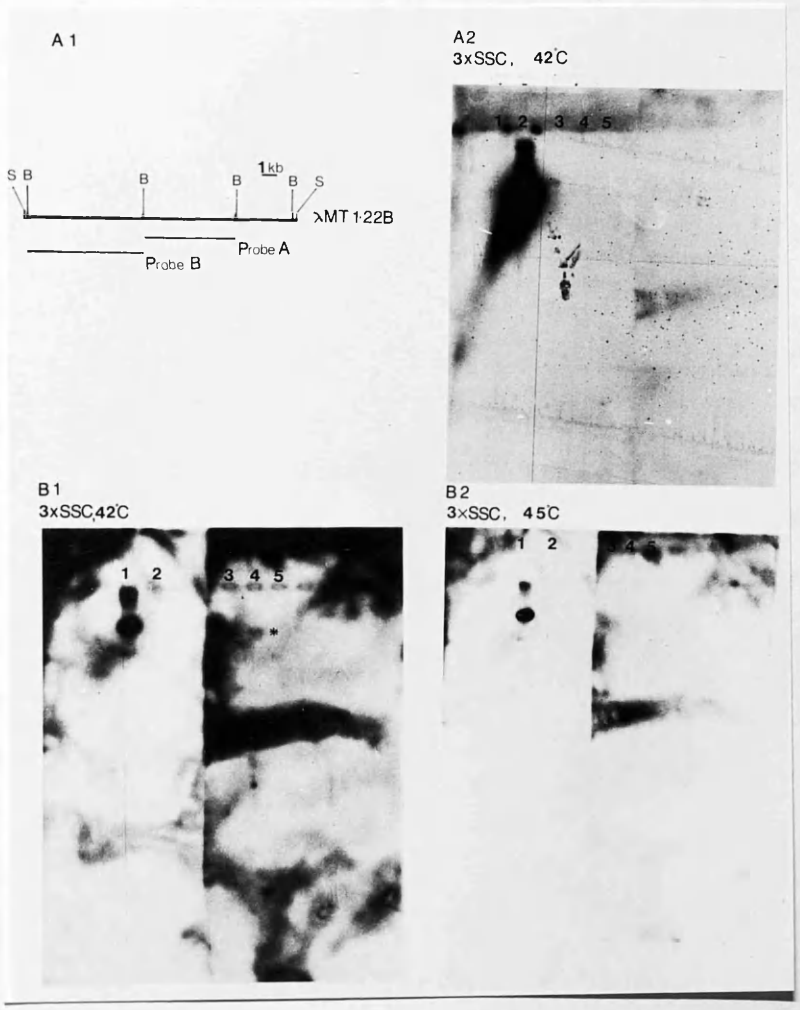
Panel A2 : Probe A analysis.

Panel B1 : Probe B analysis, 3X SSC, 42 oC.

B2 : Probe B analysis, 3X SSC, 45 oC.

Lane 1. pMT 1.22 - B3 x B.

2. pMT 1.22 - B2 x B.
3. pTg 221.8 x P.
4. pTg 221.9 x H.
5. pTg 221.12 x P


A1
S B
B
B
1kb
B S
λMT1·22B
Probe A
Probe B
A2
3×SSC, 42°C
1 2 3 4 5
B1
3×SSC, 42°C
1 2
3 4 5
B2
3×SSC, 45°C
1 2

formally possible that small sequence features common to both telomeres may have been missed at the resolution employed in these investigations. In total, these observations are quite unusual and further emphasize the singular character of these M-VSG gene BC telomeres.

One question that immediately arises from these observations is that of whether ESAG related sequences are at all detectable in the EATRO 795 genome? To answer this question, an EATRO 795 genomic Southern blot was hybridized with a 6.85 Kb stretch of pTg 221.9 containing ESAGs 2 and 3. Hybridizations were performed at 68°C for 16 hours and the filters were washed to a stringency of 0.1X SSC 68°C. The result of this hybridization is shown in Figure 4.3.3. The plasmid fragment carrying ESAGs 2 and 3 readily detects itself in the positive control (lane 1), and detects numerous bands in the genomic digest, illustrating the presence of these ESAGs in the EATRO 2340 genome. This is a strong argument in favour of the ILTat 1.22/GUTat 7.1 M-VSG gene telomere being radically different from previously characterized B-VSG gene expression site telomeres, all of which are typified by the presence of ESAGs

4.4. Northern analysis of the M-VSG gene expression sites for ILTat 1.22 and GUTat 7.1.

The basic aim of this section of experimentation was to see if any sequences present on the cloned telomeres were expressed during bloodstream infection where an ELC mechanism was being used for the expression of the respective VSG gene in either EATRO 2340 or 795. This investigation was to be extended using EATRO

795 procyclic form RNA, to gain a more complete transcriptional picture of the ILTat 1.22 BC telomere throughout the trypanosome life cycle, a strategy not feasible with the rodent adapted GUTat lines belonging to EATRO 2340.

A direct analysis of in situ expressing trypanosomes was not possible at the time of these investigations (due to the lack of available in situ expressing populations for ILTat 1.22) hence the use of the ELC expressors. This strategy, however, does allow for a number of issues to be addressed. ELC formation results in a cotransposed segment stretching 5' from within the EcoRI fragment containing the 70 bp repeat sequence, to the VSG gene being transposed to the bloodstream expression site (ref. Figure 3.6.3). As transcription occurs in the bloodstream site, BC telomeric sequences upstream of the 70 bp repeat region should therefore be transcriptionally silent, as telomeric expression within bloodstream expressors appears to be exclusive to the active expression site. Any detectable transcripts from the upstream regions of M-VSG telomeres during bloodstream expression would thus have to come from analogous sequences located elsewhere in the genome, either chromosome internally or from the active bloodstream expression site telomere; otherwise from a non-VSG transcription unit on the M-VSG telomeres.

The summated Northern data are presented in Figure 4.4. Probes used for each hybridization are illustrated in the map above the blots, save the positive control for the procyclic RNA hybridizations, that for TIM which is described in Chapter 2. In the Northern blots shown only two distinct probe fragments detected

Figure 4.3.3.

Southern analysis to assess the EATRO 795 genome for the presence of ESAGs.

Lane 1. pTg 221.9 x P.

2. EATRO 795 x P.

Marker size is shown in Kb.

RNA species in bloodstream Northerns (marked B).

The first of these, Probe 2, corresponds to the VSG gene and is shown alongside the 5' PstI/Hin dIII fragment from pTcV 7.1-14 (Probe 1), the positive control for the bloodstream Northerns, in the figure. In each instance a 1.6 Kb species is detected.

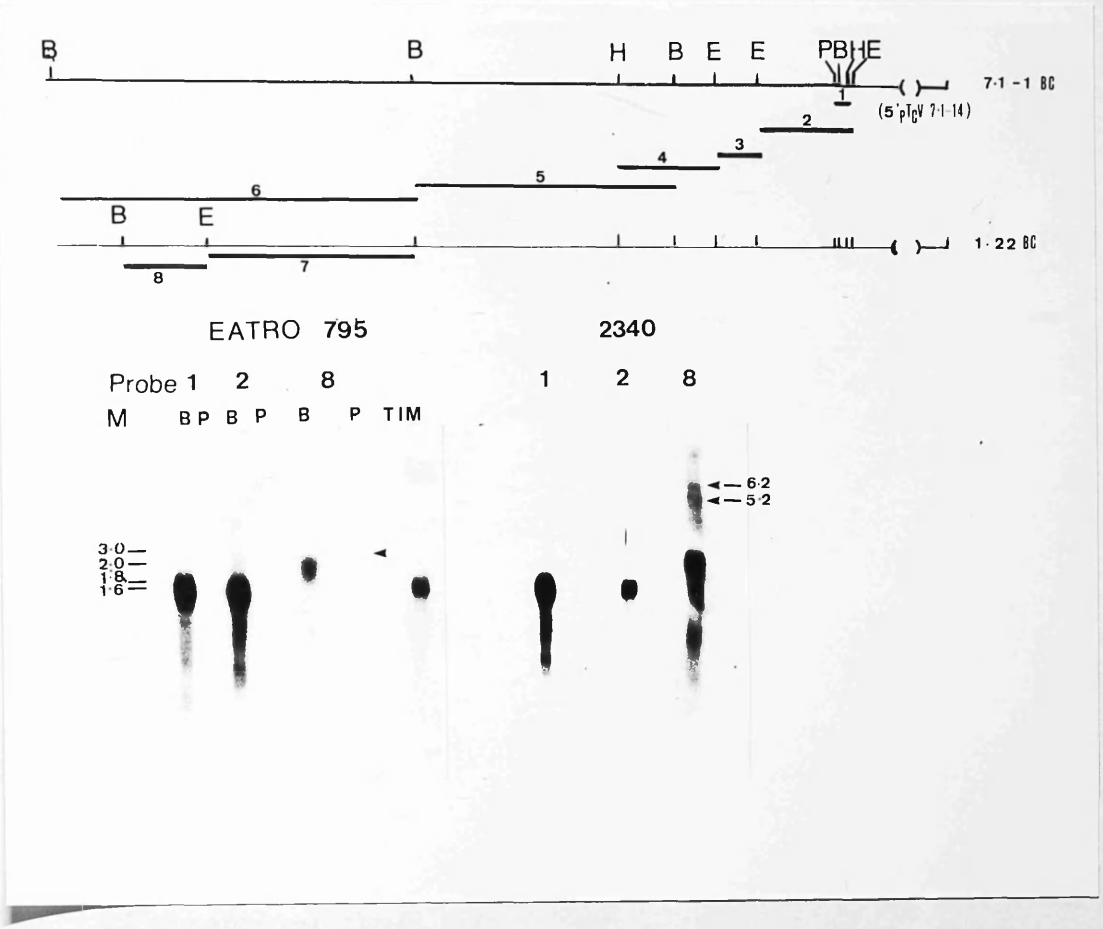
The second probe detecting an RNA species is Probe 8, common to lambda MT 7.1-2 and MT 1.22 B and corresponding to the genomic repeat sequence.

In each instance Probe 8 gives a weak signal comprising a background smear with faint bands at 2 Kb, 5 Kb and 6 Kb. As this probe detects multicopy sequence in genomic Southern blots the implication is that it may be detecting transcription from elsewhere in the genome, since the remainder of the upstream probes from each telomere failed to detect transcripts.

There is still the formal possibility that on the telomere represented by lambda MT 1.22 B and MT 7.1-2, this region does constitute a separate transcription unit, albeit one that produces a much less abundant transcript than the VSG gene, judging by the faintness of signal obtained with Probe 8, which might be considered unusual given the DNA sequence abundance of the probe sequence in the genome.

Figure 4.4

Northern analysis of the M-ES for ILTat 1.22 and GUTat 7.1.



Probe fragments derived from the respective lambda clones are delineated beneath the maps of the respective BC telomeres shown below.

The total RNA source for each analysis is indicated above the panels showing the Northern blots.

Lane. B = Bloodstream form total RNA.

P. = Procyclic form total RNA.

M. = Trypanosome poly A$^-$RNA markers.

Sizes are shown in Kb.

In the procyclic total RNA (lanes marked P) only Probe 8 detects an RNA species, at all the posthybridization stringencies employed. Once again the level of transcript abundance is low, and appears slightly less than that of the positive control, a probe for the single copy gene Triose phosphate isomerase (TIM) in lane TIM, which detected a 1.6 Kb transcript. In all instances probes of similar specific activities were used (data not shown).

Once again Probe 8 detects faint bands over a background smear, the reason for this is unclear.

4.5. Transcriptional analyses of the M-VSG gene expression site.

Due to the problems inherent in studying metacyclic form trypanosomes, as explained in the introduction to this chapter, a direct analysis of the metacyclic VSG gene transcription unit is not feasible. The model system outlined in the introduction, however, details an alternative strategy to studying these genes. The isolation of trypanosome clones derived from metacyclic forms through this model system, and expressing the VSG gene in situ, allows for transcriptional analysis of the M-VSG gene expression site through Nuclear Run-on experiments and Northern analyses.

4.5.1. Reverse Northern Analysis of GUTat 7.1.

In the absence of available EATRO 2340 metacyclic form trypanosomes expressing GUTat 7.1 in situ, it was deemed reasonable to utilise EATRO 795 RNA for a transcriptional analysis of the two

GUTat 7.1 BC telomeres in EATRO 2340, considering that GUTat 7.1-2 is an identical telomere to the ILTat 1.22 BC in EATRO 795 as judged by physical mapping.

Rather than perform a series of straight Northern hybridizations, as detailed in the previous section, it was decided to use a 'Reverse Northern' approach, synthesising ^{32}P- labelled first strand cDNA as a probe for Southern blots. This unorthodox approach was chosen firstly to overcome the limitations imposed by the small quantity of RNA available for analysis, a consequence of the difficulties still inherent when analysing M-VSG at the molecular level, and secondly to enhance detection of low abundance and any primary transcripts through cDNA:RNA hybridization, as opposed to RNA:RNA hybridization.

It must be stressed that this approach has a number of inherent errors in that there is the possibility that differential reverse transcription and labelling may lead to unequivocal results. Given the limited quantity of RNA, however, it was hoped that such an approach might give a basic indication of the size of the M-VSG transcription unit, and simultaneously the presence and abundance of ESAG transcripts from the EATRO 795 genome. ESAG transcription had been proposed as being pre-requisite for that of the VSG gene (Cully et al., 1985; Gibbs et al., 1988; Son et al., 1989) and it was still puzzling to find the absence of such sequences on the M-VSG telomeres.

The possibility still remains that such sequences are present further upstream of the region defined by each lambda clone, and

that the VSG transcription unit begins outwith the cloned region.

PolyA$^+$ RNA from EATRO 795 trypanosomes, expressing ILTat 1.22 in situ, was reverse transcribed and labelled using ^{32}P- dCTP. The resulting labelled cDNA was then hybridized with a Southern blot containing digests of lambda MT 7.1-1 and lambda MT 7.1-2, pTgB 221.8 and pTgB 221.9. The ethidium bromide stained gel used for the Southern blot is shown in Figure 4.5 Panel A. Hybridization was performed at 65°C for 48 hours, with posthybridization washing to 0.1X SSC, 68°C. Figure 4.5, Panel B depicts the autoradiographic analysis of the hybridization. A number of interesting features are immediately apparent.

Firstly, while the fragments harbouring the VSG gene on both clones (fragments 1.1 and 2.1, Panel B) give strong signals as expected, the region containing the 70 bp repeats (fragment 1.2) also gives a signal. This region is identical in both lambda clones and the signal strength is equivalent to that for the VSG coding region; though the signal for the 1.1 fragment from lambda MT 7.1-1 is weaker than the 2.1 equivalent Bam HI fragment from lambda MT 7.1-2, in the adjacent lane, both of which carry the VSG (see Panel B). The latter fragment's signal strength may be explicable in terms of the quantity of DNA in that lane being greater and the fragment size larger (1.9 Kb for the EcoRI lambda MT 7.1-1 fragment 1.1 as opposed to 3.8 Kb for the lambda MT 7.1-2 Bam HI fragment 2.1).

The area immediately 5' to the 70 bp repeat fragment in both clones gives a reduced signal relative to the downstream sequence (ref. Fragments 1.3 and 1.4 and 2.2, Panel B). In the lambda MT

7.1-2 clone clone digest (lane 5) the 5'Bam HI (fragment 2.3) gives an increased signal relative to fragment 2.2, but weaker than 2.1. This fragment contains a sequence highly repeated in the genome (Chapter 3, section 3.4) and is detectable in the Reverse Northern, but its signal strength does not equate with the abundance of this sequence in the genome. This analysis is inconclusive in this regard. While it is formally possible that transcription of the VSG gene is commencing from within fragment 2.2 (fragment 1.4 is the equivalent region on lambda MT 7.1-1), especially as the 70 bp repeat region is detectable in this analysis, a separate transcription unit may be found upstream, harbouring the repeat sequence element.

The implications of these results are discussed in detail in the discussion section of this chapter, along with an analysis of the potential errors in such an experimental approach.

The striking result from the analysis of the pTgB 221.8 and pTgB 221.9 clones is the abundance of transcription of all the ESAG sequences despite their absence on all the lambda clones. A direct comparison of the abundance of signal of these B-ES expresion site clones with those for the M-VSG BC telomeric clones is not of particular value given the potential for error through differential labelling and the factor of transcript abundance. These plasmids do, however, provide a good negative control for this experiment, as the pseudogene fragment (8.3) is not detected. The strongest hybridizing fragment in pTgB 221.9 is interestingly a 1.4 Kb band (fragment 9.1) corresponding to a region 5' of ESAG 3. This region defines a separate expression site gene (S. Graham and J.D.Barry, per. comm.) of unknown identity and function.

4.5.2

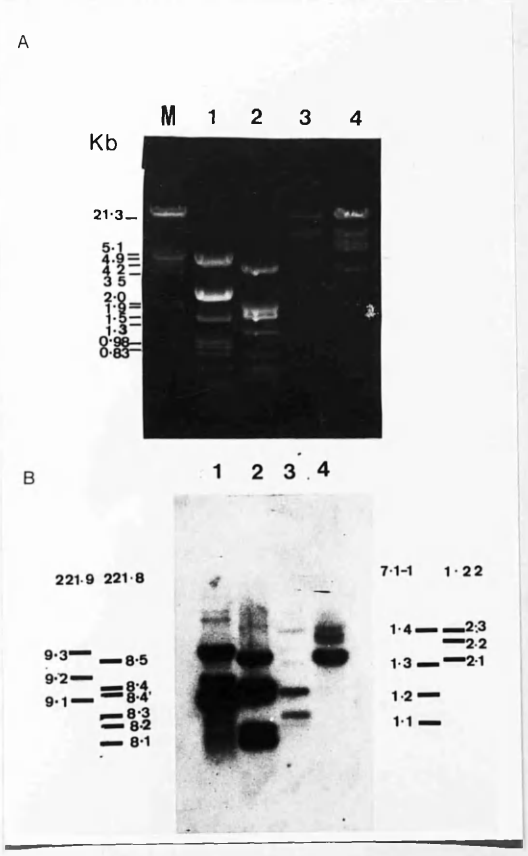
Nuclear run-on analysis for the ILTat 1.22 M-VSG gene expression site.

The second strategy used for a transcriptional analysis of the ILTat 1.22 M-VSG expression site, was Nuclear run-on. Unlike the crude approach of the 'Reverse Northern', the run-on could give a direct indication of any breaks in transcription occurring within the region of the telomere cloned in lambda MT 1.22 B, and thus define the size of the VSG gene transcription unit.

Only an initial experiment was possible, as during passaging of the trypanosomes to obtain stable in situ expression an ELC was produced. 'Run-on' reactions were performed both in the presence of 1mg/ml^{-1} alpha-amanitin, the other in 0μg ml^{-1} alpha-amanitin, a population belonging to EATRO 795 expressing ILTat 1.22 in situ (S. Graham and J.D. Barry, pers. comm.). This strategy permits some indication of the type of polymerase used in M-VSG gene transcription, as B-VSG genes have a transcription unit insensitive to the effects of alpha-amanitin. Thus, the use of a reaction in the presence of a high alpha- amanitin concentration would determine if both types of expression site used a similar polymerase activity. extension of this investigation is the inclusion of 221 expression site clones, pTg 221.8 and pTg 221.9, to allow comparison of their levels of run-on products with those from the M-VSG gene expression site.

Figure 4.5.1

Reverse Northern Analysis of GUTat 7.1.



Panel A : 1% agarose gel analysis of restricted clones for 'Reverse northern '.

M. Lambda x H/E markers.

Lane 1. pTg 221.9 x P/Pv.

2. pTg 212.8 x P/Pv.

3. Lambda MT 7.1-1 x E.

4. Lambda MT 7.1-2 x B.

Panel B : Hybridization of first strand cDNA to blot of gel shown in Panel A.

Included as further controls for the functioning of RNA polymerase I and II (the candidates for VSG transcription), were plasmid clones for the tubulin gene cluster (a Pol II transcription unit) and a region of rDNA (a Pol I transcription unit). The former should indicate whether the alpha-amanitin has inhibited RNA polymerase II, the latter polymerase I. Hybridizations following 'run-on' reactions were performed for a minimum of 48 hrs at 60°C in 3X SSC.

Posthybridization washing was to a stringency of 0.1X SSC, 68°C. In Figure 4.5.2, the Southern analysis of the 'run-on' in the presence of alpha-amanitin, yields 5 bands with the lambda MT 1.22 B clone (lane 3). The largest is 1.65 Kb (band A), corresponding to transcriptional activity at the 5' end of the clone, within the genomic repetitive region. This band partially masks a smaller 1.35 Kb band (band B) corresponding to a region at the 3' end of the clone (see map in Figure 4.5.2.1) between the 70 bp repeat region and the VSG gene. Another strong band at 410 bp corresponds to the 70 bp repeat region (band E). A fainter band at approximately 950 bp (band D) corresponds to a number of fragments. These comprise a Pst I / Pvu II fragment 5' of the 70 bp repeats (D1), a Pst I / Hin cII fragment (band D2) and a Hin cII fragment (band D3) further upstream. The final band, at 1 Kb, corresponds to the VSG gene carried on a PvuII fragment (band C). No other insert fragment yields signal.

These observations are quite surprising, in that most of the upstream fragments from this clone do not give detectable signals, in contrast to the characterized B-VSG gene expression sites which

cover approximately 40 to 60 Kb upstream of the VSG gene. Transcription of the M-VSG gene expression site does, however, also appear to be alpha-amanitin insensitive, (Figure 4.5.2.1, lane 4) though the signal is fainter than in the presence of inhibitor. The faintness of signal of the 950 bp and 1 Kb fragments (see map in figure) relative to the 70 bp repeat sequence, which yields a stronger signal from a smaller fragment, is indicative of transcription start occurring either within or in the 5' vicinity of the 3' most of these fragments (band D1) or within the 800 bp band D. Given that there is an apparent transcription gap between the two, it is likely that the VSG transcription unit is started within band D1 (see map in Figure 4.5.2.1), given that there are undetected fragments between all three fragments within the size range of band D_2. A more suitable digest of the lambda clone is needed to resolve this.

As none of the remaining upstream fragments yields detectable signal, with the exception of the genomic repetitive region at the 5' end of the clone, the VSG gene transcription 'start' must occur within the area defined by pMG 7.1-1, and 5' of the 70 bp repeat region in this clone.

'Run-on' in the absence of alpha-amanitin gives an identical result (lane 4). Panel C of this figure presents the results of the control hybridizations to plasmids carrying either ribosomal DNA (lane 1) or tubulin (lane 2). Both yield signal in the absence of alpha-amanitin, while only the former does so in its presence, indicating that both Pol II and Pol I were functioning in the 'run-ons', while only the latter functioned in the presence of the inhibitor, as deduced by the lack of tubulin signal in the presence of

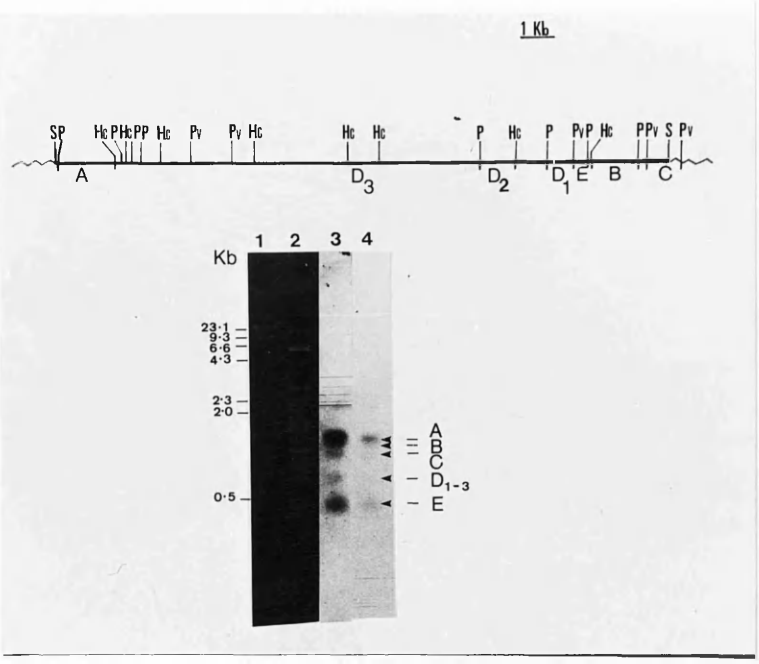
alpha-amanitin.

Figure 4.5.2.2 presents the results of the hybridization of the run-on products to plasmid clones from the 221 expression site, containing ESAGs. Though no ESAG sequences appear to be present on the ILTat 1.22 BC telomere within the region defined by MT 1.22 B, they are transcribed to high levels in the genome relative to the VSG gene transcription unit (Figure 4.5.2.2). This confirms the result of the reverse Northern analysis described in section 4.5.1. on the EATRO 2340 genome, for trypanosomes expressing GUTat 7.1.

ESAGs 1, 2 and 3, as well as the intergenic regions from the 221 expression site clones, pTg 221.8 and pTg 221.9, appear to be transcribed. The pTgB 221.8 clone contains an internal control for the 'run-on' experiments, with the presence of a VSG pseudogene (marked Ψ on the Figure, Panel A, lane 2). This is carried on a 0.9 Kb fragment, and is not detected in the 'run-on' analyses as expected.

Figure 4.5.2.1

'Nuclear Run - on' analysis of the JLTat 1.22 M-ES.



Lane 1. Lambda x H.

2. Lambda MT 1.22 B x Hc/P/Pv (1% agarose gel analysis)

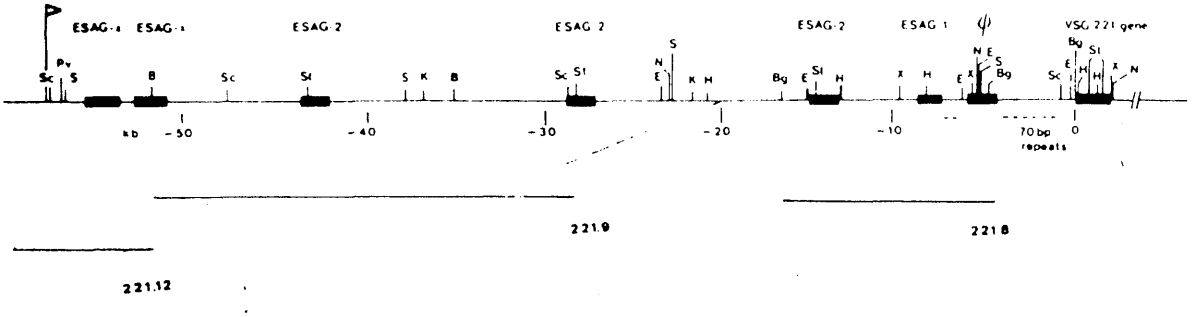
3. Lambda MT 1.22 B x Hc/P/Pv

(Southern analysis - alpha-amanitin).

4. Lambda MT 1.22 B x Hc/P/Pv

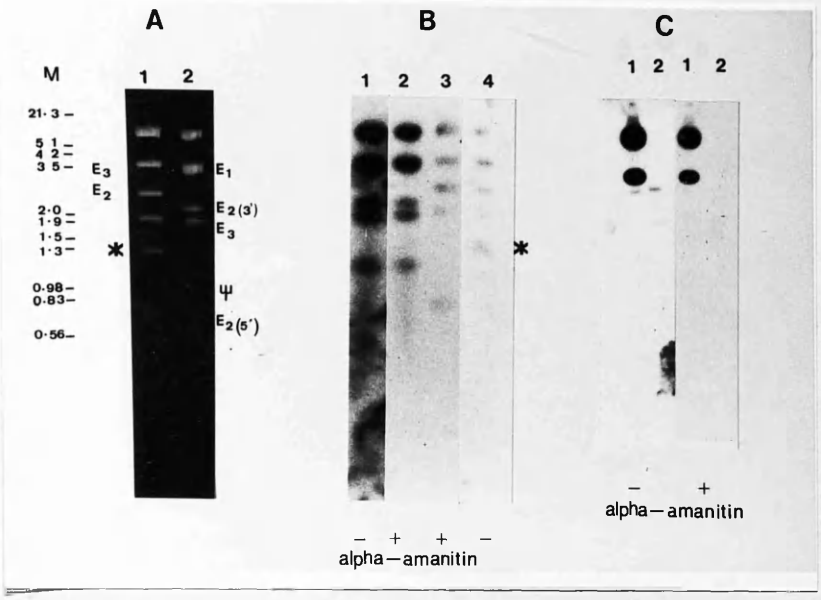
(Southern analysis + alpha-amanitin).

Physical Map of the Active VSG Gene Expression Site in Variant 221 Kooter et al.(1987)



Map of the 221 B-ES (after Kooter et al.,1987) showing the derived genomic clones used in the transcriptional analysis of the ILTat 221 M-ES.

'Run - on' analysis controls.



Panel A : 0.8% agarose gel analysis of 221 B-ES clones.

Lane 1. pTg 221.9 x C/H .2. pTg 221.8 x P/Pv.

Panel B : Southern analysis of the 221 B-ES clones.

Lane 1. pTg 221.8 x C/H (- alpha-amanitin).

2. pTg 221.8 x C/H (+ alpha-amanitin).

3. pTg 221.9 x C/H (+ alpha-amanitin).

4. pTg 221.9 x C/H (- alpha-amanitin).

Panel C : Southern analysis of rDNA and tubulin clones.

lane 1. pRIBO, (ribosomal Pol I control). 2.pTabT-1 (tubulin Pol II control).

E = ESAG . * = Alpha-amanitin sensitive fragment

The bands corresponding to fragments carrying ESAG sequences are labelled in the figure (E1 - E3). ESAG 1 is detectable on a 3.17 Kb PvuII fragment from pTgB 221.8, while ESAG 2 is readily observed on a 2.7 Kb fragment from the ClaI/HindIII digestion of pTgB 221.9. ESAG 3 is present on a 3.7 Kb fragment from the same digest.

In the presence of alpha-amanitin all the pTg 221.9 bands are visible (Panel B, lane 3) save a 1.3 Kb HindIII fragment (indicated by * in the figure, lane 4), which appears to be transcribed by an alpha-amanitin sensitive polymerase activity in EATRO 795 trypanosomes. This is in contrast to the 221 stock where its transcription is alpha-amanitin insensitive (Kooter et al., 1987).

All these observations strongly support the view that the ILTat 1.22 M-VSG gene transcription unit is quite distinct from that for B-VSG gene transcription, though both use an alpha- amanitin insensitive polymerase activity of undetermined nature. Though no ESAG sequences are present within the M-VSG gene transcription unit they are still transcribed at this stage of the life cycle, from elsewhere in the genome, and thus do not constitute a prerequisite part of the transcription unit for the VSG gene, as has previously been postulated (Cully et al., 1985); this is covered more fully later.

The observation that a 1.3 Kb HindIII fragment from the 221 expression site shows alpha-amanitin sensitivity in its transcription in trypanosomes expressing ILTat 1.22 confirms that a separate polymerase from that used for VSG genes transcribes this sequence as part of a separate transcription unit.

The 'run-on' analyses employed fail, however, to pinpoint transcription 'start' for ILTat 1.22 accurately, but do indicate the brevity of the transcription unit relative to that of a typical B-VSG gene.

4.6. Subcloning lambda MT 1.22 B.

Subcloning of lambda MT 1.22 B was effected by double digestion of the lambda clone with XbaI and Bam HI, followed by ligation of the digestion products into pUC19 and transformation of E. coli, DS 941. Recombinant clones were identified by hybridization with the 6.05 Kb Bam HI fragment from lambda MT 1.22 B. Hybridization was performed for 16 hours at 68°C, with posthybridization washing to 0.1X SSC, 68°.

Positive recombinants were streaked to purity on LA + 100ug ml^{-1} Ampicillin and subject to Birnboim and Doly miniprep analysis which revealed two distinct subclone types designated pMT 1.22-XB2 and pMT 1.22-XB2. The former contained a 3.7 Kb XbaI/Bam HI insert, the latter a 2.35 Kb double insert comprising a 2.05 Kb XbaI fragment and a 300 bp Bam HI/XbaI fragment. These were present in the correct genomic orientation in pMT 1.22-XB2, and this may have resulted from a partially digested fragment forming the recombinant, or a fortuitous ligation of two insert fragments.

Figure 4.6 presents the results of the restriction mapping of these two subclones, the insert fragments of which are shown

beneath the relative genomic map (panel A) along with the tabulated fragment sizes (Panel B).

The respective maps of each subclone indicate that pMT 1.22 -XB1 lies 5' to pMT 1.22-XB2 in terms of the genomic location of their respective insert fragments, as indicated in the maps shown in Figure 4.6 and as determined by the presence of a 300 bp XbaI/Bam HI fragment in pMT 1.22-XB2 common to pMG 7.1-1.

The subclone maps resolve the problem of the number and orientation of the small Hin dIII fragments present in this region of the telomere, indicating that the initial assumption of the 600 bp fragment being a single band (section 3.4, Chapter 3) was correct.

4.7. Sequence analysis of a putative M-VSG gene promoter region.

During the course of these investigations transcription 'start' for the ILTat 1.22 VSG gene on the BC telomere was determined to be within the 5' vicinity of the region defined by pMG 7.1-1. Studies of a series of metacyclic populations expressing ILTat 1.22 indicated that transcription start might be heterogeneous with respect to its position on the BC telomere in each population, though it generally appeared to be in the area defined by a cluster of small Hin dIII fragments (S. Graham and J.D. Barry, unpub. obs.), (see map, Chapter 3, Figure 3.3.2.1). While this area did not appear transcribed in the 'run-on' analysis in this investigation, the small sizes of the Hin dIII fragments facilitated sequencing of what appeared to be a transcriptionally important area of the telomere. Subcloning of the individual fragments into M13 mp18 and 19 sequencing phage

was achieved by isolating each fragment from a Hin dIII digest of pMT 1.22 -XB1 or pMT 1.22 -XB2 respectively, after electrophoresis on a 1.5% LMP agarose gel.

Recombinant phage were identified by hybridization with the insert fragment used as a radiolabelled probe. Hybridizations were performed at 68°C for 16 hours, with posthybridization washing to 68°C, 0.1 X SSC. Template DNA was prepared and sequenced by the Sanger chain termination method (Sanger et al., 1977).

The 300 bp fragment from pMT 1.22-XB1 which spans the XbaI cloning site and contains some vector linker sequence, was the first fragment to be sequenced, the data from which are presented in Figure 4.7.1.

The sequence indicates that the Hin dIII to XbaI region, belonging solely to the insert DNA, is 271 bp long while the total Hin dIII-Hin dIII fragment is 194 bp and contains 23 bp of vector sequence, this is indicated in the figure where the insert sequence start is labelled ⚑

A comparison of this 271 bp sequence against those contained in GENEBANK detected nothing of significant similarity, as determined by the WORDSEARCH package employed for the search. The sequence has a number of potential open reading frames shown in Figure 4.7.1. underneath the DNA sequence as putative protein sequences; these are labelled a - f. The largest of these is open reading frame f, which runs from position 142 for 174

bp and which has a potential start site outside the cloned region 5' to the Hin dIII cloning site. The second largest potential reading frame is 'a', which runs from the XbaI site at the 3' end of the insert for 113 bp. The remaining reading frames are not of any significant size, as can be seen from the figure. While no RNA species from this region of the telomere is detected in bloodstream or procyclic trypanosomes, it is formally possible that transcripts from one of these regions are present at another specific stage of the life cycle (eg:the epimastigote).

Neither of the two putative reading frames, a and f, finishing within the clone have any significant similarity with any protein sequence in GENEBANK. One sequence motif within this region is however, of note. This is a 15-mer running from position 216 to 231 which has 14/15 matches with a concensus 'ATGCTAATGACGATC' sequence (Figure 2.7.1, marked +++). The latter is common to eukaryotic viral gene control sequences and some of their host cell gene control sequences (rev. Sharp and Marciniak, 1988). Its presence in the trypanosome genome may be purely coincidental and of no functional significance. The position of this sequence within putative open reading frame f might support this view. If not, it could be of functional importance should it lie within the vicinity of the M-VSG gene promoter.

Attempts to sequence the Hin dIII fragments 3' of that in pMT 1.22 - XB1 raised an interesting artifact. Initial sequences derived from the 350 bp Hin dIII fragment of pMT 1.22 - XB2 are

Figure 4.6

Maps of the plasmid inserts for the pMT 1.22 - XB sub-clones.

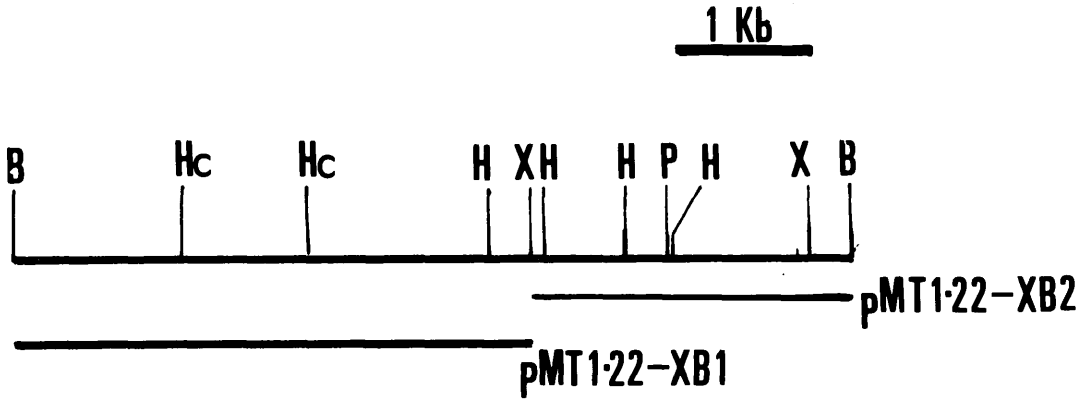


Table 4.6

Restriction mapping data from the analysis of the pMT 1.22-XB subclones.

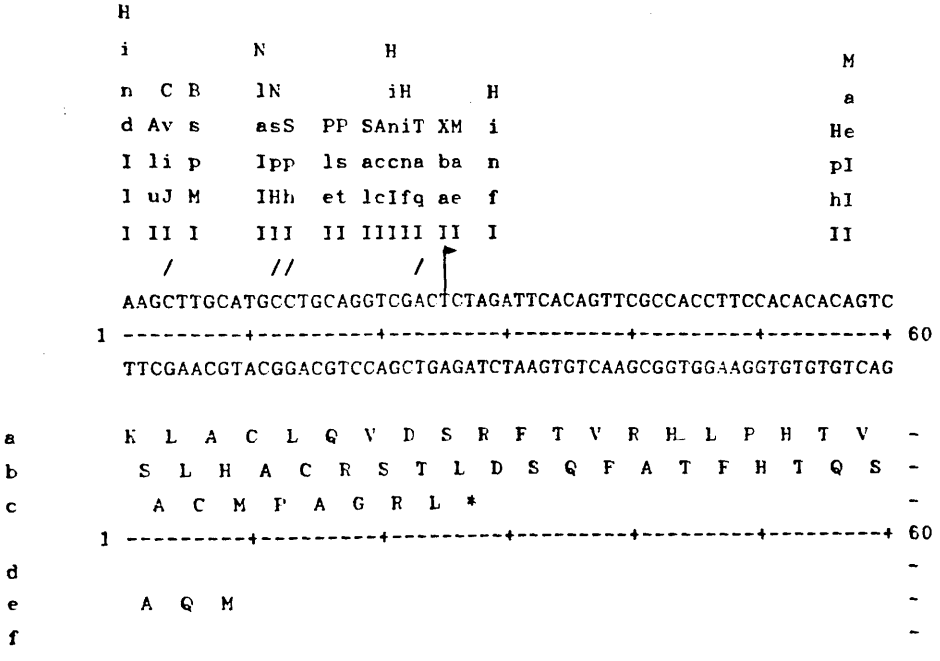
pMT 1.22-XB1	pMT 1.22-XB2
BX 3.65	BX 2.05, 0.3
BHcX 2, 1.55, 0.9	BHX 1.0, 0.6, 0.35, 0.3, 0.1
BHX 3.85, 0.3	BXP 1.05, 1.0, 0.3

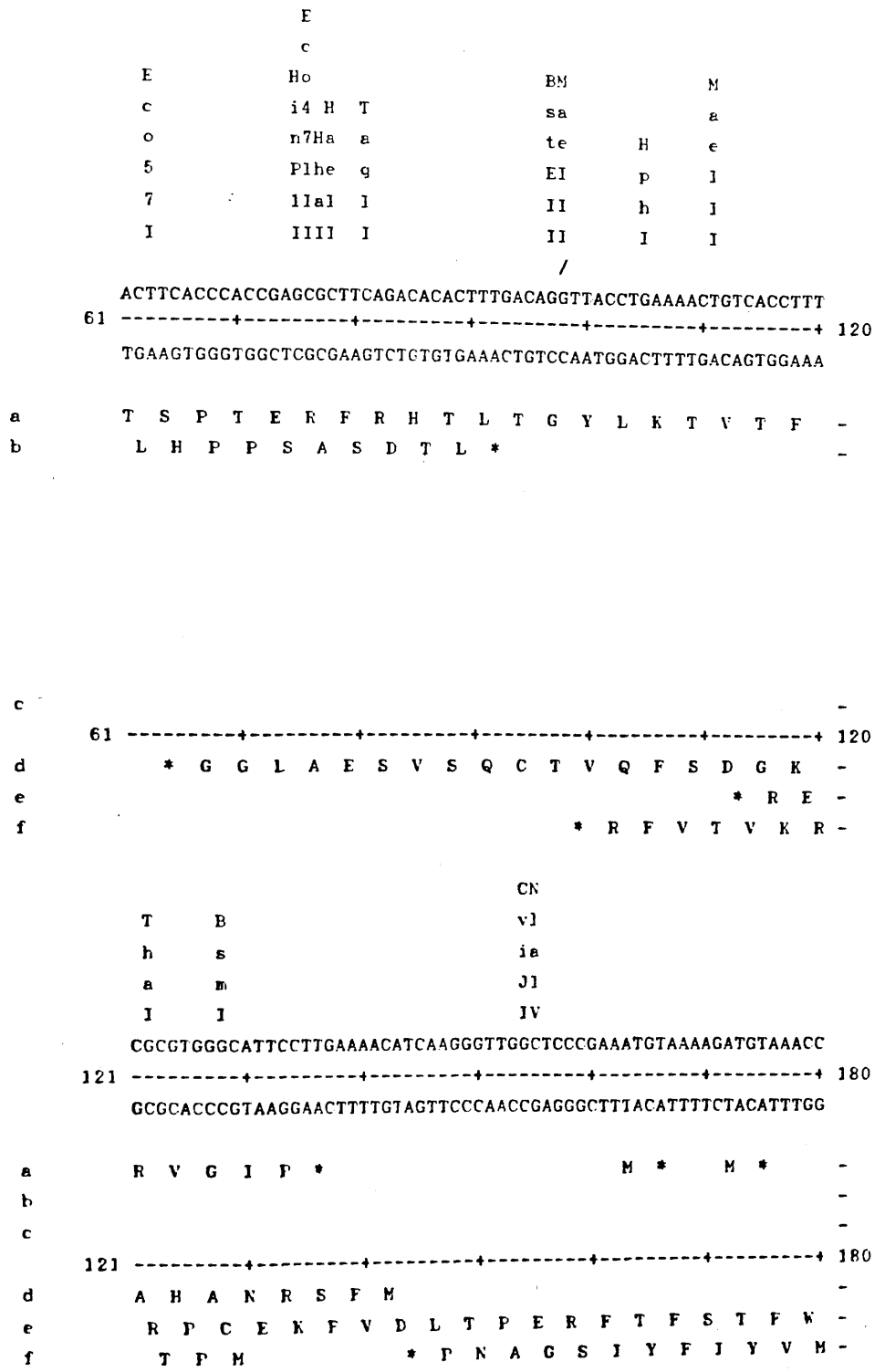
Fragment sizes are indicated in Kb.

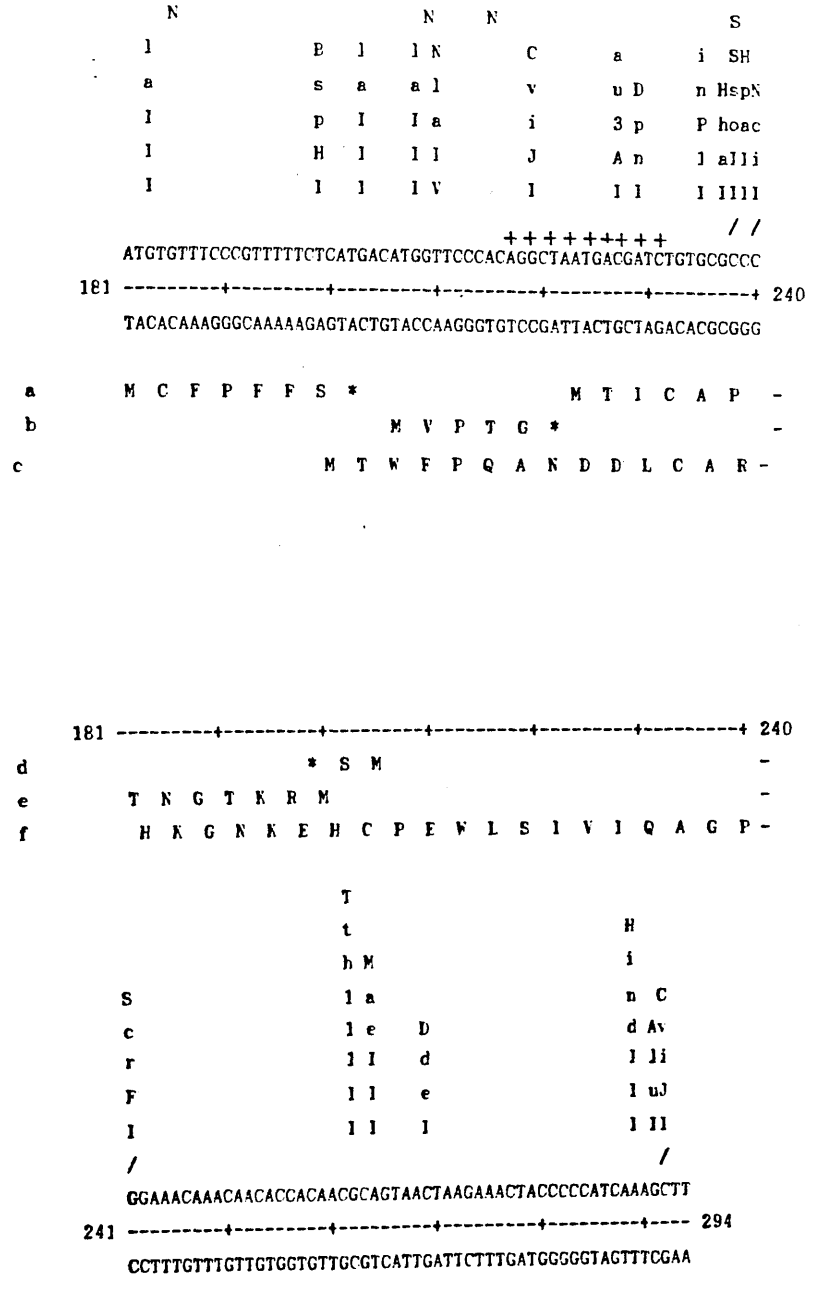
shown in Figure 4.7.2. The Hin dIII cloning site was never observed, which was initially attributed to the length of run being too long in conjunction with too little template being included in each sequencing reaction making lower bands too faint to see. Comparison of the sequences shown in Figure 4.7.2. with GENEBANK sequences detected highly similar sequences from a number of diverse sources. These are shown for comparison in Figure 4.7.3.1-3, and include (i) the pCAT 2 cloned sequence of the catalase gene from Maize (Bethards et al., 1986), (ii) the sequence of tox gene of Bacillus subtilis (Bishai et al., 1987) and (iii) cDNA fragment of a mouse gamma-globulin (Krawinkel, 1983), as well as (iv) the lac Z alpha peptide region of E. coli within the synthetic cloning vector pBGBS9B, (Spratt et al.,1986).

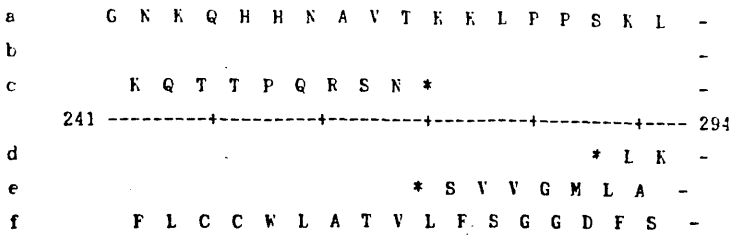
The percentage similarities between the three former sources and the pMT 1.22 - XB2 M13 subclone are very high, greater than 90% in each instance, and do not show the third base pair 'wobble' expected for coding sequences. The similarity with the pBGBS9B derived sequence is at the 89% level. The biological diversity of the former sources and the level of similarity suggest that this is not of functional significance, but rather reflects the presence of a

Figure 4.7.1 Sequence derived from pMT 1.22 - XB2 sub clone.




E
c
E
c
o
5
7
I
Ho
i4 H T
n7Ha a
Plhe g
llal I
IIII I
BM
sa
te
EI
II
II
/
M
a
e
I
H
p
h
I
I
ACTTCACCCACCGAGCGCTTCAGACACACTTTGACAGGTTACCTGAAAACTGTCACCTTT
61
120
TGAAGTGGGTGGCTCGCGAAGTCTGTGTGAAACTGTCCAATGGACTTTTGACAGTGGAAA
a T S P T E R F R H T L T G Y L K T V T F -
b L H P P S A S D T L * -
c -
61 120
d * G G L A E S V S Q C T V Q F S D G K -
e * R E -
f * R F V T V K R -
CN
T B
h s
a m
I I
vl
ia
Jl
IV
CGCGTGGGCATTCCTTGAAAACATCAAGGGTTGGCTCCCGAAATGTAAAAGATGTAAACC
121
180
GCGCACCCGTAAGGAACTTTTGTAGTTCCCAACCGAGGGCTTTACATTTTCTACATTTGG
a R V G I P * M * M * -
b -
c -
121 180
d A H A N R S F M -
e R P C E K F V D L T P E R F T F S T F W -
f T P M * P N A G S I Y F I Y V M -


ATGTGTTTCCCGTTTTTCTCATGACATGGTTCCCACAGGCTAATGACGATCTGTGCGCCC
181 ---------+---------+---------+---------+---------+---------+ 240
TACACAAAGGGCAAAAAGAGTACTGTACCAAGGGTGTCCGATTACTGCTAGACACGCGGG
a M C F P F F S * M T I C A P -
b M V P T G * -
c M T W F P Q A N D D L C A R -
181 ---------+---------+---------+---------+---------+---------+ 240
d * S M -
e T N G T K R M -
f H K G N K E H C P E W L S I V I Q A G P -
GGAAACAAACAACACCACAACGCAGTAACTAAGAAACTACCCCCATCAAAGCTT
241 ---------+---------+---------+---------+---------+---- 294
CCTTTGTTTGTTGTGGTGTTGCGTCATTGATTCTTTGATGGGGGTAGTTTCGAA



Potential open reading frames are indicated (a-f) in the right hand margin and shown beneath the DNA sequence. Nucleotide positions are indicated in both margins. Restriction sites are indicated above the DNA sequence.

contaminating DNA source. This is strengthened by the observation that the similarity with the pCAT 2 clone from Maize is outside the open reading frame, within the 3' untranslated sequence. A comparison with the same region of other catalase genes, as shown in Figure 4.7.4, indicates that the Maize sequence differs from the others significantly in this region.

The likelihood is that this region of sequence is due to a contaminant having entered the sequencing preparations, or that the M13 recombinant subclone has recombined with the lac Z region of the E.coli genome, alternatively it has been scrambled by the presence of an insertion sequence. Further plaques from the same plate used to generate the sequence 1 shown in Figure 4.7.2, produced sequence 2, 86% similar to sequence 1, which argues strongly for this being a foreign element and not part of the trypanosome genome. The percentage similarity between sequences derived from different plaques indicates the degree of scrambling that has occured during subcloning.

As a check to ensure that these sequences were not indigenous to the trypanosome genome, they were hybridized to Hin dIII digests of T.b.rhodesiense genomic DNA : both failed to show hybridization at a stringency of 0.1X SSC, 68°C (data not shown) indicating the foreign nature of the sequences. During the course of this analysis, however, it became necessary to abandon the strategy of sequencing this area of the telomere, as further analysis of in situ expressing ILTat 1.22 clones revealed that the transcription 'start' did indeed commence further 3' to this region, in agreement with the run-on analysis presented in section 4.5.2 (S. Graham and J.D. Barry unpub.

obs.). Partial sequence data from this region already existed (K. Matthews and J.D. Barry, unpub. obs.). It was therefore decided to investigate the nature of the genomic repetitive element which seemed to comprise part of a separate transcription unit outwith that of the VSG gene.

4.8. Subcloning of the genomic repetitive region of lambda MT 1.22 B.

The genomic repetitive element of lambda MT 1.22 B resided within a 1.85 Kb Eco RI/Bam HI fragment. As this is too large to clone for convenient M13 sequencing, it was decided to obtain a subclone of the 5' end of the lambda clone, to facilitate the production of a more detailed restriction map with the view to getting a more suitable fragment for sequencing. Isolation of a Bam HI subclone of the 7.6 Kb Bam HI fragment from the 5' end of lambda MT 1.22 B was achieved by screening a Bam HI plasmid bank of lambda MT 1.22 B in pUC 19 (courtesy of S. Graham) using the 1.85 Kb Eco RI/Bam HI fragment as probe. Hybridization was performed at 68°C for 16 hours, followed by posthybridization washing to a stringency of 0.1X SSC, 68°C.

Positive recombinants isolated from this screening were used to produce DNA by Birnboim and Doly minipreps and Eco RI/Bam HI double digested to isolate the fragment containing the genomic repeat. All recombinants analysed yielded a 1.4 Kb double digest fragment, not the expected 1.85 Kb fragment (data not shown). These subclones thus appeared to have a 0.45 Kb deletion in the 1.85 Kb Eco RI/Bam HI fragment from lambda MT 1.22 B.

Figure 4.7.2 Comparison of putative insertion sequences isolated during attempted sequencing of the ILTat 1.22 M-ES promoter region.

```
Percentage similarity : 89%

                .         .         .         .         .
  1 CTGTGTGAAATTGTTATCCGCTCGCAATTCCACACAACATACGAGCCGGA 50
                      ||||||  ||||||||||||||||   |      |
  1 ..................CCGCTCACAATTCCACACAACAACGGCCGGCAA 33
                .         .         .         .         .
 51 AGCATAAAGTGTAAAGCCTGGGGTGCCTAATG..AGTGAGCTAACTCACA 98
     ||||           ||||||||||||||||      ||||||||||||
 34 CGCAT......CGGTGCCTGGGGTGCCTAATGCAGTGCAGCTAACTCACA 77
                .         .         .         .         .
 99 TTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGT.GTG 147
    |||||||||||||||||||||||||||||||||||||||||||||||| |||
 78 TTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTG 127
                .         .         .         .         .
148 CCAGCTGCATTAATGAAT.GGCCAACGCGCGGGG..............  180
    ||||||||||||||||| ||||||||||||||
128 CCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGCGGCGGTTTGCGTA 177
```

Figure 4.7.3.1
Diagram of tox gene showing the relative positions of sequence similarity with the putative insertion element.

Percentage Similarity: 96.429

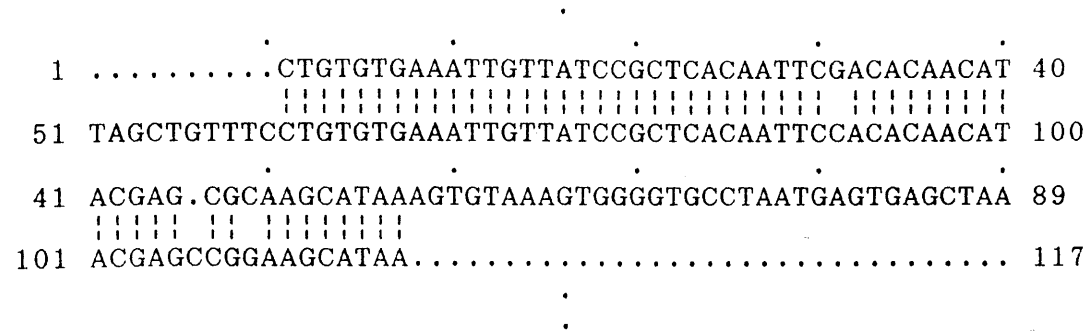


Figure 4.7.3.2.

Sequence of the maize Cat2 clone pCat2.1c compared with the foreign sequence.
Percentage Similarity: 88.827

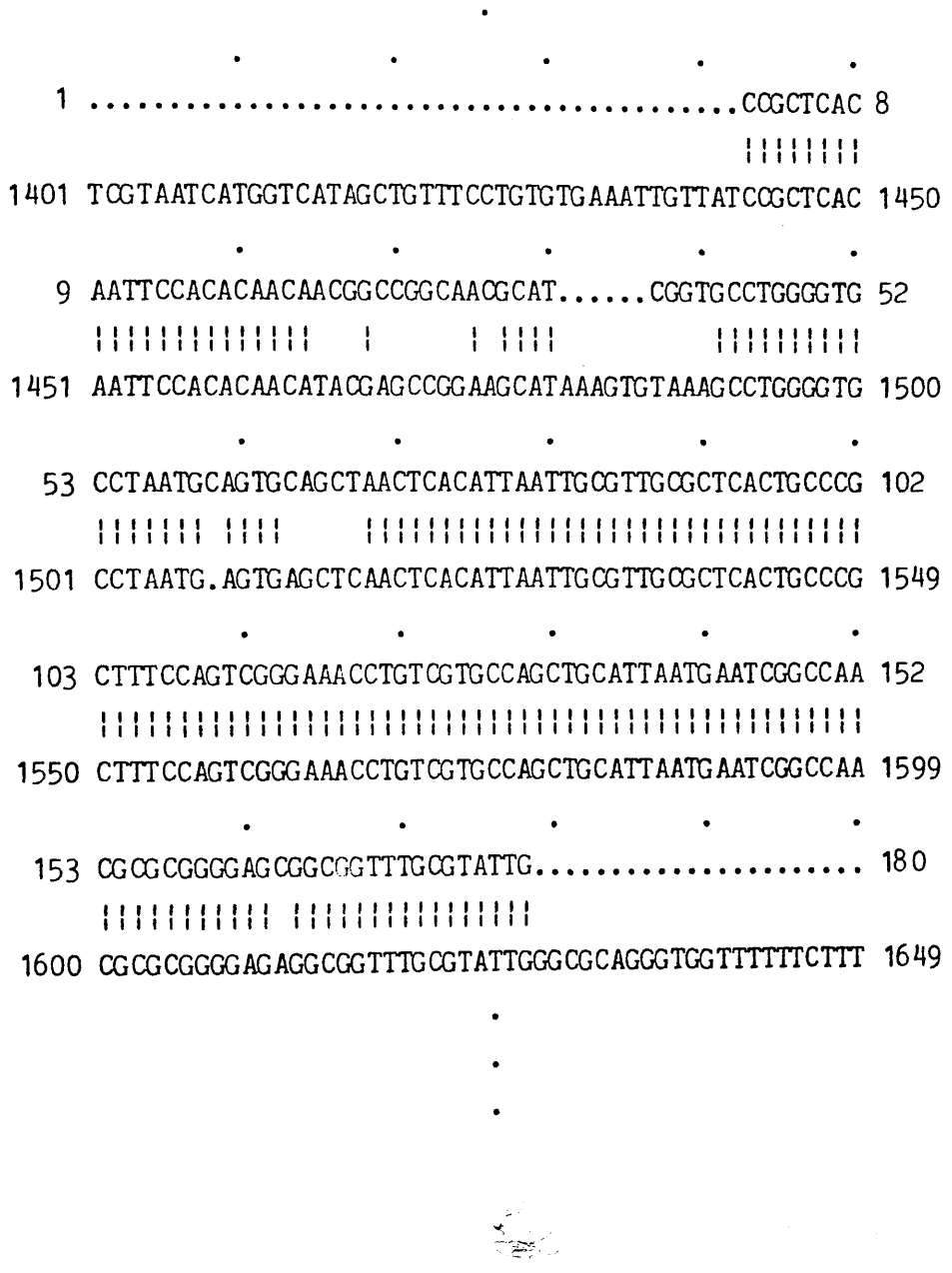


Figure 4.7.3.3
Mouse mRNA fragment (hybridizing weakly to heavy chain C-gamma-3 probe)compared with the putative insertion sequence.

Percentage similarity 98%

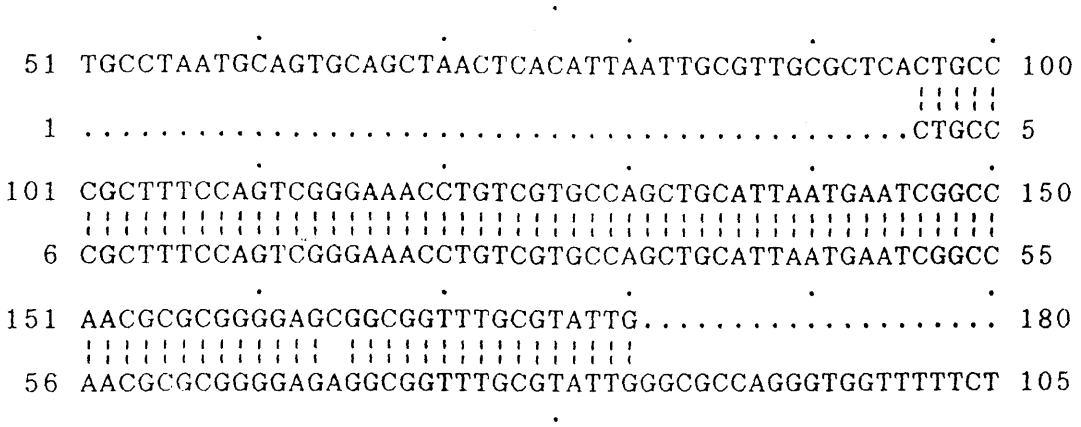


Figure 4.7.3.4
Cloning vector pBGS9- with the kanamycin resistance gene from Tn903, lacZ alpha-peptide of pUC9 and SspI fragment of bacteriophage f1 compared with the foreign insertion sequence.

Percentage Similarity: 89.326

```
201 CATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCA 250
                                    ||||||||||||||||
  1 ..................................CCGCTCACAATTCCA 15

251 CACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATG 300
    |||||||   |      | ||||          ||||||||||||||||
 16 CACAACAACGGCCGGCAACGCAT......CGGTGCCTGGGGTGCCTAATG 59

301 ..AGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 348
      ||||||||||||||||||||||||||||||||||||||||||||||||
 60 CAGTGCAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA 109

349 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGG 398
    ||||||||||||||||||||||||||||||||||||||||||||||||||
110 GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGG 159

399 GGAGAGGCGGTTTGCGTATTGGCTGCCTCGCGCGTTTCGGTGATGACGGT 448
    |||| ||||||||||||||||
160 GGAGCGGCGGTTTGCGTATTG..............................  180
```

Figure 4.7.4
Comparison of predicted amino acid sequences from the 3' end of catalase genes.

```
Amino acid position :
                  50                  60
MAIZE  : L C E I V I - R S Q F H T T - Y E P E A.
RAT    : V H P D Y G A R V Q A L L D Q Y N S Q K ....
BOVINE : V H P D Y G S R I Q A L L D K Y N E E K ....
HUMAN  : V H P D Y G S H I Q A L L D K Y N A E K ....
YEAST  : K D P K V K K R V T Q Y F G L L N E D L ....
```

To circumvent this problem a total EcoRI digest of lambda MT 1.22 B was "shotgun" cloned into pUC 19. Recombinants containing the genomic repeat sequence were isolated by transformation into NM 621, the E. coli host initially used to isolate the lambda clone and suitable for stabilizing repetitive sequences. Recombinants were identified by hybridization with the 1.95 Kb EcoRI fragment from lambda MT 1.22 B. Hybridizations were performed at 68°C for 16 hours, and washed posthybridization to 0.1X SSC, 68°C. Recombinants thus isolated were streaked to purity and checked by Birnboim and Doly minipreps. No recombinant isolated appeared to carry the 1.95 Kb EcoRI fragment.

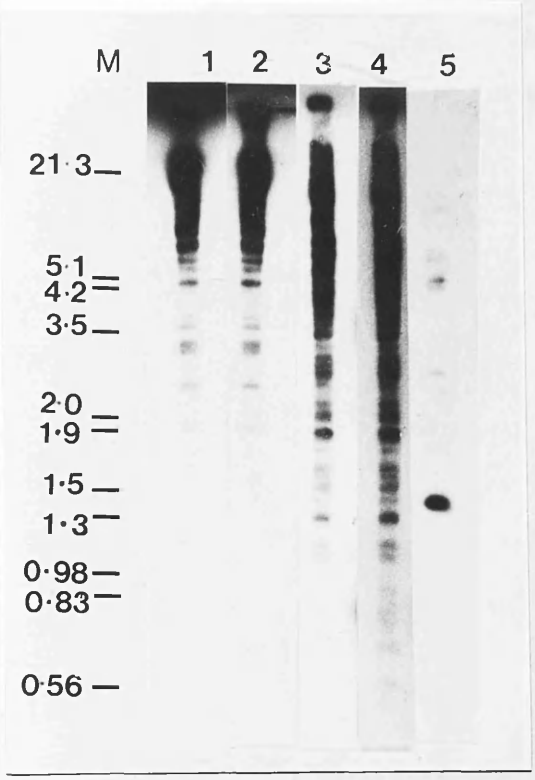
To overcome the problem of insert instability it was decided to attempt to delineate, if possible, a single unit of the repeat sequence within this region, by a combination of restriction mapping to discover sites in the repeat sequence, and comparisons with genomic Southern blots from digests with any enzyme found to be cutting within the repeat region. It was hoped this strategy would delineate the genomic repeat more finely within the 1.95 Kb EcoRI fragment and identify fragments containing the genomic repeat sequence, of a size suitable for sequencing.

One enzyme SphI, was found to cut within the region harbouring the genomic repeat sequence, liberating an 850 bp fragment from a representative subclone containing a 0.50 Kb deletion from the repeat region pMT 1.22 -RE (see Figure 4.8.1). Double digestion with Sph I and Kpn I isolated the insert as two fragments of 850 bp and 600 bp. Hybridization of these fragments against genomic Southern blots revealed that the deletion in the

subclone is within the 850 bp SphI fragment, the 600 bp Kpn I/Sph I double digest fragment detects a fragment of equivalent size in a similar KpnI/SphI genomic digest (Figure 4.8.1, lane 4). The Southern analyses in Figure 4.8.1 show genomic digests hybridized with either the 850 bp SphI fragment or the 600 bp SphI/KpnI fragment. Hybridizations were performed at 68°C for 16 hours, followed by posthybridization washing to 0.1X SSC, 68°C. It is only the 600 bp fragment that detects a fragment of equivalent size in the genomic digests. This fragment also detects a strong band at 1.3 Kb, possibly the true size of the deletion fragment of 850 bp. The 600 bp Sph I fragment gives an identical banding pattern (lanes 2 and 4) to the 850 bp probe, with the same digest; the common banding pattern indicates both contain, or are part of the same repeat unit. The 600 bp probe detects a 1.95 Kb Eco RI fragment (lanes 2) and strongly detects the positive control of the pUC 19 recombinant digested with EcoRI (lanes 5). These observations suggest strongly that the pUC 19 recombinant clone carries a 0.5 Kb deletion but that the 600 bp Sph I fragment is likely to be an unaltered fragment, judging by its detection in the Kpn I/Sph I a double digest (lanel 4). This fragment was sequenced following cloning into M13 mp 18, and utilizing the method of Sanger et al. (1977). The partial sequence of this clone is shown in Figure 4.8.2. This sequence detects nothing of significant similarity in a comparison with sequences listed in GENEBANK, and thus may comprise a novel trypanosome sequence element. As can be seen in the predicted protein sequences in Figure 4.8.2 no open reading frames of significant length are observed.

As the sequence is only partial the unit of repetition is undefined and thus further analysis is required to determine the character of this sequence.

Figure 4.8.1 Genomic Southern analysis of ILTat 1.22 BC telomere repeat sequence.



Lane 1. EATRO 795 x E vs 850 bp K/Sp probe.

2. " 600 bp " "

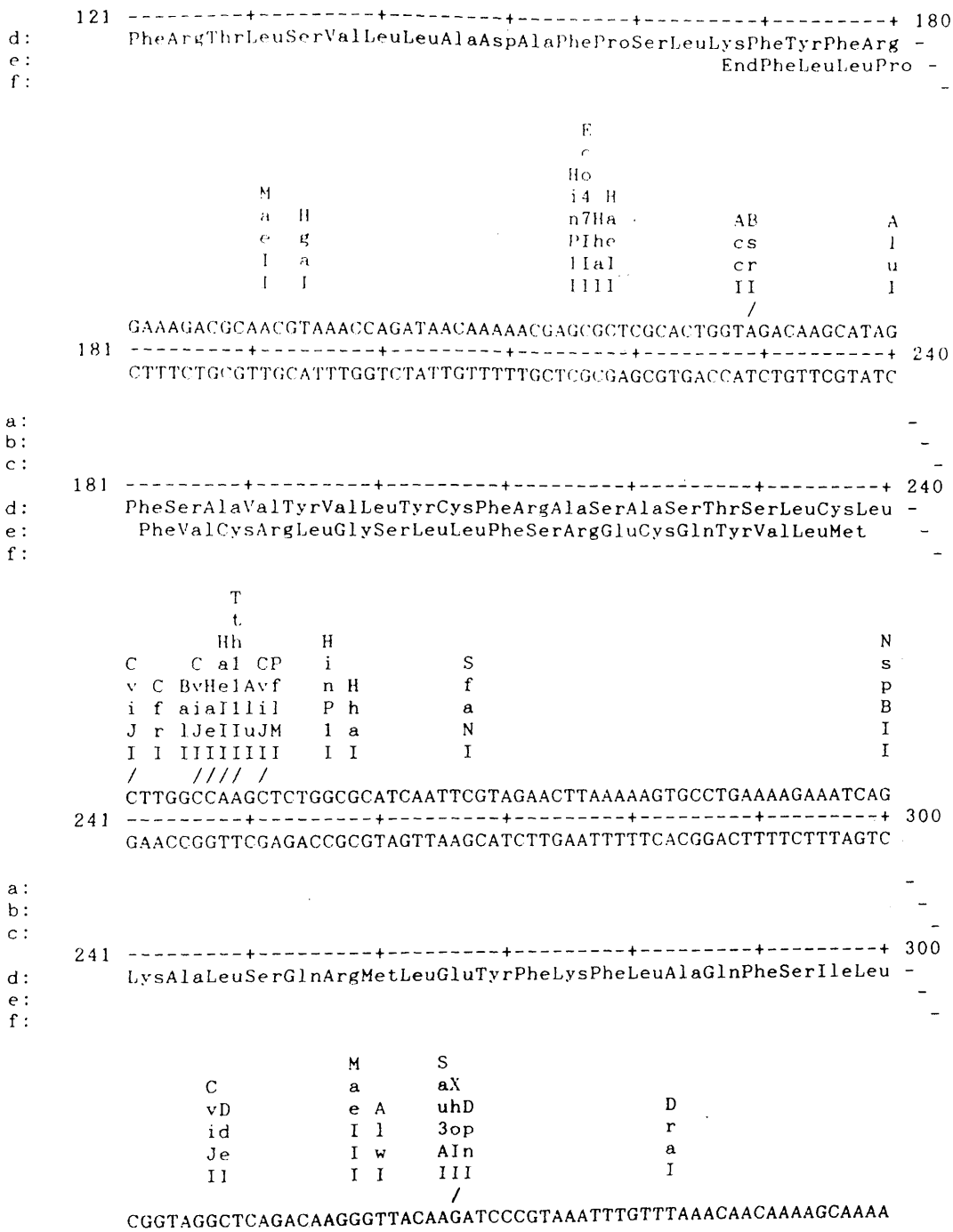
3. " K/S vs 850 bp K/S probe, 24 hour exposure.

4. " " 600 bp K/S probe, 48 hour exposure.

5. pMT 1.22-RE x E vs 600 bp K/S probe.

Figure 4.8.2 Partial sequence of pMT 1.22 - RE.

Possible open reading frames are labelled a-f in the right hand margin and shown below the DNA sequence. Nucleotide positions are indicated in each margin. Restriction sites are indicated above the DNA sequence.


121
180
d: PheArgThrLeuSerValLeuLeuAlaAspAlaPheProSerLeuLysPheTyrPheArg -
e: EndPheLeuLeuPro -
f: -
GAAAGACGCAACGTAAACCAGATAACAAAAACGAGCGCTCGCACTGGTAGACAAGCATAG
181
240
CTTTCTGCGTTGCATTTGGTCTATTGTTTTTGCTCGCGAGCGTGACCATCTGTTCGTATC
a: -
b: -
c: -
181
240
d: PheSerAlaValTyrValLeuTyrCysPheArgAlaSerAlaSerThrSerLeuCysLeu -
e: PheValCysArgLeuGlySerLeuLeuPheSerArgGluCysGlnTyrValLeuMet -
f: -
CTTGGCCAAGCTCTGGCGCATCAATTCGTAGAACTTAAAAAGTGCCTGAAAAGAAATCAG
241
300
GAACCGGTTCGAGACCGCGTAGTTAAGCATCTTGAATTTTTCACGGACTTTTCTTTAGTC
a: -
b: -
c: -
241
300
d: LysAlaLeuSerGlnArgMetLeuGluTyrPheLysPheLeuAlaGlnPheSerIleLeu -
e: -
f: -
CGGTAGGCTCAGACAAGGGTTACAAGATCCCGTAAATTTGTTTAAACAACAAAAGCAAAA

```
      301 ---------+---------+---------+---------+---------+---------+ 360
          GCCATCCGAGTCTGTTCCCAATGTTCTAGGGCATTTAAACAAATTTGTTGTTTTCGTTTT

a:                                                                    -
b:                                                                     -
c:                                                                      -
      301 ---------+---------+---------+---------+---------+---------+ 360
d:        ProLeuSerLeuCysProAsnCysSerGlyThrPheLysAsnLeuCysCysPheCysPhe -
e:                             EndLeuIleGlyTyrIleGlnLysPheLeuLeuLeuLeuVal -
f:                                                 EndValValPheAlaPheCys -

          CAG
      361 --- 363
          GTC

a:              -
b:               -
c:                -
      361 --- 363
d:        Leu -
e:         ??? -
f:           ??? -
```

4.9. How novel is the ILTat 1.22 BC telomere ?.

The outstanding feature of the ILTat 1.22 BC telomere in the genome of the EATRO 795 stock is its lack of numerous genomic multicopy sequence elements typical of B-VSG gene expression sites. This is a prominent feature of the ILTat 1.22 BC telomere in genomic Southern blots (section 3.6) used in mapping this telomere.

An immediate question arising from this observation is whether or not these same areas of the BC telomere detectable as single copy bands on genomic Southerns at high stringency (65°C, 0.1X SSC) identify more abundant copy number sequence elements at lower stringencies? Furthermore, are any of these ILTat 1.22 BC telomere sequences detectable in other trypanosome species undergoing antigenic variation? An answer to the latter question might hopefully be able to pinpoint regions of the telomere of fundamental importance to the antigenic variation process, should a positive answer be obtained.

In an attempt to investigate these questions, genomic Southern blots of various trypanosome DNAs were hybridized with lambda MT 1.22 B derived probes, and subsequently washed to a posthybridizational stringency of 3X SSC, 45°C. After autoradiography the blots were washed up in stages of increasing stringency (2X SSC, 50°C; 1X SSC, 60°C; 0.1X SSC, 65°C), with autoradiographic analysis after each step. The probes used for Southern analysis are shown in Figure 3.6.1 and correspond to Probes 4 and 5. The exclusion of Probe 6 from this investigation was due to its inherent multicopy status in the T. b. rhodesiense genome,

while Probes 1, 2 and 3 fell within the region defined by pMG 7.1-1, already investigated elsewhere (K. Matthews and J.D. Barry unpub.obs.) and were thus also excluded.

Initial investigation focused on Probe 4, a 6.05 Kb Bam HI fragment, within which region the 'start' of the VSG transcription unit was potentially defined (Section 4.5.2) hence one might expect it to contain sequence pertinent to the control of VSG gene expression and thus potentially found elsewhere in the genome and in other trypanosome species capable of antigenic variation. Figure 4.9.1. is a compilation of this investigation in which Probe 4 has been hybridized to both EATRO 795 expressor and non-expressor stocks for ILTat 1.22, EATRO 2340 expressor and non-expressor stocks for GUTat 7.1, T. congolense and T. vivax DNA. The Southern blot of the Probe 4 hybridization is shown in Figure 4.9.1 as a montage of the autoradiographs of the various stringency washes after hybridization.

Analysis of the autoradiographs presented in Figure 4.9.1. shows a strong positive signal at 6.05 Kb in lanes 1, 4, 5 and 6. Lane 5, an EATRO 795, ILTat 1.22 expressor, acts as a positive control for this investigation, as it was from this stock that lambda MT 1.22 B was cloned. No difference in the size of the fragment is observed between either both GUTat and ILTat DNAs, or between expressor and non-expressor populations. A number of background bands are visible in all these lanes at all stringencies below 0.1X SSC 68°C. The number and intensity of these bands decreases with increasing posthybridization wash temperature and decreasing SSC concentration. At 45°C, 3X SSC, at least 12 distinct bands are visible in these lanes.

The bands have been labelled a - l in the figure, and range in size from > 23 Kb to 1.8 Kb. These bands are most clearly visualized in lane 6, where they are marked and the corresponding labels shown in the right hand margin of the figure. Prominent among the background bands are a 7.8 Kb fragment (h) and a fragment which sizes at 3.8 Kb (e). The latter is more accurately sized at 3.3 Kb in other analyses with Probe 4 on better resolved gels (see section 3.3). Its size estimation in this instance may be a reflection on small salt differences between the marker digests and the Bam HI genomic digests, which accentuated the migration of the smaller marker fragments relative to the genomic digests.

The other background bands size at 1.8 Kb (a), 2.4 Kb (b), 2.5Kb (c), 3.3 Kb (d), 4.7 Kb (g), 7.8 Kb (h), 9.4 Kb (i) 11 Kb (j) 14 Kb (k) and >23 Kb (l)..

The latter three are not well resolved in lanes 5, 6 and 7, due to the strength of signal at this exposure. Bands i,j and l possibly constitute doublets as the fainter signals for these bands in lane 1 allows the discrimination of a lower band marked by the signal strength of the upper band in lanes, 4, 5, 6. These are arrowed in lane 1. Bands a, b and c, faint even at the lowest posthybridization stringency, disappear at a stringency of 2X SSC, 50°C. The remaining bands, though faint, remain until 1X SSC 60°C, but disappear at 0.1X SSC, 68°C, on the overnight (12 hr) exposure shown in Panel B. The bands at 7.8 and 3.8 Kb (e and f respectively) are faintly visible on longer exposure (data not presented).

These observations are intriguing, illustrating that the ILTat 1.22 BC telomere has sequence elements which at relaxed stringencies (e.g. 1X SSC, 55°C) can detect related elements elsewhere in the T. b. rhodesiense genome. This might not be unexpected if the telomeric region defined by Probe 4 harbours sequences involved in the control of antigenic variation or M-VSG gene expression.

Noteworthy in this respect, is the fact that the positive control signal (i.e. the 6.05 Kb BamHI fragment, f) and the background signals are identical in all the EATRO 2340 and EATRO 795 DNAs, indicating that this is a stable genomic region in the T. b. rhodesiense genome, given the temporal diversity of the origins of each DNA used (see Chapter 5) and the dynamism of the trypanosome genome. This observation lends weight to the assertion that this region harbours important sequence elements.

At low stringencies (< 1X SSC, 55°C) Probe 4 detects bands in the T. congolense (lane 3) and T. vivax (lanes 3 and 7) digests. Even at the lowest stringency employed these bands are faint, but at least 7 distinct bands are visible in lane 3 (T. congolense). These are labelled 1-7 in the figure.

Three of these bands, 1, 3 and 6 which are more permanent than the remainder, are faint even at 45°C, 3X SSC, and are clustered at between 14-23 Kb, with the exception of band 2 at 9 Kb. At 2X SSC, 55°C, the clustering of the larger bands resolves, and bands 5 and 7 become more distinct (see Panel B) while band 2 disappears. At 1 X SSC, 60°C all the bands in this lane cease to be detected.

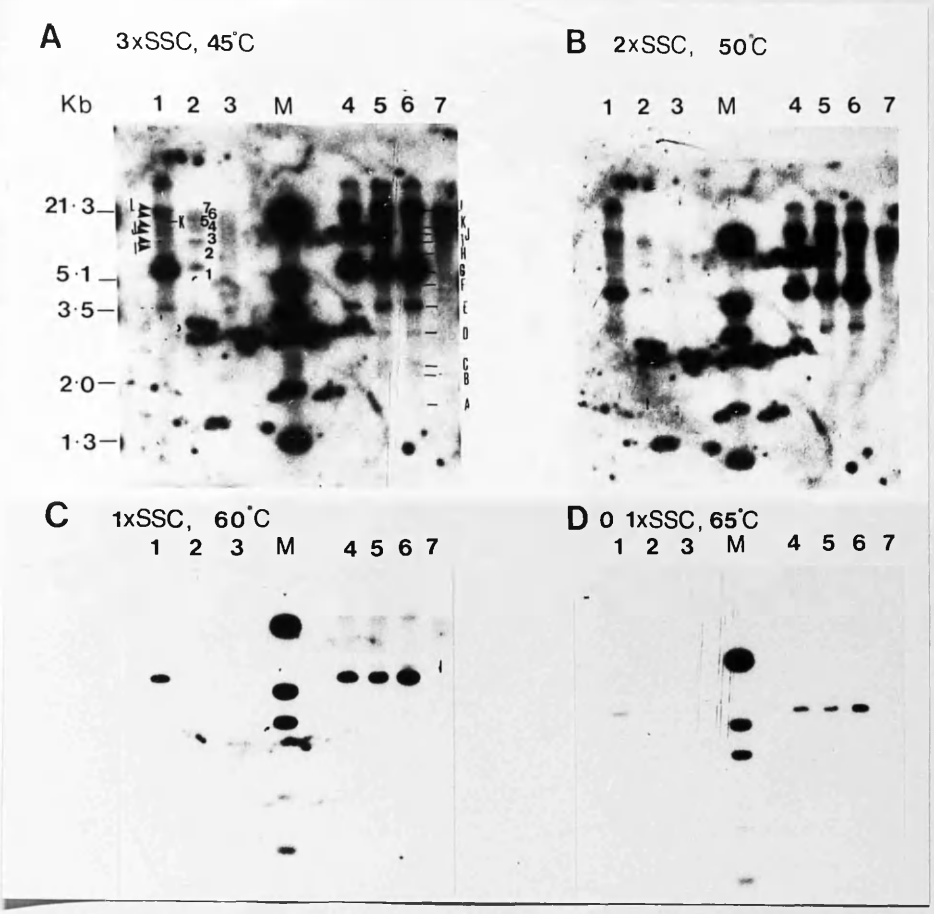
The significance of these observations is undetermined, but given the unusual physical character of this M-VSG gene expression site it is interesting to speculate that it may have elements in common with equivalent sites in T. congolense, none of which have yet been cloned.

The T. vivax digests (lanes 3 and 7) give unusual results in that no distinct bands are visible, but there is a smeared signal in both lanes at 3X SSC, 45°C. At 1X SSC, 60°C, the smearing in lane 3 (T. vivax, 'Swain') vanishes, while in lane 7 (T.vivax, 2.1) the smearing lessens and numerous faint bands are just visible, sizing from 3 Kb upwards. These all disappear at 65°C 0.1X SSC. It is difficult to attach significance to these observations save that related sequence elements to those present on the Il Tat 1.22 M-VSG expression site are also present in diverse trypanosome species undergoing antigenic variation. Although these elements appear structurally conserved to a first approximation in T.b.rhodensiense stocks, this does not allow for comparisons with T. congolense or T. vivax stocks.

Unlike Probe 4, Probe 5 does not detect a range of fragments in the EATRO 795 genome, even at the lowest stringency employed (3X SSC, 45°C). Figure 4.9.2 shows the result of hybridizing Probe 5 with Bam HI digested ILTat 1.22

Figure 4.9.1

Investigation of the status of upstream sequence on the ILTat 1.22 BC telomere : Probe 4 analysis.



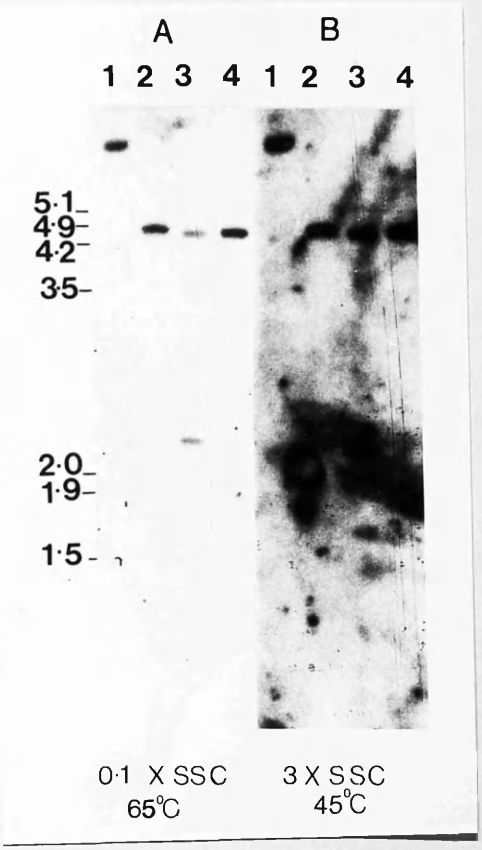
Stringencies employed for posthybridizational washing are shown above the individual panels.

Lane 1. GUTat 7.2 2. T.congolense. 3. T.vivax 'Swain'.

4. ILTat 1.21. 5. ILTat 1.22. 6. GUTat 7.1. 7. T.vivax 2.1.

Figure 4.9.2

Investigation of the status of upstream sequence on the ILTat 1.22 BC telomere continued; Probe 5 analysis.



Panel A : Posthybridization stringency, 0.1X SSC, 65 °C.

Panel B : Posthybridization stringency, 3X SSC, 45 °C.

Lane 1. ILTat 1.22 x B.

2. ILTat 1.22 x E.

3. ILTat 1.22 x H.

4. ILTat 1.22 x B/E.

genomic DNA and subsequent posthybridizational washing to 3X SSC, 45°oC and 0.1X SSC, 65°C. In both instances only a 7.6 Kb Bam HI fragment is observed. This probe failed to hybridize with T.vivax or T.congolense, (data not shown).

4.10. Attempted Chromosome Walk on the ILTat 1.22 BC Telomere.

A number of interesting questions can be posed for the ILTat 1.22 BC telomere. Does, for example it contain elements further upstream that allow it to be used as a B-VSG gene expression site? Are the ESAG sequences upstream? Where does the genomic repetitive region end? Attempts to answer these questions were initiated by trying to "walk" upstream on the telomere beyond the region cloned in lambda MT 1.22 B.

10^5 p.f.u. of the EATRO 795 genomic library were screened using the 6.05 Kb BamHI fragment from lambda MT 1.22 B. Hybridizations were performed at 68°C for 16 hours and followed by washing to a stringency of 0.1X SSC, 68°C. The probe detected 8 positive recombinants, all of which were picked and plaque purified. Restriction analysis of their DNAs failed to yield a fragment diagnostic of lambda MT 1.22 B (data not shown). The remaining two clones were identical to lambda MT 1.22 B.

The six other clones presumably correspond to regions of the genome that crossreact with the 6.05 Kb Bam HI probe from the ILTat 1.22 BC telomere, or are rearranged. The probe used to isolate these recombinants detects two prominent crossreacting

fragments in genomic Southern blots, at 7.4 and 3.3 Kb, neither of which appears to be present in the six recombinants. A larger band of > 20 Kb is detected at 1X SSC, 60°C in some genomic Southerns (Figure 4.9) and possibly these recombinants represent clones of this region. Given the number of phage screened (10 genome equivalents) it was decided not to pursue this strategy further, as it appeared another potentially unclonable genomic region was present upstream, an experience commonly encountered with trypanosome telomeres.

Discussion.

The dissection of the anatomy of the M-VSG gene telomeres presented in this chapter highlights a remarkable physical dissimilarity from previously characterized B-VSG expression site telomeres.

The distinctive nature of the BC telomeres for GUTat 7.1 and that for ILTat 1.22, relative to typical B-VSG gene expression site telomeres is emphasised by the lack of ESAGs on the former and any other common sequence, outwith the presence of telomeric and 70 bp repeat sequences and a VSG gene.

The structural individuality of the M-VSG expression site, exemplified by that for ILTat 1.22, is a direct reflection of the need for, and preservation of a distinct metacyclic-VSG gene pool and expression system, a topic discussed extensively in the Chapter 6.

The 70 bp repeat sequence of the ILTat 1.22/GUTat 7.1 telomere is defined within a 410 bp Pst I/Puv II fragment and contains just one complete 70 bp repeat (K. Matthews and J.D. Barry, unpub. obs.). This corresponds to the smallest such repeat region characterized to date.

In a Northern analysis of the three cloned telomeres only the VSG gene and the genomic repeat sequence found on lambda MT 1.22 B and GUTat 7.1-2, detected RNA species in total RNA from bloodstream trypomastigotes; none of the intervening sequences in each clone appeared represented. In total procyclic form RNA Northerns only the repeat sequence appears represented. The

detectable signal in both bloodstream and procyclic Northern blots is weak using fragments containing the repeat sequence as probe, much less than the respective positive controls, the VSG gene and TIM. The signal in each instance is present as a smear, not a discrete band, though in the bloodstream Northern three faint bands are discernable. Whether these represent mature and precursor transcripts, or a range of mature transcript sizes remains undetermined. Partial sequencing of this repeat sequence revealed nothing of similarity to anything present in GENEBANK. The difficulties encountered in subcloning the region of lambda MT 1.22 B carrying this sequence are puzzling. Subcloning using NM 621 as the E. coli host should have allayed any problems associated with insert or replicon instability. The repetitive isolation of subclones with a 0.55 Kb deletion suggests another factor at work, though it is possible that the initial lambda MT 1.22 B plaques used to generate DNA for the construction of the plasmid bands for subclone isolation, could themselves have carried the deletion. Instability of this region, however, was not observed at any other time.

Of the two approaches used for a more detailed transcriptional analysis of the metacyclic telomeres the Reverse Northern employed for investigating GUTats 7.1-1 and 7.1-2 is subject to a number of inadequacies and consequently inconclusive results. Principal among the drawbacks inherent in this technique are differential reverse transcription and differential labelling of given RNA species within the RNA pool. Nevertheless, there is still scope for drawing insight into the transcriptional activity on the GUTat 7.1 telomere and of characteristic B-VSG expression site elements within the genome of an *in situ* 7.1 expressor.

While the limited material from the in situ expressing population during the development of the model system precluded a comparative Northern analysis, the Reverse Northern results proved intriguing.

As expected the fragments carrying the VSG gene on either lambda clone were strongly detected (Figures 4.5.1), but significantly in the lambda MT 7.1-1 EcoRI digest so was the 70 bp containing fragment (band 1.1). The region 5' to this was not as strongly detected (band 1.4). As processing of the VSG gene primary transcript by trans-splicing occurs 3' to the 70 bp region, and as the 70 bp repeat sequence is not represented by a mature mRNA species, it seems reasonable to suggest that the primary transcript for the GUTat 7.1 VSG gene occurs within the 5' vicinity of the 70 bp repeat containing EcoRI fragment (band 1.1).

Signals detected upstream of this region in fragments represented in bands 1.3 and 1.4, may either be due to transcription 'start' being localized within such regions, or to the presence of another transcription unit.

In the instance of band 1.3, a 3.5 Kb EcoRI doublet, the latter seems likely, as EATRO 795 trypanosomes have only a single BC telomere for this VSG gene, that equivalent to GUTat 7.1-2, hence the observed signals for the upstream areas of GUTat 7.1-1 must be derived from another transcription unit.

In the absence of more 'Nuclear run-ons' it is not formally

possible to delineate the area of transcription 'start' for GUTat 7.1, but given the high transcript processing rate associated with trans-splicing (reviewed Borst, 1986) and the inability to detect 70 bp containing processing intermediates from B-VSG expression sites, it seems reasonable that the signal decrease within the 5' fragments adjacent to the 70 bp sequence is due to transcription commencing in the 5' vicinity of the latter (Figure 4.5.1).

The genomic repeat region of lambda MT 7.1-2 is once again detected at a low level relative to the VSG gene. Whether this is a reflection of its steady state level or is an artifact is undetermined.

A noteworthy feature of this analysis is that all the ESAGs typical of the B-VSG expression site are strongly detected in the experiment, indicating that their transcription within the EATRO 795 trypanosomes is occurring outwith the M-VSG expression site.

It is undetermined from this analysis whether their level of representation (see Figure 4.5.1) relative to that of the VSG gene is a true reflection of their steady state level, given the drawbacks inherent in the analysis. The pseudogene fragments on pTgB 221.8 are not detected in the experiment, and thus give an indication of the purity of the RNA used in the analysis and that no DNA contamination was responsible for any of the signals.

The 'run-on' analysis used for the ILTat 1.22 BC telomere clone provided a more direct approach to the transcriptional investigation of the M-VSG expression site. This latter investigation provided a clear demonstration of a transcriptional break on the

telomere and indicated once again that transcription was starting 5' of the 70 bp repeat sequence and within 5 Kb of the VSG gene. This result is quite singular given the size and complexity of the characterized B-VSG expression sites which span 60 Kb 5' of the VSG gene. The M-VSG gene transcription unit, however, did appear to be transcribed by an alpha-amanitin insensitive polymerase, though whether or not this utilized a similar promoter to that for B-VSG gene expression sites is open to speculation, as the latter was not detectable on lambda MT 1.22 B by Southern analysis. Sequence analysis, however, is needed to confirm this.

Another key feature from this investigation is that the repetitive region at the extreme 5' end of lambda MT 1.22 B is represented at a level similar to the VSG gene. This may constitute a separate transcription unit on the ILTat 1.22 BC telomere or be due to transcription from elsewhere in the genome. This latter hypothesis is relevant in discussing the B-VSG expression site sequences from 221 expressing trypanosomes. Whilst none of these are detectable on the lambda MT 1.22 B clone, they are readily detected in the 'run-on' and appear at high levels. Thus they are seemingly transcribed from elsewhere in the genome, and presumably subject to heavy post-transcriptional control, given their low transcript abundance relative to the VSG mRNA (S. Graham and J.D. Barry, unpub. obs.). A further interesting feature of this investigation of the B-VSG expression site sequences is the finding of an alpha-amanitin sensitive region defined within a 1.3 Kb Hin dIII fragment. Transcription of this fragment is alpha-amanitin insensitive in 221 expressing trypanosomes. The differential transcriptional sensitivity of this fragment between the two stocks may be due to the

fragment being mobile and its alpha-amanitin sensitivity on the 221 expression site being by virtue of its presence within the site. In ILTat 1.22 expressors, its location elsewhere would thus be envisaged as part of an alpha-amanitin sensitive transcription unit, and outwith that of the VSG gene. Support for the idea of ESAG transcription outwith the VSG gene transcription unit comes from the observation of ESAG transcription in procyclic trypomastigotes (Alexandre et al., 1988). The role of ESAG transcription on B-VSG gene expression sites, and the relationship of ESAG products to antigenic variation (Cully et al., 1986; Kooter et al., 1987; Pays et al., 1989a) in the light of all these observations requires more detailed examination.

It is the lack of any upstream sequence homology with B-VSG expression sites that distinguishes the singular character of the ILTat 1.22 M-VSG expression site. The latter, even at relaxed stringencies (3X SSC, 45°C) shows no cross homology with the 221 M-VSG expression site upstream of the 70 bp repeat region, and perhaps most surprisingly no gross structural homology with the B-VSG gene promoter. The M-VSG expression site does share alpha- amanitin insensitivity with the B-VSG counterpart, in respect of transcriptional activity, and the possibility remains that they will share small sequence features not readily detectable in the Southern analyses employed for these investigations.

The absence of ESAGs within the M-VSG transcription unit sheds doubt that their presence is required as an integral part of the VSG transcription unit (Cully et al., 1986) though an interactive role is not discounted, as they do appear to be expressed in the metacyclic form, outwith the VSG transcription unit. Any such role can not be

a universal characteristic of ESAGs given their presence in procyclic forms (Alexandre et al., 1988).

One immediate consequence of the physical dissimilarity between the M-ES and B-ES as determined in these analyses is that there is little scope for recombination between them, outwith the co-transposed unit that contains the B-VSG gene as an ELC in the B-VSG expression site. The brevity of the 70 bp repeat region in the ILTat 1.22 M-VSG gene expression site further decreases the likelihood of such an event.

This in effect isolates the M-VSG gene expression site telomere from the pool utilized during antigenic variation in chronic bloodstream infections. The dominance of M-VSG gene expression in early bloodstream parasitaemias (Turner et al., 1988) indicates the distinctive nature of M-VSG gene expression, a feature emphasised by the unique molecular structure of the ILTat 1.22 M-VSG gene expression site. The presence of the ILTat 1.22 gene in this site may result in its preferential activation in early bloodstream populations once a B-VSG gene expression mechanism has been adopted, given that an actively transcribed M-VSG telomere may have a chromatin conformation more readily accessible to recombination machinery once B-VSG expression is about to commence. This hypothesis is dissected more extensively in Chapter 6. Another consequence of the physical individuality of the ILTat 1.22 M-VSG expression site is enhanced stability of the M-VSG gene transcription unit. Given that frequency of reciprocal recombination and non-specific gene conversion events are inevitably decreased as a result of reduced cross homology with B-VSG gene telomeres, and that M-VSG gene

expression appears to be by an in situ mechanism (S. Graham and J.D. Barry, pers. comm.) thus again decreasing the likelihood of gross expression site rearrangements, it is not surprising that this M-VSG expression site is stable in comparison to the substantial alterations observed for B-VSG telomeres (Kooter et al., 1987). The ILTat 1.22 BC telomere has indeed been noted for its stability in both field and fly transmitted populations. The only observed difference for this telomere has been its duplication as opposed to an internal rearrangement (Cornelissen et al., 1985a).

The fundamental importance of a separate metacyclic repertoire (see Chapter 6) suggests that any apparatus of M-VSG gene expression harboured on the ILTat 1.22 BC telomere should lie outwith the co-transposed unit utilized in ELC formation, and thus 5' of the 70 bp repeats, and be present to some degree in related trypanosome species undergoing antigenic variation.

Hybridization of a 6.05 Kb Bam HI fragment from lambda MT 1.22 B (Probe 4, Figure 4.9) at low stringency to a series of genomic digests of T. b. rhodesiense. T. congolense and T. vivax DNA produced an intriguing result. The probe contains at its 3'end the putative area of transcription start for ILTat 1.22, and at the relaxed stringency of 2X SSC, 55°C detects multiple bands in digests corresponding to all three species, though in T. vivax the signals are faint; possibly a reflection on the evolutionary relationship between this, the most primitive of the three in terms of its developmental cycle, and T. b. rhodesiense, the most sophisticated in terms of its developmental cycle with respect to its adaptations for antigenic variation.

It is interesting to speculate if the banding patterns observed correspond to metacyclic specific sequences, as the number of bands is within the order of the number of estimated metacyclic telomeres. For T. b. rhodesiense and T. congolense the numbers, a minimum of 12 bands in the former and 7 bands in the latter, are within such limits. The complete M-VAT repertoire of one stock of T. congolense comprised 12 VATs (Crowe et al., 1983), while that of T. b. rhodesiense can be defined by no more than 27 separate antibodies (Turner et al., 1988).

No conclusions can be drawn from such an analysis, without comparison to other M-VSG telomeres. No such telomeres have been cloned for any T. congolense or T. vivax VSG gene. The ILTat 1.61 telomere in T. b. rhodesiense has now been cloned (K. Matthews and J.D. Barry pers. comm.) and a comparison with that for ILTat 1.22 will be of great significance in defining metacyclic specific sequence elements. The presence of sequences elsewhere in the T. b. rhodesiense genome and in those of T. vivax and T. congolense hybridizing with the ILTat 1.22 expression site probe, albeit at relaxed stringencies, indicates the presence of fundamentally important elements to M-VSG expression. It is thus significant to note that this region detects sequences in T. congolense and T. vivax, while 70 bp repeat sequence ubiquitous in T. brucei B-VSG gene expression site, does not (J.D. Barry, pers. comm.).

At high stringencies (0.1X SSC, 68°C) the probe detects two cross-reacting fragments on longer exposures sizing at 7.8 and 3.3 Kb. If these prove to be on other metacyclic telomeres it may provide an

explanation for the preferential interswitching between certain M-VATs during early bloodstream parasitaemias (J.D. Barry, pers. comm.). Such a hypothesis though, requires the degree of sequence homology between each telomere within the region of transcription start to be tied to a hierarchy of expression site prevalence (see Chapter 6). It is also possible that the M-VAT interswitches reflect a switch from in situ to ELC expression on the bloodstream of the host. Cloning of the relevant M-VSG expression site and comparison with that for ILTat 1.22 should give a degree of reference to the plausibility of such a scenario.

The unique physical characteristics of the ILTat 1.22 M-VSG expression site allow for an explanation of the stability of this M-VAT, and thus gives some insight into the maintenance of the M-VAT repertoire over time.

The following chapter pursues this topic coupled with the epidemiology of the ILTat 1.22 M-VAT, in order to examine the evolution of this telomere in natural populations. The stability of this expression site telomere provides an excellent opportunity to utilize it as a marker for tracing the origins of foci within epidemic infections in the field.

Chapter 5.

Comparisons of the ILTat 1.22 BC telomere in T. brucei stocks from different epidemic foci of infection.

Introduction.

African trypanosomes belonging to the T. brucei species are the agents responsible for 'Nagana' in both domestic and game animals, and 'sleeping sickness' in humans. As such T. brucei causes a major medical and consequential economic problem.

Traditionally this species comprises three sub-species, T. b. brucei, T. b. gambiense and T. b. rhodesiense, a classification based on epidemiological and parasitological criteria; principally their geographical distribution, host specificity and the pathology of their infection (Hoare, 1972). T. b. gambiense and T. b. rhodesiense constitute the human infective African trypanosomes, the former appearing better adapted to man and causing the chronic form of human trypanosomiasis, the latter causing 'acute' sleeping sickness typical of East Africa.

T. b. gambiense is found in Central to West Africa coincident with the range of habitat of 'riverine', tsetse flies (Glossina palpalis, G. tachinoides). T. b. rhodesiense is transmitted by tsetse flies of the G. morsitans group, habitual in the Savannah and lake-shore regions of East Africa (Ford, 1971). T. b brucei was initially distinguished as a separate subspecies solely on the basis of its human infectivity via the unethical practice of inoculation into volunteers, as T. b. brucei is morphologically indistinguishable from both T. b. gambiense

and T. b. rhodesiense and is sympatric with the latter.
The classification of these three groupings on the basis of host specificity is not of particular value, as all three appear to be enzootic, and have been isolated from domestic and game animals which act as reservoir hosts (Tait et al., 1984).

While isoenzyme (Tait et al., 1984), and RFLP (Paindavoine et al.,1989) analyses indicate that T. b. gambiense is a sibling or sub-species of T. b. brucei, the separate sub-species status of T. b. brucei and T. b. rhodesiense is not apparent. The latter appears to be a specific set of variants of the former which shows resistance to the lytic activity of human serum in vitro (Tait et al, 1985), and thus both appear to be variants of the same species.

T.b.rhodesiense populations within a given focus show homogeneity with respect to enzyme variants (Tait et al, 1985) and M-VAT repertoire, indicating that within a given focus of infection T. b. rhodesiense populations belong to a single serodeme (Barry et al, 1983). T. b. gambiense, somewhat in contrast, shows a great degree of homogeneity between antigen repertoires of geographically distant populations (Gray, 1972; Paindavoine et al, 1989) and most of the endemic foci in West Africa may contain one major serodeme (Babiker, 1981). This is reflected in the size of the genome, which is 70% that of T. b. brucei, and VSG gene repertoire which correspondingly is estimated to be about 50% smaller (Dero et al, 1987), a possible consequence of the evolution from an ancestral T. b. brucei human infective variant in West Africa, which became genetically isolated from other tyranosome populations in that area. Relevant to such a hypothesis is the observation that T.b.gambiense

grows more slowly than T.b.rhodesiense (J.D. Barry, pers. comm.) and hence may require fewer VSGs. Alternatively T.b. rhodesiense may be evolving much more rapidly and consequently expanding its VSG repertoire at a commensurate rate.

Diagnosis of 'acute' sleeping sickness from East Africa, however, stems from an initial report by Stephens and Fantham (1910), and is not thought to have been endemic prior to this. Sporadic epidemics of 'acute' sleeping sickness have been reported since then across the tsetse belt of East Africa, typified by the increased virulence of the disease with its Northward geographical spread (Ormerod, 1961). Currently the disease is epidemic around the Northern shore of the lake Victoria, where successive foci infection have appeared from 1940 to the present day.

Molecular evidence suggests that T. b. rhodesiense has evolved independently from T. b. brucei and not from T. b. gambiense, given the latter's reduced genome size and complexity (Dero et al., 1987). T. b. gambiense appears the most ancient of the two human infective forms, being more adapted to its human host, as is apparent from its reduced virulence and greater chronicity of infection, resulting in a better predisposition to uptake from the host on a further tsetse fly bite.

The question remains, however, as to whether the origin of T. b. rhodesiense reflects a single event and subsequent evolution of this second human infective form, or whether it has arisen by several independent events of this sort, or if such variants arise de novo, leading to the periodic epidemics associated with the disease in East

Africa and reflected in the homogeneity of enzyme variants (Tait et al., 1985) and the presence of just a single major serodeme in given foci of infection (Barry et al.,1983). Thus, it would be envisaged that there is an independently arisen deme of T. b. rhodesiense in each focus, which lies dormant (ie. in reservoir hosts) until perturbance of local conditions which affect the epidemiological balance eg. flooding, war, etc.

Evidence in favour of the latter hypothesis has come from the analysis of kDNA (Borst et al, 1981) and isoenzyme variation (Gibson et al, 1980) which indicate a disparity between stocks of trypanosomes derived from northern and southern regions of the East African section of the tsetse belt. This implies that northward spreading of the disease this century appears to be a consequence, the Northern epidemics having arisen 'de novo' and not by spread from a Southern endemic focus of infection.

The dynamism of the trypanosome genome and the substantial genomic rearrangements associated with antigenic variation (for a review see Borst, 1986) are problematic when seeking to obtain conclusions from differences observed in a comparison of geographically isolated stocks of T. b. rhodesiense. As mentioned previously, however, the M-VAT repertoire within given foci of infections tend to be homogeneous, and though this repertoire will change slowly with time (Barry et al., 1983), it provides a good marker for tracing the origin of a given trypanosome population. In this respect the ILTat 1.22 BC telomere offers a considerable advantage as a genomic marker for tracing the origin and evolution of an endemic focus. The unique physical characteristics of this

telomere give it exceptional stability within the trypanosome genome in comparison to other telomeres, as detailed in the previous chapter. As such, the use of this telomere as a comparative marker between trypanosome stocks isolated from geographically distinct foci of infection, and from human, animal and tsetse sources of a given focus, provides an opportunity to examine the origin of each focus of infection. This allows a number of questions to be posed:

(i) Does each epidemic constitute the local transmission of a de novo human infective variant of a resident infective variant, or novel introduction from a different focus?

(ii) Do all stocks from within one endemic focus of infection constitute a single serodeme or related serodemes?

(iii) Do more than one human infective strains occur within one endemic focus during epidemics?

This study addresses these questions by comparative analysis of 24 stocks from Kenya, Zambia, Uganda, Tanzania and West Africa, utilizing as the basis, the singular properties of the ILTat 1.22 BC telomere.

Results.

5.1

Comparative analysis of trypanosome stocks for the presence of the ILTat 1.22 BC telomere.

Table 5.1 presents a detailed breakdown of trypanosome stocks utilized in this analysis.

Restriction digests of DNA from individual stocks were investigated by hybridization with three individual probes from lambda MT 1.22 B, corresponding to the ILTat 1.22 BC telomere. The probes are illustrated in Figure 5.1.1. and correspond to an ILTat 1.22 VSG gene fragment on a 200 bp Bam HI fragment (Probe 1), the 6.05 Kb Bam HI fragment (Probe 2) and 4.7 Kb Eco RI fragment (Probe 3) of lambda MT 1.22 B.

Hybridization of these probes to the genomic Southern blots was performed at 68°C for 16 hours followed by posthybridization washing to a stringency of 0.1X SSC, 68°C. The result of the hybridization with Probe 1 is shown in Figure 5.1.2. panel A.
The inherent positive controls for these investigations are EATRO 174, the earliest isolate known to express EATRO 795 (ILTat 1.22) and EATRO 2340 (lane 10) the stocks best characterised for the ILTat 1.22 BC telomere.

In both instances the ILTat 1.22 BC is detected on a 2.2 Kb Eco RI fragment as expected (see Chapter 3), with the signal in the EATRO 2340 digest (lane 10) considerably stronger as this stock possesses two BC genes for ILTat 1.22. In EATROs 7 (lane 3) and

18 (lane 10) where ILTat 1.22 is not detected by phenotypic analysis of metacyclic populations (J.D. Barry pers. comm.), the probe failed to detect the presence of ILTat 1.22 in the genome.

All other Kenyan stocks present from the focus possess the gene on the same sized Eco RI fragment as the positive controls, as do two of the three Ugandan stocks. UTAR 3 (lane 12) and UTAR 4 (lane 13). UTAR 5 (lane 11) does not possess the ILTat 1.22 gene. Two lanes, 8 and 9, have extra bands these are partial digest fragments, as determined by visual analysis of the ethidium bromide stained gel of the blot (not shown) which shows incomplete cutting in these lanes.

Neither the 8 Zambian stocks nor those from Tanzania or West Africa hybridized with Probe 1, indicating the absence of this VSG gene. This is in agreement with phenotypic analysis of several of these stocks using immunological methods (J.D. Barry pers. comm).

Hybridization with Probe 2, as shown in Figure 5.1.2, panel B gives equivalent results to Probe 1 for the Kenyan (Figure 5.1.2, Panel A) and Ugandan (Figure 5.1.2, panel D) stocks. The upstream region of the telomere thus is present in the positive and absent in the negative control digests (lanes 1, EATRO 174, and lane 2, EATRO 7, respectively). o7 3

In all the stocks, save the UTARs, showing hybridization with Probe 2 a faint cross-reacting band of 3.3Kb is observed in agreement with the EATRO 795 analysis with the same probe (Chapter 4), though in the latter a 7.8 Kb fragment is also observed in long

exposures. Only three other stocks hybridized with this probe, none of which was Zambian. These constituted two stocks from Tanzania (East Africa) and a West African T. brucei stock (see Figure 5.1.2, Panel D). Interestingly, all three stocks hybridize with Probe 2, despite the extreme difference in the geographical location of the two types of stock. Three bands are apparent in the West African stock STIB 386 (lane 1, section 1,Panel D) sizing at > 23Kb, 15Kb and 12Kb, while a 15Kb band is detected in the Tanzanian stocks (lanes 2 and 3).

Table 5.1.Stocks used in the epidemiological analysis of the ILTat 1.22 BC telomere.

Eatro	Cloned	Year	Place	Host
174	C	1959	Busoga	U Human
3	C	1960	Busoga	U Fly
7	C	1961	Yimbo	K Fly
95	C	1961	Yimbo	K Human
97	C	1961	Yimbo	K Human
110	C	1961	Yimbo	K Human
116	C	1961	Yimbo	K Human
156	C	1961	Yimbo	K Human
UTAR 3	C	1981	Busoga	U Human
UTAR 4	C	1982	Busoga	U Human
UTAR 5	C	1982	Busoga	U Human
ZAMB.194	C	1982	Kasyasya	Z Human
199	C	1982	Kasyasya	Z Human
208	C	1982	Kasyasya	Z Human
210	C	1982	Kasyasya	Z Human
218	C	1982	Chibale	Z Human
222	C	1982	Kasyasya	Z Human
269	C	1983	Kasyasya	Z Dom.An.
274	C	1982	Kasyasya	Z Human
290	C	1982	Kasyasya	Z Human

Tanz. STIB 247-L	C	1971	Serengeti	T Hartebeest
EATRO 1534	C	1969	Kiboko	T* Tsetse
West African.				
STIB 386	C	1978		I Human

KEY : U = UGANDA, Z = ZAMBIA, T = TANZANIA, I = IVORY COAST.

C = CLONED, * = NOT FROM THE FOCUS AT LAKE VICTORIA.

Figure 5.1.1

Probes used in the Epidemiological investigation on the ILTat 1.22 BC telomere.

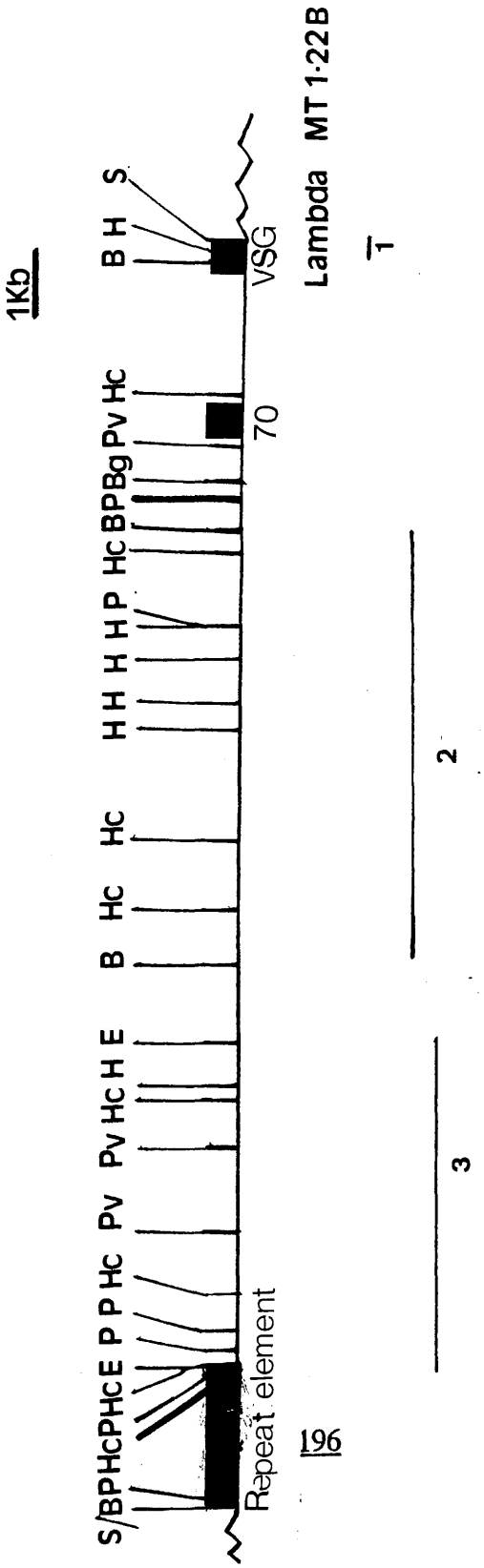


Figure 5.1.2

Southern analysis of trypanosome stocks .

Panel A:

Markers (M = Lambda x H/E) are shown in Kb.

Lane 1. EATRO 174. 2.EATRO 3, 3. EATRO 7, 4. EATRO 18, 5. EATRO 95, 6. EATRO 97, 7. EATRO 106, 8. EATRO 116, 9. EATRO 156, 10. EATRO 2340, 11. UTAR 5, 12. UTAR 3, 13. UTAR 4.

Panel B:

Lambda x H markers (M) are shown in Kb.

Lane 1. EATRO 174. 2. EATRO 7, 3. EATRO 95, 4. EATRO 97, 5. EATRO 106, 6. EATRO 116, Panel C:

Lambda x H markers (M) are shown in Kb.

Panel C:Lane 1. EATRO 174, 2. EATRO 3, 3. EATRO 7, 4. EATRO 18, 5. EATRO 2340, 6. EATRO 95, 7. EATRO 106, 8. EATRO 116, 9. EATRO 156.

Markers (M) are indicated in Kb.

Panel D:

Section 1 :

A: Probe 2 investigation.

B: Probe 3 investigation.

Lane 1. STIB 386.

2. STIB 247.

3. EATRO 1534.

Section 2 :

A: Probe 3 analysis.

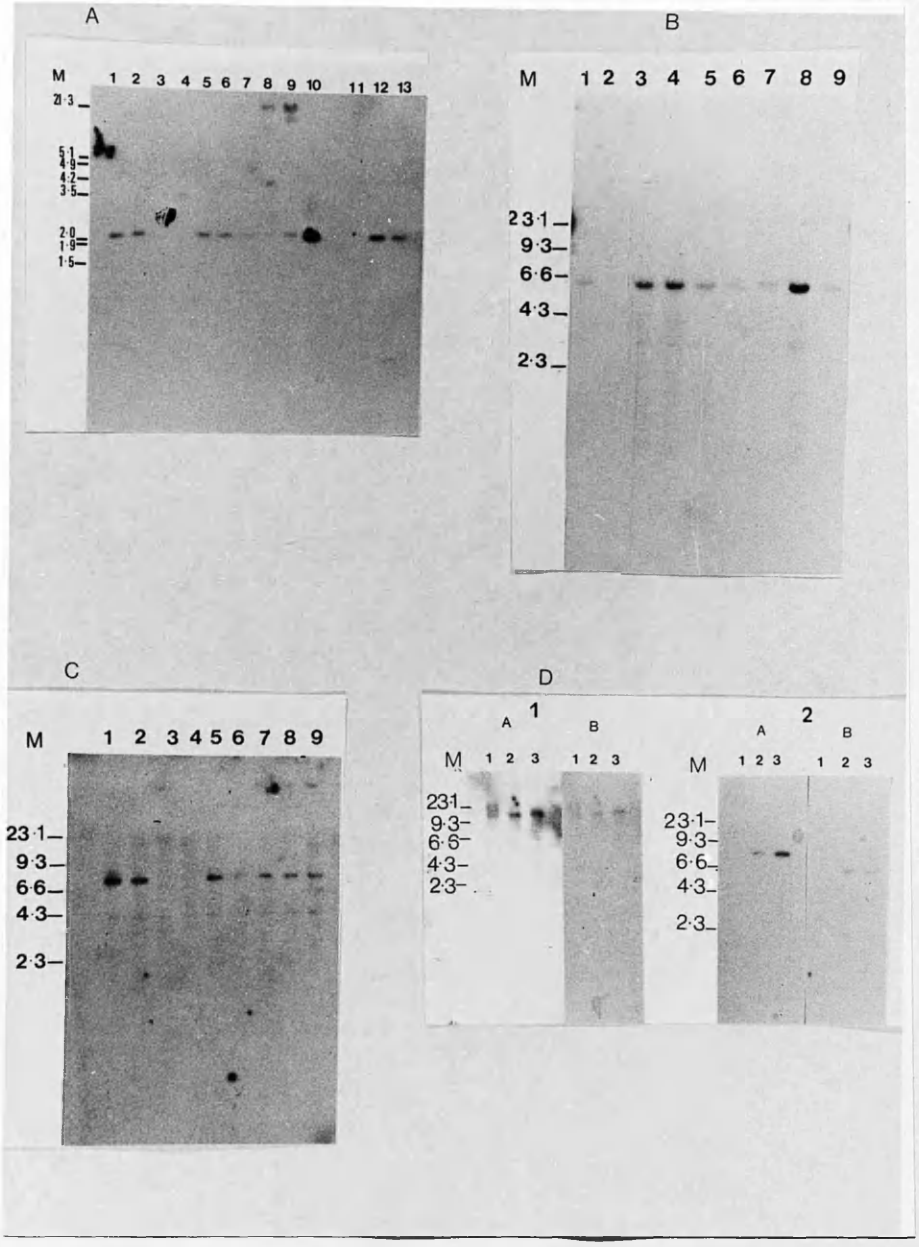
B: Probe 2 analysis.

Lane 1. UTAR 5

Lane 2. UTAR 3.

Lane 3. UTAR 4.

Marker sizes are indicated in Kb.


A
M 1 2 3 4 5 6 7 8 9 10 11 12 13
21·3
5·1
4·9
4·2
3·5
2·0
1·9
1·5
B
M 1 2 3 4 5 6 7 8 9
23·1
9·3
6·6
4·3
2·3
C
M 1 2 3 4 5 6 7 8 9
23·1
9·3
6·6
4·3
2·3
D
1
A
B
M 1 2 3 1 2 3
23·1
9·3
6·6
4·3
2·3
2
A
B
M 1 2 3 1 2 3
23·1
9·3
6·6
4·3
2·3

Probe 3 detected in 7.6Kb Bam HI band in all the Kenyan stocks (Figure 5.1.2, panel C) save the negative controls (lanes 3 and 4). The three stocks from the Tanzania and West African gave identical bands to that observed for Probe 2; these are, however, faint due to the blot having been reprobed.

DISCUSSION.

Comparative analyses between geographically distant endemic foci of T. brucei rhodesiense infection provided some intriguing results pertinent in tracing the origins and spread of this disease across Africa.

The antigenic composition of metacyclic populations has previously been used as serodeme specific markers in attempts to investigate the origins and spread of epidemics over the past thirty years (Barry et al., 1983). The metacyclic repertoire shows some instability and gradually changes over time (Barry et al., 1983), thus posing a problem in tracing relationships between foci over long periods of time. Isoenzyme analysis, however, has been used to divide T. b. rhodesiense stocks into two general groups constituting those of North-east and South- east Africa (Gibson et al., 1980), and indicates that the North- eastern epidemics are not the result of gradual spread of the disease from the South-east, but have arisen de novo from the indigenous T. b. brucei population.

The dynamism of the T. brucei genome, however, reflected in the rapid evolution of the VSGs repertoire (Kooter et al., 1988; Hide, G., Le Ray, D. Catland, P., Barry, J.D. and Tait, A., in prep.) can lead to equivocal results in the interpretations of such comparative data. Thus the use of a stable genomic tag, such as the ILTat 1.22 BC telomere, allows for a more fruitful investigation of relationships between stocks. The stability of this telomere is dramatically emphasized by this study, in that it is detected, unaltered, within the region defined by the lambda MT 1.22 B clone, in 8 Kenyan stocks

isolated over a 24 year period from 1958-1982 (see Table 5.1).

These observations correlate with previous observations of the persistence of this M-VAT in stocks from an East African focus from 1961-1977 (Barry et al., 1983) and help provide a molecular explanation for these on the basis of the structural individuality of this particular telomeric locus (see Chapter 4).

The failure to detect the ILTat 1.22 BC telomere in any of the stocks from the Zambian focus tends to indicate that the Kenyan and Ugandan epidemics may have been the result of the de novo appearance of new T. b. rhodesiense strains, and not by northward spread of a single virulent strain into Kenya. While it is formally possible that these two endemic foci are perpetuated by the same serodeme, that in Zambia having lost the 1.22 telomere, this does not agree with immunological analyses and isoenzyme and RFLP analyses of these stocks both of which favour the former postulate (J. D. Barry unpub. obs.; Hide, G., Le Ray, D., Catland, P., Barry, J.D. and Tait, A., in prep).

No trace of any VATS in common in either metacyclic or bloodstream populations has been observed, though this analysis still suffers from the limitations of the immunological techniques employed relative to a direct molecular investigation. (J.D. Barry, pers. comm.).

The situation with respect to the Ugandan stocks indicates that the telomere is present in those initially isolated in 1959 (EATRO 174) and 1960 (EATRO 3) from human and fly respectively. The presence of the telomere in these stocks and in those from the

Kenyan Focus, isolated from the Yimbo area from 1961 onwards indicates that these may belong to the same serodeme, and that in this instance the Kenyan focus may have originated from the eastward spread of the disease from Uganda, and was not due to a new virulent strain having risen in Kenya. More recent Ugandan isolates, from human hosts in the Busoga district, show that in two stocks the telomere is present unaltered (ie: in UTARs 3 and 4) but in a third stock (UTAR 5) it is not present. The absence of the telomere in this latter stock can be formally attributed to a number of causes.

(i) Loss of the telomere from the genome of this stock, from the same serodene responsible for the current epidemic through telomere conversion associated with antigenic variation.

(ii) Loss of the telomere in progeny for trypanosomes having gone through their sexual cycle.

(iii) UTAR 5 constitutes a separate human infective strain, having arisen de novo with the same focus as the previous serodeme, or having entered this focus from another endemic region. Antibody analysis suggest partial similarity amongst stocks in this epidemic. The original serodeme is widely present, though other stocks have overlapping, but not identical repertoires with it, (W. Isharaza, pers. comm). The latter feature may be an indication of sexual reproduction being prevalent amongst trypanosome populations within this focus, and merits further investigation.

As to which, if any, of these possibilities accounts for the absence of the telomere is open to speculation. M-VAT analysis of the UTAR 5 stock should give a better indication of which is the most likely explanation. Should the stock have all other M- VATS in common with UTAR's 3 and 4 or EATRO 174 or 3, then it would

likely mean that telomere conversion was responsible for the absence of the telomere in UTAR 5. Failing this, however, consideration of the latter two possibilities favours the proposal that UTAR 5 is a de novo human infective strain, as the permanent presence of the telomere in stocks from some epidemic area over the period 1959-1982, suggests that the trypanosome sexual cycle is not a prevalent factor for tryanosomes of this serodeme.

The possibility that UTAR 5 is the progeny of sexual reproduction is not formally discounted, and requires VAT analysis to be fully discounted as well as the examination of further stocks from the same focus. The observation of the unaltered presence of the 1.22 telomere in stocks isolated from human and reservoir hosts strongly suggests that sexual reproduction is not prevalent within these foci and thus transmission of the parasite and maintenance of the epidemics is by mechanical means, likely through the inadvertent transmission by other haematophagous insects. Such a situation has direct ramifications for attempts to control the spread and occurrence of trypanosomiasis through directly blocking its transmission by tsetse flies using immunological methods (J.D. Barry, in preparation), or by tsetse fly trapping and elimination (rev. Jordan, 1985). In support of this hypothesis is the observation that the current epidemic in Kenya, ongoing since 1979, T. brucei isolated from tsetse flies in this area is a very rare occurrence. Only three stocks have managed to be isolated from an exhaustive tsetse fly examination although other trypanosomes such as T. conglense and T. vivax were routinely found. Two of the three T. brucei isolates proved human infective and each was completely different to the trypanosome responsible for the epidemic, sharing only one M-VSG in common with it (J.D. Barry,

pers. comm.) Thus strategies that involve manipulation of the trypanosomes' singular biochemistry consequently seem a more relevant approach to tackling the disease.

While none of the Tanzanian and West African stocks possesses the ILTat 1.22 gene, they hybridize with the two upstream probes 2 and 3 (Figure 5.1.2). In each case the hybridizing fragments size differently from those in the stocks from Kenya and Uganda.

The presence of hybridizing fragments in the STIB 386 stock from West Africa is significant in underlining the fundamental advantage in utilizing this M-VAT telomere as a comparative marker to investigate the relationship between stocks isolated both by geography and by time. The inability to detect probes 2 and 3 in stocks from Zambia, but not a West African stock, is perhaps a reflection on the range of geographical overlap between compatible sexually reproducing populations, or is an indication of the separate origin of the East African Foci from West African T. b. brucei.

T. b. rhodesiense is thought to have originated in West Africa around the Zambezi basin, spreading via Zambia into Tanzania and further north into Kenya, Uganda and Ethiopia (Ormerod, 1961).

Isoenzyme analysis, however, suggests that Tanzanian stocks are related to those in Zambia, more so than Ugandan or Kenyan foci (Gibson et al., 1980). The evidence presented in this work, however, disputes this for a single Tanzanian stock investigated and relates this more closely to West African and North east African stocks. It is formally possible that this is a consequence of sexual

reproduction in the field, but the stability of the telomeric marker suggests otherwise and indicates a separate origin for the North eastern stocks from West African T. brucei. Further molecular analysis is needed to confirm this.

In summary, all these observations indicate that the Kenyan and Ugandan endemic foci resulted from the appearance and spread of a new virulent human-infective strain of T. b. brucei; and not from the northward spread of South east African endemic foci. An exception to his may be relevant to the observations that UTAR 5 from the recent Ugandan epidemic (1976-1982) failed to hybridize with the probes from the ILTat 1.22 M-VSG telomere. The recent epidemic is thought to have spread from the endemic foci around the north of Lake Victoria, which agrees with the finding of the ILTat 1.22 telomere in UTAR's 3 and 4. The absence of this telomere in UTAR 5 as described previously may be due to the appearance of a de novo virulent strain. Another possibility, however, is that its appearance in the Ugandan focus is due to changing socio-ecological conditions in this area (Ford, 1979), resulting in new epidemics coincident with refugee and military movements. The latter proposal has been made to account for the appearance of Zambian zymodemes within the Ugandan focus (Gibson and Gashumba, 1983).

Chapter 6.

General Discussion.

6.1. Antigenic Variation.

Antigenic variation can broadly be defined as the ability to alter the antigenic character of cell surface components, typically at a rate orders of magnitude higher than the mutation rate within the genome of the organism. This strategy is employed by a wide range of microbial pathogens (reviewed, Seifert and So, 1988) all of which show a limited capacity for variation of cell surface components. Six genes comprise the pilin repertoire in a single N. gonorrhoeae bacterium (Haas and Myer, 1986) while a simple B. hermsii bacterium can give rise to 24 distinct serotypes (Stoenner et al., 1982). The T. brucei repertoire, however, comprises $>10^3$ distinct VSG genes (Van der Ploeg et al., 1984a), thus allowing for a much more extensive capacity for circumventing destruction by the hosts' immune system.

This repertoire evolves rapidly through the mediation of gene conversion events inherent in the process of antigenic variation (Pays and Steinert, 1988), and abetted to a lesser degree by point mutation (Roth, C.W and Eisen, H., cited Pays and Steinert, 1988). It is unknown if transformation by DNA from lysed trypanosomes in bloodstream infection is also a determining factor in the process of antigenic variation and evolution of the VSG repertoire, analogous to pilin and P.II variation in N. gonorrhoea (Seifert and So., 1988). It is unlikely, if at all possible, that this could occur other than very rarely, as opposed to its major role in N.gonorrhoea : previous studies of VSG genes would have revealed this.

The pathogenic prokaryotic and trypanosome systems of antigenic variation all exhibit a similar frequency of antigenic switching, approximately 10^{-2} to 10^{-4} switches per generation in the prokaryotic system (Seifert and So., 1988) and 10^{-2} to 10^{-3} switches per generation in the trypanosome system (Turner and Barry, 1989). There are a number of important implications relative to antigenic variation in African trypanosomes arising from this observed switching rate. Firstly, the rate of antigen switching suggests that there are specific molecular mechanisms being employed to 'drive' the system. This is inferred from the fact that the observed frequency of switching is orders of magnitude higher than background recombination events and a mutation rate in these organisms of $\sim 10^{-6}$/locus (cited; Pays and Steinert, 1988). This also concurs with prokaryotic systems where the mutation rate is in the order of 10^{-6} events per cell generation (cited: Seifert and So., 1988).

The second consequence of the antigen switching rate is that it implies that there is an optimum switching rate in the course of bloodstream infection. As such the rate must reflect a balance between the dynamics of parasite growth and host responses, coupled with an economic use of the available antigen repertoire, the latter being enhanced by a hierarchical order of expression of given VSG genes. By such means the parasite;

(i) does not kill its host upon initial infection by over rapid growth.

(ii) Use of a hierarchy of VATs enables the parasite to avoid premature termination of an infection, principally through antibody mediated immune lysis resulting from previous host exposure to specific VATs.

(iii) The ability to prolong infection by antigenic switching at a rate sufficient to circumvent the VSG - specific immune responses and thus enhance the prospects for subsequent transmission by the tsetse fly vector.

(iv) A further consequence of a high rate of switching is that the host is presented with a large number of different VATs at the same time during bloodstream infection (Hajduk and Vickerman, 1981). This seemingly is counterproductive to the classically presumed aim of antigenic variation to prolong infection to enhance the probability of transmission to the tsetse fly. This function may, however, be achieved by offsetting it against a hierarchical use of individual ESs. (Liu et al., 1985) to down regulate the number of VSGs expressed by the population at any one time.

The selective pressure associated with a biochemically driven system of antigenic variation may be directly responsible for the rapid evolution of VAT repertoire (Barry et al., 1983; Bernards et al., 1986) and the associated genomic rearrangements. The question remains, however, as to what extent the observed switching rates between specific VATs is a reflection of one specific mechanism or is a composite rate derived from the number of distinct mechanisms associated with antigenic switching at the molecular level? Furthermore, to what extent are the relative prevalences and sequential expression of given VATs during the course of an infection a consequence of the factors contributing to the switching rate? VAT prevalence and temporal sequence of expression have previously been attributed to a series of factors, such as differential growth rates between trypanosomes expressing different VATs and interactions between spatially segregated populations within the host (for a review

see Seed et al., 1984; Borst, 1986). Such factors, however, may be all but redundant in view of a 'driven' system of antigenic variation. As will be discussed later, prevalence and temporal sequencing of VAT expression may be a direct consequence of the rate of switching and the molecular characteristics of VSG gene expression sites.

As such, it is reasonable to assume that both the rate of gene conversion into a given expression site and inter-expression site switching are responsible for the observed rate of $10^{-2}/10^{-3}$ switches per generation. In the metacyclic form, where VSG expression is probably in situ (S. Graham and J.D. Barry, unpub. obs) the rate is a direct reflection of de novo activation of M-ESs. It is this mode of antigenic variation that dominates bloodstream switching (Bernards et al., 1984a; Myler et al., 1984, Liu et al., 1985) and hence it may be the predominant factor determining the switching rate. It is difficult to decide the contribution of other modes of switching events to the overall rate. It is unlikely that gene conversion events associated with antigenic variation occur at 'background'levels ie. at the mutation rate of $\sim 10^{-6}$ events cell generation^{-1} locus^{-1} and that the overall rate is determined by de novo activation of another ES. This seems a reasonable assumption given the rapid evolution of the VSG gene repertoire, if this is tied with the high rate of antigenic variation.

6.2 The M-VSG Gene Repertoire.

T.b. rhodesiense has the capacity to express more than 10^3 individual VATs, though only a small subset of this repertoire, probably no more than 27 VATs, is expressed in the metacyclic form. (Turner et al., 1988). The questions thus arise as to how and why

only a specific subset is expressed at this stage of the life cycle and how is this M-VAT repertoire maintained?

To answer these questions one needs to assess the characteristics of M-VSG genes and their expression.

All M-VSG genes characterized to date exist on the largest size class of trypanosome chromosome (Cornelissen et al., 1985a, Lenardo et al., 1986; Delauw et al., 1987).

Contradictory results on the expression of these genes have been reported, with the gene either being expressed in situ (ILTat 1.22 & 1.61 - S. Graham & J.D. Barry, unpub. obs.; M-VATs 4 and 7 - Lenardo et al., 1984) or by an ELC mechanism (AnTat 1.3 - Delauw et al., 1987). Previous studies, however, suffer from a number of disadvantages in that they employ trypanosome lines suitable for tsetse fly transmission, but unstable in laboratory rodents. As such, these lines are unsuitable for the study of metacyclic expression, as by the time enough material for molecular analysis can be gathered the trypanosomes have either lost expression of the initially expressed VAT or have switched to a B-VSG gene expression system. This latter feature is not readily apparent in previous studies, as is detailed below, and consequently is the source of the contradictory observations.

As described in the work presented in this thesis, use of a model line of trypanosome with the unusual combination of fly transmissibility and rodent virulence, atypical of previous experimental lines, overcomes the drawbacks inherent in previous systems and

reveals that the M-VSG genes are initially expressed in situ for several days (S. Graham and J.D. Barry, unpub. obs.), after which a switch occurs concomitant with the appearance of ELCs, sometimes of the same, M-VSG gene. This switch in expression mechanism reflects a change to a distinct bloodstream expression system, utilising typical bloodstream expression sites.

Previous observations of M-VSG genes expressed by an ELC mechanism (Delauw et al., 1987) may thus be attributed to the analytical system used in which the observation of an ELC mechanism is a direct consequence of the delay in generating enough trypanosomes to provide a suitable quantity of material for molecular analysis.

Two M-VSG genes have, however, been reported to be expressed in situ (Lenardo et al., 1984, 1986). The polyclonal origin of each VAT in the system used by these workers (see. introduction, Chapter. 4) allows for the major criticism that the presence of multiple ELCs within the population, each with a different fragment size in genomic Southern blots, might not be detected. Significantly though, each BC gene was observed to be devoid of 70bp repeat sequence on their 5' flanks, and this feature was attributed to be responsible for the necessity to express these VSG genes in situ, by virtue of ELC formation being precluded due to lack of 70bp repeats as an essential "molecular substrate" for ELC formation (Lenardo et al., 1986). It seems clear, however, from the observation on M-VSG genes using the model systems described that in situ activations is a specific feature of metacyclic form VSG gene expression, and not a consequence of lacking 70bp repeat sequence. Indeed, the ILTat 1.22

BC M-VSG gene, while having only a single complete 70bp repeat unit on its 5' flank, participates in ELC formation at a high frequency in bloodstream infection. In each instance the breakpoint for ELC formation maps within the 70bp sequence on the BC telomere (K.Matthews and J.D.Barry, unpub. obs.).

It seems likely, based on the observations of the lack of 70bp sequence on some M-VSG gene BC telomeres (Lenardo et al., 1984, 1986), and the brevity of such sequence on others (ILTat 1.22 - this study; ILTat 1.61 - K. Matthews and J.D. Barry unpub. obs.), that this feature (little or no barren region) will be typical of M-VSG expression sites.

The structural individuality of the M-VSG expression site described in this study presents a number of important implications concerning the maintenance of such telomeric sites, the direction and mechanism of gene conversion during antigenic switching, and offers insight into the predictability of M-VAT repertoires (Barry, 1989).

(i) A direct consequence of the lack of homology between the ILTat 1.22 M-VSG gene ES and B-ESs, is its increased stability within the genome. Lack of homology results in a decreased frequency of reciprocal recombination and non-specific gene conversion events with other chromosomes. Such events typically involve sequence elements upstream of the 70bp sequence, ESAGs for example, and result in substantial genomic rearrangements (Kooter et al., 1988). The individuality of the 1.22 M-VSG expression site may allow the conservation of any important elements carried on this telomere. The observed stability of this telomere in different

stocks in the field (Barry et al., 1983; this study) and in repeated fly transmissions in the laboratory (Barry et al., 1983) provides direct support for such a hypothesis.

No internal rearrangement of this expression site has yet been observed, and the only difference noted between stocks has been the duplication of the BC telomere (Cornelissen et al., 1985a). It should be noted, however, that the stability of the ILTat 1.22 BC telomere may represent an extreme case. The key feature of this expression site, with respect to other M-VSG expression sites is that the VSG transcription unit is short, again decreasing the target area for non-specific recombination events. If this is a universal feature of M-VSG expression sites, then the upstream areas of the telomeres harbouring these sites need not necessarily show the same degree of structural individuality as the 1.22 site. Such telomeres would as a consequence be less stable than that of ILTat 1.22 BC telomere in the genome and thus more prone to rearrangement or loss from the M-VAT repertoire, hence allowing for evolution of the repertoire over time.

A possible example of this is the presence of ESAG 1 on two reported M-VAT telomeres (Son et al., 1989) and on one of the ILTat 1.61 BC copies in EATRO 795 (K. Matthews and J.D. Barry, unpub. obs.).

The presence of ESAG 1 on the M-VAT telomere (Son et al., 1989) may explain the odd observation of this telomere being able to produce an ELC in chronic bloodstream infection though lacking 70bp repeat sequence (Lenardo et al., 1984, 1986), an observation

initially unexplained by these workers. The presence of ESAG 1 would readily allow for a telomere conversion to occur and thus allow transport of this M-VAT into a B-VSG gene expression site.

(ii) The structural individuality of the ILTat 1.22 BC telomere results in gene conversion events involving this telomere being mediated through the 70bp region. The observed pattern of ELC production resulting from such events thus involves the telomere with the short region of homology acting as a donor to another with a long region of homology, with respect to 70bp repeat sequence. This agrees with the predictions and implications of the double-strand break repair model of gene conversion (Sun et al., 1989; Fincham and Oliver, 1989).

A number of models can be postulated for the control of such events with respect to the ILTat 1.22 telomere and its expression in metacyclic and bloodstream form trypanosomes. These are now discussed.

6.2.1 Model A Simple Recombinator Model.

This model envisages the B-VSG expression site acting as the recipient locus for gene conversion by preferentially binding an endonuclease required to generate the conversion. The preferential binding of the endonuclease is proposed to be a consequence of the recipient locus possessing a greater number of binding sites for the endonuclease, which in this instance are proposed as being the 70bp repeat units. Thus, the telomere with the least number of 70bp repeats will act as the donor in the conversion event. This model is

illustrated in Figure 6.1, and fits well with the observation of 1.22 ELC formation on the switch from metacyclic to bloodstream expression mode, and complies with the double-strand break model for gene conversion (Szostak et al., 1983; Fincham & Oliver, 1989; Sun et al., 1989).

This model can also be extended to explain the hierarchy of VSG gene expression sequence if the probability of expression is inversely linked with the number of 70bp repeats 5' of the VSG gene. This would be offset by a stochastic activation of individual expression sites, as observed in the salivary glands of the tsetse fly (Tetley, et al., 1986), thus engendering the sequence of VSG gene expression with a degree of unpredictability, as observed in vivo (Capbern et al., 1977; Barry, 1986). Correlation of the extent of 70bp sequence with endonuclease binding in this model also fits the observations on M-VSG telomeres lacking 70bp repeats (Lenardo et al., 1984, 1986). Such VSG genes are observed to be expressed by an ELC mechanism very rarely in chronic infections, presumably at a frequency akin to the spontaneous recombination (mutation) rate of $\sim 10^{-6}$ events per locus, and mediated via sequence homology with other expression sites, upstream of the VSG and amid the ESAG sequences typical of B-VSG expression sites.

Further evidence in support of a switching mechanism directed through the 70bp repeats comes directly from the observation of a 19-mer sequence within the repeat being used as a breakpoint consistently during ELC formation in chronic bloodstream infection (H. Eisen unpub. obs.), and indirectly from the observation that the yeast MAT switch endonuclease octameric recognition sequence is

present within the most highly conserved part of the 70bp repeat sequence (Michiels et al., 1984).

Contrasting data from examination of conversion boundaries in antigenic switches from five relapse trypanosome populations expressing the 221 VSG gene indicates that in these populations the conversion 5' boundary is 5' of the 70bp repeats (Kooter et al., 1988).

As these observations were made using conventional rodent adapted populations where switching occurs at a lower order of magnitude than in wild populations conversion events with boundaries 5' of the 70bp may thus be spontaneous recombinations, and not due to a biochemically driven switching system. Alternately the 70bp repeats are not the endonuclease binding or cut sites. The latter feature can readily be accompanied by the following model.

6.2.2 Model B. Modified Recombinator Model.

This model is based on a restriction modification model for the initiation and control of recombination in Neurospora (Catcheside, 1986). The model proposes that specific recombinator elements are masked by DNA modification, and that unmodified recombinators then act as recognition sites for an endonuclease with properties similar to a type I restriction enzyme which cuts at a distance from its recognition site, thus producing a differential set of break points over a given stretch of sequence, in a series of conversion events.

Such a model fits well with the observed gradient of

unidentified covalent modification on inactive expression site telomeres (Bernards et al., 1984b; Pays et al., 1984) and the lack of such modification on the active expression site telomere. This model is illustrated in Figure 6.2, though in this instance a type I endonuclease is not an essential feature of the model.

The active site telomere is thus proposed as acting as the recipient in the conversion event, while the donor telomere is determined by another factor, possibly the number of 70bp repeats it possesses. As such, this feature of the model is not mutually exclusive with the simple recombinator model, though the absence of 70bp repeat sequence recognizable by hybridization in T. congolese (P. Shiels and J.D. Barry, unpub. obs.) argues against these acting as recombinators, though it remains possible that a small sequence motif present in the VSG ES of each species fulfills this role.

1. An attractive feature of a model involving modification is that covalent DNA modification has been associated with transcriptional inactivity in a number of other eukaryotic systems, especially when present in the 5' vicinity of a promoter (reviewed, Nickol and Felsenfield, 1983; Doerfler, 1983; Keshet et al., 1986), though whether it is the direct cause of transcriptional inactivity is inconclusive.

2. A prediction from this hypothesis is that rearrangements involving inactive telomeric expression sites should only occur at the background genomic recombination frequency, an inherently testable prediction now that a suitable model system has been developed for studying antigenic variation (Turner and Barry, 1989).

3. A key factor in this scenario is that telomeric modification (i) can be used to direct tracking of a specific endonuclease needed for the gene conversion to the active expression site, and (ii) engender transcriptional inactivity to other expression sites.

Such a scenario, however, couples expression site switching and acquisition and loss of modification of the respective new and old active sites. One might envisage this occurring in conjunction with cell division. It is known that telomeres are replicated late in the cell-cycle (Blackburn and Szostak, 1984), an event which in this instance would result in the production of hemi-modified telomeric expression sites.

Hemi-modification may not be sufficient to maintain the transcriptional inactivity of these telomeres, and transcription complexes could then form and move along their length preventing further modification of these sites. The possibility also arises for multiple expression site activation as has been observed in vitro (Baltz et al., 1986).

4. This situation also implies that the activation of individual expression sites may be hierarchical, subject to the distribution of the substrate sequences for modification. Support for this idea comes from ;

(i) the preferential use of given B-ESs in vivo (Liu et al., 1985).
(ii) There is an apparent preferential expression of the previously

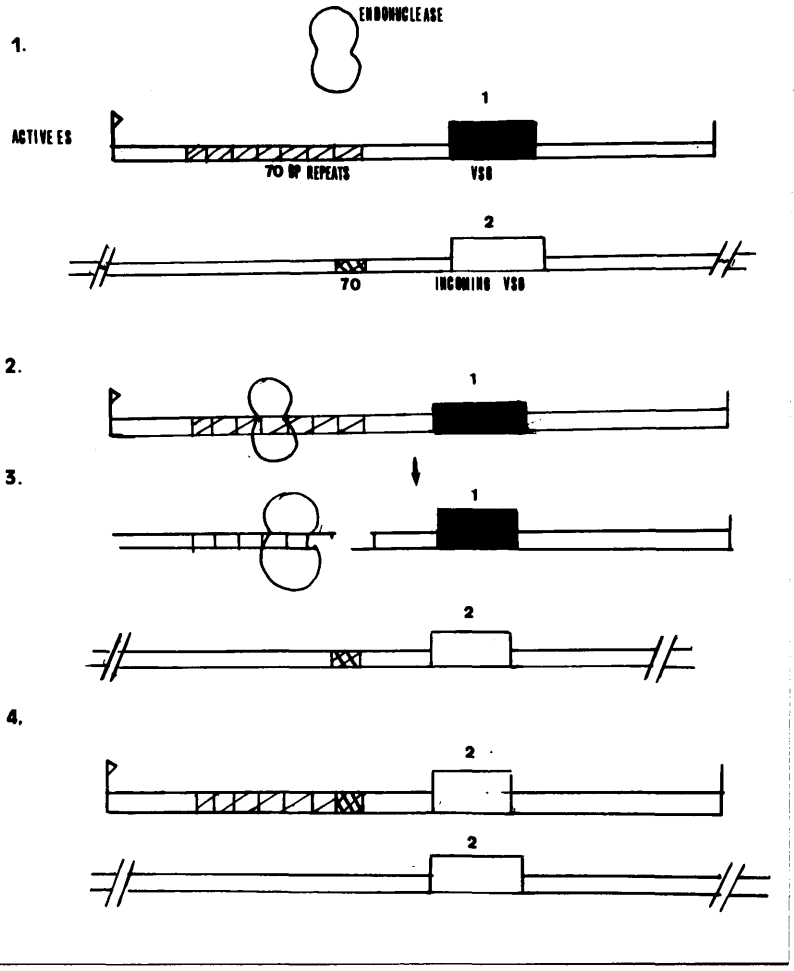
Figure 6.1

Simple recombinator model for antigenic variation.

1. Active ES prior to switch (VSG 1) and incoming VSG with smaller number of 70 bp sequence 5'(VSG 2).

2. Endonuclease binds to 70 bp repeats on the active ES and (3.) makes double strand break. The break is repaired using incoming VSG cotransposed unit as template.

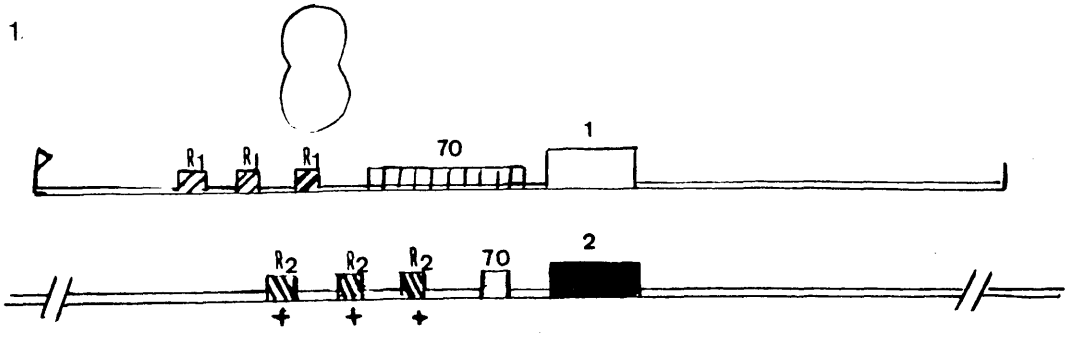
4. Duplicative transposition complete and new ELC (VSG 2) sits in the active ES.


1.
ENDONUCLEASE
ACTIVE ES
1
70 BP REPEATS
VSG
2
70
INCOMING VSG
2.
1
3.
1
2
4.
2
2

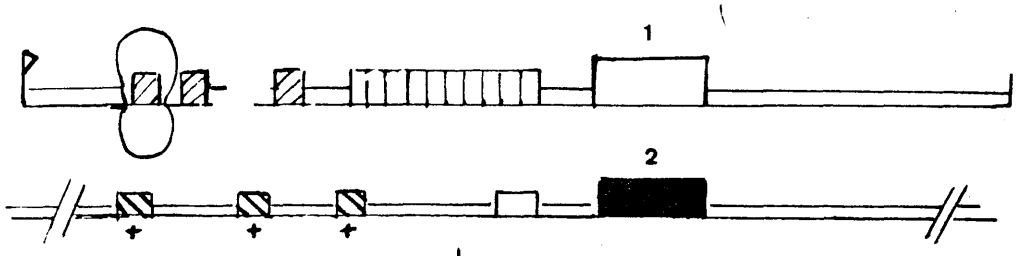
Modified recombinator model for antigenic variation.

1. Endonuclease tracks onto active ES by virtue of its unmodified status.
2. Endonuclease binds to unmodified recombinator and makes double strand break.
3. Gap repair using incoming VSG (VSG 2) as template. Smaller 70 bp repeat number 5' of incoming VSG specifies direction of gene conversion.
4. Conversion complete, VSG 2 ELC sits in active ES.

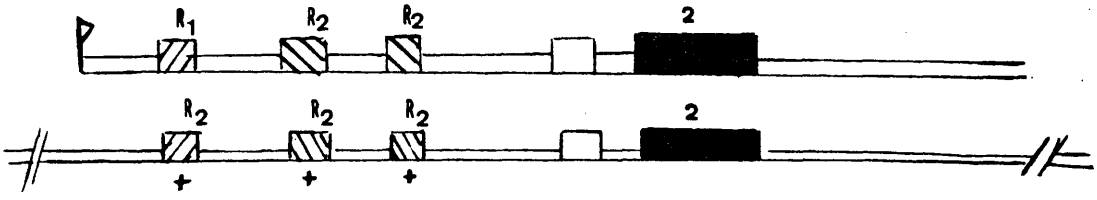
1.



2.



3.



expressed M-VSG gene when the switch to B-VSG expression occurs (S. Graham and J.D. Barry, unpub. obs). This could be due to an increased chance of it being re-expressed by virtue of the chromatin status of its telomere when ELC formation occurs, allowing recombinational machinery easier access than onto other M-VAT telomeres.

(iii) The 'anamnestic' re-expression of the last ingested B- VSG gene would be accounted for by virtue of it being at an ES high in the ranking order of ES activation in bloodstream infection, and would require no specific memory mechanism.

5. A further prediction of this model is that expression site exclusivity is not guaranteed in vitvo, as has been observed (Baltz et al., 1986) and that double expressors may occur at a discrete frequency which might be determined given an accurate knowledge of the number of expression sites and their frequency of activation in vivo. Double expressors, however, would not readily be perpetuated as they are strongly selected against in vivo, and incompatibility between different VSGs on the cell surface may result in their contra-selection in vitro, as has been observed in vitro (Baltz et al., 1986).

6. Another outcome of the features described in this situation is that expression site activation does not require specialized compartmentalization to engender transcriptional activity as proposed by Shea et al (1987), where attachment to a (nucleolar) site is responsible for expression site exclusivity and provides the necessary polymerase for transcription. This model fails to account sufficiently for multiple expressors and anamnestic re-expression.

7. One criticism of this model is that it fails to explain why all expression sites are not active in the procyclic form, given that all these telomeres are unmodified (Pays et al., 1984, Laurent et al., 1984a). Possibly this is because these specific telomeric domains are transcriptionally inactivated at the point of their respective transcription start sites, by disassociation from the nuclear matrix or by lack of availability of a given polymerase subclass. This is supported by the observation of differential alpha-amanitin resistant polymerase activity between bloodstream and procyclic trypomastigotes (Tittawella, 1988).

6.2.3 Model C. Mobile Enhancer/Activator Model.

A third model can by hypothesised to explain the observations on pattern of antigenic variation. This model envisages that stochastic ES activation is mediated by a mobile enhancer / activator element transposing in a conservative fashion between telomeres. ES hierarchy would then be determined by the target site availability for this element, and rare double expressing trypanosomes would occur if the element transposed duplicatively. The element is proposed to sit 5' to the VSG ES promoter and act in cis to alter the long range chromatin conformation of the ES.

This model envisages that all the ES telomeres are associated with the nuclear matrix, as has been observed in other eukaryotes (rev. Lima de Faria, 1983) and all, save the active ES by virtue of its chromatin structure, get modified.

The hierarchy of sequence of VSG expression would then be determined through the extent of homology blocks present on the ES, in an inverse fashion as for models A and B, as would the direction of the conversion event. Loss of VSG gene expression in procyclics would be considered to be under a separate dominant control mechanism.

This model has a major disadvantage in that no putative enhancer element has been identified in the 5' vicinity of the VSG promoter, which undergoes no DNA rearrangement when active or inactive (Zomerdijk et al., 1990; Pays et al., 1989b). The formal possibility remains that such an element exists further upstream and targets on the matrix attachment site for the whole telomeric domain. Long range mapping of a telomere such as that for GUTat 7.1, with two BC genes in EATRO 2340, might be undertaken to confirm such a hypothesis.

This hypothesis is in keeping with the observation that a RIME retrosposon is found within the active and not inactive ESs for the AnTat 1.3A gene in nearly all of a series of variants examined, which was proposed as possibly priming the ES for activation (Murphy et al., 1987a).

In one variant where the AnTat 1.3A ES is inactive a copy of RIME is still present, which suggested that it was not the only factor underlying ES activation. It is still possible that the positioning of the element in the ES is crucial for its activation, or that the presence of a second element (RIME or another mobile element) has influenced

the chromatin structure of the ES to inactivate it. Inactivation of an ES by the insertion of a mobile element has previously been described (Cornelissen et al., 1985b).

In this respect, it will be interesting to define more comprehensively the nature of the non 70bp repetitive element on the ILTat 1.22 BC telomere and cross compare this telomere with those for other M-VSG genes.

6.3 What triggers the switch from M-VSG to B-VSG expression?.

The metacyclic forms constitutes a transient, amitotic developmental stage of the life cycle. In vitro observations indicate that it develops into the bloodstream forms within six hours of entering the host (Brun et al., 1984), yet M-VSG expression persists for a few days. A number of possibilities may account for this;

(i) Repression of B-ESs, diluted out over a critical number of cell divisions.
(ii) Interaction with specific host factors.
(iii) Developmental regulation of the polymerase for VSG transcription or the switching endonuclease, or of its activity.

While the first two postulates have no experimental support, the last hypothesis is attractive with respect to the activity of the endonuclease or polymerase being developmentally programmed.

For the endonuclease this may be achieved by linking it to telomere growth. A critical number of cell divisions may be needed

to achieve a telomere of sufficient length for the endonuclease to recognise it and track in. Length may be critical if the telomere needs to form a specific secondary structure to access the endonuclease.

Support for this hypothesis comes from a number of observations;

(i) Telomere growth is developmentally regulated by other protozoa, notably Tetrahymena and Paramecium (see Farney and Blackburn, 1988).

(ii) Telomeres in trypanosomes grow over successive divisions (Bernards et al., 1983), which would readily allow for fine control of endonuclease access and permit a further means of expression sequence hierarchy for VSG gene, if telomere length is a factor in switching. Noticeably, there is a tendency for active site telomeres to increase in length over the course of switching events (Myler et al., 1988), which may offset the loss of any telomeric repeat sequence that occurs spontaneously (Bernards et al., 1983) and be a consequence of homologous recombination between telomeric repeats, 3' of any conversion involving the VSG gene.

This involvement of telomere length in the timing of the switch to B-ES activation remains speculative, but given the current development of transformation systems (Bellafoto and Cross,1989), will be readily testable.

Developmental regulation of polymerase activity is also attractive, in that in bloodstream forms, VSG transcription appears to be mediated by positive activation (Ehlers et al., 1987). Developmental regulation of this activator through for o7 3

example, a position effect could thus allow critical timing of a switch to B-ES activation. Use of a different activator for M-ES transcription similarly controlled, would then allow for separate and specific M-ES transcription. Such a system could be considered analogous to the use of multiple sigma factors in Bacillus spp to regulate developmental gene expression (Losick and Pero, 1982).

The observation that VSG ES promoters are still transcribed in procyclic forms , but transcription fails to proceed more than a few Kb into the ES (Zomerdijk et al., 1990) indicates that efficient polymerase activity may be dependent on a specific activator, in keeping with the above hypothesis. Support for this idea comes from the observation that RNA polymerase I requires a distinct activated subform, RNA polymerase 1-C, to accurately initiate transcription of mouse rDNA (Tower and Solner-Webb, 1987). Similarly, an activated subform of RNA polymerase II, termed IIb, ups the transcription of this polymerase in HeLa cells more than ten fold (Bartholomew et al., 1986).

A number of questions remain pertaining to antigenic variation, that relate directly to these models. These are listed below and discussed separately:

(i) Why the need for a separate M-VSG repertoire and expression system?

(ii) Why the large size and polycistronic character of B-ESs relative to M-ESs?

(iii) Why are ESs telomeric?

(iv) How is the VSG repertoire and the separate ESs maintained and transmitted through meiosis and mitosis in a genome which may

contain many aneuploid chromosomes?

6.4 Metacyclic expression sites; why maintain a separate repertoire?.

The M-VAT repertoire is characterized by a mixture of VATs of predictable composition, which comprise only a small subset of the total VAT repertoire of the parasite. The analysis of the 1.22 telomere suggests that this predictability is due to the simple structure of M-VSG transcription units relative to their bloodstream counterparts, which precludes their rearrangement during antigenic variation. Such telomeres will thus always act as donors during antigenic switching gene conversion events (see previous section) through lack of, or decreased, homology with other expression sites. Evolution of the M-VAT repertoire would thus be slower than that of the B-VAT repertoire, with loss of genes or rearrangements within the former occurring at a rate compatible with spontaneous recombination events in the genome. The predictability of the M-VAT repertoire is thus a direct consequence of the physical character of their respective transcription units and the mode of expression employed in the metacyclic forms. The distinctive nature of the 1.22 telomere may also reflect its location on a >2Mb chromosome, a feature common to all M-VSG genes characterized to date (Cornelissen et al., 1985a; Lenardo et al., 1984, 1986; Delauw et al., 1987).

These chromosomes contain housekeeping genes for T. brucei and thus need to be maintained in a stable fashion. Reducing homology with B-ESs aids such a situation by preventing any re-arrangements which might endanger the functioning of housekeeping genes within such a dynamic genome.

Maintenance of the separate, predictable, M-VAT repertoire seems at odds with the general strategy of antigenic variation, but this may represent a compromise between the need to use economically the VSG repertoire in the field, faced with a strong selective pressure for novelty, especially in partially immune reservoir hosts which have been previously exposed to many of the VSGs in the trypanosome's repertoire. Simultaneous expression of a limited number of VSG genes after tsetse fly transmission, however, both achieves the aim of economy in the use of the repertoire and in providing suitable diversity to initiate infection long enough for the bloodstream expression system to come into operation and prolong the infection, enabling further transmission of the parasite.

As there is no requirement for antigenic switching in metacyclic forms a simple expression system is all that is needed to operate until a 'driven' system is required as the parasite multiplies to sufficient numbers within the host and destruction by the hosts immune system is imminent.

6.5 Bloodstream Expression Sites (B-ES).

1. Why are they bigger and more complex than the 1.22 M-VSG gene expression sites?

Bloodstream expression sites appear to be between 40-60Kb long (Kooter, et al., 1987; Johnson et al., 1987; Alexander et al 1988; Gibbs & Cross 1988; Pays et al., 1989a) and contain a series of seven families of non-VSG isogenes, the ESAGs (Cully et al., 1985). The

ESAG composition and order within each B-ES appears to be predictable, though their spatial distribution and number within a given site varies.

The number of different ESAGs in the genome has been estimated to be between 14-25 isogenes, (Cully et al., 1985) present at approximately 20 copies per genome, a figure used to derive a similar estimate for the number of potential bloodstream expression sites (Kooter et al., 1987), though this estimate does not take into consideration the duplication and triplication of ESAGs with a given expression site. Initial speculation on the complexity of B-ESs has centred on the co-ordinate transcription of ESAGs with the VSG gene on a given expression site.

The function of ESAG protein products remains unknown, though sequence analysis of several of these genes suggests that they are membrane associated. The ESAG 4 predicted polypeptide shows some sequence similarity to the active region of yeast adenylate cyclase, a known transmembrane protein (Pays et al., 1989a). Significantly, this homology is in the C-terminus of both proteins; (Kataoka et al., 1985).

The homology block in ESAG 4 is situated adjacent to a putative transmembrane domain on its C terminal side, suggesting a location for the former on the inner side of the plasma membrane.

The product of one other ESAG has been identified as a glycoprotein, but its identity and function remain undefined (Cully et al., 1986).

While ESAG expression is co-ordinated with that of the VSG gene while situated at telomeric loci, this is not exclusive. Transcription of ESAGs has been detected in procyclic trypomastigotes (Alexandre et al., 1988; Zomerdijk et al; 1990; S. Graham and J.D. Barry, unpub. obs.) and subject to alpha- amanitin sensitivity during bloodstream infection (Pays et al., 1989a). These observations indicate that ESAG functions are not exclusively limited to a role in VSG gene expression and functioning as has been previously suggested (Cully et al., 1986; Kooter et al., 1987; Pays et al., 1989a). This is supported by the absence of any ESAGs on the 1.22 metacyclic expression site, though this does not preclude a trans-acting role for ESAG products, as these are present in the metacyclic form. A further consideration of these observations, however, suggests that ESAG functioning might be related to VSG expression solely in bloodstream forms. ESAG 4 may encode trypanosome adenylate cyclase, as it is thought that VSG release stimulate adenylate cyclase activity (Voorheis and Martin, 1980). It might then not be unreasonable to expect functional relationship between these genes and their co-ordinate expression in cis in bloodstream forms.

The role of VSG turnover on the cell surface, may therefore correlate with the time differentiation into bloodstream forms begins; after approximately six hours (Brun et al., 1984). Any requisite ESAG activity might conceivably be supplied in trans at this point. The VSG turnover rate is estimated at 3 hours (Ehlers et al., 1987), which fits within the observed transformation times.

The expression of ESAGs in procyclic forms may reflect a need for differential enzyme activity or functioning at stage specific sections of the life cycle, and thus require expression from a non-VSG transcription unit; this may be reflected in the isogenic character of ESAGs. The complexity of the B-VSG transcription unit may therefore mirror the need for co-ordinate expression of a given battery of genes stage specifically, to comply with the needs of a highly active switching system.

The large size and polycistronic nature of the B-VSG expression site is not atypical in T. brucei, nor in other trypanosome species. The actin (Ben Amar et al., 1988) tubulin (Imboden et al., 1987) and calmodulin (Tschudi and Ulhm 1988) genes of T. brucei and a repeated gene in T. cruzi (Gonzalez et al., 1985) are all transcribed in large polycistronic transcription units. This may be a direct consequence of the process of discontinuous transcription and if there is a need to co-ordinate the expression of functionally inter-related, or stage-specifically required genes. In a genome employing trans- splicing to produce mature mRNAs the requirement for individual promoters for each gene is negated by providing a common 5' end for all mRNAs, thus enhancing control of co-ordinately expressed genes by placing them in polycistronic transcription units initiated from a single point. This also may allow the trypanosome to respond rapidly to the extreme environmental differences and accompanying gross metabolic changes encountered in the course of its developmental cycle.

A small size and relative simplicity for M-VSG expression sites, as illustrated by that for 1.22, thus appears a consequence of the

metacyclic form being an amitotic developmental stage, where co-ordinate expression of antigenic switching machinery is not required as M-VSG interswitching relies on de novo ES activation, and a need for reduced homology with other expression sites is necessary, as detailed previously.

6.6 Why are VSG expression sites Telomeric?.

Positioning at a telomere is essential, but not sufficient, for VSG gene expression, as many BC genes are themselves telomeric. Expression is thus dependent on the active site promoter being telomeric. Conventional thought suggests that this is a consequence of VSG transcription being performed by a RNA polymerase I type enzyme, as judged by its insensitivity to alpha-amanitin (Kooter & Borst 1984, Kooter et al., 1987) .

A more likely explanation for a telomeric location of these expression sites may relate directly to the mechanics of their transcription. When active these sites appear to be under torsional stress, indicating that the expression site is anchored to the nuclear matrix (Greaves and Borst, 1987). This is not unreasonable given that transcription control sequences are associated with the nuclear matrix in higher eukaryotes (Gasser and Laemmli, 1987).

As discussed previously, however, this interaction could either be precluded by DNA modification, or occur on all telomeres, with inactivation occurring as a consequence of modification, preventing access to the transcriptional apparatus by masking recognition sequences or providing a substrate for forming an inaccessible tertiary

structure.

The latter is an attractive proposition as it allows telomeres to be attached to the nuclear matrix, thereby providing a means of controlled partitioning at cell division, as is discussed later. Furthermore, loss of telomere modification in procyclic cells would be offset by shut down of VSG synthesis.

Having the ES at a telomere may thus be a reflection of the need for an interaction with the nuclear matrix for transcription but tempered with the mechanics of the need for a high transcription rate. The mechanics of this situation can be regarded as follows: DNA unwinding in front of a moving transcription complex and rewinding behind it results in the formation of positive supercoils in front of the complex and negative supercoils behind it, as observed in vivo using yeast DNA topoisomerase I and II mutants (Tsao et al., 1989). The positioning of the expression site at the telomere thus readily allows for release to positive supercoils, by providing an end that rotates freely in space equivalent to an infinite topoisomerase II concentration (assuming that the telomere end is not attached to any matrix component).

The sensitivity of the active expression site to single stranded endonucleases (Greaves and Borst, 1987) would then be due to an accumulation of negative supercoils in the wake of the transcription complex, and maintained by swivelling of the telomere around the transcription complex while fixed to the matrix at its 5' end. Stress would thus be generated by the transcription complex, and possibly transpliceosomes, by the sheer bulk acting as a 3' anchor site, with

stress being maintained by polymerase density on the telomere. This would account for the observation that single strand endonuclease sensitivity of the active expression site does not show site preference (Greaves and Borst, 1987). This hypothesis cannot account for the observations that transcriptionally inactive and active trypanosome chromatin yields the same nucleosome repeat pattern (Greaves and Borst, 1987) and why the sensitivity to single strand endonucleases is not shared by other transcription units transcribed at high rates, such as the mini-exon or rRNA genes (P. Borst, pers. comm). These features remain baffling and indicate that the chromatin structure of these transcription units may even be atypical in the trypanosome genome.

A relevant observation in this respect is that the level of Z-DNA in mammalian cell nuclei appears to be regulated by torsional stress . Such transcription units, like that for the VSG gene, do not show endonuclease hypersensitivity (Wittig et al., 1989) and indicate that the VSG transcription unit may possess a similar DNA conformation when active.

One other reason for having VSG genes at a telomere is that it would allow the putative endonuclease involved in antigenic switching to track from the end of the chromosome onto an active expression site, especially if switching is associated with increasing telomere length during bloodstream infection. It may be significant that recipient sites in antigenic variation o7 3 generally increase their telomere length (Myler et al., 1988), as might be expected over a sequence of switches, though telomere shortening must eventually occur, which would be envisaged as occurring separately from

switching events on the active telomere.

The increased length of recipient telomeres in such switches has been proposed as a mechanism for increasing the rate of reciprocal recombination at the active site, through the mediacy of telomeric repeats (Myler et al., 1988; Pays and Steinert, 1989). This, however, is contrary to the double strand break model for gene conversion (Szostak et al., 1983) which requires the recipient locus to possess the greater degree (in terms of length) of sequence homology. Thus increasing the length of the telomere at an active expression site, while formally enhancing the chance of reciprocal recombination, is more likely to serve a tracking mechanism considering the driven nature of the switching system.

6.7 What is the ploidy of VSG genes and how is transmission of their repertoire achieved?.

The chromosomal organization of the trypanosome genome present a number of conceptual difficulties in attempting to account for the transmission of the VSG repertoire to trypanosome progeny, given that VSG genes occur on all chromosome classes, some of which appear to segregate in a non-Mendelian fashion during meiosis (Sternberg et al., 1988), a feature consistent with the observations of their being only 10 kinetochores in the trypanosome nucleus (M. Steinert & G. Steinert, pers. comm.), far too few to facilitate segregation of the total chromosome complement. The observation that some VSG genes are haploid (Bernards et al., 1984; reviewed, Donelson, 1988) in the trypanosome genome also poses a dilemma

in terms of transmission of the complete repertoire through meiosis.

These difficulties can be resolved somewhat by an examination of the distribution of trypanosome genes across the chromosome complement.

All 'house-keeping' genes and the M-VSG genes examined to date are found on the two largest classes of chromosomes, of 2Mb and >2Mb (Lenardo et al., 1984; Cornelissen et al., 1985a; Delauw et al., 1987). The majority of B-VSG genes are on smaller chromosomes, though their expression sites are on the larger chromosomes (P. Borst, pers. comm.). The mendelian character of the segregation pattern of the larger chromosomes suggests a diploid state and reflects the essential character of the genes they carry. An estimate of 10 large chromosomes in the genome is consistent with the estimation of 10 kinetochores, implying that there may be 20 available telomeres for use as expression sites. This figure may correspond to the >2Mb chromosomes on which M-VSG genes might be situated exclusively. This is currently being tested (J.D.Barry, pers comm). Should this prove to be so, then the B-ESs must reside on only a few of the >2MB and on 2Mb or smaller chromosomes depending on the size of the M-VAT repertoire, such speculation awaits confirmation of the size of the T. brucei M-VAT repertoire, and better resolution of its chromosomes.

Situating single copy (haploid) M-VSG genes on diploid chromosomes requires regarding such genes occupying the telomeres of homologous chromosomes as 'Pseudo'-alleles, thus permitting ready mitotic and meiotic segregation of the M-VAT repertoire. The

location of B-VSG genes on chromosomes showing non-mendelian segregation during meiosis may allow for rapid evolution of the VAT repertoire. This may be circumvented during mitosis and a more stable segregation achieved if segregation is mediated through specific contact with the nuclear scaffold. This is not an unreasonable suggestion as centromeres have been observed to bind the nuclear scaffold elements (Amati and Gasser, 1988) identifiable as kinetochore proteins (Earnshaw et al., 1984) in yeast, as have telomeres (Amati and Gasser 1988).

The VSG gene repertoire can thus be classed as a repetitive DNA, the repeats at the 3' and 5' flanks of the VSG coding sequence, occupying the same regions of homologous chromosomes, and removing the need for allelism of VSG genes as well as providing a substrate for the reorganisation and evolution of the repertoire and ruling out the need for allelic exclusion system. Confinement of the ESs to larger chromosomes results in maintenance of their presence in the genome and may allow for position effects concurrent with developmental programming to influence the timing of their activation. In this respect it will be interesting to determine the upstream character of the 1.22 M-VSG and B-VSG expression site telomeres. Furthermore, the distribution of B-ESs over the 2Mb and >2Mb chromosomes may well influence their hierarchy of activation relative to the timing of replication of the individual chromosomes.

ADDENDUM

Further sequence characterisation of the repetitive element on the ILTat 1.22 BC telomere reveals this element to be a copy of the

retrosposon INGI (Kimmel et al., 1987),also termed TRS (Murphy et al., 1987), (O. Shonekan and J.D.Barry, pers. comm.). This is the first example of a telomeric copy of such an element in trypanosomes. The presence of this element will allow for ready testing of a model invoking the presence of a mobile element correlated with transcriptional activity of a given ES.

This is particularly interesting with respect to the status of the GUTat 7.1 BC in EATRO 2340, one of the two copies of which carries this retrosposon.

Thalassa, thalassa.....

Xenophon.

REFERENCES

1. Agabian N., Perry K.L. and Murphy W.S. (1987) RNA joining and Trypanosome gene expression. IN, MOLECULAR BIOLOGY OF RNA, NEW PERSPECTIVES., N. Agabian ed. Acad. Pr.Inc., 151-171.

2. Alexandre S., Guyaux M., Murphy N.B., Coquelet H., Pays A., Steinert M. and Pays E. (1988) Putative genes of a variant- specific antigen gene transcription unit in Trypanosoma brucei. MOL. CELL. BIOL., 8, 2367 - 2378.

3. Aline R.F. Jnr , MacDonald G., Brown E., Allison J., Myler P.J., Rothwell V. and Stuart K. (1985a) (TAA)n within sequences flanking several intrachromosomal variant surface glycoprotein genes in Trypanosoma brucei. NUCLEIC ACIDS RES., 13, 3161 - 3177.

4. Aline R.F. Jnr , Scholler J.K. and Stuart K. (1989) Transcripts fom the co-transposed segment of variant surface glycoprotein genes are in Trypanosoma brucei polyribosomes. MOL. BIOCHEM. PARASITOL., 32, 169-178.

5. Aline R.F. Jnr , Scholler K., Nelson R.G., Agabian N. and Stuart K. (1985b) Preferential activation of telomeric variant surface glycoprotein genes in Trypanosoma brucei. MOL. BIOCHEM. PARASITOL., 17, 310-320.

6. Amati B.B. and Gasser S.M. (1988) Chromosomal ARS and CEN elements bind specifically to yeast nuclear scaffold. CELL, 54, 967-978.

7. Askonas B.A. and Bancroft G.J. (1984) Interaction of African trypanosomes with the immune system. PHIL. TRANS. ROY. SOC. LOND., 307, 41-50.

8. Babiker E.A. (1981) Studies on Trypanosoma brucei gambiense : parasitology, antigenic variation and serology. PH.D. THESIS., Univ. Louvain, Belgium.

9. Baltz T., Giroud C., Baltz D., Roth C., Raibaud A. and Eisen H. (1986) Stable expression of two variable surface glycoproteins by cloned Trypanosoma equiperdum. NATURE, 319, 602-604.

10. Banerjee A.K. (1980) 5'- terminal cap structures in eukaryotic messenger ribonucleic acids. MICROB. REV., 44, 175- 205.

11. Barbour S.D., Nagasishi H., Templin A. and Clark A.J. (1970) Biochemical and genetic studies of recombination proficiency in Escherischia coli II. Rec^{+} revertants caused by indirect supression of rec^{-} mutations. P.N.A.S.(USA), 67, 128-135.

12. Barry J.D. (1979) Capping of Variable antigens of Trypanosoma brucei and its immunological and biological significance. J. CELL SCI., 37, 287-302.

13. Barry J.D. (1986) The molecular biology of African trypanosomes. TROP. DIS. BULL., 83, R1-R25.

14. Barry J.D., Crowe J.S. and Vickerman K. (1985) Neutralization of individual variable antigen types in metacyclic populations of Trypanosoma brucei does not prevent their subsequent expression in mice. PARASITOL., 90, 79-88.

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

16. Barry J.D., Hajduk S.L., Vickerman K. and Le Ray D. (1979) Detection of multiple variable antigen types in Trypanosoma brucei. TRANS. ROY. SOC. TROP. MED. HYG., 73, 205-208.

17. Bartholomew B., Dahmus M.E. and Meares C.F. (1986) RNA contacts subunits IIo and IIc in HeLa RNA polymerase II transcription complexes. J. BIOL. CHEM., 261, 14226-14231.

18. Bartolomei M.S. and Corden J.L. (1987) Localization of an alpha-amanitin resistance mutation in the gene encoding the largest subunit of mouse RNA polymerase II. MOL. CELL. BIOL., 7, 586-594.

19. Bartolomei M.S., Haldane N.F., Cullen C.R. and Corden J.R. (1988) Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of RNA polymrase II . MOL. CELL. BIOL., 8, 330-339.

20. Bassett C.L. and Kushner S.K. (1984) Exonucleases I, III and V are required for stability of Col E1 related plasmids in Escherischia coli. J. BACT., 157, 661-664.

21. Bateman E. and Paule M.R. (1986) Regulation of eukaryotic ribosomal RNA transcription by RNA polymerase modification. CELL, 47, 445-450.

22. Bellafoto V. and Cross G.A.M. (1990) Expression of a bacterial gene in a trypanosomatid protozoan. SCIENCE, 224, 1167-1169.

23. Ben Amar M.F., Pays A., Tebabi P., Dero B., Seebeck T., Steinert M. and Pays E. (1988) Structure and transcription of the actin genes of Trypanosoma brucei. MOL. CELL. BIOL., 8, 2166- 2176..

24. Benton W.D. and Davies R.W. (1977) Screening of lambda gt recombinant clones by hybridization to single plaques in situ. SCIENCE, 196, 180-183.

25. Bernards A., De Lange T., Michels P.A.M., Liu A.Y.C., Huisman M.J. and Borst P. (1984a) Two modes of activation of a single surface antigen gene of Trypanosoma brucei. CELL, 36, 163 - 170.

26. Bernards A., Kooter J.M. and Borst P. (1985) Structure and transcription of a telomeric surface antigen gene of Trypanosoma brucei. MOL. CELL. BIOL., 5, 545-553.

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

28. Bernards A., Van der Ploeg L.H.T., Frasch A.C.C. and Borst P. (1981) Activation of trypanosome surface glycoprotein genes involves a duplication - transposition leading to an altered 3' end. CELL, 27, 497-505.

29. Bernards A., Van der Ploeg L.H.T., Gibson W.C., Leegwater P., Eygenraam F., De Lange T., Weijers P., Calafat J. and Borst P. (1986a) Rapid change in the repertoire of variable surface glycoprotein genes in trypanosomes by gene duplication and deletion. J. MOL. BIOL., 190, 1-10.

30. Bernards A., Van Harten Loosbroek N. and Borst P. (1984b) Modification of telomeric DNA in Trypanosoma brucei, a role in antigenic variation. NUC. ACIDS RES., 12, 4153-4170.

31. Bethards L.A., Skadsen R.W. and Skandalios J.G. (1986b) Isolation aand characterisation of a cDNA clone for the Cat 2 gene in maize and its homology with other catalases. P.N.A.S.(USA), 84, 6830-6834.

32. Birnboim H.C. and Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. NUC. ACIDS RES., 7, 1513-1523.

33. Bishai W.R., Rappuoli R. and Murphy J.R. (1987) High level expression of a proteolytically sensitive diptheria toxin fragment in E. coli. J. BACT., 169, 5140-5151.

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

35. Blackburn E.H. and Szostak J.W. (1984) The molecular structure of centromeres and telomeres. ANN. REV. BIOCHEM., 53, 163-194.

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

37. 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 terminals. NUC. ACIDS RES., 9, 4735-4743.

38. Borst P. (1986) Discontinuous transcription and antigenic variation in trypanosomes. ANN. REV. BIOCHEM., 55, 701-732.

39. Borst P., Fase Fowler F., Frasch A.C.C., Hoeijmakers J.H.J. and Weijers P.J. (1980) Characterisation of DNA from Trypanosoma brucei and related trypanosomes by restriction endonuclease digestion. MOL. BIOCHEM. PARASITOL., 1, 221-246.

40. 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.

41. Bouloy M., Plotch S.J. and Krug R.M. (1980) Both the 7-methyl and 2-O methyl groups of the cap of mRNA strongly influence the ability to act as a primer of influenza virus RNA transcription. P.N.A.S.(U.S.A.), 77, 3952-3956.

42. Bridgen P.J., Cross G.A.M. and Bridgen J. (1976) N-terminal amno acid sequence of a variant specific surface antigen of Trypanosoma brucei. NATURE, 263, 613-614.

43. Brown R.C., Evans D.A. and Vickerman K. (1973) Changes in oxidative metabolism and ultrastructure accompanying differentiation of the mitochondrion in Trypanosoma brucei. INT. J. PARASITOL., 3, 691-704.

44. Brun R., Hecker H., Jenni L. and Moloo S.K. (1984) A quantitative ultrastructural study on the transformation of Trypanosoma brucei brucei metacyclic to bloodstream forms in vitro. ACTA TROPICA, 41, 117-129.

45. Brun R., Jenni L., Tanner M., Schonenberger M. and Schell K.F. (1979) Cultivation of vertebrate infective forms derived from metacyclic forms of pleomorphic Trypanosoma brucei stocks. ACTA TROPICA, 36, 387-390.

46. Cadena D.L. and Dahmus M.E. (1987) Messenger of RNA synthesis in mammalian cells is catalysed by the phosphorylated form of RNA polymerase II. J. BIOL. CHEM., 262, 12468-12474.

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

48. Capbern A., Giroud C., Baltz T. and Mattern P. (1977) Trypanosoma equiperdum : Etude des variations antigenique au cours de la trypanosomose experimentale du lapin. EXP. PARASITOL., 42, 6-13.

49. Cardoso de Almeida M.L., Allan M.L. and Turner M.J. (1984) Purification and properties of the membrane form of VSGs from Trypanosoma brucei. J. PROTOZOOL., 31, 53-60.

50. Carle G.F. and Olsen M.V. (1984) Separation of chromosomal DNA from yeast by orthogonal field alternation gel electrophoresis. NUC. ACIDS RES., 12, 5647-5664.

51. Catcheside D.E.A. (1986) A restriction and modification model for the initiation and control of recombination in Neurosporra. GENET. RES., 47, 157-165.

52. Clarke L. and Carbon J. (1976) A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire E. coli genome. CELL, 9, 91.

53. Cochet-Meilhac M. and Chambon P. (1974) Animal DNA - dependent RNA polymerases. Mechanism for the inhibition of RNA polymerase B by amatoxins. BIOCHIM. BIOPHYS. ACTA, 353, 160-184.

54. Cornelissen A.W.C.A., Bakkeren G., Barry J.D., Michels P.A.M. and Borst P. (1985a) Characteristics of trypanosome variant antigen genes active in the tsetse fly. NUC. ACIDS RES., 13, 4661-4676.

55. Cornelissen A.W.C.A., Johnson P., Kooter J.M., Van der Ploeg L.H.T. and Borst P. (1985b) Two simultaneously active VSG gene transcription units in a single Trypanosoma brucei variant. CELL, 41, 825-832.

56. 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.

57. 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.

58. Cully D.F., Gibbs C.P. and Cross G.A.M. (1986) Identification of proteins encoded by variable surface glycoprotein expression site associated genes in Tyrpanosoma brucei. MOL. BIOCHEM. PARASITOL., 21, 189-197.

59. Cully D.F., Ip H.S. and Cross G.A.M. (1985) Co-ordinate transcription of variable surface glycoprotein genes and an expression site associated gene family in Trypanosoma brucei. CELL, 42, 173-182.

60. De Lange T. and Borst P. (1982) Genomic environment of the expression linked extra copies of genes for the surface antigens of Trypanosoma brucei resembles the end of a chromosome. NATURE, 292, 451-453.

61. De Lange T., Berkvens T.M., Verrman H.J.G., Frasch A.C., Barry J.D. and Borst P. (1984a) Comparison of genes coding for the common 5' terminal sequence of mRNAs in three Trypanosome species. NUC. ACIDS RES., 12, 4431-4443.

62. De Lange T., Kooter J.M., Luirink J. and Borst P. (1986) Transcription of the transposed trypanosome surface antigen gene starts upstream of the transposed segment. EMBO J., 4, 3299-306.

63. De Lange T., Kooter J.M., Michels P.A.M. and Borst P. (1983) Telomere conversion in trypanosomes. NUCLEIC ACIDS RES., 11, 8149-8165.

64. De Lange T., Liu 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 the variable surface glycoprotein genes : a multiple promoter for VSG gene transcription ? CELL, 34, 891-900.

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

66. Delauw M.F., Laurent M., Paindavoine P., Aerts D., Pays E., Le Ray D. and Steinert M. (1987) Characterisation of genes coding for two major metacyclic surface antigens in Trypanosoma brucei. MOL. BIOCHEM. PARASITOL., 23, 9-17.

67. Delauw M.F., Pays E., Steinert M., Aerts D., Van Meirvenne N. and Le Ray D. (1985) Inactivation and reactivation of a variant-specific antigen gene in cyclically transmitted Trypanosoma brucei. EMBO J., 4, 989-993.

68. Dero B., Zampetti-Bosseler F., Pays E. and Steinert M. (1988) The genome and antigen repertoire of Trypanosoma brucei gambiense are smaller than those of T.b.brucei. MOL. BIOCHEM. PARASITOL., 26, 247-256.

69. Doerfler W. (1983) DNA methylation and gene activity. ANNU. REV. BIOCHEM., 52, 93-124.

70. Donelson J.E. (1988) Unsolved mysteries of trypanosome antigenic variaton. In, THE BIOLOGY OF PARASITISM : A MOLECULAR AND IMMUNOLOGICAL APPROACH., Englund P.T and Sher A., eds. Alan R. Liss, Inc., N.Y., 371-400.

71. Dorfman D. and Donelson J. (1984) Characterisation of the 1.35 Kb DNA repeat unit containing the conserved 35 nucleotides of the VSG mRNAs in Trypanosoma brucei. NUC. ACIDS RES., 12, 4907-4920.

72. Duggan A.J. (1970) A historical perspective. In, THE AFRICAN TRYPANOSOMIASES, Mulligan, H.W. and Potts, W.H., eds. Min. Overseas Dev., Geo. Allen and Unwin., xli-lxxxviii.

73. Earnshaw W.C., Halligan B., Cooke C.A. and Rothfield N. (1984) The kinetchore is part of the metaphase chromosome scaffold. J. CELL. BIOL., 98, 352-357.

74. Ehlers B., Czichos J. and Overath P. (1987) RNA turnover in Trypanosoma brucei. MOL. CELL. BIOL., 7, 1242-1249.

75. Ehrlich and Wang (1981) 5-methylcytosine in eukaryotic DNA. SCIENCE, 212, 1350-1357.

76. Englund P.T., Hereld D., Krakow J., Doering T., Masterson W. and Hart G.W. (1988) Glycolipid membrane anchor of the trypanosome variant surface glycoprotein: its biosynthesis and cleavage. In, THE BIOLOGY OF PARASITISM, A MOLECULAR AND IMMUNOLOGICAL APPROACH. Englund, P.T. and Sher, A., eds. A.R. Liss Inc., 401-412.

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

78. 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.

79. Farney J.D. and Blackburn E.H. (1988) Developmentally controlled telomere addition in wild type and mutant Paramecium MOL. CELL. BIOL., 8, 251-258.

80. Ferguson M.A.J., Duszenko M., Lamont G.S., Overath P. and Cross G.A.M. (1986) Biosynthesis of a Trypanosoma brucei variant surface glycoprotein, N-glycosylation and addition of a phosphatidyl membrane anchor. J. BIOL. CHEM., 261, 356-362.

81. Fincham J.R.S. and Oliver P. (1989) Initiation of recombination. NATURE, 338, 14.

82. Ford J. (1979) Ideas which have influenced attempts to solve the problem of African Trypanosomiasis. SOCIOLOG. SCI. MED., 30B, 269-275.

83. Freistadt M.S., Cross G.A.M. and Robertson H.D. (1988) Discontinuously synthesised mRNA from Trypanosoma brucei contains the highly methylated cap structure m7GpppA*C(2-6)mU*A* J. BIOL. CHEM., 263, 15071-15075.

84. Freistadt M.S., Cross G.A.M., Branch A.D. and Robertson H.D. (1987) Direct analysis of the mini exon donor RNA of Trypanosoma brucei : detection of a novel cap structure also present in messenger RNA. NUC. ACIDS RES., 15, 9861-9877.

85. Gardiner P.R.T.W., Clarke M.C. and Mutharia L.M. (1987) Identification and isolation of a variant surface glycoprotein from Trypanosoma vivax. SCIENCE, 235, 774-777.

86. Gasser C.P. and Laemmli V.K. (1987) A glimpse at chromosomal order. TRENDS GENET., 3, 16-22.

87. Gibbs C.P. and Cross G.A.M. (1988) Cloning and transcriptional analysis of a variant surface glycoprotein expression site in Trypanosoma brucei. MOL. BIOCHEM. PARASITOL., 28, 197-207.

88. Gibson W.C. (1985) Molecular karyotyping. PARASITOL. TODAY, 1, 64-65.

89. Gibson W.C. and Borst P. (1986) Size fractionation of small chromosomes of Trypanozoon and Nanomonas trypanosomes by PFG gel electrophoresis. MOL. BIOCHEM. PARASITOL., 16, 231-242.

90. Gibson W.C. and Gashumba J.K. (1983) Isoenzyme characterisation of some Trypanozoon stocks from a recent trypanosomiasis epidemic in Uganda. TRANS. ROY. SOC. TROP. MED. HYG., 77, 114-118.

91. Gibson W.C., Marshall T.F. and Godfrey D.G. (1980) Numerical analysis of enzyme polymorphisms: a new approach to the epidemiology and taxonomy of trypanosomes of the sub-genus Trypanozoon. ADV. PARASITOL., 18, 175-246..

92. 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.

93. Gonzalez A., Lerner T.J., Huecas M., Sosa-Pineda B., Nogueira N. and Lizardi P.M. (1985) Apparent generation of segmented mRNA from two separate gene families in Trypanosoma cruzi. NUC. ACIDS RES., 13, 5789-5804.

94. Gorski K., Roch J.M., Prentki P. and Krisch H.M. (1985) The stability of the bacteriophage T4 gene 32 mRNA : a 5' spliced leader sequence that can stabilize mRNA transcripts. CELL, 43, 461-469.

95. Graham S.V. and Birnie G.D. (1986) DNA and RNA. In, NUCLEAR STRUCTURES, Mc Gillivray A.J. and Birnie G.D., eds. Butterworths., 182-201.

96. Gray A.J. (1972) Variable agglutinogenic antigens of Trypanosoma gambiense and their distribution among isolates of the trypanosomes collected in different places in Nigeria. TRANS. ROY. SOC. TROP. HYG. MED., 66, 263-284.

97. Greaves D. and Borst P. (1987) Trypanosoma brucei variant specific glycoprotein gene chromatin is sensitive to single strand specific endonuclease digestion. J. MOL. BIOL., 197, 471- 483.

98. Greenleaf A.L. (1983) Amanitin resistant RNA polymerase II mutations are in the enzymes largest sub-unit. J. BIOL. CHEM., 258, 13403-13406.

99. Greenleaf A.L., Basett L.M., Jiamachello P.F. and Coulter D.E. (1979) Alpha-amanitin resistant Drosophila melanogaster with an altered RNA polymerase II. CELL, 18, 613-622.

100. 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.

101. Haas R. and Meyer T.F. (1986) The repertoire of silent pilus genes in Neisseria gonorrhoea : evidence for gene conversion. CELL, 44, 107-115.

102. Hajduk S.L. and Vickerman K. (1981) Antigenic variation in cyclically transmitted Trypanosoma brucei. Variable antigen composition of the first parasitaemia in mice bitten by trypanosome-infected Glossina morsitans. PARASITOLOGY, 83, 609-621.

103. 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. PARASITOLOGY, 83, 595-607.

104. Hall T. and Esser K. (1984) Topological mapping of protective and non protective epitopes on the variant surface glycoproteins of the WRATat 1 clone of Trypanosoma brucei rhodesiense. J. IMM., 132, 2059-2063.

105. Harley J.M.B., Cunningham M.P. and Van Hoeve K. (1966) The numbers of Trypanosoma rhodesiense extruded by Glossina morsitans during feeding. ANN. TROP. MED. PARASITOL., 60, 455-460..

106. Hart R.P., Mc Devitt M.A. and Nenne J.R. (1985) Poly(A) site cleavage in HeLa nuclear extract is dependent on downstream sequences. CELL, 43, 677-683.

107. Hecker H. and Gander E.S. (1985) The compaction pattern of the chromatin of trypanosomes. BIOL. CELL, 53, 199-208.

108. Heitman J. and Model P. (1987) Site specific methylases induce the SOS repair system in Escherischia coli J. BACT., 169, 3243-3250.

109. Henderson E.R. and Blackburn E.H. (1989) An overhanging 3' terminus is a conserved feature of telomeres, MOL. CELL. BIOL., 9, 345-348.

110. Herskowitz I. (1983) Cellular differentiation, cell lineages and transposable genetic cassettes in yeast. CURR. TOP. DEV. BIOL., 18, 1-14..

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

112. Hodo H.G. a. J Hatcher F.M. (1986) Multiple DNA dependent RNA polymerases in Trypnosoma cruzi. MOL. BIOCHEM. PARASITOL., 19, 77-81.

113. Hoeijmakers J.H.J., Borst P., Van den Burg J., Weissman C. and Cross G.A.M. (1980) The isolation of plasmids containing DNA complementary to mRNA for variant surface glycoproteins of Trypanosoma brucei. GENE, 8, 391-417.

114. Holder A.A. and Cross G.A.M. (1981) Glycopeptides from variant surface glycoproteins of Trypanosoma brucei. C terminal location of antigenically crossreacting moieties. MOL. BIOCHEM. PARASITOL., 2, 135-150.

115. Honigberg S.M. and Radding C.M. (1988) The mechanics of winding and unwinding helices in recombination : torsional stress associated wtih strand transfer promoted by Rec A protein. CELL, 54, 525-532.

116. 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. NUCLEIC ACIDS RES., 15, 7357 - 7368.

117. Imboden M., Blum B., De Lange T., Braun R. and Seebeck T. (1985) The tubulin mRNAs of Trypanosoma brucei. J. MOL. BIOL., 188, 393-402..

118. Jenni L., Marti S., Schweiger 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.

119. Johnson J.G. and Cross G.A.M. (1979) Selective cleavage of the variant surface glycoproteins from Trypanosoma brucei. BIOCHEM. J., 178, 689-697.

120. Johnson P.J. and Borst P. (1986) Mapping of VSG genes on large expression site chromosomes of Trypanosoma brucei separated by PFGE. GENE, 43, 2213-2220.

121. Johnson P., Kooter J.M. and Borst P. (1987) Inactivation of transcription by UV irradiation of Trypanosoma brucei provides evidence for a multicistronic transcription unit including a VSG gene. CELL, 51, 273-281.

122. Jordan A.M. (1985) Tsetse eradication plans for southern Africa. PARASITOL. TODAY, 1, 121-123.

123. Kaiser K. and Murray N.E. (1985) The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries. In, DNA CLONING, VOL.I, Glover D., ed. IRL Pr. Oxford, 1-65.

124. Kataoaka J., Broek D. and Wigler M. (1985) DNA sequence and characterisation of the Saccharomyces cerevisiae gene encoding adenylate cyclase. CELL,43, 493-505.

125. Keshet I., Lieman-Hurwitz J. and Cedar H. (1986) DNA methylation affects the formation of active chromatin. CELL, 44, 535-543.

126. Kimmel B.E., Onesmo K., Ole Moyoi and Young J.R. (1987) INGI, a 5.2Kb dispersed sequence element from Trypanosoma brucei that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. MOL. CELL. BIOL., 7 1465-1475.

127. Kimmel B.E., Samson S., Wu J., Hirschberg R. and Yarbrough L.R. (1985) Tubulin genes of the African trypanosome Trypanosoma brucei rhodesiense: nucleotide sequence of a 3.7 kb fragment containing genes for alpha and beta tubulins. GENE, 35, 237 - 248.

128. Konarska M.M., Padgett R.A. and Sharp P.A. (1984) Recognition of cap structure in splicing in vitro of mRNA precursors. CELL, 38, 731-736.

129. Kooter J.M. (1988) Transcription of variant surface antigen genes in Trypanosoma brucei. PH.D. THESIS, Univ. Amsterdam, The Netherlands.

130. Kooter J.M. and Borst P. (1984) Alpha-amanitin insensitive transcription of a variant surface glycoprotein gene provides further evidence for discontinuous transcription in trypanosomes. NUC. ACIDS RES., 12, 9457-9472.

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

132. 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 T.brucei. CELL, 51, 261-272.

133. Kooter J.M., Winter A.J., d'Oliveira C., Wagter R. and Borst P. (1988) 5' boundries of telomere conversions in Trypanosoma brucei. GENE, 69, 1-11.

134. Krainer A.R., Maniatis T., Ruskin B. and Green M.R. (1984) Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced in vitro. CELL, 36, 993-1005.

135. Krause M. and Hirsh D. (1987) A trans-spliced leader sequence on actin mRNA in Caenorhabditis elegans. CELL, 49, 753- 761.

136. Krawinkel U. (1983) Mouse mRNA fragment hybridising weakly to a heavy chain C-gamma 3 probe. UNIV. KOELN STAFF REPORT, F.R.G.

137. Laird P.W., Kooter J.M., Loosbroek N. and Borst P. (1985) Mature mRNAs of Trypanosoma brucei possess a 5' cap structure acquired by discontinuous RNA synthesis NUC. ACIDS RES., 13, 4253-4266.

138. Laird P.W., Zomerdijk J.C.E.M., De Korte D. and Borst P. (1987) In vivo labelling of intermediates in the discontinuous synthesis of mRNAs in Trypanosoma brucei. EMBO J, 6, 1055-1062.

139. Lalor T.M., Kjeldegaard M., Shimamoto G.T., Strickler J.T., Konigsberg W.A. and Richards F.F. (1984) A polygene protein family with multiple folding patterns. P.N.A.S.(USA), 81, 998- 1002.

140. 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.

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

142. Laurent M., Pays E., Delinte K., Magnus E., Van Meirvenne N., Steinert M. (1984b) Evolution of a surface antigen gene repertoire linked to non duplicative gene activation. NATURE, 308, 370-373.

143. Laurent M., Pays E., Van der Werf A., Aerts D., Magnus E., Van Meirvenne N. and Steinert M. (1984a) Translocation alters the activation rate of a trypanosome surface antigen gene. NUC. ACIDS RES., 12, 8319-8328.

144. Leach D.R.F. and Stahl F.W. (1983) Viability of lambda phages carrying a perfect palindrome in the absence of recombination nucleases. NATURE, 305, 448-449.

145. Lenardo M.J., Dorfman D. and Donelson J.E. (1985) The spliced leader of Trypanosoma brucei has a potential role as a cap donor structure. MOL. CELL. BIOL, 5, 2487-2490.

146. Lenardo M.J., Esser K., Moone A.M., Van der Ploeg L.H.T. and Donelson J.E. (1986) Metacyclic variant surface glycoprotein genes of Trypanosoma brucei subsp. rhodesiense are activated in situ and their expression is transcriptionally regulated. MOL. CELL. BIOL., 6, 1991-1997.

147. Lenardo M.J., Rice-Ficht A.C., Kelly G., Esser K. and Donelson J.E. (1984) Characterisation of the genes specifying two metacyclic variable antigen tyes. P.N.A.S. (USA), 81, 6642-6646.

148. Lima De Faria A. (1983) Organization and function of telomeres. In, MOLECULAR EVOLUTION AND ORGANIZATION OF THE CHROMOSOME., Amsterdam, Elsevier Sci., 701-721.

149. Liu A.Y.C., Michels P.A.M., Bernards A. and Borst P. (1985) Trypanosome surface glycoprotein genes expressed early in infection. J. MOL. BIOL., 182, 383-396.

150. Liu A.Y.C., Van der Ploeg L.H.T., Rijsewijk F.A.M. and Borst P. (1983) The transcription unit of variant surface glycoprotein gene 118 of Trypanosoma brucei. Presence of repeated elements at its border and absence of promoter-associated sequences. J. MOL. BIOL., 167, 57-75.

151. Losick R. and Pero J. (1982) Cascades of sigma factors. CELL, 25, 582-584.

152. Lund B., Lindberg B., Marklund I. and Normark S. (1987) The pap protein is the a-D-galactopyranosyl-(1-4)-B-D-galactopyranose binding adhesin of uropathogenic Escherischia coli P.N.A.S. (USA), 84, 5898-5902.

153. Majumder H.K., Boothroyd J.C. and Weber H. (1981) Homologous 3'terminal regions for surface antigens of different variants of Trypanosoma brucei. NUC. ACIDS RES., 9, 4745.

154. Matthyssens G., Michiels F., Hammers R., Pays E. and Steinert M. (1981) The variant surface glycoproteins of Trypanosoma brucei have a conserved terminus. NATURE, 293, 230- 233.

155. Mc Connell J., Turner M. and Rovis R. (1983) Biosynthesis of Trypanosoma brucei variant surface glycoproteins - Analysis of carbohydrate heterogeneity and timing of post translational modifications MOL. BIOCHEM. PARASITOL., 8, 119-135.

156. Mc Neillage G.J.C., Herbert W.J. and Lumsden W.H.R. (1969) Antigenic type of first relapse variants arising from a strain of Trypanosoma (Trypanozsoon) brucei. EXP. PARASITOL., 25, 1-7.

157. Messing J. (1983) New M13 vectors for cloning. METHODS ENZYMOL., 101, 20-78.

158. Michels P.A.M., Liu A.Y.C., Bernards A., Sloof P., Van der Bijl M.M.W., Schinkel A.H., Menke H.H. and Borst P. (1983) Activation of the genes for variant surface glycoproteins 117 and 118 in Trypanosoma brucei. J. MOL. BIOL., 166, 537 -556.

159. Michels P.A.M., Van der Ploeg L.H.T., Liu A.Y.C. and Borst P. (1984) Inactivation and reactivation of an expression linked gene copy for a variant surface glycoprotein in Trypanosoma brucei EMBO J., 2, 1345-1351.

160. Michiels F., Matthyssens G., Kronenberger P., Pays E., Dero B., Van Asassel S., Darville M., Cravador A., Steinert M. and Hammers R. (1983) Gene activation and re-expression of a Trypanosoma brucei variant surface glycoprotein. EMBO J., 2, 1185-1192.

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

162. Miller A.M. and Nasmyth K.A. (1984) Role of DNA replication in the repression of silent mating type loci in yeast. NATURE, 312, 247-251.

163. Miller E.M., Allan L.M. and Turner M.J. (1984) Topological analysis of the antigenic determinants on a variant surface glycoprotein of Trypanosoma brucei. MOL. BIOCHEM. PARASITOL, 13, 67-81.

164. Miller E.N. and Turner M.J. (1981) Analysis of antigenic types appearing first relapse populations of clones of Trypanosoma brucei PARASITOL., 82, 63-80.

165. Mowry K.L. and Steitz J.A. (1987) Both conserved signals on mammalian histone pre-mRNAs associate with small nuclear ribonucleoproteins during 3'end formation in vitro. MOL. CELL. BIOL., 7, 1663-1672.

166. Muhich M.L., Hughes D.E., Simpson A.M. and Simpson L. (1987) The monogenetic kinetoplastid protozoan, Crithidia fasiculata, contains a transcriptionally active, multicopy mini exon sequence. NUC. ACIDS RES., 15, 3141-3153.

167. Murphy N.B., Guyaux M., Pays E. and Steinert M. (1987) Analysis of VSG expression site sequences in Trypanosoma brucei. IN, MOLECULAR STRATEGIES OF PARASITIC INVASION., Agabian, N., Goodman, H. and Nogueiro eds., Alan R. Liss Inc., N.Y., 449-469.

168. Murphy N.B., Pays A., Tebabi P., Coquelet H., Guyaux M., Steinert M. and Pays E. (1987) Trypanosoma brucei repeated element with unusual structural and transcriptional properties. J. MOL. BIOL., 195, 855-871.

169. 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.

170. Murray A.W. and Szostak J.W. (1983) Construction of artificial chromosomes in yeast. NATURE, 305, 189-193.

171. Myler P.J., Allison J., Agabian N. and Stuart K. (1984) Antigenic variation in African trypanosomes by gene replacement or activation of alternate telomeres. CELL, 39, 203-211.

172. Myler P.J., Aline R.F., Scholler J.K. and Stuart K. (1988) Changes in telomere length associated with antigenic variation in Trypanosoma brucei. MOL. BIOCHEM. PARASITOL., 29, 243-250.

173. Nash T.A.N. (1960) A review of the African Trypanosomiasis problem. TROP. DIS. BULL., 57, 973-1003.

174. Nelson R.G., Parsons M., Barr P.J., Stuart K., Selkirk M., and Agabian N. (1983) Sequences homologous to the variable antigen mRNA spliced leader are located in tandem repeats and variable orphons in Trypanosoma brucei. CELL, 34, 901-909.

175. Nelson R.G., Parsons M., Selkirk M., Newport G. and Agabian N. (1984) Sequences homologous to the variable antigen mRNA in Trypanasomatidae which do not undergo antigenic variation. NATURE, 308, 665-667.

176. Nickol J.M. and Felsenfeld G. (1983) DNA conformation at the 5' end of the chicken B-globin gene. CELL, 35, 467-477.

177. Nonet M., Sweetser D. and Young R.A. (1987) Functional redundancy and structural polymorphismin a large subunit of RNA polymerase II. CELL, 50, 909-915.

178. Omerod W.E. (1961) The epidemic spread of rhodesian sleepng sickness 1908-1960. TRANS. ROY. SOC. TROP. MED. HYG., 55, 525- 538.

179. 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.

180. Overath P., Czichos J., Stock U. and Nonnengaesser C. (1983) Repression of glycoprotein synthesis and release of surface coatduring transformation of Trypanosoma brucei. EMBO J., 2, 1721-1728.

181. Paindavoine P., Zampetti-Bosseler F., Coquelet H., Pays E. and Steinert M. (1989) Different allele frequencies in Trypanosoma brucei brucei and Trypanosoma brucei gambiense populations. MOL. BIOCHEM. PARASITOL., 32, 61-72.

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

183. Paule M.R., Iida C.T., Perna P.J., Harris G.H., Knoll D.A.D. and Alessio J.M. (1984) In vitro evidence that eukaryotic ribosomal RNA transcription is regulated by the modification of RNA polymerase I. NUC. ACIDS RES., 12, 8161-8180.

184. Pays E. and Steinert M. (1988) Control of antigen gene expression in African trypanosomes. ANNU. REV. GENET., 22, 107-126.

185. Pays E., Coquelet H., Pays A., Tebabi P. and Steinert M. (1989b) Post transcriptional control of variant surface glycoprotein gene expression site. MOL. CELL. BIOL., 9, 4018- 4021.

186. Pays E., Delauw M.F., Laurent M. and Steinert M. (1984) Possible DNA modification in GC dinucleotides of Trypanosoma brucei telomeric sequences; relationship with antigen gene transcription. NUC. ACIDS RES., 12, 5235-5249.

187. Pays E., Delauw M.F., Van Assel S., Laurent M., Vervoort T., Van Meirvenne N. and Steinert M. (1983) Modification of a Trypanpsoma brucei brucei antigen repertoire by different recombinational mechanisms. CELL, 39, 263-266.

188. Pays E., Guyaux M., Aerts D., Van Meirvenne N. and Steinert M. (1985) Telomere reciprocal recombination as a mechanism for antigenic variation in trypanosomes. NATURE, 316, 562 - 564.

189. Pays E., Laurent M., Delinte K., Van Meirvenne N. and Steinert M. (1983) Differential size variations between transcriptionally active and inactive telomeres of Trypanosoma brucei. NUC. ACIDS. RES., 11, 8137-8147.

190. Pays E., Lheureux M. and Steinert M. (1981a) The expression linked copy of a surface antigen gene in Trypanosoma brucei is probably the one transcribed. NATURE, 292, 265-267.

191. Pays E., Lheureux M. and Steinert M. (1982) Structure and expression of a Trypanosoma brucei gambiense variant specific antigen gene. NUC. ACIDS RES., 10, 3149-3163.

192. 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 T.brucei. CELL, 57, 835-845.

193. Pays E., Van Meirvenne N., Le Ray D. and Steinert M. (1981b) Gene duplication linked to antigenic variation in Trypanosoma brucei. P.N.A.S.(USA), 78, 2673-2677.

194. Raleigh E.A. and Wilson G. (1981) Escherischia coli K12 restricts DNA containing 5-methylcytosine P.N.A.S.(USA), 83, 9670-9674.

195. Raleigh E.A., Murray N.E., Revel H., Blumenthal R.M., Westaway D., Reith A.D., Rigby P.W.J., Elhai J. and Hanahan D. (1988) mcr A and mcr B restriction phenotypes of some Escherischia coli strains and implications in gene cloning. NUC. ACIDS RES., 16, 1563-1575.

196. Rice-Ficht A.C., Chen K.K. and Donelson J.E. (1981) Sequence homologies near the C terminal of variable surface glycoproteins of Trypanosoma brucei. NATURE, 243, 53-57.

197. Roth C.W., Longacre S.L., Raibaud A., Baltz T. and Eisen H. (1986) The use of incomplete genes for the construction of a Trypanosoma equiperdum variant surface glycoprotein gene. NATURE, 308, 370-373.

198. Sanger F., Nicklen S. and Coulson A.R. (1977) DNA sequencing with chain terminating inhibitors. P.N.A.S.(USA), 74, 5463-5467.

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

200. Seed J.R., Edwards R. and Sechelski J. (1984) The ecology of antigenic variation. J. PROTOZOOL., 31, 48-53.

201. Segal E., Billyard E., So M., Storzbach S. and Meyer T.F. (1985) Role of chromosomal rearrangements in Neisseria gonorrhoea pilus phase variation. CELL, 40, 293-300.

202. Seifert S. and So M. (1988) Genetic mechanisms of Bacterial antigenic variation. MICROB. REV., 52, 327-336.

203. Sentenac A. (1985) Eukaryotic RNA polymerases. CRTC CRIT. REV. BIOCHEM., 18, 31-90.

204. Sharp P.A. and Marciniak R.A. (1988) HIV TAR : an RNA enhancer? Cell,59, 229-230.

205. Shatkin A.J. (1976) Capping of eukaryotic mRNAs. CELL, 9, 645-653.

206. Shea C. and Van der Ploeg L.H.T. (1988) Stable variant specific transcripts of the variant cell surface glycoprotein gene 1.8 expression site in Trypanosoma brucei MOL. CELL. BIOL., 8, 854-859.

207. Shea C., Glass D.J., Parangi S. and Van der Ploeg L.H.T. (1986) Variant surface glycoprotein gene expression site switches in Trypanosoma brucei. J.BIOL. CHEM., 261, 6056-6063.

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

209. Sloof P., Bos J.L., Konings A.F.J.M., Menke H.M., Borst P., Gutteridge W.E. and Leon W. (1983) Characterization of satellite DNA in Trypanosoma brucei and Trypanosoma cruzi. J. MOL. BIOL., 167, 1-21.

210. Sogin M.L., Elwood H.J. and Gunderson J.H. (1986) Evolutionary diversity of eukaryotic small subunit rRNA genes. P.N.A.S.(USA), 83, 1383-1387.

211. Son H.J., Cook G.A., Hall T. and Donelson J.E. (1989) Expression site associated genes of Trypanosoma brucei rhodesiense. MOL. BIOCHEM. PARASITOL., 33, 59-66.

212. Sonnenborn T.M. (1970) The cytoplasm in heredity. HEREDITY, 4, 11-36.

213. Southern E. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. MOL. BIOL., 98, 503-517.

214. Spratt B.G., Hedge P.J., te Heesens S., Eldeman A. and Broome-Smith J.K. (1986) Kanamycin resistant vectors that are analogous to pUC8, pUC9, pEMBL8 and pEMBL9. GENE, 41, 337-342.

215. Stephens J.W.W. and Fantham H.B. (1910) On the peculiar morphology of a trypanosome from a case of sleeping sickness and the possibility of it being a new species (Trypanosoma rhodesiense). PROC. ROY. SOC., SERIES B, 83, 28-33.

216. Sternberg J., Tait A., Haley S., Wells J.M., Le Page R.W.F., Schweizer J. and Jenni L. (1988) Genetic exchange in African Trypanosomes : characterisation of a new hybrid genotype. MOL. BIOCHEM. PARASITOL., 27, 191-200.

217. Stoenner H.G., Dodd T. and Larsen C. (1982) Antigenic variation of Borrelia hermsii. J. EXP. MED., 156, 1297-1311.

218. Sun H., Treco D., Schultes N.P. and Szostak J. (1989) Double strand breaks at the initiation sites for meiotic gene conversion. NATURE, 338, 87-90.

219. Sutton R.E. and Boothroyd J.C. (1986) Evidence for trans-splicing in Trypanosomes CELL, 47, 527-535.

220. Sutton R.E. and Boothroyd J.C. (1988) Trypanosome trans-splicing utilizes 2'- 5' branches and a corresponding de- branching activity. EMBO J., 5, 1431-1437.

221. Swinkels B.W., Gibson W.C., Osinga K.A., Kramer R., Veeneman G.H., van Boom J.H. and Borst P. (1986) Characterisation of the gene for the microbody (glycosomal) triosephosphate isomerase of Trypanosoma brucei. EMBO J., 5, 1291-1298.

222. Szostak J.W., Orr-Weaver T.L. and Rothstein R.J. (1983) The double strand-break-repair model for recombination. CELL, 33, 25-35.

223. Tait A. (1980) Evidence for diploidy and mating in trypanosomes. NATURE, 287, 536-538..

224. Tait A., Barry J.D., Wink R., Sanderson A. and Crowe J.S. (1985) Enzyme variation in Trypanosoma brucei ssp.II. Evidence for Trypanosoma brucei rhodesiense being a set of variants of Trypanosoma brucei. PARASITOL., 90, 89-100.

225. Tait A., Eldirdiri A.B. and Le Ray D. (1984) Enzyme variation in Trypanosoma brucei ssp. I. Evidence for the sub- specification of Trypanosoma brucei gambiense. PARASITOL., 89, 311-326.

226. Takacs A.M., Denker J.A., Perrine K.G., Maroney P.A. and Nilsen T.W. (1988) A 22-nucleotide spliced leader sequence in the human parasitic nematode Brugia malayi is identical to the trans-spliced leader exon in Caenorhabditis elegans. P.N.A.S.(USA), 85, 7932-7936.

227. Tetley L., Turner C.M.R., Barry J.D., Crowe J.S. and Vickerman K. (1987) Onset of expression of the variant surface glycoproteins of Trypanosoma brucei in the tsetse studied using immunoelectron microscopy. J. CELL. SCI., 87, 363-372.

228. Tetley L., Vickerman K. and Moloo S.K. (1981) Absence of a surface coat from metacyclic Trypanosoma vivax : possible implications for vaccination against vivax trypanosomiasis. TRANS. ROY. SOC. TROP. MED. HYG., 75, 409-414.

229. Tittawella I. (1988) Variable RNA polymerase populations in the lifecycle of Trypanosoma brucei. F.E.B., 227, 122-126.

230. Tower J. and Solner-Webb B. (1987) Transcription of mouse rDNA is regulated by an activated sub form of RNA polymerase I. CELL, 50, 873-883.

231. Tsao Y.P., Wu H.Y. and Liu L.F. (1989) Transcription driven supercoiling of DNA : direct chemical evidence from in vitro studies. CELL, 56, 111-118.

232. Tschudi C. and Ullu E. (1988) Polygene transcripte are precursors to calmodulin mRNAs in Trypanosomes. EMBO J., 7, 455-463.

233. Turner C.M.R. and Barry J.D. (1989) High frequency of antigenic variation in Trypanosoma brucei rhodesiense infections. PARASITOL., 99, 67-75.

234. Turner C.M.R., Barry J.D. and Vickerman K. (1986) Independent expression of the metacyclic and bloodstream variable antigen repertoires in Trypanosoma brucei rhodesiense. PARASITOL., 92, 67-73.

235. Turner C.M.R., Barry J.D., Maudlin I. and Vickerman K. (1988) An estimate of the size of the metacyclic variable antigen repertoire of Trypanosoma brucei rhodesiense. PARASITOL., 97, 269-276.

236. Turner M.J., Cardoso de Almeida M.L., Gurnett A.M., Raper J. and Ward J. (1985) Biosynthesis, attachment and release of variant specific glycoproteins of the African trypanosomes. CURR. TOPS. IN MICROBIOL. AND IMMUNOL., 117, 23-55.

237. Van der Ploeg H.T., Valerio D., De Lange T., Bernards A., Borst P. and Grosveld F.G. (1982b) Analysis of cosmid clones of nuclear DNA from Trypanosoma brucei shows that the genes for variant surface glycoproteins are clustered in the genome. NUC. ACIDS RES., 10, 5905-5923.

238. Van der Ploeg L.H.T., Bernards A., Rijsewijk F.A.M. and Borst P. (1982a) Characterisation of the DNA duplication transposition that controls the expression of two genes for variant surface glycoproteins in Trypanosoma brucei. NUC. ACIDS RES., 10, 593-609.

239. Van der Ploeg L.H.T., Cornelissen A.W.C.A., Barry J.D. and Borst P. (1984d) Chromosomes of Kinetoplastida. EMBO J., 3, 3109-3115.

240 Van der Ploeg L.H.T., Cornelissen A.W.C.A., Michels P.A.M. and Borst P. (1984b) Chromosome rearrangements in Trypanosoma brucei. CELL, 39, 213-221.

241. 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.

242. Van der Ploeg L.H.T., Liu A.Y.C., Michels P.A.M., De Lange T., Borst P., Majumder H.K., Weber H., Veeneman G.H., and Van Boom J.H. (1982c) RNA splicing is required to make the messenger RNA for a variant surface antigen in trypanosomes. NUC. ACIDS RES., 10, 3591-3604.

243. Van der Ploeg L.H.T., Schwartz D.C., Cantor C.R. and Borst P. (1984a) Antigenic variation in Trypanosoma brucei analysed by electrophoretic separation of chromosome sized DNA molecules. CELL, 37, 77-84.

244. Vickerman K. (1969) On the surface coat and flagellar adhesions in trypanosomes. J. CELL SCI., 5, 169-193.

245. Vickerman K. (1978) Antigenic variation in trypanosomes NATURE, 273, 613-617.

246. Vickerman K. and Barry J.D. (1982) African trypanosomiasis. In, IMMUNOLOGY OF PARASITIC INFECTIONS., Cohen, S. and Warren, K.S., eds. Blackwell, Oxford., 204-260.

247. Vickerman K. and Luckins A.G. (1969) Localization of variable antigens in the surface of Trypanosoma brucei using ferritin conjugated antibody. NATURE, 224, 1125-1126.

248. Vickerman K. and Preston T.M. (1970) Spindle microtubules in the dividing nuclei of trypanosomes. J. CELL. SCI., 6, 365-384.

249. Vieira J. and Messing J. (1982) The pUC plasmids, a M13mp7 - derived system for insertion mutagenesis and sequencing with synthetic universal primers. GENE, 19, 259-268.

250. Voorheis H.P and Martin R.B. "Swell dialysis" demonstrates that adenylate cyclase in Trypanosoma brucei is regulated by O{#?0 calcium ions. Eur. J. Biochem., 113, 223-227.

251. White T.C., Rudenko G. and Borst P. (1986) Three small RNAs within the 10 Kb trypanosome ribosomal transcriion unit are analogous to domain VII of other eukaryotic 28s ribosomal RNAs . NUC. ACIDS RES., 14, 9471-9489..

252. 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. NUC. ACIDS RES., 16, 6725-6736.

253. Williams R.O., Young J.R. and Majiwa P.A.O. (1982) Genome rearrangements of Trypanosoma brucei VSG genes: presence of a mini chromosome. NATURE, 299, 417-421.

254. Williams R.O., Young J.R. and Majiwa P.A.O. (1979) Genomic rearrangements correlated with antigenic variation in Trypanosoma brucei. NATURE, 299, 847-849.

255. Wittig B., Dorbic T. and Rich A. (1989) The level of Z-DNA in mammalian cell nuclei is regulated by torsional strain. J. CELL. BIO., 108, 755-764.

256. Wyman A.R. and Wertman K.P. (1987) Host strains that alleviate under-representation of specific sequences ; an overview. IN, METHODS IN ENZYMOLOGY., Berger S.L.and Kimmel A.R., eds., Acad. Pr., 173-151.

257. Young J.R., Miller E.N., Williams R.O. and Turner M.J. (1983) Are there two classes of VSG gene in Trypanosoma brucei? NATURE, 306, 196-198.

258. Zampetti-Bosseler F., Schweiger J., Pays E., Jenni L. and Steinert M. (1986) Evidence of haploidy in metacyclic forms of Trypanosoma brucei. P.N.A.S.(USA), 83, 6063-6064.

259. Zieg J.M., Silverman M., Hilmen H. and Simon M. (1977) Recombinal switching for gene expression. SCIENCE, 196, 170-175.

260. Zomerdijk J.C.B.M., Ouellette M., ten Asbroek A.L.M.A., Kieft R., Bommer A.M.M. and Borst P. (1990) Active and inactive versions of a promoter for a variant specific glycoprotein gene expression site in Trypanosoma brucei. CELL, submitted.