Metacyclic VSG gene activation in *Trypanosoma brucei rhodesiense*.

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

by

Keith Roland Matthews

Institute of Genetics
Church Street
Glasgow

April 1990

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

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

ABBREVIATIONS.

CHEMICALS.

ATP : Adenosine triphosphate.

BSA : Bovine serum albumin.

CsCl : Caesium chloride.

ddH_2O : Distilled deionized water.

DNA : Deoxyribonucleic acid.

cDNA : Complementary DNA.

dNTP : 2' deoxy nucleotide.

DTT : Dithiothreitol.

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

EtBr : Ethidium bromide.

HCl : Hydrochloric acid.

IPTG : Isopropylthiogalactoside.

RNA : Ribonucleic acid.

mRNA : Messenger RNA.

RNAse : Ribonuclease A.

Poly A^+ : Poly adenylated.

SDS : Sodium dodecyl sulphate.

TEMED : N,N,N' N' tetramethyl ethylenediamine.

Tris : Tris (hydroxymethyl) amino methane.

TE : Tris EDTA.

X-gal : 5-bromo-4-chloro-3-indolyl-beta-D-galactoside.

Measurements.

mA : Milliamps.

bp : Base pairs.

^{o}C : Degrees centrigade

g : Centrifugal force equal to gravitational acceleration.

g : Grammes.

kb : Kilobase pairs.

kD : Kilodaltons.

Kg : Kilogrammes.

l : Litre.

M : Molar.

Mb : Megabase pairs.

mg : Milligrammes.

ml : Millilitres.

mM : Millimolar.

nm : Nanometre.

w/v : Weight / volume.

ug : Microgrammes.

ul : Microlitres.

Miscellaneous.

AnTat : Antwerp Trypanozoon antigen type.

AnTAR : Antwerp Trypanozoon antigen repertoire.

BC : Basic copy.

BoTat : Bordeaux Trypanozoon antigen type.

EATRO : East African Trypanosomiasis Research Organization.

ELC : Expression linked copy.

ESAG : Expression site associated gene.

GUTat : Glasgow University Trypanozoon antigen type.

GUG : Glasgow University, Genetics.

ILTat : ILRAD Trypanozoon antigen type.

MiTat : Molteno Institute Trypanozoon antigen type.

PFGE : Pulsed field gel electrophoresis

p.f.u. : Plaque forming units.

Serodemes : Stable, immunologically distinct strains exhibiting no cross immunity.

VAT : Variable antigen type.

B-VAT : Bloodstream VAT.

I-VAT : Ingested VAT.

M-VAT : Metacyclic VAT.

VSG : Variant surface glycoprotein.

ACKNOWLEDGEMENTS

Somewhere in this thesis I've written a disclaimer saying that the work presented here is all my own, unless otherwise stated. That is not true. None of the work would have been possible without the friendship and help of everybody in the Genetics Department and for that I want to thank them. Some more special thanks are needed, however. First, I want to thank Dave Barry for his supervision of the project and for providing the encouragement when it was needed. But go check the hot corner, Dave, 20 pence says it's a mess. Second, I have to thank Sheila for putting up with a large number of bad ideas, jokes and singing and for not shouting at me too much. It was all an attempt to provoke you into physical violence, Sheila, preferably with a blotted agarose gel. Thirdly I want to thank everone else in the lab, particularly Dave Killen early on, and more recently Jeremy and Iain for the "good tips". Paul Shiels also deserves a mention for the amateur psychology joke.

Now, the more general thanks. Mary Burke, of course, for giving me a home to go to whenever I was banished from the first floor, which unfortunately for the sixth floor was quite often. The sixth floor also put up with me outside hours on a number of occasions, as did nearly everyone else in the department and for that I am grateful. Finally I acknowledge the various aquatic, extraterrestrial and embryological friends who have kept my mind occupied and most people in the tea room amused for the last three years. I hope I didn't go on too much.

The work was supported by the Medical Research Council and the DHSS.

SUMMARY

African trypanosomes are able to evade the immune response of their mammalian host by their ability to change periodically their surface glycoprotein coat, in a process known as antigenic variation. This surface coat is first displayed in the parasites' metacyclic stage, which exists in the salivary glands of the parasites' insect vector, the tsetse fly. Previous analyses of the features of antigenic variation in the metacyclic stage have revealed that the system appears very different from the more complex and better characterized system employed in parasite populations established in the mammalian bloodstream. The work presented here has attempted to elucidate the reasons for these differences.

Firstly, the mechanism for the activation of the expression of the metacyclic variant surface glycoprotein (VSG) coat has been studied. This has been done by infecting tsetse flies with trypanosomes expressing metacyclic VSG genes as they are believed to be activated in the tsetse fly salivary gland. An analysis of the resulting metacyclic population after cyclical transmission has revealed that there is apparently no heritable genomic change associated with activation of these genes. This is distinct from what is found for bloodstream VSG genes, the activation signals for which are preserved through the tsetse fly, and suggests that the metacyclic VSG repertoire is reset with transmission.

Secondly, the structure of a metacyclic VSG gene expression locus has been investigated. This has involved the cloning, by the progressive isolation of overlapping restriction fragments, of 16Kb of the ILTat 1.61 gene environment. Analysis of the clones has revealed that the locus has just a very short upstream restriction site barren region, which contrasts with the long barren region associated with bloodstream VSG expression sites. Evidence for the presence of a region which cross hybridizes with one of the several families of expression site associated genes

found in all bloodstream VSG expression sites has also been found. An analysis of nascent transcripts from trypanosomes expressing the ILTat 1.61 gene as it is believed to be expressed in the tsetse fly, has revealed that the transcription unit is very short with respect to other examined VSG expression sites.

Finally, the reactivation of a gene encoding the metacyclic variant antigen type ILTat 1.22 in established bloodstream parasites has been investigated. This reactivation has been found always to involve the generation of a duplicate gene copy by conversion of one of a number of distinct bloodstream VSG expression sites. The upstream limit of the gene conversion process which results in the generation of this duplicate gene has been found always to occur within a region containing just 1.5 copies of the transposition associated 70bp repeat motif, of which upstream barren regions on bloodstream expression sites are composed. The sequence of the ILTat 1.22 gene and transposed segment has been determined. Neither is unusual with respect to the environment of bloodstream VSG genes.

The determined characteristics of metacyclic VSG gene activation and expression site structure have provided potential explanations for the unusual features of surface coat expression in the tsetse fly.

Chapter 1

INTRODUCTION

1.1 Trypanosome Species.

The pathogenic Salivarian trypanosomes are of major economic and medical importance throughout the African continent. Around 50 million people are exposed to these parasites, resulting in 10,000-20,000 infections per year (Allsopp et al., 1985; Goodwin et al., 1985). More important is the cost to food production; one third of all African cattle are at risk from trypanosomiasis in over 37 countries.

A number of distinct species and subspecies are responsible for the African trypanosomiases (Hoare, 1972). These include Trypanosoma vivax, T.congolense, T.simiae, T.suis, T.brucei brucei, T.b.gambiense and T.b.rhodesiense, of which only the last two are human infective. The geographical distribution of these species matches the distribution of their intermediate host, the blood feeding tsetse fly (Glossina spp.). Only in the case of two other Salivarian trypanosomes, where this vector has been dispensed with, has the range of the parasite been free to expand. Thus, the purely mechanically transmitted T.evansi and the venereally spread horse parasite, T.equiperdum, are found in South America, Asia and Europe, in addition to tropical Africa.

1.2 The Biology of Trypanosoma brucei rhodesiense.

The epidemiology of T.b.rhodesiense, causing the acute form of human sleeping sickness, has been described by Apted (1970). The parasite, identified by Stephens and Fantham (1910), is morphologically indistinguishable from both the causative agents of chronic human sleeping sickness, T.b.gambiense, and the exclusively animal infective T.b.brucei. Its principal insect vectors, dependent upon location, are the woodland and bush inhabiting Glossina morsitans, G.swinnertoni,and G.pallipides. On rare occasions the riverine G.palpalis has also been implicated in the parasites' transmission. Although T.b.rhodesiense is human infective, man is only a chance host, with the game animals being

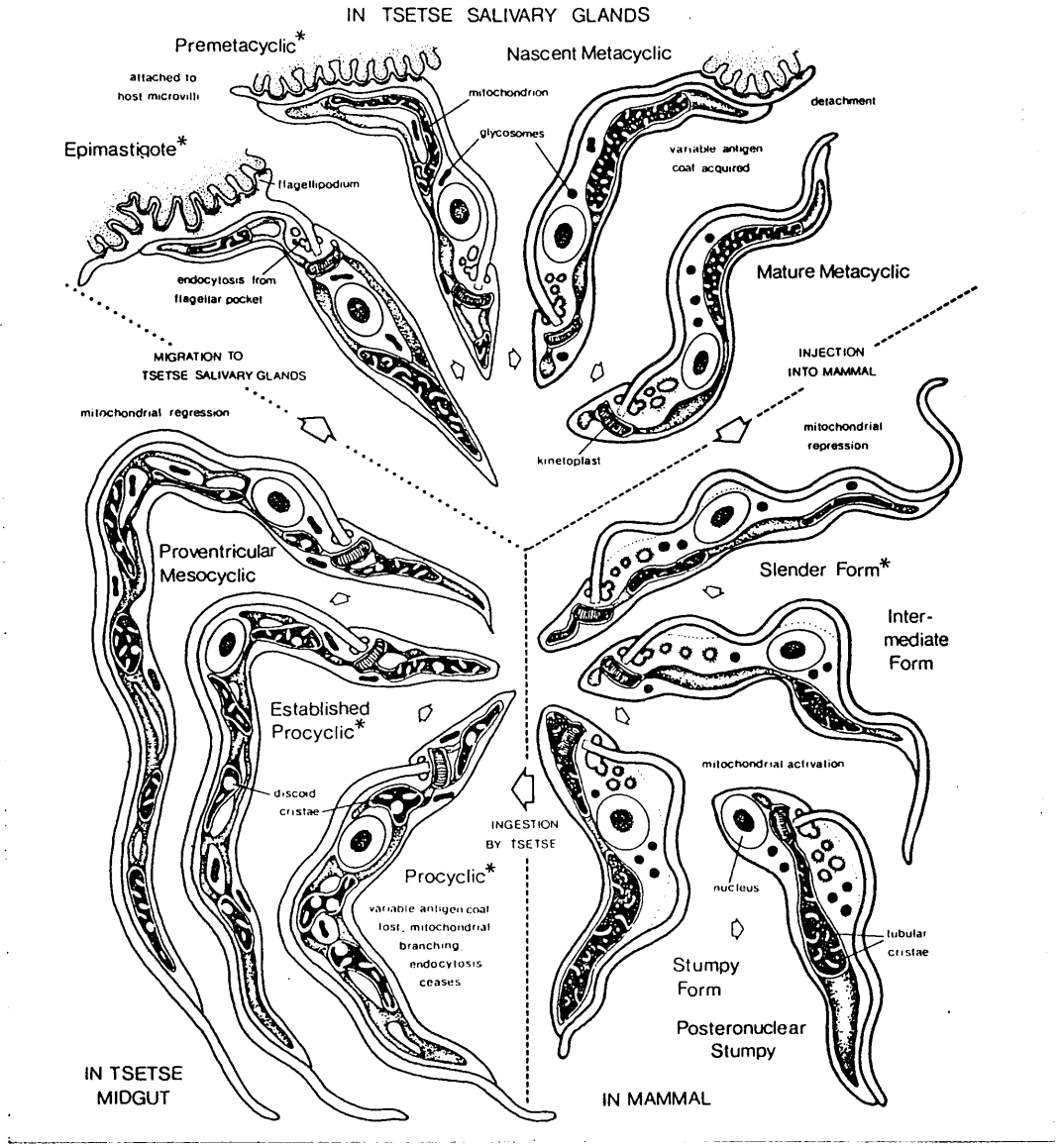


Figure 1.1; The life cycle of Trypanosoma brucei brucei (from Vickerman, 1985).

Facing page 2

the important reservoirs. In human epidemics, however, the parasite is probably more commonly spread in a man-tsetse-man cycle, with the trypanosome either undergoing development in the fly or being spread directly through infected blood in the proboscis (Apted, 1962).

1.3 Cyclical Development of T.b. rhodesiense.

Members of the T.brucei group undergo a great number of morphological, ultrastructural and biochemical changes during cyclical development (reviewed by Vickerman, 1985; Figure 1.1). In the mammalian bloodstream, the parasite displays pleomorphism (Vickerman, 1969), with long slender, intermediate and short stumpy trypomastigote forms being present. The long slender form is rapidly dividing, with a doubling time in the order of 5 to 10 hours (Myler et al., 1985; Turner and Barry, 1989), and displays its kinetoplast at a position posterior to the centrally located nucleus. The kinetoplast is contained within the parasite's mitochondrion and is unique to, and characteristic of, the kinetoplastida. It is composed of several thousand concatenated DNA minicircles which are heterogeneous in sequence, and approximately 50 identical DNA maxicircles (reviewed by Englund et al., 1982). The mitochondrial activity of long slender bloodstream form trypanosomes is fully repressed; living in a glucose rich environment, the parasite uses only glycolysis resulting in the production and excretion of pyruvate. Glycolytic reactions in the trypanosomatidae are unique in being confined to the microbody like organelle, the glycosome (Opperdoes and Borst, 1977).

Non-dividing short stumpy trypomastigotes become very abundant during the population crises characteristic of trypanosome infection and are believed to be responsible for infection of the tsetse fly (Wijers and Willet, 1960). To this end, they display the beginnings of mitochondrial reactivation. Once in the fly midgut, the parasites transform to the procyclic form, which is indistinguishable from those growing in in vitro culture media.

From the gut lumen, the procyclic trypanosomes are believed to migrate via the end of the peritrophic membrane, which forms a replaceable lining to the insect midgut, to the endoperitrophic space. Infection of the tsetse fly by trypanosomes is most successful when the insect is a newly emerged adult. One reason for this may be that at this stage the peritrophic membrane is not fully formed, making invasion of the endoperitrophic space easier. Another more recently identified possibility is that the level of trypanocidal lectins produced by the tsetse fly are quite low early on. This may be a consequence of their sequestration by high levels of glucosamine, a possible product of the chitinase activity of endosymbiotic rickettsia-like organisms present in trypanosome susceptible tsetse flies (reviewed by Maudlin and Welburn, 1988). Once in the endoperitrophic space the trypanosomes multiply extensively before penetrating the proventriculus at the point of secretion of the peritrophic membrane. They then migrate forward to the mouthparts and salivary glands, the final site of development for trypanosomes of the brucei group.

In the salivary gland, the proventicular parasites differentiate into the epimastigote stage. The epimastigote cells are attached to the salivary gland wall via flagellipodia and undergo multiple divisions, before forming the premetacyclic, and ultimately the metacyclic forms.

At some point prior to metacyclogenesis, trypanosomes can undergo sexual exchange (Tait, 1980; Jenni et al., 1986; Paindavoine et al., 1986; Sternberg et al., 1988). It has been suggested, on the basis of DNA content, that it may be the metacyclic form which is the gametic stage in this process (Zampetti-Bosseler et al., 1986). This has now been shown to be incorrect; metacyclic forms are diploid, having been observed to be both homozygous and heterozygous for some genetic markers (Tait et al., 1989).

The metacyclic stage exists free in the tsetse fly salivary gland lumen and is non replicative. It is this stage which is responsible for invasion of a new mammal host and to this end it displays a number of preadaptations for life in the bloodstream.

This entails a repression of their mitochondrial activity and the acquisition of a dense surface glycoprotein coat covering the entire parasite surface. This coat is present on all bloodstream stages of African trypanosomes as well as the metacyclics, and one of its functions is to prevent lysis of the parasite by non-specific mammalian immune mechanisms (Rifkin, 1978). Another is to assist in the evasion of specific humoral responses, which permits the parasite to survive for months, unsequestered, in the mammalian bloodstream.

1.4 Antigenic Variation: an Overview.

Trypanosome infection in mammals is characterized by the cyclical appearance and disappearance of a parasitaemia. In experimental clonal infection, the majority of members of each successive population express the same antigens contributed by the densely packed, identical variable surface glycoprotein (VSG) molecules, which mask other membrane proteins. When the host raises an antibody response, it does so against this highly immunogenic coat, such that all organisms bearing a particular variable antigen type (VAT) are destroyed. The population can be regenerated, however, because a small proportion of the parasites have switched to the expression of an antigenically distinct VSG. This exchange of surface antigens allows prolonged infection in the bloodstream and is known as antigenic variation.

The trypanosome's potential for antigenic variation is enormous: the descendents of a single trypanosome can produce at least 101 immunologically distinct coat types (Capbern *et al.*, 1977). The switch in coat types is not induced by the host immune response since trypanosomes in culture can change their surface coat (Lamont *et al.*, 1986), and is very unpredictable, at least in established infections. This, coupled with the observation that different trypanosomes in the field can display completely different VSG repertoires, has so far thwarted attempts to control the parasite using vaccine-based approaches. It has not, however, thwarted investigations into the nature of the surface molecule and the elegant and complex molecular

mechanisms the parasite uses to achieve its expression.

1.5 The Variant Surface Glycoprotein.

The VSG coat is displayed only on the bloodstream and metacyclic stages (Vickerman, 1969) and was first shown to be responsible for the trypanosome's different VATs by Vickerman and Luckins (1969). The VSG molecule is a 53,000-65,000 dalton glycoprotein (Cross, 1975) which can be cleaved by partial tryptic digestion into an N-terminal domain comprising two thirds of the molecule, and a C- terminal domain comprising the rest (Cross and Johnson, 1976). The N-terminal domain is extremely diverse in sequence, with no obvious conservation between different VSG molecules (Bridgen et al., 1976; Rice-Ficht et al., 1981). In an analysis of secondary and tertiary structural conservation, however, similarity is found. The mature N-terminal 30 residues, for example, show a particularly conserved cysteine at position 15 to 17, which is flanked by nearby hydrophobic amino acids (Olafson et al., 1984). This cysteine is believed to be associated with intramolecular disulphide linkages, commonly with a second cysteine at approximately position 150 (reviewed by Turner, 1988). This coupling is overlapping with a second pair of cysteines around positions 120 and 190, which results in the formation of a molecule folded back on itself, in the form of an upright cyclindrical structure of approximately 100x40x40 Angstroms (Freymann et al., 1984). An X-ray crystallographic analysis revealed that the N-terminus of two unrelated VSG molecules, derived from trypanosomes from distinct serodemes, have remarkably similar tertiary structure despite an absence of N-terminal sequence homology (Metcalf et al., 1986). This overall structural conservation probably reflects the need for the VSGs to form a tightly packed coat on the parasite, resistant to penetration by antibodies against common surface antigens.

Within the C-terminal 100 or so amino acids, far more primary sequence conservation is present (Majumder et al., 1981; Boothroyd et al., 1981; Rice-Ficht et al., 1981), from which a

broad classification of VSGs has been devised based on the carboxy terminal residue (Borst and Cross, 1982). VSG molecules with a C- terminal asparagine or aspartic acid residue are classified Group I whilst those with serine are Group II. The relative conservation over the C-terminal domain and the gross sequence diversity over the N-terminal 350 amino acids reflects the relative function of each portion: The N-terminal domain accounts for the variable antigenicity of the molecule on living trypanosomes, while the C-terminal domain is associated with membrane anchoring.

During maturation, the VSG undergoes a number of post translational modifications. An N-terminal signal peptide of approximately 20-40 hydrophobic amino acids is cleaved off (Boothroyd et al., 1981; McConnell et al., 1981), as is a similar C-terminal sequence of approximately 23 amino acids (Boothroyd et al., 1980; Holder and Cross, 1981; Rice Ficht et al., 1981). The C-terminal domain of the molecule is also glycosylated, with one or more asparagine- linked high mannose type modifications, whose addition is tunicamycin sensitive (Holder and Cross, 1981; Strickler and Patton, 1980,1982a; Rovis and Dube, 1981). These glycosylations are believed to be associated with the stabilization of VSG dimers on the trypanosome surface (Strickler and Patton, 1982a; 1982b). Protein modification is also found at the extreme C-terminus of the mature VSG molecule. This modification contains mannose, galactose and glucosamine and is the region on the VSG to which antibodies with reactivity to several distinct variants are able to bind, at least on soluble form VSG molecules (Holder and Cross, 1981). This is called the cross reacting determinant (CRD) and is a part of the molecule concerned with anchoring to the trypanosome membrane. Instead of an insertion of a hydrophobic C-terminal domain into the surface membrane, the VSG has been found to be linked through a myristic acid containing glycolipid tail (Ferguson and Cross, 1984; Ferguson et al., 1985). This tail is added in a preassembled form (Bangs et al., 1985; Ferguson et al., 1986; Masterson et al., 1989) during VSG biogenesis to the residue at the mature C-terminus. The membrane anchor is able to be cleaved by the

endogenous phospholipase C or VSG lipase, releasing soluble form VSG (Cardoso de Almeida and Turner, 1983; Ferguson et al., 1985; Hereld et al., 1986; Fox et al., 1986) in a process which has been suggested to be involved with VSG turnover (Ferguson and Williams, 1988).

1.6 The Trypanosome Genome.

The trypanosome has a $7x10^4$kb diploid genome (Borst et al., 1982), the chromosomes of which do not condense at any point during the cell cycle. An examination of the chromosomes has, therefore, relied upon the use of pulsed field gel electrophoresis which allows a visualization of intact nuclear DNA (Schwartz and Cantor, 1984). This form of analysis involves subjecting chromosome-sized DNA molecules to an alternating electric field, with progression through the gel being thought to rely upon the rate with which the chromosomes can reorientate in the agarose matrix, prior to continued migration. This is associated with the length of the DNA molecule, such that chromosomes become resolved on the basis of size. Van der Ploeg et al.(1984a) first applied this technology to the trypanosome and resolved the genome into four distinct size classes. Under improved separation conditions, this has recently been increased to 20 resolvable bands, accounting for at least 80% of the trypanosome genome (Van der Ploeg et al., 1989). Most characteristic of trypanosomes able to undergo antigenic variation (Borst et al., 1984) are the 100 or so "minichromosomes" of between 50 and 150 kb in length (Williams et al., 1982; Sloof et al., 1983; Van der Ploeg et al., 1984a; 1984b), although these are absent in most T.vivax. Above these run five 200-430 kb chromosomes, nine chromosomes in the size range 680 kb to 3Mb and, finally, four chromosomes of 3-5.7 Mb which have only recently been resolved. Some material is retained in the slot; this may represent chromosomes still too large to be resolved and/or a non-specific trapping of DNA by the kinetoplast DNA which is unable to enter the gel matrix (Van der Ploeg et al., 1984a; 1989). VSG genes have been observed in all

the chromosome size classes that have been resolved, at least using early apparatus.

In addition to its karyotypic complexity, the trypanosome genome has been found to be very plastic, with frequent chromosome size changes being seen during both life in the bloodstream (Van der Ploeg et al., 1984b) and cyclical development (Le Page et al., 1988). Less extensive changes in chromosome length are also accountable for by the growth and periodic shortening of trypanosome telomeres (Bernards et al., 1983; Van der Ploeg et al., 1984c; Pays et al., 1983c). This 6-10bp growth/generation is likely to be a result of telomerase-mediated extension, as seen in Tetrahymena (Greider and Blackburn, 1989), although this would not account for the extensive occasional deletions seen during multiplication in the bloodstream and associated with trypanosome stress (Bernards et al., 1983; Bernards et al., 1984b; Pays et al., 1984).

1.7 VSG Genes and the Trypanosome Genome.

By hybridization, it has been estimated that the trypanosome possesses at least 1000 VSG genes and has the potential to encode even more by gene assembly (Van der Ploeg et al., 1982; reviewed by Pays, 1989). Many appear to be located at the very ends of chromosomes, with only telomeric repeats, or degenerate telomeric repeat-containing sequences, being found to their 3' side (Williams et al., 1982; De Lange and Borst, 1982; Van der Ploeg et al., 1984c). Like other trypanosome genes, none has been found to contain introns. Of the many VSG genes, usually only one is expressed at a time, excepting during VSG switching (Esser and Schoenbechler, 1985) and during in vitro culture (Baltz et al., 1986). Those active VSG genes are always located at telomeres, giving the trypanosome a maximum potential to express 200 or so distinct coats, if every telomere could drive VSG transcription. In fact, the number of sites with the potential to express VSG genes appears far smaller, being estimated at 20-25 in number (Cully et al., 1985). The question of how the trypanosome generates the potential to utilize its thousand or so genes

exclusively is answered by their ability to move VSG genes in to, and out of, expression sites by genomic rearrangement.

1.8 The Activation of VSG Genes.

VSG genes show a loosely hierarchical probability of expression, which appears in part to be a result of their activation mechanism (Liu et al., 1985; Aline et al., 1985b). The activation mechanism is the way in which a given VSG gene comes to be located in a transcribed expression locus, and these fall into two groups, dependent upon whether or not gene duplication is involved.

(A) Non Duplicative Activation.

Many of the genes encoding predominant VATs (i.e. those expressed early in infection) reside permanently at telomeres capable of VSG gene transcription and are switched on in situ (Laurent et al., 1984a; Liu et al., 1985). The presence of several expression sites in the trypanosome genome means that these genes are not always active, although how they are activated and inactivated is unknown (see section 1.11). Being activated without duplication and residing close to recombinogenic chromosome ends result in these genes being susceptible to deletion from the trypanosome genome (Laurent et al., 1984a, 1984b; Bernards et al., 1984a; Aline et al., 1989). This is advantageous, however; it may enable the trypanosome occasionally to present new sets of predominant VATs to hosts in the field which, by continual parasite challenge, may have become resistant to formerly dominant VATs.

An apparently far less frequent mechanism of non duplicative activation is telomere reciprocal translocation (Pays et al., 1985a). This involves an exchange of chromosome ends between silent and active VSG gene loci downstream of the promoter, resulting in a change in the expressed VAT but a preservation of both VSG coding sequences.

(B) Duplicative Activation.

The loss of a telomeric VSG gene from an expression locus is most commonly associated with its replacement by another telomeric VSG gene copy. This is termed telomere conversion (De Lange et al., 1983; Pays et al., 1983a, 1983b), and results in a coupled duplication of the incoming gene and destruction of the formerly expressed gene. The deletion of the telomeric VSG 221 gene, for example, has been seen to be associated with a duplication of the MiTat 1.8 gene (Bernards et al., 1984a), as has the conversion of AnTat 1.6, a predominant member of AnTAR 1, by AnTat 1.3 (Laurent et al., 1984a). These telomere conversion events are often very extensive, running from several kilobases upstream of the VSG gene to within the telomeric or subtelomeric repeats. In two cases the 5' limit of conversion has been found to be at least 40 Kb from the VSG gene (Pays et al., 1983d; Bernards et al., 1986a).

In addition to telomeric locations, many VSG genes are clustered within chromosomes (Van der Ploeg et al., 1982). These genes are incapable of being expressed unless they enter a telomeric expression site which is active or has the potential to be activated. In doing so it is a gene copy of the chromosome internal gene which invades the expression locus, during the process of gene conversion. This is called the expression linked extra copy (ELC) (Hoeijmakers et al., 1980; Pays et al., 1981) and its generation is termed duplicative transposition. The conversion mechanism appears to rely upon specific areas of homology between the basic copy gene and the expression locus. To the 5' side of VSG genes is a repeat element which is often associated with the upstream limit of the converted segment, being found in extensive arrays in telomeric expression sites. This region was first identified by Hoeijmakers et al. (1980) who observed a region free of restriction endonuclease recognition sites ("barren" region) extending 8kb upstream of the ELC for VSGs 117 and 118. Subsequently, the nature of these repeats has been determined (Liu et al., 1983; Campbell et al., 1984) and a

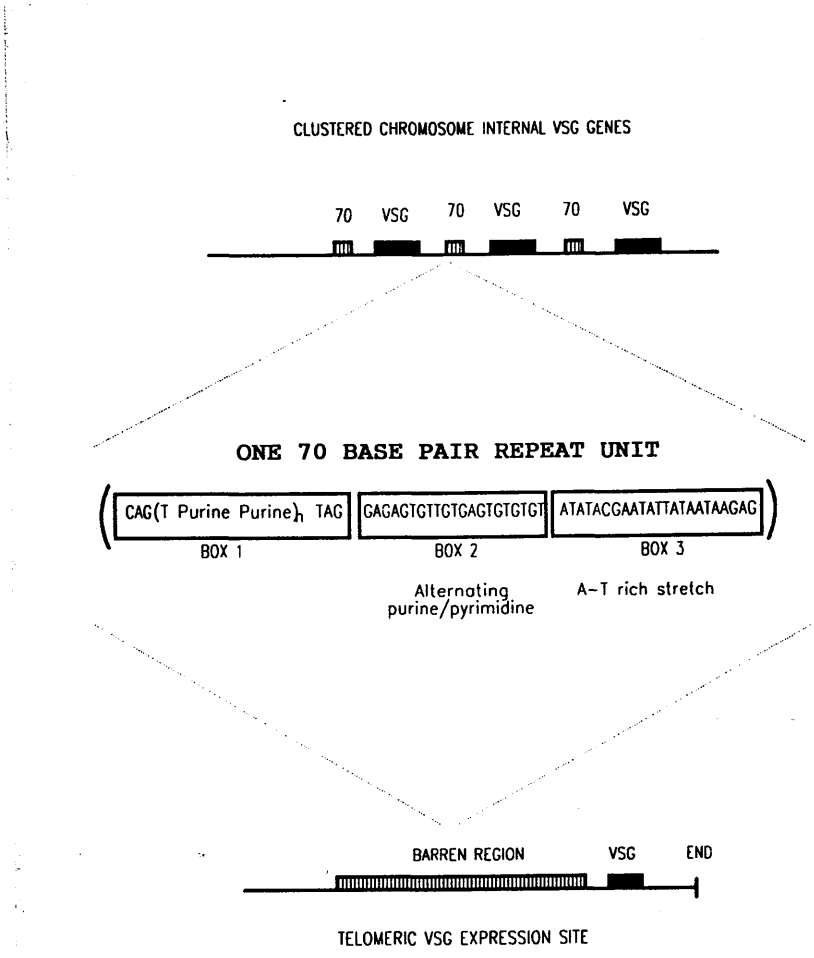


Figure 1.2; Consensus sequence for the 70bp repeat arrays found 5' to chromosome internal and telomeric VSG genes (Aline et al. 1985a). The boxes are described in the text.

Facing page 11

consensus derived (Aline et al., 1985a). This consensus repeat of approximately 70-76bp in length has been divided arbitrarily into 3 boxes (Aline et al., 1985a; Figure 1.2). The first contains a triplet repeat of T:Purine:Purine nucleotides which have been seen to extend for up to 90 repeats and are the principal source of length variation for each 70 bp repeat. These triplets were initially suggested to act as a polymerase reinitiation site because of their easy melting capacity and reiterated structure (Campbell et al., 1984), but there has been no evidence for this as yet (Shah et al., 1987). Another possibility is that they produce a potential for unusual helix conformations which may be recognised by either the trypanosome's recombination machinery or by specific factors analogous to the HO endonuclease involved in yeast mating type switching (Michiels et al., 1983; Liu et al., 1983). The potential for unusual DNA secondary structures is also found in boxes 2 and 3. Alternating purine:pyrimidine stretches may form left handed helices in box 2 and there is a possible capacity for the formation of cruciform structures between this region and the downstream AT-rich stretch in box 3. Finally, several workers have observed the high conservation of a TGTTG motif within box 2 (Liu et al., 1983; Campbell et al., 1984; Aline et al., 1985a; Florent et al., 1987; Shah et al., 1987). This shows similarity to the terminal repeats of eukaryotic transposable elements, with the Drosophila transposable element copia being most cited. It has also been noted that the TGTTG sequence is central to the yeast HO endonuclease cleavage site (Michiels et al., 1983; Pays, 1985).

To the 3' side of the VSG gene, the conversion breakpoint is less conserved between distinct switches. Some chromosome internal genes show extensive downstream homologies with the subtelomeric repeats of VSG expression loci and Aline and Stuart (1989) have used this to explain the relatively high frequency of activation of the chromosome internal VSG 118 gene. A more interesting observation has been the use of homologies within the 3' of the VSG gene itself, since this has the capacity to generate chimaeric genes by partial gene conversion (reviewed by Pays, 1989). Pays et al. (1985b) have followed what happens at

one expression locus during a number of antigen switches and observed that hybrid genes encoding antigenically distinct VSGs can result from interconversions between members of the closely related AnTat 1.1 gene family. In the case of conversions involving closely related sequences, the converted segment is often much smaller than that involved in duplicative transpositions between unrelated genes, with the start of the gene and sequences encoding the N/C terminal hinge region being favoured conversion endpoints. More recently, Thon et al. (1989) have reported that similar conversion events can allow the expression of a hybrid gene assembled from 3 distinct VSG pseudogenes, each of which is itself incapable of encoding a functional VSG. It is clear, therefore, how the reduction in the trypanosome's VSG gene repertoire by telomere conversion can be compensated by the rapid generation of new mosaic gene sequences.

1.9 VSG Gene Expression Loci.

All mechanisms of VSG gene activation, except in situ activation, act to bring a formerly silent gene into a site which is already being transcribed. This implies that the signals for VSG gene transcription are located upstream of the transposed/translocated segment (Bernards et al., 1985; De Lange et al., 1986). Since the transposed segment can be very long, this further suggests that the VSG gene promoter is located tens of kilobases upstream of the antigen gene itself. This region between the VSG gene and its promoter is called the expression site or expression locus and three have now been examined in some detail (Cully et al., 1985; Kooter et al., 1987; Shea et al., 1987; Pays et al., 1989a; Crozatier et al., submitted).

All expression loci appear to show an approximately similar architecture (see Figure 1.3). Upstream of the VSG gene, the barren region composed of 70 bp repeat arrays and, in the 221 expression site, a VSG pseudogene (Bernards et al., 1985), is found the first of several expression site associated genes (ESAGs). ESAG 1 was first identified by Cully et al. (1985) as a transcribed open reading frame with the potential to encode the

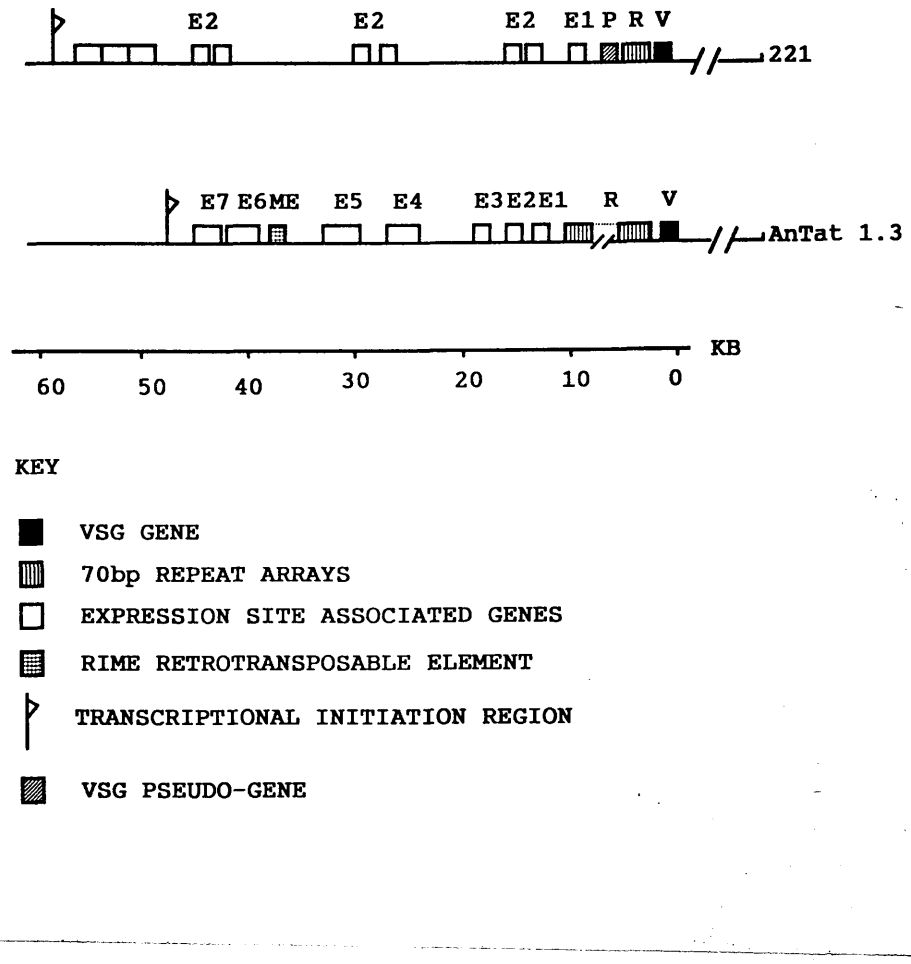


Figure 1.3; The structure of two bloodstream VSG gene expression loci (derived from Kooter et al., 1987 and Pays et al., 1989a). ME= RIME retrotransposable element, V=VSG gene, P=VSG pseudogene, R= 70bp repeat array. Open boxes represent identified open reading frames, designated expression site associated genes (ESAGs=E). Their relative numbering is in approximate order from the VSG gene. The VSG 221 expression locus contains duplications and triplications.

Facing page 13

production of a 39 KDa protein bearing an N-terminal signal sequence and C-terminal hydrophobic extension. Thus, the protein was predicted to be membrane-associated, and this has been confirmed subsequently (Cully et al., 1986). An analysis of genomic Southern blots has revealed that the ESAG 1 gene family is composed of 17 to 25 members and is diverse; the predicted proteins in the 117a and 221 expression loci show only 67% homology, even accounting for conservative amino acid changes. The positions of the several cysteine residues and N-linked glycosylation sites, however, appear relatively highly conserved (Cully et al., 1985; Alexandre et al., 1988; Son et al., 1989).

Beyond ESAG 1 are several more predicted ESAG sequences which have been numbered in their order from the VSG gene, although in the 221 expression site several regions containing these genes are duplicated or triplicated. The ESAG 2 protein, like ESAG 1, has been predicted to have the characteristics of membrane association (Kooter et al., 1988) and is of unknown function. ESAG 2, however, does appear to be associated frequently with telomere conversion events; in several cases homologies between ESAG 2 sequences in expression sites have marked the 5' conversion end point (Kooter et al., 1988; Shea et al., 1987; Pays et al., 1983a)

Recently, several other expression site sequences have been characterized in some detail (Pays et al., 1989a). This has revealed that one of the ESAGs in the AnTat 1.3 expression site (ESAG 4) shows similarity to adenylate cyclase. Also present was a retrotransposable element, RIME (Ribosomal mobile element, Hasan et al., 1984), first identified as being associated with trypanosome ribosomal genes. Large stretches of this expression locus also showed extensive homology (nearly 90%) in non-coding regions with another, distinct VSG gene expression locus (Shah et al., 1987). The very 5' end of this expression locus also contained an open reading frame which is very highly conserved among all analysed expression loci. This sequence, variously designated ESAG X, 10 or 7, appears closely linked with the VSG gene expression site promoter (Kooter et al., 1987; Pays et al., 1989a; Zomerdijk et al., submitted).

1.10 VSG Gene Expression Site Transcription.

A search for the promoter for the VSG gene expression site is complicated by the phenomenon of discontinuous transcription (reviewed in Borst, 1986). The 5' end of probably all trypanosome mRNAs possess a 39 nucleotide sequence which is not coded contiguously with the bulk of the transcript. This is termed the "mini exon" or spliced leader (SL) sequence. The mini exon RNA sequence originates as a 140 nucleoside precursor mini exon derived (med) RNA, transcribed from tandem arrays of 1.35kb genes by what appears to be RNA polymerase III (Kooter et al., 1984; Grondal et al., 1989). It is then apparently joined onto gene transcripts by trans splicing, utilizing classical cis splicing signals and resulting in the production of the fused RNA molecule and a Y shaped open lariat end product (Sutton and Boothroyd, 1986; Murphy et al., 1986). Because the spliced leader sequence prevents the identification of transcription initiation regions by normal approaches (S1 or primer extension analysis of mRNA), alternative procedures have been employed, involving an analysis of nascent transcripts. The most extensively used have been run-on transcriptional assays (Kooter and Borst 1984; Kooter et al., 1984). These involve the isolation of trypanosome nuclei in such a way that transcribing polymerases are paused on their DNA template. The polymerases are then permitted to resume transcription ("run-on") in the presence of labelled UTP, thereby producing radioactively labelled nascent transcripts. The hybridization of these transcripts to restriction digests of cloned regions of an expression site immobilized onto nitrocellulose membrane can indicate the location of the promoter, since only areas downstream of the transcriptional initiation point will be transcribed and, therefore, detected after autoradiography. In these assays, the nuclei are prepared very rapidly to prevent shutdown of VSG gene transcription associated with stress (Kooter and Borst, 1984; Bernards et al., 1985; Kooter et al., 1987), and transcript elongation carried out for a short time to limit primary transcript processing.

Four VSG gene expression loci have now been analysed using run-on transcriptional assays (Bernards et al., 1985; Kooter et al., 1987; Alexandre et al., 1988; Pays et al., 1989a; Shea et al., 1987; Gibbs and Cross, 1988). Two of these, for VSG 221 and AnTat 1.3, display transcriptional products hybridizing throughout 56 and 45kb of their expression sites, respectively. For the third expression site, however, an initial analysis suggested that the transcription unit was relatively short and began within the DNA segment that VSG 118 duplicated when transposed to its expression site (Shea et al., 1987). An analysis of the transcriptional start site for this gene revealed that it showed similarity to the trypanosome ribosomal gene promoter, which is used by RNA polymerase I. In support of this, transcription of the 118 VSG expression site, like all others, was found to be relatively insensitive to the drug alpha amanitin, a potent inhibitor of RNA polymerase II and III activity (Kooter and Borst, 1984). This prompted Shea et al. (1987) to propose a model in which RNA polymerase I was responsible for VSG transcription, with telomere activation/inactivation being mediated by a competition for nucleolar binding sites. This has been rendered unlikely, however, by a re-examination of this expression locus using ultra violet light-induced transcriptional inactivation. This technique, first applied to a VSG expression locus by Johnson et al. (1987), uses the ability of UV light to induce the formation of pyrimidine dimers in DNA (Sauerbier and Herculez, 1978). When encountered by a polymerase, these dimers cause a disengagement from the DNA template, resulting in transcriptional termination. At U.V. doses which produce few dimers, short transcription units are relatively resistant to inhibition because the polymerase has a low probability of encountering a pyrimidine dimer, while long transcription units are relatively sensitive because the chance of the transcription being terminated before completion is high. By a calibration of the system with transcription units of known length, it was estimated that the 221 and AnTat 1.3 transcription units were of a length compatible with run-on transcriptional analyses. With the 118 VSG gene expression site,

however, these two analyses were not compatible: U.V inactivation data placed the promoter 35 kb upstream of the VSG gene, outside the 118 transposed segment. It has now been shown that the proposed pol I promoter was entirely dependent for its activity on the upstream promoter, and was most probably detected by an artifact of the run-on assay (Crozatier et al., submitted). Similar apparent transcriptional gaps were detected by Alexandre et al. (1988). Furthermore, despite its alpha-amanitin insensitivity, the polymerase responsible for VSG gene expression site transcription does not appear to be polymerase I. Grondal et al. (1989) have carried out a detailed biochemical analysis of the RNA polymerases of trypanosomes and suggest that it is a modified, alpha-amanitin resistant, polymerase II which transcribes VSG gene expression sites. This is supported by the identification of a novel RNA polymerase II gene in trypanosomes, which may be unique to those that undergo antigenic variation (Evers et al., 1989)

It is therefore now apparent that all analysed VSG gene expression loci appear similar: there is a promoter, a huge polycistronic transcription unit containing several ESAGs, a long restriction site barren region, the VSG gene and the telomere.

1.11 Expression Site Activation Mechanisms

Several hypotheses have been suggested to explain how the trypanosome is able to maintain transcription of only one of its several expression loci (reviewed in Borst et al., 1990; Borst and Greaves, 1987; Pays and Steinert, 1988). The simplest, a single mobile promoter element able to hop between expression sites, now appears unlikely because two active expression loci have been observed in single trypanosomes (Cornelissen et al., 1985a; Baltz et al., 1986). In the former, two expression sites were transcriptionally active simultaneously but resulted in the expression of only one, because the other showed a truncation of the primary transcript caused by a long insertion between the promoter and the VSG gene. In the latter, cultured trypanosomes

were seen to be displaying a mixed coat stably, implying two active loci, although there is a small possibility that one expression site could contain two tandemly linked VSG genes, like the 221 VSG and associated pseudogene. Multiple coats have never been observed *in vivo*, except during antigen switching.

Another possibility is a DNA rearrangment, analogous to promoter inversion in *Salmonella* phase variation, occurring at low frequency (reviewed in Borst and Greaves, 1987). To date, however, no such genomic rearrangement has been detected which is associated with expression site activation or inactivation (Majiwa *et al.*,1982; Penncavage *et al.*, 1983; Laurent *et al.*, 1984a; Bernards *et al.*, 1984a; Myler *et al.*, 1984; Zomerdijk *et al.*, submitted).

An alternative potential mechanism of expression site control is DNA modification at inactive telomeres. This came to light when it was observed that regions proximal to VSG genes show only partial cleavage when subjected to digestion with either *Pst* I or *Pvu*II (Bernards *et al.*, 1984b; Pays *et al.*, 1984). This inhibition was found to be absent from transcribed telomeres, to decrease distally to the telomere and to be more pronounced where the extent of telomere repeats downstream of the VSG gene was greater. On this basis, Bernards *et al.* (1984b) proposed that nucleotide modification may regulate activity of the distant upstream promoter by altering the chromatin conformation of the telomere, or by altering the telomere's capacity to bind an activating site on the nuclear matrix (Van der Ploeg and Cornelissen, 1984). This model was extended by Greaves and Borst (1987) who used various indicators of chromatin structure to show that the active expression site was sensitive to enzymes recognising single stranded DNA. This was not a consequence of polymerases tracking along the template; the highly transcribed mini-exon genes showed similar sensitivity through both their transcribed and non-transcribed regions. Instead, it was proposed that active VSG expression sites are held under torsional stress by being bound to the nuclear matrix, and that this could act to stimulate promoter activity at a distance. In this case, they suggested, modified telomeres would be free and would remain

inactive unless they happened to become unmodified and, thus, able to bind the activating site(s). This demodification could be achieved through gene conversion or by exchanging chromosome tips (Pays and Steinert, 1988).

The nature of the nucleotide modification associated with inactive telomeres is unknown, being insufficiently abundant to permit analysis (Crozatier et al., 1988), but it has been suggested to be mediated by an enzyme which recognises telomere repeats and tracks internally, modifying a number of its potential sites.

1.12 Metacyclic VSGs.

The VSG expression system the trypanosome activates in the tsetse fly salivary gland, prior to mammal infection, appears somewhat different from that used by established bloodstream parasites. It was initially thought that just a single metacyclic variable antigen type (M-VAT) existed in the fly, from which all other variants followed (Gray, 1965). Subsequently, however, it was established that trypanosomes in the fly saliva showed VAT heterogeneity (Le Ray et al., 1978; Barry et al., 1979; Hajduk et al., 1981). The possibility that this resulted from rapid switching away from a basic antigen during the parasite's time in the salivary gland lumen has now been discounted; immunogold labelling of nascent metacyclics still attached to the gland wall showed that they were also heterogeneous (Tetley et al., 1987). This study also showed that VAT activation was not preprogrammed in the attached dividing stages prior to metacyclogenesis since neighbouring metacyclics were not clumped into colonies of particular M-VATs.

The number of distinct VATs produced by metacyclic trypanosomes is far more limited than the potential several hundred produced in the bloodstream. Monoclonal antibodies against T.congolense metacyclic cells established that there were 12 distinct types (Crowe et al., 1983) and a similar analysis showed that the M-VAT repertoire in T.b.rhodesiense was no more than 27 (Turner et al., 1988). These studies showed that, as

well being limited, the M-VAT repertoire was far more predictable than the bloodstream VATs, since all M-VATs from a number of distinct transmissions could be eliminated consistently using the same pool of antibodies. The M-VAT repertoire is not predictable in the long term, however, but is continually evolving. An analysis of the M-VATs from trypanosomes isolated over a 17 year period revealed that whilst some were relatively stable, others were less so; antigen types were seen to be lost and gained in the repertoire with time. The potential for a rapid generation of variability in the field has also been directly observed in the laboratory: two M-VATs were seen to be rapidly lost from repertoire, one of which disappeared after only 3 rounds of cyclical transmission and did not subsequently reappear (Barry et al., 1983)

The number of metacyclic cells obtainable from tsetse flies is insufficient for a direct molecular analysis and there is, as yet, no T.brucei culture system in which infective metacyclic organisms develop. A study of metacyclic VSG gene expression is possible, however, for two reasons. Firstly, metacyclic VATs continue to be expressed for the first 5 to 7 days in the mammalian bloodstream before switching over to the expression of predominant bloodstream VATs (Hajduk and Vickerman, 1981; Esser and Schoenbechler, 1985). Secondly, metacyclic VSG genes can be re-expressed during chronic bloodstream infection (Barry et al., 1979; Cornelissen et al., 1985a; Son et al., 1989), although at this stage they appear to be doing so using a distinct bloodstream specific mechanism, which is independent of that used in metacyclic cells (Turner et al., 1986). These two features have allowed the isolation of metacyclic VSG cDNAs and a subsequent analysis of the genomic environment of metacyclic VSG genes.

To date, seven M-VSG genes have been studied, at least to some extent, and all have been found to be located at a special type of genomic site. They are always found on the telomeres of the chromosomes which were not resolved by using early forms of pulsed field gel electrophoresis apparatus (Lenardo et al., 1986, Cornelissen et al., 1985a; Delauw et al., 1987). Metacyclic VSG

genes have been reported either to be activated in situ (Lenardo et al., 1986) or by the generation of an ELC (Delauw et al., 1987). Each study, however, has used a different indirect approach necessitated by the sparsity of metacyclic material. In one case, trypanosomes from many infected mice were pooled, enriched for particular M-VATs by immune lysis with antibodies against other M-VATs, and their DNA analysed (Lenardo et al., 1986). This indicated that M-VSG genes were expressed without duplicative transposition, although the possibility of these non-clonal populations containing multiple distinct ELCs could not be ruled out by Southern analysis. In another case, clonal trypanosome populations were used but they were collected from mice 10-12 days after leaving the fly and were unlikely to be maintaining metacyclic specific M-VSG gene expression (Delauw et al., 1987). In the most recent study, using antigenically stable trypanosome lines which retain fly transmissibility (Shiels et al., submitted; Matthews et al., 1990) two predominant M-VSG genes have been found to be activated apparently without duplication. This is further detailed in section 3.1.

The putative expression sites for four M-VSG genes have now been examined, in part. In no case has any form of barren region been found 5' to M-VSG genes and, for two, quantitative Southern analysis indicated that there was no 70 bp repeat hybridizing sequence 5' to the VSG gene (Lenardo et al., 1986). From this, it was suggested by these workers that M-VAT genes were transcribed without duplication because they were unable to escape from their locus, lacking the signals for transposition. Further upstream, Son et al. (1989) have demonstrated the existence of a gene homologous to the ESAG 1 in two putative M-VAT expression loci, although on one, this ESAG was located an unusually great distance upstream.

In no case has a reliable transcriptional analysis of a metacyclic VSG gene expression site been performed; the non clonal nature of the approach of Lenardo et al. (1984; 1986) makes the interpretation of nuclear run-on data inconclusive, and this form of analysis was not carried out be Delauw et al. (1987).

1.13 Aims.

It was the aim of the work presented here to examine the activation of metacyclic VSG genes in Trypanosoma brucei rhodesiense. This entailed;

(a) An examination of the mechanism of metacyclic VSG gene activation in the tsetse fly.

(b) A characterization of a metacyclic VSG gene expression locus, structurally and transcriptionally.

(c) An investigation of the re-activation of metacyclic VSG genes in bloodstream infection.

Chapter 2

MATERIALS AND METHODS

2.1 Animals.

BKTO mice were purchased from Bantin and Kingman Ltd, and maintained at the Institute of Virology, Glasgow University. Glossina morsitans morsitans were supplied as teneral males or as pupae by Dr I.Maudlin at the Tsetse Research Laboratory, Bristol.

2.2 Trypanosomes

A virulent cloned line of Trypanosoma brucei rhodesiense EATRO 795 was used in the majority of experiments (Turner and Barry, 1989). The trypanosomes were originally stabilated from the blood of a cow at Uhembo, Central Nyanza Province, Kenya in 1964. Stabilates GUP 810 and 871, which are described in Chapter 4 of this thesis, were gifts from Dr C.M.R.Turner, at the Protozoology Unit, Glasgow University.

2.3 List of Materials

MATERIAL	SOURCE
General chemicals and organic solvents.	B.D.H., Hopkins and Williams, Koch-Light Laboratories, May and Baker.
Media.	Davis, Oxoid.
Agar.	Davis, Difco.
Biochemicals.	Sigma.
Antibiotics.	Sigma.
Agarose.	Sigma, I.B.I.
Radiochemicals.	New England Nuclear.
Hybond-N Nylon membrane.	Amersham.
Nytran Nylon membrane.	Schleicher and Schuell.
Diethyl aminoethyl (DEAE) Cellulose DE-52	Whatman.

Citrated Sheeps' Blood.	Becton and Dickinson.
Geneclean.	Stratatech.
Circleprep Kit.	Stratatech.
Hypnorm.	Janssen.
Hypnovel.	Roche.

All enzymes were obtained from B.R.L. except the following;

Sequenase with Sequencing Kit.	U.S.B.
Nae I	Boehringer Mannheim.

2.4 Bacterial Strains.

The bacterial strains used were derivatives of Escherichia coli K-12.

Name	Genotype	Source
DH5 alpha	F^-, 80dlac M15, (lacZYA-argF)U169, recA1, endA1, hsdR17 (r_k^-,m_k^-), supE44, thi-1, gyrA, relA1.	Hanahan (1983)
HB101	F',hsdS20(r_B^-,m_B^-), supE44, ara14, galK2, lacY1,proA2, rpsL20(Sm^r), xyl5, leu, mtl1, recA13	Boyer and Roulland-Dussoix (1969)
DS941	recF143,proA7,Str31, thr1, leu6, tsx33, mtL1, his44, argE3, lacY1, galK2, ara14, lambda$^-$, $lacI^q$,lacZ M15, $lacY^+$. supE44, xyl15.	D.Sherratt

DS902	As DS941, but recA13, sup37	D.Sherratt
JM101	lac-pro, thi, supE, F', traD36, proAB, $lacI^qZ$ M15.	Messing (1983)
NM621	hsdR, mcrA, mcrB, mcrD, recD, 1009 thy^+.	Whittaker et al. (1988)
P2392	$recA^+$ (P2) hsdR514, (r_K^-, m_K^-) Sup E44, lac Y1, gal K2, galT22 metB1, trpR55	Edgerton, M.D. Unpublished data. Stratagene.

2.5 Plasmids and Bacteriophages

Plasmids and bacteriophage other than those constructed as part of this thesis are listed here. For reference, the restriction sites present in the polylinker of the plasmid vectors are given in Appendix 1.

(1) Plasmids

pUC19. pBR322 derived cloning vector. (Yanisch-Perron et al., 1985)

pMTL23. pUC/pBR322 derived cloning vector, with a more extensive polylinker sequence than pUC 18/19 (Gift from A.Bednarz)

pBluescript. phagemid cloning and sequencing vector (Short et al., 1988).

(2) Bacteriophage.

M13mp18/19 Single stranded virion, used in producing template for DNA sequencing following cloning of exogenous DNA into replicative form genome. (Messing, 1983)

Lambda Dash Replacement vector, with a T3 and T7 promoter flanking the cloning sites. Accepts inserts of 9-23 Kb (Stratagene).

2.6 E.Coli Culture Media

L-Broth: 10g tryptone, 5g yeast extract, 5g NaCl, 1g glucose, 20mg thiamine, made up to 1 litre in distilled water and adjusted to pH 7.0 with NaOH. Solid L-Agar was as L-broth, but with the addition of 12g per liter of No.3 Oxoid agar. For phage work, a supplement of 10mM $MgSO_4$ and 0.2% maltose (final concentration) was added to L-broth.

XIA Agar: 400ml L-Agar supplemented with 212ul of 50 mg/ml X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactoside), 300ul of 40mg/ml IPTG (isopropylthio-beta-D-galactoside) and 400ul of 100mg/ml ampicillin (sodium salt).

BBL Agar: 10g trypticase peptone (BBL 11921), 5g NaCl, made up to 1 litre with distilled water, adjusted to pH 7.2 with NaOH and solidified with the addition of 10g Taiyo agar.

BBL Agarose Overlay: As for BBL agar but with the addition of 2.5g $MgSO_4.6H_2O$ before solidification with 6.5g agarose (type 1 low EEO A6013).

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

M9 salts: 6g Na_2HPO_4, 3g KH_2PO_4, 1g NH_4Cl, 0.5g NaCl made up to 1 litre with distilled water.

M13 Soft Overlay: 6g Taiyo agar made up to 1 litre with distilled water. For one 9cm plate, add 25ul 40mg/ml X-gal and 20ul of 50mg/ml IPTG to 4ml of molten overlay.

2xYT medium: 16g Bacto tryptone, 10g yeast extract and 5g NaCl made up to 1 litre with distilled water.

S.O.C transformation recovery medium: 2g Bactotryptone, 0.5g Yeast extract, 1ml of 1M NaCl, 0.25ml of 1M KCl added to 97ml of distilled water and autoclaved. Supplement with 1ml of 1M $MgCl_2$/1M $MgSO_4$ and 1ml 2M glucose. Filter sterilize.

2.7 Buffer Solutions

Electrophoresis.

10x TBE Buffer: 109g Tris, 55g boric acid, 9.3g $Na_2EDTA.2H_2O$, made up to 1 litre in distilled water; pH is 8.3.

10x TAE Buffer: 48.4g Tris, 3.6g CH_3COONa, 3.6g $Na_2EDTA.2H_2O$, made up to 1 litre in distilled water, pH adjusted to 8.2 with glacial acetic acid.

Single Colony Gel Buffer: 2% Ficoll, 1% SDS, 0.01% bromophenol blue, 0.01% orange G, in 1x TAE buffer.

Agarose Gel Loading Buffer: 10% Ficoll, 0.5% SDS, 0.06% bromophenol blue, 0.06% orange G, in 1x TBE buffer.

Sequencing Gel Loading Buffer: 95% formamide, 20mM EDTA, 0.05% bromophenol blue and 0.5% Xylene cyanol FF.

Acrylamide Sequencing Gel stock: 190g acrylamide, 10g N,N' methylene bis-acrylamide made up to 500 ml with distilled water.

Sequencing gel mix: 42g Urea, 10ml of 10x TBE buffer, 14.5 ml of acrylamide sequencing gel stock, 40.5 ml distilled water. Filtered through 2 sheets of filter paper (Whatman Number 1).

DNA Manipulation.

Restriction and Ligation Buffers: Obtained from B.R.L.

dNTP Stock Solution (100mM): Dissolve 60mg dNTP in 0.8ml distilled water, adjust pH to 7.0 with 0.1M NaOH, make up to 1.0 ml with water; store at $-20^{o}C$.

TE Buffer: 10mM Tris, 1mM EDTA, pH to 8.0.

TE (10:0.1) buffer: 10mM Tris, 0.1mM EDTA, pH to 8.0.

Hin x10 buffer: 100mM Tris-HCl pH 7.5, 100mM $MgCl_2$.

S1 Nuclease Buffer: 30mM CH_3COONa pH 4.6, 50mM NaCl, 1mM $ZnSO_4$, 5% w/v glycerol.

DNA Sequencing Buffers

5x Annealing buffer: 200mM Tris-HCl, 100mM $MgCl_2$, 250mM NaCl.

dGTP labelling mix: 7.5 uM each of dGTP, dCTP and dTTP.

Termination mixes: All are 80mM each of dGTP, dCTP, dTTP, dATP; with, for the following;

Mix G: 8uM ddGTP, 50mM NaCl.
Mix C: 8uM ddCTP, 50mM NaCl.
Mix T: 8uM ddTTP, 50mM NaCl.
Mix A: 8uM ddATP, 50mM NaCl.

Enzyme Dilution Buffer: 10mM Tris-HCl pH 7.5, 5mM DTT (dithiothreitol), 0.5mg/ml BSA (bovine serum albumin).

Hybridization Solutions.

20xSSC: 3M NaCl, 300mM tri-sodium citrate, pH to 7.0.

50x Denhardt's Solution: 5g Ficoll, 5g Polyvinylpyrollidine, 5g BSA, distilled water to 500ml.

Prehybridization Solution: 6xSSC, 5xDenhardt's solution, 0.5% SDS, 10ug/ml denatured herring sperm DNA.

Denaturing Solution: 1.5M NaCl, 0.5M NaOH.

Neutralizing Solution: 1.5M Tris, 1mM EDTA, pH to 7.2.

Radiolabelled Probe preparation solutions

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

Solution A: 1ml solution theta, 18ul beta-mercaptoethanol (14.3M), 5ul 100mM dCTP, 5ul 100mM dGTP, 5ul 100mM dTTP.

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

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

Reaction Mix-A: 20ul Solution A, 50ul Solution B, 30ul Solution C. Store at -20^{o}C.

DNA Extraction, Purification and General Purpose Solutions.

Phenol: All phenol used in the purification of DNA was redistilled and contained 0.1% 8-hydroxyquinoline. Phenol was buffered against 1M Tris pH 8.0 and then stored under TE buffer.

Phenol/Chloroform: Prepared as a 50% v/v mixture.

SM Buffer: Used for bacteriophage storage and dilution:- 5.8g NaCl, 2g $MgSO_4$, 2% gelatin, 1mM Tris, pH to 7.5 then made up to 1 litre with distilled water.

Birnboim and Doly Buffer I: 50mM glucose, 25mM Tris, 10mM EDTA, pH to 8.0.

Birnboim and Doly Buffer II: 0.2M NaOH, 1% SDS, prepared immediately prior to use.

Birnboim and Doly Buffer III: 5M CH_3COOK pH 4.8; mix equal volumes of 3M CH_3COOK and CH_3COOH, pH should be 4.8.

Geneclean NEW buffer: 0.2M Tris-HCl pH 7.5, 1M NaCl, 20mM EDTA in 25ml; dilute with 225ml of dH_2O and 250ml of ethanol before use.

Trypanosome DNA extraction solutions

10x PBS: 134.8g Na_2HPO_4, 7.8g NaH_2PO_4, 42.5g NaCl , pH to 7.5 and make up to 1 litre with distilled water.

PSG (6:4): 0.488g $NaH_2PO_4.2H_2O$, 2.55g NaCl, 8.08g Na_2HPO_4 and 15g D-glucose made up to 1 litre with distilled water and adjusted to pH 8.0.

NET Buffer: 100mM NaCl, 100mM EDTA and 10mM Tris in distilled water.

Pulse field gel electrophoresis lysis buffer: 0.5M EDTA pH 9.5, 1% sodium N-laurylsarcosinate, prepared in distilled water with Proteinase K added to 0.25 mg/ml.

Run-on transcription buffers

2xElongation Buffer:

Stock solution	Volume used	Final concentration
1M Tris pH8.0	100ul	100mM
5M NaCl	10ul	50mM
4M KCl	12.5ul	100mM
1M $MgCl_2$	2ul	2mM
1M $MnCl_2$	4mM	4mM
1M DTT	2ul	2mM
0.15M Spermine	1ul	0.15mM
0.1M Spermidine	5ul	0.5mM
1M Creatine phosphate	10ul	10mM
0.1M GTP	20ul	2mM
0.1M CTP	20ul	2mM
0.1M ATP	20ul	2mM
Glycerol	250ul	25%
RNase Free dH_2O	43ul	

Run-on hybridization solution: 3xSSC, 0.1% SDS, 10xDenhardt's solution, 20mM sodium phosphate, 100ug/ml denatured herring sperm DNA and 50ug/ml E.coli tRNA.

2.8 E.coli Growth Conditions

Liquid cultures for transformation or plasmid and phage DNA preparations were routinely grown in L-Broth at 37°C with vigorous shaking. Growth on plates was normally carried out on L-Agar with antibiotics being added as required. For plating out of lambda phage, phage particles were mixed with plating cells grown in 10mM $MgSO_4$, 0.2% maltose and stored in 10mM $MgSO_4$. The phage were then incubated at 37°C to allow adsorption. This suspension was then added to 3.0ml of BBL agarose overlay (at 50°C) and poured immediately over a 9cm diameter petri dish containing BBL agar. After the soft agar had set, plates were inverted and incubated overnight at 37°C. For plating lambda libraries, 30,000 p.f.u. were adsorbed to 1ml of plating bacteria, added to 30ml of BBL agarose overlay (at 50°C) and then poured over 20cm^2 BBL agar plates. M13 transformations were plated by adding 4ml of M13 soft overlay agar and pouring over L-Agar or minimal medium+glucose plates.

2.9 Transformation of E.coli

Competent Cells

A single bacterial colony was picked into 2.5ml of L-Broth and shaken vigorously at 37°C overnight. 500ul of this culture was used to inoculate 100ml of L-Broth and the culture shaken at 37°C for approximately 2 hours, or until an $O.D_{600}$ of 0.3-0.4 had been reached. At this point, cells were transferred into 50ml Falcon tubes and harvested by centrifugation at 4360g for 10 minutes at 4°C. The resultant supernate was discarded and the pellet resuspended by flicking the tubes, after which 25ml of

cold 50mM $CaCl_2$ was added to each. After storage on ice for 30 to 60 minutes, the cells were harvested as above, and were resuspended in 2.5 ml of 50mM $CaCl_2$ at 4^oC. These cells were then stored at 4^oC until use. The cells were found to be maximally competent for transformation after 24 hours on ice, after which their competence rapidly fell.

Transformation

For each transformation, 100-200ul of competent cells were added to the transforming DNA (at a concentration of less than 0.1mg per transformation, and in a volume no more than 1/5 of the cell suspension volume) in a 15 ml Falcon tube. After mixing by gentle shaking, the cells were incubated at 4^oC for 30 minutes and heat shocked at 42^oC for 2 minutes. The cells were then returned to ice for a further 2 minutes for plasmid transformations, or were kept at room temperature for 10 minutes for M13 transformations. M13 transformations were then immediately added to M13 soft overlay and poured over MM+G plates. Plasmid transformations had 1ml of S.O.C. transformation recovery medium added and were then incubated at 37^oC for 30 to 60 minutes to allow the expression of antibiotic resistance. The cells were then plated in 100-200ul amounts over XIA plates.

2.10 Plasmid Preparation.

The plasmid content of E.coli transformants was routinely screened through a combination of single colony gel electrophoresis and restriction analysis following plasmid isolation.

Single Colony Gel Electrophoresis

This is a rapid procedure for isolating circular plasmid molecules directly from a transformed colony. A single colony was streaked to produce a bacterial patch approximately 1cm in length following overnight growth under antibiotic selection. The patch

was then scraped with a sterile toothpick, which was placed into 200ul of single colony gel buffer and left at room temperature for 15 minutes. Cell debris were then removed by centrifugation in a microcentifuge for 30 minutes at $4^{o}C$, and 50ul of the supernate analysed by electrophoresis on an agarose gel.

Birnboim and Doly Small Scale Plasmid Preparation Procedure

During the cloning of genomic DNA fragments into plasmids, it is convenient to perform small scale plasmid preparations which can confirm, by restriction analysis, that the plasmid insert contains restriction sites which have been predicted from the mapping of genomic DNA. To do this, a single transformed colony was picked into 1.5ml of L-Broth containing appropriate antibiotics and grown overnight. The cells were then transferred to a microcentrifuge tube and harvested by centrifugation. The supernate was then removed completely, the cell pellet resuspended in 100ul of Birnboim and Doly solution I and left at room temperature for 5 minutes. Then, 200ul of freshly prepared Birnboim and Doly solution II was added, the solution mixed by inversion and stood at room temperature for 5 minutes. Finally, 150ul of Birnboim and Doly solution III was added and the tube placed on ice for 10 minutes. After this time, the bacterial debris was removed by centrifugation in a microcentrifuge, and the plasmid DNA in the supernate precipitated by the addition of 1ml of absolute ethanol, which had been stored at $-20^{o}C$. The DNA was recovered by centrifugation in a microcentrifuge for 10 minutes, the pellet washed with 70% ethanol and dried at $65^{o}C$. The pellet was resuspended in 50ul of TE, to yield between 2 and 10ug of plasmid DNA.

Birnboim and Doly Large Scale Plasmid Preparation Procedure.

A modification of the alkaline-SDS extraction of Birnboim and Doly (1979) was used for large scale plasmid DNA preparation.

Plasmid bearing cells from a 250ml overnight culture were pelleted by centrifugation at 5520g for 5 minutes at 4°C. The pellet was resuspended in 4ml of Birnboim and Doly solution I and incubated on ice for 5 minutes. 8ml of freshly prepared Birnboim and Doly solution II was added and mixed by gentle inversion; the lysate was then left on ice for 4 minutes. 6ml of cooled Birnboim and Doly solution III was added and again gently mixed until the viscosity was reduced. The white floccular precipitate was centrifugated at 17400g at 4°C for 15 minutes and the supernate poured through a filter with a glass wool plug into a tube containing 12 ml of isopropanol. The tubes were mixed by inversion and then DNA allowed to precipitate at room temperature for 15 minutes before being centrifuged 17400g, 10 minutes at 20°C. The pellet was then washed with 2ml of 70% ethanol, dried under vacuum and resuspended in 4.6ml of TE by gentle rolling for 30 to 60 minutes.

To the resuspended pellet was added 8.64ml of CsCl solution in TE (50g CsCl added to 30ml TE) and 0.54 ml of 15mg/ml ethidium bromide solution. This solution was cleared by centrifugation at 4360g for 10 minutes and then transferred to a 12ml polypropylene ultracentrifuge tube and centrifuged at 267000g at 20°C for 16 hours. The resulting plasmid band was removed with a hypodermic syringe and multiple extractions with water saturated butanol carried out to remove ethidium bromide. Three volumes of TE were then added to the solution, followed by two volumes of absolute ethanol stored at -20°C. The tubes were then left on ice for 1 hour. Finally, the DNA was precipitated by centrifugation at 35000g, 4°C, washed with 70% ethanol, dried and resuspended in 500ul TE buffer.

During the latter stages of this study, the above method was replaced by the Stratatech "Circleprep" procedure. This involves a Birnboim and Doly alkaline lysis of cells grown overnight in 50 ml of "Circlegrow" medium, after which plasmid DNA is purified from RNA by using lithium chloride precipitation and Glassmilk. The methodology followed was exactly as described by the manufacturers of the kit.

2.11 Bacteriophage Lambda DNA Preparation.

Bacteriophage plaques were picked into 1ml of SM and left at $4^{o}C$ overnight or $37^{o}C$ for 1 hour. 5ul of this phage suspension was incubated for 20 minutes at $37^{o}C$ with 100ul of NM621 stored in 10mM $MgSO_4$ to allow the phage to adsorb and then added to 5ml of supplemented L-Broth and grown overnight with vigorous shaking at $37^{o}C$. Cellular debris was then removed by centrifugation at 12100g for 10 minutes.

For large scale phage DNA preparation, 5 ml of an overnight NM621 culture were inoculated into 250 ml of L-broth supplemented with magnesium and maltose and were grown with shaking for 3 hours. The bacteria were then collected by centrifugation at 9820g for 10 minutes at room temperature and were resuspended in 250 ml of L-broth supplemented with magnesium, but not maltose. To this was then added the 5 ml phage supernate and the culture grown until full lysis had occurred (usually >8 hours). After this time, 1 ml of chloroform was added to the culture which was then shaken for a further 15 minutes. Unlysed cells and cellular debris were removed by centrifugation (9820g, 10 minutes), after which DNAse I and RNAse A were each added to the supernate to a final concentration of 1 ug/ml and incubated at $37^{o}C$ for 1 hour. Solid sodium chloride was then added to a concentration of 1 M, dissolved by swirling, and the supernate left on ice for 1 hour, following which the solution was centrifuged at 5520g for 15 minutes. To the supernate was added polyethylene glycol 8000 to a concentration of 10% w/v and the phage precipitated at $4^{o}C$ for 1 hour. The phage were then pelleted by centrifugation at 5520g for 30 minutes at $4^{o}C$ and the supernate fully drained off and discarded. The dried pellet was resuspended in 6ml of SM solution, transferred to a 40 ml polypropylene tube and extracted with an equal volume of chloroform. Following centrifugation at 5520g for 15 minutes, the aqueous phase was

recovered and EDTA (to a final concentration of 20mM), proteinase K (to 50ug/ml) and SDS (to 0.5%) added. Protein digestion was continued for 1 hour at 50°C. The solution was then extracted with an equal volume of phenol, phenol/chloroform and chloroform, the aqueous phase being separated at each stage by centrifugation at 12100g for 10 minutes at room temperature. Finally, 1/10 volume 3M sodium acetate was added to the aqueous phase and the phage DNA precipitated with two volumes of ethanol for 30 minutes on ice, pelleted by centrifugation at 12100g for 10 minutes at 4°C, washed with 10 ml of 70% ethanol, dried and resuspended in 500ul of TE. Recovery was quantitated by gel electrophoresis.

2.12 Bacteriophage M13 Single Stranded DNA Preparation.

To prepare M13 single stranded DNA as a template for DNA sequencing, a modification of the method of Miller (1987) was used.

A single phage plaque was picked into 20ml of 2xYT medium containing 200ul of an overnight culture of JM101 and grown, with vigorous shaking, at 37°C for 5 to 6 hours. The bacteria were then pelleted at 12100g, 10 minutes. To the supernatent was added 4ml of 2.5M NaCl, 20% PEG and this was left at room temperature for 10 minutes. The bacteriophage were then pelleted by centrifugation at 35000g, 10 minutes, 4°C after which the supernate was thoroughly drained. The phage were then resuspended in 1ml of TE (10:0.1) and reprecipitated by the addition of 200ul of 2.5M NaCl, 20% PEG, standing at room temperature for 10 minutes and centrifugation in a microfuge for 10 minutes. The supernate was drawn off completely and resuspended in 110ul of TE (10:0.1). The phage DNA was then recovered by extraction with an equal volume of phenol, chloroform and ethanol precipitation. The dried DNA pellet was finally brought up in 50-100ul TE (10:0.1)

2.13 Extraction of High Molecular Weight DNA from Bloodstream Form Trypanosomes (Bernards *et al*. 1981).

Approximately 2ml of trypanosome infected mouse blood was

collected by cardiac puncture through a 25G needle into a 2ml syringe containing 0.4ml of 2% sodium citrate solution to inhibit coagulation. The trypanosomes were then purified on a PSG pH 8.05 equilibrated DEAE-Cellulose column at least 5 times the volume of the infected blood (Lanham and Godfrey, 1970). After pelletting by centrifugation at 3020g, the supernate was removed and the parasites resuspended in 1/10 their packed cell volume in NET buffer ($1x10^9$ trypanosomes have a 100ul packed cell volume). The parasite suspension was then lysed by the addition of 1/10 volume 30% sodium N-lauryl sarcosinate in NET buffer, followed by gentle inversion. Proteinase K was added to a final concentration of 100ug/ml, and incubated at 50^oC for 1 hour, or at room temperature overnight. After this time, NET buffer was added to a volume of 3ml and the lysate extracted with an equal volume of phenol/chloroform, followed by centifugation at 3020g for 5 minutes. This step was then repeated two times. Following this, the crude DNA extract was precipitated by the addition of 1/10 volume 3M sodium acetate and two volumes of absolute ethanol. The DNA was recovered by spooling, washed in 70% ethanol, and resuspended in 500ul of TE overnight. To this was then added RNAse A to 50ug/ml and the DNA incubated at 37^oC for 1 hour, after which proteinase K was added to a concentration of 100ug/ml and incubated at 50^oC for a further 30 minutes. Following this, the DNA was extracted three times with phenol/chloroform, ethanol precipitated, resuspended in 500ul of TE and dialysed against 2 litres of TE for 48 hours in a Sartorius microcollodion cup. The DNA recovery was quantitated by gel electrophoresis.

2.14 Restriction Endonuclease Digestion of DNA

Plasmid and bacteriophage DNA digestion reactions were routinely performed in a total volume of 20ul containing 0.2 to 2 ug of DNA, 2ul of the 10x concentrate reaction buffer recommended by the endonuclease supplier and 10 to 20 units of the appropriate enzyme. The reaction volume was made up to 20ul with ddH_2O, and incubated at the appropriate temperature for 1-2

hours. For trypanosome genomic DNA, 1-3ug were digested in a total reaction volume of 40ul for 3-16 hours, the enzyme being added in two 15-20 unit amounts at the beginning and half way through the incubation period. Serial digestions involving enzymes with different buffer optima were carried out sequentially, with either incubation at 65^{o}C or phenol extraction and ethanol precipitation being used to inactivate the first enzyme prior to adjustment of the buffer conditions.

2.15 Agarose Gel Electrophoresis

Genomic, phage and plasmid DNA restriction digests were separated by agarose gel electrophoresis. The agarose concentration in the gel was adjusted to suit the predicted size of fragments generated by digestion. DNA bands were visualised by soaking the gel after electrophoresis in 1x TBE containing 10ul of 10mg/ml ethidium bromide per 100ml of buffer for 15 minutes. The gel was then rinsed in 1xTBE without ethidium bromide and photographed with 203nm U.V. transillumination using a Polaroid camera loaded with Polaroid 4x5 land film (no.57) or a Pentax 35mm SLR loaded with Ilford HP5 film; both were fitted with a Kodak Wratten Filter no.9 (red).

2.16 DNA Extraction From Agarose.

Two methods were used for the isolation of DNA; "Gene-clean" and extraction from low melting point agarose. The "Gene-clean" procedure was followed according to the manufacturers instructions (Stratagene): digested DNA was electrophoresed on an agarose gel run in TAE buffer and the desired fragments excised. The gel slice was then dissolved in 2.5 volumes of 6M sodium iodide at 55^{o}C for 10 minutes, after which 3-5ul of "glassmilk" was added and the DNA allowed to adsorb for 5 minutes on ice. The "glassmilk" was washed 3 times with NEW buffer before the DNA was recovered by adding 50-100ul of TE, incubating at 55^{o}C for 10 minutes and finally centrifuging in a microfuge for 30 seconds. The DNA in solution was pipetted to a fresh tube, being careful

to avoid transferring any "glassmilk". For library construction, the solution was again incubated at 55°C and recovered by centrifugation to ensure all "glassmilk" had been removed. This was found to be essential when very limiting amounts of DNA were involved.

For the extraction from low melting temperature agarose, DNA was electrophoresed in TBE or TAE buffer, and the desired fragment excised. Five volumes of TE were added and the gel slice melted at 65°C for 10 minutes. After this time, the DNA was extracted twice with phenol, once with chloroform, and then recovered by ethanol precipitation and resuspension in 50ul TE.

2.17. Phosphatase Treatment of DNA and Subsequent Ligation

The treatment of vector DNA with calf intestinal alkaline phosphatase (CIAP) results in the removal of the 5' terminal phosphate group and prevents self ligation, thereby favouring the insertion of foreign DNA. To achieve this, 1 unit of CIAP was normally incubated with vector DNA during the last 30 minutes of its digestion with restriction enzymes. The enzyme activity was then removed by phenol and chloroform extraction followed by ethanol precipitation. The inability of the vector to ligate was normally assayed by ligation followed by agarose gel electrophoresis of the reaction products.

DNA fragments with compatible ends were ligated using T4 DNA ligase. In order to favour the generation of chimaeric molecules, a vector to insert ratio of 5:1 was used for phosphatased vector DNA, while a ratio of 1:5 was used where the vector DNA was unphosphatased. Typically, a total of 200ng of DNA was included in a 20ul reaction volume containing 4ul of 5xBRL ligation buffer. For DNA fragments with cohesive ends, 0.5u of T4 DNA ligase was used and the reaction incubated at 16°C overnight. For ligation of blunt ended molecules 1 unit of ligase was used, and the reaction allowed to proceed at room temperature for at least 4 hours.

Following the ligation of phage arms to insert, during library construction, an *in vitro* packaging reaction was carried

out using Gigapack Gold (Stratagene). This entailed adding 5ul of the ligation mix to a Freeze/Thaw mix, adding 15ul of Sonic extract and leaving the tubes at room temperature for two hours. 500ul of SM buffer was then added, and the tubes stored at $4^{o}C$ with 15ul of chloroform. The library was then transfected as described (section 2.8).

2.18 Nucleic Acid Hybridization

Radiolabelled Probe Preparation

Probes were prepared by the hexanucleotide priming method of Feinberg and Vogelstein (1985). A DNA restriction fragment to be labelled was excised from low melting temperature agarose and either used immediately or stored at $-20^{o}C$. The gel slice was melted at $65^{o}C$ for 10 minutes and then 10ul (approximately 10ng) removed. This 10ul was then placed in a fresh eppendorf tube and incubated in a boiling water bath for 10 minutes. The tube was then spun briefly in a microfuge and 28ul of distilled water and 10ul of Reaction mix-A added. The reaction was then begun by the addition of 30uCi of [alpha-^{32}] dATP (800 Ci/mMole) and 2 units of the Klenow fragment of DNA polymerase I (either sequencing or labelling grade). After incubation for 2-16 hours at room temperature, the reaction was heated to $65^{o}C$ for 5 minutes and the labelled fragments isolated through a Pharmacia G-50 Sephadex "Nick" column equilibrated and eluted in TE.

Southern Blotting

Following electrophoresis, agarose gels containing DNA fragments larger than 8Kb were soaked in 0.25M HCl for 30 minutes in order to partially depurinate the DNA. For gels with only small DNA fragments this step was omitted. The gel was then transferred into denaturing solution and soaked with occasional agitation for 30 minutes, after which the gel was soaked in neutralizing solution for a further 30 minutes. The gel was then inverted onto a strip of filter paper (Whatman 3MM) overlaying a

perspex plate supported above a reservoir of 20xSSC which soaked both ends of the filter. Onto the gel was overlaid a sheet of Nytran (Schleicher and Schuell) or Hybond-N (Amersham) nylon hybridization membrane and a further two sheets of 20xSSC soaked filter paper, each of which had been cut to the size of the gel. Finally, a stack of paper towels or two nappy pads were placed over the gel followed by a glass plate and a 700g weight. DNA transfer was allowed to continue overnight, after which the nylon membrane was dried at 80^{o}C, and subjected to U.V. irradiation for 2.5 minutes in order to covalently attach the DNA to the membrane. Blots were stored at room temperature, wrapped in filter paper.

Southern Hybridization

Southern hybridization was performed according to the procedures detailed by Schleicher and Schuell for use with Nytran membrane. Filters were sealed into either plastic bags or glass cylinders in the presence of 3-10ml of prehybridization solution, depending on the size of the blot. Prehybridization was then carried out for 1 hour or more at the desired hybridization temperature (usually 65^{o}C) in either a water bath or a Bachofer hybridization oven. After this time, prehybridization solution was replaced with fresh solution and the radiolabelled probe. The probe had been previously denatured together with 100ul of 10mg/ml herring sperm DNA by boiling for 10 minutes. The hybridization was continued for at least 16 hours at the appropriate temperature. After hybridization, the probe was discarded and the membrane washed twice for 30 minutes in 2xSSC, 0.1% SDS at the hybridization temperature, and then to the final stringency, which was usually 0.1xSSC, 0.1% SDS, 65^{o}C for a further 30 minutes. The membrane was then sealed into a plastic bag and covered with a sheet of Kodak X-omat S1 film which was allowed to expose in presence of two intensifying screeens at -70^{o}C.

Colony and Plaque Screening by Hybridization.

To determine whether individual bacterial colonies harboured plasmids containing desired inserts, Nytran membrane disks were laid over their agar plate and allowed to wet naturally. After one minute, the membrane was lifted off with forceps and placed, colony side up, onto a sheet of filter paper (Whatman 3MM) soaked in denaturing solution for 10 minutes. The disk was then transferred for 5 minutes each onto two filter sheets soaked in neutralizing solution. Finally the disk was rinsed in 2xSSC and dried at 80^{o}C and the DNA fixed by U.V. transillumination for 2 minutes. Plaque lifts were treated identically, with the exception that denaturation was only carried out for 5 minutes. Fixed filters were hybridized exactly in the same manner as Southern blots.

2.19 DNA Sequencing

DNA sequencing was performed using the T7 polymerase (Tabor and Richardson, 1987) "Sequenase" sequencing system according to the manufacturers' instructions (USB).

7ul of M13 template DNA, 2ul of 5x Annealing buffer and 1ul of 0.5pmol/ul M13 universal primer were mixed and incubated at 65^{o}C in a Grant small heating block for 2 minutes. The block was then turned off and allowed to cool slowly to below 35^{o}C to permit efficient primer:template annealing. To the annealing mixture was then added 1ul dithiothreitol (0.1M), 2ul dGTP labelling mix diluted to 1:15, 0.5ul [a-^{35}S]dATP (1000Ci/mmol), 2ul of Sequenase enzyme prediluted 1:8 in Enzyme Dilution Buffer and the mixture left at room temperature for 5 minutes. 3.5 ul of this mix was then added to 2.5ul of dideoxy NTP termination mixes in separate tubes and the reaction incubated at 37^{o}C for 10 minutes. The reactions were then stopped by the addition of 4 ul sequencing gel loading buffer and either stored at -20^{o}C for up to 2 weeks or immediately denatured by heating to 85^{o}C for three

minutes and loaded onto a sequencing gel.

Sequencing Gel Electrophoresis

Sequencing reaction products were electrophoresed on 0.4-0.8mm 6% Acrylamide: bis-acrylamide (38:2), 6M Urea wedge gel prepared in 1xTBE buffer. A 100ml gel mix was polymerized by the addition of 800ul of 10% w/v ammonium persulphate and 40ul of TEMED (N,N,N',N'-tetramethylethylenediamine). Gels were run using the BRL S2 sequencing gel apparatus. Prior to addition of the sequencing products, the gel was pre-run at 65 watts for 1 hour to warm the gel to 50^{o}C. Once at temperature, the gel was loaded and run at 65 watts for 3 hours, for separation of 1-200 nucleotides, or 6-10 hours for the separation of fragments 100-500 nucleotides in length. At the completion of a run, the gel was fixed in 10% acetic acid 10% methanol for 15 minutes and then dried at 80^{o}C under vacuum using a Bio-Rad model 583 gel drier. The gel was then covered with Kodak S1 X-omat film and exposed at room temperature overnight.

2.20 Cloning of progressive intermediates in DNA synthesis.

This procedure is capable of producing deletion clones of existing single stranded recombinant M13 DNA suitable for DNA sequencing or radiolabelled probe construction (Burton et al., 1988). A diagrammatic representation of the principles of the method are given in Figure 2.1. 2ug of single stranded recombinant M13 DNA was added to 19ul of distilled water, 4 ul of Hinx10 buffer and 2ul of M13 universal oligonucleotide sequencing primer (USB). The primer and template were then annealed by placing the eppendorf tube in a Grant small heating block at 65^{o}C for 2 minutes, which was then switched off and allowed to cool to below 35^{o}C.

After annealing, 1 unit of the Klenow fragment of DNA polymerase 1 was then added to the reaction tube and the mixture incubated at room temperature for 2 minutes after which 8ul of a

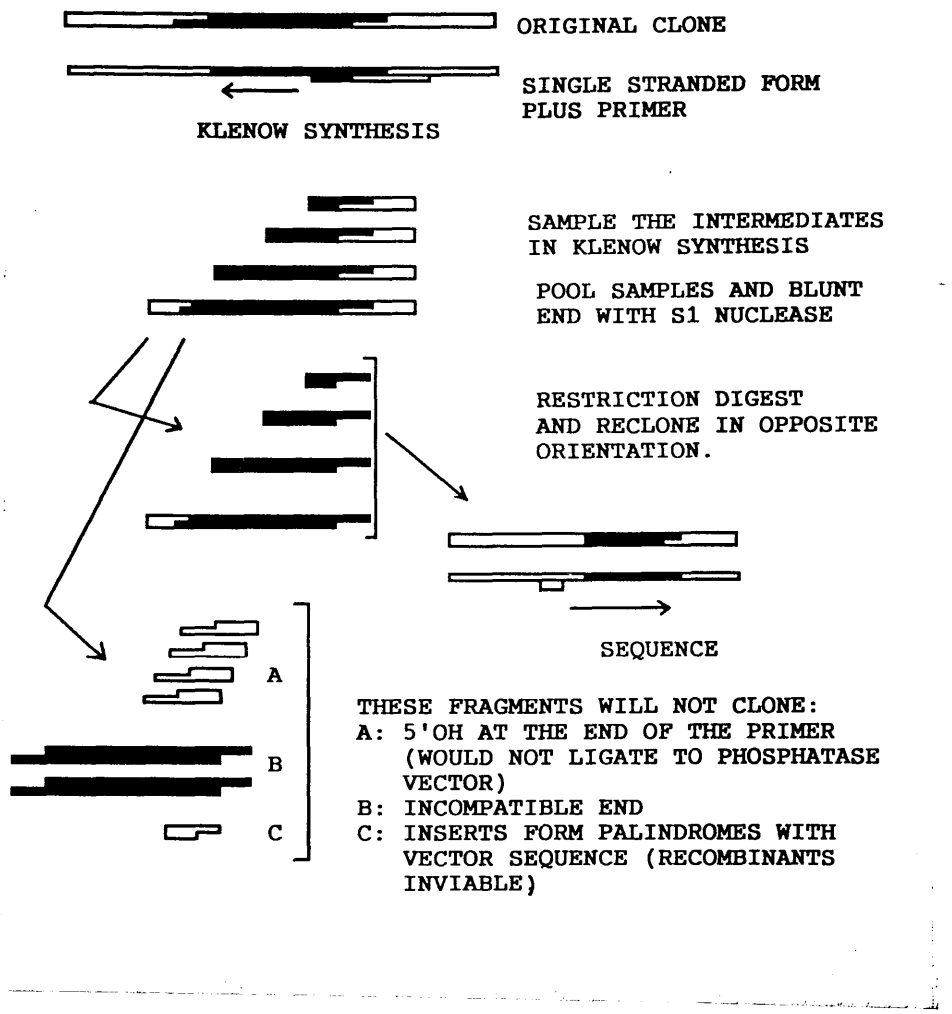


Figure 2.1; Procedure for the cloning of intermediates in M13 complementary strand synthesis (adapted from Burton et al., 1988). Black segments represent the insert to be subcloned, open segments represent M13 vector sequence. Double strand DNA is shown as double thickness, single stranded DNA as single thickness. Note that the original and final clones are circular molecules, other intermediates are linear.

Facing page 45

nucleotide mix was added (0.25mM dGTP, 0.25mM dATP, 0.25mM dCTP, 0.25mM dTTP). 10ul fractions were then withdrawn at 45 second intervals into 160ul of distilled water in a dry ice ethanol bath. This solution was then thawed and incubated at 70°C for 15 minutes following which the DNA was precipitated by the addition of 20 ul of 3M sodium acetate and 600ul of ethanol. The mixture was incubated in a dry ice ethanol bath for 30 minutes. DNA was then recovered by centifugation in an eppendorf microcentrifuge for 30 minutes at 4°C, washed in 70% ethanol and dried at 65°C. Once dried the DNA was resuspended in 50ul of S1 nuclease buffer, 10 units of S1 nuclease added and the mixture incubated at 37°C for 30 minutes. This acts to trim away all sections of the initial single stranded template DNA, leaving only the variously sized double stranded DNA products of the extension reaction. The reaction was terminated by snap freezing in a dry ice ethanol bath, followed by phenol and chloroform extraction and ethanol precipitation. The DNA was then resupended in 50ul of dH_2O. From this, 5ul was withdrawn and stored, so that it could be used as a control during transformation: transformants from this material either result from incomplete S1 nuclease digestion or the presence of double stranded replicative form M13 DNA contaminating the original single stranded template DNA preparation. To the remaining 45ul was added 5ul of 10x restriction endonuclease buffer and 10 units of an appropriate restriction enzyme, which cleaves between the primer hybridization site on the original M13 template and the start of the original insert producing cohesive ends. This results in molecules with one cohesive end (nearest the original primer site) and one blunt end, resulting from S1 cleavage (at the limit of the extension reaction).

Digestion was allowed to continue for 4 hours at 37°C, after which the DNA was recovered by ethanol precipitation and resuspended in 15ul of distilled water. This DNA was then ligated to M13 replicative form digested such that there was blunt end nearest to the primer hybridization site and a more distal cohesive end compatible with that on the insert DNA. This results in molecules of M13 DNA with inserts of variable sizes in which

the heterogenous end is nearest the primer hybridizaion site. The ligated DNA was then tranformed into E.coli JM101, as was the 5ul of DNA stored after S1 digestion. Single stranded DNA was then prepared from the resulting plaques and the insert size determined by DNA sequencing.

2.21 Tsetse Fly Infection and Maintenance.

Adult Glossina m.morsitans were supplied in groups of 50-100 flies. Immediately upon delivery, flies were chilled at 4^{o}C until they were stunned (approximately 20 minutes). They were then transferred, whilst still at 4^{o}C, into individual fly tubes (plastic universal tubes with an open end covered by a linen grill). They were then placed in an incubator maintained at 28^{o}C, >80% relative humidity for 4 hours. Prior to infection, a 1ml sample of Trypanosoma brucei rhodesiense EATRO 795 stored in liquid nitrogen was thawed quickly at 37^{o}C. To this was added 3ml of sterile PSG buffer in a dropwise manner, and the infected blood centrifuged at 1830g for 5 minutes to pellet the trypanosomes, after which the supernate was discarded. The pellet was gently resuspended, the parasites counted using an improved Neubauer haemocytometer and adjusted to 1-5x10^{6}/ml in citrated sheeps blood. This blood was then poured onto a 41^{o}C hotplate and covered by a sterile silicone membrane onto which the tsetse flies were placed to allow them to feed. Those tubes containing flies which did not feed quickly were tapped onto the bench and replaced onto the membrane. This encouraged the flies to feed, as did executing the procedure in darkness. As flies were seen to feed, their tubes were labelled and they were replaced in the incubator.

Following infection, flies were given the opportunity to feed as above, but on uninfected sterile citrated sheeps' blood, on every Monday, Wednesday and Friday.

2.22 Examination of flies for the presence of metacyclic trypanosomes (Hajduk et al. 1981).

Twenty five days after infection the fly tubes were placed on microscope slides maintained at 41°C and left for 20 minutes. The slides were then examined for the presence of tsetse fly salivary probing. Those slides which contained saliva were marked to note which fly had probed, and then examined at x400 magnification using phase contrast microscopy for metacyclic and proventricular form trypanosomes, which were distinguished morphologically. Those tubes containing flies producing metacyclic parasites were marked, and the flies allowed to feed on 30g BKTO mice anaesthetised intraperitoneally with 0.25ml of a mixture of 1 part Hypnorm, 1 part Hypnovel and 2 parts water.

2.23 Trypanosome Cloning

Three days after fly bite, mouse tail blood was collected into a Hawkins haematocrit capillary tube, centrifuged in a haematocrit centrifuge for 2 minutes, and the buffy coat examined for the presence of trypanosomes using an inverted microscope. If present, the tube was snapped at the boundary between the red blood cells and the buffy coat, and the buffy coat flushed into a microcentrifuge tube. A paper clip was then used to transfer three microdroplets to individual wells in a humidified Terasaki plate (Lux). These were then examined for the presence of a single motile trypanosome. When discovered, 20ul of Guinea Pig serum was added to the well and the complete contents removed and placed into 150ul of Guinea Pig serum. This was then injected intraperitoneally into a mouse. Further droplets were then produced and examined for the presence of further trypanosome clones.

The parasitaemia was monitored from day 5 after infection using the rapid comparison method of Herbert and Lumsden (1976) and trypanosomes collected usually on day 6-7 by cardiac puncture.

2.24 Immunofluorescence

The variable antigen type expressed by bloodstream trypanosomes was determined by the indirect immune fluorescence antibody test (IFAT). A blood smear on a microscope slide was prepared, allowed to air dry and was fixed in acetone for five minutes. The slide was then allowed to air dry for 5 minutes, after which two 0.5cm diameter circles were painted onto the slide using a Texpen ball point paint marker (Mark-Tex Corporation, New Jersey) and dried for 30 minutes on the bench. To the circles were then added either 30ul of mouse anti-ILTat 1.61 ascites fluid diluted 1:500 in PBS, or undiluted mouse anti-ILTat 1.22 hybridoma cell culture supernate, both kindly donated by Dr C.M.R Turner at Glasgow University Protozoology Unit. This antibody was allowed to adsorb for 30 minutes in a humid chamber, after which the slides were washed extensively with PBS. Excess PBS was then removed with a tissue, avoiding the marked areas. To the marked circles was then added 50ul of a 1:50 dilution of rabbit anti-mouse IgG-fluorescein isothiocyanate conjugate (Sigma) containing 0.25ul/100ul of 10mg/ml ethidium bromide. The latter was included to allow a visualization of the nucleus and kinetoplast of unstained trypanosomes, which permitted a calculation of the number of labelled trypanosomes as a percentage of total trypanosomes. After incubation with the second antibody for 30 minutes in the humid chamber, the slide was again washed extensively with PBS, excess PBS removed, and drops of 50% Glycerol:PBS added to the marked areas. A cover slip was then applied, and the slide viewed with Ploem illumination using the FITC filter cube on a Leitz Orthoplan largefield microscope.

Immunofluorescence upon fly salivary probes was carried out exactly as above, with the position of groups of metacyclic organisms being marked with a diamond pencil prior to acetone fixing.

2.25 Cryopreservation of bloodstream trypanosomes.

Trypanosomes in infected blood were cryopreserved using a procedure communicated to us by Dr I.Maudlin of the Tsetse Research Laboratory, Bristol. An infected mouse with a parasitaemia of between $2.5x10^8$ and $1x10^9$ trypanosomes/ml was exsanguinated by cardiac puncture into a syringe containing 0.4ml of 2% sodium citrate solution as an anticoagulant. To the blood was then added, dropwise, an equal volume of a 14% glycerol solution in PSG and the blood allowed to stand at room temperature for 5 to 10 minutes. After this time, the blood was dispensed in 0.25-0.5 ml volumes into sterile 1 ml screw cap plastic vials (Nunc). These were wrapped in cotton wool, placed in a polystyrene box and frozen overnight at -70^oC before being transferred to liquid nitrogen for long term storage. For the preparation of stabilates for future mammal infection the PSG:Glycerol equilibrated blood was introduced into a 60 cm length of 0.63 mm inside diameter translucent vinyl tubing (Portex). This was then cut into 2cm lengths, transferred into a 1 ml screw cap plastic vial and frozen in the same way as above.

2.26 Isolation of intact trypanosome nuclei and run-on transcription

Trypanosome nuclei were isolated from the blood of infected mice as soon as a sufficient parasitaemia had developed. For infection with either metacyclic clones or trypanosome clones derived 2-3 days after fly bite, this was typically 6-7 days post inoculation. Mice were exsanguinated by cardiac puncture and 1.75 ml of the collected blood introduced into a Stansted cell disrupter, yielding intact trypanosome nuclei, which were cryopreserved. From the remaining 0.25 ml of infected blood, trypanosomes were isolated over a DEAE cellulose column and high molecular weight DNA prepared. These procedures were routinely carried out by Dr S.Graham, exactly according to the methodology of Kooter et al. (1984).

To perform run-on transcription, $1x10^9$ cryopreserved nuclei

were thawed, and the sample split into halves. To one half was added 20ul of 10mg/ml alpha amanitin in methanol, the solution mixed thoroughly and nuclei incubated on ice for ten minutes. The remaining half was kept on ice. Both fractions were then centrifuged in a microcentrifuge for 10 minutes and treated identically in all following steps. The supernate was removed and the nuclei resuspended in 50 ul of 2x elongation buffer. To this was then added 20ul of dH_2O, 2ul of RNase inhibitor and 30ul of [a-^{32}P] dUTP (3000 Ci/mmole) and the reaction allowed to proceed at 37^{o}C for 5 minutes, after which it was stopped by placing the tube at 70^{o}C for 5 minutes. The reaction was then cooled to 37^{o} for 5 minutes after which RNase-free DNase 1 added to a final concentration of 10ug/ml and the tube incubated at 37^{o}C for a further 5 minutes. Following this, the reaction mixture was adjusted to 10mM Tris-HCl pH 8.0, 4mM EDTA pH8.0, 1% SDS and 10ug/ml proteinase K and incubated at 37 ^{o}C for 5 minutes. To act as a carrier in precipitation procedures, E.coli tRNA was added to a 100ug/ml final concentration and the reaction mixture extracted with phenol/chloroform saturated with NET buffer. The organic phase was then back extracted with an equal volume of NET buffer and the aqueous phase pooled with the aqueous phase from the first extraction. This was then adjusted to 0.3M sodium acetate, 2.5 volumes of ethanol added and the RNA allowed to precipitate on ice for 30 minutes. After this time, the RNA was pelletted by centrifugation for 20 minutes in a microfuge, the ethanol removed and the pellet drained and resuspended in 100ul TE. Run-on transcription products were then separated from unincorporated nucleotides over a TE equilibrated Pharmacia G-50 sephadex "Nick" column. Labelled nascent RNA was then hybridized with appropriate Southern blots in 1ml of run-on hybridization solution for a minimum of 48 hours at 60^{o}C.

2.27 Intact Chromosome Preparation for Pulsed Field Gel Electrophoresis.

The isolation of intact trypanosme chromosomal DNA and its subsequent separation by pulsed field gel electrophoresis (PFGE)

was carried out using the methods of Bernards et al.(1986a) and Van der Ploeg et al. (1989).

Blood was collected from a trypanosome infected mouse by cardiac puncture and the parasites isolated over a DEAE cellulose column. The trypanosomes were counted in an improved Neubauer haemocytometer, and adjusted to $1x10^8$ organisms/ml with PSG buffer. This was incubated at 37^oC for 5 minutes to warm the suspension and then an equal volume of 1% low melting temperature agarose in PSG buffer at 37^oC added. This was transferred to ten 100ul sample holders on ice. Once the blocks had solidified, they were transferred into 40 ml of lysis buffer and deproteinized at 50^oC for 18 to 48 hours. Blocks were then stored in lysis buffer at 4^oC.

Prior to electrophoresis, blocks were equilibrated in 0.5x TBE, and then half of a block inserted into a well in a preformed $20cm^2$ 1% agarose gel. The block was sealed in position with molten 1% agarose, excess agarose trimmed away, and the gel run on a Biometra Rotaphor pulsed field gel electrophoresis system under the conditions specified in the text.

Chapter 3

Metacyclic VSG gene activation in the tsetse fly

3.1 Introduction

When the bloodstream VSG system is reactivated in trypanosomes which have emerged from a tsetse fly, there can be a preferential activation of the VAT that was being expressed immediately before cyclical transmission. This is called the ingested VAT (I-VAT) "memory" effect (Hajduk et al., 1981; De Lauw et al., 1985; Turner et al., 1986). One explanation for this effect was suggested by observations upon the reactivation of certain VSG genes when trypanosomes are syringe passaged from one mammal host to another. If a VSG gene was activated by duplicative transposition, but then silenced by an inactivation of its expression site, it was found to be left lingering in its now dormant telomeric location (Laurent et al., 1984b; Buck et al., 1984a,b; Michels et al., 1984). With passage to a new host, this "lingering ex-ELC" was found to be re-expressed with a higher than normal frequency, presumably because it had bypassed the need for transposition prior to expression (Laurent et al., 1984b). A similar system could operate with passage through the tsetse fly: in this case the genes would be left lingering because all expression sites are inactivated on entering the fly (Overath et al., 1983). These genes would then be re-expressed with an elevated probability upon re-entering the bloodstream, being already resident in an expression site with the potential for activation.

More recently, another mechanism for the ingested VAT effect has been suggested which is not restricted to genes normally activated by duplication (Pays et al., 1989b). By the inhibition of nascent transcript processing using U.V. light, it was found that the formerly active VSG expression site promoter was not fully inactivated in the tsetse fly. Instead, VSG expression seemed to be repressed by primary transcript truncation close to the promoter. When alleviated by re-entry into bloodstream mode VSG expression, this might result in a preferential re-expression of the ingested VAT. If so, this observation both provides a

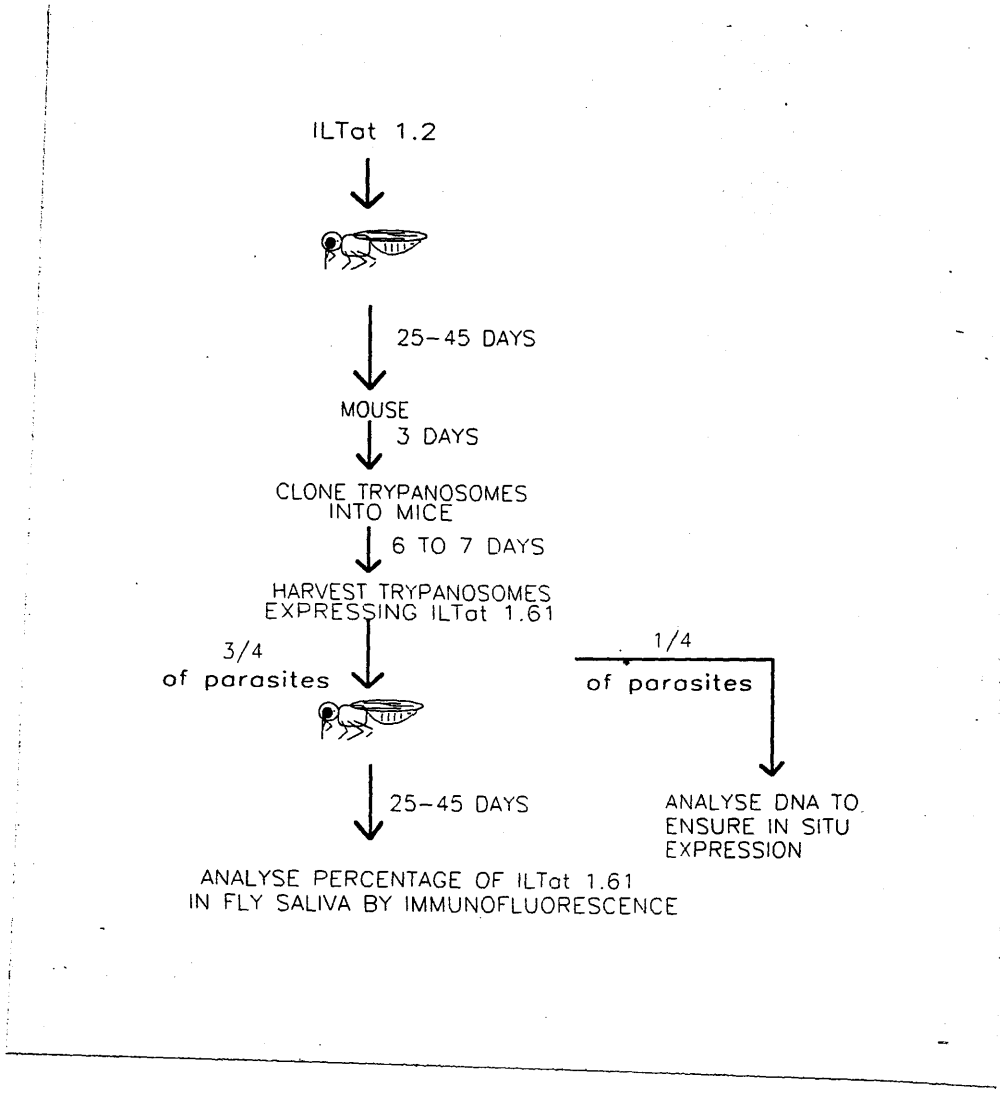


Figure 3.1; Design for the experiment described in this chapter. ILTat 1.2 is a bloodstream VAT whose gene is expressed without duplication (Penncavage et al., 1983). ILTat 1.61 is an M-VAT in the EATRO 795 stock. Tsetse flies indicate cyclical transmission.

Facing page 54

potential explanation for an I-VAT effect and indicates that the activation signals for expression site transcription are preserved with passage through the tsetse fly.

Hajduk et al. (1981) analysed the I-VAT memory effect to investigate the relationship between M-VAT and B-VAT (Bloodstream Variable Antigen Type) gene activation and found that the ingested VAT does not appear in the metacyclic repertoire. This was exploited by Turner et al. (1986), who infected tsetse flies with bloodstream trypanosomes expressing M-VATs, almost certainly from an expression linked copy (Graham and Barry, unpublished observations). Again, it was found that there was no effect on the M-VAT repertoire and it was concluded that the metacyclic and bloodstream activation signals acted independently.

Subsequently, Turner and Barry (1989) have derived a virulent line of Trypanosoma brucei rhodesiense EATRO 795 which combines the features of antigenic stability and fly transmissibility. This is atypical; syringe passaged parasites display great antigenic stability but are unable to be transmitted through tsetse flies, probably because they are monomorphic (Ashcroft, 1960). In contrast, fly transmissible lines do show pleomorphism, but are antigenically very unstable (Le Ray et al., 1978). Shiels et al. (submitted) have utilized the stability of M-VAT expression in the virulent line to investigate the site of M-VAT transcription in the tsetse fly. By isolating cloned metacyclic or very early bloodstream organisms still expressing M-VATs, it has been found that two predominant M-VAT genes are expressed without generating an ELC (i.e in situ). After around 7-10 days the parasites were then found to abandon use of this expression site and instead generate an ELC (Chapter 5, this thesis). This is likely to reflect a switch between independent expression mechanisms; one specific to metacyclics (and perhaps very early bloodstream parasites) and the other specific to established bloodstream parasites, as had been predicted (Hajduk et al., 1981; Cornelissen et al., 1985a; Turner et al., 1986).

The ability to obtain relatively large numbers of trypanosomes expressing M-VSG genes in the putative metacyclic

mode prompted a re-examination of the potential for an I-VAT effect in the metacyclic system. In order to do this, the experiment outlined in Figure 3.1 was conceived. In brief, the aim was to infect tsetse flies with clonal populations of trypanosomes still expressing a particular M-VSG gene as it is believed to be activated in the fly (i.e. from the basic copy location). In doing this, it was hoped to obtain information on the M-VSG gene activation signals: a DNA rearrangement or other heritable change in the expression site might be preserved through the fly, such that the proportion of metacyclic cells expressing the infed M-VAT in the transmitted population would be significantly elevated over normal frequencies. It was also hoped to provide proof that the parasites expressing M-VATs in the virulent line of EATRO 795 were truly still using the metacyclic system for VSG gene activation.

In the design of the experiment, it was decided to use trypanosomes expressing ILTat 1.61. This was because this M-VAT is a predominant member of the metacyclic repertoire in the EATRO 795 stock (thereby making the isolation of expressor clones relatively simple) and had previously been observed to maintain in situ expression for longer than another predominant M-VAT, ILTat 1.22 (Graham and Barry, unpublished observations). This was important because of the necessity in this experiment to obtain organisms still expressing the M-VSG gene from its basic copy location in sufficient numbers for the infection of tsetse flies and the isolation of DNA.

3.2 Isolation of clonal ILTat 1.61 populations.

Trypanosoma brucei rhodesiense, at a concentration of $1x10^6$ organisms/ml in citrated sheeps' blood, were fed to 68 Glossina morsitans morsitans of a trypanosome susceptible line (Maudlin and Dukes, 1985) at their teneral feed. After 30 days maintenance on citrated sheeps' blood, the flies were allowed to salivate onto microscope slides warmed to $41^{o}C$ and these were examined for the presence of parasites. In total, 5 flies were found to be producing metacyclic organisms, two of which were fed onto an

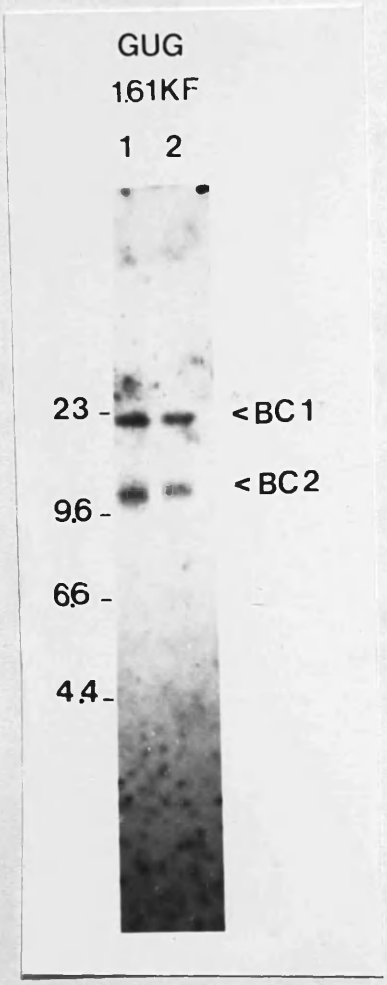


Figure 3.2; HindIII digested GUG 1.61 KF 1 and 2 DNA, hybridized with the insert of pTcV21-15. 0.1xSSC; 65°C. BC=Basic Copy of the ILTat 1.61 genes. The numbering is arbitrary.

Facing page 56

anaesthetised 16 week old mouse. Three days after fly bite, tail blood was taken from these mice and trypanosomes isolated from the buffy coat. Individual trypanosomes were then cloned from microdroplets into a further 12 mice and the infection allowed to proceed for seven days. After this time, two mice developed a parasitaemia, one of which contained trypanosomes expressing the ILTat 1.61 M-VAT at a level of 71%. This mouse was exsanguinated by cardiac puncture and 1.5 ml of its blood cryopreserved in liquid nitrogen in 0.5 ml aliquots, suitable for fly infection. This trypanosome population was named GUG 1.61 KF1. Trypanosomes in the remaining 0.3ml of blood from the infected mouse were purified on a DEAE-cellulose column and their DNA isolated.

Subsequently, a further clonal trypanosome population expressing ILTat 1.61 to a level of 85% was isolated in the same way as described above. This population was cryopreserved on day 6 after cloning and named GUG 1.61 KF2. As before, DNA was also isolated from these trypanosomes.

3.3 DNA analysis of GUG 1.61 KF1 and GUG 1.61 KF2.

The EATRO 795 stock has two basic copies of the ILTat 1.61 gene (Cornelissen et al., 1985a), with the genomic environment of each appearing to be allelic because they are identical in map for all restriction enzymes tested for at least 22 kb upstream of the VSG genes (Chapter 4, this thesis). Prior to the infection of tsetse flies, it was important to verify that GUG 1.61 KF1 and 2 contained just this basic copy number for the ILTat 1.61 gene, thereby establishing in situ expression. Therefore, 0.5 ug of DNA from these populations was digested with HindIII, which generates fragments which run from within the two basic copy VSG genes to the telomere ends (Cornelissen et al., 1985a), a distance which varies due to the nature of trypanosome telomeres (see section 1.6). This DNA was then size fractionated by agarose gel electrophoresis and a Southern blot hybridized with the ILTat 1.61 specific cDNA clone, pTcV21-15 (Cornelissen et al., 1985a). Although weak, due to the very limiting amounts of DNA available, Figure 3.2 shows that both populations contain

INFECTING POPULATION	CONCENTRATION OF TRYPANOSOMES IN FLY FEED.	NUMBER OF FLIES FED	NUMBER OF FLIES PRODUCING METACYCLIC CELLS
1.61 KF 1	6×10^6 /ml	50	5
1.61 KF 2	6×10^6 /ml	117	6
ILTat 1.2	1×10^6 /ml	41	5

TABLE 3.1 Experimental details of the cyclical transmission of the trypanosome populations 1.61 KF 1, 1.61 KF 2 and the control population, ILTat 1.2.

just the basic copy number of telomeres housing the ILTat 1.61 gene. The probe does not detect the HindIII fragment extending upstream of the gene; the cDNA clone from which the probe was derived extends only approximately 50 bp 5' to the HindIII site in the gene (Cornelissen et al., 1985a), a length insufficient to result in detectable hybridization on Southern blots under the conditions used.

3.4 Infection of Tsetse flies with GUG 1.61 KF1 and 2.

Since the two isolated ILTat 1.61 expressing populations appeared still to be transcribing their VSG genes from the basic copy locations, both were used independently to initiate tsetse fly infection. The details of these transmissions are given in Table 3.1, as are those for a transmission of parasites bearing ILTat 1.2, which is not an M-VAT, and acted as a negative control, being the population from which GUG 1.61 KF 1 and 2 had been derived . Those flies identified as harbouring metacyclic infection were allowed to salivate onto warmed microscope slides and the percentage of metacyclic cells expressing either ILTat 1.61 or ILTat 1.22 determined by the immune fluorescence antibody test. Two M-VATs were assayed to permit a comparison of the ILTat 1.61 and 1.22 levels in the transmitted populations. This was potentially advantageous, because a large elevation in the number of parasites expressing ILTat 1.61 would probably be reflected in a depressed level of other prominant M-VATs, such as ILTat 1.22. Because of the small number of metacyclic cells obtained per salivary probe (often less than 50), ILTat 1.61 and 1.22 levels from a single fly were determined on different days. The results should be comparable, however; Hajduk et al. (1981) report no variation in the proportion of distinct M-VATs produced by flies with time. Occasionally, a slide which had been stained for one M-VAT was subsequently restained for another to ensure that antibody was able to penetrate the cellular debris and faecal matter often associated with fly salivary probes. In all cases, the slide was viewed using phase contrast microscopy, to determine which of the non-fluorescent trypanosomes were

A: ILTat 1.61

POPULATION INITIATING INFECTION	NUMBER OF DISTINCT FLIES PROBED	TOTAL NUMBER OF METACYCLICS COUNTED	PERCENT EXPRESSING ILTat 1.61 (+/- range)
1.61 KF 1	5	534	39.4 (33-51)
1.61 KF 2	5	706	32.1 (10-45)
ILTat 1.2	3	499	43.7 (35-50)

B: ILTat 1.22

POPULATION INITIATING INFECTION	NUMBER OF DISTINCT FLIES PROBED	TOTAL NUMBER OF METACYCLICS COUNTED	PERCENT EXPRESSING ILTat 1.22 (+/- range)
1.61 KF 1	2	241	11.5 (10-50)
1.61 KF 2	3	860	27.0 (14-76)
ILTat 1.2	1	390	15.0 (0)

TABLE 3.2 Analysis of the percentage ILTat 1.61 and ILTat 1.22 expressors in the metacyclic populations derived from flies infected with 1.61 KF 1, 1.61 KF 2 and ILTat 1.2.

metacyclic stage and which were the non coated epimastigote and proventricular forms often regurgitated by flies.

Table 3.2 shows the proportions of ILTats 1.61 and 1.22 in the metacyclic populations of the three transmitted trypanosome populations. In this Table, the mean level for each M-VAT was calculated from a summing of all organisms counted, rather than by summing the percentage expressing each VAT from each fly probe. This was done to prevent those salivary probes with very few metacyclic organsims having an unrepresentative weighting in the calculated mean. Using this approach, it is not possible to calculate a meaningful standard deviation and so, instead, the range for each M-VAT is given, to indicate the degree of variability between samples. This was relatively great; the small number of metacyclic organisms per salivary probe results in a large founder influence.

It is clear from Table 3.2 that the ILTat 1.61 predominance is not greatly elevated where in situ expressors of this VAT are used to initiate tsetse infection. Instead, the negative control, ILTat 1.2 transmission, shows the highest prevalence of 1.61 expressors, although this is not greatly different from the GUG 1.61 KF 1 or 2 transmissions. The levels of both ILTat 1.61 and 1.22 are higher in all populations than is typical for these predominant M-VATs (Turner et al., 1986); this is probably a consequence of the difficulty in identifying non fluorescent organisms in salivary probes.

3.5 The fate of the transmitted populations in mice infected by fly bite.

In addition to the metacyclic populations, the levels of ILTats 1.61 and 1.22 were followed during their first week in a mammal host. In order to do this, infected flies were fed onto immune competent BKTO mice and trypanosomes examined by immune fluorescence upon either a buffy coat (for most infections 5 days from the fly) or a smear of infected blood (for days 6-7 from the fly). Too few trypanosomes were present in the blood 3 or 4 days after infection to allow a reliable quantitation of VAT

A: ILTat 1.61

POPULATION INITIATING INFECTION	DAYS FOLLOWING FLY BITE: DAY 5 %	DAY 5 RANGE	DAY 6 %	DAY 6 RANGE	DAY 7 %	DAY 7 RANGE
1.61 KF 1	44	(8-65)	28	(0-80)	14	(5-21)
1.61 KF 2	73	(71-90)	71	(42-95)	50	(13-93)
ILTat 1.2	44	(10-76)	50	(0-72)	19	(0-82)

B: ILTat 1.22

POPULATION INITIATING INFECTION	DAYS FOLLOWING FLY BITE: DAY 5 %	DAY 5 RANGE	DAY 6 %	DAY 6 RANGE	DAY 7 %	DAY 7 RANGE
1.61 KF 1	17	(12-27)	30	(0-50)	27	(1-66)
1.61 KF 2	12	(5-27)	3	(0-10)	1	(0-3)
ILTat 1.2	37	(10-49)	24	(11-46)	20	(0-79)

TABLE 3.3 Percentage of ILTat 1.61 and ILTat 1.22 expressors in the first patent parasitaemia in mice following cyclical transmission of 1.61 KF 1, 1.61 KF 2 and ILTat 1.2.

prevalence. The results of these analyses are given in Table 3.3. As above, mean values were calculated from the total number of organisms counted, rather than from the summed percentage for each VAT in each mouse on each day. These show that the levels of ILTat 1.61 were generally higher than ILTat 1.22 (as they were in the metacyclic populations), and that the relative prevalences of these M-VATs was similar in infections derived from flies infected with GUG 1.61 KF1 and ILTat 1.2. For GUG 1.61 KF2 derived infections, however, the level of the ILTat 1.61 was very high with respect to the other two transmissions and the ILTat 1.22 level depressed. These values were confirmed by a re-examination of duplicate blood smears prepared at the same time as those for the initial determination of VAT prevalences.

Discussion

The experiments described here have examined the effect upon the metacyclic repertoire of infecting tsetse flies with clonal populations of trypanosomes expressing the M-VAT ILTat 1.61 as it is believed to be activated in the fly. It was predicted that a genomic alteration (i.e DNA rearrangement or nucleotide modification) associated with M-VAT activation or inactivation might be preserved such that the proportion of the M-VAT in the transmitted metacyclic population would be significantly elevated. This appears not to be the case; ILTat 1.61 formed a similar proportion of the experimental transmitted metacyclic population as in a control transmission. This absence of an effect may be a consequence of either the M-VAT activation/inactivation mechanism or an inevitable limitation of the experimental system used.

1. Limitations and Assumptions of the Experimental System.

The foremost limitation in these experiments was the small number of infected tsetse flies and metacyclic organisms available. Despite the use of flies selected for susceptibility to trypanosome infection (Maudlin and Dukes, 1985) only approximately 5-10% became positive for the production of metacyclic form organisms. This, coupled with a fly mortality of approximately 1% per day over the 30-50 days needed for trypanosome cyclical development, meant that only 5 flies infected with GUG 1.61 KF1, 7 flies infected with GUG 1.61 KF2 and 5 flies infected with ILTat 1.2 were obtained. Furthermore, from each of these flies only 5 to 100 trypanosomes were available for immunofluorescence per salivary probe, as opposed to the 1,000-10,000 seen by others (Harley et al., 1966; Hajduk et al., 1981).

These limitations were inevitable. Firstly, the levels of mature metacyclic infection and fly survival obtained here were

as high as have been achieved by others (Hajduk et al., 1981) and could only have been increased significantly by initiating infection in far greater numbers of flies. This was impossible; the necessity to harvest the trypanosomes as soon as possible after leaving the fly meant that insufficient were available to infect further fly batches. Secondly, the small number of metacyclic organisms delivered per salivary probe was a consequence of using a trypanosome line compromising fly transmissibility for stability of M-VAT in situ expression. The more transmissible line used by Hajduk et al. (1981) shows a loss of M-VAT expression only 5 days from leaving the fly, even in the absence of antibody removal (Hajduk and Vickerman, 1981). This would not have permitted the isolation of antigenically relatively homogeneous populations in sufficient numbers for DNA analysis.

The primary assumption of the experiments is that the parasites used to initiate fly infection were truly still expressing M-VATs in the way that they were activated in the fly. This assumption is based upon the necessarily indirect evidence of Shiels et al. (submitted), using the model system used in this study. It is formally possible that instead of in situ expression, true metacyclic organisms express M-VATs via a metacyclic specific ELC which is abandoned in the bloodstream before there are enough parasites for DNA analysis. This seems unlikely, however; the ELC would remain lingering in the metacyclic expression locus after the switch to bloodstream specific in situ M-VAT expression. Such an ELC, which would be readily detected in Southern analysis, has never been seen for either ILTat 1.61 or another M-VAT, ILTat 1.22 (Graham and Barry, unpublished).

A second assumption is that trypanosomes still expressing M-VATs as they are activated in the fly are able to differentiate into stumpy form organisms and, thus, initiate fly infection (Ashcroft, 1960; Wijers and Willet, 1960). If this were not so, then it may have been the small percentage of non ILTat 1.61 expressors in the GUG 1.61 KF populations which were responsible for fly infection and no ingested VAT effect in the metacyclics

would be predicted. This also does not seem likely. Although not quantified, the percentage of non M-VAT expressing trypanosomes in GUG 1.61 KF2 was lower than the percentage of parasites judged to be stumpy forms in blood smears and, after immunofluoresence, organisms bearing the ILTat 1.61 coat were seen to have stumpy morphology.

A final assumption in these experiments is that the bloodstream and metacyclic expression of ILTat 1.61 cannot operate from the same telomere using distinct activation signals. If this were not the case, then it would be expected that the basic copy loci for the ILTat 1.61 genes would have features resembling a bloodstream expression locus in addition to metacyclic specific features. This is not unprecedented; the basic copy loci for two M-VAT genes in T.b.rhodesiense have ESAG 1 (Son et al., 1989), although in neither has this site been demonstrated to be the metacyclic expression locus site. A potential for an indefinite boundary between bloodstream and metacyclic repertoires has also been noted by Laurent et al. (1984a), although unlike ILTat 1.61, this involved a VAT (AnTat 1.6) which was expressed at a very low level very early in the bloodstream and may not be a true M-VAT (J.D.Barry, personal communication). Although difficult to exclude without a structural and transcriptional characterization, an expression locus for the ILTat 1.61 gene, able to act both in the insect and when established in the bloodstream, would be hard to reconcile with the tendency for this gene to generate an ELC after 10 or so days in the blood (Shiels et al., submitted; one such expressor is examined in section 4.3).

A bifunctional expression site may also not affect the outcome of the experiment: the flies were infected with clonal populations of trypanosomes which were almost certainly expressing the M-VAT to a high level continually after cloning 3 days from the fly (Shiels et al.; submitted). Since, when cloned, these trypanosomes will almost certainly have been expressing by the metacyclic mechanism, any M-VSG gene activation signals may be as likely to be preserved in the developmental switch from M-VAT to B-VAT specific expression as they are moving from M-VAT

expression directly into the fly. Thus, there may be no actual difference between fly infection initiated with either B-VAT mode or M-VAT mode expressors, provided the origin of both populations was clonal.

2. Possible implications of this analysis for the mechanism of M-VAT activation/inactivation.

If the assumptions above are valid then these experiments indicate that there is no "memory" in the metacyclic VSG system of the M-VAT gene active before transmission. This might be due to one of a number of possibilities;

(a) Any genomic alteration associated with metacyclic VSG activation is reversed when the system is inactivated.

This hypothesis is difficult to address. The detection of any genomic rearrangements in the ILTat 1.61 expression site might require extensive cloning and sequencing of the locus in metacyclic, bloodstream and procyclic stages and even a gross analysis by Southern hybridization is precluded by a lack of available DNA from the trypanosomes still expressing M-VATs. There has been no evidence of any such rearrangements of bloodstream expression loci when transmitted through the tsetse fly (Delauw et al., 1985).

The presence of nucleotide modification is also difficult to identify unless it results in endonuclease inhibition as has been seen for inactive bloodstream expression loci (Bernards et al., 1984b; Pays et al., 1984; Crozatier et al., 1988; section 1.11, this thesis). It has been reported that M-VAT gene loci do not show nucleotide modification (Lenardo et al., 1986), although there may be very tentative evidence for endonuclease inhibition at the 1.61 locus in non expressor bloodstream parasites (Chapter 4, this thesis).

(b) Metacyclic VSG gene activation is not mediated through genomic alteration, but instead involves binding of metacyclic specific factors to one of several metacyclic VSG gene expression loci in a way that is not preserved through the tsetse fly.

This model is compatible with some very recent observations on bloodstream VSG gene activation. Zomerdijk et al. (submitted) have sequenced the transcription initiation region for the VSG 221 gene locus in its active and inactive form and have found no difference. Although there is a possibility of distal rearrangements influencing the VSG gene promoter, it is simpler to envisage the existence of an (in)activation "complex" capable of binding and regulating VSG expression sites (Bernards et al., 1984b; Van der Ploeg and Cornelissen, 1984; Borst and Greaves, 1987; Pays and Steinert, 1988). The alternate activation of distinct bloodstream expression sites could be achieved by the infrequent disassociation and relocation of such a complex. Zomerdijk et al.(submitted) have further shown that inactive expression sites in the bloodstream show no transcriptional engagement, which contrasts with the I-VAT inactivation in the fly, where expression site transcription appears to be initiated, but truncated (Pays et al., 1989b). This seems to imply that there is a specific mechanism of expression site inactivation in the fly which preserves the activation signal during transmission. The absence of such a specific system in metacyclic cells would result in the observed lack of an I-VAT effect, as labile M-VAT activation signals would be lost either on entering B-VAT expression, or in this experiment, the tsetse fly. The avidity of metacyclic expression sites for an activation complex might explain the relative predominance of ILTat 1.61 over ILTat 1.22 in the metacyclic repertoire since the former is present at twice the copy number of the latter in the genome of these trypanosomes. Other factors which might effect VAT levels, such as growth rate (Seed et al., 1984; Miller and Turner, 1981; Myler et al., 1985) are not applicable in the non-dividing metacyclic

form.

3.Continued M-VAT expression in the bloodstream.

The proportion of trypanosomes expressing the M-VATs ILTat 1.22 and 1.61 were followed during the first patent parasitaemia in mice. In all cases the range between the samples is relatively great. Early on, this is probably contributed to by the small number of metacyclics delivered by the flies, which can result in quite dramatic founder effect differences between individual mice. Later, on days six to seven from the fly, some mice seem to control infection better than others, so that a given VAT can vary between quite high levels (antibody level too low) to a complete absence (high antibody level). Because these differences should be equal between different transmissions, however, the relative VAT predominances can be interpreted, albeit tentatively.

During the course of the first patent parasitaemia derived from the ILTat 1.2 and GUG 1.61 KF1 iniated transmissions, the mean level of ILTat 1.61 was approximately similar to the levels observed in the metacyclic populations, until around day 7. The same was also true of ILTat 1.22 levels, although there was a slight increase in prevalence during growth in the bloodstream. This might suggest that trypanosomes bearing these VATs may show similar growth rates, as suggested by Turner and Barry (1989).

In contrast, following GUG 1.61 KF2 transmission, ILTat 1.61 increased from 33% in the metacyclic population to around 70% of bloodstream trypanosomes 5-6 days later, and ILTat 1.22 levels were depressed with respect to the other transmissions. Whilst the day 5 figures may be somewhat inaccurate because of the small number of trypanosomes available for immunofluoresence at this time, the day 6 figures were consistent with this pattern in many (but not all) replicates and in infections initiated from different flies. These levels may be due to the clonal origin of the trypanosomes. In their original isolation, the GUG 1.61 KF2 reached a parasitaemia sufficient for DNA analysis a day earlier than GUG 1.61 KF1, although both populations were originated from

a single trypanosome and were expressing predominantly the same VAT. This might suggest that the trypanosome clone used to initiate the GUG 1.61 KF2 population had undergone some change such that it was able to grow faster when bearing ILTat 1.61 than other trypanosomes. This would result in an increased ability of ILTat 1.61 expressors to outgrow other VATs in the first patent parasitaemia after infection, explaining both its prevalence and the depressed level of ILTat 1.22 expressors. An alternative explanation concerning preferential switching to ILTat 1.61 during M-VAT inter-switching is unlikely because of the relative antigenic stability of these trypanosomes. Similarly, the result is unlikely to be a consequence of a normal bloodstream I-VAT effect operating from the 1.61 telomeres acting as a bifunctional expression site (see above) because this would be emphasised in populations isolated later, rather than earlier, after cloning as is seen here.

The observed increase in prevalence of ILTat 1.61 after GUG 1.61 KF2 transmission is remarkable, but not unprecedented. Turner et al.(1986) observed an increase in the proportion of trypanosomes bearing GuTat 7.13 (an M-VAT in the same serodeme as ILTat 1.61) from 10% in the metacyclic population to 30-40% after 4-5 days in the blood and Hajduk and Vickerman (1981) noted an increase in AnTat 1.30 from 11-20% of metacyclics (Barry et al., 1979) to 30% of day 4 bloodstream trypanosomes.

4. Possible explanations for the presence of an ingested VAT memory effect in bloodstream, but not metacyclic, trypanosomes.

It is possible to speculate on the reasons, if any, that the trypanosome displays an ingested VAT effect in the bloodstream, but not metacyclic, VSG system. The simplest explanation would be that the effect is a consequence of the activation or inactivation mechanism involved at each stage and confers no advantage on the trypanosome. Thus, on entering the fly VSG gene expression may be shut down by the transcription truncation mechanism identified by Pays et al.(1989b) simply because this is a very rapid mechanism of control, while in the metacyclics no

such system is necessary.

Alternatively, the I-VAT effect may be of definite selective advantage to the parasite as it begins the infection of a new host. In the bloodstream, the VATs which appear after M-VAT expression are quite predictable because some VATs appear to have higher probability of activation than others (Liu et al., 1985). In the field, game animals continually exposed to trypanosomes are likely to gain a degree of immunity to these dominant VATs such that there is the potential for trypanosome transmission to be blocked. Although minor VATs are also likely to be expressed in the first patent parasitaemia, these may not be in sufficient numbers to establish infection. By employing the ingested VAT effect trypanosomes can introduce novel VATs into their predominant VAT repertoire at each transmission which are dependent only on the VATs that were being expressed in the former host. These ingested VATs, together with any VATs encoded by lingering ex-ELCs in the transmitted trypanosomes, will present hosts with sufficient numbers of antigenically unfamiliar trypanosomes to establish infection (Michels et al., 1984).

In metacyclic cells, an ingested VAT effect would be potentially disastrous. The work here and the work of others have demonstrated that some M-VAT prevalences can increase very early in the bloodstream before the M-VAT system is shut down. If there were a memory of the M-VAT being expressed, then with passage through tsetse flies, the most prevalent M-VATs would be present at a higher level in the transmitted metacyclic population than in the former transmission. This might not just apply to the artificial system used here; a memory could be preserved from the metacyclics cells, into the bloodstream parasites and then into the fly. This would result in a self reinforcing cycle, with a tendency towards the expression of just a single predominant M-VAT in the metacyclic populations, against which a field immunity might easily develop.

This effect could also provide a partial explanation for why predominant M-VATs are expressed in situ. Even in the absence of a specific I-VAT effect, any system employing duplicative transposition would be prone to the generation of lingering ELCs.

Since it is, in part, developmental inactivation, rather than immune selection, which removes M-VATs from the bloodstream, there might be a greater tendency for any gene that can more readily convert the expression site, or encodes a VSG permitting faster growth rate, to remain lingering when the system is inactivated. As above, with repeated cycles, this would lead to a progressive reduction in the M-VAT repertoire. By maintaining each M-VAT gene in its own expression locus the probability of activation of each is fixed and a heterogeneous M-VAT repertoire is ensured.

Thus, both the presence and absence of any memory effect in bloodstream and metacyclic organisms can have the same effect: to promote antigenic heterogeneity in the infecting population and thereby enhance the probability of trypanosomes being able to invade new hosts.

Chapter 4

Cloning of the ILTat 1.61 metacyclic VSG gene expression locus.

4.1 Introduction.

In the previous chapter, and from studies by others, it had been seen that the VSG system expressed during the metacyclic stage of the trypanosome might be very different to that seen in more established bloodstream infection. The system is far more limited in repertoire (Crowe et al., 1983; Turner et al., 1988), more predictable in the short term (Barry et al., 1983), appears to be expressed by an independent mechanism (Turner et al., 1986) and shows no evidence of an ingested VAT memory effect upon inactivation and subsequent reactivation after fly transmission (Chapter 3, this thesis). This prompted the questions;

(i) what features of the ILTat 1.61 basic copy gene environment enable it to function as a metacyclic VSG expression locus?

(ii) In what way are these different from those seen in a bloodstream expression site?

In order to address these questions, a cloning of the telomeres harbouring the two basic copies of this gene has been attempted. The simplest approach to this would be the screening of a pre-existing trypanosome genomic DNA library constructed by the insertion of Sau3A partial digestion products into lambda EMBL 4 (Shiels, 1990). A probe for this was available; the ILTat 1.61 specific cDNA, pTcV21-15 (Cornelissen et al., 1985a). This approach, however, had already been attempted by Shiels (1990) and resulted in the isolation of no stable recombinant phage which hybridized with the probe. Although this may have been due to the underrepresentation of telomeric fragments in libraries created by partial digestion, a simultaneous screening had resulted in the successful isolation of clones derived from the telomeric ILTat 1.22 basic copy gene locus. An alternative possibility, therefore, was that clones containing inserts from the ILTat 1.61 basic copy locus were non-viable or unstable in

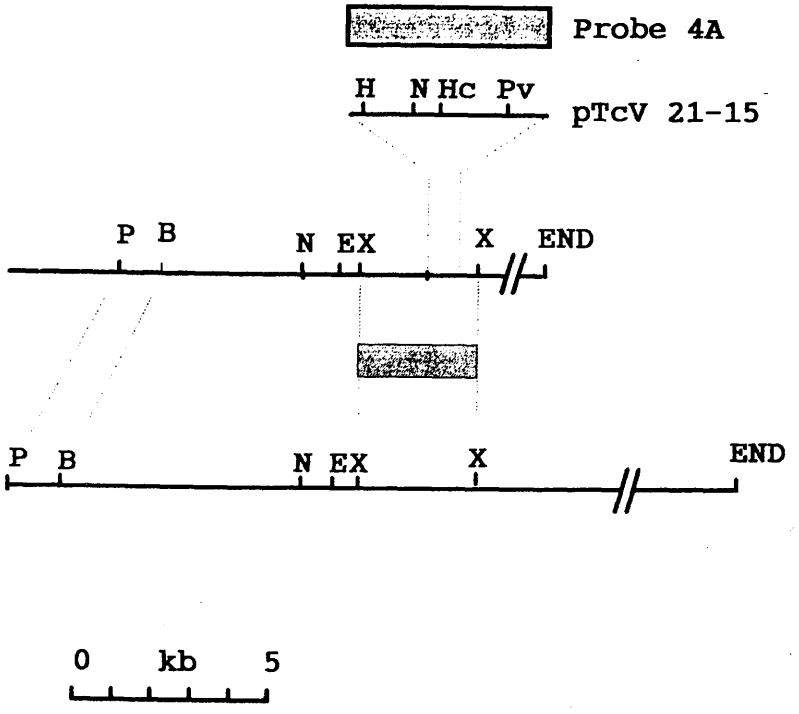


Figure 4.1. Restriction map for the gene environment of the ILTat 1.61 genes in Trypanosoma brucei rhodesiense EATRO 795 (derived from Cornelissen et al., 1985a). P=PstI, B=BamHI, N=NaeI, E=EcoRI, X=XbaI, H=HindIII, Hc=HincII, Pv=PvuII, END=telomere. pTcV21-15 is a cDNA clone for the 1.61 genes. The boxed region between the two chromosome maps represents an XbaI fragment identified as being suitable for cloning.

lambda libraries.

In view of these difficulties, I decided to attempt the cloning of the ILTat 1.61 locus by an alternative approach; the cloning of specific overlapping restriction fragments in plasmid vectors: a form of very limited chromosome walk. I hoped that this might allow the cloning and analysis of areas unstable in long lambda inserts. Although the number of recombinants which can be isolated from plasmid libraries is generally far less than from phage libraries, this was not considered a serious limitation here. The haploid trypanosome genome is approximately $4x10^4$ kb (Borst et al., 1982) which, when digested with an endonuclease with a six base pair recognition sequence, should generate around 10,000 distinct DNA fragments. By fractionating these on agarose gels and carefully purifying gel slices containing specific restriction fragments of interest (spanning 1/50-1/100 of the genomic smear) it was estimated that after ligation and transformation 1/100 to 1/200 colonies should contain the desired genomic insert. Whilst this was obviously a crude estimate, it implied that the probability of cloning specific restriction fragments by this approach was sufficient to be practical.

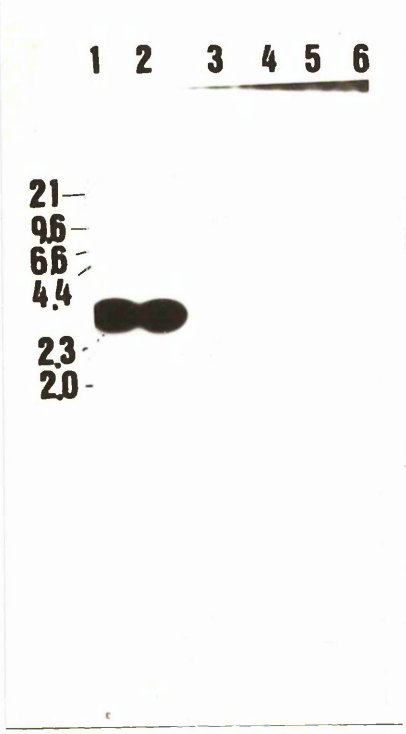
4.2 Cloning of pMT1.61-1

A limited pre-existing restriction map for the two 1.61 gene basic copy loci in the EATRO 795 serodeme indicated that both were identical for at least 5 kb 5' to the VSG gene, beyond which one appeared to contain a 2 kb insertion (Cornelissen et al., 1985a; Figure 4.1). This map was analysed for the presence of a clonable restriction fragment which spanned the region hybridizing to the cDNA for the 1.61 gene and extended for a reasonable distance upstream. This revealed that the most suitable clonable restriction fragment would be generated by XbaI digestion, which produces a 2.8 kb fragment, extending from 0.5 Kb downstream of the VSG gene to 1.5 Kb upstream (represented by a box in Figure 4.1). To clone this fragment, 50ug of 1.22j' DNA

was digested with XbaI. This DNA is derived from the virulent line of Trypanosoma brucei rhodesiense EATRO 795 used in Chapter 3, and expresses the M-VAT gene encoding ILTat 1.22 via an ELC. In order to monitor the degree of digestion of this DNA, 1/20 of the digest was run on an agarose gel. After Southern transfer and hybridization with the 1.61 gene cDNA, pTcV21-15 (probe 4A, Figure 4.1) it was clear that the digestion was complete (Figure 4.2.1). The remaining 19/20 of the digest was, therefore, size fractionated on an agarose gel prepared and run in TAE buffer and eight DNA fractions in the size range 2.5 to 3.5 kb excised and purified using "Geneclean". One tenth of the resulting size enriched DNA was subsequently run onto an agarose gel, blotted and hybridized with probe 4A. This revealed that fractions 1 and 2 of the purified DNA contained the desired restriction fragment (Figure 4.2.2). Fraction 2 was, therefore, ligated into dephosphorylated XbaI digested pUC19 and introduced into E.coli DS 941 made competent for transformation by the calcium chloride procedure. The cells were then grown on XIA plates. In total, 1060 colonies transformed to ampicillin resistance were recovered, of which 864 were recombinant as assessed by their inability to become blue in colour in the presence of X-gal and IPTG. Of these, 720 recombinant colonies were picked into duplicate 90 stab arrays and grown overnight on fresh plates. The resulting colonies were then transferred to Nytran membrane and screened for the presence of probe 4A hybridizing colonies.

After washing to a stringency of 0.1xSSC, 65^{o}C, seven strongly hybridizing bacterial stabs were detected, which were restabbed and rehybridized (Figure 4.2.3). Large scale plasmid preparations from these hybridizing colonies were then performed on four of the colonies and the resultant purified DNA mapped with a number of restriction endonucleases. The mapping gel for one, named 7A42, is shown in Figure 4.2.4. The derived maps indicated that the fragment had been cloned in each orientation and was in full agreement with that predicted from the cDNA insert restriction map (Figure 4.2.4). One difference between the map predicted by Cornelissen et al. (1985a) from genomic mapping and that observed here was the presence of an EcoRI site within

4.2.1 422 4.2.3



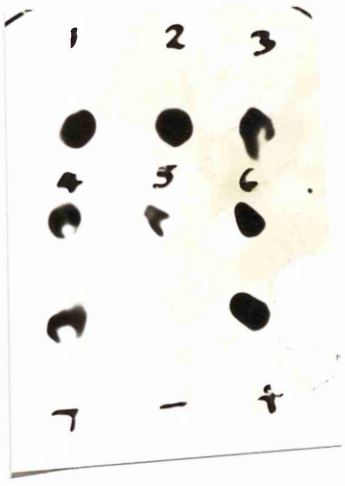


Figure 4.2.1; A fraction of the bulk 1.22j'trypanosome DNA XbaI digest hybridized with probe 4A.The hybridizing 2.8 Kb fragment is indicated. 0.1xSSC, 65°C.

Figure 4.2.2; 1.22j' trypanosome DNA digested with XbaI, size fractionated on an agarose gel and fractions of approximately 2.8 Kb eluted. One-tenth of the resulting purified DNA is shown hybridized with probe 4A. 0.1xSSC,65°C.

Figure 4.2.3; Restabbed recombinant colonies hybridized with the probe 4A. 1-7 are colonies containing XbaI fragments of 1.22j' DNA of approximately 2.8kb, (-)= pUC 19 containing bacteria, (+)= pTcV21-15 containing bacteria.

Facing page 74

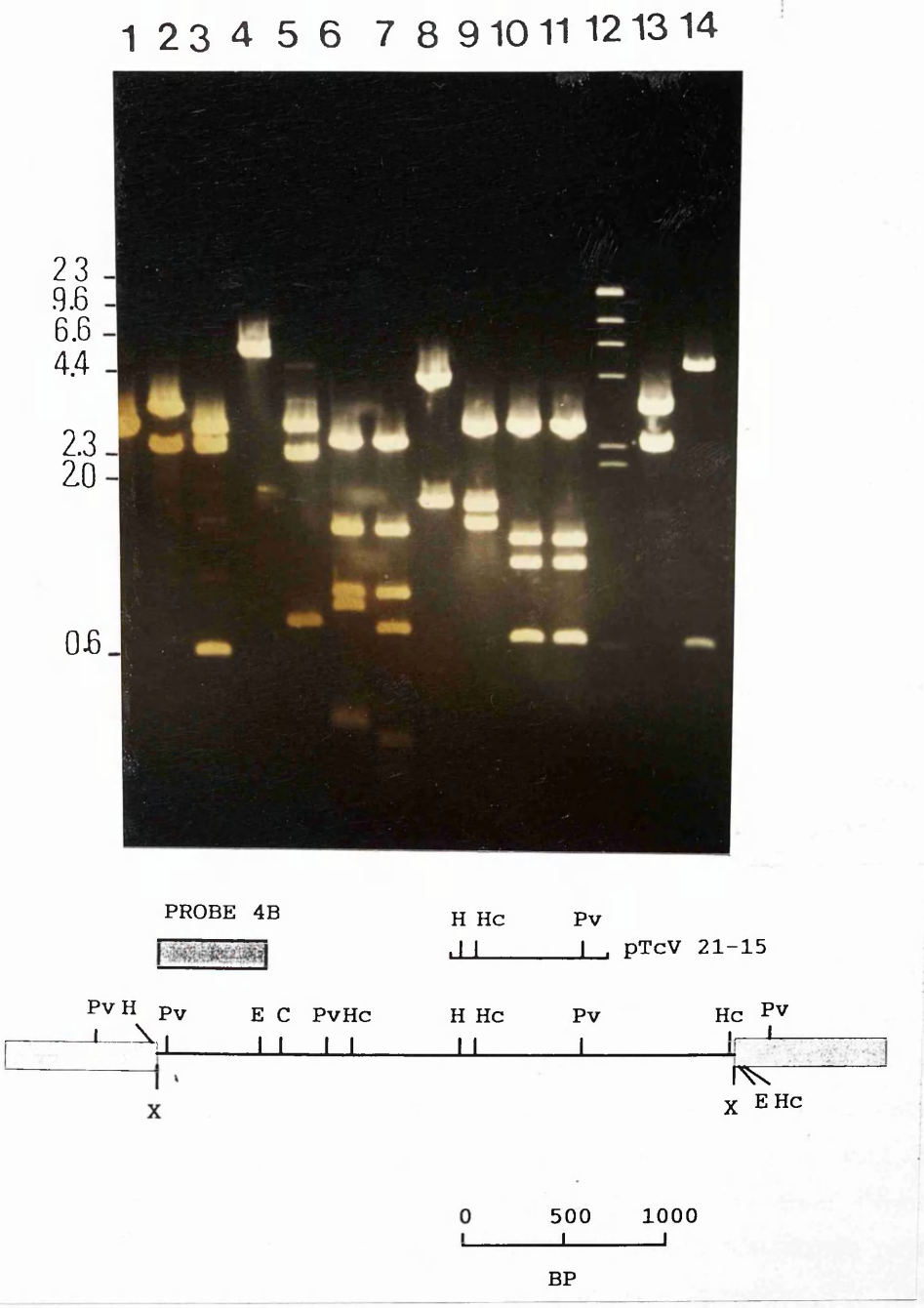


Figure 4.2.4; Mapping gel for one probe 4A hybridizing plasmid. Lanes 1-11 are, repectively, 7A42 digested with XbaI, EcoRI, EcoRI/XbaI, ClaI, ClaI/XbaI, PvuII, PvuII/XbaI, HindIII, HindIII/XbaI, HincII, HincII/XbaI. Lane 12 is HindIII digested lambda DNA size markers. Lanes 13 and 14 are two 4A hybridizing plasmids, 7A42 and 4A61, digested with EcoRI to show that they are in the reverse orientation. Below the gel is shown the derived restriction map for the 7A42 insert. Above the map is shown the corresponding pTcV21-15 insert hybridizing region and the area from which probe 4B is derived. Pv=PvuII, E=EcoRI, X=XbaI, Hc=HincII, C=ClaI, H=HindIII.

the 5' end of the insert (compare Figure 4.2.4 with the corresponding region on Figure 4.1). This is probably because of the difficulty in positioning restriction sites from genomic Southern blots when hybridizing a probe which detects fragments which run all the way to the telomere end, as was the case for EcoRI digestion in the work of Cornelissen et al. (1985a).

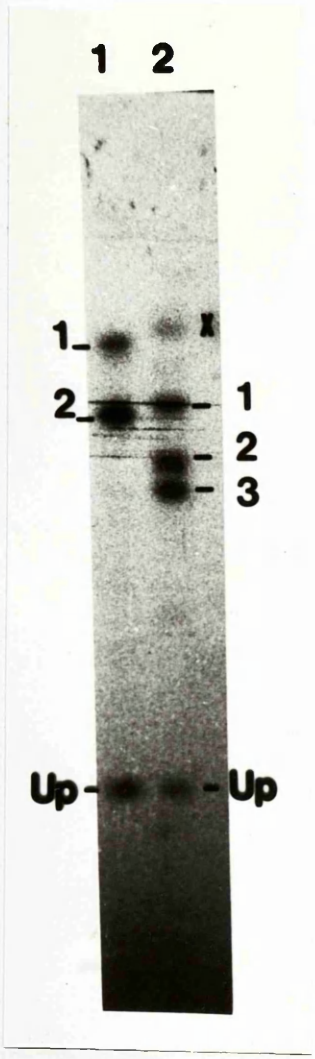
The plasmids derived in this section were named pMT (Metacyclic Telomere) 1.61-1, for the orientation of 7A42, and pMT1.61-2, for the reverse orientation.

4.3 Confirmation of the reality of pMT1.61-1.

When the 1.61 gene is expressed in bloodstream trypanosomes, an extra copy of the gene is generated at a bloodstream expression site (Cornelissen et al., 1985a; Graham and Barry, unpublished) in addition to the two basic copies of the gene. This observation was employed to ensure the reality of one pMT1.61-1 clone; its insert was hybridized to genomic DNA derived from trypanosomes with and without an ELC for the 1.61 gene. If real, the clone on a HindIII digestion would be expected to hybridize to two telomeres bearing the 1.61 gene for a non-expressor and three for an expressor with an ELC. Both DNAs should also show an additional band of unknown size representing hybridization to the HindIII fragment extending beyond the 5' limit of the XbaI clone. Figure 4.3.1 reveals that the above predictions are fulfilled by pMT1.61-1; there are three prominent bands in the expressor and two in the non-expressor, in addition to the upstream fragment on the telomeres (marked "Up" in Figure 4.3.1). The band marked "x" in this Figure is probably a product of partial DNA digestion, although it is possible that this band (or one of the other supposed telomeric fragments, labelled 1,2 or 3) is, in fact, derived from upstream region of the 1.61 gene ELC. This would be possible if the restriction maps of the basic copies and ELC copy began to diverge within the area detected by the probe.

The pMT1.61-1 insert was also hybridized to XbaI, XbaI/HindIII and HindIII digested 1.22j' DNA to ensure that

4.3.1



4.3.2

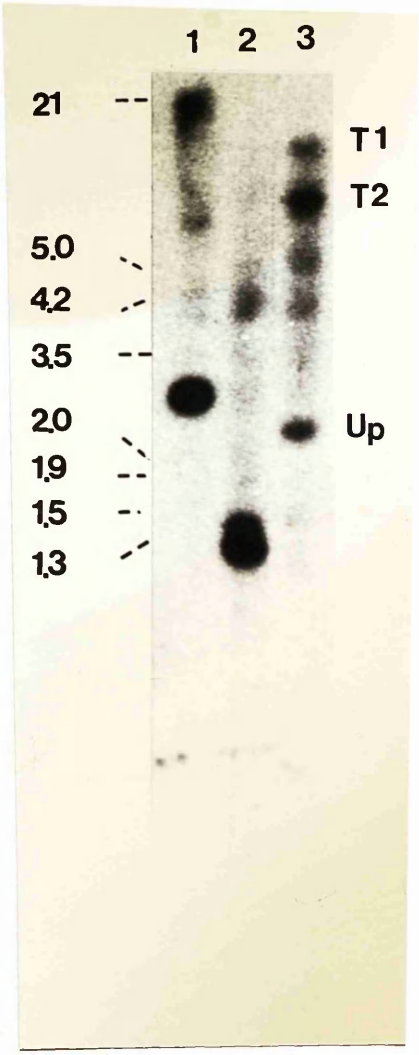


Figure 4.3.1; HindIII digested 1.22j' DNA, a 1.61 non expressor, (lane 1) or 1.61i, a 1.61 expressor (lane 2) hybridized with the pMT1.61-1 insert. To the side of the lanes are indicated the telomere fragments and the upstream HindIII fragment for the 1.61 basic copy loci (Up). "x" represents either a partial digestion product or an upstream fragment derived from the ELC copy of the 1.61 gene (see text). 0.1xSSC, 65^{o}C.

Figure 4.3.2; 1.22j' DNA digested with XbaI (lane 1), XbaI/HindIII (lane 2) or HindIII (lane 3) hybridized with the pMT1.61-1 insert. T1, T2 and Up shown beside lane 3 represent hybridization to the bands detected in Figure 4.3.1, lane 1. 0.1xSSC, 65^{o}C.

fragments predicted from the plasmid mapping were detected. Figure 4.3.2 demonstrates that this is the case; there is hybridization to a 2.8 kb XbaI fragment (lane 1), 1.5 and 1.3 kb XbaI/HindIII fragments (lane 2) and to the telomeres and upstream fragment on the HindIII digest (lane 3). In addition, a number of other high molecular weight bands are detected on this blot, albeit weakly. These bands are probably detected as a consequence of the pMT1.61-1 insert extending either into the region of the VSG gene which shows homology with other VSG genes (Majumder et al., 1981; Boothroyd et al., 1981; Rice-Ficht et al., 1981), or into the subtelomeric repeats downstream of VSG genes at chromosome ends (Blackburn and Challoner, 1984; Van der Ploeg et al., 1984c; Aline and Stuart, 1989).

Thus, from two analyses, pMT1.61-1 and pMT1.61-2 appeared to be faithful clones of the genomic environment around the ILTat 1.61 gene.

4.4 Genomic mapping of restriction sites 5' to the area represented on pMT1.61-1

In order to progress further upstream of pMT1.61-1 it was necessary to map restriction endonuclease cleavage sites 5' to the limit of this clone. To do this, 1.22j' genomic DNA was digested to completion with ClaI, EcoRI, HincII/EcoRI, HincII, HindIII or PstI. Once size fractionated, these digests were Southern blotted and hybridized with the 500 bp EcoRI/XbaI fragment at the 5' end of the pMT1.61-1 insert (probe 4B, Figure 4.2.4). The resulting autoradiograph and derived restriction map is shown in Figures 4.3.3 and 4.3.4. The map resolves no differences between the two basic copy loci, which is contrary to the results of Cornelissen et al. (1985a). This is addressed further in section 4.13.

From the restriction map it was apparent that a suitable restriction fragment to attempt to clone would be either the 6.1 Kb EcoRI/PstI fragment or the 6.2 Kb ClaI/PstI fragment extending 5' to the area represented by pMT1.61-1 (represented by line A on Figure 4.3.4). Of these, it was decided to select the latter

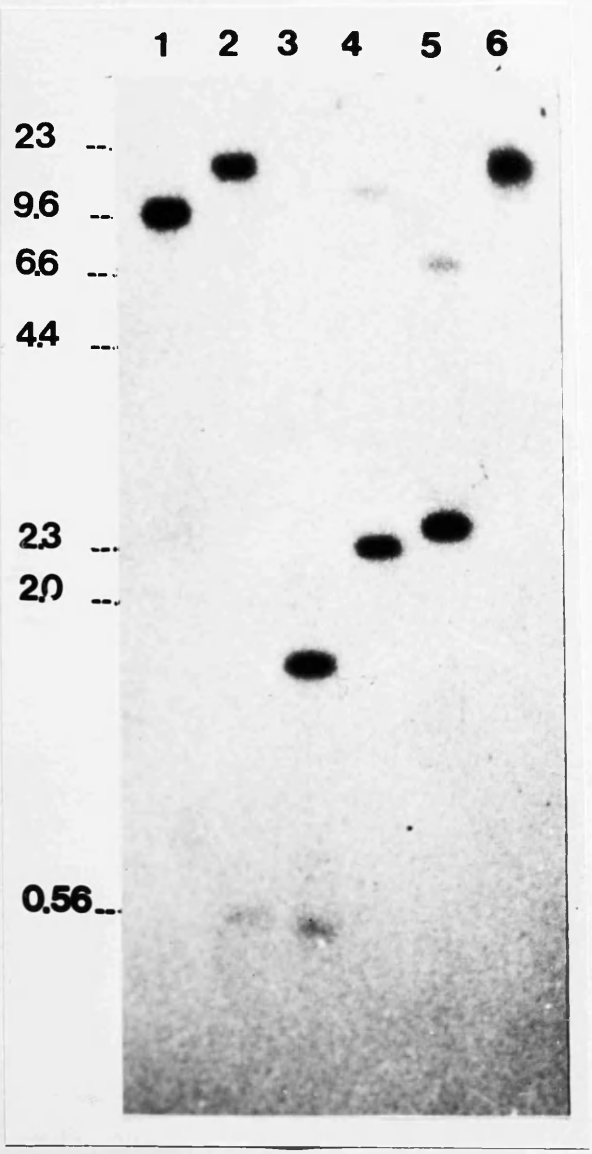


Figure 4.3.3 Genomic mapping of restriction sites upstream of the pMT1.61-1 insert. Lanes are 1.22j' DNA digested with (1) ClaI (2) EcoRI (3) HincII/EcoRI (4) HincII (5) HindIII (6) PstI and hybridized with probe 4B. 0.1xSSC, 65°C.

Facing page 77

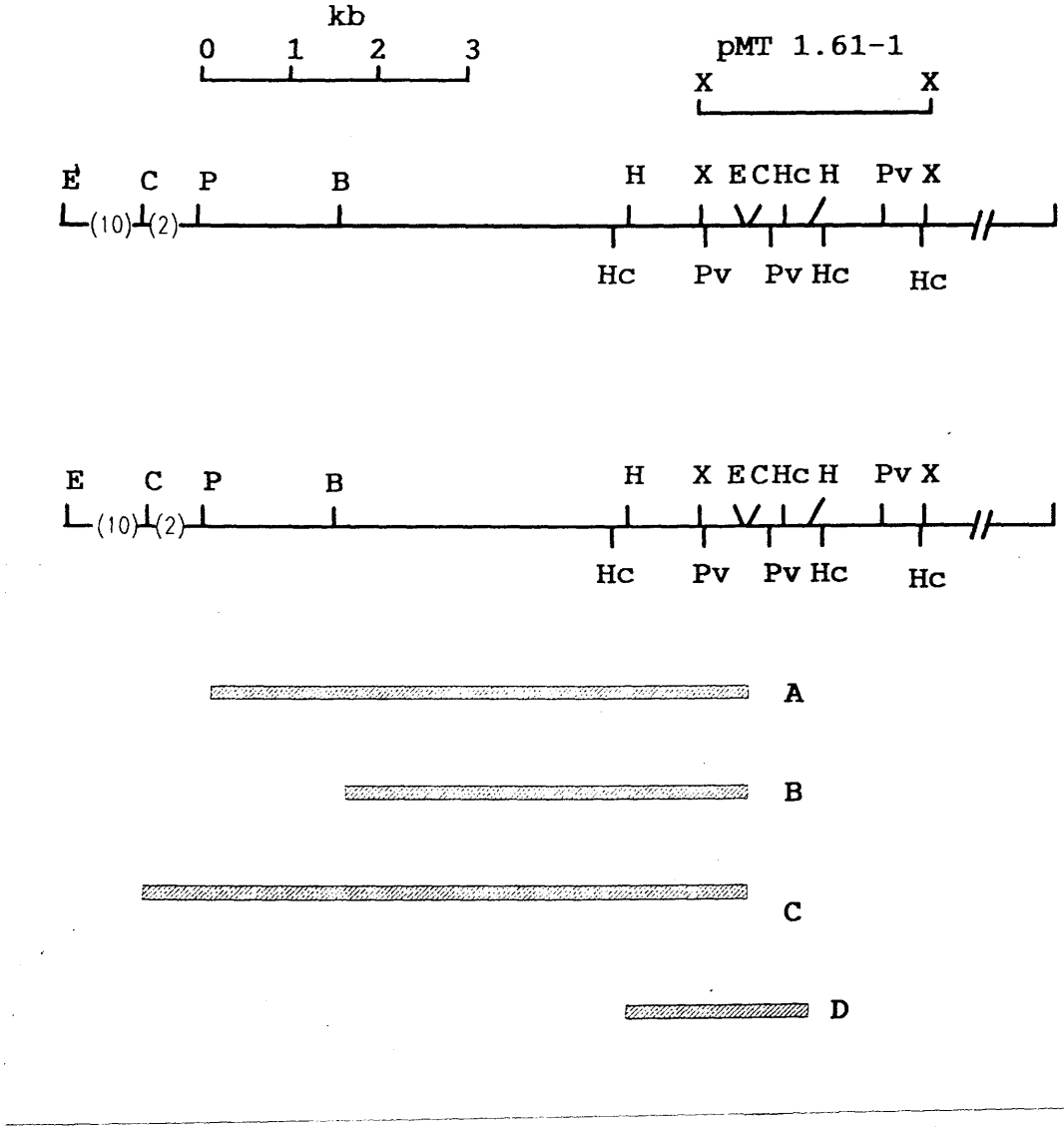


Figure 4.3.4 Restriction map for the ILTat 1.61 basic copy loci. E=EcoRI, C=ClaI, P=PstI, B=BamHI, Hc=HincII, Pv=PvuII, H=HindIII, X=XbaI. Lines below the map represent fragments selected for the construction of the libraries described in sections 4.5, 4.7 and 4.8.

because problems had been experienced in digesting 1.22j' DNA with PstI and EcoRI in combination, while PstI and ClaI had been seen to digest well (not shown).

4.5 Cloning of pMT1.61-3

The cloning of pMT1.61-3 was carried out essentially as described for pMT1.61-1. 50ug of 1.22j' DNA was digested sequentially with PstI and ClaI, size fractionated on a TAE agarose gel and segments between 4.4 kb and 6.7 kb excised. Once the purified DNA fractions had been hybridized to determine which contained the desired restriction fragment (Figure 4.4.1), DNA from the appropriate fraction was ligated into the plasmid vector pMTL23. This was used because pMTL23, unlike pUC19, has a ClaI restriction site within its polylinker sequence. Following transformation into E.coli DS 941, 770 colonies were obtained, of which 752 were recombinant. Duplicate arrays of these colonies were produced as stabs and screened using probe 4B. After hybridization, and washing of the filters to a stringency of 0.1xSSC, 65^{o}C, seven hybridizing stabs were identified as being clearly and specifically positive (Figure 4.4.2).

A large scale plasmid preparation of one of the colonies was then prepared and digested in order to ensure those fragments predicted from genomic mapping were present within the insert. When this was done, however, it was observed that although the insert size was correct (Figure 4.5.1, lane 1), and a fragment approximately corresponding to the size of the ClaI/XbaI fragment of pMT1.61-1 was present (600bp fragment, lane 8), other restriction patterns were incompatible with those predicted from genomic mapping. Furthermore, hybridization of the blot with the probe 4B failed to produce the expected hybridization pattern (Figure 4.5.2); the small 600 bp fragment generated in an ClaI/XbaI digestion, which should generate a fragment spanning the probe, failed to hybridize (lane 8), while other hybridizing fragments were incompatible with the expected pattern (Table 4.1). These results indicated that pMT1.61-3 was either a highly

4.4.1

4.4.2

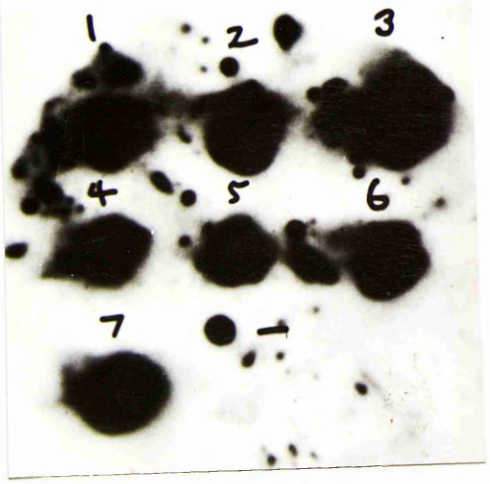


Figure 4.4.1; 1/10 of a size fractionation of PstI/ClaI digested 1.22j' DNA hybridized with probe 4B. 0.1xSSC, 65°C.
Figure 4.4.2; Restabbed colonies hybridizing with probe 4B. 1-7 represent recombinant colonies from the PstI/ClaI library. (-) is a stab of bacteria harbouring pMTL23. 0.1xSSC, 65°C.

Facing page 79

SIZE HYBRIDIZING (KILOBASES)	RESTRICTION ENZYMES USED								
	P/C	C/B	B/P	E	P/E	Hc	X	X/C	PX
PREDICTED	6.2	4.8	7.4	8.8	6.1	$\leq$4.1	8.8	0.6	$\leq$5.6
OBSERVED	6.2	3.2	3.2	4.8 0.7	1.7 0.7	10	4.8	4.8	4.8

TABLE 4.1 A comparison of the predicted hybridization pattern for probe 4B onto pMT 1.61-3 with the observed pattern. Sizes for both observed and predicted take into account the position of restriction sites within the vector.
P=PstI, C=ClaI, B=BamHI, E=EcoRI, Hc=HincII, X=XbaI.

rearranged derivative of the desired fragment, or that an unlinked fragment of a very similar size to, and showing a high level of sequence similarity with, the desired insert had been cloned. In order to determine which of these two possibilities was correct, probe 4B and one of its partial homologues, the EcoRI 720 bp fragment of pMT1.61-3 (see Figure 4.5.2, lane 4), were hybridized to duplicate Southern blots of 1.22j' DNA digested with XbaI, HincII and PstI/ClaI. If the pMT1.61-3 insert were a rearranged derivative of the 1.61 expression locus then this should generate an identical hybridization pattern in each case. If, however, the pMT1.61-3 insert were instead a faithful copy of a comigrating fragment showing homology to the 1.61 locus then, in each case, two bands would be expected to hybridize of a predictable size (see Figure 4.3.4 and Table 4.1). Figure 4.6 shows that the autoradiogram obtained fully conformed to the predictions for the pMT1.61-3 insert being derived from a location in the trypanosome genome other than the 1.61 basic copy loci. On an XbaI digestion, a 4.8 kb band derived from pMT1.61-3 and the 2.8 kb band cloned into pMT1.61-1 hybridized, while with HincII digestion a 2.3 kb 1.61 derived fragment was detected as was a 1.61-3 fragment of greater than 6.2 Kb (there is no HincII site within the pMT1.61-3 insert). Finally, the PstI/ClaI digestion resulted in just a single detected band, representing hybridization to the comigrating fragments derived from both the 1.61 and 1.61-3 loci.

For each blot, the intensity of the two hybridizing bands differs. This is because the 720 bp EcoRI fragment derived from pMT1.61-3 covers only a part of the region to which probe 4B hybridizes (refer to Table 4.1 and Figure 4.5.2, lane 4).

Having determined that pMT1.61-3 did not represent the desired insert, the remaining clones from the primary transformation which hybridized to probe 4B were examined. A small scale plasmid preparation was performed on each and the resultant DNA digested with EcoRI and XbaI. After electrophoresis, it was seen that none had the 500bp EcoRI/XbaI fragment seen in pMT1.61-1 (Figure 4.7, lane 1), but instead produced fragments compatible with pMT1.61-3. Thus, from seven

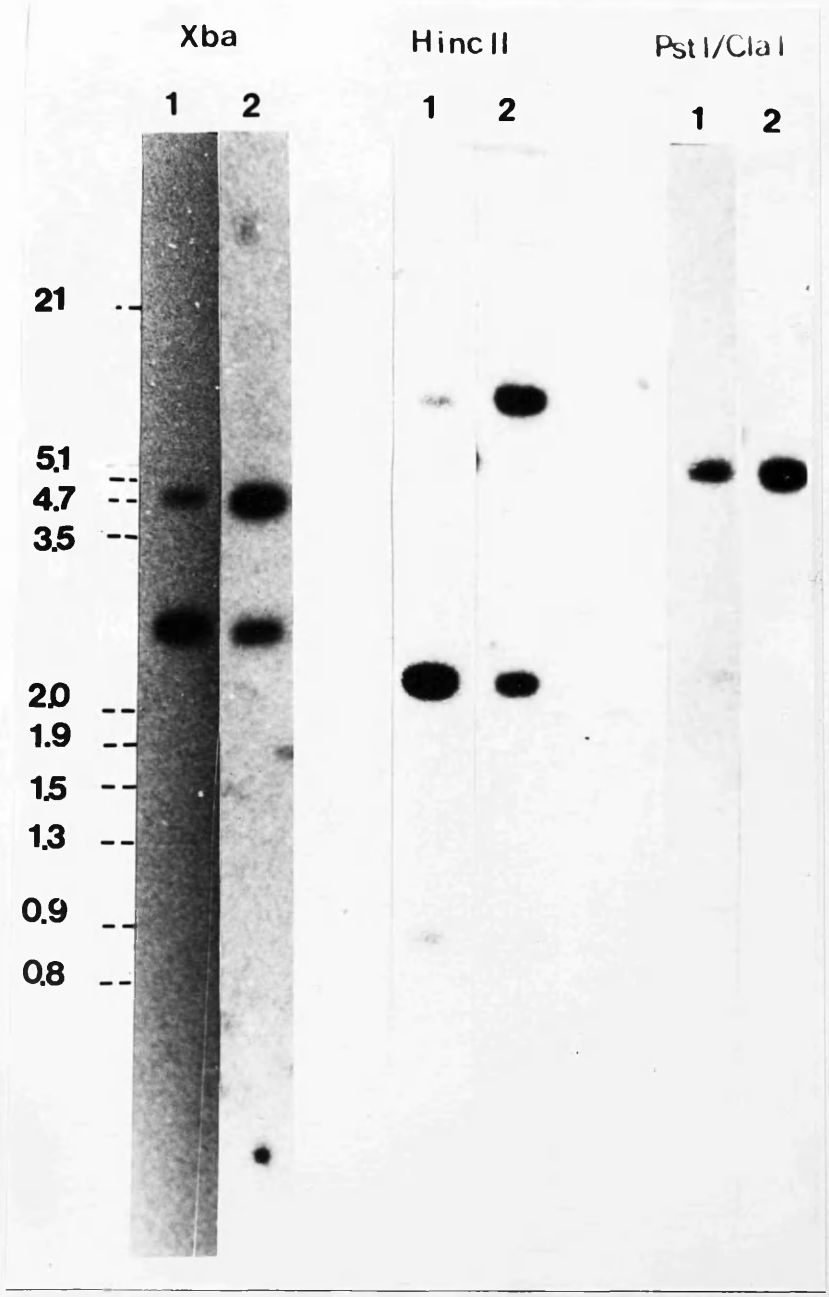


Figure 4.6; 1.22j' trypanosome genomic DNA digested with XbaI, HincII or PstI/ClaI and hybridized with either probe 4B (lanes 1) or a 720bp EcoRI fragment of pMT1.61-3 (lanes 2). 0.1xSSC, 65^{o}C.

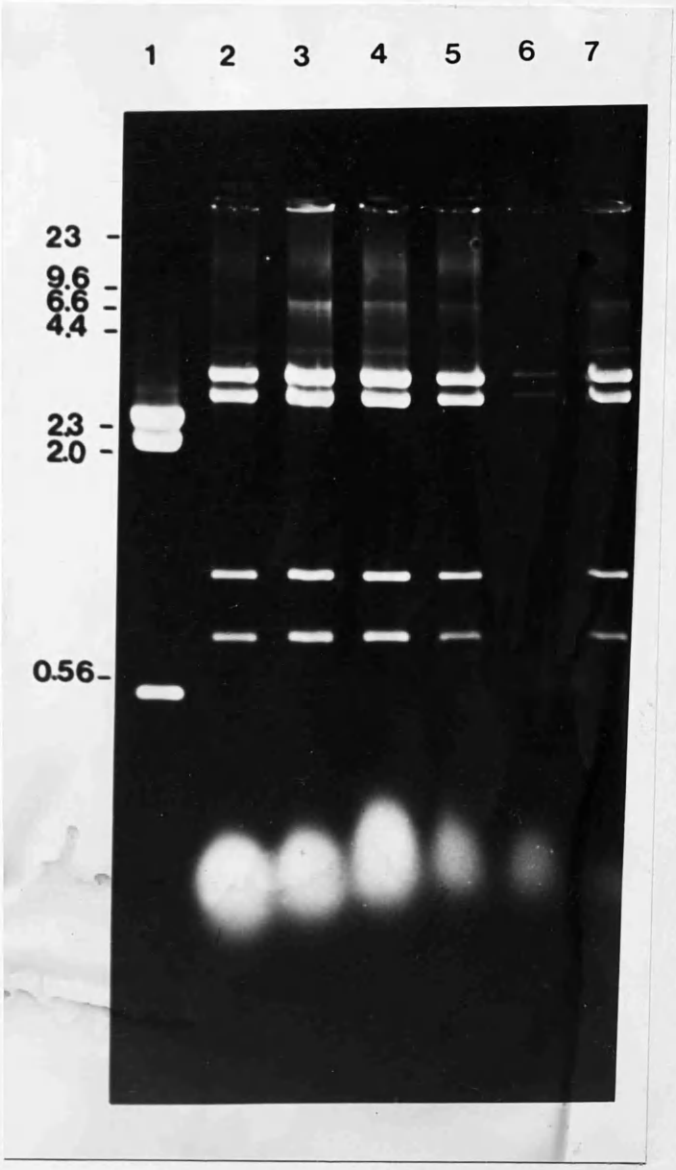


Figure 4.7; EcoRI/XbaI digested pMT1.61-1 (lane1) or probe 4B hybridizing plasmids isolated from the PstI/ClaI library construction (lanes 2-7). None of the 4B hybridizing plasmids yields a 500bp EcoRI/XbaI fragment predicted for a 1.61 loci derived clone.

```
          EcoRI
          ------    10    ***              30                    50
161       GAATTCAGTTGTACAATTTTTTTGCTTTATACTTTTAATTTAGGATAATTTCTATATGTT
          ||||||||||||||||| 0 ||||||||||||||||||||||||||||||||||||||||
163       GAATTCAGTTGTACACAATTTTTGCTTTATACTTTTAATTTAGGATAATTTCTATATGTT

                    70                    90                   *110
161       TCCAGAAACAATTAGAATGTATTGTTTTCGCTCTATTCTGTTACTTTTTAAATTTACATT
          ||||||||||||||||||||||||||||||||||||||||||||||||||1|||||||||
163       TCCAGAAACAATTAGAATGTATTGTTTTCGCTCTATTCTGTTACTTTCTAAATTTACATT

                    130                   150             *     170
161       GTCCGTTTTCACTCCGACCAACCAGCGTCACGCGTCGCTTCACCAGGATACGATTCATTT
          ||||||||||||||||||||||||||||||||||||||||||||||||2|||||||||||
163       GTCCGTTTTCACTCCGACCAACCAGCGTCACGCGTCGCTTCACCCGGATACGATTCATTT

                    190                   210                   230 **
161       TTTGCTTTCGCTTTAGATTCTTTTGTTCCGCTATAATTCGCTGTTTCGTAAGGCAGAATA
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||3 |||||
163       TTTGCTTTCGCTTTAGATTCTTTTGTTCCGCTATAATTCGCTGTTTCGTAAGGATGAATA

                  250  *       * *     *
161       ATTTGGTTTCCGGAAGCTCCTTACCTATCTGCTTTTTT.....
          ||||||||||||4|||||||5|6|||||7|||||||||
163       ATTTGGTTTCCGAAAGCTCCCT.CCTATATGCTTTTTT.....
```

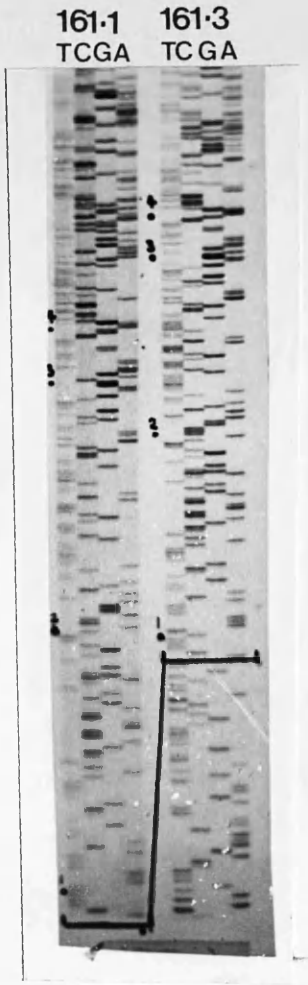


Figure 4.8; Partial sequence for the cross reacting region of pMT1.61-1 and pMT1.61-3. Differences are shown between the sequences and marked with an asterisk. The numbered differences correspond to the indicated differences shown on the sequencing ladders in the right hand panel.

independent clones, none had been detected which was derived from the 1.61 basic copy loci. This seemed to indicate that this area might be unclonable or underrepresented in plasmid libraries, in addition to its observed absence from lambda genomic libraries (Shiels, 1990).

4.6 Sequence of the pMT1.61-1/pMT1.61-3 cross reacting region.

It was of interest to determine the sequence of the region of homology shared by the pMT1.61-1 and 3 inserts since they might contain a conserved and functionally important motif or open reading frame. Therefore, both the pMT1.61-3 720bp EcoRI and the pMT1.61-1 500bp EcoRI/XbaI fragments were subcloned into the bacteriophage M13 and partially sequenced. Figure 4.8 shows the determined nucleotide sequence. It demonstrates a very high level of sequence identity between the two clones over the region analysed, with base substitutions rather than insertions or deletions comprising the majority of differences. Although only one strand of each insert was sequenced, the differences were consistently reproducible between independently isolated M13 subclones and the autoradiograms were unambiguous (see Figure 4.8). A computer search of the Genbank database using the Wilbur and Lipman (1983) algorithm implemented on the University of Wisconsin GCG sequence analysis package (Devereux et al., 1984) revealed no significant homologues, and an analysis of the nucleotide sequence using the "Translate" program revealed that neither contained an extensive open reading frame.

4.7 Attempted cloning of two upstream fragments on the 1.61 telomere.

The failure to detect the 6.2 Kb PstI/ClaI fragment of the 1.61 expression locus prompted an attempted cloning of two distinct fragments, a BamHI/ClaI fragment extending approximately 4.8 kb upstream of pMT1.61-1 (Figure 4.3.4, line B) and a ClaI/EcoRI fragment extending approximately 10 Kb upstream

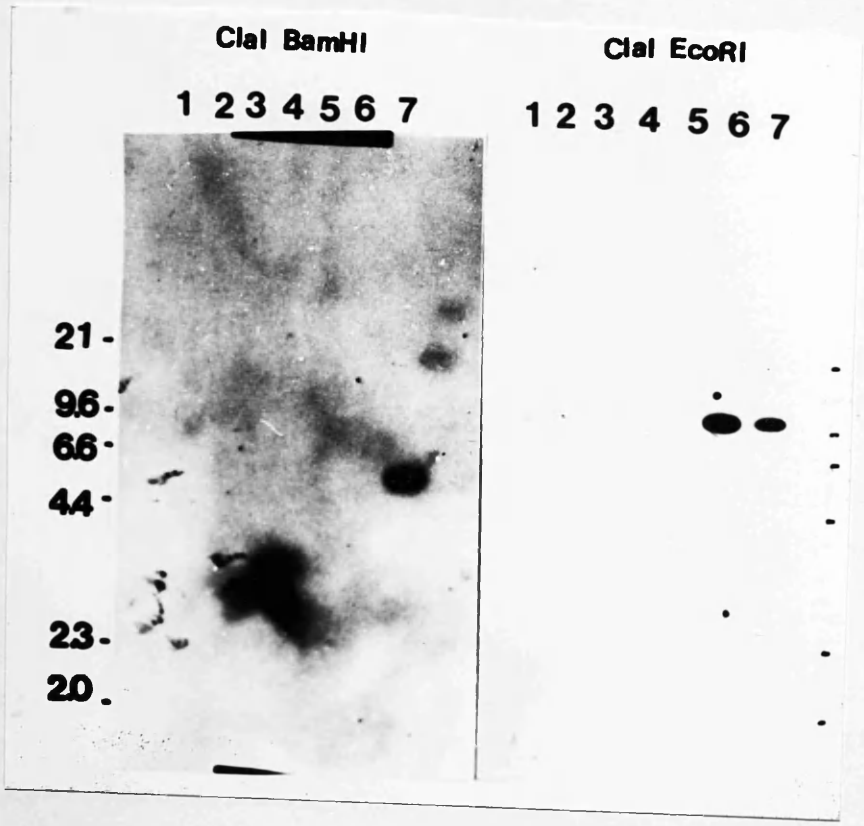


Figure 4.9; 1.22j' trypanosome genomic DNA digested with either ClaI/BamHI (lanes 1-7, left panel) or ClaI/EcoRI (lanes 1-7, right panel), size fractionated and hybridized with probe 4B. 0.1xSSC, 65^{o}C.

Facing page 84

(Figure 4.3.4, line C). These sites were chosen to generate fragments both upstream and downstream of the first PstI site on the 1.61 telomere, because it was possible that the sequence around this site, when in combination with the downstream ClaI site, would be unable to produce a stable insert. Both enzyme combinations would also eliminate the possibility of cloning a pMT1.61-3 linked fragment.

The libraries were constructed simultaneously and generated by the methodology described for the pMT1.61-1 cloning except that, as with the cloning of pMT1.61-3, the plasmid vector pMTL23 was used because of its possession of a suitable ClaI cloning site. Size selected fractions of a large scale digestion of 1.22j' DNA were hybridized with probe 4B, and the hybridizing fractions (Figure 4.9) ligated into pMTL23 and transformed into E.coli DS941. In total, 125 recombinants for the ClaI/EcoRI library and 600 recombinants for the ClaI/BamHI library were recovered, which were picked into arrays, hybridized with probe 4B and washed at high stringency. This revealed that neither library contained a hybridizing colony. To ensure that the white colonies were truly recombinant, single colony lysates were performed on ten randomly chosen colonies from each library. This revealed that all seemed to contain inserts, although there was evidence for a high level of instability in the ClaI/EcoRI library, since the uncut plasmid DNA ran at a very variable size between clones (not shown). All plasmids from the ClaI/BamHI library were of a relatively uniform size. Because of the potential for deletion when fragments around 10 Kb were inserted into pMTL23, it was not attempted to generate more recombinant clones than the relatively small number obtained from the ClaI/EcoRI library transformation. The ClaI/BamHI library was retransformed, however, to generate more recombinants. This transformation utilized the rec A^- host strain E.coli DS902 and was plated onto L-Broth plates supplemented with 0.2M glucose and ampicillin. Glucose was included here to repress the plasmid's lac promoter which may have been resulting in the fortuitous production of a toxic transcript or protein from the 1.61 telomere derived insert. For this reason, colour selection driven

from this promoter was not possible in this transformation. None of the 736 transformed colonies (of which 95% were predicted to be recombinant from the original transformation with colour selection) was found to contain an insert which hybridized to probe 4B.

4.8 Cloning of pMT1.61-4

There seemed to be two potential explanations for the failure to progress further upstream using three distinct libraries: (1) a region in the expression locus between the 5' limit of the pMT1.61-1 insert and the upstream BamHI site was unclonable or highly underrepresented in plasmid and lambda vectors, or (2) the ClaI or adjacent EcoRI site represented in pMT1.61-1 contained sequence which would not produce a viable clone when at one terminus of the insert. In order to avoid these two potential difficulties, it was decided to attempt the cloning of the 2.4 Kb HindIII fragment immediately upstream of the cDNA hybridizing region on the 1.61 telomere, extending 900 bp beyond the 5' limit of pMT1.61-1 (see Figure 4.3.4, line D). The methodology used was again similar to that used in the previous size enriched library constructions; the size selected fraction (Figure 4.10) of a HindIII digestion of 1.22j' DNA was ligated into pUC19 and transformed into E.coli HB101. A total of 4604 recombinants were recovered, 430 of which were picked into duplicate stabs and hybridized with the probe 4B of pMT1.61-1. After washing to high stringency (0.1xSSC, 65^{o}C), 2 positives were identified. Small scale plasmid DNA preparations were performed on these, which revealed that both clones were the same, and the restriction map of one was determined from digestion of a large scale plasmid DNA preparation (Figure 4.11.1). This map conformed to that predicted for the desired fragment (compare Figures 4.11.2 and 4.3.4); there is a 500 bp EcoRI/XbaI fragment, a 1.0 Kb EcoRI/HindIII fragment, and both a downstream 1.5 Kb and upstream 900 bp XbaI/HindIII fragment. Thus, the plasmid clone appeared to be derived from the 1.61 telomere and was named pMT1.61-4.

In order to confirm the reality of this insert and to

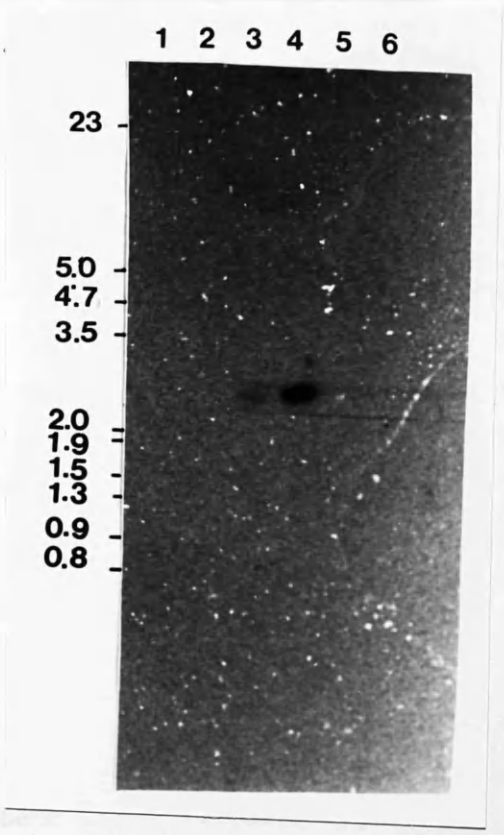
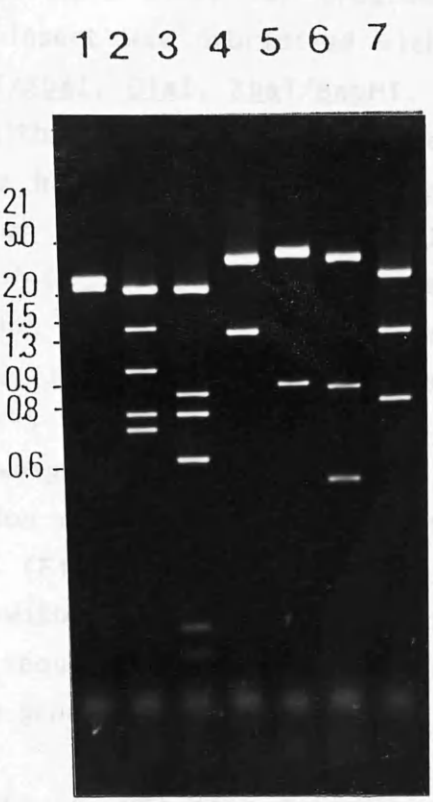


Figure 4.10.1; 1.22j' DNA digested with HindIII, size fractionated and 1/10 of the eluted DNA fractions hybridized with probe 4B. 0.1xSSC, 65°C,



4.11.1

4.11.2

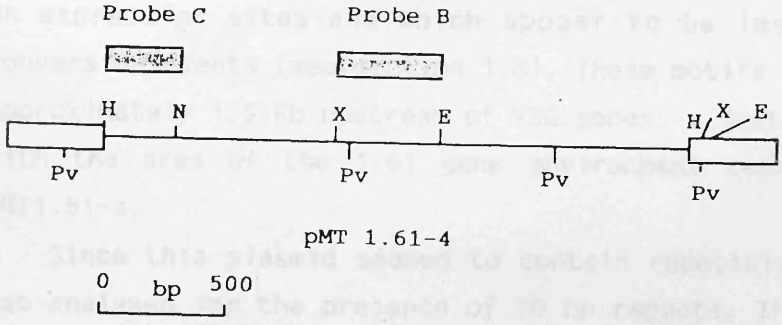


Figure 4.11.1; Mapping gel for pMT1.61-4. Lanes are, respectively, (1) HindIII (2) PvuII (3) HindIII/PvuII (4) XbaI (5) EcoRI (6) EcoRI/XbaI (7) HindIII/XbaI.
Figure 4.11.2; Derived restriction map of pMT1.61-4. The indicated probes are described in the text. Pv=PvuII, H=HindIII, N=NaeI, X=XbaI, E=EcoRI.

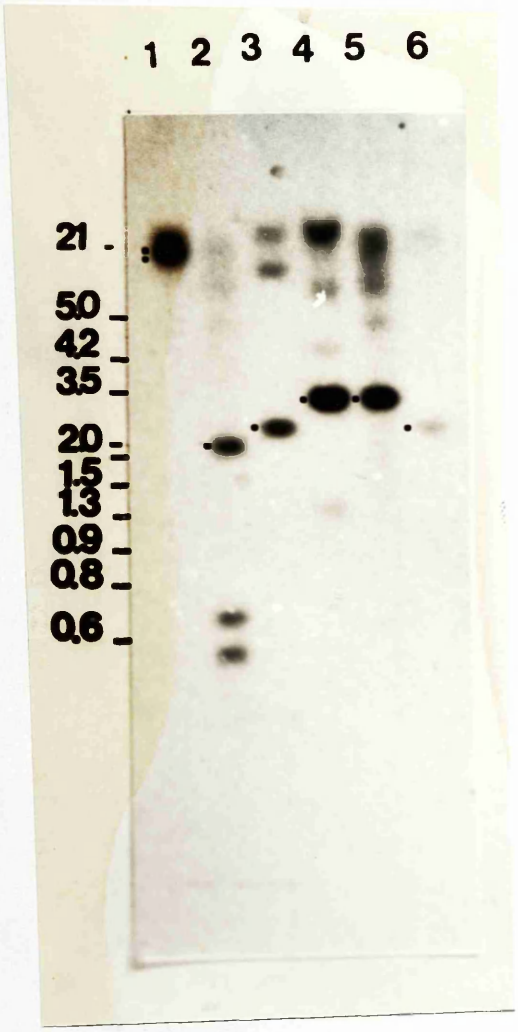
identify suitable sites for progressing further upstream, the pMT1.61-4 insert was hybridized with 1.22j' DNA digested with PstI, ClaI/XbaI, ClaI, XbaI/BamHI, XbaI and HindIII (Figure 4.12.1). Although the upstream fragments hybridize quite poorly (the probe had been made in two parts, with the specific labelling of the 900 bp HindIII/XbaI fragment being much less than that of the 1.5 kb fragment) these digests revealed specific hybridization to bands of the predicted size (marked with dots on Figure 4.12.1). Additionally, the fragments cross reacting with the pMT1.61-3 insert are detected, as are a number of other high molecular weight bands. These appear as a smear on a distinct hybridization of the pMT1.61-4 insert with ClaI/EcoRI digested 1.22j' DNA (Figure 4.12.2; dots indicate those bands fully homologous with the probe used). This suggested that pMT1.61-4 contained sequence reiterated extensively elsewhere in the trypanosome genome.

4.9 Analysis of pMT1.61-4 for the presence of 70 bp repeat sequence

Many VSG gene loci contain a repeat sequence approximately 70 bp in length, which can form long restriction site barren regions on expression sites and which appear to be involved in gene conversion events (see section 1.8). These motifs are often found approximately 1.5 Kb upstream of VSG genes; a distance comparable with the area of the 1.61 gene environment represented within pMT1.61-4.

Since this plasmid seemed to contain repetitive sequence, it was analysed for the presence of 70 bp repeats. This was done by hybridizing a PstI/ TaqI fragment of pTg 221-1 (Bernards et al., 1985), containing an extensive array of 70 bp repeats, to pMT1.61-4 doubly digested with HindIII and XbaI. This revealed that, indeed, pMT1.61-4 does hybridize with the repeats over the 900 bp XbaI-HindIII fragment entirely 5' to the region represented on pMT1.61-1 (Figure 4.13).

4.12.1



4.12.2

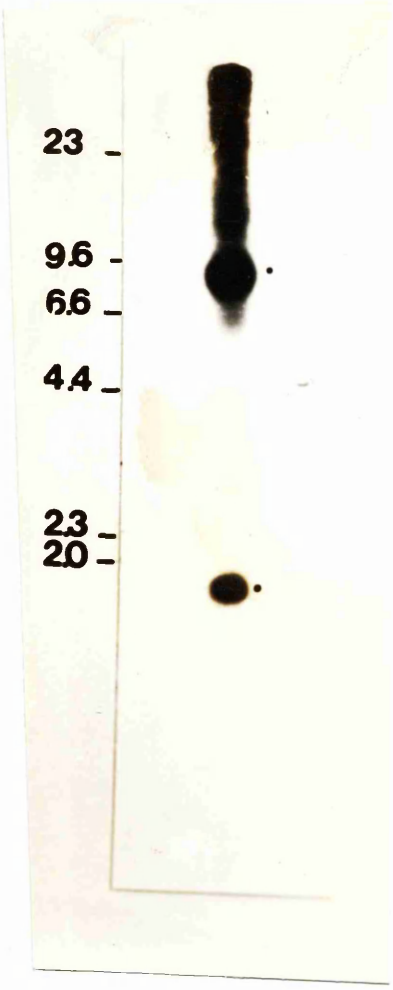
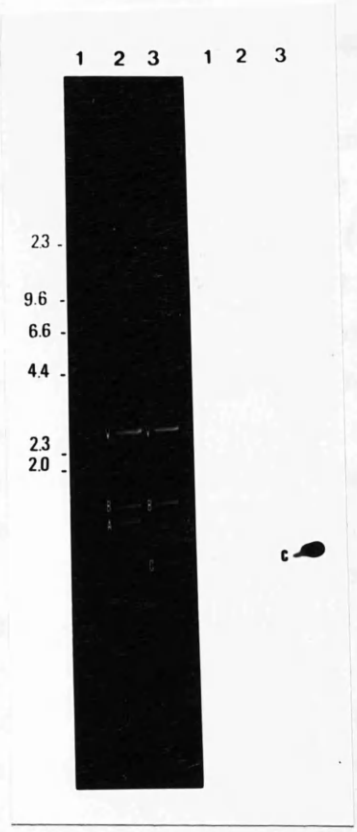


Figure 4.12.1; 1.22j' DNA digested with PstI (lane 1), ClaI/XbaI (lane 2), ClaI (lane 3), XbaI/BamHI (lane 4), XbaI (lane 5) or HindIII (lane 6) and hybridized with the pMT1.61-4 insert. Dots indicate those bands derived from the 1.61 loci. Fragments running upstream of pMT1.61-4 are not marked, as they hybridize weakly and are not easily distinguishable from cross reacting bands.0.1xSSC, 65^{o}C.

Figure 4.12.2; ClaI/EcoRI digested 1.22j' DNA hybridized with the pMT1.61-4 insert. Dots mark those bands fully homologous to the probe, as predicted from Figure 4.3.4. 0.1xSSC, 65^{o}C.

Facing page 89



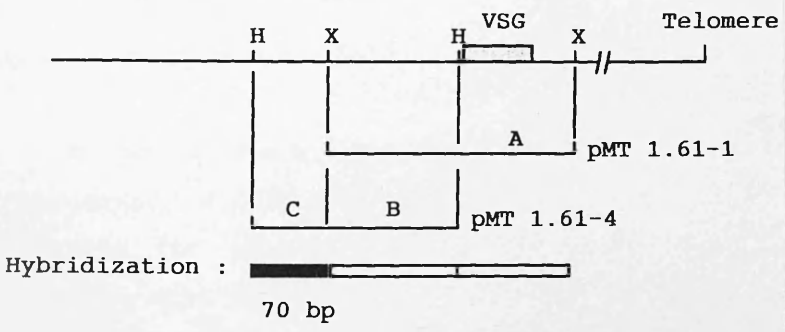


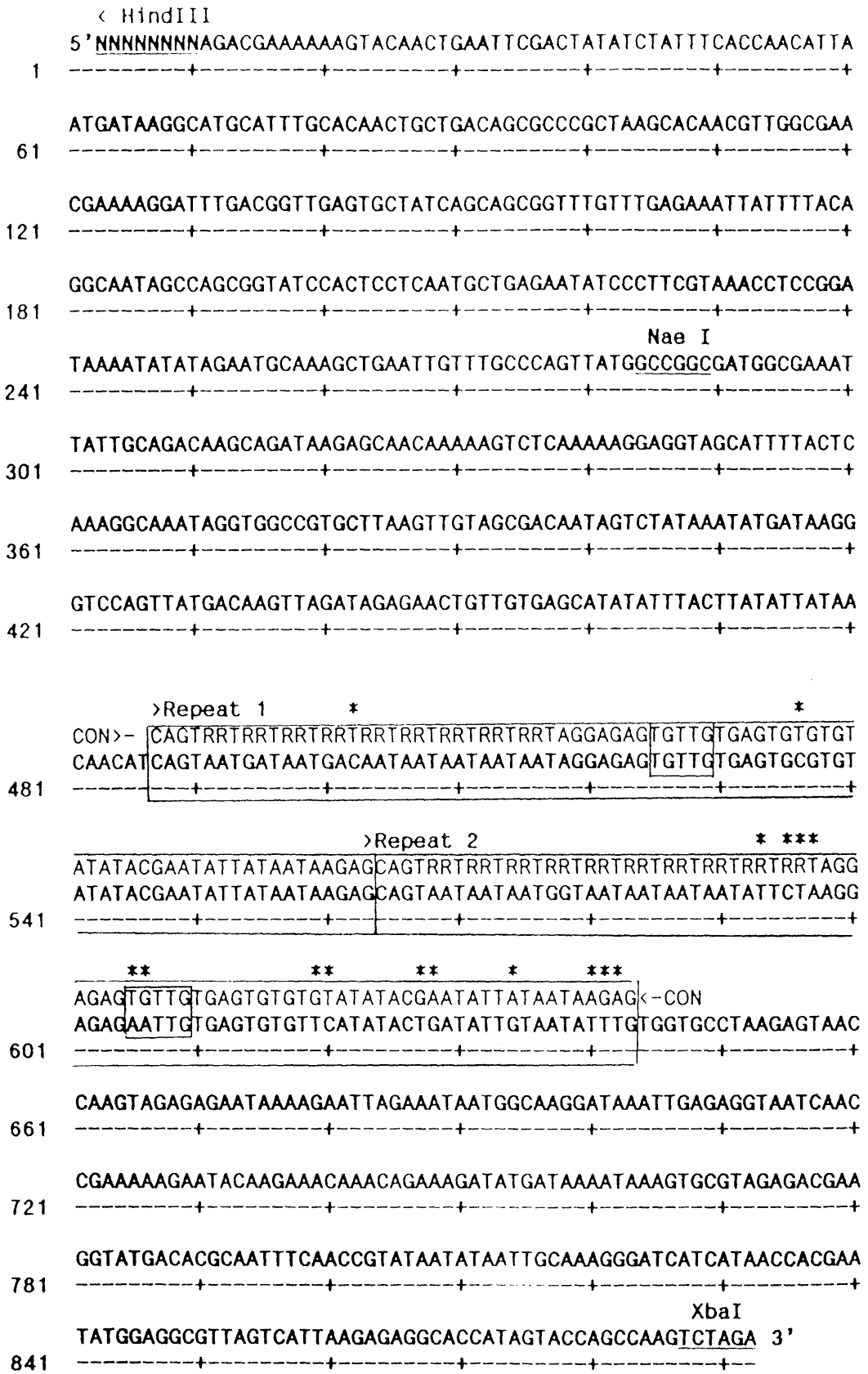
Figure 4.13; pMT 1.61-4 contains 70bp repeats. Lambda DNA size marker (lane 1), XbaI/HindIII digested pMT1.61-1 (lane 2) and pMT 1.61-4 (lane 3) hybridized with 70bp repeats from the VSG 221 locus (Bernards et al., 1985). 0.1xSSC, 65^{o}C. The line drawing below summarizes the results, with H=HindIII, X=XbaI. V=Vector band.

4.10 Sequence of the 70 bp repeat containing region on pMT1.61-4

The 900bp HindIII-XbaI fragment of pMT1.61-4 was subcloned into bacteriophage M13mp19 and 18 and sequenced (Figure 4.14). This revealed that there were two full copies of the 70 bp repeat sequence located 250 bp 5' to the XbaI site on pMT1.61-4. Although the sequence for the majority of the fragment has only been determined from a single strand, and the 5' HindIII site was not identified (this area had run off the gel, being very close to the primer annealing site), the 70 bp repeat region sequence has been determined on two strands, and using independent clones. The repeats are relatively conserved with respect to the consensus sequence, although a number of the most conserved features are disrupted. Firstly, neither repeat has an uninterrupted (T Purine Purine) sequence; in each there are T or C nucleotides at unusual positions. Secondly, only one of the repeats has an intact TGTTG (boxed in Figure 4.14), the most conserved feature of all characterized 70 bp repeats (Aline et al. 1985; Shah et al., 1987).

4.11 Cloning of pMT1.61-5

Although there were only 2 copies of the 70 bp repeats represented within pMT1.61-4, it was felt that their potential capacity for secondary structure formation (Liu et al., 1983) might be contributing to the failure to clone fragments spanning this region. It was decided, therefore, to attempt to clone a restriction fragment whose 3' terminus was upstream of the repeats found within pMT1.61-4. By sequencing from the 5' HindIII site on pMT1.61-4, a NaeI/HindIII fragment of approximately 275 bp (probe 4C; the position of the NaeI site is shown in Figure 4.14) had been identified which appeared non-repetitive in the genome and detected a 2.7 Kb NaeI fragment, extending 2.45 Kb upstream of the 5' limit of pMT1.61-4 (Figure 4.15.1). In order to clone this fragment, therefore, 50ug of 1.22j' DNA was digested with NaeI and fragments of this size isolated from the gel. Fraction 3, which contained the desired fragment, as



Facing page 91

Figure 4.14; Nucleotide sequence of the 900bp HindIII/XbaI fragment of pMT1.61-4. The sequence does not fully extend to the HindIII site as is the string of "Ns". R=Purine nucleotide. Asterisks indicate divergence of the 70bp motif sequence from the consensus of Aline et al. (1985a). The conserved element, TGTTG, is boxed.

4.15.1 4.15.2 4.15.3

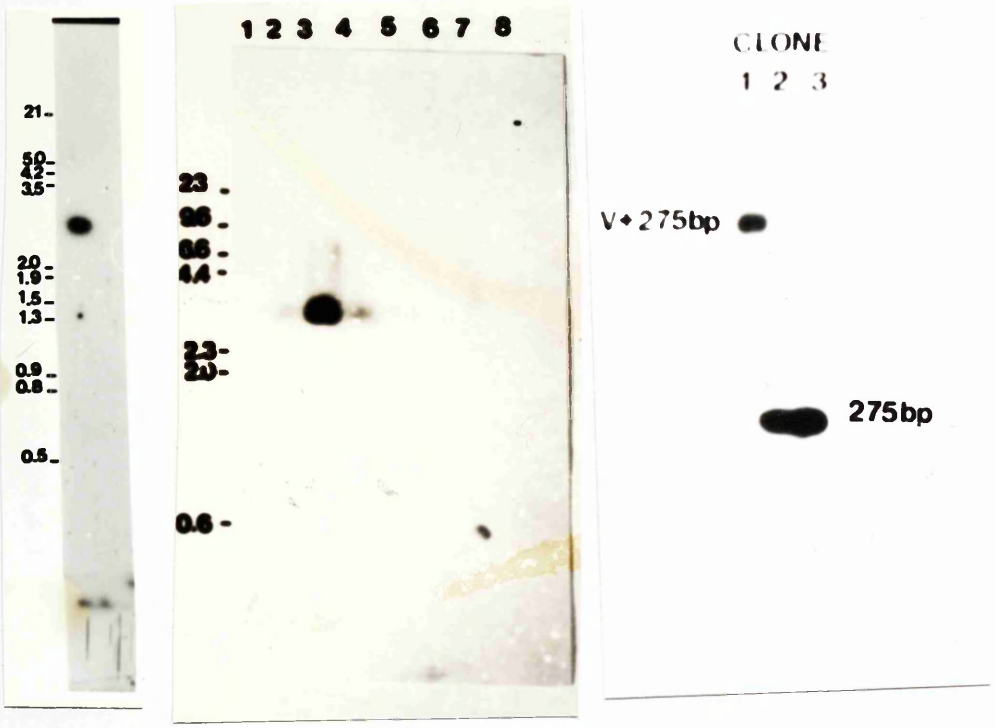
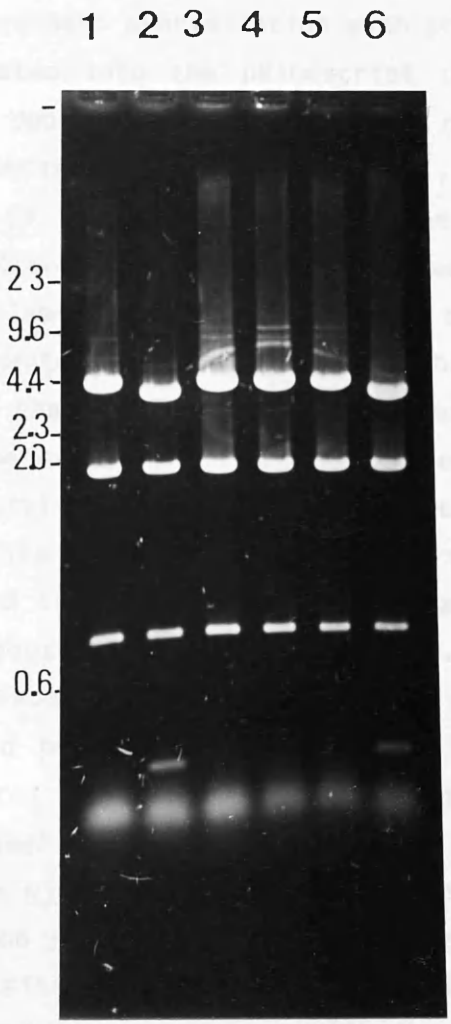


Figure 4.15.1; Probe 4C of pMT1.61-4 hybridized with NaeI digested 1.22j' DNA. A 2.7kb fragment is detected. 0.1xSSC, 65°C.
Figure 4.15.2; NaeI digested 1.22j' DNA size fractionated and 1/10 of these fractions hybridized with probe 4C. 0.1xSSC, 65°C.
Figure 4.15.3; Small scale plasmid preparation of the 3 positive colonies hybridized with probe 4C to indicate that clone 1 is in the reverse orientation to clones 2 and 3: V=vector sequence. 0.1xSSC, 65°C.

Facing page 92



4.16.2

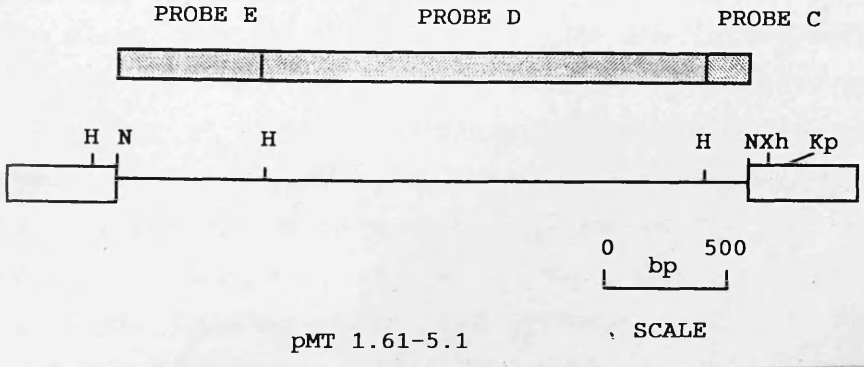


Figure 4.16.1; Mapping gel for pMT1.61-5.1. Lanes are, respectively, pMT1.61-5 digested with (1) HindIII (2) HindIII/XhoI (3) HindIII/PstI (4) HindIII/BamHI (5) HindIII/EcoRI (6) HindIII/KpnI.

Figure 4.16.2; Restriction map for pMT1.61-5.1. H=HindIII, N=NaeI, Xh=XhoI, Kp=KpnI. The probes are described in the text.

assessed by Southern hybridization with probe 4C (Figure 4.15.2), was then ligated into the pBluescript plasmid vector digested with HincII. Upon transformation, 2600 colonies transformed to ampicillin resistance were recovered, of which 1312 were recombinant. Of these, 450 were screened with probe 4C which detected 3 strongly hybridizing colonies, at high stringency. Small scale plasmid DNA preparation of these positive colonies were then digested with HindIII and hybridized with probe 4C. This revealed that all three had the same insert, clone 1 being in the reverse orientation to clones 2 and 3; the HindIII site within the polylinker sequence of the vector being adjacent to the HindIII site represented at the extremity of pMT1.61-4 in the latter and therefore produced a fragment of around 275 bp, fully homologous to probe 4C (Figure 4.15.3). A large scale plasmid DNA preparation was then carried out on clone 1 and the plasmid, named pMT1.61-5, mapped with a number of restriction enzymes (Figures 4.16.1 and 2). Although the insert could not be mapped with NaeI directly, as this site was lost by blunt end ligation to a HincII cut vector, by using restriction sites adjacent to the HincII site in the vector polylinker, distal to the HindIII site, there was seen to be a 275 bp XhoI (= NaeI)/HindIII fragment, and an absence of cleavage sites for PstI, EcoRI and BamHI. This conformed to the predicted map for this clone (compare Figure 4.16.2 with the corresponding region on Figure 4.3.4). The clone was also hybridized onto genomic DNA to ensure that it detected bands of the sizes predicted from mapping of the plasmid and previous genomic mapping (Figure 4.17). This was the case; hybridization of the 1900 bp HindIII fragment of pMT1.61-5 (probe 4D) to HindIII, PstI, XbaI and XbaI/BamHI digested 1.22j' DNA produced bands of 1900bp, the telomeric fragments, 7400bp and 4200 bp, repectively. These fragments were all of the predicted size (although the 7400 bp XbaI had not been previously mapped).

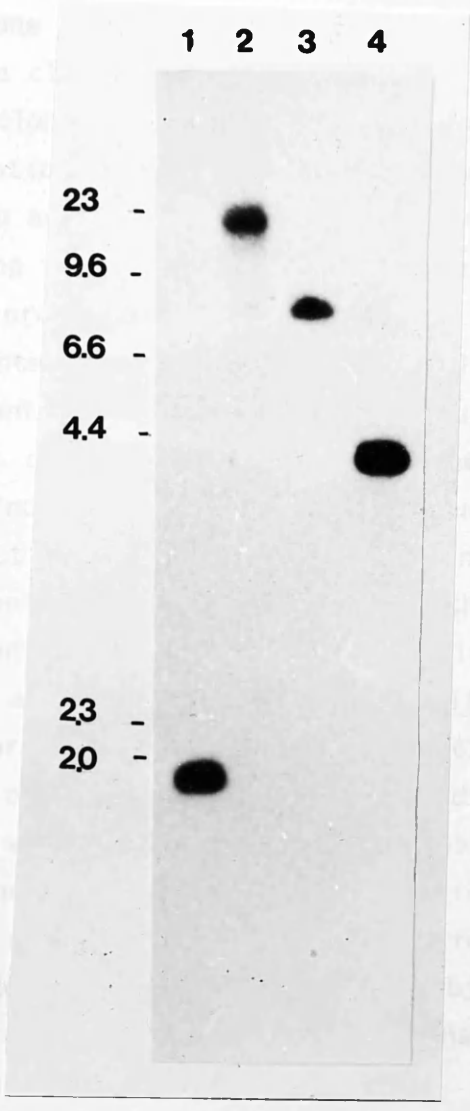


Figure 4.17; Genomic confirmation of pMT1.61-5. 1.22j' digested with HindIII (lane 1), PstI (lane 2), XbaI (lane 3) or XbaI/BamHI (lane 4), hybridized with probe 4D. 0.1xSSC, 65^{o}C.

4.12 Cloning of pMT 1.61-6

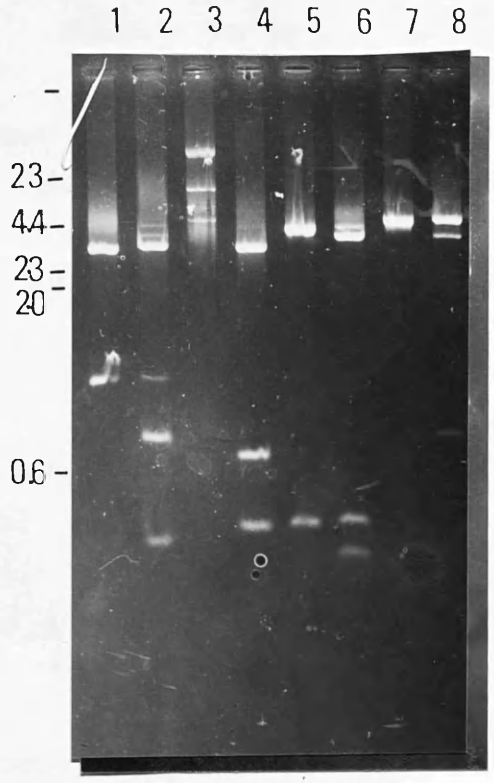
The only restriction fragment available for walking 5' of pMT1.61-5 was a 900bp HindIII fragment, extending just 300 bp upstream (not shown). Although short, it was decided to attempt to clone this nearby fragment to allow it to be used as a probe in the cloning of the next NaeI fragment on the telomere, 5' to that cloned in pMT1.61-5. Therefore, 1.22j' DNA was digested to completion with HindIII and fragments in the size range between 600 bp and 1200 bp excised and purified by extraction using low melting temperature agarose, rather than "Geneclean" as had been used previously. This was used because problems had been encountered when attempting to ligate "Geneclean" purified fragments at this time. It was thought that these problems were a result of fines of "glassmilk" remaining with the purified DNA and binding it during low temperature ligation.

Fraction 4, assessed to contain the desired 900bp HindIII fragment after hybridization of the 600bp HindIII/XhoI (=NaeI) fragment of pMT1.61-5 (probe 4E; Figure 4.18.1), was ligated into pUC19 and used to transform E.coli DH5-alpha, resulting in the recovery of 944 colonies of which 622 were recombinant. Duplicate stabs of these colonies were hybridized with Probe 4E of pMT1.61-5 and washed to a stringency of 65^{o}C, 0.1x SSC. After exposure, just one hybridizing colony was identified upon which a full scale plasmid preparation was performed, using the "Circleprep" procedure. The clone was then subjected to restriction mapping (Figure 4.18.2 and 3) to ensure that the expected fragments were present. This was the case.

The insert of this plasmid, termed pMT1.61-6, was also hybridized to genomic DNA digested with various enzymes to ensure it detected the correct genomic fragments. Figure 4.19 demonstrates that whilst this was so, the insert also hybridized to a cross reacting fragment. This is marked for each lane in Figure 4.19 by an arrowhead.

4.18.1

4.18.2



4.18.3

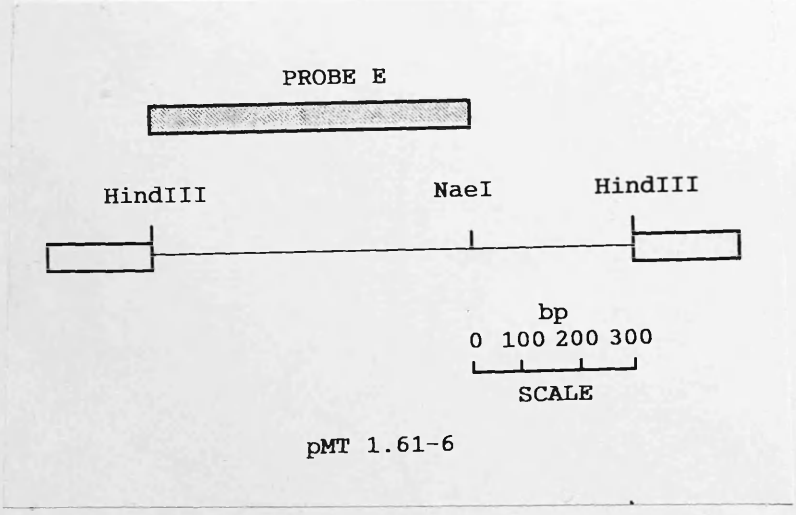


Figure 4.18.1; HindIII digested 1.22j' DNA size fractionated and 1/10 hybridized with probe 4E. 0.1xSSC, 65°C.
Figure 4.18.2 Mapping gel for pMT1.61-6. Lanes are, respectively, pMT1.61-6 digested with HindIII, NaeI/HindIII, NaeI, HincII/HindIII, HincII, HincII/NaeI, BamHI and BamHI/NaeI.
Figure 4.18.3; Restriction map for pMT1.61-6. HincII sites are not shown here, having not been mapped for other 1.61 loci derived clones.

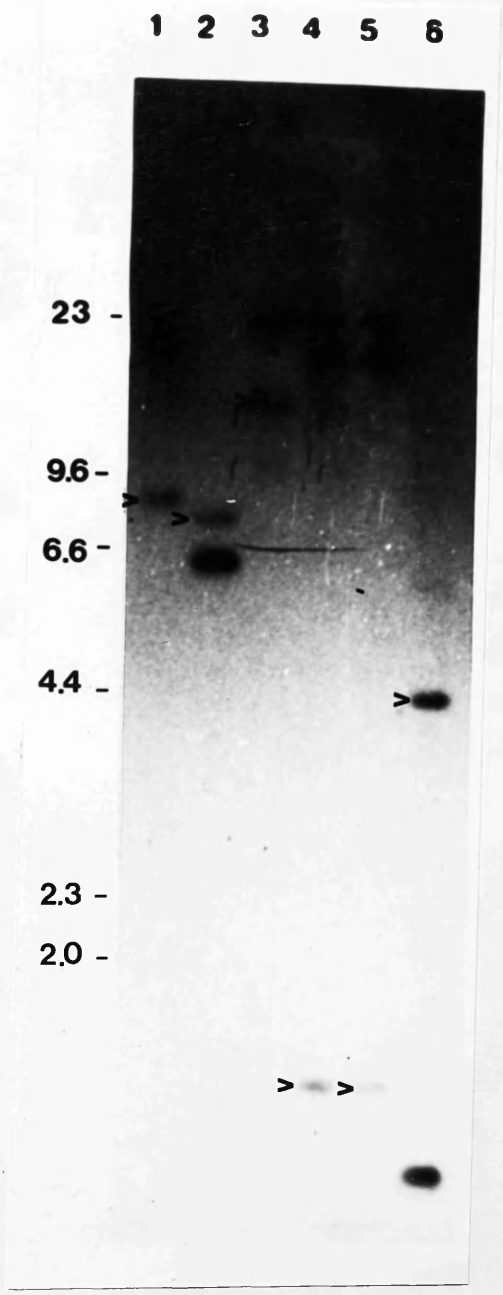


Figure 4.19; 1.22j' DNA digested with PstI (lane 1), PstI/EcoRI (lane 2), EcoRI (lane 3), PstI/BamHI (lane 4), BamHI (lane 5) and HindIII (lane 6) and hybridized with the pMT1.61-6 insert. 0.1xSSC, 65^{o}C. Arrowheads indicate bands cross reacting with the probe.

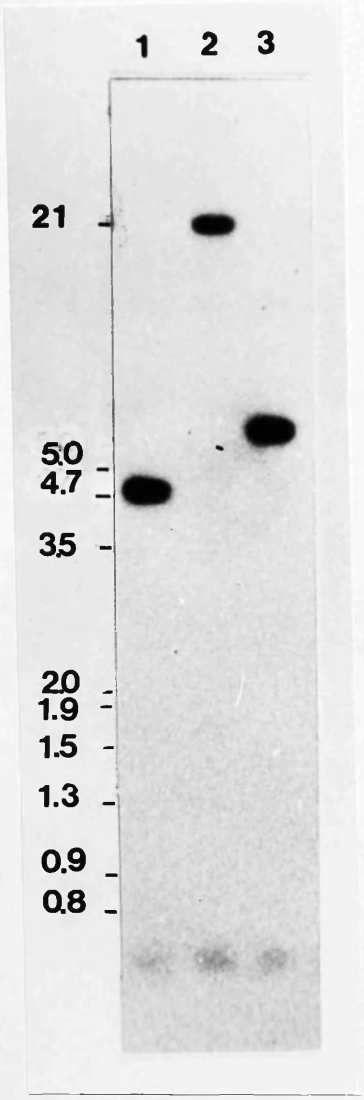


Figure 4.20; 1.22j'trypanosome genomic DNA digested with BamHI/EcoRI (lane 1), EcoRI (lane 2) and PstI/EcoRI (lane 3) and hybridized with probe 4B. There is no evidence of divergence between the loci containing the two ILTat 1.61 gene copies. The weak band detected at approximately 700bp represents cross rection of probe 4B with the fragment cloned by mistake as pMT1.61-3 (see section 4.5).

Facing page 97

4.13 Mapping of the predicted basic copy loci divergence

The pre-existing map for the ILTat 1.61 gene loci constructed by Cornelissen et al. (1985a) had predicted the presence of an insertion in one locus downstream of the first BamHI site from the end of the telomere (shown in Figure 4.1). Because the telomere had been cloned to within 400bp of this site and no divergence had been yet detected, it was decided to analyse genomic DNA for the presence of this insertion, using a probe upstream of the cDNA hybridizing region. Therefore, 1.22j' DNA was digested with BamHI/EcoRI, EcoRI, and PstI/EcoRI and hybridized with probe 4B, derived from pMT1.61-1. The resulting autoradiograph (Figure 4.20) clearly detects no such insertion in the 1.61 loci either downstream of the BamHI site, as had been predicted, or within the long EcoRI fragment of approximately 22kb, extending upstream from within the area represented in pMT1.61-1.

4.14 Cloning Upstream of pMT1.61-6

In an analysis of the bloodstream expression loci for the ILTat 1.22 gene, a lambda library had been constructed from a complete EcoRI digestion of EATRO 795 DNA at a time coincident with this stage in the 1.61 expression site cloning. This library construction is detailed in section 5.9. I decided to screen this library for the presence of fragments derived from the 1.61 gene expression loci because it was known these contained an EcoRI fragment of a size compatible with the insert range of lambda Dash (approximately 22 kb, see Figure 4.20), with its 3' terminus represented within pMT1.61-1. Although this area had proved to be absent in plasmid libraries and in lambda libraries containing Sau3A partial digestion products, it was thought possible that a library constructed from complete digestion products in low concentration might produce a far higher representation of the desired fragment, given the vector's inherent selection for fragments greater than 10 kb in length (Frischauf et al., 1983).

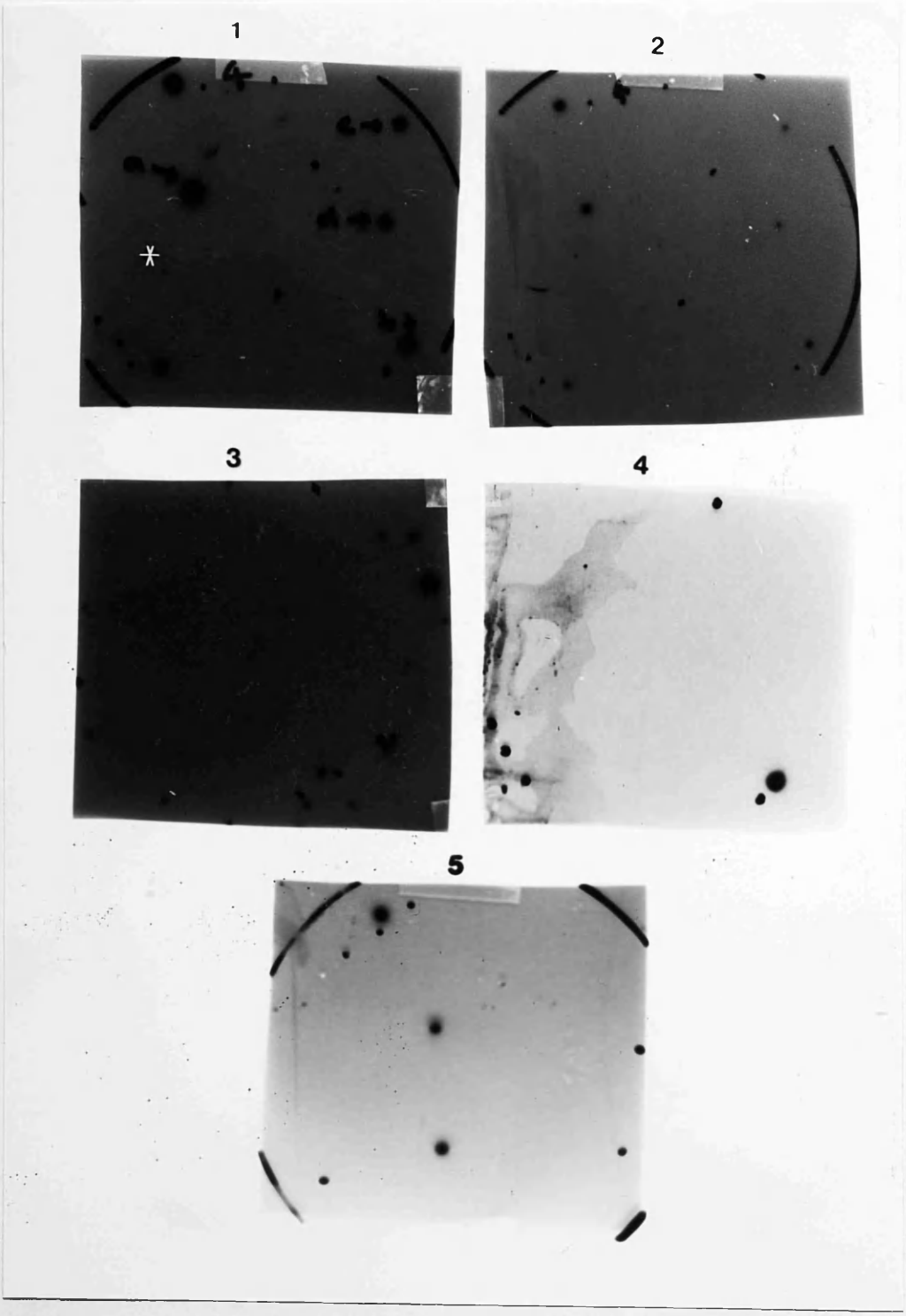


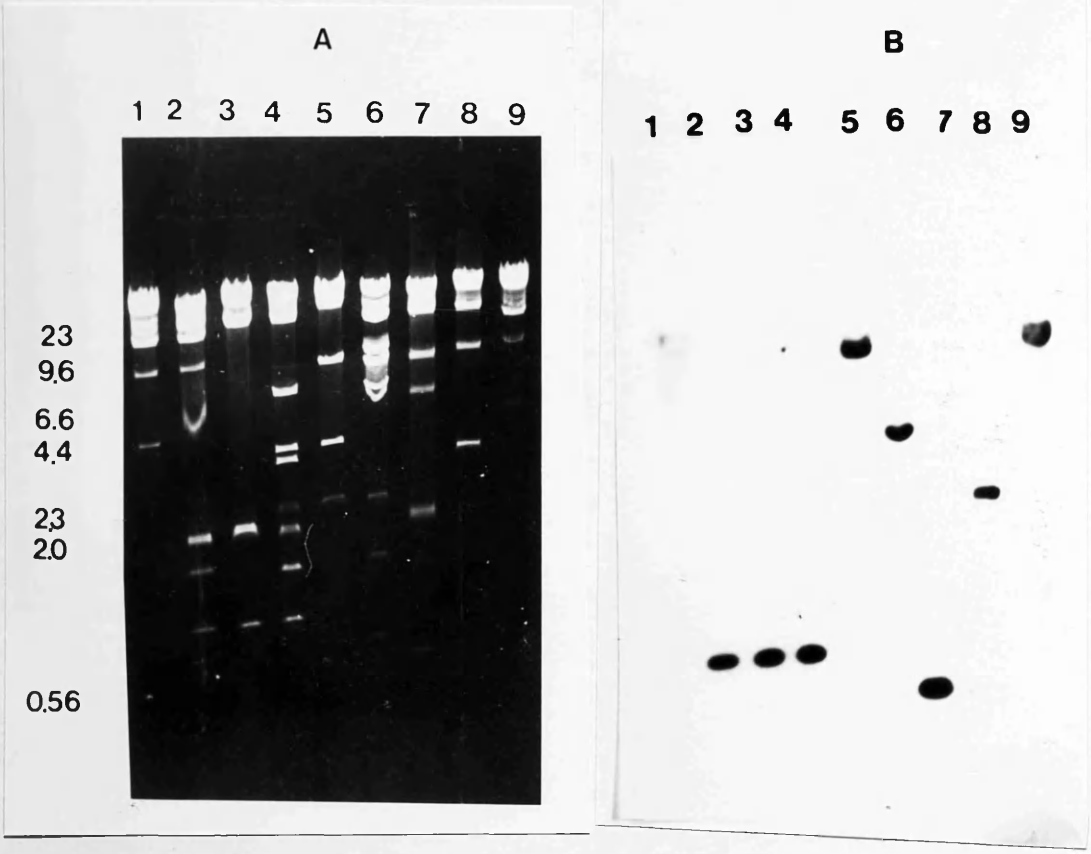
Figure 4.21; Evidence for instability of the isolated lambda clones. The respective panels are described in the text. The asterisk in panel 1 indicates a plaque hybridizing weakly with the probe. The letters indicate plaques which were selected for purification. All hybridizations were washed to a stringency of 0.1xSSC, 65^{o}C.

Facing page 98

Furthermore, the isolation of pMT1.61-4,5 and 6 had provided probes which could, in combination, distinguish cross reacting fragments from true 1.61 telomere derived fragments.

Phage from the library construction were plated onto the spi selective host strain E.coli P2392, producing 60,000 plaques over two $20cm^2$ BBL plates. spi selection operates in hosts lysogenic for phage P2 and relies on the replacement of the lambda gam gene with inserted foreign DNA. Phage which are gam^- (i.e recombinants) grow in these hosts while gam^+ phage (non recombinants) do not, thereby ensuring that all plaques are recombinant in a single plating step. This selection was necessary here because titration of the library on spi selective and non selective hosts had indicated the library contained just 10% recombinants.

Duplicate lifts of the plated recombinant phage plaques onto Nytran membrane were hybridized with the 1900bp HindIII fragment of pMT161.5 (probe 4D), which was known to detect no fragments in the trypanosome genome, other than those derived from the 1.61 basic copy loci (see Figure 4.17). After washing to a stringency of 65^oC, 0.1xSSC, a total of 8 plaques were identified which hybridized intensely in duplicate, in addition to a number of more weakly hybridizing plaques. Four of these plaques were picked into SM buffer and transfected into E.coli NM621, which is more tolerant of repetitive DNA sequence than the former host, P2392, which is recombination competent ($recA^+$). The resultant plaques were then hybridized with two probes from distict areas on the 1.61 telomere, both upstream and downstream from the 1900 HindIII fragment: probe 4E from pMT1.61-5 (see Figure 4.16.2) and probe 4C from pMT1.61-4 (see Figure 4.11.2). When washed to high stringency, both probes produced duplicate hybridization patterns (Figure 4.21.1 and 2), although probe 4C, having been labelled to lower specific activity, gave a less intense hybridization. In total, approximately 30% of the plaques on each plate hybridized, with the remainder hybridizing either very weakly (marked by an asterisk in Figure 4.21.1) or not at all. To achieve plaque purity, well isolated plaques were picked and replated. Hybridization of these plaques revealed that, once



Facing page 99

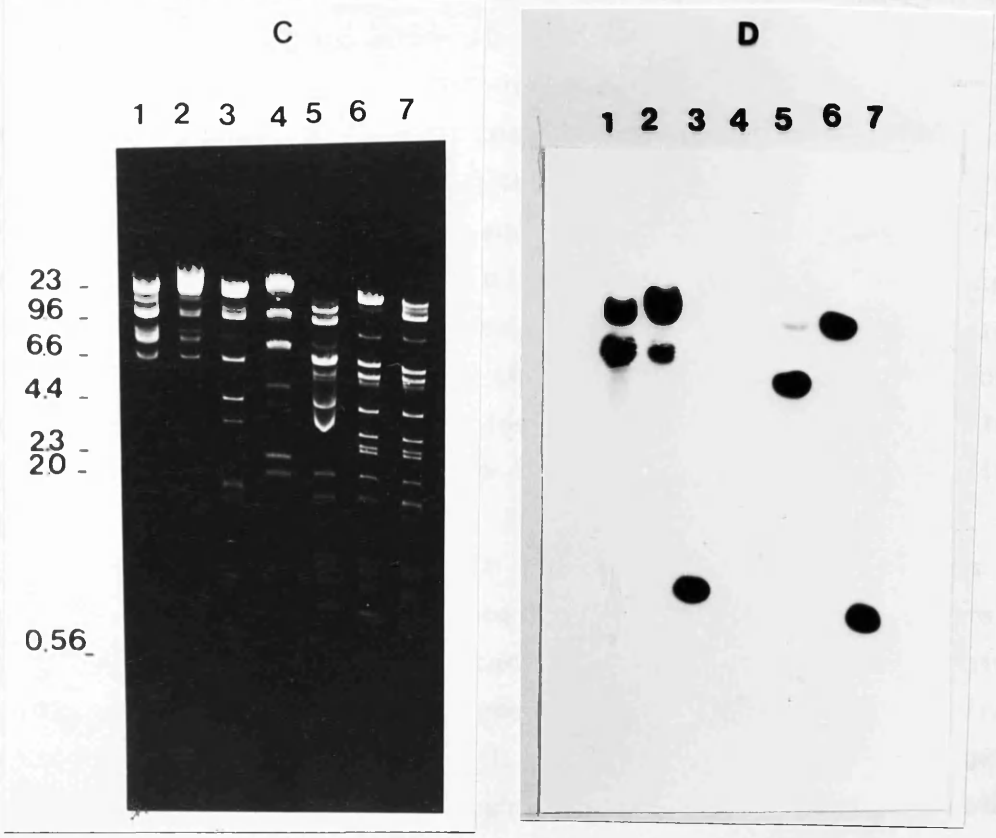


Figure 4.22; A: Lanes 1-7, A large scale DNA preparation of one isolated lambda clone digested with, repectively, EcoRI, HindIII/EcoRI, HindIII, BamHI/HindIII, BamHI, BamHI/EcoRI, NaeI/HindIII, NaeI, NaeI/EcoRI
C: Lanes 1-3 and 5-7, the same DNA digested, respectively, with EcoRI/XbaI, XbaI, XbaI/HindIII, PstI/EcoRI, PstI and PstI/HindIII. Lane 4 contains lambda DNA size markers.
B: The gel in panel A hybridized with probe 4D. 0.1xSSC, 65^{o}C.
D: The gel in panel C hybridized with probe 4D. 0.1xSSC, 65^{o}C.

again, only a small percentage of plaques hybridized strongly, with many hybridizing only weakly or not at all (Figure 4.21.3 is representative). With a third plaque purification step, it was found that all plaques, except one, no longer hybridized with the 1.61 telomere derived probe 4D (Figure 4.21.4). This, together with the previous difficulties in trying to clone this region by other approaches, suggested that the clone might be prone to deletion of its insert, or of the probe-hybridizing region. The one remaining hybridizing plaque, when replated, produced 3 plaques all of which hybridized, albeit weakly, perhaps indicating that this contained a somewhat more stable phage (Figure 4.21.5). A large scale phage DNA preparation of the clone was, therefore, carried out using a large initial phage inoculum to minimise culture lysis time in order to limit the possibility of any loss of stability.

In order to determine if the hybridizing clone was a derivative of the 1.61 telomere and whether it was extensively rearranged, its DNA was digested with a number of enzymes which would produce restriction fragments of a size predictable from existing genomic maps of the 1.61 telomere. The resulting gels (Figure 4.22, panels A and C) showed evidence of either partial digestion or deletion. Most notably, the EcoRI digested lane (Figure 4.22A, lane 1) not only produced a 22 Kb band (somewhat obscured due to the amount of DNA loaded onto this gel), but also bands running at approximately 4.4 kb and 9 kb. These fragments cannot be explained by either the fragments derived from the lambda arms or lambda insert and may, therefore, have represented either extensively deleted or rearranged versions of these. The number of unexpected bands appeared too few for them to be a result of EcoRI star activity. Other lanes also show similar evidence of either partial digestion or limited insert deletion, with several bands of anomalous intensity being present. In order to determine which of these two possibilities was responsible, the gel was blotted and hybridized with probe 4D from pMT1.61-5. The resultant autoradiograph (Figure 4.22, panels B and D) demonstrated that, in all enzyme digestions, the fragments expected from genomic mapping hybridized and that there was no

evidence of partial digestion, excepting with NaeI/EcoRI and XbaI digestion (lanes 1 and 2, panel D). Thus, the bands of anomalous intensity appear to be derived from deleted or rearranged phage which retain the more downstream regions of the insert intact (no unexpected bands hybridize with probe 4D).

To minimise further instability in the lambda clone, henceforth named lambda 1.61, and to allow an interpretable map to be constructed, it was decided to attempt to subclone fragments of the phage insert into plasmid vectors. The strategy chosen was to generate two banks of recombinant plasmids containing inserts derived from either BamHI or XbaI digestion of lambda 1.61 and then to shuttle between the two banks to allow a progressive walk along the lambda insert. This was relatively simple because of the large amount of information known about the restriction map of the 1.61 telomere derived from genomic mapping using the clones pMT1.61-1,4,5 and 6. To generate the plasmid banks, 1 ug of lambda 1.61 DNA was digested to completion with BamHI or XbaI, and ligated into either BamHI digested pBluescript or XbaI digested pUC19, both of which had been dephosphorylated by CIAP. After transformation, recombinants from the XbaI bank were hybridized with probe 4B derived from pMT1.61-1, to detect colonies harbouring plasmids with the most 3' 7.4 kb XbaI fragment (see lane 2, Figure 4.22D; mapped onto genomic DNA in Figure 4.17, lane 3).

The 5' BamHI-XbaI fragment of one of these clones was then hybridized to the BamHI bank, and a hybridizing colony detected. The plasmid from this colony was used to rescreen the XbaI bank and a futher walk achieved. In this way, the plasmids shown in Figure 4.23 were isolated and subsequently mapped.

Although these clones did not represent all the subclones obtained, the remainder were found not to hybridize to trypanosome genomic DNA, but did hybridize to lambda DNA size markers. These clones were taken to be derived from the scrambled 5' end of the lambda clone since the arms of lambda Dash contain no BamHI or XbaI sites which could allow these fragments to be cloned from an unrearranged phage.

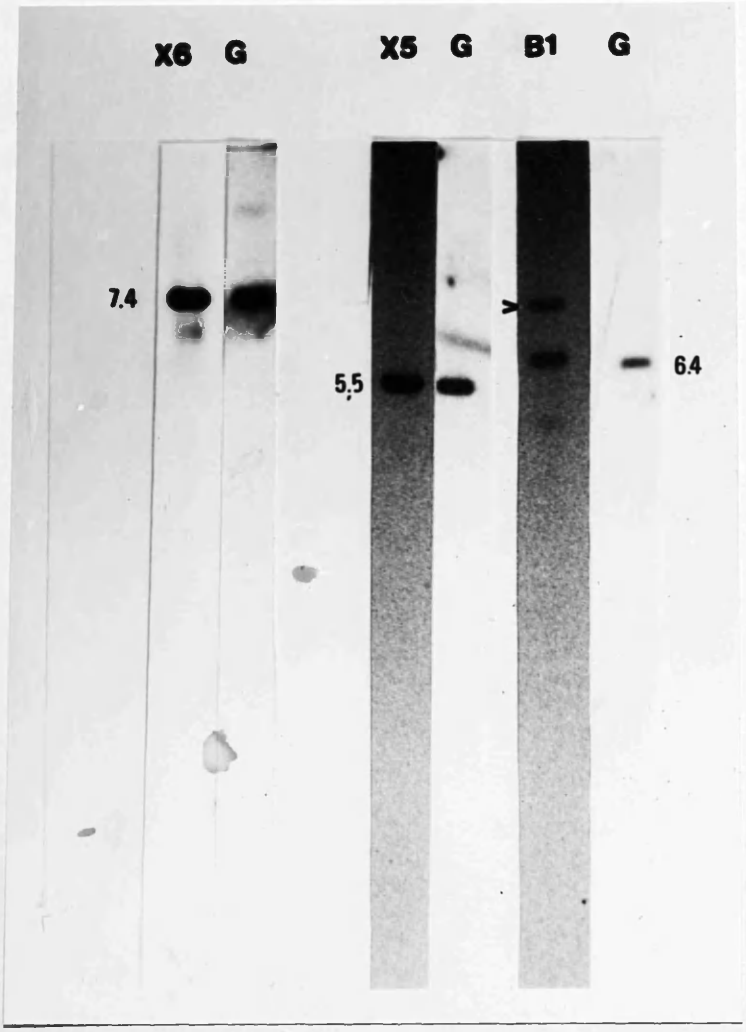


Figure 4.24; Genomic verification of the plasmid subclones of lambda 1.61, plambda X6, X5 and B1. In each case plasmid DNA is in the left hand lane, genomic DNA is in the right hand lane (lane G). Digests used and the respective probes are described in the text.
Plasmid lanes were exposed for approximately 30 minutes, Genomic lanes were exposed overnight.
The arrowhead indicates a partial digestion product of plambda B1, clearly visible on the ethidium stained gel (not shown).

Facing page 102

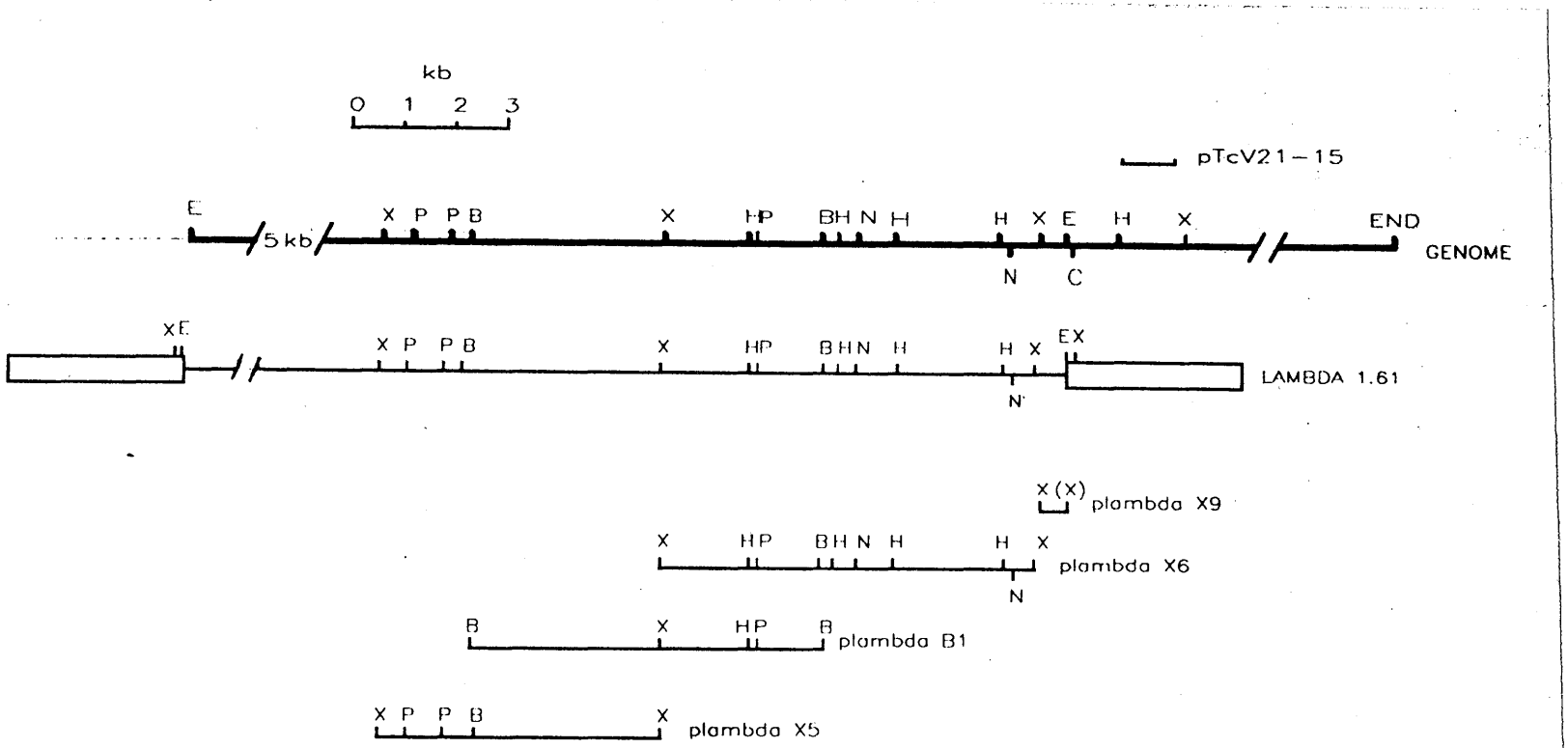


Figure 4.23; Derived subclones of lambda 1.61. E=EcoRI, X=XbaI, P=PstI, B=BamHI, H=HindIII, N=NaeI, C=ClaI. Filled boxes represent the arms of the lambda vector.

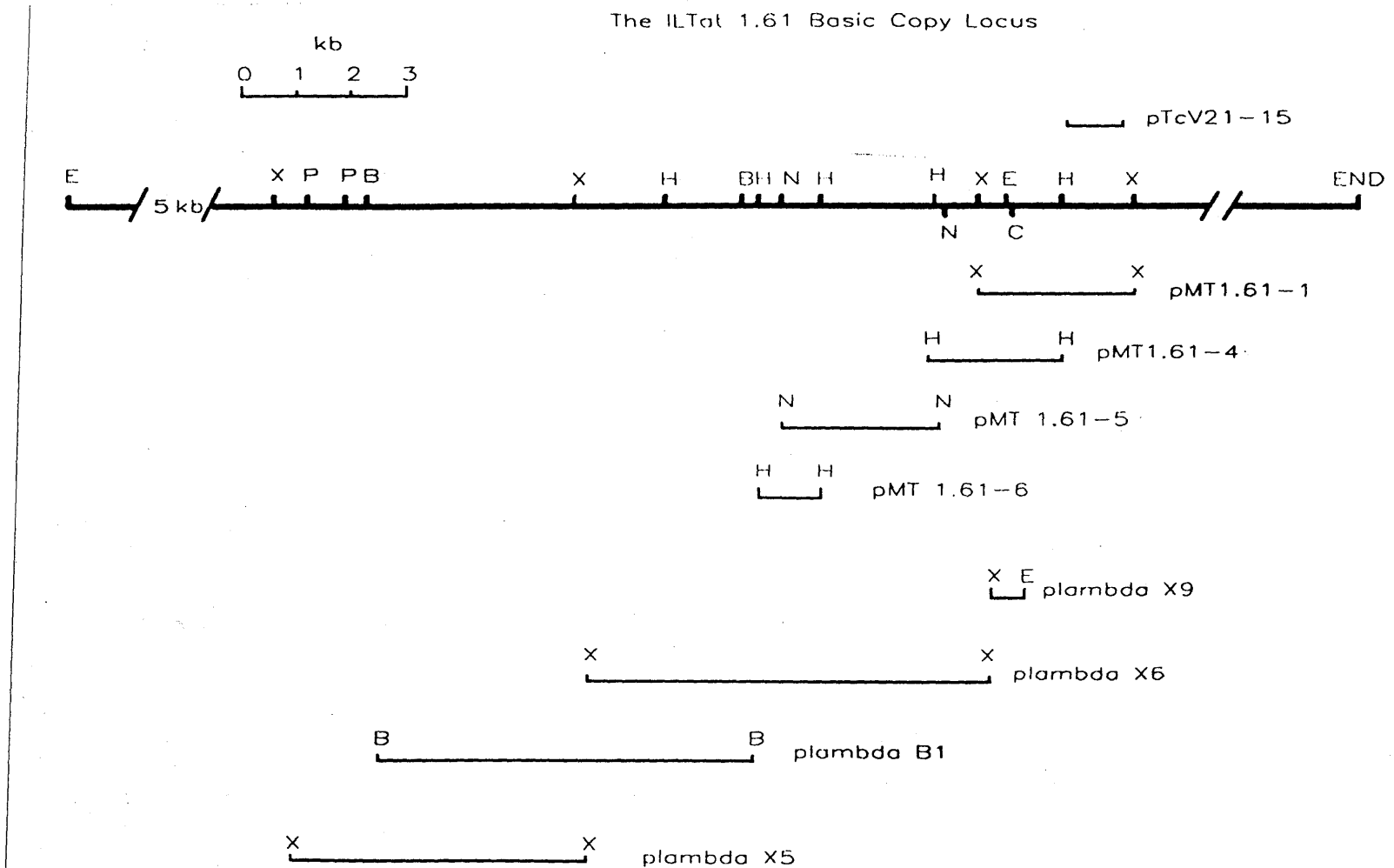


Figure 4.25; Summary of the derived clones from the ILTat 1.61 gene locus. E=EcoRI, X=XbaI, P=PstI, B=BamHI, H=HindIII, N=NaeI, C=ClaI.

Thus, in total, 13 Kb of the 22 kb EcoRI fragment cloned in lambda 1.61 had been successfully subcloned in plasmids and restriction mapped.

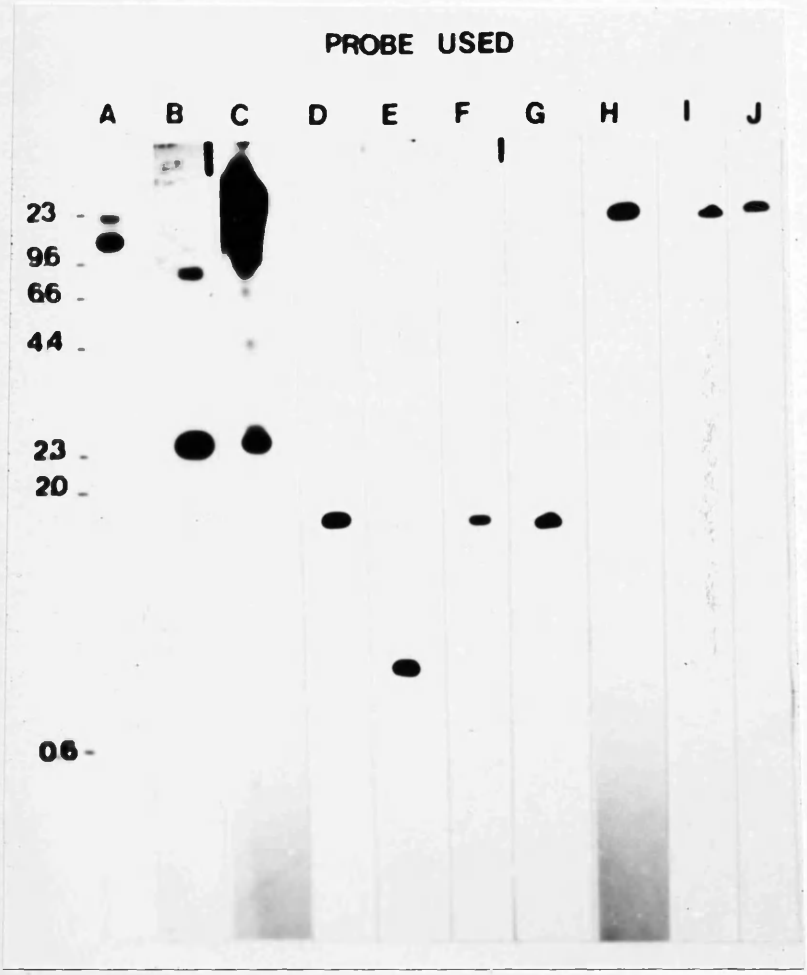
4.15 Genomic fidelity of the lambda 1.61 derived plasmid clones.

Given the observed instability of lambda 1.61, it was important to verify that the plasmids isolated from this phage were of the same size as their genomic equivalents. To test this, 1.22j' DNA was digested to completion with either XbaI or BamHI and electrophoresed on the same gel (but separated from) the plasmids plambda X6, plambda X5 and plambda B1 digested to excise their inserts. The digests were then hybridized with either probe 4D derived from pMT1.61-5 (to detect plambda X6 and its genomic equivalent), or the 3.5kb BamHI/XbaI fragment of plambda B1 (to detect both plambda B1 and X5 and their genomic equivalents). Figure 4.24 demonstrates that each plasmid clone is of an identical size to its genomic equivalent, indicating that they are probably not deleted or rearranged.

A summary of the full map of the 1.61 telomeres and the corresponding plasmid and lambda derived clones is given in Figure 4.25.

4.16 Analysis of the genomic repetitiveness of the 1.61 basic copy locus and verification of the clone origins.

Analyses of bloodstream VSG loci have revealed that the great majority of the expression site is composed of sequence which is reiterated extensively elsewhere in the genome, even when hybridizations were carried out at high stringency (Kooter et al. 1987; Kooter et al., 1988; Alexandre et al., 1988; Gibbs and Cross, 1988; Pays et al., 1989a). To determine whether the same was true of the ILTat 1.61 gene environment, all sections of the available clones were hybridized to 1.22j' genomic DNA digested with HindIII. The probes used are labelled in Figure 4.26. In



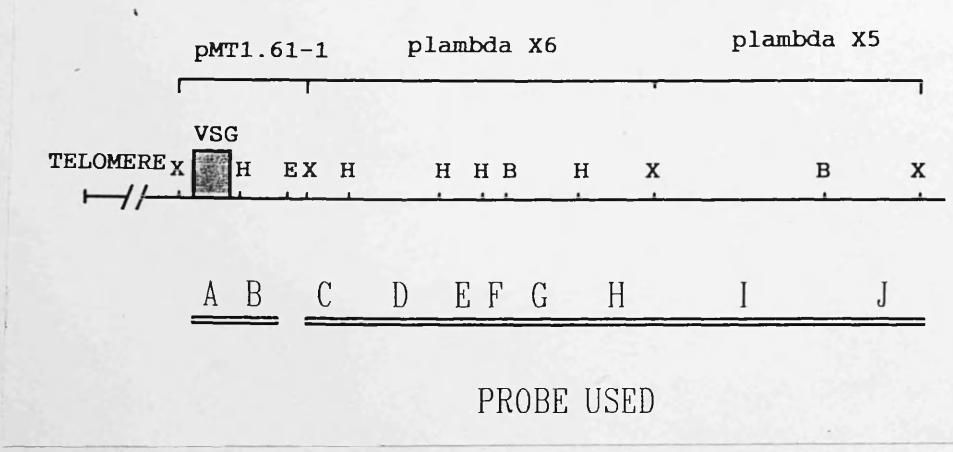


Figure 4.26; 1.22j' DNA hybridized with the probes indicated. These probes are shown below the autoradiogram. For the telomere map; X=XbaI, H=HindIII, E=EcoRI, B=BamHI. 0.1xSSC, 65^{o}C.

Facing page 104

several cases these probes have been used under different names earlier in this chapter; they are renamed here for clarity in these experiments. Figure 4.26 shows that at high stringency the majority of probes detect just a single major band of the size predicted from mapping of the clones, a result compatible with the analysis of the ILTat 1.22 locus (Shiels, 1990), but very different to the highly repetitive nature of bloodstream expression loci. One area of the telomere was clearly repetitive, however, as revealed by the hybridization of probe C. This area was expected to be reiterated in the genome; it contains the region represented within pMT1.61-4 which had previously been found to contain 70bp repeats (see sections 4.9 and 4.10).

The probes have also been hybridized to three distinct trypanosome populations. The first was 1.22j' trypanosomes, as used earlier, which are from the EATRO 795 stock and contain two basic copies of 1.61 gene (Cornelissen et al., 1985a). The other two, GUP 810 and GUP 871, had one and zero basic copies of the 1.61 genes respectively. These populations are from the EATRO 2340 stock, which normally has one copy of the 1.61 gene; the GUP 871 population lost this copy during cyclical transmission in the laboratory (Barry et al., 1983; Cornelissen et al., 1985a). DNA from these populations were hybridized with the available probes for two reasons. Firstly, it was considered important to verify that it was truly the 1.61 loci which had been cloned and not another locus, cross reacting with the probes used. If the probes were derived from the 1.61 loci, it was expected that the intensity of hybridization to 1.22j' DNA would be approximately twice that seen on GUP 810 DNA, and there would be no hybridization to the homologous fragment on the GUP 871 DNA. Secondly, it was hoped that the limit of the region of the 1.61 gene locus which had been deleted in GUP 810 and 871 might be contained within the available clones. If so, then it would be predicted that at some point along the telomere the probes used would hybridize to the DNA from all three populations with equal intensity. If identified, probes upstream of the deleted region could be employed in future studies to identify the gene, if any, which had replaced 1.61. It could be then determined whether

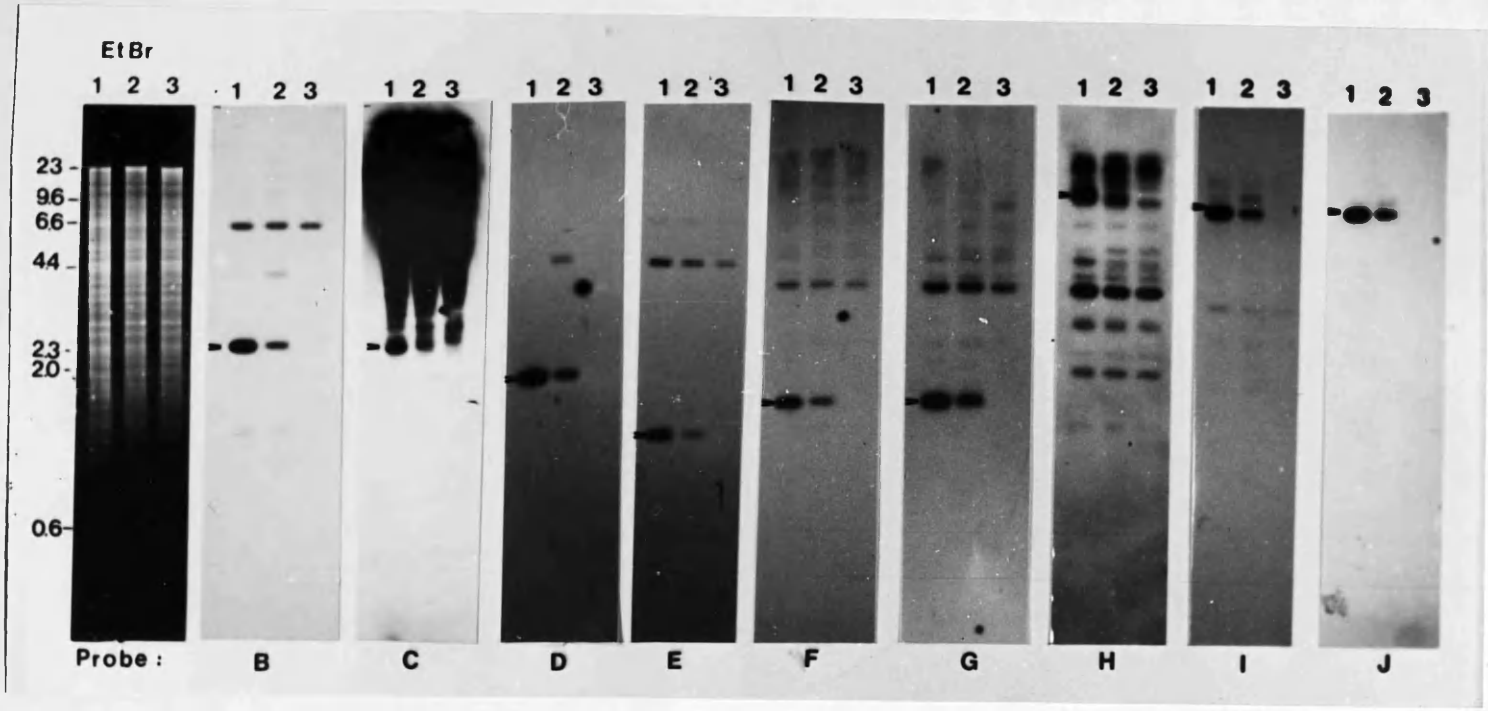


Figure 4.27; All regions of the clones derived from the 1.61 gene loci hybridized to DNAs from 1.22j' (lane 1), GUP 810 (lane 2) and GUP 871 (lane 3) trypanosome populations, each digested with HindIII. The probes used are shown on Figure 4.26. The blots were washed to a stringency of 2xSSC, 65^{o}C. Arrowheads indicate bands fully homologous with the probes used.

Facing page 105

these genes had become members of the metacyclic VSG gene repertoire.

Figure 4.27 shows the results of this analysis using a hybridization stringency lower than that used above. Firstly, the Figure reveals that for all probes, the 2:1:0 ratio of hybridization intensity to the 1.61 derived sequence is present. This indicates that the boundary of the deleted region of the 1.61 telomeres in the GUP 810 and 871 populations has not been crossed.

Secondly, at the lower stringency used in this experiment, it can be seen that the clones still show little evidence of extensive reiteration in the genome, although a major cross reacting band is seen with probes B,E,F and G. Probes F, G and especially probe H also seem to detect a more highly reiterated sequence in the genome, which was not detected previously. This is investigated in the next section.

Thirdly, it can be seen that probe H also detects a relatively prominent band of approximately 5 kb in the DNA from 1.22j', but not the other populations. This is unlikely to represent a divergence between the basic copy sites for the 1.61 gene in this population because this would have been detected in the EcoRI digest of this DNA in the experiment described in section 4.13. It would also have resulted in equivalent hybridization intensities between the 18kb HindIII band detected with probes H, I and J on 1.22j' and the GUP 810 DNAs. A more likely explanation, then, is that this band represents hybridization to a cross reacting fragment unlinked to the 1.61 gene environment which is polymorphic between the populations. The repetitive nature of the sequences detected with this probe make this quite likely, particularly if the sequence is associated with other VSG gene telomeres susceptible to recombination (see section 4.17).

Finally, the experiment shows that with GUP 810 DNA there is hybridization of the probes B and D to a band not detected in either of the other two populations. These bands are probably due to incomplete digestion of this DNA, because on long exposures the 1.22j' DNA also contains these, albeit weakly. This, however,

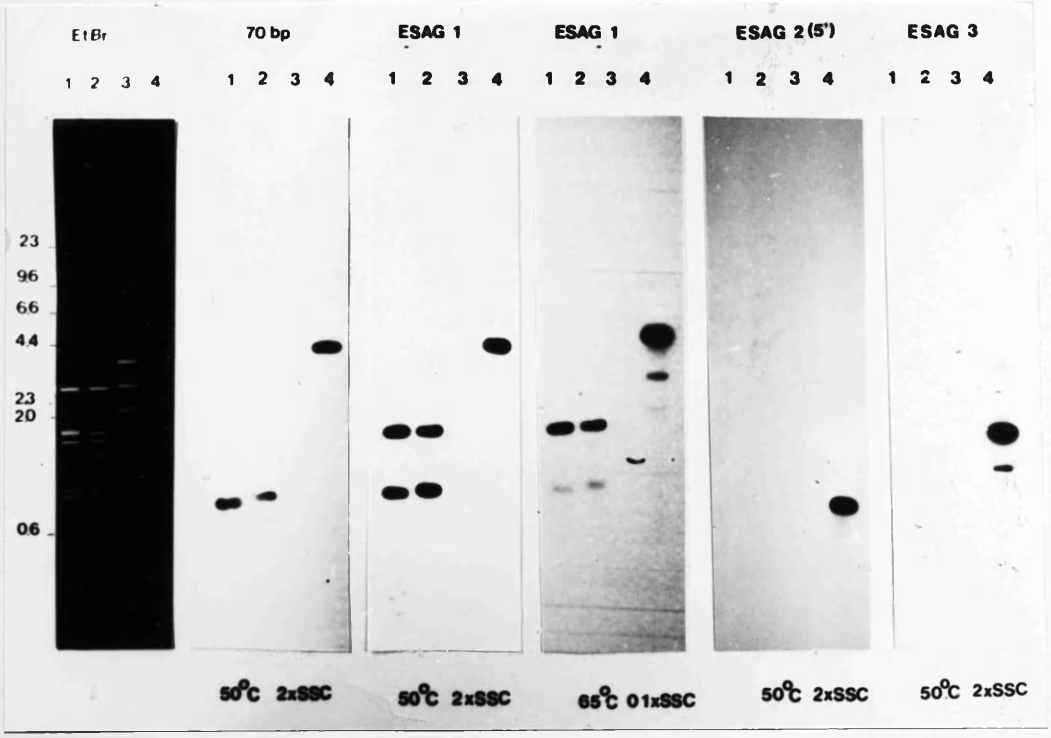


Figure 4.28; Examination of repetitive regions on the 1.61 locus. Lanes are, respectively, (1) plambda X6 digested with XbaI/HindIII, (2) plambda X6 digested with BamHI/HindIII, (3) plambda X5 digested with BamHI/XbaI, (4) pTg221-8 (Kooter et al., 1987) digested with PvuII. For each panel the probes used and stringency employed are shown. ESAG probes were derived from pTg221-8. Each ESAG probe contains the complete gene, excepting the probe for ESAG 2, which is derived from the 5' end of the gene.

Facing page 106

cannot fully explain the bands, because they are not seen with all the probes used. This was surprising because the individual lanes on the gel containing GUP 810 DNA had been loaded from a single bulk DNA digestion. The blots have also been rehybridized with the insert of a plasmid containing part of the expression locus for ILTat 1.22, another M-VSG gene there was no sign of partial digestion (not shown). An alternative possibility, then, is that some sites on the telomere are less susceptible to cleavage in this population. That this extra band is most prominent with those probes derived from the regions of the expression site most proximal to the telomere might imply that a nucleotide modification similar to that seen on inactive bloodstream expression loci is present here (Bernards et al., 1984; see section 1.11). The modification would need to be different, however; although the HindIII cleavage site contains a GC dinucleotide in common with PstI and PvuII (enzymes previously seen to be subject to inhibition by telomere modification), inhibition of HindIII has not been previously detected. Why there would be no partial cleavage of the ILTat 1.22 locus is also not clear as this, too, is an M-VSG gene expression locus and was probably not active in the GUP 871 trypanosomes used for DNA preparation, those being derived from established bloodstream infection.

4.17 Examination of the repetitive sequences on the 1.61 clones.

The observation that some of the probes used in section 4.16 hybridized to other sequences in the genome prompted an analysis of the available clones for features characteristic of bloodstream VSG expression loci. Therefore, the lambda 1.61 subclones plambda X6 and plambda X5 (spanning from 1.5kb 5' of the VSG gene to 14 kb upstream) were hybridized with the first four features seen in the VSG 221 expression site; 70bp repeat sequence and ESAGs 1, 2 and 3. (Bernards et al., 1984; Kooter et al., 1987).

The resulting autoradiograms show that both 70bp repeats and ESAG 1 appear to hybridize to the clones (Figure 4.28). The area

hybridizing to the 70bp repeats is the same as that as had been detected previously in the cloning of pMT1.61-4. The intensity of hybridization to this fragment, which was known to contain just two 70bp repeat motifs, and the low stringency employed (2xSSC, 50^{o}C), indicates that there is probably a complete absence of these motifs on the remainder of the cloned area. The 221-8 plasmid (lane 4, Figure 4.28; the amount of DNA is less in this lane than the others, making its ethidium image rather weak) also hybridizes to 70bp repeat sequence. This is as expected; there are a small number of 70 bp repeats between the ESAG 1 and the pseudo VSG gene in the 221 expression site (Bernards et al., 1985).

The ESAG 1 probe also detects sequence on the 1.61 clones, in the region from which the repetitive probe H had been derived. The autoradiogram also reveals hybridization to the fragment on the 1.61 telomere containing the 70 bp repeats. As above, this is due to the presence of these repeats within the region of the 221 telomere used to construct the ESAG 1 probe. This hybridization, however, is much reduced at high stringency (0.1xSSC, 65^{o}C). The region of the 1.61 telomere showing ESAG 1 homology also shows a less intense hybridization at high stringency. Together with the lack of apparent cross hybridization on genomic DNA of this fragment at high stringency (Figure 4.26, probe H), this seems to indicate that the ESAG on the 1.61 telomere might be quite diverged from other such genes in the trypanosome genome.

Neither ESAG 2 or ESAG 3 hybridized on the 1.61 clones, even at low stringency (2xSSC, 50^{o}C). The presence or absence of other identified ESAGs on bloodstream VSG expression loci has not been assayed. The lack of genomic repetitiveness of the cloned regions beyond the region which hybridized to ESAG 1, even at moderate stringency (65^{o}C, 2xSSC; Figure 4.27), and the conservation of ESAG gene order on bloodstream expression loci makes their presence unlikely.

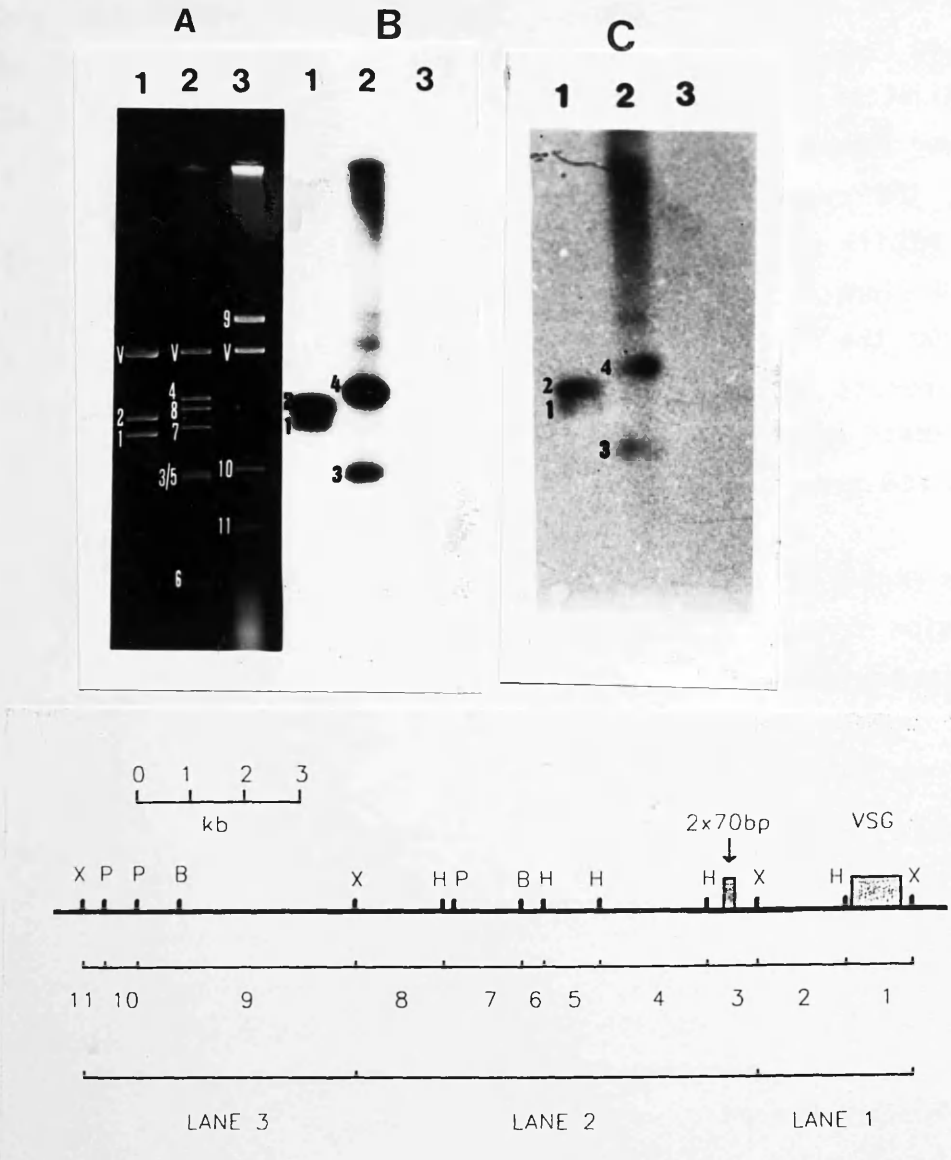


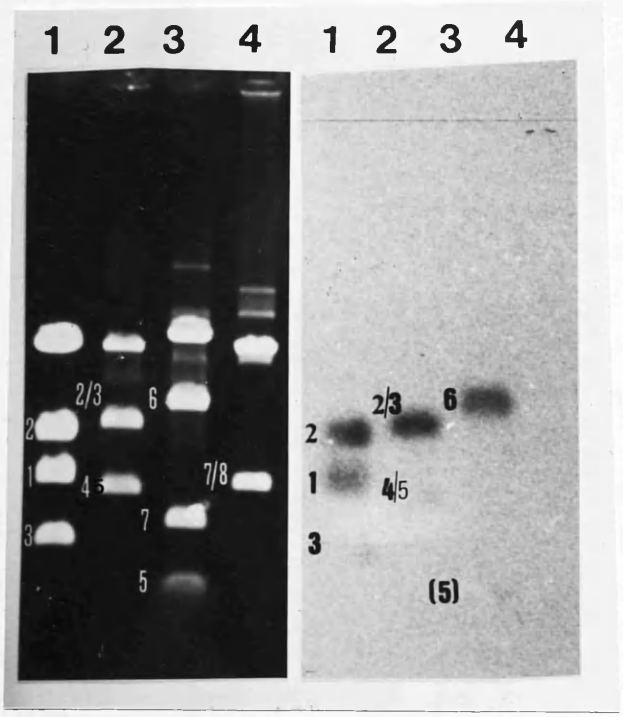
Figure 4.29; An analysis of nascent transcripts from the ILTat 1.61 gene locus in metacyclic derived organisms. Panel A shows an ethidium stained gel of pMT1.61-1 digested with HindIII/XbaI (lane 1), plambda X6 digested with HindIII/BamHI (lane 2) and plambda X5 digested with XbaI/PstI. Panel B is the same gel hybridized with nascent transcripts obtained in the absence of the drug alpha-amanitin. Panel C is the nascent transcripts derived in the presence of 500ug/ml alpha-amanitin hybridized with a duplicate blot to that used in panel B. Below is shown a map to indicate the positions of the various bands labelled on panels A, B and C with respect to the ILTat 1.61 VSG gene. 0.5xSSC, 65^{o}C.

4.18 Transcriptional Analysis of the ILTat 1.61 Locus in Trypanosomes Expressing this Gene as it is Activated in the Fly.

Using the model system described in chapter 3 of this thesis, it is possible to obtain just sufficient metacyclic clone derived trypanosomes to perform an analysis of metacyclic form VSG transcription. The trypanosomes for this were cloned directly from tsetse fly saliva by Dr J.D Barry and grown for seven days in mice. From these mice a population, termed 1.61r, was identified which expressed the ILTat 1.61 gene to a high level. Nuclei were prepared from these parasites according to the method of Kooter et al. (1984) and cryopreserved by Dr S.Graham.

The cryopreserved nuclei were subjected to a "run-on" transcriptional assay both in the presence and absence of 500ug/ml alpha-amanitin. Filters containing digested pMT1.61-1, plambda X6 and plambda X5 (representing the entire cloned area of the 1.61 gene locus) were hybridized at 65^{o}C for 72 hours and then washed to a stringency of 65^{oC}, 0.5xSSC. The need to isolate trypanosomes in as short a time as possible after leaving the tsetse fly means that very limiting amounts of material are available for "run-on" transcriptional analyses. For this reason, it was necessary to use blots of as small a size possible and to incubate distinct blots with the same hybridization solution consecutively.

Thus, after hybridization, the hybridization solution was retained and rehybridized with blots containing plasmid digests of cloned trypanosome tubulin and ribosomal genes. These control for alpha-amanitin activity; tubulin genes are transcribed by RNA polymerase II, which is relatively sensitive to the drug, while ribosomal genes are transcribed by the relatively resistant RNA polymerase I (Kooter and Borst, 1984). Thus, the ribosomal transcript hybridization should be the same between the two sets of conditions, while tubulin transcript hybridization should be greatly depressed in the run-on carried out in the presence of alpha-amanitin. In fact, neither of these blots produced any signal after exposure, probably due to a failure in blotting.



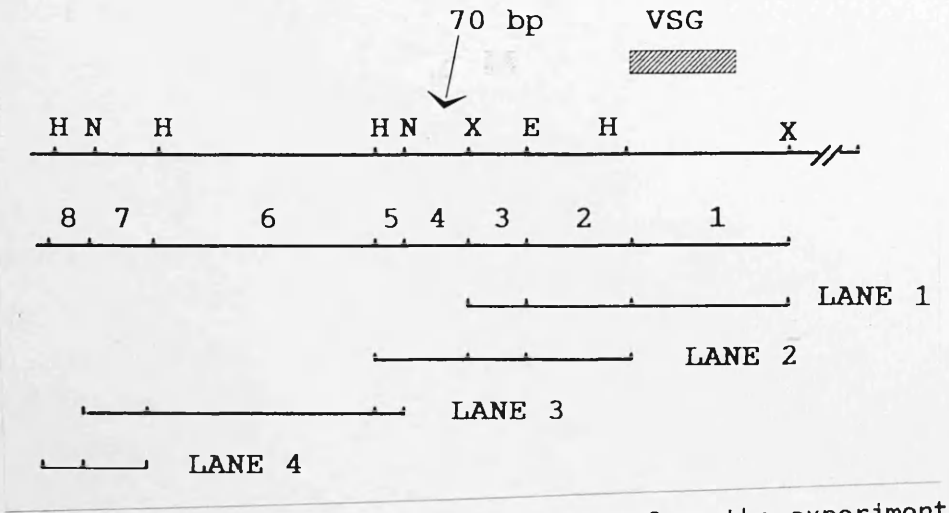


Figure 4.30; Run on transcripts derived from the experiment shown in Figure 4.29, panel B, rehybridized with the 1.61 locus derived plasmid clones to resolve the hybridization to bands 3 and 5 in Figure 4.29. Lane I is HindIII/EcoRI digested pMT1.61, lane 2 is pMT1.61-4 digested with HindIII/XbaI, lane 3 is pMT1.61-5 digested with HindIII/NaeI, lane 4 is pMT 1.61-6 digested with HindIII. The position of the bands labelled in the ethidium picture with respect to the ILTat 1.61 gene are shown below. Unlabelled bands in the ethidium picture are vector derived. 0.5xSSC, 65^{o}C.

Nevertheless, the hybridization solution was again retained and rehybridized with a further blot of the plasmid clones of the 1.61 telomere; pMT1.61-1, 4, 5 and 6. This was performed in order to resolve any hybridization to fragments which co-migrated on the original blot. Filters were hybridized for 72 hours and washed to 65^{o}C, 0.5x SSC.

Figures 4.29 shows that there is hybridization to the cloned areas of the 1.61 telomeres, and that this terminates within the available clones. To the 3' side, the level of hybridization on the 1.3 kb fragment of pMT1.61-1 (fragment 1) is approximately half the intensity of the 1.5kb fragment above (fragment 2). Since the 1.3 kb fragment covers the VSG gene and downstream, this may indicate transcription termination beyond the gene, rather than a progression of transcription to the limit of the cloned region. The upstream limit of transcription falls within the area represented on plambda X6. On the initial blot, co-migration of fragments made it difficult to determine whether transcription proceeded into the more upstream 900bp HindIII fragment (fragment 5) because this co-migrated with the 900 bp fragment of plambda X6 which contains the 70bp repeats (fragment 3). Rehybridization with the plasmid blot (Figure 4.30) revealed no hybridization to pMT1.61-6 (containing the upstream 900bp fragment), however, indicating that transcription probably begins within the 1900 bp HindIII fragment immediately downstream (fragment 4, Figure 4.29; Fragment 6, Figure 4.30). The weakness of hybridization to the plasmid blot, particularly on smaller fragments makes this equivocal, however. Despite this, the failure to detect hybridization to the available 1.61 gene locus clones beyond band 5 suggests that the 1.61 transcription unit is unusually short, probably extending no more than 4.3 kb upstream of the cDNA hybridizing region in the expression site.

There was no evidence of hybridization to the region containing the ESAG 1 hybridizing region, even at prolonged exposures.

The pattern of hybridization is the same for alpha-amanitin treated and untreated nuclei, although it is much weaker in the former. This may imply sensitivity to the drug, although the

failure of the control blots makes this equivocal. Previous analyses of transcription of the ILTat 1.61 gene using the cDNA clone have indicated that it is alpha amanitin insensitive (S.Graham, unpublished).

A full summary of the information derived about the ILTat 1.61 gene loci is summarised in Figure 4.31.

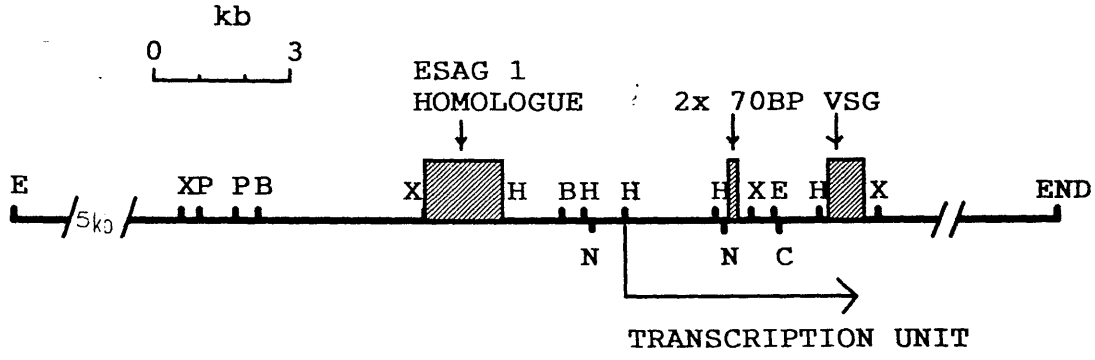


Figure 4.31; A summary of the information derived from the cloning and characterization of the ILTat 1.61 basic copy gene loci. E=EcoRI, X=XbaI, P=PstI, B=BamHI, N=NaeI, C=ClaI, END= telomere, 70bp= 70 bp repeat motif.

Discussion

1. Difficulties in Cloning the ILTat 1.61 Expression Loci in Plasmids

This chapter has described the cloning and characterization of a metacyclic expression locus for the gene encoding ILTat 1.61. Clones for this locus had never been previously isolated from trypanosome genomic libraries (Shiels, 1990) and difficulties in their isolation have been encountered here. Specifically, I have failed to isolate as a single insert a region extending from an EcoRI site represented within pMT1.61-1 to a BamHI site 4.7 kb upstream, although this region was successfully subcloned later from lambda 1.61.

Difficulties in cloning VSG expression sites have been frequently encountered by others. Van der Ploeg et al. (1982) screened cosmid banks for the presence of the ELC of several VSG genes and found none, despite the ready isolation of basic copy clones. This was not due to the paucity of restriction sites between VSG genes and the chromosome end resulting from the repetitive nature of telomeric sequences (Blackburn and Challoner, 1984); the cloning of specific restriction fragments also failed. This was probably due, instead, to the presence of a large number of 70 bp repeat motifs in expression sites. The presence of a long repeat array was not the problem in the cloning described here, however; only a short 70 bp repeat hybridizing region was identified on the telomere and this area was readily cloned in pMT1.61-4. This length of repeats has also not presented cloning difficulties in other analyses (Aline et al., 1985; Florent et al., 1987; Shah et al., 1987).

Another important determinant in the ability to clone specific DNA fragments is the competence for recombination of the bacterial host. In the experiments described here, the recF host strain E.coli DS941 was used early on, while recA strains were

used subsequently (E.coli HB101, DS902, DH5 alpha). Although recF hosts are more prone to rearranging inserted sequences than rec A hosts, this is unlikely to have resulted in the failure to clone the desired plasmid insert in the constructed libraries: the lambda 1.61 subclones of the same areas were readily generated and propagated in E.coli DS941 and the ClaI-BamHI library was replated onto E.coli DS902, which is a rec A mutant. An alternative possibility, therefore, concerns the restriction status of the host strains used. This describes the ability of the bacterial host to recognise exogenous DNA on the basis of nucleotide modification at particular sequences (reviewed by Raleigh, 1987). Sequences containing such sites are cleaved and are, thus, absent from libraries containing foreign genomic DNA.

This is also unlikely to have been responsible for the failure to clone in plasmids part of the 1.61 expression loci, however: all plasmid strains used were defective for either the EcoK or EcoB restriction systems. Similarly, the alternative restriction systems, mcr and mar (modified cytosine restriction, modified adenosine restriction) cannot have operated on the fragments being cloned: whilst the mcr and mar status of the hosts used in the isolation of plasmid clones was generally unknown, the lambda 1.61 clone was isolated from a host strain fully mcr and mar competent.

Finally, the failure to isolate the desired clones was probably not due to the isolation of insufficient recombinants; the constructed PstI-ClaI library contained 7 identical clones derived from a cross reacting fragment. The degree of homology that this fragment showed with the fragment of pMT1.61-1 which detected it and its relative intensity after Southern hybridization, indicated that this fragment was probably of similar copy number in the genome as the desired fragment. The number of recombinants isolated from the ClaI-BamHI library should also have been sufficient to clone the 1.61 derived fragment, when transformed into both DS941 and DS 902. The ClaI-EcoRI library, however, probably resulted in insufficient recombinants to expect to clone the 10kb ClaI-EcoRI fragment of the 1.61 locus and even these were unstable.

2. Instability of lambda 1.61.

There was additional evidence of instability in the isolated lambda clones with a tendency for deletion of the probe hybridizing region and circumstantial evidence for apparent rearrangement at one end of the clone. The initial instability in the library may have been due to its primary plating on the rec A^+ host E.coli P2392. This was essential, however; the low recombinant level in the library necessitated selection against non recombinants by using spi selection system which requires a recombination competent host (Sorge, 1988). In any case, this may not have mattered because the phage showed continued instability when propagated in the recombination disabled E.coli NM621 and the libraries of Shiels (1990) were constructed using this host.

Upstream sequences in the expression site may have caused the instability of the lambda insert and this has precedent in the cloning of other VSG gene loci. Kooter et al. (1987) failed to isolate a clone from a region in the 221 expression site around 30 kb from the VSG gene and Zomerdijk et al. (submitted) have discovered a series of 50bp repeats refractory to cloning 5' of the VSG promoter in the same expression site. These 50 bp motifs, like the 70bp repeat region, form restriction site barren regions and produce a hybridization smear on Southern blots. There was no evidence of these repeats within the clone plambda X5, the most upstream stable plasmid clone, however; hybridization of its 5' terminus did not reveal it to be extensively reiterated in the genome.

3. Structure of the ILTat 1.61 basic copy locus.

Approximately 16kb of the ILTat 1.61 basic copy loci have been cloned. Progressive restriction mapping during the isolation of plasmid clones and the results in section 4.13 have not identified the 2 kb insertion in one of the two basic copy telomeres reported by Cornelissen et al. (1985a). This is simple to explain. Cornelissen et al. (1985a) used the pTcV21-15 cDNA

clone to map sites upstream on the two basic copy loci. Where an insertion in one locus is reported, it is only with enzymes which do not cut within, or downstream of the region hybridizing to the cDNA probe. These fragments, therefore, run all the way from the restriction sites down to the chromosome ends, a distance between 18 and 25 kb. In these experiments it would not have been possible to distinguish the difference in telomere length between the two copies from an upstream insertion between restriction sites. The relatively high resolution Southern blot shown in Figure 4.20, and the use of clones spanning the region of the predicted divergence clearly resolves no difference between the two telomeres.

The absence of restriction map differences between the two basic copy loci means that the areas cloned are probably derived at random from parts of both, rather than being representative of one or other. This should not matter, provided the two are as homologous as they appear to be by mapping. Any divergence at the level of single nucleotides will affect neither the genomic structural analysis or the transcriptional analysis, which both rely on hybridization. Future analyses on promoter sequences, however, may need to account for this.

The analysis of the structure of the 1.61 loci by hybridization has revealed that they are largely composed of sequence in low copy number in the genome which contrasts strongly with what has been found for bloodstream VSG expression sites at stringencies higher than were used here (Kooter et al., 1987; Pays et al., 1989a). The 1.61 clones, however, do show four repetitive regions.

Firstly, the 3' half of pMT1.61-1 hybridized to multiple bands at high stringency (Figure 4.3.2). This is probably due to the clone bearing the subtelomeric sequences found downstream of many VSG genes (Van der Ploeg et al., 1984c). The cross hybridization is not due to homologies within the 3' half of the VSG gene itself, because the cDNA insert does not detect these multiple bands at the same stringency (Figure 4.26).

Secondly, around 1.5kb upstream of the cDNA hybridizing region is the 70bp repeats which, as mentioned above, cover a

very short area. Their length is very unusual, being more similar to the extent of repeats upstream of chromosomal internal VSG genes (Van der Ploeg et al., 1982) than the 5-30 kb extent of repeats often associated with telomeric expression sites (Van der Ploeg et al., 1982; Laurent et al., 1984a). It is in agreement, however, with the observed absence of a barren region on the M-VAT basic copy loci examined by Lenardo et al. (1984), although in those cases there were no repeats whatsoever (Lenardo et al., 1986).

Thirdly, beyond the 70 bp repeats was found a short region, apparently double copy in the trypanosome genome (section 4.12). This region does not indicate a divergence between the 1.61 basic copy loci; as above, probes hybridizing to fragments spanning this region detect no divergence. Instead, then, this is probably simply a region with cross homology in the genome, similar to that which resulted in the erroneous cloning of pMT1.61-3. Many other probes for the 1.61 locus also seem to pick up one or two major cross reacting bands. These do not seem to be associated with the 1.61 locus as they are present in the GUP 871 trypanosome line which has no copies of this gene. They could, however, represent hybridization to another expression site or a limited number of sites with which the 1.61 locus shows partial homology over some distance. For example, it is possible that the 1.61 locus is susceptible to infrequent partial conversions, or that the locus was derived from another locus by a large single conversion event after which the sequences have diverged. Such large conversion events have been seen previously in bloodstream expression sites (Pays et al., 1983c; Bernards et al., 1986a).

A gene conversion process may have resulted in the appearance of the fourth repetitive region on the clones, which shows similarity to Expression Site Associated Gene 1. The use of DNA probes for ESAG 1 result in a strong hybridization to 17-25 bands in trypanosome genomic DNA from several subspecies, even at high stringency (Cully et al., 1985; Kooter et al., 1987). With the ESAG 1 homologue on the 1.61 locus, however, hybridization at the same stringency is far less intense. This might indicate that the ESAG here is diverged and even a pseudogene. Although this area

would need to be sequenced to test this, there is precedent for a crippled ESAG in the telomeric basic copy locus for the AnTat 1.3 gene. The environment for this gene resembles a bloodstream VSG expression site but is apparently unable to be used for VSG transcription (Alexandre et al., 1988).

Other studied metacyclic basic copy gene loci have been found to contain an ESAG 1 sequence uniterrupted by stop codons (Son et al., 1989) excepting that encoding ILTat 1.22 (Shiels et al., submitted). This latter expression locus also shows less cross homology than the 1.61 locus with other sequences in the genome.

No other ESAGs have been detected in the 1.61 locus, although from bloodstream expression site analyses they might be predicted to be present within the 5.5 kb upstream of the ESAG1 related sequence. Alexandre et al. (1988) found three distinct ESAGs clustered within a 6kb region on the AnTat 1.3 expression site, as did Kooter et al. (1987), in an 8kb stretch of the VSG 221 locus. Chromosome walking upstream of the existing 1.61 clones would be necessary to determine if an absence or unusual spacing of ESAGs were a feature of the metacyclic VSG gene environment.

In the populations 1.22j', GUP 810 and GUP 871 there are 2, 1 and 0 copies of the ILTat 1.61 gene. Hybridization of the DNAs of these populations has revealed that no probe derived from the cloned area of the locus detects the three populations with equal intensity. This indicates that limit of deletion of the gene environment has not been crossed. The deletion of the 1.61 loci in the GUG 810 and 871 populations might, therefore, be explained by a long conversion of the locus (Pays et al., 1983; Bernards et al., 1986). An alternative possibility is that the locus has been lost through sexual processes in these parasites. The M-VAT genes are on the trypanosomes largest chromosomes which, unlike the smaller chromosomes, appear to conform to normal mendelian principles during trypanosome genetic exchange (Sternberg et al., 1988). Since self fertilization has been observed in these trypanosomes (CMR Turner, personal communication), the appearance of organisms with 2, 1 and 0 copies of the 1.61 locus is compatible with the F2 progeny of an F1 heterozygote. This possibility has been tested very recently and been found to be

probably incorrect; the three trypanosome populations show no evidence of selfing, being heterozygous for several genetic loci in independent clones (CMR Turner, personal communication).

There was some evidence of a potential telomere modification of the 1.61 locus in GUP 810 trypanosomes, which have just one 1.61 gene. If this is so, it is the first case in which restriction enzyme cleavage inhibition has been seen for a metacyclic VSG gene locus (Lenardo et al., 1986). Why such modification would be needed in this population and not in 1.22j' trypanosomes is not clear; in both populations the expression site is inactive. This result would need further confirmation before any firm conclusions could be drawn.

4. Expression Site Transcription.

A run-on transcriptional analysis of the 1.61 locus indicates that transcription begins approximately 1 to 2.5 kb upstream of the 70bp repeats, producing a primary transcript of about 5kb. This assumes that transcription terminates 3' to the VSG (see section 4.18), in agreement with previous analyses (De Lange et al., 1986; cited in Borst et al., 1986) but in contrast to Rudenko and Van der Ploeg (1989) who suggested transcription extended into the telomeric repeats. This transcription, however, was not proven to be associated with VSG transcription.

A short metacyclic expression locus agrees with a similar analysis by S.Graham (unpublished observations) on the ILTat 1.22 locus but is in sharp contrast to the bloodstream VSG transcription unit which may be between 40 and 59 kb in length (Kooter et al., 1987; Pays et al., 1989a; Crozatier et al., submitted). In two cases a short transcription unit has been seen for a bloodstream locus (Shea et al., 1987; Alexandre et al., 1988) but in both the apparent transcription starts were within 1 Kb of the most 5' region of the available expression site clones. The isolation of further clones revealed that the start was, in fact, merely a short gap possibly produced by polymerase pausing or lethargy in this region or very rapid primary transcript processing. In the M-VAT gene locus studied here, the region

showing an absence of hybridization to run-on transcripts extends 10 kb upstream, a length unlikely to be explicable by the features described above. There is, however, one alternative to the locus having a short transcription unit which is compatible with the data obtained here: a polymerase able to begin transcription at a distant promoter and then re-initiate close to the VSG gene (Campbell et al., 1984). Although unlikely, this possibility could be tested by using primer extension methodology to examine the 5' ends of the nascent transcript. These experiments would be technically demanding, however, considering the very limiting amounts of material available for an analysis of metacyclic transcription.

The run-on experiment showed no evidence of transcription of the ESAG 1 homologue, although these genes do appear to be transcribed in the metacyclic model system trypanosomes (S.Graham, unpublished). This further supports the divergence of this region of the 1.61 locus from other ESAGs and suggests that ESAGs need not always be transcribed co-ordinately with a VSG in their own expression site. This is in agreeement with the analysis of the ILTat 1.22 gene expression, where there was no ESAG or ESAG homologue within 18 kb of the VSG gene, but where there were abundant ESAG primary transcripts (Graham and Barry,in preparation). Similarly, uncoupled ESAG and VSG transcription has been seen in procyclic form parasites (Alexandre et al., 1988; Pays et al., 1989a). Assuming transcribed ESAGs in metacyclic derived parasites are in another VSG expression locus, it is apparent that they can be expressed using a different control mechanism to the VSG gene itself, at least at this point in the life cycle. This may be achieved through the use of either distinct promoters or transcription factors or may employ transcription termination to prevent downstream VSG expression. The long arrays of 70bp repeats on bloodstream expression loci may assist in this transcriptional isolation of distinct regions of the expression site, by their AT rich stretches acting as transcription terminators, for example, or by their potential ability to delineate chromatin domains. Such AT-rich stretches have been implicated in both mammalian (LeMeur et al., 1984) and

yeast (Henikoff and Cohen, 1984) transcriptional termination, and are also targets for enzymes contributing to chromatin organization (Adachi et al., 1989). Either process might permit uncoupled ESAG and VSG gene transcription, but would need to be under developmental control, to permit full expression site transcription in the bloodstream VSG system.

5.The M-VAT Expression Site and the M-VAT Repertoire.

A simple transcription unit and expression site structure can explain many of the peculiarities of the M-VAT system. In the previous chapter, it was seen that it might be advantageous for the trypanosome to keep its M-VAT expression loci distinct from those used in the bloodstream system. This was because this could prevent the metacyclic cells becoming prone to frequent expression of dominant B-VATs, against which hosts in the field might have developed immunity. The relatively non-repetitive nature of the M-VAT expression loci could assist in this by limiting their susceptibility to conversion by bloodstream VSG genes. Most notable is the extent of 70bp repeats on the telomere, since these repeats often seem to be involved in conversions at bloodstream expression sites. By having a very short stretch of these repeats, as is seen here, or a complete absence of them, as has been seen by others (Lenardo et al., 1986), the M-VAT loci may be less susceptible to invasion by bloodstream expression site sequences, particularly if conversion events were specifically initiated in these repeats. This is more fully discussed in chapter 5 of this thesis.

Although the 70bp repeats might be the most frequently used signals for recombination, other expression site homologies are likely to be susceptible to more generalized recombination, albeit in the longer term. It is this that may be responsible for the relative instability of ILTat 1.61 as a member of the metacyclic repertoire: Barry et al. (1983) found that this M-VAT was lost from the metacyclic VSG repertoire after only three rounds of cyclical transmission in the laboratory. Perhaps, then, by having an ESAG related sequence and other repeated areas, the

expression site is prone to conversion by related sequences which do not possess the signals for M-VAT gene activation. A similar situation may have been observed in trypanosomes from other stocks; the MVAT 4 gene locus has a gross architecture similar to that of the 1.61 gene and has also been reported to be lost from the genome (Son et al., 1989; Lenardo et al., 1986). In contrast to this, the ILTat 1.22 gene locus appears very non repetitive within the genome (Shiels et al., submitted; Chapter 5, this thesis) and accordingly is very stable in the metacyclic repertoire (Barry et al., 1983). From this, it appears that M-VAT expression sites may not show the apparent high level of similarity seen between bloodstream expression sites (Alexandre et al., 1988; Crozatier et al., submitted), but may instead show a spectrum of relatedness, ranging between complete disimilarity and partial similarity with other M-VAT loci and bloodstream expression sites. This might be possible in this system because the transcription unit appears to begin very close to the VSG gene, thereby allowing the non-coding upstream environment to drift in sequence. This expression site heterogeneity might permit the trypanosomes both to maintain a separate and predictable M-VSG gene pool and allow the potential for change in repertoire in the long term.

The brevity of the transcription unit for the ILTat 1.61 locus is in striking contrast to other VSG expression loci except that for another M-VSG gene, encoding ILTat 1.22 (S.Graham and P.Shiels, unpublished observations), and serves to emphasise the distinction between the bloodstream and metacyclic VSG systems. There are a number of potential reasons why the metacyclic system should be so different. Firstly, by keeping ESAGs out of the transcription unit, its relative recombinational isolation can be maintained, as described above. Secondly, it is possible that a complex transcription unit is not required in metacyclics. The function of no identified ESAG is known; they may have no role at the metacyclic stage. They may be needed early on in the bloodstream, however, and this may explain why the parasite possibly recruits ESAG expression from elsewhere in the genome at this time. Finally, it may be energetically inefficient to

maintain a huge transcription unit in the metacyclic stage. The energy availability in the fly salivary gland is likely to be far more limited than in the bloodstream; perhaps the parasite needs to balance preadaptation for life in the mammal with its ability to survive for a sufficient time in the salivary gland to permit transmission. This is hard to reconcile, however, with the large number of rapidly dividing epimastigote forms which co-exist with metacyclic cells in the tsetse fly salivary gland.

From a preliminary investigation by Dr S.Graham, the transcription of ILTat 1.61 does not appear to be alpha-amanitin sensitive (although this needs verification) and may therefore be transcribed by the same polymerase used for bloodstream expression sites. This presents no difficulties for the functional isolation of the M-VAT system; an alpha-amanitin insensitive polymerase also seems to transcribe the insect form surface molecule, procyclin (Konig et al., 1989; Rudenko et al., 1989). The different specificities of the polymerase may be mediated by different components of the holoenzyme. It would be interesting to observe whether the metacyclic system displayed the sensistivity to heat shown by the bloodstream system (Kooter and Borst, 1984, Kooter et al., 1987). It would be predicted they might not, as M-VAT expression needs to be maintained during the temperature fluctuations experienced in the tsetse fly.

Chapter 5

The activation of metacyclic VSG genes in the mammalian bloodstream.

5.1 Introduction

In the experiments described in Chapters 3 and 4 of this thesis and in the experiments of Shiels *et al.*(submitted) it has been seen that the genes for two predominant M-VATs in the EATRO 795 stock, ILTats 1.61 and 1.22, are expressed without duplication very soon after leaving the fly and, by extrapolation, in the fly salivary gland. When re-expressed, however, in a chronic bloodstream infection, it is found that both these genes become activated only after generation of an expression linked copy (Cornelissen *et al.*, 1985a).

The unusual features of the expression sites for the ILTat 1.61 and 1.22 genes (Chapter 4, this thesis; Shiels *et al.*, submitted) prompted an examination of how this duplicative transposition was achieved; the 1.61 locus was found to have only two of the transposition-associated 70 bp repeats and there was no obvious restriction enzyme "barren region" in the 1.22 gene locus (Cornelissen *et al.*, 1985a).

In order to examine the reactivation of M-VAT genes in the bloodstream, I chose to investigate the duplicative activation of the 1.22 gene. This gene was chosen, in preference to the 1.61 gene, because a relatively large number of trypanosome populations were available which expressed this gene stably in mice. These populations were derived from recloned populations of trypanosomes which formerly had been expressing the 1.22 gene *in situ*, having been isolated from tsetse fly salivary exudates. These recloned populations have been found invariably to be expressing the 1.22 gene via an ELC, as they do in more chronic infection (S.Graham and J.D. Barry, unpublished).

The experiments required an initial partial characterization of a region proximal to the 1.22 gene, represented in a genomic plasmid clone, pMG 7.1-1 (Cornelissen *et al.*, 1985a). This was carried out prior to the more extensive characterization of this expression locus by Shiels *et al.* (submitted).

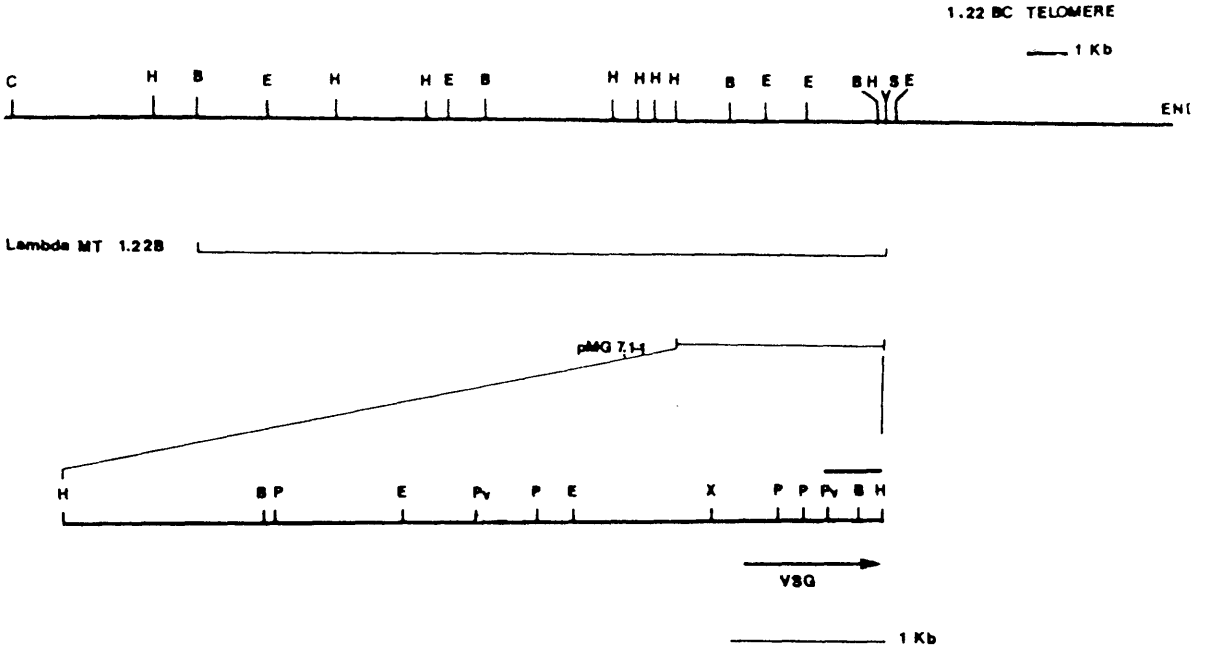


Figure 5.1; A: Restriction map of the ILTat 1.22 basic copy gene locus and the derived clones pMG7.1-1 and Lambda 1.22B(lambda MT 1.22B) (Cornelissen et al., 1985a; Shiels et al., submitted). H=HindIII, P=PstI, X=XbaI, B=BamHI, C=ClaI, E=EcoRI, S=SphI, Pv=PvuII.

Facing page 125

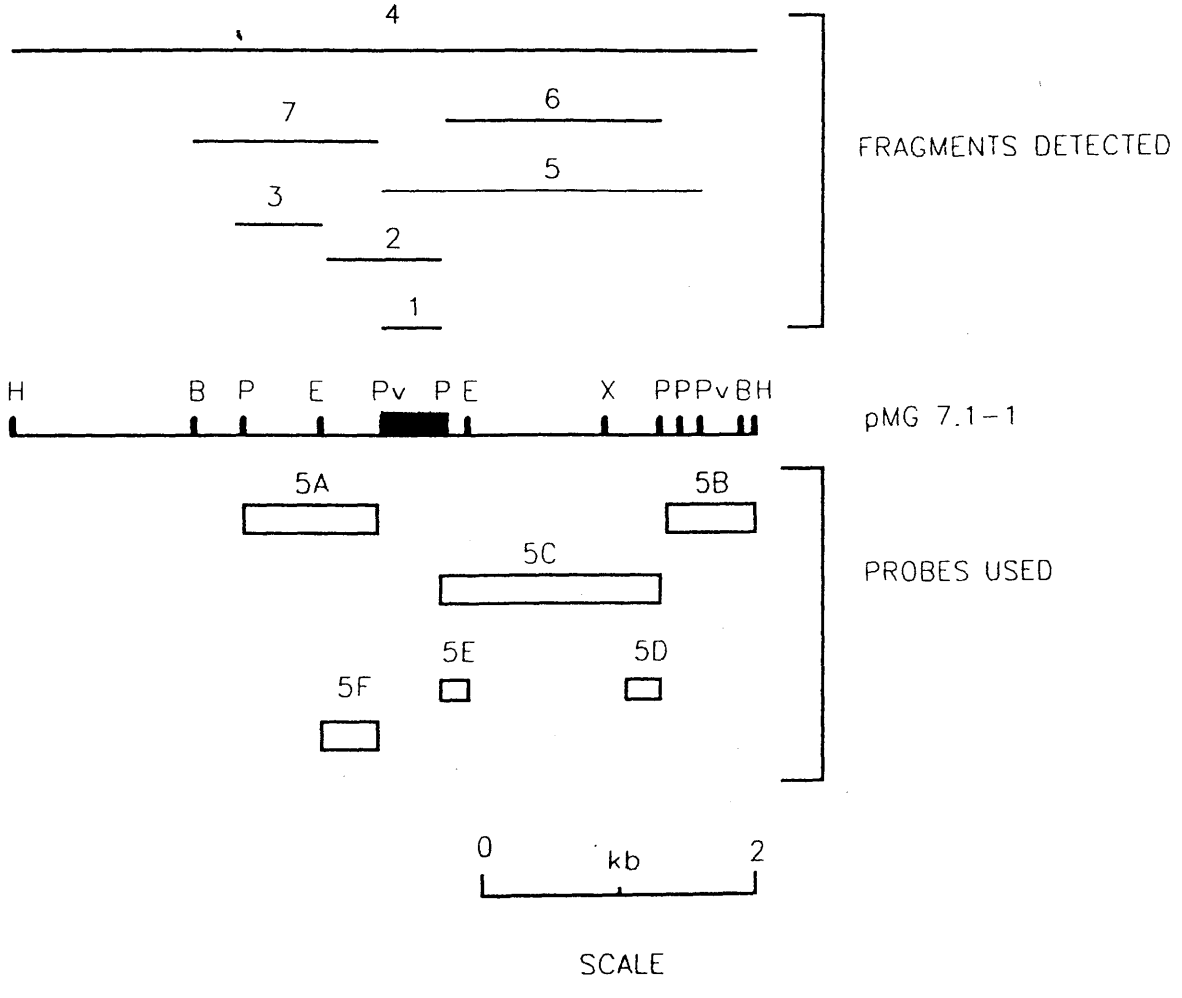


Figure 5.1; B: pMG7.1-1, showing probes used and fragments described throughout this Chapter.
H=HindIII, B=BamHI, P=PstI, E=EcoRI, X=XbaI, Pv=PvuII
The filled box represents the region containing 70bp repeats, identified in section 5.2 of this Chapter.

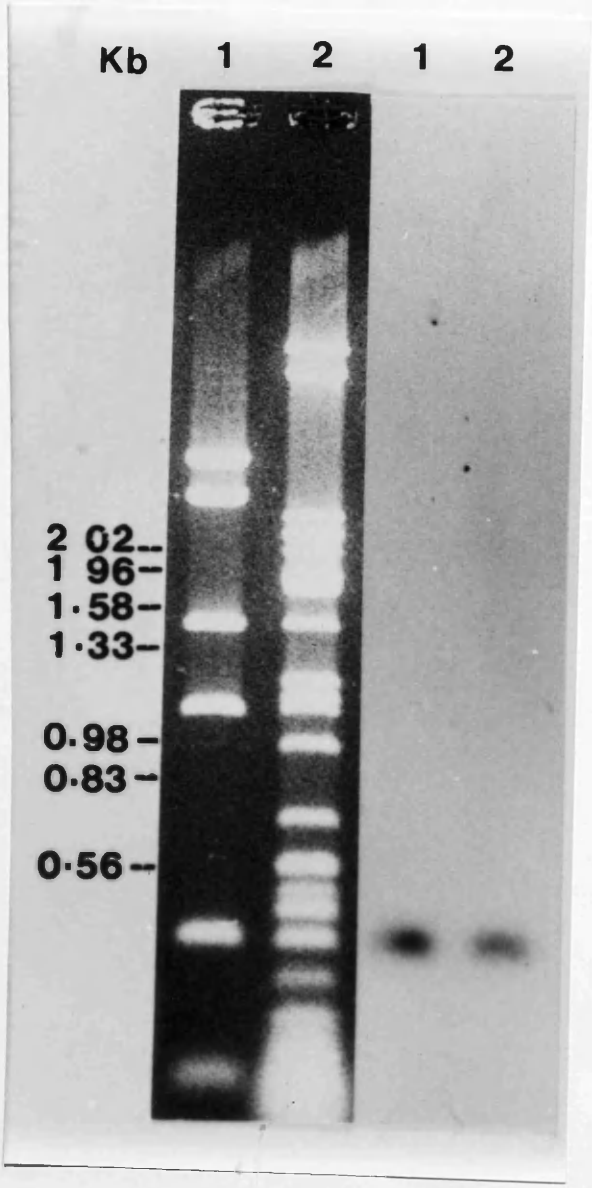


Figure 5.2; 70 bp repeats on the ILTat 1.22 basic copy locus. Lanes are PstI/PvuII digested pMG7.1-1 (lane 1) and lambda 1.22B (lane 2). The left panel shows the ethidium stained gel; the right panel the DNAs hybridized with 70bp repeats from the VSG 221 gene expression locus (Bernards et al., 1985). 50^{o}C, 2xSSC.

Facing page 126

5.2 Characterization of pMG 7.1-1

The generation of an expression linked copy seems to rely upon homologies between VSG loci, such as the 70bp repeats (Michels et al., 1983; Liu et al.,1983) and, for telomeric genes, other expression site homologies (Kooter et al., 1988; Lee and Van der Ploeg, 1987). Whilst there was no apparent barren region in the restriction map upstream of the 1.22 gene (Figure 5.1; updated from Cornelissen et al., 1985a) there remained the possibility of a very short 70bp repeat array as seen on the ILTat 1.61 locus (Chapter 4, this thesis).

To determine if there were any repeats present within the ILTat 1.22 expression locus, pMG7.1-1 and lambda 1.22B, a clone of the locus extending 17 kb upstream of the 1.22 gene (Shiels et al., submitted), were doubly digested with PstI and PvuII and hybridized with a 4.5 Kb PstI/TaqI fragment of pTg221-1, which is entirely composed of tandem 70 bp repeats derived from the VSG 221 expression site (Bernards et al., 1985). Figure 5.2 shows that there is specific hybridization of this probe to a 420 bp PstI/PvuII fragment, approximately 2100bp from the 3' HindIII site of the plasmid insert (fragment 1, Figure 5.1). No other region of either pMG7.1-1 or lambda 1.22B hybridizes with the 70 bp repeats probe, even at very prolonged exposures and at low stringency (3xSSC, 42^{o}C; not shown). This indicates that although the 1.22 basic copy locus displays no barren region, 70 bp repeat sequence is present on the telomere, albeit confined to a very short stretch.

Identical digests of pMG7.1-1 have also been tested for the presence of expression site associated genes (ESAGs), and none has been found. This is not presented, having been subsequently repeated and presented in Shiels et al. (submitted).

5.3 Genomic verification of the extent of the 70 repeat hybridizing region

Barren regions upstream of trypanosome telomeric VSG genes, being composed of tandemly linked direct repeats, are potentially

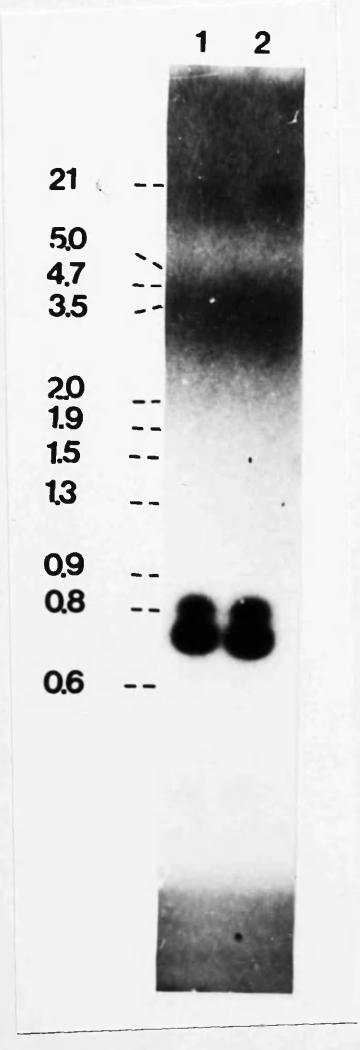


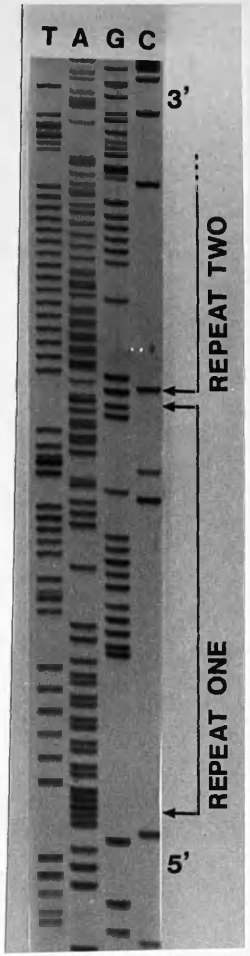
Figure 5.3; Verification the pMG7.1-1 has not deleted any 70bp repeats during cloning. Lanes are PstI/EcoRI digested 1.22j' trypanosome genomic DNA (lane 1) or the non-expressor, C2, DNA (lane 2) hybridized with probe 5A. 0.1xSSC, 65 °C.

Facing page 127

susceptible to extensive deletion during cloning procedures. In consequence, it was important to verify that the extent of the 70bp repeats on pMG 7.1-1 was representative of the corresponding region in the trypanosome genome. To do this it was necessary to use a probe which did not contain the repeats themselves, because this would result in hybridization to many VSG loci, but which would instead hybridize to a fragment spanning this region. Therefore, a 1250 bp PstI/PvuII fragment of pMG7.1-1 (probe 5A, Figure 5.1) was used as a probe onto a Southern blot of 1.22j' trypanosome genomic DNA doubly digested with PstI and EcoRI and size fractionated by gel electrophoresis. This DNA, as described in Chapter 4 of this study, is derived from established bloodstream parasites which express the ILTat 1.22 gene via an expression linked copy. If pMG 7.1-1 was an undeleted clone of the equivalent genomic stretch, then two bands were predicted to hybridize on genomic DNA; an 830 bp fragment (fragment 2, Figure 5.1), containing the 70 bp repeat region and a 730 bp fragment (fragment 3, Figure 5.1), immediately upstream. Figure 5.3, lane 1, shows that the two bands detected are of exactly the expected size with their hybridization intensity reflecting the extent of their homology with probe 5A: the upper band hybridizes over only approximately half of its length while the lower band hybridizes over its entire length. Thus, to a resolution of less than one repeat unit, this experiment confirms the fidelity of the cloned 70 bp repeat region from the ILTat 1.22 gene locus with respect to the trypanosome genome.

Lane 2 of Figure 5.3 shows the same experiment performed on the DNA of a trypanosome population distinct from that used for lane 1. In this population, the 1.22 gene was not being expressed and there is no ELC (not shown). It is clear from the Figure that the size of the hybridizing fragments are the same as those seen for 1.22j' DNA. This indicates that there is no change in the number of 70bp repeats on the ILTat 1.22 gene telomere associated with either activity or inactivity of the gene.

```
      PvuII
   1  CTGGTGAACTCGCTATAAATAGTAGGTTAATTACCTCTTTTGCATGTCAGAGTTACGACG

  61  TTGTATATAGCAAAAATAATAATAATAATAATAGGAGAGTGTTGTGAGTGTGTGTATATA
      CONSENSUS->CAGTRRTRRTRRTRRTRRTAGGAGAGTGTTGTGAGTGTGTGTATATA

 121  CGAATCTTATAATAAGAGAAGCAGTAATAATAATAATAATGATAATAATGATGATGATGA
      CGAATATTATAATAAGAG---CAGTRRTRRTRRTRRTRRTRRTRRTRRTRRTRRTRR

 181  TGATAGTAATACGAGGAAAGTGTGTTGTTAGTGCAAAAGAATGGACAAGCCACAGCAGAC
      TRRTRRTRRTAGGAG---AGTG--TTGTGAGTGTGTGTATATACGAATATTATAATAAGA

 241  GCCTTCCAACAATTTAGTGGCCCCAGGAAGGGAACCAAAAACTGAGAGAGCGCGAGGAAT
      G<-

 301  GTTAAAGGGAAAAGACAATAAGACAGAGGAATTGAGAAGGAGATTTGAATATATGCAGGT

 361  TTCGCAAGACAAAACGATACTGTAAATTCTCTTGCACGCTGTAAATTTCTCTTGCACGAA

                  Pst I
 421  ACAAAAGCGGCACGGCGCTGCAG
```



Facing page 128

Figure 5.4; Sequence of the 70bp repeat hybridizing region on pMG7.1-1. Only one half of the 5' PvuII site is shown because the subclone used generated a hybrid site by blunt end ligation. The consensus sequence shown is derived from Aline et al. (1985).
R=Purine nucleotide. Underlined nucleotides indicate divergence from the consensus. Gaps have been introduced to maximize alignment.
The right hand panel shows the sequencing gel of the the 70bp repeat region on pMG7.1-1. The sequence is unambiguous.

5.4 Sequence of the 70 bp repeat containing region.

In order to determine the extent of the 70 bp repeats within the 420 bp PstI/PvuII fragment of pMG7.1-1, this region was cloned in each orientation into the bacteriophage M13 mp 18 and 19 and sequenced. Aline et al.(1985a) have defined a 70 bp repeat motif as divided into 3 distinct regions; a (T Purine Purine) stretch, an alternating purine/pyrimidine stretch and an A-T rich stretch (see section 1.8). The sequence in Figure 5.4 shows that the 420 bp PstI/PvuII fragment contains just one complete 70 bp repeat motif with an additional (T Purine Purine) stretch and part of the alternating purine/pyrimidine region. This is shorter even than the 70bp repeat stretch on the ILTat 1.61 locus and is very much shorter than the extent of repeats associated with telomeric bloodstream VSG genes (Van der Ploeg et al., 1982; Liu et al., 1983; Bernards et al., 1985a; Shah et al., 1987 Alexandre et al., 1988).

5.5 Are the 70 bp repeats functional in gene conversion when 1.22 generates an ELC?

The short extent of repeats on the 1.22 telomere raised the question of whether the region could function as the limit of conversion when the 1.22 gene is activated by duplicative transposition. In order to test this, the conversion endpoint was determined for 1.22j' trypanosomes. Initially, to confirm the presence of an ELC in this population, 1.22j' DNA was digested with HindIII, as was the DNA of a trypanosome population, 1.61i, which did not express the 1.22 gene. After agarose gel electrophoresis and Southern blotting, the DNAs were hybridized with the 5' portion of the 1.22 specific cDNA clone insert, pTcV7.1-14 (Cornelissen et al., 1985a; probe 5B, Figure 5.1). This detects a HindIII fragment on the 1.22 locus which extends from within the VSG gene to 5.3 kb upstream; the region equivalent to the pMG 7.1-1 insert (fragment 4, Figure 5.1). Figure 5.5, panel A, reveals that whilst the probe detects this 5.3kb band in both 1.22j' and 1.61i, there is also a band of

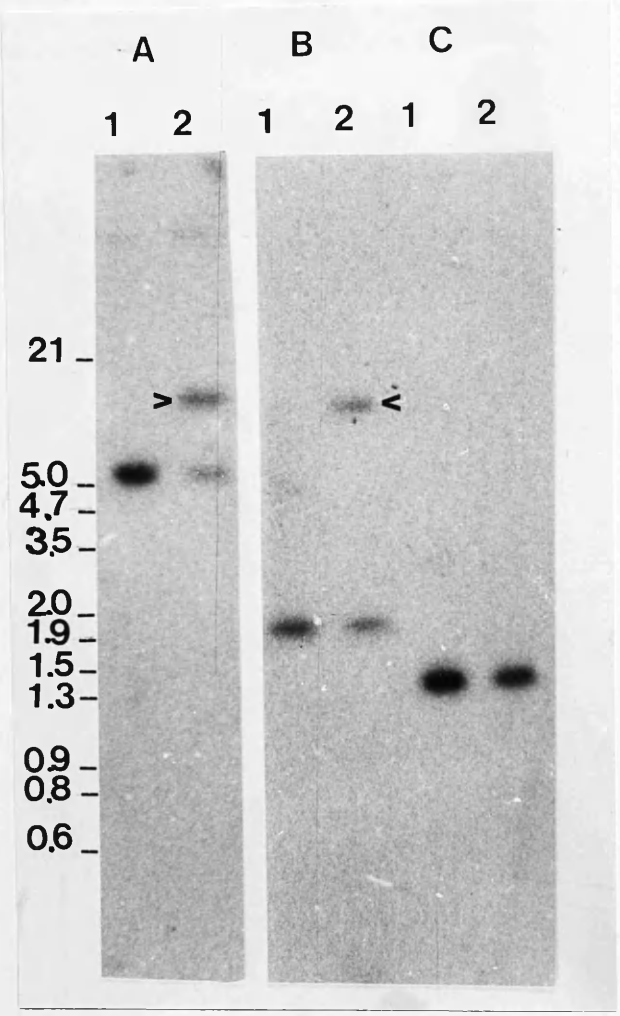


Figure 5.5; Determination of the limit of the 1.22 gene locus duplicated segment in the ELC expressor, 1.22j'. Panel A, HindIII digested 1.61i trypanosome genomic DNA (lane 1) and 1.22j' DNA (lane 2) hybridized with probe 5B. Panels B and C, the same DNAs digested with BamHI/PvuII and PstI, respectively, and hybridized with probe 5C. Arrows indicate the ELC derived bands. 0.1xSSC, $65^{o}C$.

Facing page 129

approximately 14 kb in the 1.22j' track not present in the 1.61i lane (marked by an arrowhead). This confirms the presence of an ELC in the 1.22j' population and indicates that the restriction maps of the basic copy and expression linked copy gene loci begin to differ within the 5.3 kb HindIII fragment.

To determine if the 1.22 telomere 70 bp repeats were the limit of the duplicated segment, 1.22j' and 1.61i DNAs were doubly digested with BamHI and PvuII and hybridized with a 1540 bp PstI fragment of pMG7.1-1 (probe 5C, Figure 5.1). On a BamHI/PvuII digest this probe detects the 1.22 telomere from within the VSG gene to just 5' of the 70bp repeat containing area (fragment 5, Figure 5.1). Figure 5.5, panel B, reveals that the probe detects two bands in 1.22j', indicating, as above, hybridization to a region on the 1.22 telomere within which the restriction maps for the basic copy and ELC loci begin to differ (i.e. the limit of the converted segment). To map the ELC limit more precisely, the 1.22 j' and 1.61i DNAs were also digested with PstI and hybridized with probe 5C. Figure 5.5 panel C shows that in this case there is hybrization to just one band, representing hybridization to an area over which the restriction maps of the 1.22 BC and ELC are the same (fragment 6, Figure 5.1). This demonstrates that in 1.22j' trypanosomes the limit of duplication is within the 420 bp PstI/PvuII fragment represented within pMG7.1-1, the area containing the 70 bp repeats.

5.6 Does the 1.22 BC telomere always involve the 70 bp repeats in ELC generation?

Kooter et al.(1988) have analysed ELC generation for the 221 VSG gene which, like the 1.22 gene, is telomeric. They have found that while this gene can use its 70bp repeats in ELC generation, other sequences in the expression site can mark the conversion limit. To determine, therefore, if the 1.22 gene could also use other areas of its basic copy locus as a limit of conversion, the experiment in section 5.5 was repeated, using the DNA from a number of independent trypanosome populations expressing the 1.22 gene. All, excepting 1.22i, were cloned

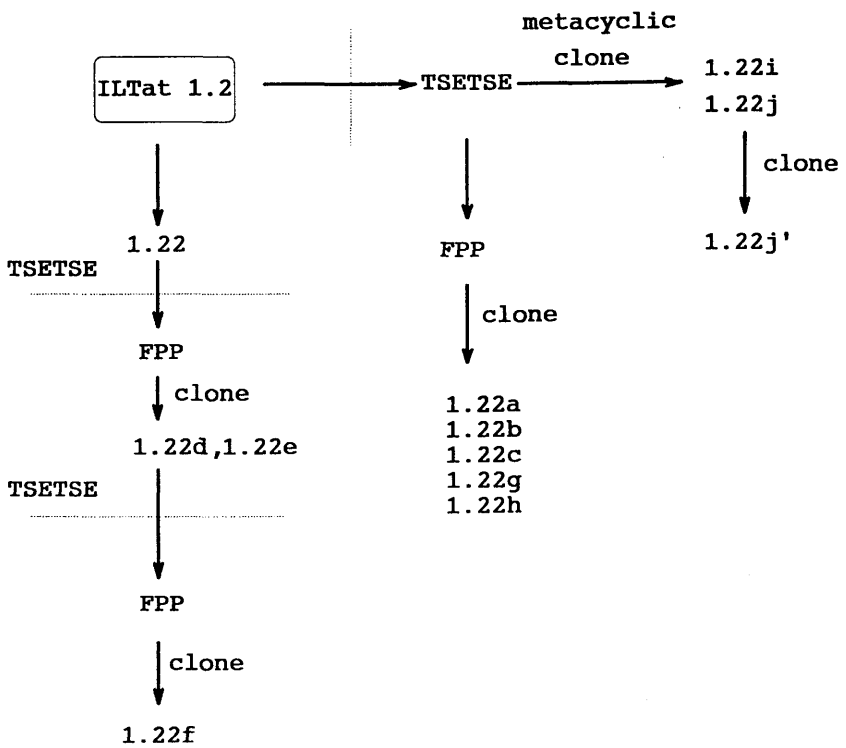


Figure 5.6; Genealogy of various ILTat 1.22 expressor populations. Dotted lines indicate tsetse transmission, FPP= first patent parasitaemia in mice following infection by tsetse fly bite, clone=infection of a single trypanosome into a mouse. Note that, for clarity, the clones 1.22a,b,c,g and h are grouped together, despite their independent isolation from distinct tsetse transmissions. All clones were derived by J.D.Barry.

Facing page 130

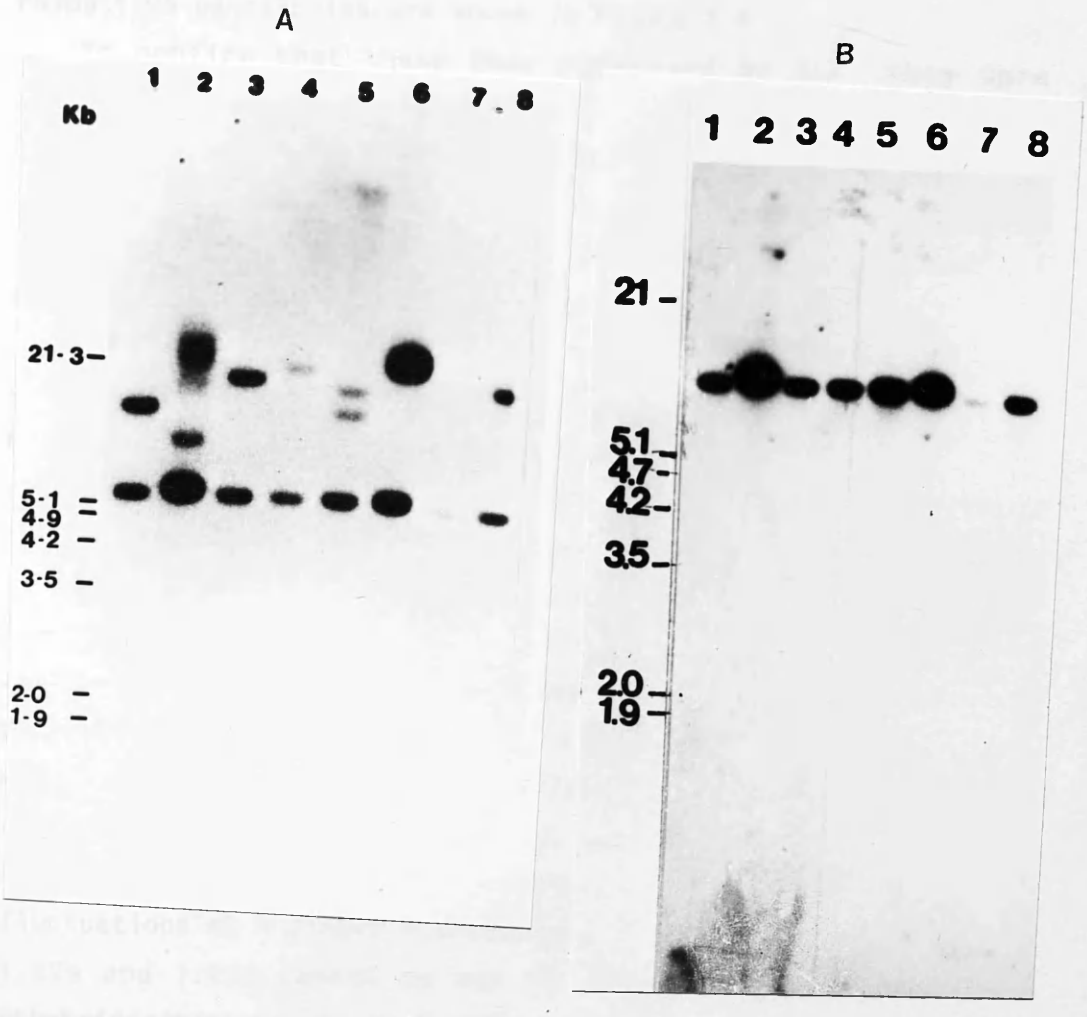
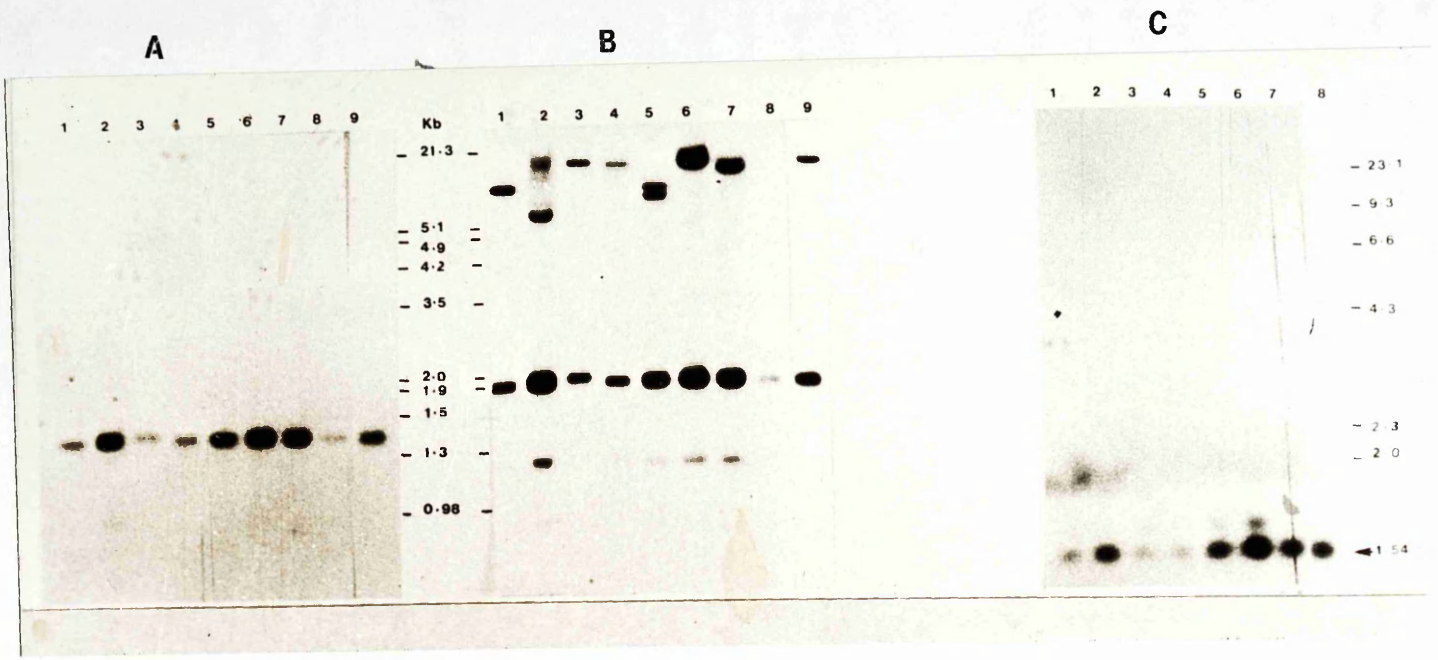


Figure 5.7; ELC expressors of the ILTat 1.22 gene. Panel A, Lanes 1-8, repectively, HindIII digested 1.22a, 1.22b, 1.22c, 1.22d, 1.22e, 1.22f, 1.22i, 1.22j' DNA, hybridized with probe 5B. 1.22i is a metacyclic derived clonal population, all others were cloned from bloodstream form trypanosomes in the first patent parasitaemia in mice. Panel B is the same blot, stripped and rehybridized with a cDNA probe for the triose phosphate isomerase gene. 0.1xSSC, 65^{o}C. Note that DNA from 1.22g trypanosomes, which is used in subsequent experiments, was not included here due to there being limited material.

populations derived from trypanosomes in the first parasitaemia in a mouse after its infection by tsetse fly bite. Their respective genealogies are shown in Figure 5.6.

To confirm that these DNAs possessed an ELC, they were firstly digested with HindIII and hybridized with probe 5B, as had been done for 1.22j' in section 5.5. Figure 5.7, panel A, shows that in each case the predicted 5.3 Kb basic copy fragment is detected, as is an additional fragment of variable size, representing hybridization to an ELC. Exceptions are the control 1.22i, which has no ELC (even at very long exposures of the autoradiogram shown in Figure 5.7, panel A), being derived from a metacyclic clone displaying expression of the 1.22 gene in situ (see Figure 5.6), and 1.22e which shows apparently two ELCs. This could have arisen if the population was non clonal, although it is possible that two independent conversion events have generated the two ELCs, one or both of which may be transcriptionally active. Although this has not been tested, recloning of this population or either DNAse 1 or S1 nuclease digestion could distinguish between these possibilities. In 1.22b, there is an apparent smear above the distinct ELC derived band. This may represent multiple ELCs, or inter-population barren region size fluctuations at a single expression site. These multiple bands in 1.22e and 1.22b cannot be due to incomplete DNA digestion; rehybridization of the blot with a probe for the gene for triose phosphate isomerase produces just a single hybridizing band for all DNAs (Figure 5.7, panel B). Thus, these experiments demonstrated that all of the populations, excluding the control in situ expressor 1.22i, contained at least one ELC for the 1.22 gene.

The DNAs for these 8 populations were then subjected to a similar experiment to that described in section 5.5 in order to determine the limit of the duplicated segment for each. Panel A of Figure 5.8 represents probe 5A of pMG 7.1-1 hybridized to the various DNAs doubly digested with BamHI and PvuII. In each case a single band hybridizes. This reveals that in none of the populations does the duplication limit lie within the region



Facing page 132

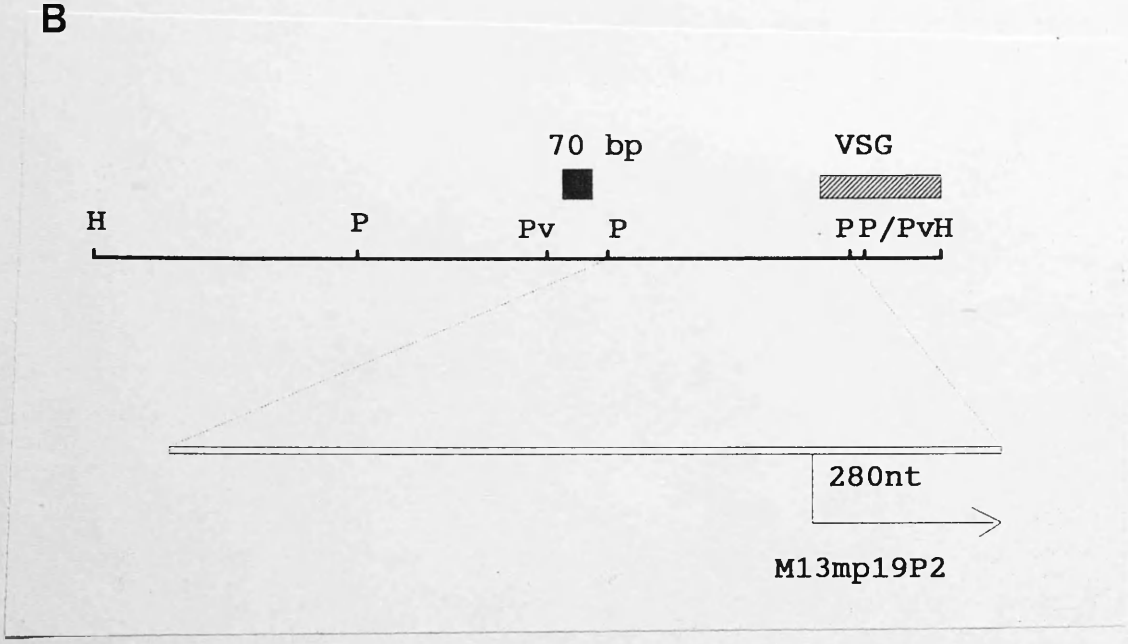
Figure 5.8; The 5' limit of the 1.22 gene duplicated segment always falls within the region on the basic copy locus containing 70bp repeats. Lanes 1-9 (panel A, B) are, respectively, 1.22a, 1.22b, 1.22c, 1.22d, 1.22e, 1.22f, 1.22g, 1.22i, 1.22j' DNAs digested with BamHI/PvuII and hybridized with probe 5A (panel A) or probe 5C (panel B). Lanes 1-8 (panel C) are, respectively, 1.22a, 1.22b, 1.22c, 1.22d, 1.22e, 1.22f, 1.22i and 1.22j' DNAs digested with PstI and hybridized with probe 5C. Note that on panel C DNA from 1.22g trypanosomes is missing due there being limiting material. 0.1xSSC, 65^{o}C.

of the basic copy locus immediately upstream of the 70 bp repeats (fragment 7, Figure 5.1). It also indicates that in all cases the DNAs were digested to completion.

Panel B is the same blot stripped of the original probe and rehybridized with probe 5C. In this case, all populations, with the exception of 1.22e, 1.22b and 1.22i, display two bands; one of 2.0 kb representing the basic copy locus, and another of variable size representing the ELC. As before , 1.22i has only its basic copy while 1.22e and 1.22b have additional ELC bands. This shows that, as with 1.22j', the duplication limit falls within either the 70 bp repeats or downstream.

Panel C of Figure 5.8 shows the DNAs (except 1.22g; insufficient was available for this blot) digested with PstI and hybridized with probe 5C of pMG 7.1-1. In this case, for each DNA, just a single band is detected, as had been seen for 1.22j' alone (section 5.5). This represents hybridization of the probe over an area of the 1.22 gene locus where the basic copy and expression linked copies have the same restriction map. From this it is possible to conclude that for all ELC expressors examined the limit of the 1.22 gene duplicated segment lies upstream of the area equivalent to probe 5C and downstream of the area equivalent to probe 5A; the area containing the 70 bp repeats.

On panels B and C of Figure 5.8, probe 5C can be seen to detect an additional band of 1250 bp and 1600 bp respectively, which was not seen in Figure 5.5, panel B, because the specific activity of the probe was higher in this experiment. These bands were felt to be a result of the probe containing sequence partially homologous with another VSG transposed segment, and not due to there being a third copy of the 1.22 gene environment in these trypanosomes. In order to test this, a probe was constructed which contained just the 3' 280 nt of the fragment from which probe 5C was derived (probe 5D, Figure 5.1). This was generated by the cloning of intermediates in M13 complementary strand synthesis (section 2.20; Materials and Methods). In brief, the universal sequencing primer was used to prime synthesis of M13mp18 DNA containing the 1540bp PstI fragment of pMG7.1-1



Facing page 133

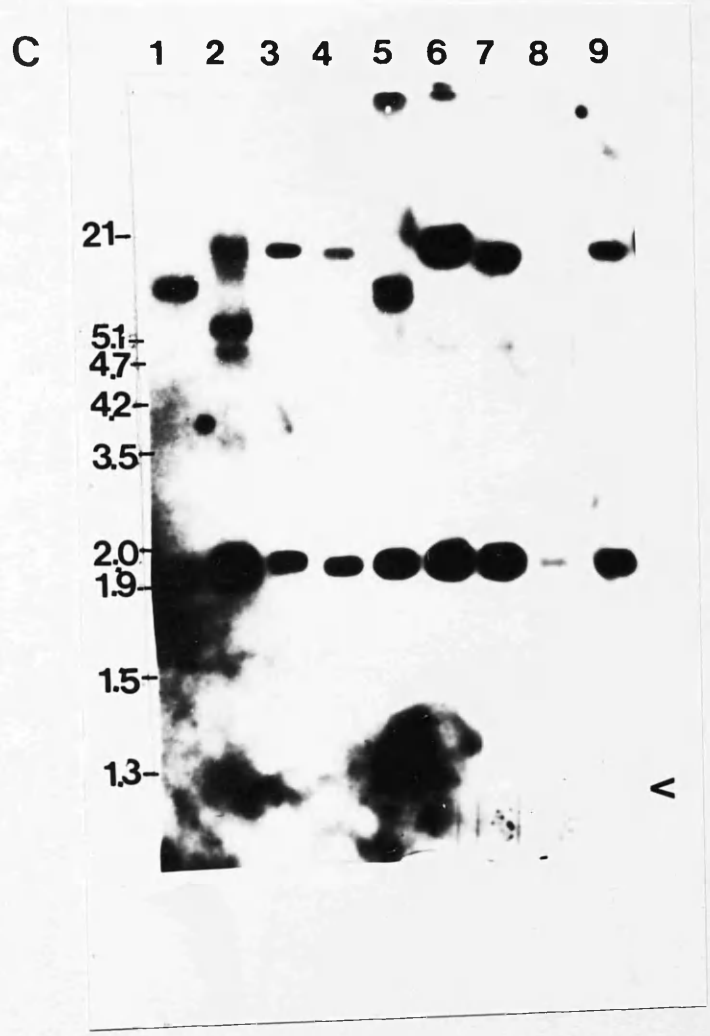
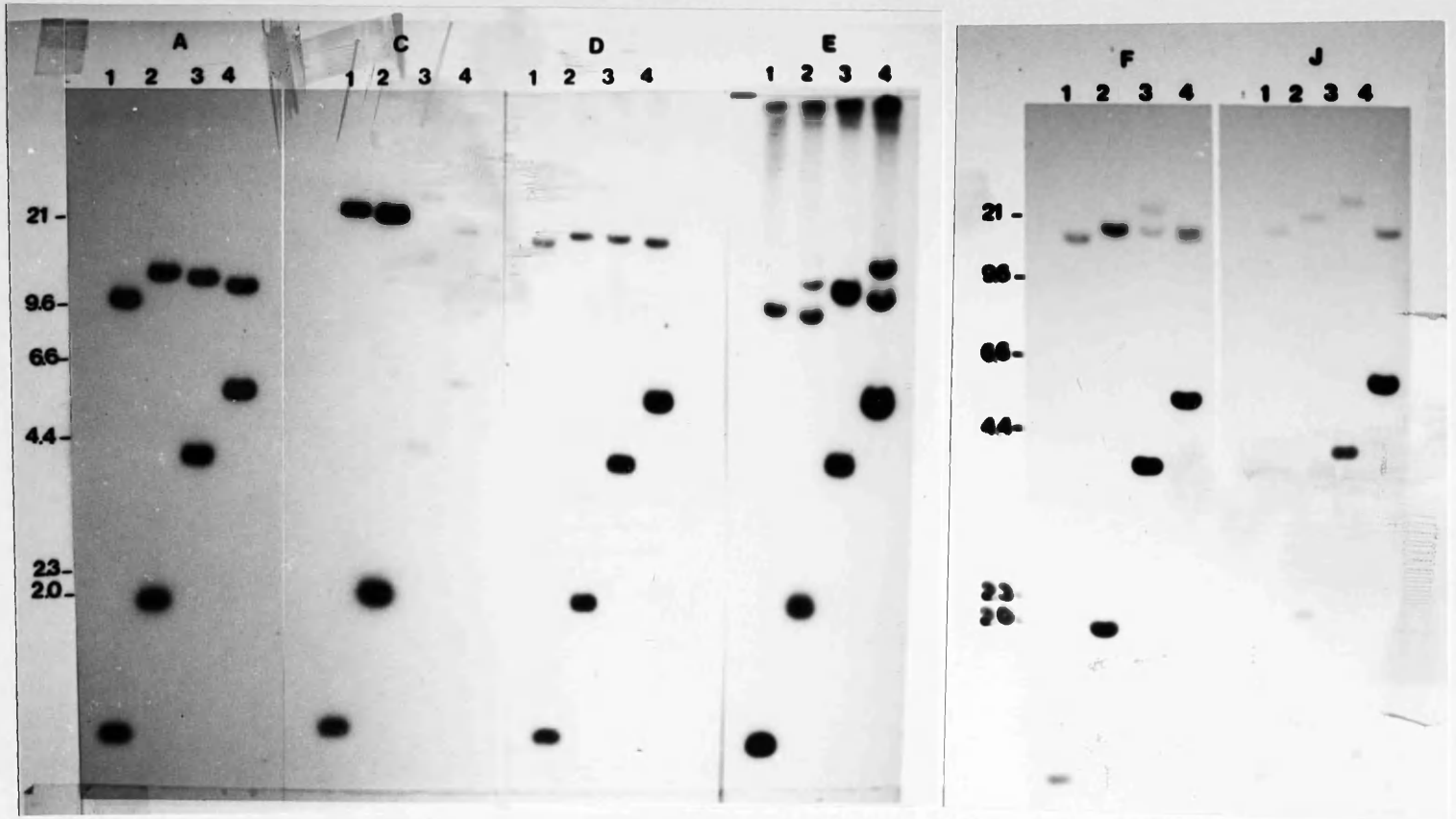


Figure 5.9; Panel A; Sampling of the intermediates in synthesis during the construction of probe 5D. Lane 1 contains the pooled double stranded products of extension after digestion with S1 nuclease (with the DNA populations at each time point being labelled 1-5). Lane 2 is the pooled DNAs, digested and religated into M13, according to the methodology outlined in Figure 2.1. The blot was hybridized with probe 5C, 0.1xSSC, 65^{o}C.
Panel B; representation of the region of the 1540bp PstI fragment of pMG7.1-1 subcloned into M13mp19, as M13mp19P2 and used as probe 5D.
Figure 5.9; Panel C; M13mp19P2 hybridized with the blot used in Figure 5.8, panel A and B.The lanes are as described in the legend for Figure 5.8. The cross reacting band of 1.2kb seen in Figure 5.8, panel B, is no longer detected (its former position is marked by an arrowhead).



Facing page 134

Figure 5.10; Determination of the expression site map for the 1.22 gene in various ELC expressors. The trypanosome DNAs used were, panel A=1.22a, , panel C=1.22c, panel D=1.22d, panel E=1.22e, panel F=1.22f and panel j=1.22j'. The DNAs were digested with EcoRI (lane 1), PvuII (lane 2), BamHI (lane 3) and HindIII (lane 4). All DNAs were hybridized with probe 5E. 0.1xSSC, 65^{o}C.

(equivalent to probe 5C), and this was allowed to proceed for differing times. The resultant partial duplex molecules were subjected to S1 nuclease treatment, digested with PstI and ligated into PstI/HincII digested M13mp19. The reaction products were monitored before and after ligation by visualisation on a gel, and subsequent Southern hybridization with the 1540bp PstI fragment (Figure 5.9.A). After transformation, one resulting clone, M13mp19P2, was sequenced to determine the extent of its insert, and then radiolabelled and hybridized to the blot used in Figure 5.8 panel A and B, which had been stripped of its original probe. Figure 5.9.c reveals that in this case only the basic copy and ELC are detected, with there being no evidence of hybridization to the cross reacting band (its position when this blot had been hybridized with probe 5C is indicated by an arrowhead in Figure 5.9.2)

5.7 Does the 1.22 gene always transpose to the same expression site or can multiple expression loci activate the gene?

Having discovered that the 1.22 gene was able to undergo gene conversion events, it was of interest to determine how many expression sites were acting to accept and activate the 1.22 gene. To determine this, six of the DNAs from populations expressing 1.22 via an ELC were digested with four different restriction endonucleases and Southern blots hybridized with a 250 bp PstI/EcoRI fragment derived from pMG7.1-1 (probe 5E, Figure 5.1B). This probe, for all restriction digests, would detect both a single basic copy derived band of known size and an ELC derived band of unknown size, allowing a restriction map for the ELCs to be constructed. Thus, for each DNA shown in Figure 5.10, the bands of a constant size between the different DNAs were basic copy derived and of the size predicted from previous mapping (not shown), while the variable bands are ELC derived. The restriction maps generated (Figure 5.11) all display a barren region for the endonucleases tested of at least 7 Kb, implying the existence at these expression sites of extensive 70bp repeat

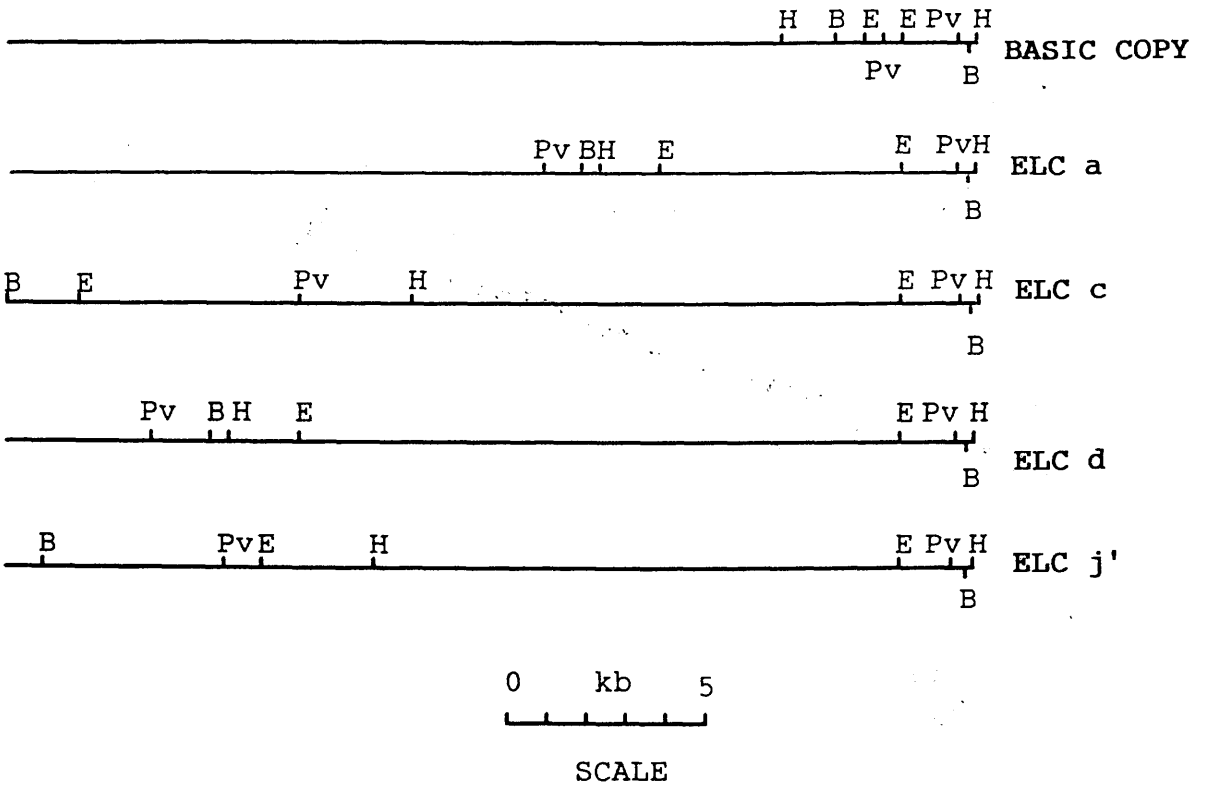


Figure 5.11; Derived restriction map for the ILTat 1.22 expression loci. E=EcoRI, B=BamHI, Pv=PvuII, H=HindIII. ELC=Expression Linked Copy.

Facing page 135

arrays. Because of this long barren region on each ELC, it was difficult to size accurately these fragments for mapping. Instead, then, the respective order of restriction sites on each was used to determine the presence of distinct expression sites. This is potentially prone to error; it is possible that every population is using the same expression site, but that their map differs due to restriction site degeneracies within the barren region introduced by successive conversions and abortive conversions. The approach has been used extensively by others, however, and, in cases where the same expression site appears to have been used, there is little discrepancy in the maps (Michels et al., 1983; Pays et al., 1983a; Myler et al., 1984; 1988).

Beyond the barren region, the restriction maps fall into at least three distinct patterns, one of which appears to be used in two populations, 1.22a and 1.22d. The expression site for the ELCs in the population 1.22e could not be determined, because of the presence of 2 ELCs. It was clear, however, that the two ELCs could not represent use of the same expression locus with a different sized barren region in a non clonal population; there was not a constant size difference between the two bands in each distinct digest (Figure 5.10). In 1.22f there were also two ELC bands, at least in lanes containing BamHI and HindIII digested DNAs (lanes 3 and 4; Figure 5.10, Panel F). These are probably not the products of partial DNA digestion, as there is no evidence of partial digestion of the basic copy. These bands were not seen in Figure 5.8 and may, therefore, represent the presence of 2 ELCs which comigrated on BamHI/PvuII digestion. This is not unlikely, even in a clonal population. The 1.22f population was cloned after cyclical transmission initiated by 1.22d and might, therefore, have both a lingering ELC from the 1.22d trypanosomes and a newly activated one (see Figure 5.6). Although this is difficult to determine, the pattern of bands in Panel F is compatible with this hypothesis.

Thus, from an analysis of six ILTat 1.22 expressor populations, established in the bloodstream, there appears good evidence that a number of distinct expression sites can be used to activate the 1.22 gene.

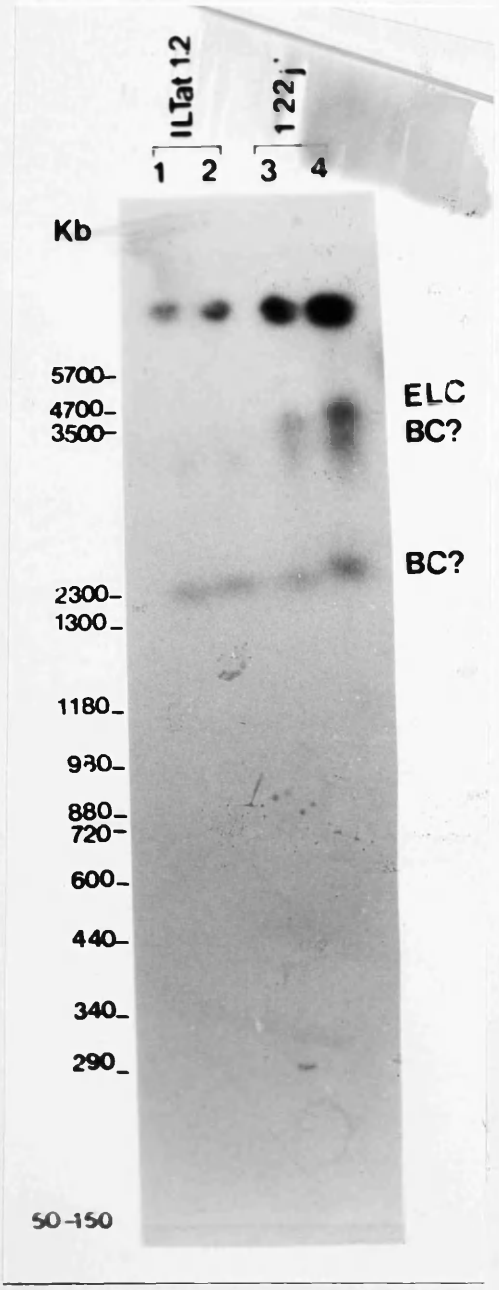


Figure 5.12; The basic copy and expression linked copy chromosomal locations of the 1.22 gene in 1.22j' trypanosomes. The blot shows ILTat 1.22 (lanes 1 and 2) and 1.22j' (lanes 3 and 4) chromosome sized DNA separated by PFGE and hybridized with probe 5C. BC?=Basic copy chromosome, ELC= Expression linked copy chromosome. 0.1x SSC, 65^{O}C.

Facing page 136

5.8 Chromosomal location of the 1.22j' expression linked copy

In previous analyses of the chromosomal location of the ELCs for metacyclic VSG genes, it has been found that, like the basic copies, they are all on the parasites' large chromosomes (Cornelissen et al., 1985a; Delauw et al., 1987). It was of interest, therefore, to determine the chromosome, or chromosome size class, which the 1.22 gene converted when undergoing duplicative activation. It was also hoped that by following similar conditions to those used by Van der Ploeg et al. (1989) for the improved separation of large chromosomes, it would be possible to resolve the basic copy and ELC gene harbouring chromosomes away from the slot material.

Therefore, chromosome blocks were prepared from ILTat 1.2 trypanosomes (not expressing the 1.22 gene) and from 1.22j' trypanosomes, the ELC expressor examined in section 5.5. These samples were then separated on a 0.7% agarose gel, using a Rotaphor rotating field electrophoresis system with a ramped pulse time decreasing on a log scale from 4500 seconds to 800 seconds. The gel was run for 96 hours at 46 volts, 11^{o}C. The gel was then re-run for a further 20 hours at 220 volts, 13^{o}C using a pulse interval ramped from 80 to 20 seconds. This was carried out by J.D.Barry.

After blotting, the filter was hybridized with probe 5C, which would detect both the basic copy and ELC harbouring chromosomes. The resulting autoradiogram shown in Figure 5.12, shows that the chromosome harbouring the 1.22 gene basic copy has, for the first time, probably been resolved at approximately 2 or 3.5 Mb (labelled BC? on Figure 5.12), although there is still hybridizing material in the gel slot. Since two bands hybridize, it is not possible to determine which is the true BC and which is result of cross reaction with the probe 5C. Unfortunately, an attempt to resolve this question by rehybridization of the blot with a distinct probe (probe 5A), which does not detect cross reacting bands, failed.

In lanes containing the chromosomes of 1.22j' trypanosomes a

band is detected at approximately 4Mb, which is not seen ILTat 1.2 DNA. This probably represents hybridization to the expression linked copy chromosome in these trypanosomes, which, as before, has not been previously resolved from the slot material.

5.9 Attempt to clone a 1.22 expression linked copy gene locus.

The unusual nature of the basic copy locus for the 1.22 gene (Shiels et al., submitted) prompted an examination of the bloodstream expression locus for this gene, for the purpose of comparison. It was decided to attempt this using DNA from the 1.22e population, since this contained two expression loci. This was advantageous because a single representative genomic library might allow the isolation and characterization of two expression loci. Additionally, it has been found in the past by others that expression sites can be unclonable, probably because of the repetitive nature of the barren region (Van der Ploeg et al., 1982, discussed in chapter 4 of this thesis). By producing a library from DNA with two loci containing ELCs of the 1.22 gene, it was hoped that the probability of success would be increased; if one site were refractory to cloning, the other might be present. The library was constructed by ligating 100ng of EcoRI digested 1.22e DNA into EcoRI digested lambda DASH. This vector will accept DNA inserts of between 9 and 22Kb, an insert size compatible with the derived EcoRI map for both 1.22e ELCs (Figure 5.10). Because there was insufficient DNA for size selection, it was possible that the basic copy 1130bp EcoRI fragment would clone as a multiple insert in addition to the ELC fragments, producing a false positive when the library was hybridized with probe 5E of pMG7.1-1. To minimize this, the ligation of the library was carried out in a relatively large volume (30ul), with the genomic DNA at low concentration with respect to the vector. This would favour the recovery of inserts derived from larger fragments. Additionally, the library was screened with two probes, probe 5E and a 400 bp EcoRI-PvuII fragment of pMG7.1-1 (corresponding to probe 5F, Figure 5.1). Probe 5E would detect the region immediately 3' to the transposition limit and would

detect both basic copy and ELC derived inserts, while probe 5F, being beyond the limit of the transposed segment, would detect only basic copy derived fragments. Thus, plaques hybridizing with both probes would contain only the 1130 bp basic copy fragment and could be rejected, while plaques positive with probe 5E, but not with probe 5F, should be ELC derived.

Using the spi selection system for lambda replacement vectors (see section 4.14), a total of 60,000 recombinants were plated and screened, in the manner described above. This revealed no plaques which hybridized to one, but not the other probe, although 4 plaques were detected that hybridized to both probes and presumably represented a cloning of the basic copy fragment, in combination with others (not shown). The failure to detect an ELC-derived recombinant was probably not due to the screening of insufficient plaques because screening of the same filters with the EcoRI/XbaI fragment of pMT1.61-1 had detected more than 8 copies of the 22kb EcoRI fragment in the 1.61 locus (section 4.14). This indicates that both 1.22 ELCs were absent from, or underrepresented in, the library. The potentially highly repetitive nature of a 1.22 ELC extending for greater than 10 Kb discouraged further attempts to clone these loci.

5.10 Confirmation that the expression linked copy was the one transcribed.

Despite the presence of an ELC, there was a possibility that the 1.22 gene was still being expressed from its basic copy locus in the populations analysed above. This would not be unprecedented: the conversion of transcriptionally silent expression loci has been noted frequently by others (Myler et al., 1988; Aline et al., 1989). To test this possibility, a run-on transcriptional analysis was performed upon nuclei derived from 1.22j' trypanosomes expressing ILTat 1.22 in established bloodstream infection. The result, presented in Figure 5.13, indicates that it is, indeed, the ELC that is transcribed. The hybridization observed extends from the VSG gene until the 70bp repeats, but not beyond. Thus, fragment 1 is positive (though

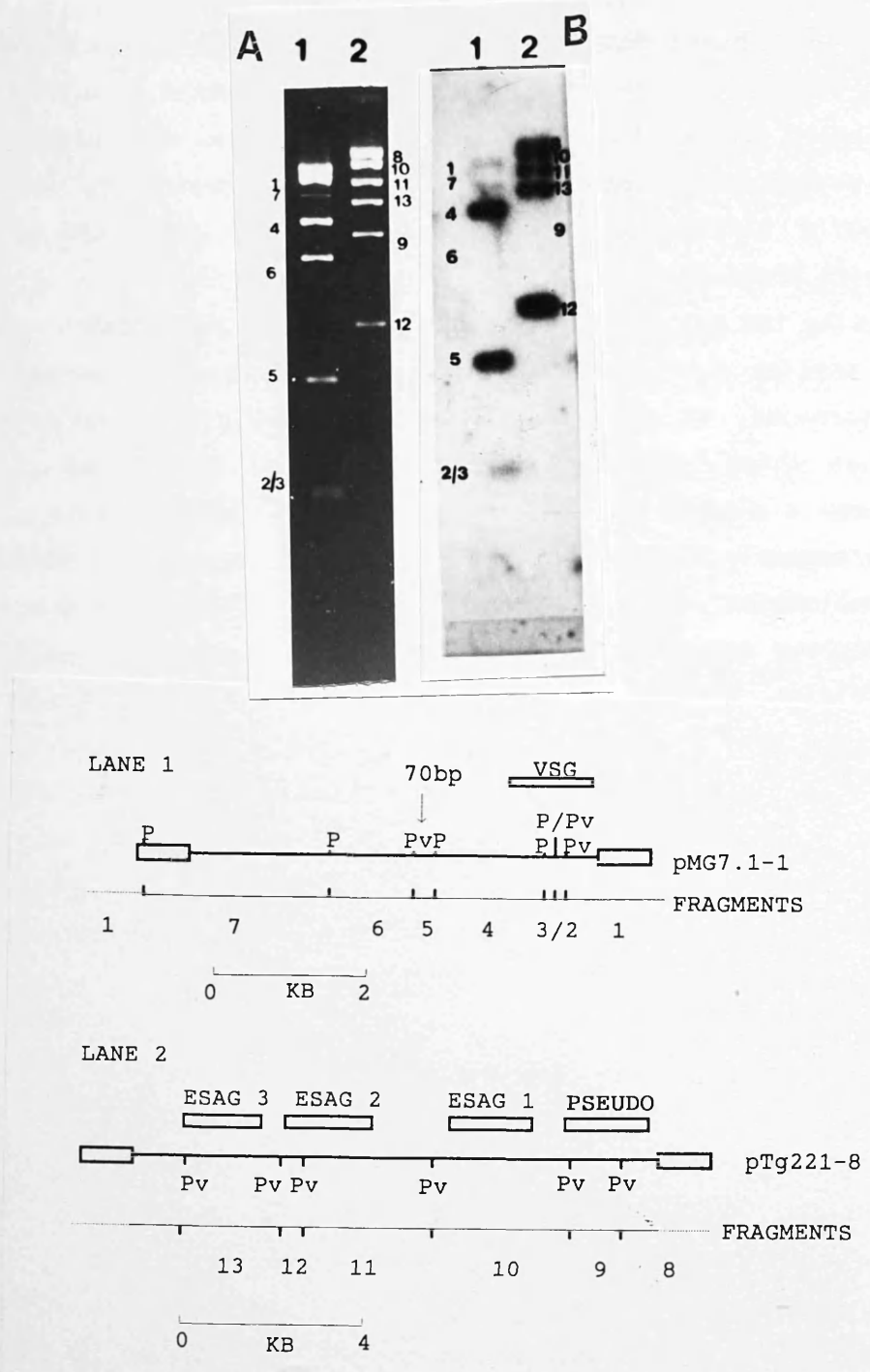


Figure 5.13; The ELC of the 1.22 gene in 1.22j' trypanosomes is probably the one transcribed. Panel A shows the ethidium stained gel, panel B is this gel blotted and hybridized with nascent transcripts from 1.22j' trypanosomes. Lane 1 is pMG7.1-1 digested with PstI and PvuII. Lane 2 is pTg221-8, a plasmid derived from a bloodstream expression locus containing ESAGs 1,2 and 3 (Kooter et al., 1987). Below is shown the relative positions of the DNA fragments on either pMG7.1-1 or pTg221-8.

weak; it has only 400bp of trypanosome derived DNA with the rest being vector), as are the co-migrating fragments 2 and 3, fragments 4 and 5. Beyond fragment 5, however, which contains the 70 bp repeats, there is no hybridization, this being most clear for band 6. This is what would be expected for ELC transcription, since the transcribed locus only shares homology with the basic copy locus downstream of the limit of transposition. This contrasts with the situation with expression from the basic copy gene locus in metacyclic derived clones, where transcription is seen to extend 2 kb upstream of the 70bp repeats, and hybridization to band 6 is clearly detected (S.Graham, unpublished). Lane 2 of Figure 5.13 represents hybridization of the run-on products to pTg221-8 from the VSG 221 expression site (Kooter et al., 1987). The hybridization on this plasmid indicates that ESAGs 1, 2 and 3 are transcribed (bands 10, 11, 12 and 13 are positive), while there is no hybridization to the pseudo VSG gene from the 221 locus (band 9), as would be expected.

Since the 1.22j' ELC had a barren region of approximately 12 Kb (Figure 5.11), most of which is probably composed of 70 bp repeats (Shah et al., 1987), it would be expected that a very intense hybridization of run-on products would be detected in Figure 5.13 on the 70 bp containing region of pMG7.1-1. This is not seen, the intensity of hybridization to fragment 5 being, instead, approximately equivalent to the intensity of band 4, which is 1.5kb in length. This apparently poor transcription of the 70bp repeats is possibly an artifact of the run-on assay. Alexandre et al. (1988) noted a similar effect and ascribed it to great instability of transcripts from the barren region, although they used a far longer period for "running on" which may have allowed more primary transcript processing than in the shorter incubation period used here. An alternative explanation is that the polymerases inefficiently transcribe this very repetitive sequence under the conditions used for the run-on assay.

Lane 2 of the blot also shows an abnormally intense hybridization intensity to the fragment containing the 5' end of ESAG 2 (band 12). This might be due to one, or both, of two

possibilities. The first concerns the conservation of ESAGs between distinct trypanosome stocks and distinct expression sites. The plasmid insert in lane 2 of Figure 5.13 was derived from a different stock of trypanosomes (S427) and from a different expression site (VSG 221) to that used for this run-on experiment. Since ESAGs appear to be quite diverse (probes for ESAG 1 detect 17-25 bands of varying intensity; Cully et al. 1985), it is possible different ESAGs on the 1.22 bloodstream expression site will show variable similarity with genes in the VSG 221 site. Thus, the greater hybridization to ESAG 2 might reflect a higher degree of similarity between the 1.22 and 221 expression sites for this gene than for other expression site genes. This is in agreement with a previous analysis, where ESAG 2 has been found to show relative conservation between expression sites in the S427 stock (Kooter et al., 1988). The alternative possibility is that the ESAG 2 gene is transcribed from more than one expression locus in the 1.22j' population, thereby increasing the production of transcripts able to hybridize with the corresponding region on the plasmid clone. Although this contrasts with models for a single active expression site in trypanosomes, an uncoupling of ESAG 2 from VSG transcription has been observed by others, at least in insect stages of the parasite (Alexandre et al., 1988).

5.11 Sequence of the transposed segment

The entire nucleotide sequence of the 1.22 gene basic copy locus has been determined from the HindIII site within the VSG gene (at the 3' limit of pMG7.1-1) to the PvuII site upstream of the 70 bp repeat containing region. This approximately represents the portion of the expression site cloned in pMG7.1-1 which is duplicated when the 1.22 gene converts bloodstream expression loci. This sequencing project was initiated at a time when it was thought that the start of transcription for the 1.22 expression site used in metacyclic derived trypanosomes was within the area of the 1.22 basic copy locus which is transposed into a bloodstream expression locus during established infection (i.e

within, or downstream of, the 70bp repeats). Thus, it was hoped to identify the signals responsible for M-VSG gene activation. Although this was subsequently found not to be the case, the project was completed with the following aims;

(i) To derive information about the predicted VSG protein; i.e was its predicted primary and secondary structure consistent with those derived for other, typical bloodstream, VSGs?

(ii) To determine whether there were any open reading frames within the transposed segment other than the VSG itself which could potentially encode protein.

It was also felt that an examination of the nucleotide composition of the 1.22 gene and its environs might provide some, albeit tentative, information concerning the susceptibility to conversion by other bloodstream VSG genes. For example, I wondered whether the relative stability of the 1.22 gene in the field over long periods (Barry et al., 1983) might have permitted a drift in the codon usage toward that of housekeeping genes and away from that of bloodstream VSG genes, which show a complete absence of bias, possibly due to their rapid evolution (Michels,1986).

The strategy chosen for the sequencing project was the cloning of specific restriction fragments of pMG7.1-1 into bacteriophage M13 vectors. These could then be used to produce single stranded DNA template for DNA sequencing using the dideoxy method, modified for use with T7 polymerase (Tabor and Richardson, 1987). This was chosen in preference to other approaches, such as the production and sequencing of nested deletion fragments of the region, because of the large amount of restriction site information available for pMG 7.1-1.

The overall strategy for the project is summarized in Figure 5.14, with the sequence being assembled using the GCG sequence analysis programmes (Devereux et al., 1984). The area has been sequenced on both strands throughout, with the exception of one area, represented by an open box in Figure 5.14. The lack of

restriction sites upstream of the XbaI site in this area had necessitated the use of synthetic oligonucleotides to allow polymerase priming from within an existing clone spanning this region. These olionucleotides were synthesised by Dr V. Math at the Department of Biochemistry, Glasgow University. Two oligonucleotides were constructed which were the complement of each other, thereby allowing sequence to be derived both upstream and downstream of the region to which the oligonucleotide hybridized (marked "oligo" in Figure 5.14). Whilst easily interpretable data were derived from the oligonucleotide priming the sequencing reaction running toward the upstream Dra I site (sequence 12, Figure 5.14), sequencing in the opposite direction produced uninterpretable data. This appeared to be because the oligonucleotide was able to anneal to two sites; the expected site and an additional one, either in another part of the phage insert or within the vector sequence. Further attempts were not made to attempt to sequence this downstream region; the sequence for the area had been determined on the other strand without difficulty (sequence 13, Figure 5.14).

5.12 The ILTat 1.22 VSG gene nucleotide and predicted amino acid sequence.

Analysis of the nucleotide sequence (Figure 5.15) on the 1.22 transposed segment reveals only one extensive open reading frame beginning at nucleotide 1463 (numbering from the PvuII hybrid site at the start of the sequence) and extending down to the terminal HindIII site in pMG7.1-1. The first methionine codon begins 187bp downstream of the beginning of this open reading frame which then extends for a further 886 bp, predicting a protein of 295 amino acids or larger (the open reading frame extends beyond the limit of the sequenced region). Comparison of the sequence of a cDNA for the ILTat 1.22 mRNA (kindly supplied by Dr M.Carrington, University of Cambridge) reveals that this open reading frame is equivalent to the VSG gene coding stretch (lower sequence, Figure 5.15). The cDNA sequence is in full

agreement with the genomic sequence except at its very 5' terminus, where the two sequences diverge (boxed in Figure 5.15). This is due to the presence of part of the mini exon sequence on the 5' of the cDNA. This sequence is spliced *in trans* onto apparently all trypanosome transcripts (see section 1.10), a fact which was used in the preparation of the cDNA clone; complementary oligonucleotides to the spliced leader and a 14 nucleotide sequence relatively conserved between VSG gene 3' ends were used to prime cyclical amplification of the VSG cDNA (M.Carrington, personal communication). The divergence between the cDNA and genomic sequence is at an AG dinucleotide, the consensus 3' splice acceptor for *trans* splicing. Upstream of this site is none of the branch point sequences characteristic in mammalian or yeast *cis* splicing, in agreement with analyses of other trypanosome genes (Mount *et al.*, 1982). There is, however, a relatively pyrimidine-rich stretch preceding the splice site (overlined in Figure 5.15), which has been suggested to function in message splice site selection in trypanosomes (Layden and Eisen 1988). There is also an appropriately spaced motif which is partially complementary to a region on the trypanosome small ribonucleoprotein U2 which might be involved in base pairing during splice site recognition (Pazelt *et al.*, 1989). This site also contains the sequence TTTCA which resembles the TTTCR (where R is purine) motif which appears upstream of many splice sites (reviewed in Laird, 1989).

There is the potential for a second splice site upstream of the one seen here, which is located at nucleotide 1491 (over crossed in Figure 5.15). This potential splice site has the conserved C/TNNAG (where N is any nucleotide) splice site seen frequently at other trypanosome splice sites and also has a polypyrimidine tract at a similar spacing to that on the more 3' splice site. The potential for alternate splicing of VSG gene transcripts has been seen by others (Layden and Eisen, 1988; Coquelet *et al.*, 1989). There is no evidence of introns within the 1.22 gene, in agreement with all other trypanosome genes studied to date.

The 1.22 gene nucleotide sequence and the predicted protein

have been analysed for relationship to sequences in the Genbank and NBRF (National Biomedical Research Foundation) databases using a modified Wilbur and Lipman (1983) algorithm implemented on the GCG sequence analysis programmes. As expected for the predicted N-terminal two thirds of a VSG protein, no significant homologies were found.

There is no available protein analysis for ILTat 1.22 and so the mature N-terminus of the protein can only be estimated. However, the relative conservation of the position of the first cysteine in analysed VSG proteins (Olafson et al., 1984) allows an estimate for the N-terminal amino acid. It has been found in all VSG sequences analysed that there is a cysteine residue located approximately 15-17 amino acids from the mature N terminus. Analysis of the ILTat 1.22 predicted sequences reveals a cysteine residue 31 amino acids down from the first methionine and no other for a further 112 amino acids. If the alanine residue 16 amino acids upstream of this cysteine marked the mature N-terminus, then this would predict a loss of 16 amino acids from the protein during maturation. This stretch of relatively hydrophobic amino acids is likely to represent a cleaved N-terminal signal sequence, as has been seen for other VSGs, although in these cases the stretch is normally longer, extending between 20 and 30 residues (McConnell et al., 1981).

The predicted amino acid sequence has a number of other features in common with published VSG primary sequences. Using an arbitrary numbering system, placing the first cysteine at residue 16, the predicted sequence displays hydrophobic residues at positions 11, 12, 15 and 25 , which is similar to other VSGs (Olafson et al., 1984). Also, there are cysteine residues at positions 16, 128, 142, 200, 243 and 263 in ILTat 1.22 (numbered below the predicted protein sequence in Figure 5.15). The position of the first three is very similar to that seen for other predicted VSG sequences (reviewed in Turner, 1988). There is the potential for the formation of a 126 amino acid loop between the first and third cysteine residues, which is consistent with the approximately 140 amino acid loop predicted for many other VSG amino acid sequences (Turner, 1988). The


PvuII
1 CTGGTGAACTCGCTATAAATAGTAGGTTAATTACCTCTTTTGCATGTCAGAGTTACGACG
61 TTGTATATAGCAAAAATAATAATAATAATAATAGGAGAGTGTTGTGAGTGTGTGTATATA
REPEATS->::
121 CGAATCTTATAATAAGAGAAGCAGTAATAATAATAATAATGATAATAATGATGATGATGA
::
181 TGATAGTAATACGAGGAAAGTGTGTTGTTAGTGCAAAAGAATGGACAAGCCACAGCAGAC
::
241 GCCTTCCAACAATTTAGTGGCCCCAGGAAGGGAACCAAAAACTGAGAGAGCGCGAGGAAT
:<-
301 GTTAAAGGGAAAAGACAATAAGACAGAGGAATTGAGAAGGAGATTTGAATATATGCAGGT
361 TTCGCAAGACAAAACGATACTGTAAATTCTCTTGCACGCTGTAAATTTCTCTTGCACGAA
Pst I
421 ACAAAAGCGGCACGGCGCTGCAGGAAGAGGAGGCCTCGCTTAACCCGGGTGGAGTAGCCA
481 TAAAATCTGCGCCTATAAAAGCAGTAGTCGACATAGAAGGGAAGGCAGATGTAAATGAAA
541 TGAAATGACAAAATAATAAACATACCATCACTAAAGAAAATGGTACGATAATAATTAATA
601 ATGCTCCGCATTTGACATAAAAGGGATAAATTCCACCCTTGAAACAATGGTCGCAACTGC
EcoR I
661 AATTATTAATTTTATGGCGAATTCTAATATGCAACTGAAAGAAAAACAAACATAAAAAAA
721 TATTTGATAAGTGTGATGTAATTTCGAAGTATTCTTATCTAAAATTTTTTTAATGTTCAA
.Dra I
781 CATTTTTTGTAAAACATATAACATTATATGAAGAAAATCTTTTAAAATATTACAACTATT
Dra I
841 GAATTTTATGTGAATAGAAAAAAGAATTAAAACGCATGAAATGCCTAAAATAGCCCATTT
901 TAAAAGTTGAGAAAGAACCCTATAAAGTGCATGAGGTTTGTAAATGACACTCCGCAGACG
961 CGCTATTGGAGCAAGTGACAGTCCCTGCATCCATGGCACAGCAGCTAATTCACGTGAATA

```
1021 CATACTTCCGAGGCATTGATGTAGCATATTAATCATGTGGAAAAAAATCAGAATGCAGTT

1081 CGAACTAAAATGCCTAAATCTTGAGTGCAAACAGACGACACGCACCAAGGATATAAAAAT

1141 GTAATGAAAGGAAAGCCATGTCTGTGACGAAGAGCAACTATACCTGTATATGTGGCCAGA

1201 ATATAATTGATAATTGTCGAAAAACAAGAGCAAATACCATAAGGTTTGTGTTTTTACACA

1261 ATAAAATGTGTAAGATATTTGAACAGATGCGTGGAAGCGAGTTGCTTACAGTGAAGACAG

1321 ACGCAATACGAAAATGGAAACACGACATCAGTGTGATCATGCTAGAAGTAATAATAGGTG

1381 AATCACTGTAAAAAGTTCAACAAAAGCAGCAGAAATGCTGTTTTTGTAGACAAATCACAA

              Xba I                                    ++++++
1441 CGAAGGGCTCTAGAAATTCTAAAAACTTTTATTTATACGTGTACTAGTTTCTGTAGCATC
                              > ORF START      Alternate splice site?

1501 CGCAGTTAAAGACTGTCGTCTACAATGCCGCGGGCGTTTCACTCGCTTAATACTTACCCC
                                          [CAATGAGACTA]
                          Splice junction
1561 TCACAAACTTAATGAGGCAAAGACGCTATCTATCGCCATTTTAGCCACCTACACACGGTG
cDNA> TTTCTGTACTATATTGACGCTATCTATCGCCATTTTAGCCACCTACACACGGTG
          mini exon

1621 GCACGCAACACAAGCAGTAAATTCTCTAATGGACACAGCCCAGGTTTTCGCCCTTTTTTA
     GCACGCAACACAAGCAGTAAATTCTCTAATGGACACAGCCCAGGTTTTCGCCCTTTTTTA
                                MetAspThrAlaGlnValPheAlaLeuPheTy
                                -16-15-14-13-12-11-10-9 -8 -7 -6

1681 CATGGCGACCGTCATGGCAGCAGGGACAAAAAAACAAAGCGTCTCAAGCAGTGTCAGACCC
     CATGGCGACCGTCATGGCAGCAGGGACAAAAAAACAAAGCGTCTCAAGCAGTGTCAGACCC
     rMetAlaThrValMetAlaAlaGlyThrLysAsnLysAlaSerGlnAlaValSerAspPr
      -5 -4 -3 -2 -1  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15

1741 TTGCAGCGAAATTCATTTCGACGAGCAACTAGCGAACTACTTCGAAAACGAAGTTTCCGC
     TTGCAGCGAAATTCATTTCGACGAGCAACTAGCGAACTACTTCGAAAACGAAGTTTCCGC
     oCysSerGluIleHisPheAspGluGlnLeuAlaAsnTyrPheGluAsnGluValSerAl
      16 17 18 19 20 21 22 23 24 25

1801 GGCAACGACGCAGCTCGACGAGAATCAGAACTTCGAACGAAGCTGGAAGCTGCTGCAATA
     GGCAACGACGCAGCTCGACGAGAATCAGAACTTCGAACGAAGCTGGAAGCTGCTGCAATA
     aAlaThrThrGlnLeuAspGluAsnGlnAsnPheGluArgSerTrpLysLeuLeuGlnTy
```

```
      Pst I
1861 TCTGCAGATGGATCACCAAAAGTCAAAAGGCGCGGCAGCGTTAGCTGCATACGCATCTAC
     TCTGCAGATGGATCACCAAAAGTCAAAAGGCGCGGCAGCGTTAGCTGCATACGCATCTAC
     rLeuGlnMetAspHisGlnLysSerLysGlyAlaAlaAlaLeuAlaAlaTyrAlaSerTh

1921 GATAAACATCAGAACGGCTGCAAACGTAAAAGCAGCGAGCGGAGAGCTGCTGACGGCAGC
     GATAAACATCAGAACGGCTGCAAACGTAAAAGCAGCGAGCGGAGAGCTGCTGACGGCAGC
     rIleAsnIleArgThrAlaAlaAsnValLysAlaAlaSerGlyGluLeuLeuThrAlaAl

                                          PvuII/PstI
1981 TAGCCTACTCCGGCAGCGAGCAGCAAATGTCTCAGCCGCGTTTCAGCTGCAGGGCCAAGG
     TAGCCTACTCCGGCAGCGAGCAGCAAATGTCTCAGCCGCGTTTCAGCTGCAGGGCCAAGG
     aSerLeuLeuArgGlnArgAlaAlaAsnValSerAlaAlaPheGlnLeuGlnGlyGlnGl
                    potential N-glycosylation site

2041 TGTCATCAAGCTGGGCACACCAGATATCGACAACGGAGCGAAGTCGATAACGCATGCAGA
     TGTCATCAAGCTGGGCACACCAGATATCGACAACGGAGCGAAGTCGATAACGCATGCAGA
     yValIleLysLeuGlyThrProAspIleAspAsnGlyAlaLysSerIleThrHisAlaAs

2101 CGCCGGCTGCAACTACGCAGCCATTAGCAAGACGGTTCCAACACAGCGATGCACACCGCC
     CGCCGGCTGCAACTACGCAGCCATTAGCAAGACGGTTCCAACACAGCGATGCACACCGCC
     pAlaGlyCysAsnTyrAlaAlaIleSerLysThrValProThrGlnArgCysThrProPr
          128                                              142
                    PvuII
2161 GCAACAGCAAGCTGACACGATTACAGCTGCGGACATGCAACCAGACAAGCTAGACGAGCT
     GCAACAGCAAGCTGACACGATTACAGCTGCGGACATGCAACCAGACAAGCTAGACGAGCT
     oGlnGlnGlnAlaAspThrIleThrAlaAlaAspMetGlnProAspLysLeuAspGluLe

2221 TCAACTGATAACAGAAGCCTACACGACAACGATAACAATAGCAGCGAGTGCTTACAGCAA
     TCAACTGATAACAGAAGCCTACACGACAACGATAACAATAGCAGCGAGTGCTTACAGCAA
     uGlnLeuIleThrGluAlaTyrThrThrThrIleThrIleAlaAlaSerAlaTyrSerLy

2281 AGGGACACCGGCGACAGGACACACCGTATACACTTACGGCAACTGCCAAAGCACAGGTGG
     AGGGACACCGGCGACAGGACACACCGTATACACTTACGGCAACTGCCAAAGCACAGGTGG
     sGlyThrProAlaThrGlyHisThrValTyrThrTyrGlyAsnCysGlnSerThrGlyGl
                                                  200

                                        BamHI
2341 GAGCGCTTCGGCACAGCTAGGTGACACCCATGCACTCGGGATCCATGTCAAGACGATTGG
     GAGCGCTTCGGCACAGCTAGGTGACACCCATGCACTCGGGATCCATGTCAAGACGATTGG
     ySerAlaSerAlaGlnLeuGlyAspThrHisAlaLeuGlyIleHisValLysThrIleGl
```

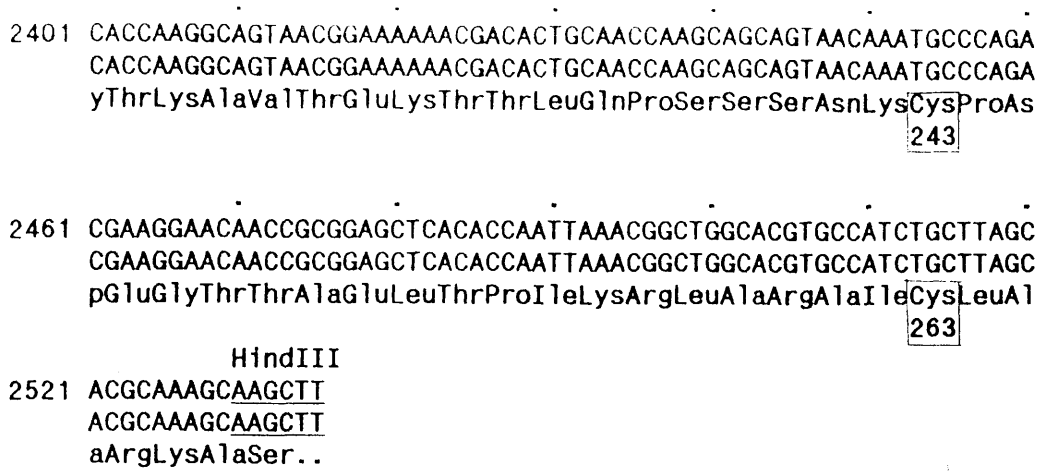


Figure 5.15; Sequence of the ILTat 1.22 gene transposed segment. Restriction sites are shown as underlined. The 70 bp repeats are indicated by colons below the sequence and delineated by arrows. "ORF" marks the beginning of the 1.22 gene open reading frame. Features potentially concerned with addition of the mini-exon sequence to the mRNA are boxed, as are cysteines in the predicted amino acid sequence. The sequence in parentheses indicates a potential branch point recognition region for the trypanosome U2 small ribonuclear particle (Patzelt et al., 1989). The predicted amino acid sequence is numbered using the first cysteine as an arbitary position 16. The nucleotide sequence is numbered from the PvuII hybrid site. Throughout, the upper line is genomic sequence, and, where present, the second line is the sequence of the cDNA for the ILTat 1.22 gene. The lowest lines give the three letter codes for the predicted amino acid sequences.

GENE	ANALYSIS	BASE COMPOSITION (PERCENT)			
		A	T	G	C
1.22	TOTAL NUCLEOTIDES	31	16	24	28
1.22	THIRD BASE FREQUENCY (ALL CODONS)	31	13	24	31
1.22	THIRD BASE FREQUENCY (ISOACCEPTING CODONS)	37	16	27	19
OTHER VSGs	TOTAL NUCLEOTIDES	36	15	24	26
OTHER VSGs	THIRD BASE FREQUENCY (ALL CODONS)	38	12	20	29
HOUSE-KEEPING	TOTAL NUCLEOTIDES	12	25	27	36
HOUSE-KEEPING	THIRD BASE FREQUENCY (ALL CODONS)	23	23	28	26
1.22 TRANSPOSED SEGMENT (NOT VSG)	TOTAL NUCLEOTIDES	38	26	19	18

Table 5.1 Nucleotide composition and codon frequency of the ILTat 1.22 gene and upstream environment compared to the average of three other VSG gene sequences and eight housekeeping genes (Michels, 1986)

Facing page 149

the electron withdrawing inductive effect of the chlorine atom enhancing chain scission. Main chain scission proved especially awkward, however, when attempts were made to prepare the monochlorinated derivative as the products obtained in these cases were oils.

When the reaction was carried out in air, the IR spectra showed no significant increases in the carbonyl stretching absorption (1775 cm^{-1}) relative to that for the ether C-O-C asymmetric stretching band (1120 cm^{-1}). This suggests that under these experimental conditions oxygen does not enhance chain scission via the formation of carbonyl compounds. Kagiya et al[111] investigating radiation induced degradation of PEO in the atmosphere of chlorine compounds observed the appearance of a variety of carbonyl absorption bands in the IR spectrum when PEO powder was subjected to γ radiation in air. These included ester type carbonyl (1750 cm^{-1}), aldehydic (1733 cm^{-1}), ketonic (1721 cm^{-1}), and carbonyl (1715 cm^{-1}) absorptions. In addition, hydroxyl bands were observed (3400 cm^{-1}). Similar oxidative degradation occurred when UV radiation was employed. In similar experiments in CCl_4 little change in the IR spectrum was noted apart from the appearance of an acid chloride absorption. Chlorine appeared to have the greatest effect on the degradation of PEO when the polymer was irradiated with UV light. This was seen by the decrease in molecular weight and the presence of IR bands at 1760 cm^{-1} and 800 cm^{-1} due to acid chloride and C-Cl ansorptions respectively in the irradiated product. Also it was noted that when UV irradiation of PEO was carried out in vacuum, cross linking took place. Thus the results obtained in the present investigation

nucleotide composition in favour of A and T and against C.

The nucleotide compositions for both the coding and non coding stretches of the 1.22 transposed segment were determined using the "Composition" program in the GCG sequencing analysis software. The codon usage for the portion of the ILTat 1.22 gene was determined using the "Codon Frequency" program. Table 5.1 reveals that the 1.22 gene is remarkably similar in both overall and third base nucleotide composition to other VSG genes. The nucleotide composition of the transposed segment, however, seems particularly rich with A and T with respect to the coding regions of housekeeping genes.

Discussion

This chapter has examined the structure of the locus for the gene encoding ILTat 1.22 and has studied the activation of this gene in established bloodstream infection. In a previous analysis this gene has been found to be expressed *in situ* in metacyclic clones using the model line of trypanosomes employed in Chapter 3 of this thesis (Shiels *et al.*, submitted). Shiels *et al.* (submitted) have also investigated the 1.22 gene environment; most of it was less repetitive in the genome than the ILTat 1.61 expression locus (Chapter 4, this thesis) and far less reiterated than bloodstream VSG expression sites (Kooter *et al.*, 1987; Alexandre *et al.*, 1988; Pays *et al.*; 1989a). The experiments described here have confirmed earlier analyses which indicated that when the gene is expressed later than about 7 days after leaving the tsetse fly, it only does so following duplicative transposition (Cornelissen *et al.*, 1985a); the gene converts what is presumably a normal bloodstream VSG expression site. This is compatible with the independence of M-VAT and B-VAT expression mechanisms: because the metacyclic expression site has little in common with a bloodstream expression site, the 1.22 gene requires to be transposed for activation at this time. Analysis of the expression locus for the presence of the transposition associated 70bp repeat motifs has revealed that, like on the ILTat 1.61 gene locus (Chapter 4, this thesis), the extent of these repeats is very short. This is in contrast to what has been found for other M-VAT genes where none has been detected (Lenardo *et al.*, 1986) and strikingly different from telomeric bloodstream VSG expression loci where several kilobases of repeats are often present. Despite their brevity, however, the very short region containing the 70bp repeats has been found to act always as the 5' conversion limit when the 1.22 gene undergoes duplicative transposition. This, together with the observed stability of ILTat 1.22 in the M-VAT repertoire of the EATRO 795 stock with time (Barry *et al.*, 1983), has implications for the mechanism of expression site conversion.

1. M-VAT Switching Mechanisms

It has been proposed that VSG gene conversions may occur by employing a non-specific mechanism because the observed switching frequency (10^{-6} to 10^{-7} per cell per generation; Lamont et al 1986) is approximately equivalent with the levels of background mitotic recombination in yeast (Pays 1985; Borst and Greaves, 1987; Timmers et al., 1987). Using this model, homologies between VSG gene loci would initiate conversion, with regions showing extensive homology being more frequently involved than those with less homology. Thus, the high degree of similarity between distinct bloodstream expression sites at telomeres could allow their frequent interconversion, while the homologies between chromosomal internal genes and expression sites, being restricted to the VSG gene 3' region and upstream flanks (70bp repeats and regions of the transposed segment), would be less frequently involved in recombination.

The telomeric nature of expression sites might also enhance the initiation of such a non-specific conversion mechanism. For example, the double strand breaks at the ends of chromosomes might be recombinogenic (Szostak et al., 1983), as might the nicked strands on replicating telomeres (Blackburn and Challoner, 1984; Meselson and Radding, 1975). Inter-telomeric recombinations might also be promoted by the non-Watson and Crick base pairings which can form within the telomere repeats (Henderson et al., 1987), or be stabilized by the possibilities for inter duplex association, mediated by Hoogsteen bonding between runs of guanosines (Williamson et al., 1989). This bonding, which can result in the formation of a four stranded DNA molecule, has been suggested to play a role in systems involving DNA recombination, with interactions between the "S-sequences" involved in immunoglobulin heavy chain gene rearrangement being one example (Sen and Gilbert, 1988). With respect to VSG genes, such telomeric events could promote conversion progressing upstream, away from the chromosome end and terminating in homologous regions further along the expression site such as 70bp

repeat motifs or ESAGs (Florent et al., 1987).

Although compatible with interconversion between highly related bloodstream expression sites, the above scenario is incompatible with the frequency and directionality of conversion events involving the 1.22 gene. Firstly, homology alone does not seem to govern duplicative transposition of the gene. The ILTat 1.22 basic copy locus has only 1.5 70bp repeat motifs in common with bloodstream expression loci but generates an ELC consistently after switching away from in situ expression. This contrasts with the M-VAT 4 and 7 gene loci which are also telomeric, have greater similarity to bloodstream expression sites (they have a conserved ESAG), but have no 70bp repeat motif, and have never been reported to generate an ELC (Son et al., 1989). Similarly, chromosomal internal VSG genes without any repeats do not seem to be able to undergo duplicative activation (Pays et al., 1983d). This implies that a single intact 70 bp repeat motif might be necessary and sufficient for efficient ELC generation, regardless of the presence other homologies, and in contrast to previous proposals in which the extent of repeats associated with the gene conversion donor was suggested to determine its transposition frequency (Timmers et al., 1987).

Secondly, a conversion event initiated at the chromosome telomere cannot explain the stability of the 1.22 gene in the M-VAT repertoire: if telomeric events 3' of the VSG gene were driving the conversion process, the telomeric 1.22 gene would be itself converted as frequently as it converts other telomeres. This is not so; it can be displaced from its bloodstream expression locus (J.D. Barry, personal communication), but has not been seen to be replaced in the metacyclic expression locus telomere (exept rarely, where the whole locus appears to have been absent; Shiels, 1990). An argument against a 3' to 5' conversion orientation has also been raised recently by others; Scholler et al.(1989) have observed that a newly converted chromosome has a telomere length identical to that of the conversion donor. They explain this in terms of an upstream initiation, conversion migration and then termination at the telomere.

The observation that the 1.22 gene readily acts as a conversion donor employing its 70bp repeats, but is itself relatively resistant to conversion, can be explained in terms of a more directed conversion mechanism than background mitotic recombination processes (Barry, 1989). In this model, the long arrays of 70bp repeats on bloodstream expression sites might be recognised by a specific endonuclease, analogous to the HO endonuclease in yeast mating type switching (Kostricken et al., 1983), producing a double strand break. This could then be converted by the donor sequence (the ILTat 1.22 tranposed segment). In this process the 1.22 gene locus, like chromosomal internal VSG genes, would be relatively resistant to conversion because the probability of their very short 70bp repeat motif being recognised and cut would be low.

A conversion mechanism of this type would be advantageous to the trypanosome; as well as keeping the M-VAT VSG system apart from the bloodstream system (Chapter 3, this thesis; Shiels et al., submitted), it would limit the likelihood of bloodstream expression sites becoming internalised into chromosomes. Also, by using a specific conversion initiation mechanism reliant upon a homology block found with the vast majority of VSG genes, the trypanosome could avoid the possibility of becoming trapped into a continual cycle of conversions between very related genes (Timmers et al., 1987).

Support for this form of specific conversion mechanism has come from recent analyses on the frequency of antigen switching. In trypanosome lines able to be transmitted through flies and, thus, more closely resembling those in the field than the rodent adapted lines previously analysed, it has been found that switching frequencies may be as high as 10^{-1} to 10^{-3} per cell per generation (Turner and Barry, 1989). Similarly, an analysis of transposition events occurring without a change in the expressed VSG gene has implied that conversion frequencies may be higher than had been supposed, even in adapted lines (Myler et al., 1988; Aline et al., 1989). Whilst these elevated frequencies could be accounted for by a system reliant on simple homology (background telomere recombination frequencies may be very high;

Pluta and Zakian, 1989) it could not easily account for the reduction of this frequency with adaptation to rodent hosts.

A relatively specific system is not contradicted by the observed frequency with which sequences other than 70bp repeats mark the transposition limit when telomeric genes are activated by conversion (Kooter et al., 1988; Pays et al., 1983b; Lee and Van der Ploeg, 1987). Firstly, it is quite possible that, after the initiation of conversion, the great similarities between bloodstream expression sites could result in extensive branch migration of a Holliday intermediate in the process, resulting in resolution of the conversion event at a point distant from its initiation. This migration could occur both sides of the conversion initiation point; the double strand break conversion model (Szostak et al., 1983) proposes the involvement of two Holliday intermediates at either side of the double strand gap which initiates conversion. This could permit bi-directional conversion as has been observed at the yeast MAT locus (McGill et al., 1989). Although the extent of the conversion migration at one side of the HO endonuclease cut site is limited in that system, this appears to be due to the presence of a specific mechanism for ensuring the polarity of the conversion process (White and Haber, 1990).

Secondly, sequences outwith the 70 bp repeats might also mark the duplication limit, if there were a degree of overlap between the frequencies of specific conversions and non specific conversions. This is not unlikely, because any recombination initiating machinery would need to be at a low level to prevent the trypanosome surface being in continual antigenic flux.

Finally, the proposal of a relatively specific conversion mechanism does not preclude the existence of other recombination processes contributing to VSG gene switching. For example, the endpoint for a number of conversion events during antigen switching has been seen to be close to, or within, the ESAG 2 sequence on bloodstream expression sites (Kooter et al 1988; Shea et al., 1987). This gene bears great similarity to both the so-called "companion sequence", which accompanies the AnTat 1.1 gene during duplicative transposition (Pays et al., 1983a), and

the region close to the proposed RNA polymerase I promoter on the VSG 118 expression site (Shea et al., 1987). Although this pol I promoter-like sequence does not appear to be the promoter controlling VSG transcription (Crozatier et al., submitted), its subsidiary activity may act to promote conversion in an analogous way to the HOT 1 sequence in Saccharomyces cerevisae (Keil and Roeder, 1985). This sequence is closely linked to the promoters for ribosomal gene expression and appears to promote the initiation of mitotic conversion, provided its locus is transcriptionally active (Voelkel-Meiman et al., 1987).

A system for inducing gene conversions at VSG expression sites need not be as specific as the highly regulated yeast MAT locus (reviewed in Herskowitz, 1989). Indeed it might be better for the trypanosome to have a more sloppy system for controlling antigen switching as this could contribute to the relative unpredictability of the sequence of VATs produced by a population. Thus, instead of a strict target sequence, the system may simply show a preference for 70bp repeats because of their capacity for secondary structure or relative exposure in chromatin. For example, and as has been noted previously (Liu et al, 1983), the TRR stretch (where R is purine) and alternating purine:pyrimidine motifs of the 70 bp repeat motif is similar to many other sequences associated with genes known to be able to undergo gene conversion and recombination. Examples include the TCA, TCC repetitive triplets seen at the flanks of the variable heavy chain genes encoding immunoglobulins (Cohen et al.,1982), the TGGGG motifs associated with the S elements of heavy chain constant region genes (Shimizu and Honjo, 1984) and the $(CT)_{27}$ "zipper" repeat adjacent to sea urchin histone genes (Kedes, 1979). As pointed out by Kedes (1979), the repetitive nature of such motifs can act to increase greatly the frequency of inter-gene base pairing which apparently precede recombination processes. Long arrays of the A-T rich 70bp repeats may also promote sequence recognition by a conversion initiating enzyme by their being relatively excluded from compaction into chromatin. For example, the Z DNA and cruciform structures such as have been predicted for the repeats (Liu et al., 1983), have been

observed to limit the assembly of histones onto DNA (Nickol et al., 1982), and can be recombinogenic (reviewed in Holliday, 1989). Similarly, the alternating purine: pyrimidine stretches which comprise one section of the 70bp repeats, when under torsional stress might form unusual helix conformations and, thus, be less readily assembled into chromatin (Eissenberg et al., 1985). Torsional stress associated with transcription might further enhance the possibilities for recombination at the repeats, if they are cleaved during the relief of this stress by topoisomerases (Tsao et al., 1989). Topoisomerase II, for example, generates transient double strand breaks in transcribed DNA and prefentially binds to both AT-rich stretches and alternating purines and pyrimidine arrays (Spitzner et al. 1990). Although this could be advantageous to the trypanosome by preferentially directing new genes into an active expression site, it could less readily account for the frequent conversions at silent VSG gene loci (Myler et al., 1988; Aline et al., 1989).

2. The Timing of ELC generation for the 1.22 gene.

Why would the 1.22 gene generate an ELC upon the inactivation of the M-VAT system, but not before? There might be a number of reasons. Firstly, the specific conversion mechanism supposed above might not become active until the bloodstream VSG system is in operation, not being required beforehand. This possibility could be addressed by looking for evidence of silent VSG gene rearrangements while the metacyclic VSG system was still extant, or during in vitro culture of procyclic trypanosomes.

Secondly, trypanosomes generating an ELC during the first few days in the mammal host might not be selected until the bloodstream system is activated and the ELC is able to be expressed. These organisms would be unlikely to increase in the population with respect to those without an ELC and, therefore, might not be detected.

Finally, with movement into bloodstream mode VSG expression, the previously active M-VSG gene might act as a preferential donor in gene conversion events because of its nuclear location.

If specific sites exist within the nucleus for M-VAT or B-VAT gene transcription or both, the formerly expressed M-VAT gene might more readily convert a newly activated bloodstream expression locus because of their close spatial proximity at the switch from one system to another. Later on, M-VAT genes might less readily convert bloodstream expression loci, being silent, and potentially compacted into chromatin.

3. The acceptor expression sites for the 1.22 gene ELCs.

Section 5.7 investigated, by restriction analysis, the recipients for conversion by the 1.22 gene and found that a number of expression sites could act as acceptors (assuming that the restriction map is indicative of the site used). Although only a very small sample was analysed, there was only one case where the same expression site seemed to be used (excluding the possible lingering ELC in 1.22f). This indicates that the gene is not being preferentially targetted to its most related bloodstream locus, and supports the hypothesis that homologies outwith the 70bp repeats are not responsible for directing VSG gene conversions. It also suggests that there is not a tight restriction on the interactions of the 1.22 telomere, which may have been one explanation for its stability in the genome.

There may still, however, be a restriction on the class of chromosome with which the locus can interact, at least at the switch from M-VAT to B-VAT mode expression. Although only one case has been looked at, the 1.22j' ELC was found to be on a very large chromosome. This is in agreement with the work of Delauw *et al*. (1987) who found, as here, that a number of sites could accept M-VAT genes and that these were confined to large chromosomes at least early in infection. It is also in agreement with the analyses of Cornelissen *et al*. (1985a). This might imply that, at the switch from M-VAT to B-VAT mode activation, M-VAT genes can only interact with expression sites on other large chromosomes. Although this obviously needs further study, this might limit the potential for these genes to be converted, as most other VSG genes are harboured in smaller chromosomes (Van

der Ploeg and Cornelissen, 1984). Experiments to localize the ELCs for the remaining expressor populations analysed in this chapter are underway (O. Shonekan, personal communication).

4. A Comparison With Other M-VAT Loci.

It is interesting to observe that both of the M-VAT loci examined in this study have been found to a have a very short 70bp repeat sequence, while those analysed by Lenardo et al. (1986) had none.

The reason for this is possibly a consequence of how the genes in each case have been selected for study. Cornelissen et al. (1985a) isolated their cDNAS from bloodstream organisms which had independently activated M-VATs, using a procedure developed by Barry et al. (1979). This involved isolating those trypanosomes expressing VATs in chronically infected rabbits immediately prior to the detection of M-VAT reactivity in their serum. This is important, because those most readily able to enter a bloodstream expression site will be preferentially selected; i.e those with the signals for transposition, 70bp repeats. In contrast, Lenardo et al. (1984) isolated their M-VAT cDNAs from populations only five days from the fly, before inactivation of the metacyclic VSG expression system. This will not have selected for those M-VAT genes most able to undergo duplicative transposition and these , accordingly, may lack 70bp repeats. An analysis of the basic copy locus for the M-VAT genes studied by Delauw et al.(1987) did not involve a search for 70 bp repeats, although no barren region was detected. These genes were expressed from an ELC, however, 10 days after cyclical transmission and may, therefore, have had a very short "barren region" as on those studied in this thesis.

5. The Transposed Segment

The section of the 1.22 locus transposed segment represented within pMG 7.1-1 has been sequenced. Most notable is the open reading frame containing the ILTat 1.22 gene coding sequence. In

all respects this gene appears like other characterized VSG genes. This is as would be expected; although the gene is first expressed in the tsetse fly salivary gland, this is only as a preadaptation for life in the mammalian bloodstream. Although the predicted primary amino acid sequence appears unrelated to other VSGs in the databases, it shows all the features which have been noted by others to show a degree of conservation in VSG sequences. Thus, the predicted VSG has an N-terminal hydrophobic signal sequence, closely followed by a cysteine with nearby hydrophobic residues, and a pair of cysteines at residues 128 and 142. The conservation of these motifs is compatible with the observed overall similarity in tertiary structure which has been seen between unrelated VSG molecules (Metcalf et al 1986). This probably reflects structural constraints on VSGs in spite of their selection for diversity. They may need sufficient conservation in shape to allow them to pack effectively on the trypanosome surface and thereby form an effective barrier to non specific immune mechanisms, for example. Similarly, a relative conservation between the secondary and tertiary structure of VSG molecules is likely to assist the trypanosome in maintaining the integrity of its coat when two VSG types are on the trypanosome surface, after a switch in the expressed VSG gene.

An analysis of the nucleotide composition for the region of the gene sequenced revealed no apparent codon bias despite the abundance of the VSG protein in trypanosomes. Both the nucleotide composition and absence of bias at the third base in the codon is in full agreement with other VSG gene analyses (summarised in Michels, 1986). For other VSG genes, it was suggested that this lack of bias is a consequence of the very rapid evolution of these sequences resulting from frequent transpositions, partial conversions and reciprocal recombinations. This is not incompatible with the absence of bias in the 1.22 gene, which is apparently quite unsusceptible to these processes (see above). In the longer term it is very likely that new genes will be brought into the metacyclic repertoire from the bloodstream repertoire by even improbable conversion events or reciprocal recombinations. Although potentially rare, the timescale of these events would be

far less than that needed for the development of a detectable codon bias.

An analysis of the non coding region of the transposed segment between the 70bp repeats and the VSG gene reveals that the nucleotide sequence is very rich in adenosine and thymidine; the GC content is 36.5% compared to 55% over the VSG gene. This has also been noted by Michels (1986) who suggested that the bias may have a role at VSG telomeric expression sites. For example, it was suggested that an A-rich sequence might be related to the requirements for a particular chromatin structure on the telomere. Alternatively, an A/T-rich nucleotide sequence might assist in the displacement of successsive transposed segments from the expression site; the higher melting capacity of A/T-rich DNA duplices could act to promote progression of a conversion event from the 70bp repeats down into the VSG gene and beyond. These interpretations are speculative, however, as more recent analyses of non coding stretches in the trypanosome genome away from VSG expression sites have revealed that they, too, are quite A-T rich (Dr P.Michels, personal communication).

Chapter 6

Concluding Remarks.

This thesis has examined two metacyclic VSG genes when activated either in the bloodstream or, indirectly, in the tsetse fly. I have found that one metacyclic VSG gene locus appears very different to the more complex expression loci used by trypanosomes established in the bloodstream. The expression site structure appears very simple, with just a remarkably short restriction site "barren" region and a sequence hybridizing with Expression Site Associated Gene 1. An analysis of the nascent transcripts derived from the expression locus has revealed that the expression site does not include this upstream ESAG related sequence, but instead appears to begin approximately 5 kb in front of the VSG gene. This is in striking contrast to the huge transcription units observed in bloodstream expression sites, which may extend for 55 kb in front of the VSG gene. I have also examined the potential signals for M-VSG gene activation by infecting tsetse flies with trypanosomes expressing M-VSG genes as they are believed to be activated in the fly and this has revealed, in contrast to the bloodstream system, that there is no preservation of the signals for expression. Finally, I have examined the reactivation of another M-VSG gene during more established bloodstream infection and have found that this always entails the generation of an additional gene copy, mediated through the very short barren region on the M-VSG gene telomere.

Taken together, these findings have emphasised the distinctions between the bloodstream and metacyclic VSG systems and provided clues to the reasons for these. The most striking features of the M-VAT system, its limited nature and predictability, for example, can be explained by the apparent absence of a duplicative activation mechanism and the simplicity of the expression site. Thus, the system might be limited because the M-VSG genes are activated without duplication, thereby limiting the number of potential M-VATs to the number of metacyclic expression sites. Alternatively, duplicative transposition might be possible in the metacyclic system, but only between a particular VSG gene subset. The predictability of the M-VAT repertoire could be a consequence of the simplicity of these expression sites: being relatively non repetitive, they are

relatively insusceptible to invasion from without and therefore relatively static in the genome.

This thesis has also suggested that the reason the parasite maintains a distinct bloodstream and metacyclic VSG system is that at each stage there are different priorities. In the metacyclics it is more important to ensure a heterogeneous repertoire than a very diverse one because this stage is concerned with the invasion of new hosts. Although in the laboratory, the bloodstream VSG system could accomplish this efficiently, in the field the situation is very different. The continual exposure of game animals to these parasites can lead to the development of a tolerance to infection. This tolerence permits the development of immunity to the trypanosomes, with the predominant VAT types most frequently challenging and, thus, reinforcing this immunity. In the absence of a separate pool of metacyclic VSGs, it is these VATs which would be expressed first of all in a new host and there would be a high probability of their elimination. By maintaining the metacyclic repertoire distinct from that employed in the bloodstream, the host is presented with a parasite population against which it is unlikely to have maintained immunity, because these metacyclic VSG genes are rarely activated in chronic bloodstream infection. This hypothesis is not contradicted by the apparent frequency with which bloodstream expressors of the ILTat 1.22 gene were generated in chapter 5 of this thesis. These MVAT genes were being moved into a bloodstream site at a special time; immediately following their expression *in situ*

Once established in the bloodstream, it is the diversity of the VSG system which is most important. If the metacyclic system were still extant after some time in the blood, the parasite would rapidly exhaust its VSG repertoire and its probability of being transmitted to a new tsetse fly vector would be reduced. Instead, then, it is advantageous at this stage to activate a distinct repertoire which shows far less predictability. Whilst this will inevitably result in a high frequency of activation of more predominant VATs, against which immunity will quickly

develop and be maintained, the population will be in sufficient numbers in the bloodstream to ensure that even VATs with a very small probability of expression can appear and evade immune responses. These minor VATs, essential to the long term survival of the parasite, might only be produced sufficiently frequently in the presence of a specific mechanism for the initiation of their conversion of expression loci. This continual presentation of new VATs, as well as the constant reactivation of former VATs, is likely to contribute to the host immune response becoming overwhelmed during infection (Turner and Barry, 1989).

Future Experiments.

The work presented here immediately suggests a number of quite straightforward future experiments, many of which are currently underway. The most immediate concerns the apparent transcription initiation region for the ILTat 1.61 gene. That this area indeed represents the transcription start will need to be confirmed in independent clonal populations, as will the relative sensitivity of this transcription unit to inhibition by alpha-amanitin. Assuming confirmation, sequence analysis of the initiation region must be performed . Once obtained, this should be compared with the analogous region on another M-VAT gene locus, such as the ILTat 1.22 , where the transcription unit has also been delineated. If this revealed related motifs between the two loci, then an analysis for factors binding such sequences might be carried out. These experiments would be relatively difficult because of the limiting amounts of material available from such trypansomes.

The hypothesis that duplicative activation of bloodstream VSG genes can be mediated by a specific factor may be addressed by examining trypanosome extracts for the presence of proteins able to bind and cut 70 bp repeat sequence. Although this has been tried unsuccessfully by other groups (J.D.Barry, personal communication), this has been invariably carried out using the

rodent adapted trypanosome lines which may have far lower levels of such a molecule. This search, then, should be carried out with nuclear extracts from fly transmissible parasites and using an extensive array of 70bp repeats as a template. An extensive array would be advantageous over the short arrays cloned in this thesis because a low level specific protein may need to bind co-operatively to its template to ensure cleavage of the long arrays at expression sites, rather than the larger number of dispersed repeats associated with chromosome internal genes. A suitable control would be extracts from rodent adapted trypanosome lines. These experiments are underway.

A recently opened route to the investigation of many of features revealed in this thesis has been provided by the development of transfection technology for the kinetoplastida (Bellofatto and Cross, 1989; Laban et al., 1990; Kapler et al.,1990). Using this technology it may be possible to design suitable vectors for "trapping" recombination stimulating pathways, should they exist. For example, it may be possible to co-transfect exogenous extrachromosal elements, containing sequences which, if recombined at a 70bp repeat target on each, would be selected, or lethal. This might permit the identification of either hyper or hypo-switching mutants.

Transfection will also be invaluable in the investigation of the promoter elements for M-VAT genes. Firstly, in vitro manipulation of identified promoter regions will permit an analysis of sequences regulating M-VAT gene expression. Secondly, it may be possible to unambiguously verify the link between the proposed M-VSG promoter studied in the early bloodstream organisms used here and true metacyclics, by placing some exogenous gene under its control. Activation of the promoter might then be "tagged" in true metacyclics using in situ hybridization to identify expression of the foreign transcript. This would rely on sufficiently stable transformation to permit cyclical transmission, which might be difficult to achieve in the absense of selection in the fly. Finally, and in the shorter term, it might be possible to follow the transformation from metacyclic to bloodstream mode VSG expression by using vectors

containing promoters for both systems controlling different selectable markers. This might also be manipulated to provide a selection for trypanosomes showing prolonged use of the metacyclic system, or even developmental mutants, unable to progress to use of the bloodstream VSG repertoire.

Whilst many of the experiments described above are very speculative, they demonstrate the potential for investigations within what now appears to be a manipulable genetic system.

References.

Adachi Y., Kas E. and Laemmli U.K. (1989) Preferential, cooperative binding of DNA topoisomerase II to scaffold associated regions. EMBO J., 8, 3997-4006.

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.

Aline R.F. jnr and Stuart K. (1989) Trypanosoma brucei: Conserved sequence organization 3' to telomeric variant surface glycoprotein genes. EXP. PARASITOL., 68, 57-66.

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.

Aline R.F. jnr, Myler P.J. and Stuart K. (1989) Trypanosoma brucei: frequent loss of a telomeric variant surface glycoprotein gene. EXP. PARASITOL., 68, 8-16.

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.

Allsopp R., Hall D.H. and Jones T. (1985) Fatal attraction for the tsetse fly. NEW SCIENTIST, 108, 40-43.

Apted F.I.C. (1962) Sleeping sickness in Tanganyika, past, present and future. TRANS. R. SOC. TROP. MED. HYG., 56, 15-23.

Apted F.I.C. (1970) The epidemiology of Rhodesian sleeping sickness. In THE AFRICAN TRYPANOSOMIASES. H.W.Mulligan, Ed. George Allen and Unwin Ltd., pp645-660.

Ashcroft M.T. (1960) A comparison between a syringe-passaged and a tsetse transmitted line of a strain of Trypanosoma rhodesiense ANN. TROP. MED., 54, 44-53.

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.

Bangs J.D., Hereld D., Krakow J.L., Hart G.W. and Englund P.T. (1985) Rapid processing of the carboxy terminus of a trypanosome variant surface glycoprotein. PROC. NATL. ACAD. SCI., 82, 3207-3211.

Barry J.D. (1989) African trypanosomes: an elusive target. In NEW STRATEGIES IN PARASITOLOGY, McAdam K.P.W.J. ed, Churchill Livingston, London.

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

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.

Bellofatto V. and Cross G.A.M. (1989) Expression of a bacterial gene in a trypanosomatid protozoan. SCIENCE, 224, 1167-1169.

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 single surface antigen gene of Trypanosoma brucei. CELL, 36, 163-170.

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.

Bernards A., Kooter J.M., Michels P.A.M., Moberts R.M.P. and Borst P. (1986a) Pulsed field gradient electrophoresis of DNA digested in agarose allows the sizing of the large duplication unit of a surface antigen gene in trypanosomes. GENE, 42, 313-322.

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

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. (1986b) Rapid change in the repertoire of variable surface glycoprotein genes in trypanosomes by gene duplication and deletion. J. MOL. BIOL., 190, 1-10.

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

Bernards A., Van Harten Loosbroek N. and Borst P. (1984b) Modification of telomeric DNA in Trypanosoma brucei, a role in antigenic variation. NUCLEIC ACIDS RES., 12, 4153-4170.

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

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

Boothroyd J.C., Cross G.A.M., Hoeijmakers J.H.J. and Borst P. (1980) A variant surface glycoprotein of Trypanosoma brucei synthesized with a C-terminal hydrophobic tail absent from purified glycoprotein. NATURE, 288, 624-626.

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

Borst P. (1986) Discontinuous transcription and antigenic variation in trypanosomes. ANNU. REV. BIOCHEM., 55, 701-732.

Borst P. and Cross G.A.M. (1982) Molecular basis for trypanosome antigenic variation. CELL, 29, 291-303.

Borst P. and Greaves D.R. (1987) Programmed gene rearrangements altering gene expression. SCIENCE, 235, 658-668.

Borst P., Bernards A., Van der Ploeg L.H.T., Michels P.A.M., Liu A.Y.C., De Lange T. and Sloof P. (1984) Gene rearrangements controlling the expression of surface antigen genes in trypanosomes. In MOLECULAR BIOLOGY AND HOST PARASITE INTERACTIONS, Agabian N. and Eisen, H. eds, pp205-217.

Borst P., Van der Ploeg L.H.T., Van Hoek J.F.M., Tas J. and James J. (1982) On the DNA content and ploidy of trypanosomes. MOL. BIOCHEM. PARASITOL., 6, 13-23.

Borst P., Zomerdijk J., Ouellette M., Crozatier M. and Ten Asbroek A. (1990) Control of variant-specific surface glycoprotein expression sites in Trypanosoma brucei. In PARASITES: MOLECULAR BIOLOGY, DRUG AND VACCINE DESIGN. Agabian N. and Cerami A. eds. Alan R. Liss, New York. In Press.

Boyer H.W. and Roulland-Dussoix D. (1969) A complementation analysis of the restriction and modification of DNA in Escherichia coli. J.MOL.BIOL., 41, 459-467.

Bridgen P.J., Cross G.A.M. and Bridgen J. (1976) N-terminal amino acid sequence of a variant specific surface antigen of Trypanosoma brucei. NATURE, 263, 613-614.

Buck G.A., Jacquemot C., Baltz T. and Eisen H. (1984a) Re-expression of an inactivated variable surface glycoprotein gene in Trypanosoma equiperdum. GENE, 32, 329-336.

Buck G.A., Longacre S., Raibaud A., Hibner U., Giroud C., Baltz T., Baltz D. and Eisen H. (1984b) Stability of expression linked surface antigen genes in Trypanosoma equiperdum. NATURE, 307, 563-565.

Burton F.H., Loeb D.D., McGraw R.A., Edgell M.H., and Hutchison C.A. III (1988) A directed nucleotide-sequencing approach for single stranded vectors based on recloning intermediates of a progressive DNA synthesis reaction. GENE, 67, 159-168.

Campbell D.A., Van Bree M.P. and Boothroyd J.C. (1984) The 5' limit of transposition and upstream barren region of a trypanosome VSG gene: tandem 76 base pair repeats flanking (TAA)90. NUCLEIC ACIDS RES., 12, 2759-2774.

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.

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

Cohen J.B., Effron K., Rechavi G., Ben-Neriah Y., Zakut R. and Givol D. (1982) Simple DNA sequences in homologous flanking regions near immunoglobulin VH genes: a role in gene interaction? NUCLEIC ACIDS RES., 10, 3353-3370.

Coquelet H., Tebabi P., Pays A., Steinert M. and Pays E. (1989) Trypanosoma brucei: enrichment by UV of intergenic transcripts from the variable surface glycoprotein gene expression site. MOL. CELL. BIOL., 9, 4022-4025.

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. NUCLEIC ACIDS RES., 13, 4661-4676.

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.

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

Cross G.A.M. and Johnson J.G. (1976) Structure and organization of the variant-specific surface antigens of Trypanosoma brucei. In BIOCHEMISTRY OF PARASITES AND HOST PARASITE RELATIONSHIPS, Van den Bossche, H. ed. Elsevier, Amsterdam, pp413-420.

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

Crozatier M., Jan de Brij R., Den Engelse L., Johnson P.J. and Borst P. (1988) Nucleoside analysis of DNA from Trypanosoma brucei and Trypanosoma equiperdum. MOL. BIOCHEM. PARASITOL., 31, 127-132.

Crozatier M., Van der Ploeg L.H.T., Johnson P.J., Gommers-Ampt J., Borst P. Structure of a telomeric expression site for variant specific surface antigens in T.brucei. Submitted.

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 Trypanosoma brucei. MOL. BIOCHEM. PARASITOL., 21, 189-197.

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

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.

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.

De Lange T., Kooter J.M., Michels P.A.M. and Borst P. (1983) Telomere conversion in trypanosomes. NUCLEIC ACIDS RES., 11, 8149 - 8165.

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.

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.

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

Eissenberg J.C., Cartwright I.L., Thomas G.H. and Elgin S.C.R. (1985) Selected topics in chromatin structure. ANNU. REV. GENET., 19, 485-536.

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

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

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.

Feinberg A.P. and Vogelstein B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. ANAL. BIOCHEM., 132, 6-13.

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

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

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.

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

Florent I., Baltz T., Raibaud A. and Eisen H. (1987) On the role of repeated sequences 5' to variant surface glycoprotein genes in African trypanosomes. GENE, 53, 55-62.

Fox J.A., Duszenko M., Ferguson M.A.J., Low M.G. and Cross G.A.M. (1986) Purification and characterisation of a novel glycanphosphatidylinositol specific phospholipase C from Trypanosoma brucei. J. BIOL. CHEM., 261, 15767-15771.

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

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

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.

Goodwin L.G. (1985) Trypanosomes: An Introduction. BRIT. MED. BULL., 41, 103-104.

Gray A.R. (1965) Antigenic variation in a strain of Trypanosoma brucei transmitted by Glossina morsitans. J. GEN. MICROBIOL., 41, 195-214.

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.

Greider C.W. and Blackburn E.H. (1989) A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. NATURE, 337, 331-337.

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

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. PARASITOL., 83, 609-621.

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

Hanahan D. (1983) Studies on the transformation of Escherichia coli by plasmids. J. MOL. BIOL., 166, 557-580.

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..

Hasan G., Turner M.J. and Cordingley J.S. (1984) Complete nucleotide sequence of an unusual mobile element from Trypanosoma brucei. CELL, 37, 333-341.

Henderson E., Hardin C.C., Walk S.K., Tinoco I. and Blackburn E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs CELL, 51, 899-908.

Henikoff S. and Cohen E.H. (1984) Sequences responsible for transcription termination on a gene segment in Saccharomyces cerevisiae. MOL. CELL. BIOL., 4, 1515-1520.

Herbert W.J. and Lumsden W.H.R. (1976) Trypanosoma brucei: a rapid "matching" method for estimating the host's parasitaemia EXPTL. PARASITOL., 40, 427-431.

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

Herskowitz I. (1989) A regulatory hierarchy for cell specialization in yeast. NATURE, 342, 749-757.

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

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

Holder A.A. and Cross G.A.M. (1981) Glycopeptides from variant surface glycoproteins of Trypanosoma brucei. C-terminal location of antigenically cross reacting moieties. MOL. BIOCHEM. PARASITOL., 2, 135-150.

Holliday R. (1989) Untwisting B-Z DNA. TRENDS GENET., 5, 355-356.

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

Johnson J.G. and Cross G.A.M. (1977) Carbohydrate composition of the variant specific surface antigen glycoproteins from Trypanosoma brucei. J. PROTOZOOL., 24, 587-591.

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.

Kapler G.M., Coburn C.M., and Beverley S.M. (1990) Stable transfection of the human parasite Leishmania major delineates a 30-Kb region sufficient for extrachromosomal replication and expression. MOL. CELL. BIOL., 10, 1084-1094.

Kedes L.H. (1979) Histone genes and histone messengers. ANNU. REV. BIOCHEM., 48, 837-870.

Keil R.L. and Roeder G.S. (1984) Cis-acting recombination-stimulating activity in a fragment of the ribosomal DNA of S.cerevisiae. CELL, 39, 377-386.

Konig E., Delius H., Carrington M., Williams R.O. and Roditi I. (1989) Duplication and transcription of procyclin genes in Trypanosoma brucei. NUCLEIC ACIDS RES., 17, 8727-8739.

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. NUCLEIC ACIDS RES., 12, 9457-9472..

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

Kooter J.M., van der Spek H.J., Wagter R., d'Oliveira C.E., van der Hoeven F., Johnson P.J. and Borst P. (1987) The anatomy and transcription of a telomeric expression site for variant specific surface antigens in T.brucei. CELL, 51, 261-272.

Kooter J.M., Winter A.J., d'Oliveira C., Wagter R. and Borst P. (1988) 5' boundaries of telomere conversions in Trypanosoma brucei. GENE, 69, 1-11.

Kostricken R., Strathern J.N., Klar A.J.S., Hicks J.B. and Heffron F. (1983) A site-specific endonuclease essential for mating type switching in Saccharomyces cerevisiae. CELL, 35, 167-174.

Laban A., Tobin J.F., de Lafaille M.A.C. and Wirth D.F. (1990) Stable expression of the bacterial neo^{r} gene in Leishmania enriettii. NATURE., 343, 572-574.

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

Lamont G.S., Tucker R.S. and Cross G.A.M. (1986) Analysis of antigen switching rates in Trypanosoma brucei. PARASITOLOGY, 92, 355-367.

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.

Laurent M., Pays E., Delinte K., Magnus E., Van Meirvenne N., and Steinert M. (1984a) Evolution of a surface antigen gene repertoire linked to non duplicative gene activation. NATURE, 308, 370-373.

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.

Laurent M., Pays E., Van der Werf A., Aerts D., Magnus E., Van Meirvenne N. and Steinert M. (1984b) Translocation alters the activation rate of a trypanosome surface antigen gene. NUCLEIC ACIDS RES., 12, 8319-8328.

Layden R.E. and Eisen H. (1988) Alternate trans splicing in Trypanosoma equiperdum: implications for splice site selection. MOL. CELL. BIOL., 8, 1352-1360.

Lee M.G.S. and Van der Ploeg L.H.T. (1987) Frequent independent duplicative transpositions activate a single VSG gene. MOL. CELL. BIOL., 7, 357-364.

LeMeur M.A., Galliot B. and Gerlinger P. (1984) Termination of the ovalbulmin gene transcription. EMBO J., 3, 2779-2786.

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.

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 types. PROC. NATL. ACAD. SCI., 81, 6642-6646.

Le Page R.W.F., Wells J.M., Prospero T.D. and Sternberg J. (1988) Genetic analysis of hybrid Trypanosoma brucei. In CURRENT COMMUNICATIONS IN MOLECULAR BIOLOGY: MOLECULAR GENETICS OF PARASITIC PROTOZOA, Turner, M.J. and Arnot, D. eds. Cold Spring Harbor, pp 65-71.

Le Ray D., Barry J.D. and Vickerman K. (1978) Antigenic heterogeneity of metacyclic forms of Trypanosoma brucei NATURE, 273, 300-302.

Le Ray D., Barry J.D., Easton C. and Vickerman K. (1977) First tsetse fly transmission of the "AnTat" serodeme of Trypanosoma brucei ANN. SOC. BELGE MED. TROP., 57, 369-381.

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.

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.

Majiwa P.A.O., Young J.R., Englund P.T., Shapiro P.Z. and Williams R.O. (1982) Two distinct forms of surface antigen gene rearrangements in Trypanosoma brucei. NATURE, 297, 514-516.

Majumder H.K., Boothroyd J.C. and Weber H. (1981) Homologous 3'-terminal regions of mRNAs for surface antigens of different antigenic variants of Trypanosoma brucei. NUCLEIC ACIDS RES., 9, 4745-4753.

Maniatis T., Fritch E.F. and Sambrook J. (1982) MOLECULAR CLONING: A LABORATORY MANUAL. Cold Spring Harbor Laboratory. U.S.A.

Masterson W.J., Doering T.L., Hart G.W. and Englund P.T. (1989) A novel pathway for glycan assembly: biosynthesis of the glycosyl-phosphatidylinositol anchor of the trypanosome variant surface glycoprotein. CELL, 56, 793-800.

Matthews K.R., Shiels P., Graham S.V. and Barry J.D. (1990) Structure and expression of trypanosome metacyclic VSG expression sites. In PARASITES; MOLECULAR BIOLOGY, DRUG AND VACCINE DESIGN. Agabian, N. and Cerami, A. eds. Alan R. Liss, New York. In Press.

Maudlin I. and Dukes P. (1985) Extrachromosomal inheritance of susceptibility to trypanosome infection in tsetse flies I. Selection of susceptible and refractory lines of Glossina morsitans morsitans. ANN. TROP. MED. PARASITOL., 79, 317-324.

Maudlin I. and Welburn S.C. (1988) Tsetse immunity and the transmission of trypanosomiasis. PARASITOL. TODAY, 4, 109-111.

McConnell J., Gurnett A.M., Cordingly J.S., Walker J.S. and Turner M.J. (1981) Biosynthesis of Trypanosoma brucei variant surface glycoprotein. I. Synthesis, size and processing of an N-terminal signal peptide MOL. BIOCHEM. PARASITOL., 4, 225-242.

McGill C., Shafer B. and Strathern J. (1989) Coconversion of flanking sequences with homothallic switching. CELL, 57, 459-467.

Meselson M.S. and Radding C.M. (1975) A general model for genetic recombination. PROC. NATL. ACAD. SCI., 72, 358-361.

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

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

Michels P.A.M. (1986) Evolutionary aspects of trypanosomes: analysis of genes. J. MOL. EVOL., 24, 45-52.

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.

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.

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

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.

Miller H. (1987) Practical aspects of preparing phage and plasmid DNA: growth, maintenance, and storage of bacteria and bacteriophage. METHODS IN ENZYMOLOGY, 152, 145-170.

Mount S.M. (1982) A catalogue of splice junction sequences. NUCLEIC ACIDS RES., 10, 459-427.

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.

Myler P.J., Aline R.F.jnr , Scholler J.K. and Stuart K. (1988) Multiple events associated with antigenic switching in Trypanosoma brucei. MOL. BIOCHEM. PARASITOL., 29, 227-241.

Myler P.J., Allen A.L., Agabian N. and Stuart K. (1985) Antigenic variation in clones of Trypanosoma brucei grown in immune-

deficient mice. INFECT. IMM., 47, 684-690.

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.

Nickol J., Behe J. and Felsenfeld G. (1982) Effect of the B-Z transition in poly (dG-m5dC).poly (dG-m5dC) on nucleosomal formation. PROC. NATL. ACAD. SCI., 79, 1771-1775.

Olafson R.W., Clarke M.W., Keilland S.O., Barbet A.F.,and McGuire T.D. (1984) Amino terminal sequence homologies amongst variant surface glycoproteins of the African trypanosomes. MOL. BIOCHEM. PARASITOL., 2, 287.

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

Overath P., Czichos J., Stock U. and Nonnengaesser C. (1983) Repression of glycoprotein synthesis and release of surface coat during transformation of Trypanosoma brucei. EMBO J., 2, 1721-1728.

Paindavoine P., Zampetti-Bosseler F., Pays E., Schweizer J., Guyaux M., Jenni L. and Steinert M. (1986) Trypanosome hybrids generated in tsetse flies by nuclear fusion. EMBO J., 5, 3631-3636.

Patzelt E., Perry K.L. and Agabian N. (1989) Mapping of branch sites in trans spliced pre-mRNAs of Trypanosoma brucei. MOL. CELL. BIOL., 9, 4291-4297.

Pays E. (1985) Gene conversion in trypanosome antigenic variation. PROG. NUCLEIC ACID RES. MOL. BIOL., 32, 1-26.

Pays E. (1989) Pseudogenes, chimaeric genes and the timing of antigen variation in African trypanosomes. TRENDS GENET., 5, 389-391.

Pays E. and Steinert M. (1988) Control of antigen gene expression in African trypanosomes. ANNU. REV. GENET., 22, 107-126.

Pays E., Coquelet H., Pays A., Tebabi P. and Steinert M. (1989b) Post transcriptional control of a variant surface glycoprotein gene expression site. MOL. CELL. BIOL., 9, 4018-4021.

Pays E., Dekerck P., Van Assel S., Babiker E.A., Le Ray D., Van Meirvenne N. and Steinert M. (1983e) Comparative analysis of a Trypanosoma brucei gambiense antigen gene family and its potential use in epidemiology of sleeping sickness MOL. BIOCHEM. PARASITOL., 7, 63-74.

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. NUCLEIC ACIDS RES., 12, 5235-5249.

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

Pays E., Guyaux M., Aerts D., Van Meirvenne N. and Steinert M. (1985a) Telomere reciprocal recombination as a mechanism for antigenic variation in trypanosomes. NATURE, 316, 562 - 564.

Pays E., Houard S., Pays A., Van Assel S., Dupont F., Aerts D., Huet-Duvillier G., Gomes V., Richet C., Degand P., Van Meirvenne N. and Steinert M. (1985b) Trypanosoma brucei: the extent of conversion in antigen genes may be related to the DNA coding specificity. CELL, 42, 821-829.

Pays E., Laurent M., Delinte K., Van Meirvenne N. and Steinert M. (1983c) Differential size variations between transcriptionally active and inactive telomeres of Trypanosoma brucei. NUCLEIC ACIDS. RES., 11, 8137-8147.

Pays E., Lheureux M. and Steinert M. (1981) The expression linked copy of a surface antigen gene in Trypanosoma brucei is probably the one transcribed. NATURE, 292, 265-267.

Pays E., Lheureux M. and Steinert M. (1982) Structure and expression of a Trypanosoma brucei gambiense variant specific antigen gene. NUCLEIC ACIDS RES., 10, 3149-3163.

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.

Pays E., Van Asssel S., Laurent M., Darville M., Vervoort T., Van Meirvenne N. and Steinert M. (1983b) Gene conversion as a mechanism for antigenic variation in trypanosomes. CELL, 34, 371-381.

Pays E., Van Assel S., Laurent M., Dero B., Michiels F., Kronenberger, Matthyssens G., Van Meirvenne N., Le Ray D. and Steinert, M. (1983a) At least two transposed sequences are associated in the expression site of a surface antigen gene in different trypanosome clones. CELL, 34, 359-369.

Penncavage N.A., Julius M.A. and Marcu K.B. (1983) An African trypanosome variant surface glycoprotein gene whose expression is not activated by duplication. NUCLEIC ACIDS RES., 11, 8343-8357.

Pluta F. and Zakian V.A. (1989) Recombination occurs during telomere formation in yeast. NATURE, 337, 429-433.

Raleigh E.A. (1987) Restriction and modification in vivo by Escherichia coli K12. METHODS ENZM., 152, 130-141.

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.

Rifkin M.R. (1978) Identification of the trypanocidal factor in normal human serum: high density lipoprotein. PROC. NATL. ACAD. SCI., 75, 3450-3454.

Rovis R. and Dube L. (1982) Identification and characterisation of the two N-acetylglucosaminyl-transferases associated with Trypanosoma brucei microsomes. MOL. BIOCHEM. PARASITOL., 5, 173-187.

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

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

Sauerbier W. and Herculez K. (1978) Gene and transcription unit mapping by radiation effects. ANNU. REV. GENET., 12, 329-363.

Scholler J.K., Myler P.J. and Stuart K.D. (1989) A novel telomeric gene conversion in Trypanosoma brucei. MOL. BIOCHEM. PARASITOL., 35, 11-20.

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

Seed J.R., Edwards R. and Sechelski J. (1984) The ecology of antigenic variation. J. PROTOZOOL., 31, 48-53.

Sen D. and Gilbert W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. NATURE, 334, 364-366.

Shah J.S., Young J.R., Kimmel B.E., Jams K.P. and Williams R.O. (1987) The 5' flanking sequence of a Trypanosoma brucei variable surface glycoprotein gene. MOL. BIOCHEM. PARASITOL., 24, 163-174.

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.

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.

Shiels P.G. (1990) Analysis of metacyclic telomeres in Trypanosoma brucei PhD THESIS, University of Glasgow.

Shiels P., Graham S.V., Matthews K.R. and Barry J.D. Distinct, stage specific, activation mechanisms of a trypanosome VSG gene in a novel telomeric locus. Submitted.

Shimizu A. and Honjo T. (1984) Immunoglobulin class switching. CELL, 36, 801-803.

Short J.M., Fernandez J.M., Sorge J.A. and Huse W.D. (1988) Lambda Zap: a bacteriophage lambda expression vector with in vivo excision properties. NUCLEIC ACIDS RES., 16, 7583-7600.

Sloof P., Menke H.H., Caspers P.M. and Borst P. (1983) Size fractionation of Trypanosoma brucei DNA: localization of the 177bp repeat satellite DNA and a variant surface glycoprotein gene in a mini-chromosomal DNA fraction. NUCLEIC ACIDS RES., 11, 3889-3901.

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.

Sorge J.A. (1988) Bacteriophage lambda cloning vectors. in VECTORS: A SURVEY OF MOLECULAR CLONING VECTORS AND THEIR USES, Rodriguez, R.L. and Denhardt, D.T. eds. Butterworths, Stoneham, MA, U.S.A., p43-60.

Spitzner J.R., Chung I.K. and Muller M.T. (1990) Eukaryotic topoisomerase II preferentially cleaves alternating purine-pyrimidine repeats. NUCLEIC ACIDS RES., 18, 1-11.

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.

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.

Strickler J.E. and Patton C.L. (1980) Trypanosoma brucei brucei : Inhibition of glycosylation of the major variable surface glycoprotein by tunicamycin. PROC. NATL. ACAD. SCI., 77, 1529-1533.

Strickler J.E. and Patton C.L. (1982a) Trypanosoma brucei: effect of inhibition of N-linked glycosylation on the nearest neighbor analysis of the major variable surface glycoprotein. MOL. BIOCHEM. PARASITOL., 5, 117-131.

Strickler J.E. and Patton C.L. (1982b) Trypanosoma brucei: nearest-neighbor analyses on the major variable surface glycoprotein- cross linking patterns with intact cells. EXPTL. PARASITOL., 53, 117-132.

Sutton R.E. and Boothroyd J.C. (1986) Evidence for trans splicing in trypanosomes. CELL, 47, 527-535.

Szostak J.W., Orr-Weaver T.L., Rothstein R.J. and Stahl F.W. (1983) The double strand-break-repair model for recombination. CELL, 33, 25-35.

Tabor S. and Richardson C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. PROC. NATL. ACAD. SCI., 84, 4767-4771.

Tait A. (1980) Evidence for diploidy and mating in trypanosomes. NATURE, 287, 536-538..

Tait A., Turner C.M.R., Le Page R.W.F. and Wells J.M. (1989) Genetic evidence that metacyclic forms of Trypanosoma brucei are diploid. MOL. BIOCHEM. PARASITOL., 37, 247-256.

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 fly studied using immunoelectron microscopy. J. CELL. SCI., 87, 363-372.

Thon G., Baltz T. and Eisen H. (1989) Antigenic diversity by the recombination of pseudogenes. GENES DEV., 3, 1247-1254.

Timmers H., Th M., De Lange T., Kooter J.M. and Borst P. (1987) Coincident multiple activations of the same surface antigen gene in Trypanosoma brucei. MOL. CELL. BIOL., 194, 81-90.

Tsao Y.P., Wu , H-Y and Liu L.F. (1989) Transcription-driven supercoiling of DNA: direct biochemical evidence from in vitro studies. CELL, 56, 111-118.

Turner C.M.R. and Barry J.D. (1989) High frequency of antigenic variation in Trypanosoma brucei rhodesiense infections. PARASITOL., 99, 67-75.

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.

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.

Turner M.J. (1988) Trypanosome variant surface glycoprotein. in THE BIOLOGY OF PARASITISM., Englund, P.T and Sher, A. eds., Alan R. Liss, New York., p349-369..

Van der Ploeg L.H.T., Valerio D., De Lange T., Bernards A., Borst P. and Grosveld F.G. (1982) Analysis of cosmid clones of nuclear DNA from Trypanosoma brucei shows that the genes for variant surface glycoproteins are clustered in the genome. NUCLEIC ACIDS RES., 10, 5905-5923.

Van der Ploeg L.H.T. and Cornelissen A.W.A.C. (1984) The contribution of gene rearrangement to antigenic variation in Trypanosoma brucei. PHIL. TRANS. ROY. SOC. (LOND.) B, 307, 13-26.

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.

Van der Ploeg L.H.T., Liu A.Y.C. and Borst P. (1984c) The structure of the growing telomeres of trypanosomes. CELL, 36, 459-468.

Van der Ploeg L.H.T., 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.

Van der Ploeg L.H.T., Smith C.L., Polvere R.I. and Gottesdiener K.M. (1989) Improved separation of chromosome-sized DNA from Trypanosoma brucei, stock 427-60. NUCLEIC ACIDS RES., 17, 3217-3227.

Vickerman K. (1969) On the surface coat and flagellar adhesions in trypanosomes. J. CELL SCI., 5, 169-193.

Vickerman K. (1985) Developmental cycles and biology of pathogenic trypanosomes. BR. MED. BULL., 41, 105-114..

Vickerman K. and Luckins A.G. (1969) Localization of variable antigens on the surface of Trypanosoma brucei using ferritin conjugated antibody. NATURE, 224, 1125-1126.

Voelkel-Meiman K., Keil R.L. and Roeder G.S. (1987) Recombination-stimulating sequences in yeast rDNA correspond to sequences regulating transcription by RNA polymerase I. CELL, 48, 1071-1079.

White C.I. and Haber J.E. (1990) Intermediates of recombination during mating type switching in Saccharomyces cerevisiae. EMBO J., 9, 663-673.

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. NUCLEIC ACIDS RES., 16, 6725-6736.

Wijers D.J. and Willet K.C. (1960) Factors that may influence the infection rate of Glossina palpalis with Trypanosoma gambiense: the number and morphology of the trypanosomes present in the blood of the host at the time of the infective feed. ANN. TROP. MED., 54, 341-349.

Wilbur W.J. and Lipman D.J. (1983) Rapid similarity searches of nucleic acid and protein data bases. PROC. NATL. ACAD. SCI., 80, 726-730.

Williamson J.R., Raghuraman M.K. and Cech T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. CELL, 59, 871-880.

Williams R.O., Young J. and Majiwa P. (1982) Genomic environment of Trypanosoma brucei VSG genes: presence of a minichromosome. NATURE, 299, 417-421.

Yanisch-Perron C., Vieira J. and Messing J. (1985) Improved M13 cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. GENE, 33, 103-119.

Zampetti-Bosseler F., Schweiger J., Pays E., Jenni L. and Steinert M. (1986) Evidence of haploidy in metacyclic forms of Trypanosoma brucei. PROC. NATL. ACAD. SCI., 83, 6063-6064.

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., submitted.

pUC 19

```
                                        HincII
                                        AccI
                                        SalI
    EcoRI       KpnI      BamHI         ------        SphI
    ------      ------    ------                      ------
....GAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTT.....
          ------    ------      ------      ------    ------
           SacI      SmaI        XbaI        PstI      HindIII
                     XbaI
```

pMTL 23

```
          EcoRV     NdeI    HincII    PstI  HindIII  EcoRI  SstI    KpnI     SphI   ClaI  BglII  StuI
                                                ------         ------
....GCCGGCGATATCGGATCCATATGACGTCGACGCGTCTGCAGAAGCTTCTAGAATTCGAGCTCCCGGGTACCATGGCATGCATCGATAGATCTCGAGGCCTCGCGA.....
    ------    ------  ------ ------             ------          ------ ------                 ------   ------
     NaeI      BamHI    AatI   MluI              XbaI            SmaI   NcoI                  XhoI     NruI
```

pBluescript

```
                                                                                  HincII
                                                                                  AccI              ApaI
    SacI    SacII    EagI    XbaI   SpeI  BamHI SmaI  PstI  EcoRI EcoRV HindIII ClaI   SalI  XhoI     DraII KpnI
    ------  ------   ------  ------ ------ ------ ------ ------ ------ ------ ------   ------ ------   ------ ------
....GAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACC...
```

Appendix I: Polylinker restriction sites for the plasmid vectors pUC19, pMTL23 and pBluescript. Only enzymes with a 6bp recognition sequence are shown.