Chromosomal environment of a trypanosome metacyclic VSG gene expression site

Kerri Kobryn
Wellcome Unit of Molecular Parasitology
The Anderson College
University of Glasgow

Submitted for the degree of Doctor of Philosophy

May, 1996.

Acknowledgements

Special thanks to J.D. Barry and to my funding body: the M and D foundation.

Contents

Abstract.....2

Chapter 1

Introduction.....3
1.1 A general introduction to *Trypanosoma brucei*.....4
1.2 Introduction to the VSG coat.....6
1.2.1 Focus on the Variant Surface Glycoprotein.....6
1.2.2 Focus on the VSG coat.....7
1.3 Genetic mechanisms of antigenic variation.....8
1.4 Antigen switching and the course of a bloodstream infection.....10
1.4.1 Order of VAT presentation.....10
1.4.2 Switching rates.....11
1.4.3 Recombination.....12
1.5 The metacyclic VSG repertoire.....13
1.6 Gene expression.....14
1.6.1 Gene organisation.....14
1.6.2 RNA processing.....14
1.6.3 RNA polymerases.....14
1.7 The control of gene expression in bloodstream VSG expression sites...16
1.7.1 The bloodstream expression site.....16
1.7.2 Life cycle stage-specific VSG expression.....19
1.7.3 Silencing and *in situ* activation.....20
1.8 Metacyclic VSG expression site control.....22

Chapter 2

Methods and Materials.....24
2.1 Abbreviations.....25
2.2 Strains used.....26
2.3 Microbiological media.....27
2.3.1 Media for growth of yeast.....27
2.3.2 *E. coli* growth media.....28
2.4 Electrophoresis.....28
2.5 Standard protocols.....29
2.5.1 *T. brucei* genomic DNA preparation.....29
2.5.2 Yeast genomic DNA preparation.....29
2.5.3 Rapid protocol for the preparation of chromosome-length yeast DNA.....29
2.5.4 Preparation of high grade chromosome-length DNA.....30
2.5.5 Plasmid minipreps.....30
2.5.6 Preparation of inserts for subcloning.....31
2.5.7 Labelling and purification of DNA probes.....31
2.5.8 Preparation of λ DNA.....32

2.5.9 Transformation of *E. coli*.....32
2.5.10 Nucleic acid transfer.....33
2.5.11 Hybridisation.....33
2.6 Specialised protocols.....35
2.6.1 Spheroplasting *S. cerevisiae*.....35
2.6.2 Transformation of *S. cerevisiae* for YAC library construction.....36
2.6.3 YAC library amplification.....36
2.6.4 Storage of primary YAC library transformants.....36
2.6.5 Colony hybridisation screening of YAC libraries.....37
2.6.6 Transcriptional run-on analysis.....37
2.6.7 Slot blotting.....38
2.6.8 Gelase treatment.....39
2.7 Referenced kit protocols.....40
2.7.1 Qiagen plasmid preparations.....40
2.7.2 Gelase (Cambio).....40
2.7.3 Riboprobe synthesis kit (Stratagene).....40
2.7.4 UBI Sequenase kit.....40

Chapter 3

Cloning of *T. brucei rhodesiense* sequences utilising Yeast Artificial Chromosomes.....41
3.1 Introduction: Theorectical considerations.....42
3.2 Introduction: Practical considerations.....46
3.3 A Rationale.....47
3.4 Determination of the size of *T. brucei* genomic DNA to be used in YAC cloning protocols.....49
3.5 Nucleic Acid Chemistry for the construction of a Telomere Complementation library using the pYAC4-*Neo-Not* I vector system.....50
3.6 Construction of a *T. brucei rhodesiense* YAC library utilising the pYAC4 vector system.....55
3.7 Gridding, Amplifying and Screening.....57
3.8 Karyotype Analysis of the Y1.22 'clones'.....58

Chapter 4

Analysing a YAC clone.....62
4.1 Introduction: Cloning and contig assembly from YACs.....63
4.2 Karyotype Analysis of the ILTat 1.22 M-VSG basic copy gene.....64
4.3 Plasmid end-rescue: an attempt to confirm the karyotypic location of the clones Ya and Y1.....65
4.4 *In situ* YAC mapping.....67
4.5 Cloning the YACs into phage λ.....69
4.6 Screening the Ya genomic library for YAC-specific clones.....70
4.7 Transcriptional analysis of LλYa and Yλ122.....71
4.8 Some mapping of the λ clones.....73

4.9 Probing the problem....76
4.10 Karyotype Analysis of the Y1.22 probe....80
4.11 Recloning the 1.22 locus....81
4.12 Restriction mapping λMT1.22-2....83
4.13 Identifying the transcription unit in Yλ122....85
4.14 Concluding remarks....87

Chapter 5
Characterising the YAC library....88
5.1 Introduction....89
5.2 Examining Ya and Y1....90
5.3 Examination of 20 clones from the YAC library....93

Chapter 6
Defining the ILTat 1.22 metacyclic domain....97
6.1 Introduction....98
6.2 Cloning more of the 1.22 locus....98
6.3 Restriction mapping λMT1.22-3....100
6.4 Subcloning novel 1.22 sequences/and verification of the clone inserts....102
6.5 Southern analysis of the new 1.22 sequences....105
6.6 The 1.22 metacyclic domain in STIB 386....113
6.7 Transcriptional analysis of the novel 1.22 sequences derived from λMT1.22-3....118
6.7.1 Gene hunting in pE2.4....124
6.8 Sequencing pE2.4....125
6.9 Are M-VSG expression sites remodelled B-VSG expression sites?....125
6.10 Northern analysis of pE2.4....127
6.11 Postscript....133
6.12 **Perspectives**....134
Summary....134
Thoughts about the metacyclic domain....134
Models of gene expression....135
Futures....136
Appendix....138

References....139

Figures List

1.1 *T. brucei* life cycle....4
1.2 VSG switching mechanisms....9
1.3 Structure of bloodstream VSG expression sites....16
1.4 The ESAGs....18

3.1 pYAC4....44
3.2 pYAC4-*Neo*-*Not* I....46
3.3 Sizing *T. brucei* genomic DNA....49
3.4 *Sau* 3AI partial digests....51
3.5 Examing the libraries....54
3.6 *Eco* RI partial digests....55
3.7 The 1.22BC gene locus....57
3.8 The Y1.22 clones....58
3.9 Clonality of the Y1.22 clones....60

4.1 1.22BC karyotype assignment....65
4.2 Plasmid end-rescue....66
4.3 *In situ* YAC mapping....68
4.4 Run-on analysis of LλYa and Yλ122....72
4.5 Mapping data....74
4.6 Restriction maps....75
4.7 Southern/karyotype analysis of LλYa....76
4.8 Southern/karyotype analysis of Yλ122....77
4.9 Southern/karyotype analysis of RλYa....78
4.10 Karyotype analysis of the probe Y1.22....80
4.11 Origin of the library screening probes....82
4.12 Mapping data....84
4.13 Restriction maps of λ1.22B and λMT1.22-2....85
4.14 The Yλ122 transcription unit....86

5.1 Detection of vector/insert junction abnormalties....90
5.2 Analysis of Ya and Y1....91-92
5.3 Southern/karyotype analysis of 20 clones....93-94

6.1 Origin of the library screening probe....99
6.2 Overlap gel....99
6.3 λMT1.22-3 mapping data....101
6.4 The updated 1.22BC locus....102
6.5 Assessing the fidelity of the 1.22 clonings....103-104
6.6 Relationship between isolates....106
6.7 Southern analysis of pHB....107

6.8 Southern analysis of pH1.2.....108
6.9 Southern analysis of pEH2.1.....109
6.10 Southern analysis of pE2.4.....110
6.11 More southern analysis of pE2.4.....111
6.12 Metacyclic domain (R5) southern analysis.....114
6.13 Southern analysis for the 1.22BC gene.....115
6.14 Proposed evolution of the 1.22 locus.....117
6.15 'X sequence' run-ons.....119-120
6.16 'X sequence' slot blot run-ons.....122-123
6.17 70 bp repeat probing of the 'X sequences'.....126
6.18 ESAG 1K northerns.....128
6.19 A model of the 'X sequences'.....131

ABSTRACT

The gene encoding the metacyclic Variant Surface Glycoprotein (VSG) ILTat 1.22 is expressed *in situ* from a monocistronic, telomeric transcription unit in metacyclic-derived trypanosomes (Graham and Barry *Mol. Cell. Biol.* 15(11):5945-56, 1995). The genomic environment upstream of the 1.22 basic copy gene (1.22BC) is composed of single copy, transcriptionally silent sequence. This sequence occurs in an area which, in bloodstream VSG gene expression sites, is thought to be subject to the influence of a developmentally regulated position silencing effect (Horn and Cross *Cell* 83:555-61, 1995; Rudenko *et al. Cell* 83:547-53, 1995). For this reason, the silent area upstream of the 1.22BC is designated as a 'metacyclic domain'. The metacyclic domain is defined as any single copy, silent sequence linked with the M-VSG gene.

In a step towards understanding the nature of the control of M-VSG gene expression, efforts to define the extent of the metacyclic domain for the ILTat 1.22 M-VSG gene were undertaken. Initially, YAC cloning technology was employed to accomplish the required cloning. This proved only partially successful and ultimately unnecessary. λ cloning, genomic southern analysis and transcriptional studies were employed to define the extent of the 1.22 metacyclic domain. The metacyclic domain was found to end 21 kb upstream of the 1.22BC gene. Sequence upstream of the metacyclic domain appears to be diploid and transcriptionally active, especially in metacyclic-derived trypanosomes. A gene candidate in this region was partially sequenced and found to have sequences highly homologous to a gene,ESAG 1, that occurs in bloodstream VSG gene expression sites.

Chapter 1

Introduction

1.1 A general introduction to *Trypanosoma brucei*

Trypanosoma brucei is a flagellated protozoan parasite. Three morphologically indistinguishable variants are identified by their differences in host range and associated pathologies. *T. brucei brucei* infects game animals and cattle causing Nagana (in cattle) while its host range variant *T. b. rhodesiense* is able to infect humans and cause disease. *T. b. rhodesiense* causes an acute form of the disease, Sleeping Sickness, in East Africa, while *T. b. gambiense,* a second subspecies, causes a chronic form of sleeping sickness. In all cases the transmission of the parasite is cyclical, alternating between the trypanosome's tsetse fly and mammalian hosts. *T. brucei* is found in the bloodstream and connective tissue fluid of the mammalian host. It multiplies in the bloodstream of the mammalian host and is transmitted by the bite of the bloodsucking tsetse fly.

The essential features of the life cycle of *T. brucei* are depicted in **Fig. 1.1**, after Vickerman (1985).

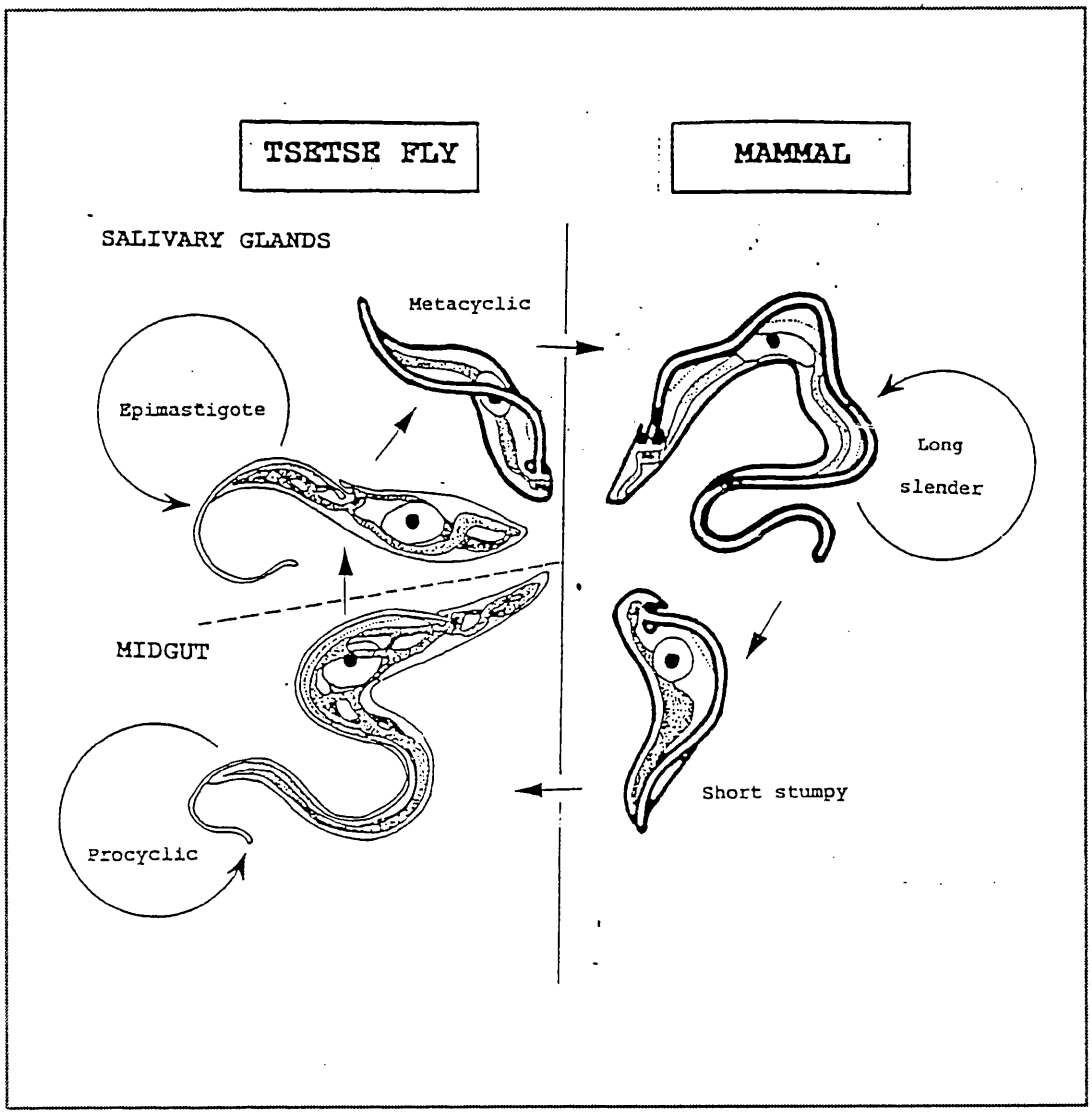


Figure 1.1 *T. brucei* life cycle modified by J.D. Barry from (Vickerman 1985)

As can be seen from **Fig. 1.1**, during the course of its life cycle *T. brucei* differentiates into many different forms. This cyclical progression of developmental forms is accompanied by complex changes in morphology, cellular structure, metabolism and surface coat composition. Two of these forms are crucial to the cyclical transmission of the parasite: the stumpy and metacyclic forms. The transmission from the mammal to the tsetse fly vector is accomplished by the nondividing 'stumpy' form of the parasite. Only the stumpy form allows for the continuation of the life cycle (Vickerman 1985). After they are ingested by the fly their mitochondrion becomes fully activated, an event which heralds a rapid switch in the metabolic pathways used by the parasite. Bloodstream trypanosomes use glucose as their sole carbon source and metabolise it glycolytically. Trypanosomes in the fly, on the other hand, use amino acids, notably proline, as their energy source in an oxidative metabolic cycle. Transmission variants exist that have been generated by sequential syringe-mediated passage through laboratory animals. These 'monomorphic' lines have lost the ability to be cyclically transmitted. This correlates with the inability to form stumpy trypanosomes and to activate the mitochondrion (Hajduk and Vickerman 1981a). The transition from fly back to the mammalian host is effected by the non-proliferative metacyclic stage of the parasite. The metacyclic form seems preadapted for survival in the mammalian host.

One of the changes that occurs throughout the life cycle that particularly interests us is the change in the trypanosome's surface coat. For its entire life cycle, *T. brucei* is coated by glycolipid anchored glycoproteins (Cross 1975; Roditi and Pearson 1990). Dense coats of Variant Surface Glycoprotein (VSG) and procyclic acidic repetitive protein (parp) envelop the parasite protecting it from serum components (and tsetse molecules?), in the relevant life cycle stages.

The VSG coat of the ingested bloodstream trypanosomes is released as parp molecules are incorporated into a transiently mixed coat. It has been proposed that the peculiar acidic repeat (Glu-Pro) of parp makes it a long rodlike protein that interdigitates between VSG molecules (Ziegelbauer and Overath 1990). The parp N-terminal epitopes are accessible to antibody, indicating that parp's tertiary structure is more elongated than that of the VSG. The production of a transiently mixed coat and the gradual release of VSG ensure that the trypanosome is never 'naked' (Ziegelbauer and Overath 1990; Roditi and Pearson 1990). Before the premetacyclic stage detaches from the epithelium of the tsetse fly salivary gland the parp coat is replaced by a VSG coat (Vickerman 1985). After detachment, the resultant metacyclic trypanosome is ready, upon a subsequent tsetse bloodmeal, to infect its mammalian host. The genetic mechanisms that control trypanosome surface coat switches have been a subject of keen interest to many investigators and students alike.

1.2 Introduction to the VSG coat

In the bloodstream, the trypanosome is covered by ~10^7 molecules of a single species of VSG, resulting in a thick protective coat (Cross 1975). The VSG in the surface coat represents the antigenic 'face' of the trypanosome. Antibodies to other trypanosome antigens are raised, probably from dead parasites, but they offer no protective immunity (Vickerman 1985). The antigenic face of the trypanosome infection changes as trypanosomes switch the VSG in their surface coat. Switching of the VSG coat in bloodstream trypanosomes, antigenic variation, is generally accepted as the principal means by which *T. brucei* evades the host immune system producing the chronic infection that increases its likelihood of cyclical transmission (see **1.2.2**). Study of the bloodstream trypanosome's surface coat has produced a great deal of interesting literature, any overview of which logically starts with the VSG molecule itself.

1.2.1 Focus on the Variant Surface Glycoprotein

VSG is translated as a preprotein with a variable hydrophobic leader and a well conserved tail. Protein processing yields a glycosylated, membrane-associated, mature protein lacking its C and N-terminal extensions. The leader is presumed to be a target signalling sequence. Biochemical purifications of VSG noted two distinct, purified forms: soluble VSG (sVSG) and mfVSG or membrane-form VSG (Cardosa de Almeida and Turner 1983). This finding presaged the discovery of a novel form of protein-membrane association known as glycosylphosphatidylinositol (GPI)-anchoring. This involves the modification of the preprotein in the endoplasmic reticulum, in a process that releases the conserved tail and covalently links the VSG to the glycolipid moiety: glycosylphosphatidylinositol (Ferguson *et al.* 1988). The mature mfVSG is exocytosed to the cell surface in the flagellar pocket, the sole site of (endo)exocytosis in *T. brucei* and from there laterally diffuses to form the coat (Overath *et al.* 1994).

Trypsin digestion has helped separate the VSG into two domains. The larger N-terminal domain defines 66% to 75% of the mature sequence and is extremely variable in primary structure between the large numbers of different possible VSGs. N-terminal domains share only 13-30% sequence identity and contain the epitopes accessible to antibodies *in vivo* (Carrington *et al.* 1991; Miller *et al.* 1984). Randomly aligned polypeptides produce sequence identity values of 10-20% (Doolittle 1986).The smaller C-terminal domain is more conserved between VSG species. The trypsin digestion that separates these two domains occurs in the so called 'hinge' region of the protein whose flexibility has made crystallization of the whole protein problematical (Cross 1984).

In contrast to the sequence diversity in immunoglobulins, the sequence variability of VSGs is distributed throughout the protein, especially in the N-termi-

nal domain. This fact makes it very difficult to align VSG sequences. Carrington *et al.* (1991) however, did align the amino acid sequences of 19 VSGs. They propose, by aligning these sequences in the context of conserved patterns of cysteine residues and glycosylation site distribution, that VSGs be considered as combinations of different classes of N and C-terminal domains. Three classes of N-terminal domain can be identified and designated as classes A, B and C. Four classes of C-terminal domains can be identified (classes 1-4). In this system VSGs are classified as combinations of different domains (e.g. ILTat 1.22 as A2 and ILTat 1.23 as B3 etc.).

Despite the extremely low levels of sequence similarity in the N-terminal domains, VSGs of the same domain type probably share similar tertiary and quaternary structures. The conserved cysteines point towards this possibility. The determination of high resolution structures for two class A N-terminal domains, by X-ray crystallography, provides dramatic confirmation. MITat 1.2 and ILTat 1.24 share only 16% sequence identity but 60% of the sequence forms the same structure. Remarkably, one small α-helix in ILTat 1.24 is substituted in MITat 1.2 by an N-linked oligosaccharide that occupies roughly the same space as the proteinaceous α-helix. The authors of this study contend that '...antigenic variation, therefore occurs by sequence variation and limited conformational modification and not by gross structural alteration...'(Blum *et al.* 1993).

1.2.2 Focus on the VSG surface coat

The VSG coat seems to serve as a protective barrier against specific and nonspecific immune responses of the mammalian host. Coated trypanosomes do not activate complement in the absence of specific antibody and are ingested by macrophages only in the presence of Variable Antigen Type (VAT)-specific antibody (*i.e.* by opsonization). Procyclic form trypanosomes and bloodstream forms whose VSG coats have been enzymatically removed activate complement in the absence of antibody-antigen interactions (alternate pathway) and are readily ingested by macrophages without opsonization (Vickerman 1985). When antigenic variation is considered, a model emerges in which the VSG coat is seen as an antigenically variable filter shielding invariant surface antigens. The filter allows small molecules like trypsin and metabolites to penetrate the coat but blocks the access of the larger antibody molecules to buried invariant epitopes (Overath *et al.* 1994). The structural findings discussed in the preceding section raise the possibility that much of the sequence variability of VSGs could have arisen to thwart the development of T-cell mediated enhancement of humoral immunity, rather than to generate surface coats of dramatically different structures (Blum *et al.* 1993).

1.3 Genetic mechanisms of antigenic variation

Antigenic variation is the prerogative of the proliferative slender bloodstream stage of the life cycle. The two non-proliferative life cycle stages coated with VSG, the metacyclic stage and the stumpy form, are thought not to switch their coats. The activation of VSG in metacyclic trypanosome populations is polyclonal, with fixed proportions of VSGs in the metacyclic repertoire, within a given serodeme, arising in the salivary glands of the tsetse fly (Barry *et al.* 1979; Hajduk *et al.* 1981b). This suggested very early on that the metacyclic VSG repertoire has a distinct mode of activation and serves a different function from its bloodstream counterparts. This will be discussed elsewhere.

With the arrival of molecular biological techniques much of the *T. brucei* research in the 1980's focused on elucidating the rules and mechanisms behind antigenic variation. The results of this research are best summarized as a list of rules with an accompanying diagram of the mechanisms driving antigenic variation.

The following rules apply to antigenic variation:

1) A single VSG gene is expressed at a time from a repertoire of around 1000 genes (Van der Ploeg *et al.* 1982). Most of these 1000 genes are not expressible in their native genomic location and must be duplicatively transposed into special locations where they can be expressed. The silent version of the VSG gene is referred to as the basic copy gene (BC) and the transposed, expressed version is called the expression-linked copy (ELC).

2) The expressed VSG gene is always transcribed from a telomeric location in a bloodstream VSG expression site (B-ES). B-ESs are large polycistronic loci that have the VSG at the telomere proximal end, eight intervening expression site associated genes (ESAGs) and the B-ES promoter, 45 to 60 kb upstream of the VSG gene.

3)10 to 20 such bloodstream expression sites exist, only one of which is normally active at a time. Besides the recombinationally driven antigen switches, Variable Antigen Type (VAT) switches can be effected by changing the expression site that is active (*in situ* activation).

Fig. 1.2 summarizes the mechanisms that have been observed to account for antigen switches. This is taken from Pays *et al.* (1994). This information, in turn was derived from the following sources: reciprocal telomere exchange (Pays *et al.* 1985); duplicative activation and *in situ* activation (Myler *et al.* 1985); segmental gene conversion (Longacre and Eisen 1986; Thon *et al.* 1990; Pays *et al.* 1983).

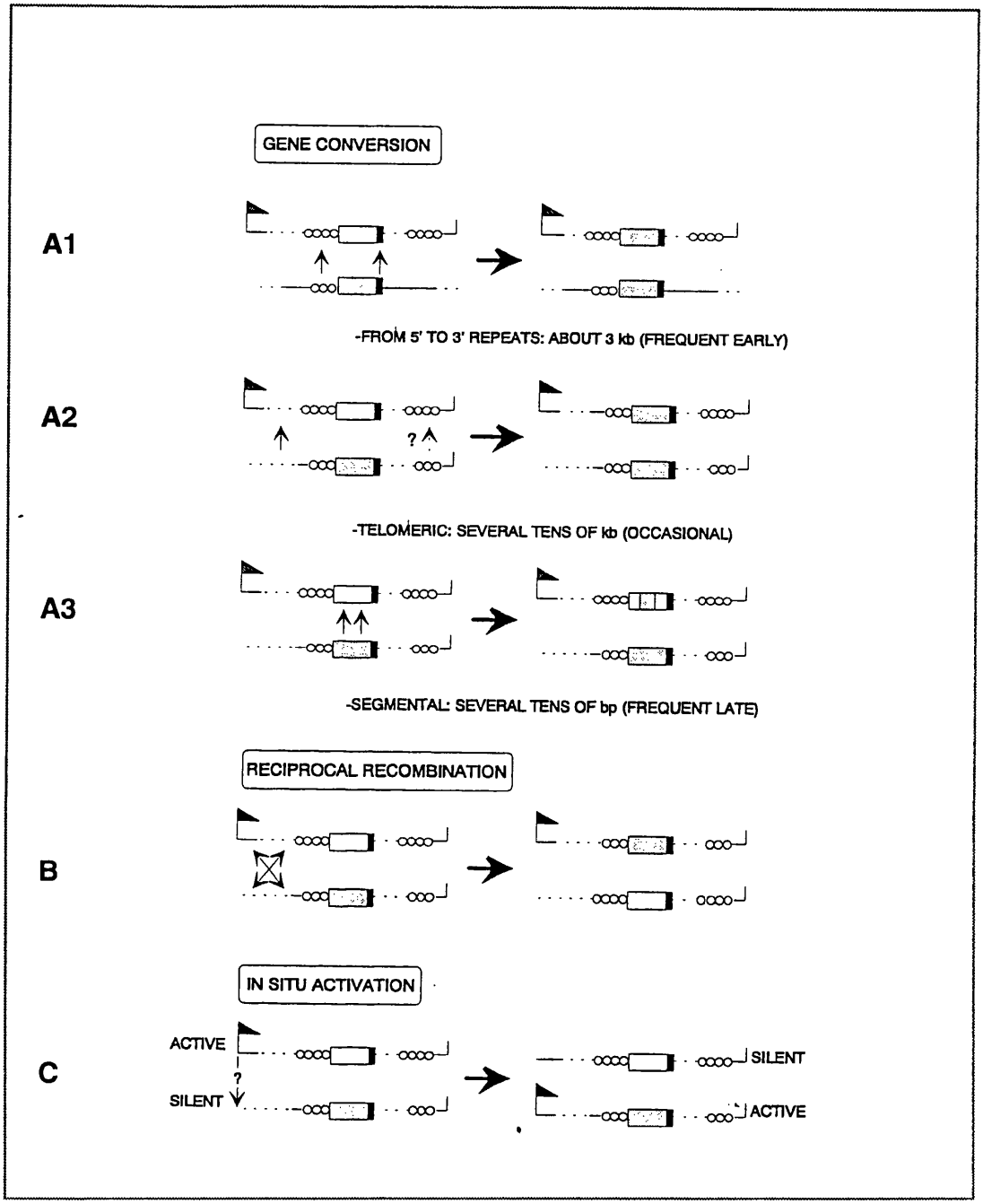


Figure 1.2 Mechanisms of antigenic variation (Pays *et al.* 1994). Flags represent VSG promoters, boxes represent VSG genes with shading indicating sequence homologies, small circles on the 5' end of the VSG genes represent 70 bp repeats, while the circles on the 3' end represent telomere repeats. in A1-3 the vertical arrows indicate the probable boundaries of conversion. In B the crossed arrows indicate a reciprocal recombination event. In C the arrow with the question mark represents a possible cross talk between the expression sites that switch their activity. A1: this represents the norm, accounting for the majority of VAT switch events. A2: in telomeric conversions, association of the telomeres may be important in making this a relatively common event, since inverted telomeric VSGs contribute much less frequently to this kind of conversion. A3: this kind of event is uncommon and therefore its products appear late in the course of a chronic infection. Silent ES are almost certainly used in the multistep process of forming chimaeric genes. B: reciprocal recombination is puzzlingly rare in VAT switches (only one case documented) despite the fact that it should not be deleterious to the trypanosome. C: the process of *in situ* activation is largely an enigma.

1.4 Antigen switching and the course of a bloodstream infection

Bloodstream *T. brucei* infection is characterised by successive parasitaemic waves of different VAT composition. Individual trypanosomes are coated with a single species of VSG but the population itself in the parasitaemic waves is antigenically heterogenous. The host immune system mobilizes an effective immune response to the VATs in a particular parasitaemic wave, eliminating those parasites. Switch variants expand and characterise the next fever relapse. The mammalian host need not present an immune challenge for *T. brucei* to exhibit antigenic variation (Myler *et al.* 1985). However, it is unknown what relationship antigenic variation of cloned trypanosome lines grown in immune-deficient lab animals might have to antigenic variation in the field.

1,4.1 Order of VAT presentation

One of the vital characteristics of the VAT presentation system that makes chronic trypanosome bloodstream infection possible is the loosely 'programmed' temporal order of VAT presentation that occurs. The general order of VAT presentation is essentially independent of host factors, being similar for *T. vivax* infections of five different host species (Barry 1986). Molecular characterisation of expression sites, VSG gene pools and switch events has shown that this temporal ordering of VATs is primarily a function of the probability that a particular VSG will be activated. Several causative factors have been proposed, including genomic location, orientation, sequence of the flanks and finally the sequence of the VSG gene itself.

The VSG genes with the highest probability of activation are those residing at telomeres. Early in an infection, VSG genes already in expression sites are frequently activated by a change of the expression site that is active; an enigmatic process called *in situ* activation. Another easily activated class of VSG genes are telomeric genes outwith expression sites. Their high probability of activation, provided they are in an obverse orientation with respect to the telomere repeats, indicates that interactions of telomeres aid their recombination into active ESs (Van der Werf *et al.* 1990). In turn, these recombination events probably account for the unusual increase in telomere length observed for recipient telomeres involved in telomere conversion events (Myler *et al.* 1985). Later, a single 'dominant' expression site drives expression with subsequent antigen switches occurring by gene replacement (Liu *et al.* 1985). Later in the course of an infection, antigen switches are accomplished by less frequent events, most importantly the activation of VSG genes from internally situated tandem arrays. Here, the probability of activation has been proposed to depend upon the sequence of the flanks. The usual boundaries of conversion fall within the 5' imperfect 70 bp direct repeats and the conserved 3' extreme of the VSG coding frame and the 3' UTR. The number (and by inference) the specific se-

quence of 70 bp repeats present in the 5' flank of the potential donor gene seems to be especially important in the probability of activation (Laurent *et al.* 1984; Pays *et al.* 1994). Some internal VSG genes and some telomeric M-VSG genes lack these 70 bp repeats. These genes (the internal ones at least) contribute to antigenic variation in the latest stages of infection by contributing to segmental gene conversion events that construct functional hybrid genes in the active B-ES. Such multiple, segmental conversions can also use pseudogenes as their donors (Kamper and Barbet 1992; Thon *et al.* 1990). It has even been reported that a M-VSG gene lacking 70 bp repeats may be duplicated, under prolonged selection, in an error-prone manner (Lu *et al.* 1993). It has been proposed that such chimaeric and mutant genes, once created and 'inherited' can become important new members of the VSG repertoire.

To the student of antigenic variation, one of the most mystifying aspects of the subject is how to reconcile this probabilistic mechanism of VSG gene activation with a loose ordering of antigen presentation characterised by parasitaemic waves. Ideas about the rapid clonal expansion of trypanosomes in which a beneficial antigen switch event has occurred are oversimplified. Timmers *et al.* (1987) found that though the 118 VSG gene consistently appears around day 20 in rabbits infected with strain 427 of *T. brucei*, the molecular structure around the 118 expression linked copy (ELC) is heterogenous. Therefore, multiple independent activation events of 118 contribute to the appearance, on day 20, of the VSG encoded by this gene. The picture of a complex, probabilistic mechanism for antigenic variation emerges. This machinery produces population variation that is qualitatively predictable while being quantitatively unpredictable. When this is combined with immune selection which eliminates VATs for which there is memory and against which antibody usually persists for the course of the infection, this system eventually exhausts the VSG repertoire in a hierarchical manner.

1.4.2 Switching rates

Different studies determining the switch rate of the VSG coat have produced estimates ranging from 10^{-2} to 10^{-7} switches per trypanosome per generation (Turner and Barry 1989; Cornelissen *et al.* 1985a & b; Lamont *et al.* 1986). Laboratory-adapted trypanosome lines that have been extensively syringe passaged produce antigenically very homogeneous (~99%) populations that switch infrequently compared to more immediately field-derived lines that exhibit less antigenic homogeneity (60-88%) and much higher switching rates (Turner and Barry 1989). This is an important issue because the numbers derived from the different lines lead to different predictions about the nature of the recombination system responsible for antigenic variation. A low switching rate implies a system that utilises the regular gene conversion machinery, that in other eukaryotes

operates at a mutation rate of 10^{-6} per cell per generation (Jackson and Fink 1981). A switch rate orders of magnitude greater than this basal conversion rate implies the presence of a specific system driving antigenic variation (Turner and Barry 1989).

As has been pointed out, laboratory-adapted lines exhibit switching rates consistent with basal conversion rates reported for other organisms. A model can be envisioned in which a VSG-specific recombination system, that operates at a higher level than the general homologous recombination system, is active in more immediately field-derived trypanosome lines. This system would be operative in recombinations involving 70 bp repeats. The laboratory-adapted switching lines, under this model, have failed to activate a VSG-specific recombination system due to the disruption of the parasite's cell cycle (*i.e.* they are dividing constitutively).

1.4.3 Recombination

It is worthwhile to point out some of the peculiarities of the recombination events that account for the bulk of antigenic variation. Generally, these gene replacement events are referred to as duplicative transpositions. This terminology is imprecise; no evidence supports the existence of transposition *per se*. Instead, the consensus is that gene replacement occurs by gene conversion. Indeed, some aspects of duplicative transposition are reminiscent of the Szostak formulation of general recombination, in which double strand breaks (DSB) initiate recombination (Szostak *et al.* 1983). The donor polarity of duplicative transposition events involving genes of the M-VSG repertoire is a case in point. Metacyclic VSG genes reside in metacyclic expression sites (M-ES). M-ES have a simple structure. One hallmark of the M-ES is the small number or total absence of 70 bp repeats. This simple structure correlates with the relative stability and, by inference, the recombinational isolation of the metacyclic VSG repertoire from the bloodstream repertoire. The relative recombinational isolation of the metacyclic telomeres prevents frequent invasion of this repertoire by genes from the bloodstream pool, while the M-VSGs, themselves, readily replace the VSG in the active bloodstream expression site. The paucity of 70 bp repeats in the M-ESs, in combination with a DSB conversion mechanism, is conjectured to be responsible for donor polarity (Matthews *et al.* 1990). Additionally, the ELC mutations reported by Lu *et al.* (1993), provided they arise during the duplication event, demonstrate that both of the newly synthesized DNA strands produced in this gene replacement event go to the recipient copy, (the ELC). This result is only compatible with a gene conversion event initiated by a double strand break (DSB). It should be noted that the proposal that these mutations arise during the gene replacement has been challenged (Graham and Barry, submitted).

However, some features suggest that trypanosomes are operating recombinational machinery with a difference:
1) the gene conversion machinery operates at a level far above that reported for general recombination in other eukaryotes
2) conversion events between internal VSGs and the ELC it displaces occur with a large bias of gene conversion over reciprocal recombination events

A great deal of heteroduplex DNA must be created in such asymmetric, continuous conversion events. Over the threshold of approximately 1 kb of heteroduplex DNA, the incidence of reciprocal recombination in budding yeast rises to 50% (Ahn and Livingston 1986). Reciprocal recombination could conceivably effect an antigen change without damaging viability but has not been observed, except in one case between telomeres (Pays *et al.* 1985). The implicitly assumed branched intermediates (Holliday junctions) might be resolved by an alternative mechanism not involving cleavage and reunion. This could account for the bias in the observed products.

The unusual features of VSG gene replacement listed above make the system responsible for the mating type switch of *S. cerevisiae* (Haber 1992) seem more analogous to VSG gene replacement than gene conversion operating in the context of general recombination.

1.5 The metacyclic VSG repertoire

Metacyclic trypanosomes are coated with VSG. This stage has a small (<28) VSG repertoire that is expressed independently of the bloodstream VSG repertoire (Esser *et al.* 1982; Crowe *et al.* 1983; Turner *et al.* 1986; Turner *et al.* 1988). The activation of this small metacyclic repertoire is polyclonal (Tetley *et al.* 1987) with the various VATs arising in fixed proportions in a particular serodeme (Hajduk *et al.* 1981). While the metacyclic population is antigenically heterogenous, individual trypanosomes are coated with a single species of VSG.

Polyclonal activation of a predictable subset of VSGs is a very different situation from the antigenic variation that occurs in bloodstream forms after the bloodstream VSG repertoire has been activated (around day 6 of infection). The trypanosomes in the bloodstream from day 1 to 5 are bloodstream forms referred to as metacyclic-derived trypanosomes; these trypanosomes are still expressing M-VSGs *in situ* from metacyclic expression sites. Whereas the role of the switching bloodstream forms is to prolong infection, the metacyclics and the metacyclic-derived trypanosomes must initiate and establish an infection in a host that may have extensive VAT memory; an antigenically heterogenous population would be more effective for this purpose than a monoclonal population (Barry *et al.* 1990).

The relative predictability of the small M-VSG repertoire raised hopes that it may serve as a target for vaccine creation. Those hopes were short lived as it was found that, over time with sequential tsetse transmission, the M-VSG repertoire gradually changes (Barry *et al.* 1983).

1.6 Gene expression

1.6.1 Gene organisation

T. brucei has an almost eubacterial organisation of its genes; intronless and polycistronic. Genes within a polycistron may be tandemly arranged copies of related genes, as is found in the locus for the *T. brucei* glucose transporters or they may be unrelated genes, as is the case in the bloodstream VSG expression sites (Bringaud and Baltz 1993; Pays *et al.* 1989a). The only known exceptions to this polycistronic gene organisation are the metacyclic VSG genes; they are expressed *in situ* from monocistronic, telomeric loci (Graham and Barry 1995). Gene organisation has implications for the control of gene expression. There is very little evidence (the exceptions will be considered elsewhere) for significant control through the life cycle at the level of transcription (*i.e.* most loci are constitutively transcribed). Despite this and the coordinate transcription of many genes, there is significant life cycle stage-specific control of gene expression as well as intralocus differential gene expression (a non-B-ES example of intralocus differential control is found in Bringaud and Baltz 1993). Posttranscriptional control of gene expression is, therefore, implicated in many cases.

1.6.2 RNA processing

The RNA processing reactions of *T. brucei* have some special features. Despite the polycistronic organisation of most *T. brucei* genes, it is highly unlikely that unitary pre-mRNAs are ever produced. Even as transcription proceeds the RNA processing machinery acts upon the nascent RNA, breaking it up into its constituent messengers (Ullu *et al.* 1993). Rapid *trans* splicing (*i.e.* cotranscriptional splicing) cleaves the precursor RNA and then polyadenylation is effected (Ullu *et al.* 1993). Both processes seem to be under the control of an intergenic polypyrimidine tract (Vassella *et al.* 1994). In *trans* splicing, the 5' ends of nascent RNA molecules are processed by the addition of a small 39 nt splice leader RNA derived from a reaction with an 140 nt mini-exon donor RNA (Murphy *et al.* 1986). As the splice leader is supplied by independent loci this phenomenon, first described in *T. brucei*, is called *trans* splicing. All *T. brucei* mRNAs are *trans* spliced, receiving with their splice leader a 5' cap (Comelissen *et al.* 1986). *trans* splicing is thought to be the key to the apparent break, in *T. brucei*, from the strict association between RNA polymerase II (RNAP II) transcription and the production of mRNA, found in the rest of Eukarya (Zomerdijk *et al.* 1991).

1.6.3 RNA polymerases (RNAPs)

Like other eukaryotes, *T. brucei* has three classes of RNAP (Cornelissen *et*

al. 1991). The biochemical observation of outstanding interest in regards to the transcription of the loci containing the major surface antigen genes (parp and VSG) is that this transcription is resistant to very high concentrations of α-amanitin. The degree of α-amanitin sensitivity of transcription is a now classical method for determining which class of polymerase is responsible for that transcription. The insensitivity of the transcription of the parp loci and the VSG expression sites to this drug suggests that they are transcribed by RNAP I, the enzyme complex associated with the transcription of ribosomal DNA (rDNA). There is a growing consensus that RNAP I could indeed be responsible for the transcription of the protein encoding genes in the parp loci and VSG expression sites (this is reviewed by Chung *et al.* 1992). However, rDNA and VSG transcription, when assayed by nascent RNA 'run-on' analysis, display different degrees of α-amanitin sensitivity in the presence of Mn^{2+} (citation of unpublished results in Pays *et al.* 1994). The resolution of this question will probably depend upon the *in vitro* reconstitution of VSG and/or parp transcription.

1.7 The control of gene expression in bloodstream VSG expression sites

VSG expression and bloodstream expression sites are modulated at multiple levels. With the change in environment that precipitates the differentiation of bloodstream trypanosomes into the procyclic form, comes the rapid switch-off of VSG expression. VSG genes are not transcribed in established procyclic cells as transcriptional elongation aborts far upstream of the VSG gene. Once the bloodstream VSG repertoire is engaged, VSG is again expressed from a single B-ES, the other expression sites being silent. There is substantial regulation of VSG expression; life cycle stage-specific control of expression as well as ES-specific expression control in bloodstream forms. What follows is a précis of some of the noteworthy developments in the study of such regulation.

1.7.1 The bloodstream expression site

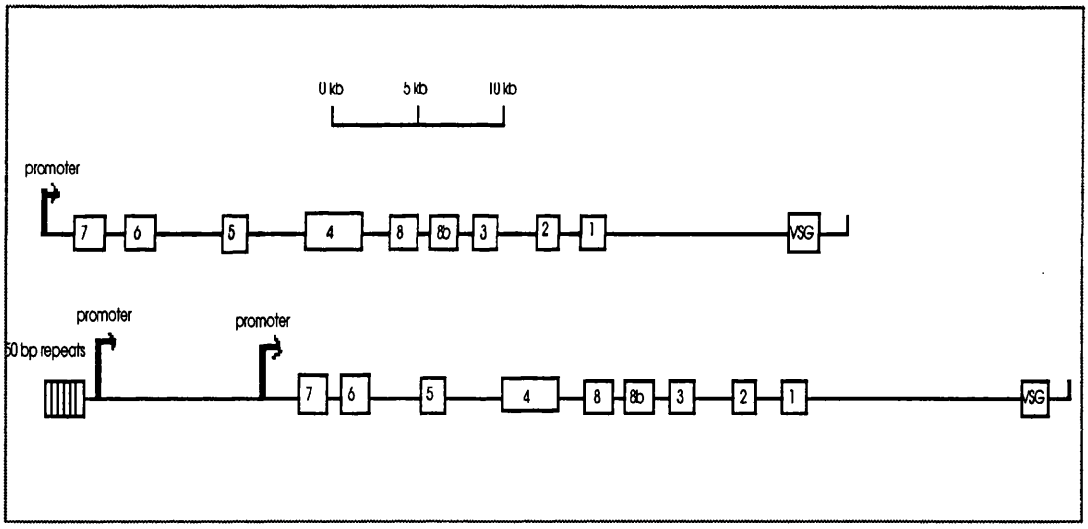


Figure 1.3 The canonical structure of B-ESs. The boxes represent genes. The VSG gene is indicated, while the expression site associated genes (ESAGs) are represented with a number. The long space between ESAG 1 and the VSG gene is populated by a variable but large number of 70 bp repeats.

Fig. 1.3 summarises the polycistronic structure of B-ESs. The nature of RNA processing in trypanosomes plus the sheer size of the B-ES, delayed the isolation of the promoter that drives VSG expression (Zomerdijk *et al.* 1990; Zomerdijk *et al.* 1991). There seem to be two forms of promoter configuration: a single promoter unassociated with repeats and two tandemly repeated promoters separated by 13 kb of intervening sequence, associated with upstream 50 bp repeats. In the ESs with a tandem promoter arrangement, the upstream promoter and the intervening sequence are dispensable. An active expression site with a tandem promoter structure has been observed to delete the upstream pro-

moter along with the intervening sequence in a relapse switch in which the ES is inactivated. This ES was readily reactivated without any further recombination (Gottesdiener *et al.* 1991). Such deletion events could be stimulated by transcription itself (Voelkel-Meiman *et al.* 1987).

Figure 1.4 presents a schematic summary of what is known about the structure and function of the products encoded by the ESAGs. Expression studies of bloodstream expression sites and of ESAGs have clearly demonstrated that there are ESAGs outwith expression sites, and that ESAG expression is not confined to the bloodstream stage of the life cycle (Pays *et al.* 1989a; Graham and Barry 1991). The fact that some ESAGs are expressed outside the context of bloodstream expression sites and that the expression of various ESAGs are detected in all life cycle stages, raises the possibility that the products of ESAGs are required in the procyclic and metacyclic stages of the life cycle.

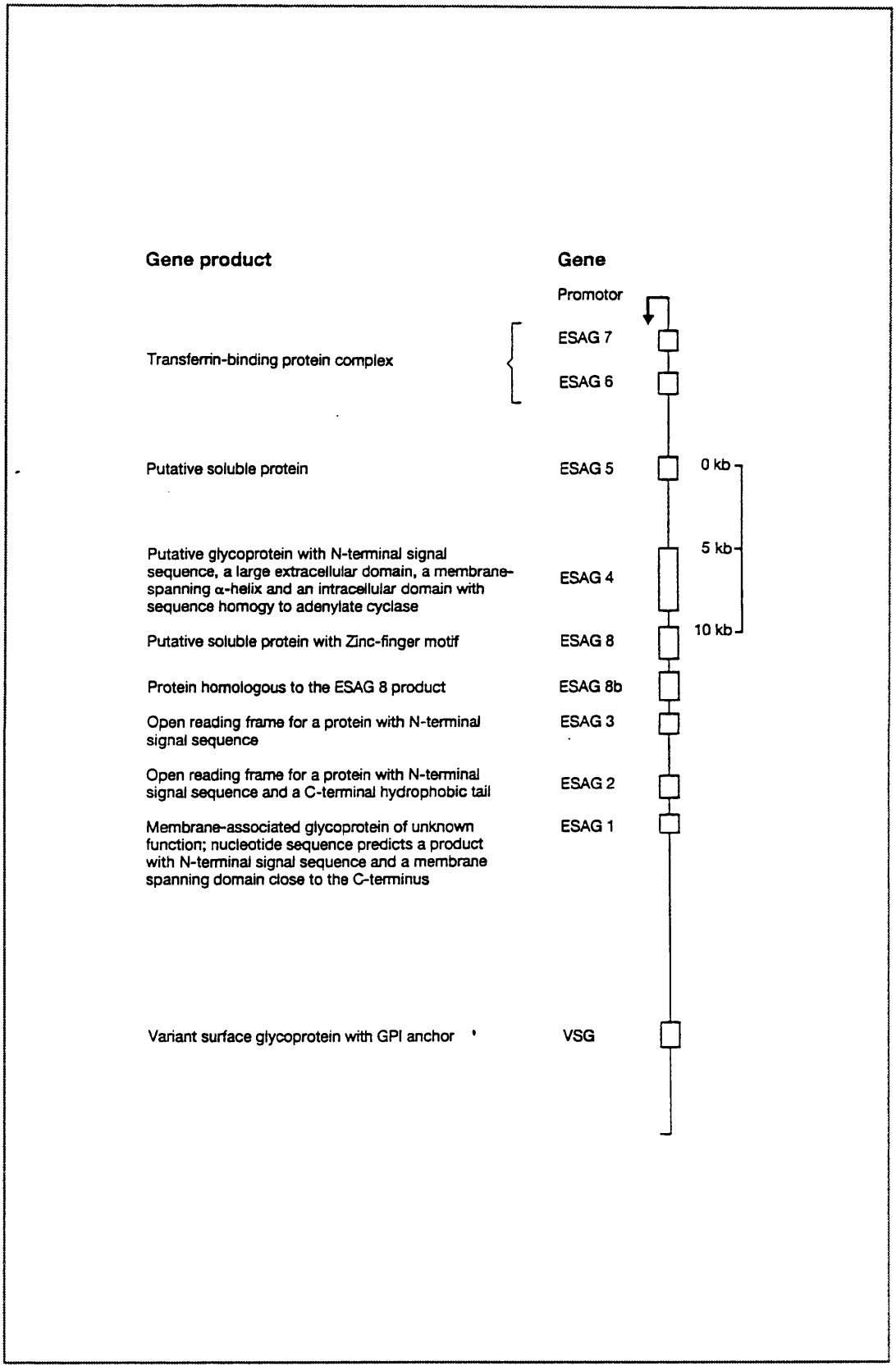


Figure 1.4 The ESAGs (Overath *et al.* 1994).

1.7.2 Life cycle stage-specific VSG expression

Transient transfection studies of the B-ES promoter show that it is constitutively active in the genetic context of a reporter construct (Zomerdijk *et al.* 1990; Jefferies *et al.* 1991). Furthermore, transcriptional run-on assays of a B-ES promoter indicate strong promoter activity exists in the procyclic form. None of the genes in a B-ES is transcribed because the transcription initiating at a B-ES promoter is aborted ~500 bp downstream from the promoter (Zomerdijk *et al.* 1990; Rudenko *et al.* 1994). The latter study demonstrates that the B-ES promoter activity observed in procyclic stage transcriptional run-ons is due to the low level activity of many B-ESs. This discredits the notion of a single constitutively active B-ES promoter; a hypothesis that sought to explain the apparent memory of VSG expression. In short, the VAT expressed just prior to differentiation into the procyclic form re-emerges as soon as the bloodstream VSG repertoire is reactivated following the fly transmission (Hajduk and Vickerman 1981 b). Pays *et al.* (1989 b) proposed that the constitutive activity of a single B-ES during the developmental progression from ingested bloodstream trypanosomes through the procyclic and metacyclic stages to the initial bloodstream forms, could account for this apparent VAT memory. Instead, it seems that this appearance of memory may be caused by more prosaic factors like the probability of reactivation of a given B-ES.

The nature of the aborted transcription downstream of the B-ES promoters is unclear. If it is classical attenuation then the *cis*-acting sequences are unable to attenuate RNAP I transcription driven by a ribosomal promoter replacing an endogenous B-ES promoter (Rudenko *et al.* 1994). If this aborted transcription is a failure of elongation, it acts only on transcription complexes specific to the B-ESs. It is tempting to postulate the existence of a cold sensitive elongation factor that cannot assemble with other proteins to form an operative transcription complex at the reduced temperatures found in the tsetse midgut or in the laboratory culture flask (27°C). Indeed, if α-amanitin sensitivity of VSG transcription was not due to RNAP I transcription, one could foist the responsibility for drug sensitivity onto this imaginary elongation factor.

What is clear from the Rudenko study is that there is a substantial level of transcriptional control at B-ESs in the procyclic stage of the life cycle. This control, mostly down regulation, is both promoter and position-dependent. A B-ES promoter is derepressed when placed in the ribosomal spacer region while a ribosomal promoter's activity is unimpaired when it replaces the endogenous B-ES promoter (Rudenko *et al.* 1994). It must be stressed that the perplexities of such context dependent changes in promoter activity (position effects) should be considered in the light of our knowledge of the organism's unusual RNA processing. *trans*-splicing supplies a 5' cap to all RNA's with the appropriate signal sequence, irregardless to which polymerase class the promoter driving its expression happens to belong. This may allow the B-ES promoter, when moved by investigators into the ribosomal gene cluster (*i.e.* the nucleolus) or into

a plasmid, to recruit RNAP I for its transcription. This may never happen in the unmanipulated genome. On balance, I believe that the Rudenko results indicate that VSG bloodstream expression sites are not *normally* transcribed by the same transcriptional machinery as is rDNA.

An additional mechanism ensuring the life cycle stage-specific expression of VSG genes operates at the posttranscriptional level. Transient and stable transfection experiments demonstrate that a 97 nt sequence fragment upstream of the poly(A) addition site in the 3' UTR controls mRNA abundance without affecting transcription. In procyclic form trypanosomes this region causes a reduced efficiency of mRNA maturation, while in bloodstream forms it causes an increase in mRNA stability (Berberof and Vanhamme *et al.* 1995). The authors of that study suggest that this extra level of control tightens the regulation of VSG expression and is especially crucial in the early stages of differentiation from bloodstream to procyclic forms. Indeed, VSG mRNA can be detected for 13 hrs. after differentiation is triggered (Pays *et al.* 1993), a period corresponding to at least one cell cycle. It is during this period that it is presumed that posttranscriptional control would be most important. The 3' UTR of parp is demonstrated in the same study to confer stage specificity in the converse pattern to its VSG counterpart.

Once the bloodstream VSG repertoire is again engaged, a single expression site, out of a potential pool of 10-20, is activated. The ES that is active can change in a process referred to in the literature as *in situ* activation, but the rule is that only a single ES is active at a time. This suggests that a mechanism for silencing the other expression sites exists in bloodstream trypanosomes.

1.7.3 Silencing and *in situ* activation

It is unknown how a B-ES is activated. Most of the hypotheses about B-ES expression concern the silencing of the remaining ESs. This work centres on the phenomena of DNA modification and position effect.

Comparisons of active and silent expression sites which seek to find key differences have failed to find any clearly causal 'smoking gun'. There has been a consistent failure to find any clear causal recombination events associated with ES inactivation or reactivation. Chromatin structure comparisons have not revealed any difference in nucleosomal packaging. However, the same study notes a marked difference in sensitivity to single-strand-specific endonucleases between active and silent expression sites. In particular, *Hinf* I sites are completely resistant to digestion in silent expression site chromatin (Greaves and Borst 1989). As this sensitivity, in the active VSG gene, is not due to preexisting single strand nicks, DNA modification or to transcription *per se*, the authors propose that the chromatin of the active ES is torsionally stressed. *Hinf* I diges-

tion does not prevent adjacent *Hinf* I sites from being digested, so if the chromatin is torsionally stressed, the stress is constrained into small units. Such a difference in chromatin structure perhaps would go undetected in a micrococcal nuclease assay.

Another fascinating difference between active and silent ESs occurs at the DNA level. Silent B-ES DNA in bloodstream trypanosomes contains an unique modified nitrogen base: β-D-glucosyl-hydroxymethyluracil ('J') (Gommers-Ampt *et al.* 1993). This modified base is present in silent subtelomeric VSG genes in bloodstream trypanosomes while the active VSG gene and internal VSG genes lack J. Base J is undetectable in procyclic form trypanosomes. In bloodstream forms, the occurrence of J correlates very tightly with the silence of subtelomeric VSG genes. It seems, however, that modification with J probably does not cause silence since the correlation is with the silence of the VSG gene and not with the transcriptional status of the B-ES. This is known because a variant of *T. brucei*, in which two B-ESs are simultaneously active, has been studied. This variant expresses only a single VSG, due to a 30 kb insertion just upstream of the VSG coding unit, in one expression site, that disrupts the transcription of the VSG gene without affecting the transcription of the rest of the ES. The inactive VSG gene in this mutant ES is modified with J (Cornelissen *et al.* 1985 c). An alternative way of thinking about the possible role of DNA modification in silencing phenomena is to consider it as a 'flag' that marks chromatin for assembly into some epigenetically heritable state.

This brings the discussion into the realm of position effect. It was discovered, through stable transfection experiments into silent expression sites, that a promoter-independent stage-specific position effect operates on reporter constructs integrated just upstream of the VSG gene (Horn and Cross 1995). The position effect operates only in the bloodstream stage of the parasite. It is tempting to think of this as a generalized telomere position effect but this seems inconsistent with the silencing of a very strong promoter that operates over >60 kb. In *S. cerevisiae* telomere position effect only operate typically over a distance of 3.5 kb or less, depending on the strength of the promoter tested. Causing the overexpression of a silencing information repressor gene (*SIR3*) by supplying it on a multicopy plasmid (2μ) can only extend the silencing effect over an additional 14 kb (Renauld *et al.* 1993). Additionally, it is known that one of the promoters repressed in Horn and Cross (1995) (rDNA promoter) is fully active adjacent to a telomere in a minichromosome (Zomerdijk *et al.* 1992). Another kind of 'position effect' could be special nuclear localisation required for activation of a B-ES. Chung *et al.* (1990) have shown that the organisation of the *T. brucei* interphase nucleus is non-random. Perhaps the organisation of the genetic material impacts upon its expression.

This author believes that chromosome topology and nuclear geometry are likely to play a role in regulating the expression of loci driven by promoters lacking strong specific repression. One can imagine a specific nuclear location to which a subtelomeric region must be brought, in order to be transcribed by a B-

ES specific transcription complex. Subtelomeric VSG genes that are not being transcribed are detected by enzyme machinery that tracks down from the telomere repeats and modifies their DNA. The degree of DNA modification with 'J' follows a gradient that decreases distal to the telomere (Bernards *et al.* 1984). This modification could signal the packaging of this DNA into a chromatin configuration that is silenced in a promoter-independent fashion, perhaps by a loop-out encompassing the ES that leaves the rest of the chromosome expressible (Hoffman *et al.* 1989). Strong silencing is absent in the procyclic stage but down regulation now occurs by processes that affect elongation and RNA processing. Needless to say, this is a speculative synthesis.

1.8 Metacyclic VSG expression site control

Metacyclic VSG genes are expressed from subtelomeric positions, like their bloodstream counterparts, but their expression sites are simpler. The genomic architecture of the M-ES consists of a promoter, a very small number (0-2) of 70 bp repeats, the VSG coding region and the repeats that characterise the telomere (Lenardo *et al.* 1986; Matthews *et al.* 1990). M-ESs are restricted to the telomeres of the largest chromosomes and are activated specifically in metacyclic trypanosomes *in situ* (Cornelissen *et al.* 1985 a; Lenardo *et al.* 1986; Graham *et al.* 1990). Upstream of these monocistronic transcription units, the first described in *T. brucei*, is low copy number sequence and at most a single ESAG (Matthews *et al.* 1990; Son *et al.* 1989). Whether this genomic environment looks similar or dissimilar to a B-ES seems to depend upon the perspective of the investigator (Son *et al.* 1989; Graham and Barry 1995). What is clearly different is the unambiguous assignment of the (metacyclic) VSG promoter to a region very close to the telomere; thereby defining a monocistronic transcription unit (Alarcon *et al.* 1994; Graham and Barry 1995). A key difference between the promoters examined in these studies is noted in their activity in procyclic transient transfection experiments. The promoter of the MVAT 4 ES, in a variant which is expressed *in situ* in bloodstream trypanosomes, is active in such experiments. The promoter of the 1.22 M-ES is silent in procyclic transient transfection experiments. If one assumes that the promoters from both studies are genuine, developmentally regulated promoters, then the studies lead one to make different inferences about how transcriptional control of M-ESs occurs. Position effect must be invoked, as for B-ES promoters for the MVAT 4 promoter and classical *trans*-acting repression for the 1.22 promoter. Further characterisation of the 1.22 promoter has identified an element similar to that described by Alarcon *et al.* (1994) upstream of the 1.22 metacyclic promoter, that in transient transfection assays drives reporter gene expression in procyclic trypanosomes, suggesting that the M-VAT4 'metacyclic promoter' is artifactual (S.V. Graham, personal communication).

Regardless of how the transcriptional control is effected there is a fundamen-

tal similarity to the bloodstream system; only one ES is expressed out of a pool of potential expression sites. Once again the interesting aspects of this problem lie with the mechanisms of silencing. No promoter replacements into silent M-ESs have been performed, so it is not known if metacyclic expression sites are subject to a developmentally regulated position effect.

The resemblance of the genomic environment of the M-ESs studied in J.D. Barry's laboratory, to that found in B-ESs, is slight. The area upstream of the M-ES, that in B-ESs would be populated with ESAGs, is mostly conserved, single copy, transcriptionally silent sequence (Graham *et al.* 1990). It is interesting that such sequence is found in an area over which a position effect is presumed to act in bloodstream expression sites. We propose that this extensive area (> 13 kb) be designated as the 'metacyclic domain'. Any single copy, silent sequence linked to the M-VSG gene, would be considered as part of the metacyclic domain. In order to understand the nature of the control of M-VSG gene expression it would be fruitful to define the extent of the 'metacyclic domains' for the ILTat 1.22 and 1.61 M-VSG genes and to determine if position effects are operative over these 'metacyclic domains'. The work reported in this thesis is an effort to define the metacyclic domain of the ILTat 1.22 M-VSG gene.

Chapter 2

Materials and Methods

2.1 Abbreviations

ARS: autonomously replicating sequence
BC: basic copy
B-ES: bloodstream VSG gene expression site
DMSO: dimethyl sulfoxide
DTT: dithiothreitol
EATRO: East African Trypanosomiasis Research Organisation
ELC: expression linked copy
ESAG: expression site associated gene
EtOH: ethanol
FIGE: field inversion gel electrophoresis
ILTat: International Laboratory for Research on Animal Diseases *Trypanozoon* antigen type
IPTG: isopropylthiogalactoside
LMP: low melting point (agarose)
M-ES: metacyclic VSG gene expression site
parp: procyclic acidic repetitive protein (procyclin)
PEG: polyethylene glycol
PFG(E): pulsed-field gel electrophoresis
pfu: plaque forming unit
PMSF: phenylmethylsulfonyl fluoride
R/T: room temperature
SS: sorbitol selection
SDS: sodium dodecyl sulphate
SSC: sodium citrate, sodium chloride
STIB: Swiss Tropical Institute, Basel
TE: Tris-HCl (8.0), EDTA
TTE: Tris base, taurine, EDTA
TBE: Tris base, boric acid, EDTA
TAE: Tris base, acetic acid, EDTA
VAT: variable antigen type
VSG: variable surface glycoprotein
YCD: yeast nitrogen base, casein, dextrose
YPD: yeast nitrogen base, peptone, dextrose

2.2 Strains used

S. cerevisiae strains

YPH 252: *MATa, ura3-52, lys2-801 amber, ade2-101 ochre, trp1-Δ1, his3-Δ200, leu2Δ1*

YPH 250: *MATα*, ...

E. coli strains

XL1-Blue MRA: *Δ(mcrA)183, Δ(mcrCB-hsdSMR-mrr)173, endA1, supE44, thi-1, gyrA96, relA1, lac*

All λ work was done in this strain background which is restriction deficient, endonuclease deficient and designed to stabilise methylated DNA. The fact that this strain is recA positive is important for robust growth which yields higher phage titres.

XL1-Blue MRF': *Δ(mcrA)183, Δ(mcrCB-hsdSMR-mrr)173, endA1, supE44, thi-1, recA1, gyrA96, relA1, lac, λ-[F', proAB, lacI^{q}ZΔM15, Tn10, (tet^r)]*

This strain was used for all standard subcloning work as the F' episome allows for blue/white selection.

Trypanosomes

Three trypanosome stocks were used in this work: EATRO 2340 (Cornelissen *et al.* 1985 a), EATRO 795 (Turner and Barry 1989), and STIB 386 (Felgner *et al.* 1981). Most work was done with various cloned expressors derived from the EATRO 795 stock. These trypanosomes were originally stabilated from the blood of an infected cow at Uhembo, Kenya in 1964. Trypanosomes were grown and 'harvested' from CFLP mice or Wistar rats, depending upon the amount of material required, as outlined by Hajduk and Vickerman (1981 b).

2.3 Microbiological media

2.3.1 Media for growth of yeast

Sorbitol Selection (SS) agar (for YAC library plating):
top agar:
0.17% (w/v) yeast nitrogen base without amino acids
0.5% (w/v) ammonium sulphate
2.5% (w/v) Bacto agar
3% (w/v) dextrose
18.2% (w/v) sorbitol
+ supplement

bottom agar:
0.17% (w/v) yeast nitrogen base without amino acids
0.5% (w/v) ammonium sulphate
2.0% (w/v) Bacto agar
3% (w/v) dextrose
18.2% (w/v) sorbitol
+ supplement

100X supplement (100 mls):
1% ade (hemisulphate)
0.2 g arg
0.1 g his
0.6 g ile
0.4 g lys
0.1 g met
0.6 g phe
0.5 g thr
0.6 g leu
1.6 g val

YCD broth and (agar):
2% glucose
0.17% (w/v) yeast nitrogen base without amino acids
0.5% (w/v) ammonium sulphate
1% Cas(ein) amino acids
15 mg/L adenine hemisulphate
pH to 5.8 with acetic acid
(2% Bacto agar)

YPD broth and (agar):
2% glucose
1% yeast extract
2% peptone
pH to 5.8 with acetic acid
(2% Bacto agar)

2.3.2 *E. coli* growth media:

All *E. coli* work was done with Luria Broth/Agar (LB) as reported in Sambrook *et al.* (1989). λ work was done with LB bottom agar and BBL top agar supplemented with magnesium sulphate (10mM) and maltose [0.2% (v/v)]. Liquid lysates were set up in LB supplemented only with $MgSO_4$. BBL agar is an inferior medium to LB resulting in slower *E. coli* growth and larger plaque size.

2.4 Electrophoresis

Sequencing gels:

BRL vertical gel systems were used under glycerol-tolerant conditions. This essentially means that TTE running buffer was used. TTE stands for Tris-Taurine-EDTA.

Pulsed Field Gel Electrophoresis (PFGE):

A functional Biometra gel system was available through most of the work. This system has a circular electrode configuration that rotates above a submerged horizontal gel system to produce the 'pulsed' field. Liquid samples were cooled for several minutes on ice prior to loading to prevent their buoyancy from carrying the samples out of the wells. Chromosome blocks were loaded by placing the blocks on the teeth of a gel comb and then the gel was cast around the comb.

2.5 Standard protocols

2.5.1 *T. brucei* genomic DNA preparation

1) Lyse the trypanosomes in 1-5 mls of **NET buffer.**
2) Add 1/10 th volume of 30% sarcosine in NET buffer.
3) Add proteinase K to 100 μg. ml^{-1} and incubate this at 37°C for 30 min. Leave this at room temperature (R/T) till required.
4) Add an equal volume of phenol/chloroform, agitate, spin at 4960 g for 5 min at R/T and extract the aqueous phase.
5) Add 2-5 mls of NET buffer to the remaining interphase, extract again with phenol.
6) Pool aqueous phases.
7) Add Na acetate to 0.3M and then add 2.5 volumes of ethanol.
8) Spool and dissolve the precipitated DNA in 0.5 ml of TE buffer.
9) Add RNase to 50 μg. ml^{-1}, incubate at 37°C for 10 min.
10) Treat with proteinase K at 100 μg. ml^{-1} for 30 min.
11) Repeat the organic extraction process.
12) Dissolve the precipitated result in 0.5-1.0 ml of TE.
13) Store at 4°C.

NET buffer: 100mM NaCl, 10mM Tris-HCl (8.0), 1mM EDTA

2.5.2 Yeast genomic DNA preparation

The protocol of Sherman *et al.* (1986) was followed with the exception that spooling was replaced with centrifugation in the ethanol precipitation step. This very simple protocol involves cell lysis, a salt precipitation of protein, ethanol precipitation of nucleic acids, and an RNase treatment. This protocol is ideal for handling many small samples that will be quickly used. If a large amount of genomic DNA is required for long-term use, a protocol that uses organic extraction should be used.

2.5.3 Rapid protocol for the preparation of chromosome-length yeast DNA

The referenced protocol of Riley *et al.* (1992) is best described as a chromosome block miniprep. The cells are spheroplasted by enzymatic disruption of their cell wall, embedded in low melting point agarose and then directly lysed in the blocks by treatment with a lithium dodecyl sulphate buffer. This kind of preparation is not suitable if the DNA must be enzymatically manipulated before being run on a gel but is adequate if a regular karyotype gel is required.

2.5.4 Preparation of high grade chromosome-length DNA

This protocol was used if high grade yeast chromosome blocks were desired for enzyme manipulations. This protocol should also be used for all preparations of chromosome-length DNA from *T. brucei* as the VSG coat presents a formidable amount of protein to eliminate by any means except proteinase K treatment.

1) Harvest cells by centrifugation and perform any pretreatments needed to make the cells prone to lysis (*i.e.* spheroplasting, if you are using yeast).
2) Mix cells 1:1 in low melting point (LMP) agarose and cast into chromosome blocks.
3) Transfer plugs into 4-5 volumes of **YLB** and incubate at room temperature for 1 h.
4) Replace with 10 volumes of **YLB** and incubate at 45°C, overnight.
5) Rinse plugs in TE several times, letting them sit in the TE for 5 min on each rinse.
6) Place the chromosome plugs in **NDS** solution with 1 mg. ml^{-1} of proteinase K, incubate this at room temp. for 30 min, then at 50°C overnight.
7) Repeat step 6) until the plugs become transparent, one to two days will suffice for yeast but trypanosome blocks can require up to 5 days of proteinase K treatment.
8) Pour off the **NDS**/protease K solution and add **wash solution**, letting the blocks sit in this for 15 min; repeat twice.
9) Decant and add **wash solution** with 1mM PMSF (protease inhibitor) and incubate at room temp. for 1 h.
10) Wash away the PMSF by washing the blocks as in step 8) in **storage solution.**
11) Store at 4°C.

YLB: 1% lithium dodecyl sulphate, 100mM EDTA, 10mM Tris-HCl (8.0)
NDS: 1% N-lauryl sarcosine, 0.5M EDTA, 10mM Tris (9.5)
wash solution: 20mM Tris-HCl (8.0), 50mM EDTA
storage solution: 2mM Tris-HCl (8.0), 5mM EDTA

2.5.5 Plasmid minipreps

The classic protocol of Birnboim and Doly (1979) was occasionally employed, but a faster and simpler method was favoured. Again, as in **2.5.2**, this DNA preparation uses a salt precipitation to rid the lysate of proteinaceous material. Despite the apparent crudeness of the technique, it generally yields DNA good enough to be sequenced.

1) Pellet 1.5 ml of overnight bacterial culture for ~10 s, pour off the medium

and resuspend the pellet in the residual medium.
2) Add 300 µl of **TENS** buffer and vortex for 5 s.
3) Add 150 µl of 3M Na acetate (pH 5.2) and vortex for 5 s.
4) Centrifuge for 2 min and transfer the supernatant to microfuge tubes containing 900 µl of ice-cold,100% EtOH.
5) Centrifuge for 2 min to precipitate the plasmid DNA.
6) Resuspend in TE with a trace of RNase (<1 µg. ml^{-1}), leave at room temp. for 5 min and then store the plasmid at -20°C.

TENS: TE with 0.1N NaOH and 0.5% SDS

2.5.6 Preparation of inserts for subcloning

Two methods were routinely employed for the preparation of insert DNA for plasmid subcloning. The organic extraction of DNA from LMP agarose gel fragments reported in Sambrook *et al.* (1989) was used occasionally but when available the less hazardous and faster Qiaquick gel extraction kit was used as suggested in the manufacturer's protocol (Qiagen). In this protocol regular TAE agarose preparative gels were used, thus affording better resolution. It was found that the quality of the agarose used was critical. Best results were abtained with SeaChem agarose from (FMC). The gel fragment was chemically depolymerized and the dissolved material centrifuged (12 000 g) through a Qiaquick centrifuge filter containing a DNA-binding matrix in its filter. The fixed DNA is washed with an EtOH-containing solution and the fragment is eluted with TE. This eluate could be stored for long periods at 4°C and served as an excellent substrate for random priming reactions. An ethanol precipitation was used to concentrate the sample when the fragment was required for subcloning.

2.5.7 Labelling and purification of DNA probes

Without exception DNA probes were labelled by random priming, using the Stratagene random labelling kit. These labelling reactions were carried out for 10 min to 1 h at 37°C. The features of the kit that allow for such rapid labelling were the 8-mer random oligonucleotides (as opposed to the usual 6-mers used in other protocols) and a version of the Klenow fragment lacking its native proofreading exonuclease function, that allows for incubation at 37°C.

The resultant probes were purified away from unincorporated nucleotides by Stratagene's NucTrap columns. These columns are G50 Sephadex gel filtration columns designed to be used under pressure, thus expediting the purification procedure. Additionally, the NucTrap columns are continually shielded by perspex throughout the procedure, reducing the laboratory worker's radiation exposure.

2.5.8 Preparation of λ DNA

Large amounts of bacteriophage DNA can be obtained from 10 ml of a successful lysate.
1) Prepare a plate with 10-100 plaque forming units (pfus) from the relevant clone.
2) Aspirate a single plaque with a pasteur pipette and preincubate it with 500 µl of cells suspended in 10mM $MgS0_4$ (OD_{600} = 0.5) for 30 min at 37°C.
3) Add this to 40 ml of preheated LB broth supplemented with $MgSO_4$ to 10mM.
4) Incubate at 37°C with agitation for 6 to 8 h.
5) Add 500 µl of chloroform to the lysate and shake gently at 37°C for a further 20 min.
6) Clear the lysate by centrifuging the sample at 13 800 g for 10 min at room temperature.
7) Take 10 ml of the lysate for DNA preparation and store the rest at 4°C as a storage stock.
8) Add RNase and DNase to the lysate to a final concentration 10 µg. ml^{-1}, incubate at room temp. for 30 min.
9) Add 0.4 volumes of a 33% PEG/3.3M NaCl solution to the lysate, incubate on ice for 30 min.
10) Precipitate the phage particles at 4°C, 13 800 g for 10 min.
11) Decant the supernatant and drain it off completely by keeping the tubes inverted for 5 min, which prevents PEG contamination of the ensuing the nucleic acid precipitation.
12) Resuspend the phage pellet in 2 ml of phage dilution buffer.
13) Add an equal volume of phenol/chloroform, extract and extract again with chloroform to remove residual traces of PEG from the supernatant.
14) Precipitate and wash the phage DNA with EtOH.
15) Resuspend the phage DNA in 100 to 500 µl of TE supplemented with 1 µg. ml^{-1} RNase.
16) Store the phage DNA at -20°C.

2.5.9 Transformation of *E. coli*

Competent XL1-Blue MRF' cells purchased commercially (Stratagene) were used in most subcloning exercises.
1) Add 25 µl of competent cells with about 40 ng of ligation mix (never more than 1 µl of ligation mix), stand this on ice for 30 min.
2) Heat shock at 42°C for 45 s (60 s for most other strains).
3) Put the cells back on ice for 2 min.
4) Add 1 ml of prewarmed LB broth to the cells and incubate at 37°C for 1 h.
5) Before this incubation period is finished prepare selection plates for blue/

white selection, if appropriate, by adding 50 µls of 2% X-gal and 6 µl of 50 mg. ml^{-1} IPTG to the surface of the plate and spread this over the surface of the plate [tetracycline (40µg. ml^{-1})/ampicillin (100 µg. ml^{-1})].
6) Add 50-200 µl of cell suspension to the surface of the plate and spread.
7) Dry at 37°C for 20 min and then invert and incubate overnight at 37°C.
8) The next day the plates should be stored at 4°C for a few hours to allow the blue colour of non-recombinant colonies to develop.

2.5.10 Nucleic acid transfer (blotting)

Nucleic acid transfers were conducted onto nylon transfer membranes. Transfer from PFGE gels was a particular problem giving very inconsistent results until a high quality, batch tested, positively charged nylon membrane was used: Zetaprobe GT membrane from Biorad. Transfer of chromosome sized DNA fragments absolutely requires acid depurination (0.25M HCl; 15min).

Other routine DNA transfers were undertaken using Hybond N and N+ (Amersham). Transfer of RNA was performed onto Genescreen (Dupont). When DNA was transferred onto Hybond N+ alkali (0.4M NaOH) was used to transfer the DNA. When Hybond N or Zetaprobe GT was used, the transfer protocol used was the classic southern method reported in Sambrook *et al.* (1989).

Nucleic acids were fixed to the nylon membranes by a combination of UV crosslinking and baking (60 mJ and 80°C baking for 2 h).

Hybridised blots were stripped with 0.4M NaOH for 20 min at room temp. with subsequent washing in 2X SSC. According to the manufacturer Hybond N filters should be stripped with boiling 0.5% SDS but the alkali stripping procedure does not seem to harm the filters. Northern blots must be stripped with boiling SDS.

2.5.11 Hybridisation

Two hybridisation techniques were employed for DNA blots: 1)a Na phosphate/SDS based hybridisation and 2) a SSC/Denhardts' system.
1)Na_2HPO_4 (7.2), 7% SDS
2) 6X SSC, 5X Denhardts' solution, 0.5% SDS, 100 µg. ml^{-1} of sheared and denatured herring sperm DNA

50X Denhardts' stock is 1% (w/v) of: Ficoll, polyvinylpyrrolidone and bovine serum albumin (Fraction V)

Buffer 1's blocking agent is the high concentration of SDS and this works well unless there is a large excess of target material on the blot. In my experience, this hybridisation system was found to be inappropriate for hybridisations of blots with plasmid or λ clones, as too much target material seems to over-whelm the blocking activity of the SDS.

All the ingredients in buffer 2, aside from the SSC, are blocking agents. This buffer system is appropriate for hybridisation experiments involving common and rare sequences.

High stringency hybridisations with both buffer systems were conducted at 65°C. Posthybridisation washes were done with 2X SSC/0.1% SDS at room temp. as a rinse step and then with 0.1X SSC/0.1% SDS for 30 min at 65°C, for the high stringency wash.

2.6 Specialised protocols

2.6.1 Spheroplasting *S. cerevisiae*

This protocol weakens and/or removes the yeast cell wall enabling the cells to be easily lysed for biochemical extractions or to be fused in transformation experiments. The protocol listed here is for the production of a large set of spheroplasts or 'ghosts' appropriate for use in YAC library construction.

1) Isolate a single colony and set up a 1 ml overnight YPD culture (30°C), use this culture the next day to seed a 50 ml culture from which the spheroplasts will be made, incubate overnight at 30°C.
2) Take an absorption spectrophotometry reading, at 660 nm, to determine the density of the culture. A reading of 2-4 is ideal indicating, a culture density of ~3-5X10^7 cells. ml^{-1}, well short of stationary phase when the cells become difficult to spheroplast.
3) Pellet cells by centrifigation (510 g, 5 min, room temp.).
4) Wash cells once by vortexing with 25 ml of dH_2O.
5) Pellet cells again as in 3).
6) Resuspend cells in 25 ml of 1M sorbitol and incubate at room temp. for 2 min.
7) Repellet the cells as in 3).
8) Resuspend pellet by vortexing in 15 ml of **SPE**.
9) Add 75 µl of 2M dithiothreitol (DTT).
10) Dilute a sample of this suspension 1:10 in H_2O and take an absorption spectrophotometry reading to establish a baseline A660 value. A660 values decrease as more cells become spheroplasted so this is used as an assay of the progress of the 'lyticase' reaction.
11) Add 1000 units of lyticase resuspended in **SPE**.
12) Incubate the reaction tube at 30°C with occasional gentle swirling.
13) Remove samples every 15 min for A660 readings. Stop the reaction when the reading are 30-40% of the baseline reading determined in step10).
14) When adequate spheroplasting is achieved, gently pellet the culture by centrifugation (280 g, 4 min, at room temp.).
15) Resuspend the the pellet in 15 ml of **STC** by gentle aspiration.
16) Repellet, as in 14), wash the pellet with 15 ml of **STC**.
17) Repellet the cells and resuspend the pellet in 2 ml of **STC**.
18) Add the cryoprotectant, dimethylsulfoxide (DMSO), to a final 7%, v/v concentration, mix and dispense the cells in 100 µl aliquots for storage at -80°C.

SPE: 1M sorbitol, 0.1M Na Phosphate (7.5), 10mM EDTA
STC: 1M sorbitol, 10mM Tris-HCl (7.5), 10mM $CaCl_2$

A simplified version of this protocol is used for spheroplasting in preparation for biochemical extractions. Cells are simply resuspended in SPE with lyticase, incubated at 37°C for several hours, gently pelleted and then lysed.

2.6.2 Transformation of *S. cerevisiae* for YAC library construction

1) Pour and dry a 22 cm^2 supplemented sorbitol selection (SS) bottom agar plate, keep at 30°C in order to preheat before starting the transformation.
2) Melt SS top agar, cool to 50°C, add the supplement and keep the top agar in a 50°C water bath.
3) Add a 100 µl aliquot of spheroplasted (YPH 252) cells to 10 ml test tube, add DNA for transformation and incubate at room temp. for 10 min.
4) Add 1ml of **PEG solution** and incubate at room temp. for a further 10 min.
5) Gently pellet the cells by centrifugation (280 g, 4 min, at room temp.).
6) Gently resuspend the cells in 200 µl of **SOS**, incubate at 30°C for 30 min.
7) Add the cells to the top agar, swirl to mix and quickly pour onto a preheated SS library plate. Let this set and then incubate the plate at 30°C for 5 to 10 days.

PEG solution: 20% PEG 8000, 10mM Tris-HCl (7.5), 10mM $CaCl_2$
SOS: 1M sorbitol, 6.5mM $CaCl_2$, 0.25% yeast extract, 0.5% bactopeptone

2.6.3 YAC library amplification

Individually patching colonies into grids is extremely laborious, so after a representative sample of the library has been gridded, the remainder of the library can be amplified safely. After 10 days growth at 30°C the top agar has dried sufficiently for this layer, containing all the recombinants, to be peeled off the plate by rolling the top agar into a cylindrical shape. Place this rolled agar into a sterile 50 ml bolus and force the agar through the narrow opening into a sterile flask. Add YCD broth to the flask and incubate at 30°C with agitation for about 3 hours. Collect the 'supernatant', bring it to 7% with DMSO and store 1 ml aliquots at -80°C. Obviously, one does not amplify the YAC library in this way until all the grids desired are already patched and stored.

2.6.4 Storage of primary YAC library transformants

Grids of primary transformants established by picking colonies out of the top agar and patching them onto YCD plates are used as templates for colony lifts for storage and screening purposes. Colony lifts are taken onto Hybond N

filters which are then placed colony side up on YCD/20% glycerol plates and incubated for 2 days at 30°C. This allows the colonies on the Hybond N filters to grow and become impregnated with the cryoprotectant glycerol. The filters are then lifted off the plates and sandwiched between sterile circles of 3M Whatman paper. The filters are stacked, wrapped in aluminium foil and stored at -80°C. These filters provide a long term template for establishing gridded plates of the YAC library for future screening purposes.

2.6.5 Colony hybridisation screening of YAC libraries

The key to yeast colony hybridisation protocols is successful colony lysis. A short growth period (overnight) is important to obtain good lysis since spheroplasting of stationary phase cells is difficult.

1) Place the colony lift filters, colony side up, into petri dishes with a backing of 3M Whatman papers wetted with SPE/10mM DTT/100 units of lyticase.
2) Seal the petri dish with parafilm and incubate the plate overnight at 37°C.
3) Incubate the plate with filter at -80°C for 10 min to lyse the cells.
4) The filters are now cycled through the following processing steps on wetted Whatman papers:
5 min on 10% SDS, 10 min on 0.5M NaOH, 5 min on 2X SSC, 5 min on 2X SSC, 5 min on 2X SSC.
5) Dry the filters briefly on a dry Whatman paper and fix the DNA to the filters (60 mJ UV; 80°C baking for 2 h).
6) It is definitely worthwhile to test the results of this protocol by undertaking a control hybridisation with pBR322 sequences. This gives the investigator an indication of the signal intensity that should be expected for a genuinely positive result.

2.6.6 Transcriptional run-on analysis

Nuclei are prepared by a Stansted cell disrupter apparatus and stored, as described in Kooter and Borst (1984) and Kooter *et al.* (1987). The transcriptional run-on is essentially a radioactive transcriptional elongation reaction followed by the extraction of RNA and its use in a subsequent hybridisation onto cloned sequences. The extraction is performed with Trizol from BRL. This is a variation on the theme of a single step RNA extraction described by Chomcynski and Sacchi (1986).

1) Following disruption of trypanosomes, pellet an aliquot of nuclei (10^9) by centrifugation in a microfuge for 4 min. If the preparation is from metacyclic-

derived trypanosomes, fewer nuclei (10^8) are available. If α-amanitin is being used in the experiment, preincubate the nuclei, on ice, with the required amount of the drug for 10 min.
2) Resuspend the pellet in 50 µl of 2X **elongation buffer** (with α-amanitin if required), add 2 µl RNasein (Promega), 5 µl of creatine kinase (20 mg. ml^{-1}), 30 µl of ^{32}P-labelled UTP (3000 Ci. $mmol^{-1}$) and bring the reaction volume up to 70 µl with H_2O.
3) Incubate the reaction tube at 37°C for 5 min (27°C; 5 min for procyclic nuclei).
4) Add 800 µl of Trizol (BRL), and incubate at room temperature for 5 min.
5) Add 160 µl of chloroform, agitate vigourously for 15 s (place eppendorf tube inside a plastic bijou bottle, which is in turn placed inside a lead 'pig' container).
6) Incubate at room temperature for 3 min.
7) Centrifuge the reaction for 15 min at 4°C (12 000 g).
8) Extract the aqueous phase and add this to a microfuge tube containing 400 µl isopropanol with 8 µg of *E. coli* tRNA.
9) Incubate the precipitation tube at room temp. for 10 min.
10) Precipitate the RNA by centrifugation (10 min; 4°C; 12 000 g).
11) Wash the pellet with 70% EtOH.
12) Dry the pellet and resuspend it in 50 µl H_2O.
13) 'Denature' the probe (80°C for 10 min), add to prewarmed **hybridisation solution**, add this solution to the experimental filter and hybridise at 55°C for 48 h.
14) Wash and expose the filters as required.

Safety tips: 1) All manipulations must be conducted in leak-proof screwtop microfuge tubes.
2) When α-amanitin is used, all materials that come into contact with the drug, especially the powdered form, should be decontaminated by submersion in strong alkali (1M NaOH).

2X elongation buffer: 100mM Tris-HCl (8.0), 50mM NaCl, 100mM KCl, 2mM $MgCl_2$, 4mM $MnCl_2$, 2mM DTT, 0.15mM spermine, 0.5mM spermidine, 10mM creatine phosphate, 2mM GTP, 2mM CTP, 2mM ATP, 25% glycerol.
hybridisation buffer: 3X SSC, 0.1% SDS, 10X Denhardts' solution, 20mM Na phosphate, 100 µg. ml^{-1} sheared and denatured herring sperm DNA, 50 µg. ml^{-1} *E. coli* tRNA.

2.6.7 Slot blotting

Plasmid substrates were slot blotted for one of the transcriptional run-on experiments.

1) Cut transfer membrane and Whatmans to fit the slot blotting apparatus. Presoak the transfer membrane (Biodyne) in 1X TAE for 30 min.
2) Aliquot 5 µg of the plasmid DNA smples and bring the volume up to 120 µl with H_2O.
3) Add 40 µl of 1M NaOH, vortex and incubate at room temp. for 10 min to denature the plasmid.
4) Add 160 µl of 10X TAE to the denatured plasmid samples, vortex, spin down (12 000 g) for 2 min and then incubate the microfuge tubes on ice.
5) Assemble the blotting apparatus with two presoaked Whatman papers backing the transfer membrane.
6) Load the plasmid samples, apply the vacuum and let samples load completely.
7) Add a drop of 5X TAE to each sample well and filter under vacuum pressure.
8) Disassemble the blotting apparatus and briefly air dry the transfer membrane.
9) Fix the plasmid DNA to the membrane (60 mJ UV; 2 h baking at 80°C).

2.6.8 **Gelase treatment** (Cambio)

This is an enzymatic method for depolymerizing agarose. This method is employed whenever YAC DNA is isolated directly from a preparative gel for use in cloning or labelling experiments. After the depolymerization reaction the DNA is selectively precipitated in an ammonium acetate/EtOH precipitation. Although the resultant pellets must be handled with extreme care to prevent their loss, this protocol has the benefit of not precipitating any appreciable amount of agarose. The products of this precipitation are adequate for enzyme manipulations, obviating the need for organic extraction.

2.7 Referenced kit protocols

2.7.1 Qiagen plasmid preparations

The Qiagen plasmid preparation system is a column purification system. As is usual with column purifications, the column is composed of a DNA binding resin. When the lysate is applied to the column, degraded RNA, proteins and metabolites pass through the column to the discarded eluate. Contaminants bound to the column are washed away with 1M NaCl. When the DNA is eluted with 1.25 M NaCl containing buffer, the DNA is then desalted and concentrated by isopropanol precipitation followed by an ethanol wash.

2.7.2 **Gelase** (Cambio)

This is an enzyme that depolymerizes low melting point agarose. Once the agarose is depolymerized the DNA is precipitated under conditions that do not precipitate the agarose as well (ammonium acetate/ethanol precipitation).

2.7.3 **Riboprobe synthesis kit** (Stratagene)

T3 and T7 RNA polymerases drive the *in vitro* production of riboprobes from small fragments generated by cleavage with enzymes with a 4 bp recognition moiety. Probes are purified by Stratagene's NucTrap purification columns.

2.7.4 UBI Sequenase kit

The kit employed used was Sequenase version 2.0. This is a modified T7 DNA polymerase from which the 3' to 5' exonuclease activity has been removed. This kit provides GTP in the modified deaza-GTP form, to reduce the incidence of GC-compressions. The double stranded plasmid substrates used were denatured for sequencing by the glycol/heat method outlined in the manufacturer's protocol. This necessitated the use of a glycerol tolerant gel, run in TTE electrophoresis buffer (with taurine rather than boric acid).

Chapter 3

Cloning of *T. brucei rhodesiense* sequences utilising Yeast Artificial Chromosomes

3.1 Introduction: Theoretical Considerations

When molecular biologists decided to tackle the question of what components of a eukaryotic chromosome are required for stable mitotic segregation, they studied an organism that was amenable to exhaustive biochemical and genetic analysis; *Saccharomyces cerevisiae*. With the advent of transfection in budding yeast and the consequent merging of molecular cloning and genetical techniques, identification of *cis* elements important for replication and stable segregation of segments of DNA was accomplished in the early 1980s. ARS (Autonomously Replicating Sequence) elements allow replication of colinear DNA in *cis* and much evidence suggests that they are genuine chromosomal origins of replication (Stinchcomb *et al.*, 1979; Zakian and Scott, 1982; Fangman *et al.*, 1983; Celnicker and Campbell, 1982). Centromeres ensure the stable segregation of the chromosomes in which they reside. Centromeric DNA in budding yeast has an 'element II' which is an 82-89bp region of very AT-rich (>90%) DNA, in addition to a highly conserved 11bp sequence ('element III'). Direct evidence of their role in centromeric function was obtained for one such element, as deletion of elements II and III from chromosome III produced an exceedingly unstable chromosome (Clarke and Carbon, 1983). These sequences were cloned and studied in circular episomes; *in vivo* they are part of linear molecules. One further sequence element is required for the replication and maintenance of linear chromosomes. Due to the discontinuous nature of lagging strand DNA synthesis, linear DNA molecules must have a special mechanism for replicating the sequences at their ends. Blackburn's group cloned and characterised these 'telomeric elements' in budding yeast (Shampay *et al.*, 1984; Szostak and Blackburn, 1982). This completes the summary of *cis* elements required for replication and stable segregation of linear yeast chromosomes.

In a pair of landmark papers Hieter *et al.*(1985) and Koshland *et al.*(1985) independently described colony colour assays in *S. cerevisiae* that could be used to measure chromosomal stability. Yeast adenine mutants (*ade 2*) are defective in 5-aminoimidazole ribonucleotide carboxylase and are red due to the accumulation of phosphoribosylaminoimidazole. Wild type cells produce colonies that are white (Roman, 1957). The former group cloned an ochre-suppressing form of a $tRNA^{TYR}$ gene, *SUP11*, that confers a dose dependent suppression of the colour phenotype in homozygous diploid *ochre ade2* mutant strains. Absence of *SUP11* gives the full mutant red colony colour, one copy of *SUP11* yields pink colonies and two or more copies completely suppress the colour phenotype, resulting in white colonies. This therefore, in diploid strains, provides a means of estimating gene copy number.

A simple quantitative assay of mitotic chromosome stability from first division events is easily derived from transfection of DNA molecules with a single copy of the *SUP11* gene into *ochre ade2* mutant strains. The prevalence of red sectoring in the resultant colonies reflects the frequency of mitotic chromosome loss. The number of half-sectored colonies over the total number of colonies derived from cells carrying a single copy of the SUP11 containing DNA, gives the frequency of mitotic chromosome loss. Most significantly, because of the dose dependence of suppression, 1:0 segregation can be distinguished from 2:0 segregation; the former gives pink/red colonies, while the latter yields white/red sectoring. The Hieter *et al.*paper demonstrated that linear molecules with telomeres, an ARS element and a centromere exhibited dramatically enhanced mitotic stability as the size of the linear molecule was increased. This stabilising effect, 3 loss events per 20, 000 divisions in a 137 kb molecule as opposed to 11 loss events per 100 divisions for a 15 kb molecule, is not approaching an asymptote at the largest size range tested (137 kb). Aside from the elegance of these genetic experiments, this work established the conceptual foundation for a new generation of cloning vectors: YACs (Yeast Artificial Chromosomes).

In molecular genetics there are two kinds of map; genetic maps, derived from cytogenetic and linkage data, and physical maps, obtained by analysis performed with recombinant DNA. Prior to the late 1980s the difference in resolution between these maps was greater than 100-fold, the former being limited, in complex genomes, to a resolution of megabase-pair regions of DNA and the latter to segments of recombinant DNA measuring only tens of kilobase-pairs (Burke *et al.*, 1987). This important gap in resolution between the genetic and physical maps was bridged by Burke, working in the lab of Olson, who described 'the cloning of large segments of exogenous DNA into yeast by means of Artificial Chromosome vectors'. These vectors allow much larger fragments of DNA to be cloned (>50 kb) and provide enhanced stability of these sequences over cloning systems that rely upon *E. coli* for their propagation and amplification (Schlessinger 1990).

In my view YACs are essentially an improved form of somatic cell hybrid-type technology. Somatic cell hybrid technologies suffer from significant drawbacks: they must be generated *in vivo* by cell fusion methods and the analysis of the target chromosome or chromosome fragment must be conducted against a complex genetic background. YACs represent a vast improvement: the DNA is 'cloned' *in vitro* in more manageable sizes and is introduced into an organism with a simple genome. The replicon/host relationship between the YAC and *S. cerevisiae* qualifies this technology as a cloning system. However, the large insert size causes their analysis to be fraught with some of the same difficulties encountered in somatic cell hybrid technologies.

The combination of an autonomously replicating unit and a host with a relatively simple genetic background was the major rationalisation for Burke *et al.* (1987) choosing a linear YAC replicon and a *S. cerevisiae* host. The YAC vectors were designed to include all their functional units on a single plasmid. They are chimaeric vectors, having the elements required for their propagation as plasmids in *E. coli* as well as having most of the requirements of linear *S. cerevisiae* chromosomes (see **Fig. 3.1**). Plasmid propagation in *E. coli* allows large amounts of the vector to be purified for *in vitro* manipulation and is accomplished by the incorporation into the YAC vector of a bacterial origin of replication, oriC, and the ß-lactamase gene that confers ampicillin resistance. The yeast sequences present on the vector split into two functional classes: those required for the maintenance of yeast chromosomes and those required for the efficient genetic selection of transformants with the desired genotype.

In the first functional class are the telomeres, centromere and ARS element. The only missing feature is sufficient molecular size for mitotic stability. This is provided by the large fragments of exogenous DNA from the organism under study, in this case *T. brucei*, that are prepared for the *in vitro* construction of the YACs. The telomere repeats are furnished by fragments containing *Tetrahymena* telomere repeats, that can be healed *in vivo* in yeast into functional telomeres. The kinetochore (centromere) and chromosomal replicator functions are provided by CEN3 and ARS1 (Clarke and Carbon, 1980; Stinchcomb *et al.*, 1979).

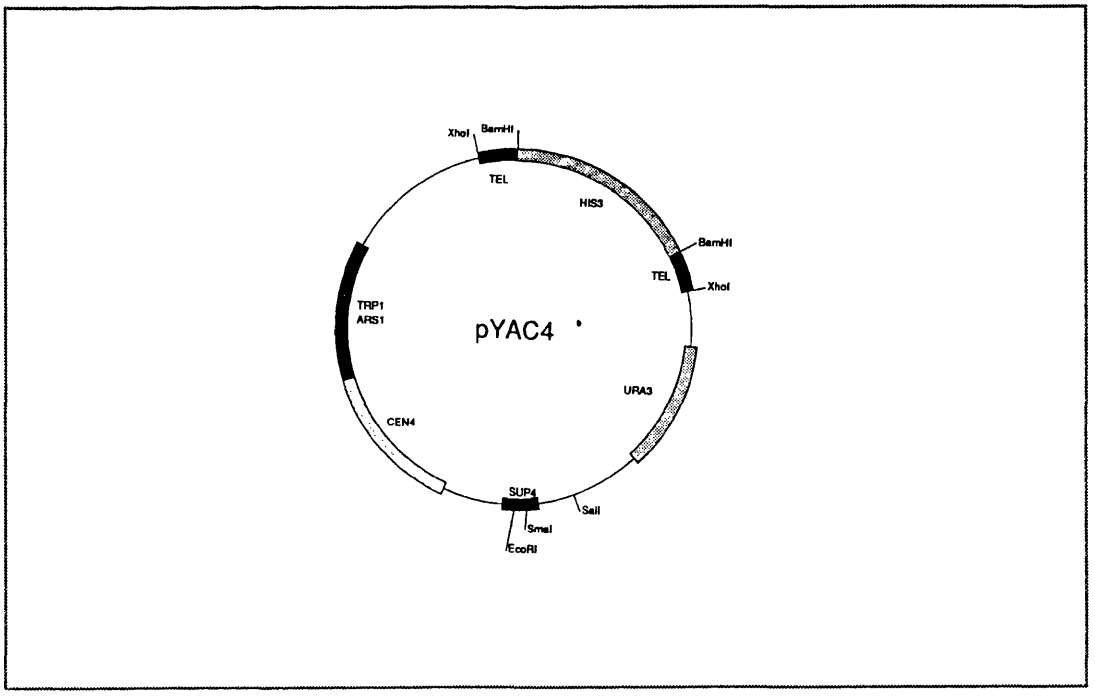


Figure 3.1 A schematic representation of the vector pYAC4. Boxed regions of the vector represent yeast sequences while the thin line represents the pBR322-derived segments of the vector. The plasmid sequences that allow for plasmid propagation in *E. coli* lie between TRP1 and one of the sets of telomere repeats. *Eco* RI and *Sma* I are the available cloning sites in the SUP4 gene of the vector. Taken from Riley *et al.* (1990).

The second functional class of yeast sequences present on the YAC vector system allow for genetic selection of cells carrying *bona fide* clones. The first requirement is for a visible marker that indicates the presence of an insert. This is achieved in pYAC4 by the presence of the *Eco* RI cloning site in the small intron of the yeast $SUP4_o$- $tRNA^{TYR}$ gene. Disruption of this gene results in the loss of suppression of any *ochre ade2* alleles that may be present in the yeast host cell and thereby produces a red colony colour. Hence, if the YAC has an insert the colony colour will be red. Selection of successful transformation events in *S. cerevisiae* by a YAC requires the presence of both vector arms, as each arm carries a gene to complement a different auxotrophic deficiency of the host. *TRP1* on the left arm complements *trp1* mutants and *URA3* on the right arm complements *ura3* mutants. This double complementation selection allows only cells transformed with structurally intact YACs to grow in the specified medium used to plate transformants. Yeast strains suitable for YAC library construction must have *trp1*, *ura3* and *ochre ade2* genotypes.

The final sequence element present in the YAC vector is the *S. cerevisiae HIS3* gene contained on a *Bam* HI fragment that serves as a stuffer fragment between the *Tetrahymena* telomeres and renders the vector circular. This stuffer is released by *Bam* HI digestion to linearise the vector in preparation for the *in vitro* construction of the YAC and to expose the *Tetrahymena* telomere sequences (Burke *et al.*, 1987; Shampay *et al.*, 1984).

YAC cloning methodology also allows direct cloning of telomeric sequences from other organisms. This variation of YAC cloning capitalises on the observation of Blackburn and Szostak that telomeric sequences from another organism (*i.e. Tetrahymena thermophila*) can 'seed' the addition of yeast telomeres (Shampay *et al.*, 1984). Such seeding occurs because *S. cerevisiae*'s telomerase activity is promiscuous in its substrate requirements. The essential modification of the YAC vector for cloning telomeres involves the removal of one of the telomere ends, thus rendering the vector a 'half-YAC'. When a 'half-YAC' vector is linearised and YAC clones are constructed *in vitro*, only those YACs that have acquired a fragment containing a telomere from the target organism (*T. brucei*) can stabilise the half-YAC and allow this to complement the auxotrophic deficiencies of the host strain.

The 'half-YAC' vector system discussed here is that developed as pYAC4-*Neo-Not* I by Cross and colleagues (1989) to clone human telomeres by complementation in yeast. **Figure 3.2** shows the pYAC4-*Neo- Not* I vector system in schematic form. All the features present in pYAC4 are also present here with three significant alterations. Most important is the absence of one of the *Tetrahymena* telomere repeats. When pYAC4-*Neo-Not* I is linearised it only has one healable end and must acquire the other from the digested *T. brucei* DNA. A

significant drawback of this vector is the loss of the visible colour phenotypic marker that indicates that the YAC has accepted an 'insert'. This is due to the cloning site no longer being in the *SUP4* gene as it is in the parent pYAC4 vector. There is also a neomycin resistance cassette in the 'half-YAC' vector but this is not germane to the cloning chemistry.

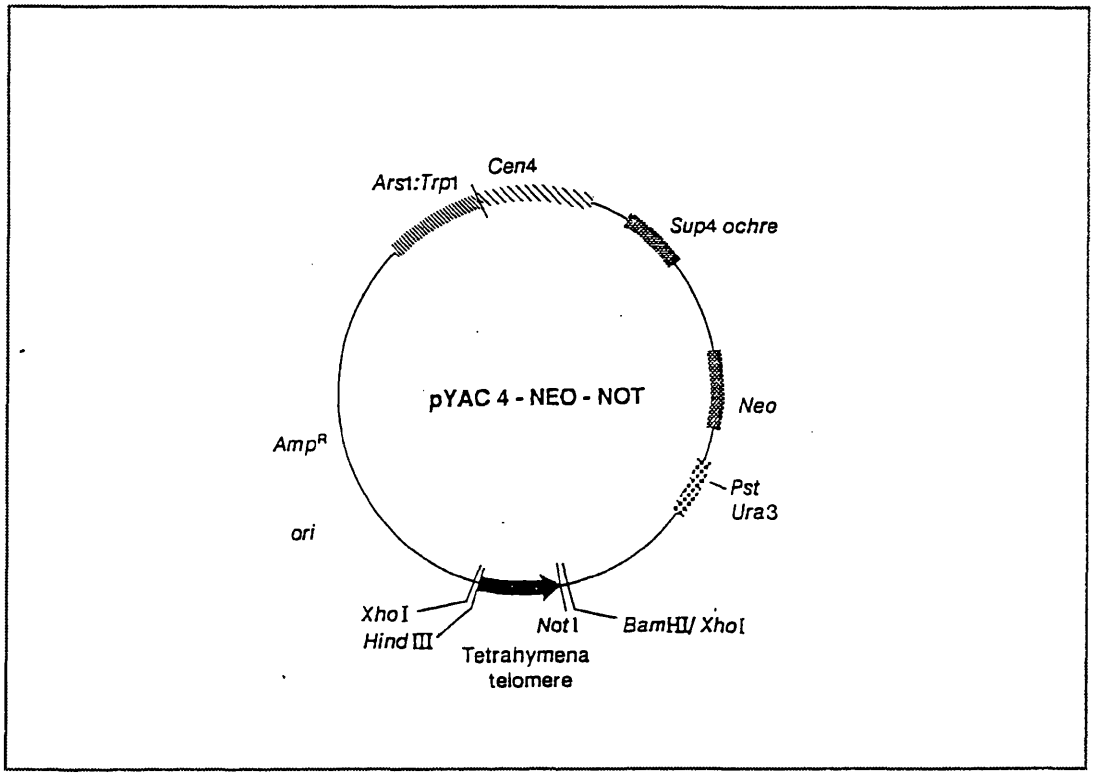


Figure 3.2 A schematic representation of the vector pYAC4-*Neo-Not* I. Taken from Crosset *al.* (1989).

3.2 Introduction: Practical Considerations

As experience with YAC cloning has accrued, limitations of this technology have been uncovered. Structural instability of YACs can and does occur. Cells carrying more than one YAC are relatively common as is the more distressing phenomenon of chimaerism. Documentation of mostly anecdotal reports suggests that YAC libraries made from higher eukaryotes have as much as 42% of their clones afflicted by one of these problems (Schlessinger 1990). About 10% of the clones in a library will have more than one YAC. Additionally, 10-30% of the clones will be chimaeric due to the co-cloning of disparate fragments or because of recombination between YACs in cells carrying more than one YAC. Finally, 2% of the clones will be demonstrably unstable, with multiple deletion derivatives being present in a single cell, presumably generated by recombination carried out by the host. As yet, completely recombination deficient host strains have not been developed (Schlessinger 1990; Vilageliu and Tyler-Smith 1992).

Retransformation studies have identified another problem: instability induced by the transformation protocols. Apparently stable YACs show a high frequency of rearrangement, 1-2%, usually by deletion, when they are retransformed into yeast (Albertsen *et al.*, 1990; Neil *et al.*, 1990; Connelly *et al.*, 1991; Vilageliu and Tyler-Smith 1992).

> "These rearrangements must occur before the establishment of the YAC in its new host, and may be due to the process of introducing a naked DNA molecule into the cell via the cell membrane and cytoplasm; repair or recombination pathways may be stimulated."
> (Vilageliu and Tyler-Smith 1992)

The 'half-YAC' telomere cloning protocols have their own problems in addition to those previously outlined for YAC cloning in general. The loss of the colour selection available in pYAC4 is particularly problematical, as even a dephosphorylated YAC vector can recircularise inside the yeast cell and act as a circular minichromosome (Dobson and Brown 1992). In human DNA, and presumably in *T. brucei* DNA as well, there exist non-telomeric sequences that can fortuitously prime the addition of functional yeast telomeres (Dobson and Brown 1992). Lastly, if the telomeric fragment from the donor organism is derived from an acrocentric or telocentric chromosome the ligated DNA may contribute to the formation of a functionally dicentric chromosome. Dicentric chromosomes are subject to breakage, deleting one or other of the two centromeres. These breakage events, if they occur in cells with multiple pieces of exogenous DNA, may elevate the occurrence of co-cloning artifacts (Dobson and Brown 1992).

The work in this chapter will be discussed with these considerations in mind.

3.3 A Rationale

As has been stated earlier in this thesis, one of the major objectives of the project was to clone and characterise the 'metacyclic domain' of the 1.22 M-VSG telomere. Previous cloning experiences with metacyclic VSG telomeres, utilising λ phage, have been encumbered by difficulties. Scrambled clones, demonstrably unstable clones and severe underrepresentation of metacyclic VSG telomeric sequences, have all been encountered (details can be found in the PhD. theses of P.G. Shiels 1990 and K.R. Matthews 1990). A current chromosome walk upstream of the 1.61 M-VSG gene has suggested that the silent area (*i.e.* the metacyclic domain) may extend as far as 30kb upstream of the chromosome end (Vincent S. Graham personal communication). As a minority of this metacyclic domain is present elsewhere in the genome (*i.e.* there are multicopy

sequences present in the 1.61 M-VSG domain) there is a real danger of 'walking' onto the wrong chromosome. This combination of prior cloning difficulties and the prospect of a long and labour intensive chromosome walk prompted us to choose YAC cloning as the first step in the assembly of a contig of λ clones spanning the metacyclic domain of the 1.22 locus. This was the first attempt, that we know of, to utilise YAC technology in *T. brucei*. Consequently, there were concerns about representation and YAC stability that needed to be considered when working with 'virgin' libraries.

In principle, the cloning of the 1.22 domain in YACs, followed by the subsequent assembly of a λ contig has several advantages. The assembly of a λ contig is expedited by the nature of the YAC vector and by the isolation, provided by YAC cloning, of potentially multi-copy sequences from their native genetic background. Once a contig of suitable length is obtained, the λ clones comprising the contig can be characterised simultaneously. Herein, though lies the risk; the fidelity of the initial YAC cloning step cannot be confirmed with confidence until a great deal of effort has been expended obtaining and characterising the λ clones.

3.4 Determination of the size of *T. brucei* genomic DNA to be used in YAC cloning protocols

The construction of YAC libraries requires the preparation of large inserts. In *T. brucei* there are approximately 100 minichromosomes that range in size from 50-150 kb and are not a repository of genes, being composed primarily of a satellite DNA sequence (Borst *et al.*, 1983). It was desired to have insert sizes larger than the range occupied by the *T. brucei* minichromosomes to minimise their contribution to the library and therefore to improve the overall genomic representation of housekeeping genes.

Genomic DNA prepared by O. Shonekan utilising a standard liquid handling method described in **Section 2.5.1** was examined by Pulsed Field Gel Electrophoresis (PFGE). **Figure 3.3** shows this analysis conducted on EATRO 795 (clone 1.22j') DNA. The DNA appears to be of very high molecular weight (>50 kb), with the minichromosomes presumably contributing to the more intense fluorescence observed in the 50-150 kb size range. It was decided that this genomic DNA sample was adequate for the generation of fragments 90 kb or larger.

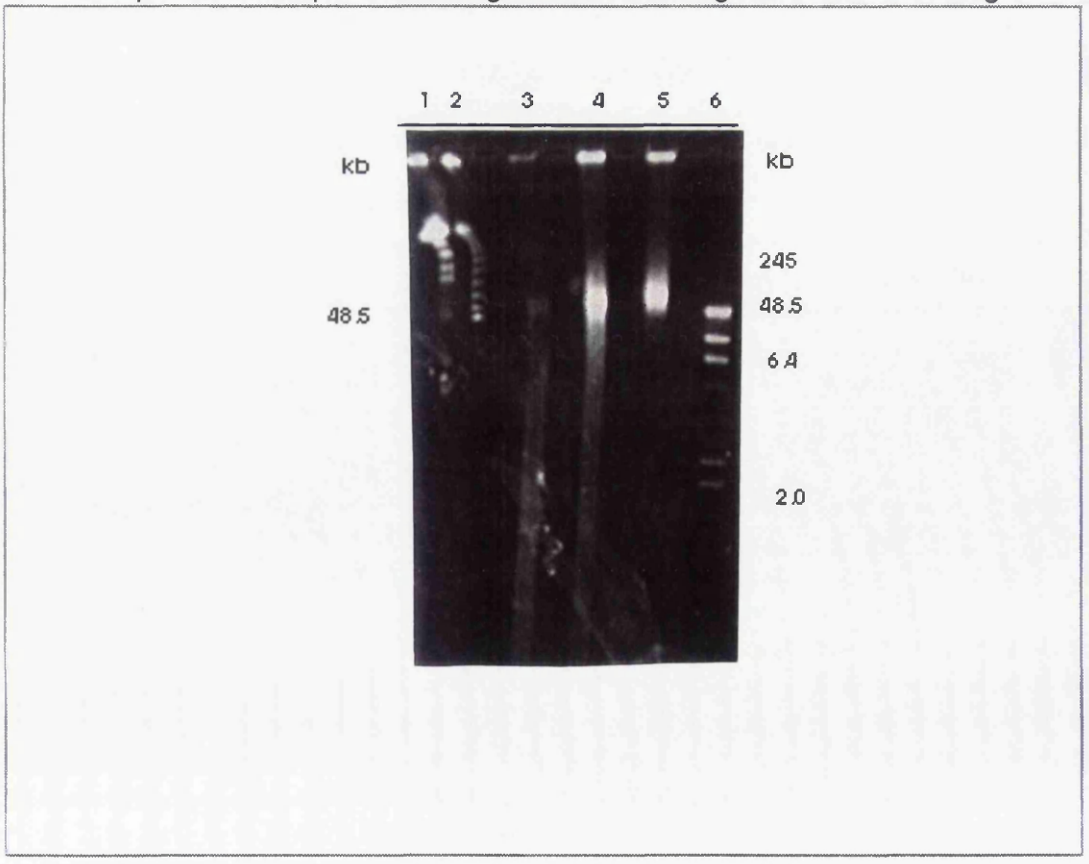


Figure 3.3 Determination of the size of EATRO 795 DNA prepared by liquid handling. Lane 1, *S. cerevisiae* chromosomes(245, 290, 370, 460kb are resolved); lane 2, multimers of λ(48.5, 97, 145.5, 194kb etc. resolved up to 436.5kb); lanes 3 and 4, EATRO 795 DNA partially digested with 2.5 units of *Sau* 3AI with different loadings; lane 5, uncut EATRO 795 DNA; lane 6, I *Hin* dIII fragments. The gel is1% agarose in1X TB0.1E running buffer run on a Biometra OFAGE system. Interval: 30-3sec linear ramp; Angle: 120-95° linear ramp; Voltage: 10V/cm; Temp.: 13°C; Duration: 11 hrs.

Figures 3.4 and **3.6** show the results of PFGE separation of *Sau* 3AI and *Eco* RI partial digests of this EATRO 795 sample. The *Sau* 3AI partial digestion experiment, **Fig. 3.4**, was conducted to determine the appropriate digestion conditions for the preparation of 'inserts' for the pYAC4-*Neo-Not* I 'half-YAC' vector and hence for the construction of a telomere library. The *Eco* RI partial digestion experiment documented in **Fig. 3.6** was conducted to determine the appropriate digestion conditions for the preparation of inserts for the pYAC4 vector and hence for the construction of a conventional YAC library.

3.5 Nucleic Acid Chemistry for the construction of a Telomere Complementation library using the pYAC4-*Neo-Not* I vector system

In the preparation of the genomic DNA, *Sau* 3AI partial digestion was chosen in preference to *Bam* HI partial digestion and also to complete *Not* I digestion. I reasoned that a restriction enzyme with a 4 bp recognition moiety would generate fragments of more uniform size than the 6 bp cutter *Bam* HI or the 8 bp cutter *Not* I both of which could suffer from more dramatic manifestations of any sequence bias in the *T. brucei* genome than *Sau* 3AI.

pYAC4-*Neo-Not* I was prepared utilising the Qiagen 'midiprep' system as described in the manufacturer's protocol. 100 µg of the vector was linearised by complete restriction with *Not* I (100 units, 1 h at 37°C). This reaction was then cooled to 0°C and a sample was analysed by agarose gel electrophoresis. *Not* I was removed by a single phenol/chloroform (organic) extraction followed by an ethanol precipitation. This *Not* I linearised pYAC4-*Neo-Not* I DNA was dissolved in 180 µl of 10mM Tris-Cl (pH 8.0). Treatment (1 h, 37°C) with 1.5 units of Alkaline Phosphatase (Boehringer Mannheim) was utilised to dephosphorylate the linearised vector. This enzyme was inactivated by heating the reaction tube to 68°C for 10 minutes. Test ligations analysed by agarose gel electrophoresis revealed the dephosphorylation to have been satisfactory (data not shown). The vector preparation was completed by total digestion with *Bam* HI (100 units; 37°C for 3 h). *Bam* HI was removed by an organic extraction followed by an ethanol precipitation. The vector was now linear, dephosphorylated at the *Not* I site and ready to be ligated to fragments of *T. brucei* genomic DNA generated by *Sau* 3AI partial digestion.

The experiment documented in **Fig. 3.4** was performed to determine partial digestion conditions for the restriction enzyme *Sau* 3AI on EATRO 795 genomic DNA. 0.00075 units of *Sau* 3AI/µg of EATRO 795 (1.22j') genomic DNA was used in a one hour incubation at 37°C in a scaled up reaction performed on 30µg of genomic DNA, in a total volume of 100 µl. *Sau* 3AI was then heat inactivated by incubation at 65°C for 10 minutes.

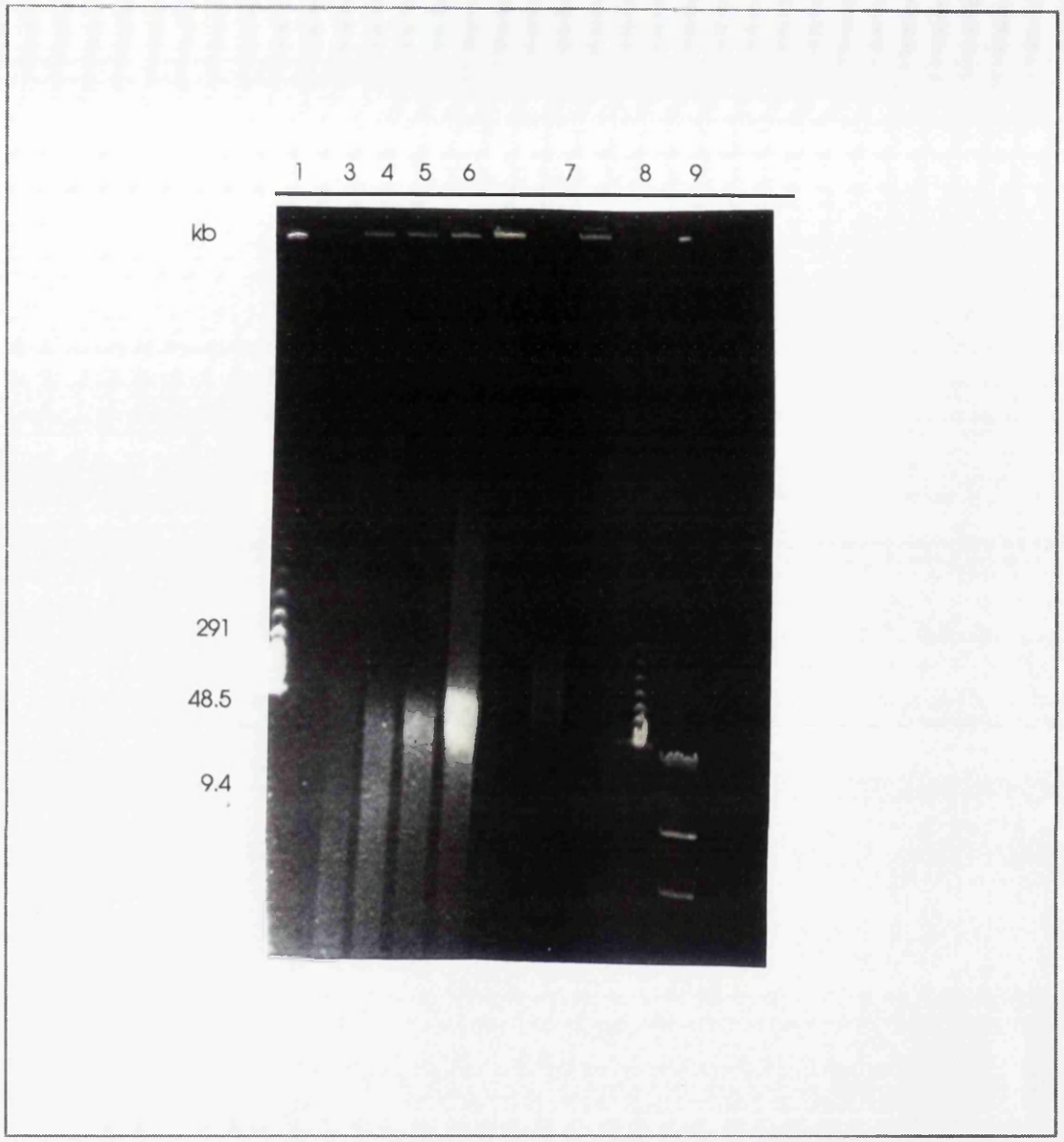


Figure 3.4 Pilot experiment to determine *Sau* 3AI partial digestion conditions for telomere library construction. This is an enlargement of the most informative lanes of the PFGE separation of the *Sau* 3AI treated EATRO 795 DNA. Lane 1, multimers of λ resolved up to 436.5kb; lane 2, 0.012 units *Sau* 3AI/μg DNA; lane 3, 0.006 units of *Sau* 3AI/μg DNA; lane4, 0.003 units *Sau* 3AI/μg DNA; lane 5, 0.0015 units *Sau* 3AI/μg DNA; lane 6, 0.0008 units of *Sau* 3AI/μg DNA; lane 7, uncut EATRO 795 DNA; lane 8, multimers of λ; lane 9, λ *Hin* dIII fragments. The PFGE parameters are the same as those employed in Fig. 3.3.

Excess amounts of vector and genomic DNA were prepared for ligation, as the transformation efficiency was expected to be low. The reported rate using pYAC4 and spheroplasted AB1380 cells ranges from 300 to 1000 transformants/ μg of ligation mix (Riley *et al.*, 1992). A large scale ligation reaction was carried out with 95 μg of treated vector and 24 μg of EATRO 795 *Sau* 3AI fragments at a final concentration of 0.68 mg. ml^{-1} at 4°C overnight. This provided a 23-fold molar excess of ligatable vector ends to insert, a concentration that favours the addition of vector arms to the large genomic fragments.

The *S. cerevisiae* strain YPH 252 was spheroplasted as described in **Section 2.6.1**. The efficacy of the spheroplasting was assessed by lysis of subsamples and by vector transformation experiments. These spheroplasts were deemed suitable and were used in an attempt to generate a telomere library. The transformation was carried out as reported in **Section 2.6.2**, with the inclusion of an undigested pYAC4-*Neo-Not* I positive control plate and a mock transformation negative control plate. 7 µg of the ligation mix and an equal amount of herring sperm DNA were pooled and used in the transformation. The herring sperm DNA is used to increase the viscosity of the transformation mix; this apparently increases the transformation efficiency (J. Ajioka, personal communication). The positive control gave a transformation efficiency of 7200 colonies/µg, while the experimental transformation yielded only 324 colonies/µg, producing 2300 colonies in total. The colonies were grown up embedded in agar, a requirement for the generation of colonies from spheroplasted cells, and took 10 days to reach a reasonable size. The colonies were white, as expected from the lack of colour selection, and there was great variation in the size of the colonies, with the largest being approximately three times the diameter of the smallest. Due to the absence of the visible selection system from this vector, it was impossible at this point to say whether recombinant clones had really been produced.

Thirteen colonies were randomly sampled and their karyotypes analysed. Separate and pooled clones were analysed by being grown in YCD broth for two days at 30°C and then being processed for chromosome-length DNA set in agarose blocks. Chromosomal karyotypes of two individual clones, two pools of 4 clones and one pool of 3 clones were obtained by PFGE separation and Southern blotting onto Zetaprobe GT nylon membrane (Biorad). pBluescript SK+ (Stratagene) was radiolabelled by random priming and used as a YAC-specific hybridisation probe, detecting the *E. coli* plasmid sequences that are not present in the yeast genome. Neither ethidium bromide staining of the karyotypes (**Fig. 3.5** A) nor hybridisation with pBluescript (**Fig. 3.5** B lanes 1-5) identifies anything in the karyotypes indicative of a Yeast Artificial Chromosome. The hybridisation signals present are weak and diffuse compared with the results of a similar analysis detailed in **Fig. 3.5** B lanes 6-8, the same hybridisation on the same filter to putative YAC clones constructed in pYAC4 . Additionally, these hybridisation signals are in a much higher molecular weight range than was anticipated. The auxotrophic complementation of the ura3 an trp1 genotype means that YAC sequences are present in these cells but the evidence in **Fig. 3.5** B lanes 1-5, suggests that these sequences are being contributed either by recircularised vector or by YACs that have recombined into the yeast genome. I suspected that these spurious transformants were more likely to be the result of religation of the vector than of ubiquitous recombination with the host; all thirteen samples represented were false clones.

In order to gain some insight into whether recircularisation of the vector *in vivo*

was giving rise to the spurious transformants, I endeavoured to repeat the ligations after removing all linearised and uncut vector from the ligation mix. 20μg of the ligation mix was loaded onto a 1X TAE Low Melting Point (LMP) agarose gel, the loading slot sealed and the gel run at 50V for 2 h to separate away any unligated vector. The slot material was excised, treated with Gelase (Cambio) (see **Section 2.6.8**) organically extracted and used as described before in a transformation experiment. In addition to the experimental set a positive control of uncut vector was included. No transformants were retrieved from the experimental set. This negative result could be due to several other technical factors not controlled for in this experiment but it lends credence to the notion that vector religation was the problem in the initial 'telomere library'. This matter was not pursued any further and the work on a telomere library was terminated due to promising developments in the other line of YAC cloning.

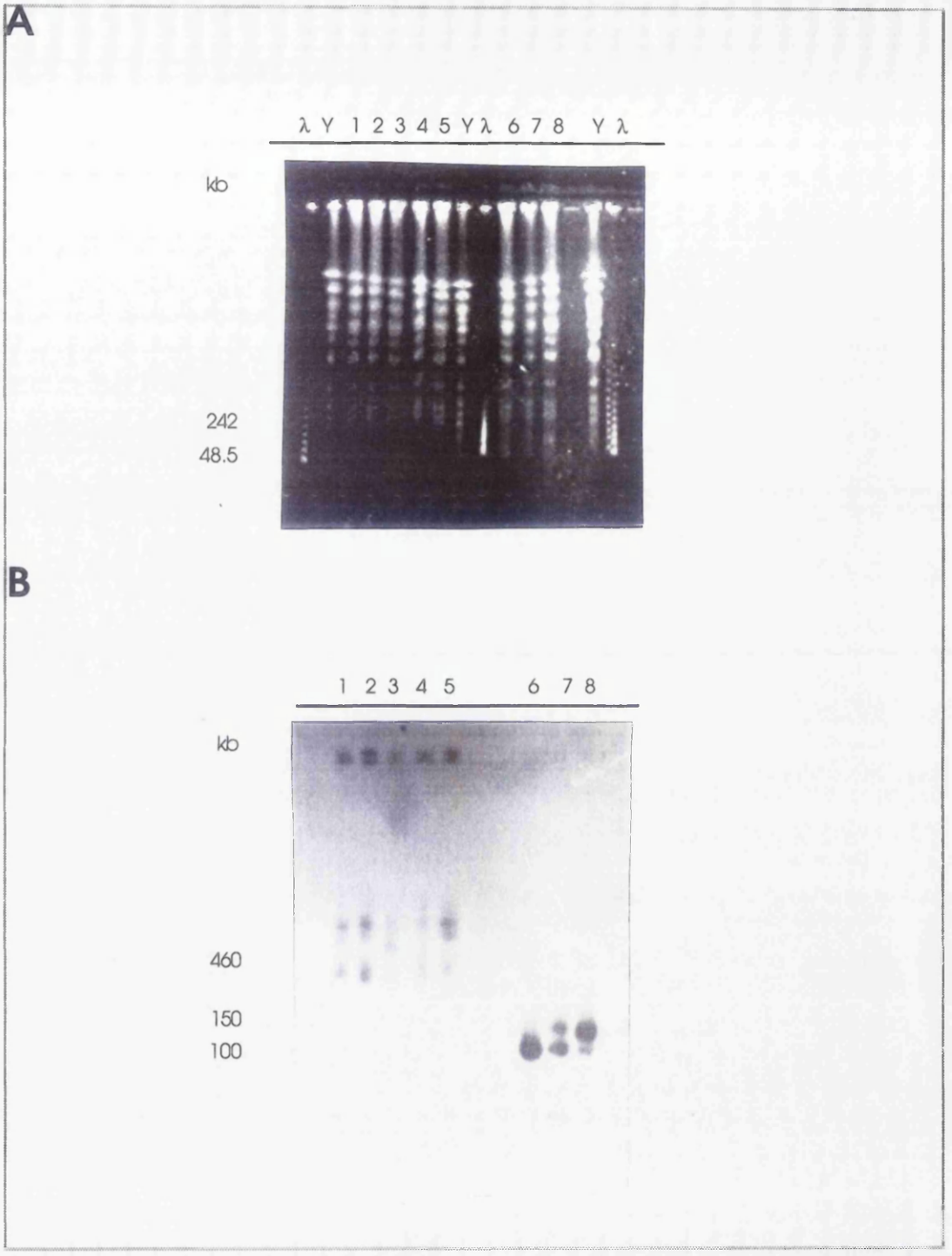


Figure 3.5 Karyotype analysis of randomly picked transformants from telomere and conventional YAC libraries. Panel A is an ethidium bromide stained gel, while Panel B is the southern blot of that gel probed with pBluescript at high stringency (h.s.). λ, multimers of λ; Y, *S. cerevisiae* chromosome standards; lane 1, a single transformant from the telomere library; lane 2, a single transformant from the telomere library; lane 3, four pooled transformants from the telomere library; lane 5, three pooled transformants from the telomere library; lanes 6, 7 and 8, pools of four transformants from the conventional YAC library. The gel is 1% agarose in1X TB0.1E running buffer separated on a Biometra OFAGE system. Interval: 100-20 sec linear ramp; Angle: 110-100° linear ramp; Voltage: 11V/cm; Temp.: 13°C; Duration: 24 hrs.

3.6 Construction of a *T. brucei rhodesiense* YAC library utilising the pYAC4 vector system

The experiment documented in **Fig. 3.6** was performed to determine partial digestion conditions for the restriction enzyme *Eco* RI on EATRO 795 (1.22j') genomic DNA. 0.025 units of *Eco* RI/μg of EATRO 795 genomic DNA was used in a scaled up reaction (37°C, 1 h) performed on 30μg of genomic DNA in an 100μl volume. *Eco* RI was partially inactivated by incubation at 65°C for 10 minutes. The expected residual restriction activity was removed by a single phenol/chloroform extraction followed by an ethanol precipitation.

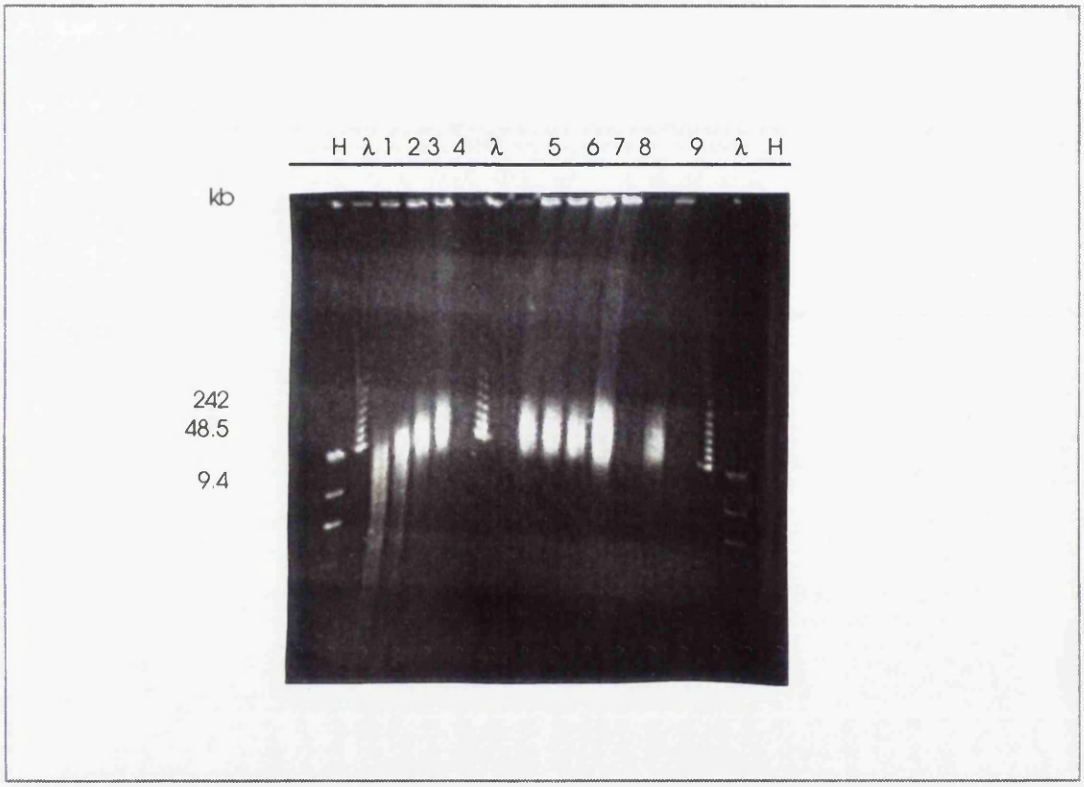


Figure 3.6 Pilot experiment to determine *Eco* RI partial digestion conditions for YAC library construction. This is an ethidium bromide stained PFGE separation of the *Eco* RI treated EATRO 795 DNA. H, *Hin* dIII digested λ; λ, multimers of λ; Lane 1, 0.4 units of *Eco* RI/μg DNA; lane 2, 0.2 units/μg; lane 3, 0.1 units/μg; lane 4, 0.05 units/μg; lane 5, 0.025 units/μg; lane6, 0.0125 units/μg; lane 7, 0.006 units/μg; lane 8, 0.0015 units/μg; lane 9, uncut EATRO 795 DNA. The PFGE parameters are the same as those employed in Fig. 3.3.

pYAC4 was purified, 'maxiprepped', by the CsCl method as outlined in Sambrook *et al.* (1989). 100μg of vector was linearised by complete digestion with *Bam* HI (100 units; 1 h, 37°C). A sample of this reaction was analysed by agarose gel electrophoresis to check the extent of digestion. The salt conditions were then adjusted and the vector was digested with *Eco* RI (100 units; 37°C, 3 h). The restriction enzyme activities were removed by a single organic extraction followed by an ethanol precipitation. Once again the enzyme digest was checked by electrophoretic analysis of a sample from the reaction tube. This *Bam* HI/*Eco* RI digested pYAC4 DNA was then dissolved in 180μl of 10mM Tris-Cl (pH 8.0). 2.5 units of alkaline phosphatase (Boehringer Mannheim) was used

to dephosphorylate the vector fragments (37°C; 1h). The phosphatase activity was heat inactivated (68°C; 10 min.). Test ligations examined by agarose gel electrophoresis revealed the dephosphorylation to have been successful. The vector preparation was completed by removing the alkaline phosphatase by an organic extraction followed by an ethanol precipitation.

A large scale ligation reaction was carried out with 95μg of treated vector and 24μg of EATRO 795 *Eco* RI fragments at a final concentration of 0.68 mg. ml^{-1} at 4°C, overnight. This provided a 46-fold molar excess of ligatable vector arms to insert, a concentration that favours the addition of vector arms to the large genomic fragments.

The YPH 252 spheroplasts prepared for the attempted construction of a telomere library were again used for the transformation. All manipulations for transformation were identical to those used in **Section 3.5**. The positive control of uncut pYAC4 gave a transformation efficiency of 6500 colonies/μg, while the experimental transformation yielded 1700 colonies/μg, producing about 12 000 transformants in total. After 10 days growth almost all the colonies developed a red colour, indicating the presence of inserts. Unlike the transformants observed from the telomere cloning attempt, these colonies had a very uniform size and appearance.

Twelve colonies were randomly sampled and their karyotypes examined. Three pools of 4 colonies each were grown in YCD broth for 2 days at 30°C and they were processed for chromosome-length DNA. Chromosomal karyotypes were obtained by PFGE separation and Southern blotting onto Zetaprobe GT nylon membrane (Biorad). pBluescript SK+ (Stratagene) was radiolabelled by random priming and used as a YAC-specific hybridisation probe, detecting the *E. coli* plasmid sequences that are not present in the yeast genome. **Fig. 3.5**, lanes 6-8, details the results of this analysis. Intense hybridisation of the pBluescript probe to DNA of 80 to 130 kb suggests the presence of YACs. Based on the 12 samples examined it appears that theYACs fall into two size classes; 80-90 kb and approximately 130 kb, with the former size class being prevalent. A library of YACs of this size is ideal in several respects. First, the contribution of the trypanosome minichromosomes should be relatively minor. Additionally, the largest YACs are still 115 kb smaller than the smallest yeast chromosome and can therefore be more readily gel purified from the yeast genome than if they overlapped in size with the yeast chromosomes. Finally, the smaller the YAC the easier it is to characterise. Encouraged by the results of this analysis I established YAC grids and amplified stocks theoretically representative of the trypanosome genome.

3.7 Gridding, Amplifying and Screening

Regeneration of yeast transformants from spheroplasts requires the growth of colonies embedded in agar. This precludes the direct lifting and screening of colonies. Therefore, the establishment of a YAC library that can be stored and screened requires the primary transformants to be excised from the agar and arrayed in grids upon new plates.

Primary transformants were 'patched' onto agar plates containing YCD, a medium that lacks uracil and tryptophan, and grown at 30°C for 2 days. 3.5 haploid genome equivalents, or 1400 colonies (assuming 100kb inserts) were 'patched' in this way. This was calculated to give a 97% probability of a given single copy sequence being represented in the library, assuming random cloning (Clarke and Carbon 1976; Sambrook *et al.*, 1989). These 1400 colonies were transferred to Hybond N filters in duplicate, one set was grown on YCD/20% glycerol for storage of the filters at -80°C, and the other set was grown on YCD for colony hybridisation. The set of filters for storage was grown at 30°C for 2 days to produce robust colonies. The colonies grown for colony hybridisation were allowed only one day's growth at 30°C, as yeast become more difficult to lyse when they reach stationary phase.

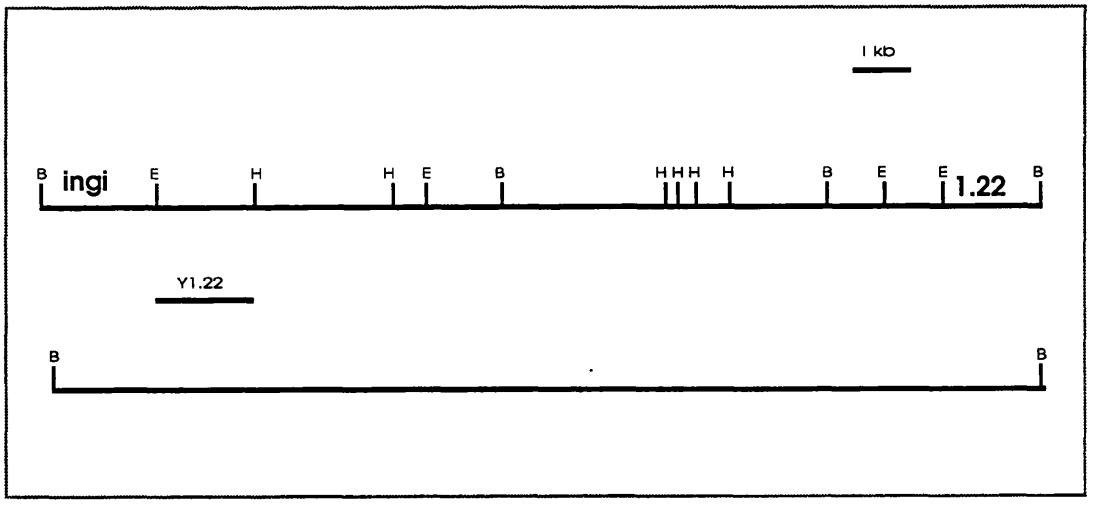


Figure 3.7 This is a representation of the restriction map of the 1.22BC gene locus. The top line represents the genomic map while the bottom, represents the length encompassed by λ1.22B. H is *Hin* dIII; E, *Eco* RI; C, *Cla* I; B, *Bam* HI. There are partial sequences of the retroposon-like element 'ingi' at the 5' end and of the 1.22 M-VSG gene at the 3' end. The fragment Y1.22 was used for screening the YAC library constructed using the pYAC4 vector system. Modified from the PhD. thesis of K.R. Matthews (1990) though the cloning was originally reported in the PhD. thesis of P.G. Shiels (1990).

Fig. 3.7 summarises, in schematic form, the structure of the ILTat 1.22 M-VSG Expression Site (ES) and its upstream region. The important features are the 4 *Eco* RI sites between the chromosome end and the fragment designated as YAC probe 1.22 (Y1.22). YAC cloning of identifiably 1.22-specific sequences must occur from one of these *Eco* RI sites. Y1.22 is the most upstream, single

copy sequence available for screening the YAC library for '1.22-specific' sequences. Hence, Y1.22 was the DNA fragment used to make probes for the subsequent library screen.

The calculations yielding a 97% probability of a single copy sequence being present in the YAC library do not take into consideration the impact telomeric position will have upon a given sequence's representation. An *Eco* RI library with 100kb inserts will only have 1 out every 25 *Eco* RI sites cleaved. In the Y1.22 screening scenario there are only 4 *Eco* RI sites between the chromosomal terminus and the screening probe. The chance of cutting at one of these 4 sites is additive, hence there should be an 6.25-fold underrepresentation of Y1.22 positive sequences within the YAC library. Screening of primary transformants would have required patching more than 10 000 transformants, an excessive task. I therefore amplified the YAC library for screening and storage purposes. **Sections 2.6.3**, **2.6.4**, and **2.6.5** describe the amplification, storage, and colony hybridisation protocols, respectively. Approximately 20 000 amplified colonies were screened with Y1.22 and 6 putative positives were isolated from a secondary Y1.22 screen. These 6 putative positives were investigated further.

3.8 Karyotype Analysis of the Y1.22 'clones'

Chromosome-length DNA was prepared for the 6 clones and chromosomal karyotypes obtained by PFGE separation. A karyotype gel was Southern blotted onto Zetaprobe GT nylon membrane (Biorad) and analysed by hybridisation with Y1.22 (**Fig. 3.8**). The Y1.22 analysis was performed as a tertiary level of confirmation of the positive hybridisation signals and to confirm that the yeast phenotype was being conferred by the presence of a YAC. Hybridisaton of pBluescript (Stratagene) onto the higher resolution filter helps to size the clones, to confirm their structural integrity and to assess their clonality.

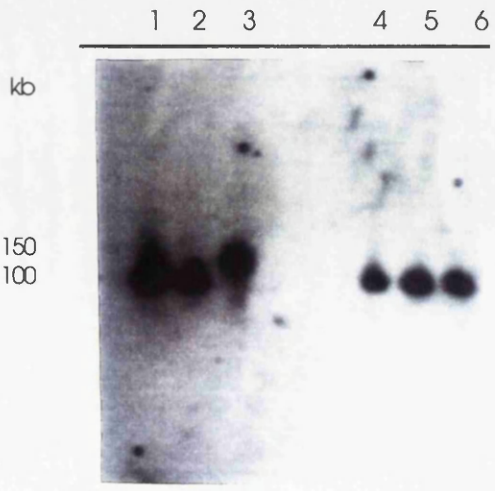


Figure 3.8 This is an autoradiogram of six clones obtained from a Y1.22 screen of the conventional YAC library. The hybridisation (h.s.) was conducted with random primed Y1.22. Gel and PFGE separation parameters are as per Fig. 3.5.

Fig. 3.8 details the results of the Y1.22 analysis. All six putative positives have a minichromosome that hybridises with Y1.22. This confirms the presence of YACs and serves as a tertiary screen. The resolution of this PFGE separation is modest and it is possible to conclude only that five of the six YACs are about 100 kb in size, while the remaining clone is approximately 150 kb. This is similar to the size distribution observed in the random karyotype analysis conducted in **Fig. 3.5**.

A higher resolution gel was run and Southern blotted, as before, and this filter was hybridised with pBluescript SK+ (Stratagene) (**Fig. 3.9**).

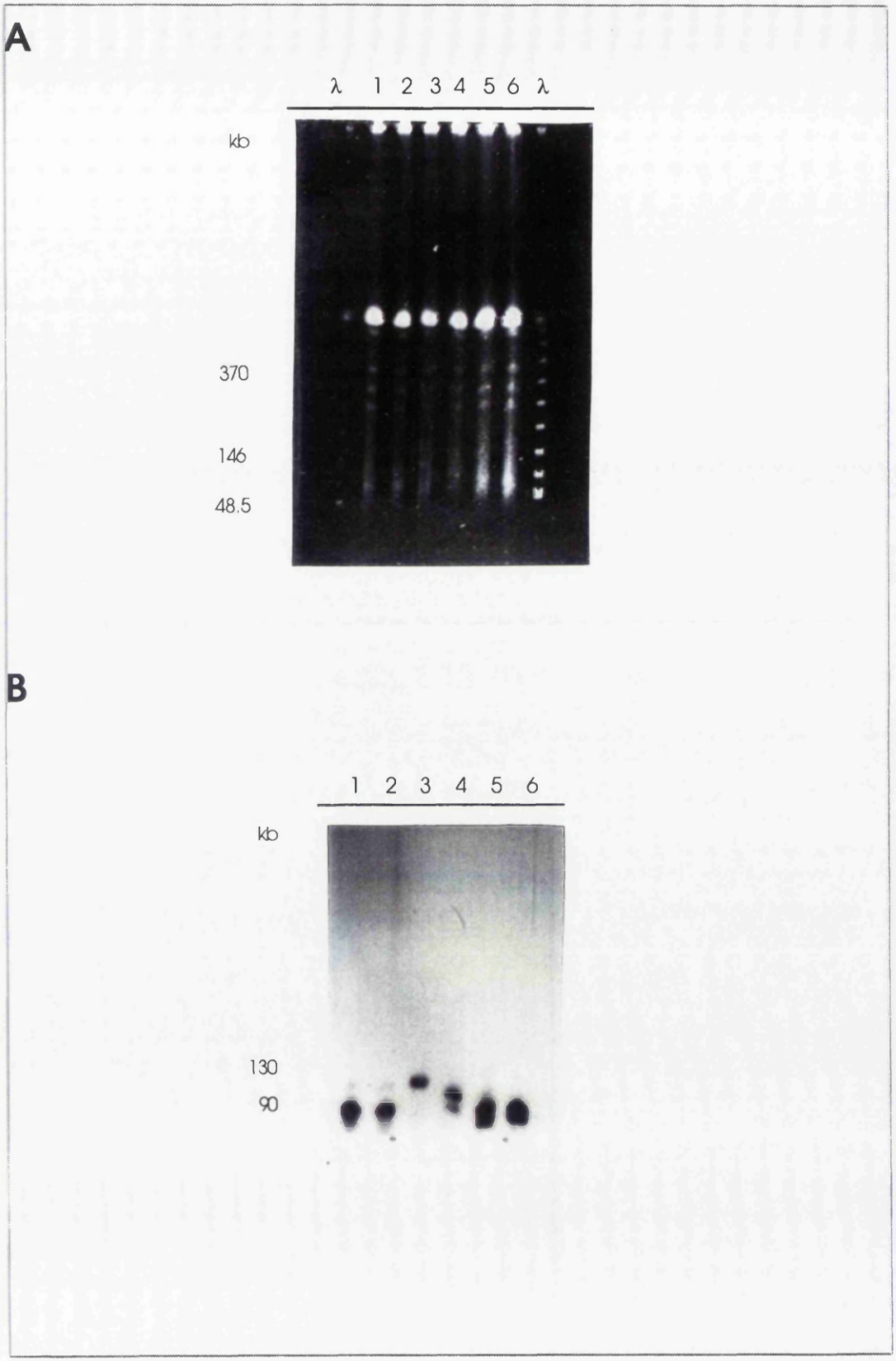


Figure 3.9 Karyotype analysis of the Y1.22 clones utilising plasmid sequences. Panel A is an ethidium bromide stained gel and the Panel B is the southern blot and pBluescript hybridisation (h.s.) of that gel. λ are multimers of λ.The gel is 1% agarose in1XTB0.1E running buffer separated on a Biometra OFAGE system. Interval: 30-3 sec linear ramp; Angle: 120-95° linear ramp; Voltage: 10V/cm; Temp.: 13°C, Duration: 30 hrs.

Fig. 3.9 reveals the results of the pBluescript (Stratagene) hybridisation. The Y1.22 positive YACs hybridised with the plasmid probe, suggesting the presence of YAC arms. However, in the last three tracks there is pronounced hybridisation to other YACs that are not apparent in the Y1.22 experiment. In lane 4 the extra YAC is larger and in lanes 5 and 6 the extra YACs are smaller than the Y1.22 positive minichromosome. These extra bands could be the result of several different problems: YAC instability, the presence of multiple YACs in a single cell or a lack of clonality of the cells. Instability seems unlikely as other Y1.22 clones appear stable and the extra band in lane 4 is larger than the Y1.22 band; not an expected hallmark of instability. The frequency of the problem, 3/6, makes the scenario of multiple YACs within a single cell seem improbable. Some of the extra bands hybridise as intensely as the Y1.22 band, an observation best explained by the YACs being present in the same cell. This scenario cannot be ruled out until randomly picked clones from the library undergo a similar analysis. This is most likely to be a clonality problem. The establishment of the secondary screening plates clearly should have involved the streaking of the primary transformants to obtain single colonies, followed by the patching of single colonies onto gridded arrays. The direct patching of the primary transformants onto gridded arrays was conducted to save time, but this analysis suggests that this approach may have given rise to a clonality problem. In any case two apparently clonal Y1.22 YACs were selected, based on this work, to be studied further.

The clone from lane 1 was chosen as the primary YAC of interest due to its moderate size (90 kb) and the larger clone (130 kb) from lane 3 was picked as a backup. The 90 kb YAC will be referred to hereafter as Ya and the backup clone as Y1.

Chapter 4

Analysing a YAC clone

4.1 Introduction: cloning and contig assembly from YACs

YACs were developed to help link the physical and genetic maps of complex genomes. As more investigators began using this technology in more purpose-based projects (*i.e.* gene cloning and transcriptional analysis) it became necessary to have all the sequences present in a YAC fragmented in conventional bacterial clones to facilitate their detailed analysis. There are two fundamentally different ways of fulfilling this requirement: screening cosmid and λ libraries with YAC probes and by cloning directly from YAC containing DNA. The first approach is more appropriate in organisms with good physical maps, well established genome projects and a relatively low proportion of repetitive sequences in the genome. In *C. elegans*, with its low proportion of repeated sequences (17%), such a probing is particularly powerful, as the cosmid resources exist in ordered arrays. A YAC probing therefore 'establishes linkage and reveals the position and extent of the YAC with respect to the cosmid contig' (Coulson *et al.* 1988). The second approach is called for when these resources are lacking. At the time of writing, for *T. brucei*, such resources are still lacking. A modest genome project has been launched recently in the laboratory of J. Blackwell, but the time when one can send away for ordered grids of λ or cosmid clones has not yet arrived. One could, of course, screen an unordered λ library (a readily available resource) with the YAC clone and isolate clones that cover the YAC but this would negate the advantage that the YAC cloning step gave in the first place; reducing concerns about walking up the wrong chromosome. Addressing this concern in a YAC probing of a trypanosome library would involve the notorious method of probe subtraction (Landegent *et al.* 1987). Even then, sorting through the resultant clones would be an enormously tedious and time consuming venture. For my purposes the only sensible approach is to clone directly from YAC containing DNA.

There exist two alternatives for 'subcloning' the entire YAC insert into λ: constructing λ libraries from total yeast DNA of a YAC clone; and constructing λ 'sublibraries' directly from gel purified YAC DNA. The theoretical and practical aspects of both approaches will be considered in this chapter.

4.2 Karyotype Analysis of the IlTat 1.22 M-VSG basic copy gene

Prior work by K.R. Matthews assigned the 1.22 BC to a 2 Mb or a 3.5 Mb chromosome (Matthews, 1990). A definitive chromosome assignment was, however, not obtained. Erstwhile, Vincent S. Graham optimised the parameters for PFG separation of the large *T. brucei* chromosomes (V.S. Graham, personal communication). **Figure 4.1** is one such PFG karyotype gel; a separation of the large chromosomes of the EATRO 795 trypanosome stock.

The top and bottom-most bands represent electrophoretic mobility compression zones and therefore cannot be usefully thought of as single karyotypic units. The bottom band in particular is a composite of many units; the minichromosomes. The compression zone at the top may also resolve into more 'chromosomes' under different electrophoretic parameters. The bands between these two compression zones will be regarded for convenience as single karyotypic units. On this basis I assigned them 'chromosome numbers'; the top compression zone being I and the bottom band with the minichromosomes, XI.

Vincent S. Graham southern blotted this gel onto Zetaprobe GT nylon membrane (Biorad) and I used a section of this filter in a hybridisation with probe A1.22, derived from the plasmid pMG 7.1-1 (Cornelissen et al. 1985). This probe is 1.22BC gene-specific and should allow an unambiguous assignment of the 1.22BC gene to a 'chromosome' in this separation. Probe A1.22 hybridises to karyotypic unit VII (**Fig. 4.1**). Band VII is bigger than the largest *S. cerevisiae* marker of 2.5 Mb. This suggests that it is the 3.5 Mb chromosome, to which K.R. Matthews also observed hybridisation, that is the genuine karyotypic location of the 1.22BC. The fact that the 1.22BC maps to a single, resolvable, karyotypic unit provides a powerful tool for the confirmation of clones derived from this locus.

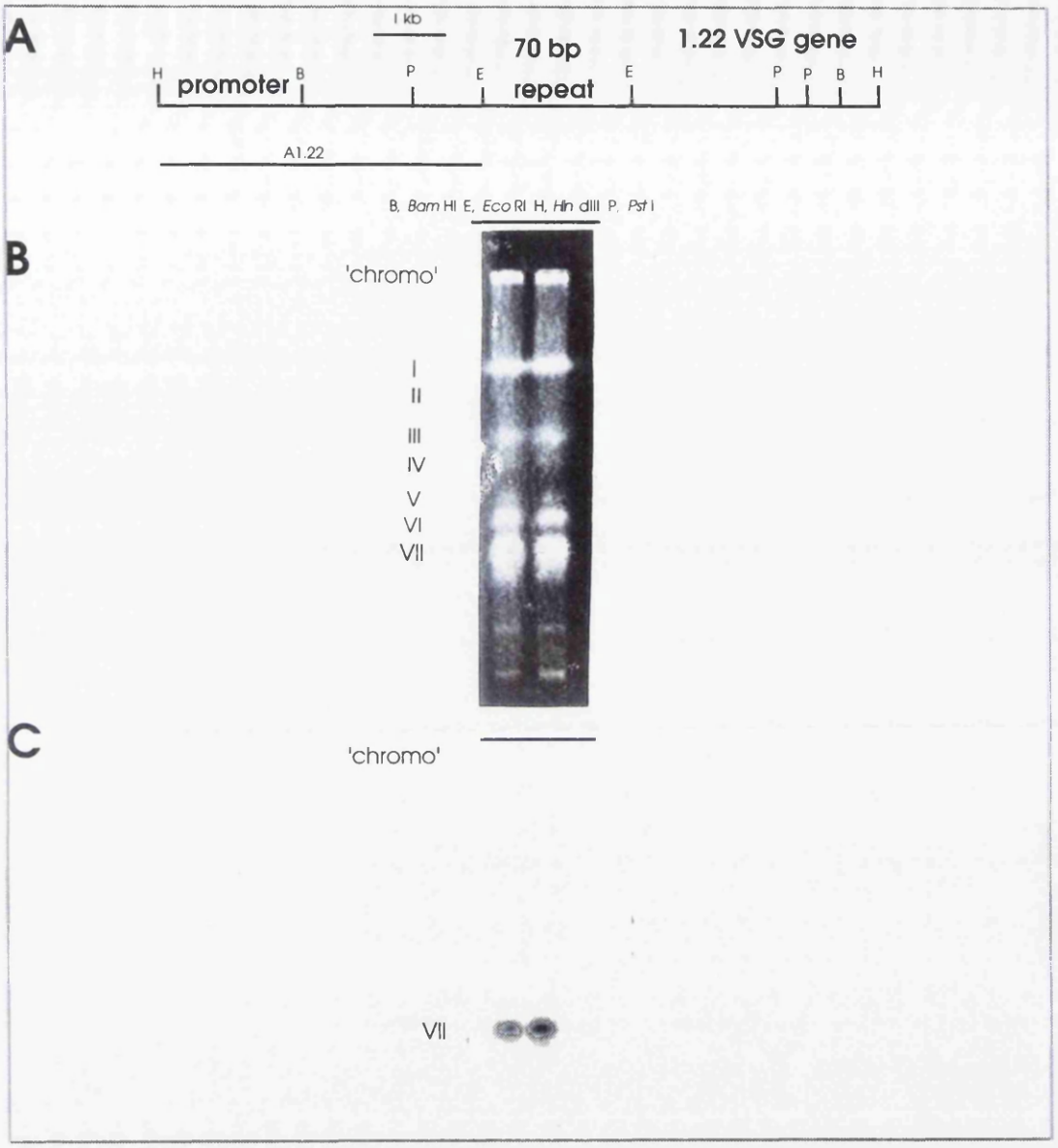


Figure 4.1 Assignment of the IlTat 1.22 M-VSG to a karyotypic unit. Panel A is a map of the insert of pMG7.1-1. Panel B is the ethidium stained karyotype gel. Panel C is the A1.22 hybridisation (h.s.) of a southern blot of the karyotype gel. The probe A1.22, derived from the promoter containing *Hin* dIII/*Eco* RI fragment of pMG7.1-1, is entirely 1.22BC specific. Eleven bands can be discerned from this PFG separation of chromosome-length EATRO 795 DNA. Along the side of the gels 'chromosome numbers' are indicated (see section 4.2) This gel is 0.7% agarose in 1XTB0.1E running buffer separated on a Biometra OFAGE system. Interval: 2000-700 sec. logarithmic ramp; Angle:115-95° linear ramp; Voltage: 4V/cm; Temp.:12°C; Duration: 120 hrs.

4.3 Plasmid end-rescue: an attempt to confirm the karyotypic location of the clones Ya and Y1.

The assignment of the 1.22BC to 'karyotypic unit VII' prompted me to attempt a rapid confirmation of the karyotypic positions of the YAC clones Ya and Y1.

The derivation of probes from the YAC inserts for this purpose was sought by plasmid rescue protocols available for rescuing sequences adjacent to the left arm of the pYAC4 vector. Plasmid end-rescue offers a quick way to obtain insert DNA without any prior analysis of the YAC clone.

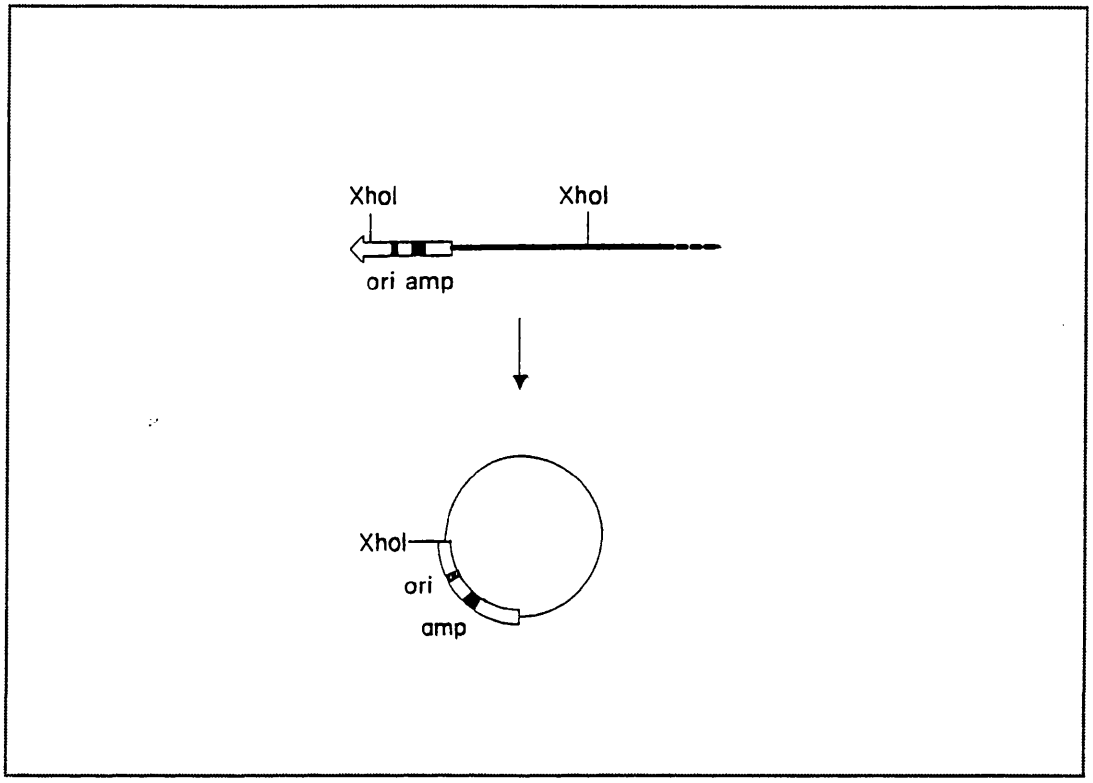


Figure 4.2 A schematic representation of the cloning chemistry involved in left arm plasmid rescue protocols. Lines represent *T. brucei* DNA in both the YAC and in the plasmid. The arrowhead represents telomere tracts, while the boxes represent vector DNA. The open boxes represent yeast sequences and the shaded boxes are the *E. coli* origin of replication and a β-lactamase gene. Circularised *Xho* I fragments with these YAC left-arm sequences can be propagated as plasmids. Taken from Bentley, (1992).

Examination of **Figure 4.2** reveals that *Xho* I digestion of the genomic DNA from YAC positive yeast will cleave away the YAC telomere tract and cut at the first *Xho* I site in the insert, yielding an *Xho* I fragment that can be rescued as a plasmid. This *Xho* I digested DNA can be ligated at low concentrations that favour the intramolecular reaction. Transformation of these ligation products into *E. coli* will result in transformants when the molecules taken up are circular and possess the left arm sequences (Amp^R and oriC) required for plasmid selection and replication.

Genomic DNA was prepared as described by Sherman *et al.* (1986) from Ya and Y1 containing yeast. These DNAs were digested with *Xho* I, ethanol precipitated and resuspended in TE. A series of ligations with different final concentrations of *Xho* I fragments were set up under appropriate buffer conditions in the presence of 3 Weiss units of T4 DNA ligase (Promega) per reaction. Ligation reactions were incubated overnight at 16°C for the following concentrations of *Xho* I digested Ya and Y1 DNAs: 24 ng/μl, 12 ng/μl, 6 ng/μl, 4 ng/μl, 1 ng/μl, 800 pg/μl, 160 pg/μl and 32 pg/μl. These ligations were ethanol precipitated and then resuspended in small volumes of TE.

E. coli transformations were performed as described in **Section 2.5.9** with 50-100 nanogram amounts of the experimental ligation mixes in addition to a pYAC4 positive control. Only the positive control yielded transformants.

There are several possible reasons for a negative result in this kind of experiment. On the purely technical side, problems with substrate purity may be impacting upon the efficacy of the ligation reactions. Perhaps too little DNA was used in the transformations; the end plasmid will account for only a small proportion of the total reaction products. However, theoretical contingencies must be considered as well. The first *Xho* I site in the YAC insert may occur too far from the junction with the left arm to make plasmid end-rescue even possible. As the ultimate goal here is contig assembly and because of the uncertain prospects of continued plasmid rescue attempts, this method was discontinued in favour of cloning the entire YAC in phage λ.

4.4 *In situ* YAC mapping

It is possible to derive restriction site mapping data for YAC clones in the absence of sufficient probes for exhaustive southern analysis. Left and right arm-specific vector probes can be used in conjunction with partial digestion of the YAC positive yeast genomic DNA to yield a restriction map. This is an adaptation, for YAC work, of the technique described by Smith and Birnstiel (1976). This kind of analysis, when applied to large entities such as YACs, provides useful 'landmarks' to help orient the investigator in later, more detailed, analyses of the clone. The resolution of this mapping technique depends upon how well the partial digests can be titrated and, more importantly, upon the resolution of the electrophoresis used to separate the fragments.

Owing to the large size of the clones to be examined, agarose-embedded chromosome-length DNA was prepared as described in **Section 2.5.4**. This meant that the conditions for partial digestion had to be perfected for DNA digested in agarose blocks. The agarose blocks were washed five times in ice-cold TE over a 2 hour interval. The blocks were then preequilibrated, on ice, for 30 min. with 0.5 ml of the relevant restriction buffer. The buffer was changed for 60 µl of the restriction buffer supplemented with gelatin and DTT to concentrations of 100 µg/ml and 1 mM, respectively. Enzyme was added at this point and the reaction tubes incubated, on ice, for a further 30 min. to allow the enzyme to permeate the agarose block. The reactions were then incubated overnight at 37°C. Several PFG gels were run in enzyme titration experiments to establish the optimal conditions for partial digestion for each restriction enzyme for which mapping data were desired. Once this was accomplished, mapping gels were run and southern blotted onto Zetaprobe GT nylon membranes (Biorad). The 2.5 kb *Pvu* II /*Bam* HI fragment of pBR322, a left arm-specific probe, was prepared in a low melting point agarose gel. This fragment was radiolabelled by random priming and hybridised, at high stringency, to the mapping filters. The results are presented in **Figure 4.3.**

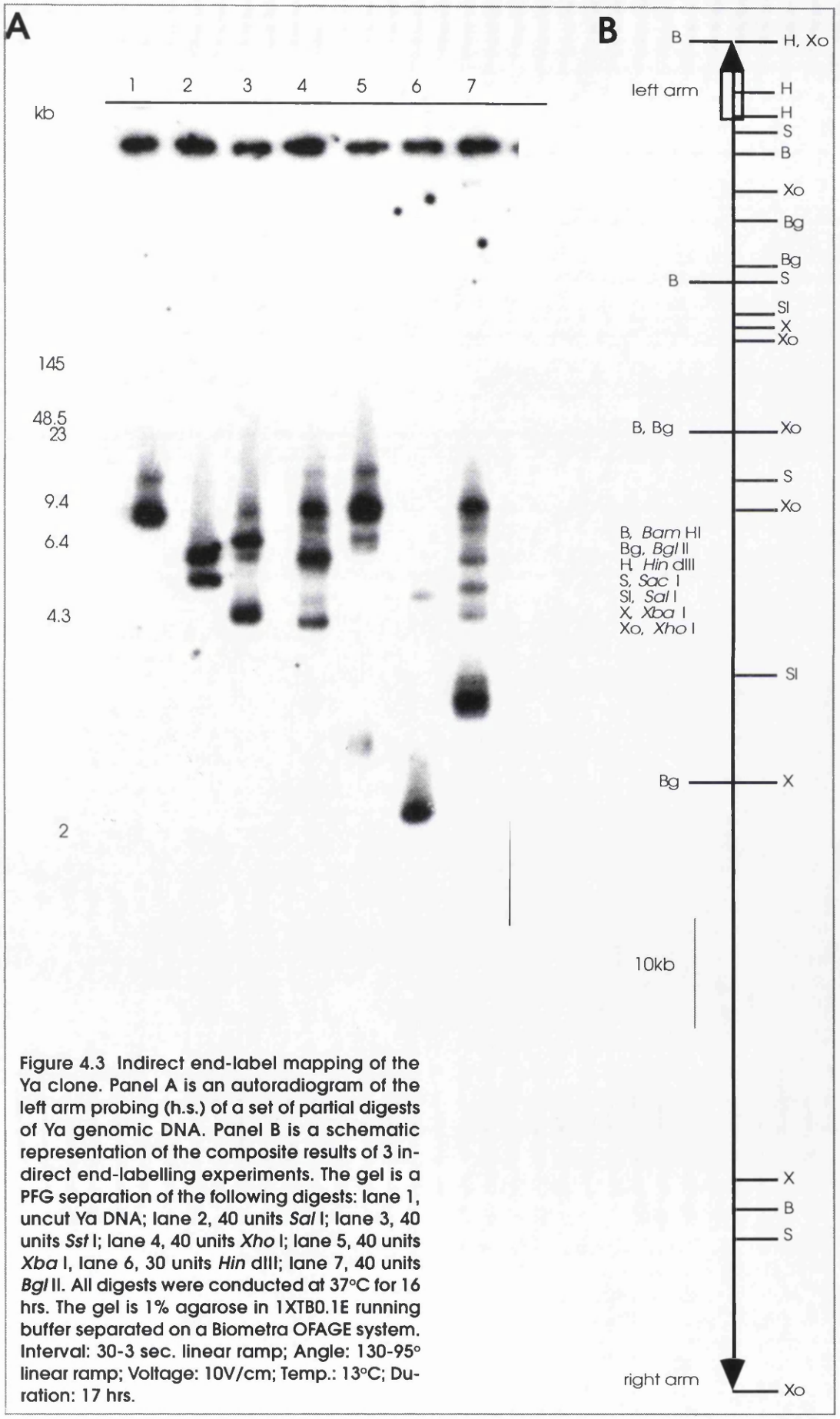


Figure 4.3 Indirect end-label mapping of the Ya clone. Panel A is an autoradiogram of the left arm probing (h.s.) of a set of partial digests of Ya genomic DNA. Panel B is a schematic representation of the composite results of 3 indirect end-labelling experiments. The gel is a PFG separation of the following digests: lane 1, uncut Ya DNA; lane 2, 40 units *Sal* I; lane 3, 40 units *Sst* I; lane 4, 40 units *Xho* I; lane 5, 40 units *Xba* I, lane 6, 30 units *Hin* dIII; lane 7, 40 units *Bgl* II. All digests were conducted at 37°C for 16 hrs. The gel is 1% agarose in 1XTB0.1E running buffer separated on a Biometra OFAGE system. Interval: 30-3 sec. linear ramp; Angle: 130-95° linear ramp; Voltage: 10V/cm; Temp.: 13°C; Duration: 17 hrs.

The stripping and reprobing of these filters with the right arm-specific 1.6 kb *Pvu* II/*Bam* HI fragment of pBR322 produced signals too faint to be informative.

This analysis was impeded by the poor resolution of the PFG gels used to separate the partially digested genomic fragments. Very high resolution separation of DNAs sized between 0.1 kb and 150 kb is afforded by Field Inversion Gel Electrophoresis (FIGE); this laboratory had no appropriate FIGE facilities at the time these experiments were performed.

4.5 Cloning the YACs into phage λ

Phage λ is easy to use and accommodates inserts up to 22 kb in size. This makes it the obvious tool of choice for the investigator faced with the prospect of a chromosome walk. The initial YAC cloning step has mitigated the usual concerns with walking onto an unlinked region of the genome, by removing the sequence cloned from the context of the rest of the *T. brucei* genome. Two approaches were employed, simultaneously, to clone the YACs into phage. First, it was attempted to clone directly from gel isolated YAC DNA into λ using the relatively esoteric method reported by Whittaker *et al.* (1993). Secondly, the simpler but ultimately tedious method of constructing genomic λ libraries from Ya and Y1 positive yeast was utilised.

It was attempted to construct a phage 'sublibrary' directly from the YAC DNA, as the generation of 50 or more recombinants would virtually assure full coverage of the YACs. This method would greatly expedite contig assembly by obviating the necessity for a full library screen and the subsequent plaque purifications at each step. This method employs special adaptations of conventional library generation protocols to make library construction possible from nanogram amounts of input DNA. Partial digests are conducted by a *Dam* methylase/*Mbo* I competition strategy developed by Hoheisel *et al.* (1989). The resultant *Mbo* I overhangs are partially end-filled and then ligated to λ phage prepared so as to have complementary 2 bp overhangs. This method removes most nonrecombinant background and chimaerism without the need for physical fractionation methods that require large amounts of input DNA.

The attempt to use this technology was unsuccessful. 50 plaques were obtained. These plaques represent background as none of them hybridised with a probe labelled from the entire YAC. The following is a cursory overview of the procedures undertaken.

1) LMP PFG preparative gel runs and excision of the YAC DNA
2) gelase (Cambio) treatment of the agarose slices and ethanol precipitation of the YAC DNA
3) titration of *Dam* methylase/*Mbo* I partial digestion conditions on unmethylated λ DNA (possible because the degree of digestion is independent of DNA concentration)
4) experimental *Dam* methylase/*Mbo* I partial digestion of Ya DNA

5) partial end-filling by Klenow with A and G of the products of step 4
6) ligation of the prepared Ya DNA to partially end-filled *Xho* I digested arms of λFIXII (Stratagene)
7) *in vitro* packaging and plating
8) filter lifts and hybridisation with total Ya probe

The major deviation from Whittaker *et al.* (1993) was the use of standard grade LMP agarose (BRL) in the preparative gels, rather than the recommended high grade SeaPlaque (FMC) agarose. Perhaps inhibitors in the agarose used contributed to my inability to generate recombinant clones.

Inserts were prepared from Ya and Y1 positive yeast genomic DNA for ligation to partially end-filled *Xho* I arms of λFIXII. *Sau* 3AI was titrated to determine the optimal partial digestion conditions for genomic library construction. The optimal conditions for the partial digestion of Ya genomic DNA was found to be $3.3(0.5)^{11}$ units. μg^{-1} DNA (1 h; 37°C). Optimal conditions for Y1 were $2(0.5)^{10}$ units. μg^{-1} DNA (1 h; 37°C). 50 µg of each Ya and Y1 were *Sau* 3AI digested in scaled up, optimal, partial digest reactions in 200 µl final volumes. Aliquots from these reactions were examined by agarose gel electrophoresis and the resultant fragments were found to be within the expected size range. 1 µg of partially end-filled *Xho* I arms of λFIXII was dephosphorylated with 0.025 units of alkaline phosphatase (Boehringer Mannheim) at 37°C for 1 h. The alkaline phosphatase activity was heat inactivated (68°C; 10 min.). The vector arms were then organically extracted, ethanol precipitated and resuspended in 2 µl of TE. Insert partial end-filling, ligation to the vector arms, *in vitro* packaging (Gigapack II Gold from Stratagene) and plating of phage on a lawn of XL-1 Blue MRA *E. coli* cells were all performed in accordance with Stratagene's recommendations in its λFIXII *Xho* I arms cloning kit.

4.6 Screening the Ya genomic library for YAC-specific clones

Only the λ library made from Ya genomic DNA was plated for screening. A 99% probability of a sequence's occurrence was sought for screening purposes. An average insert size of 16 kb was assumed. According to the formula of Clarke and Carbon (1976) 9200 plaque forming units (pfu) must be screened for this level of representation.

Approximately, 10 000 pfu were plated onto a lawn of the XL-1 Blue MRA strain of *E. coli* (Stratagene) on a 22 cm^2 plate of BBL agar. Duplicate plaque lifts were taken onto Hybond N nylon filters (Amersham), as described by Stratagene, that were subsequently baked at 80°C for 2 h. These filters were alternately screened with the three available YAC-specific probes: the 2.5 kb *Pvu* II/*Bam* HI fragment of pBR322 (left arm YAC sequence), the 1.6 kb *Pvu* II/*Bam* HI fragment of pBR322 (right arm YAC sequence) and the Y1.22 probe used to isolate the YACs (see **Fig. 3.7**). Clonally pure stocks of each of the duplicate positives from the library screens were obtained by taking them through at least

two more rounds of plating on 90 mm petri plates, and screening with the relevant probes. One positive clone from each screening was chosen for further analysis. The left arm derived clone was named LλYa, the Y1.22 derived clone Yλ122 and the right arm containing clone as RλYa.

4.7 Transcriptional analysis of LλYa and Yλ122

I obtained three YAC-specific λ clones and may have, thereby, cloned as much as 50 kb of the 90 kb Ya clone. I was not aiming for contig assembly as an end in itself but rather to use it as a means of defining the extent of the transcriptionally silent 'metacyclic domain' of the 1.22BC locus. It was desirable to obtain transcriptional maps of these λ clones. This information would assist me in deciding whether to clone more of the YAC and what clones not to analyse any further. V.S. Graham was simultaneously studying the ILTat 1.61 M-VSG BC locus, including transcriptional mapping. We therefore collaborated on transcriptional run-on experiments.

When information on the transcriptional status of a large stretch of cloned DNA is required there are two methods of choice: transcriptional run-ons and 'reverse northerns'. Both protocols use southern or slot blotted clones as substrates in hybridisation experiments indicative of gene expression. Reverse northerns use a probe of radiolabelled, first strand cDNA, synthesized from poly(A)+ selected RNA. Run-ons use a probe of short radiolabelled nascent RNA transcripts produced in cell-free nuclei; this RNA is extracted and hybridised to the southern or slot blot substrate. The two techniques yield different information. A reverse northern, in principle, should tell whether the cloned sequence gives rise to steady state RNA in the cell. Run-ons, on the other hand, should tell whether the cloned sequence was being transcribed at the time that the nuclei were isolated. Because of the nature of gene expression in trypanosomes, reverse northerns, the preferred technique applied to other organisms, are not a satisfactory method of searching for an expressed gene. A sequence that is transcribed may not give rise to steady state RNA in the life cycle stage from which the poly(A) RNA was extracted. Unless transcriptional run-ons are undertaken, large areas of transcribed DNA may appear negative.

Bloodstream nuclei from the 1.22d clone of ILTat 1.22 (EATRO 795) were isolated by passing trypanosome infected rat's blood through a cell disruptor. A radioactive transcriptional elongation reaction was incubated for 5 min. at 37°C. The RNA was then isolated by BRL's Trizol protocol and used as a hybridisation probe onto a filter with experimental and control clones. Details of the run-on protocol can be found in **Section 2.6.6.** The results of this transcriptional analysis are presented in **Figure 4.4.**

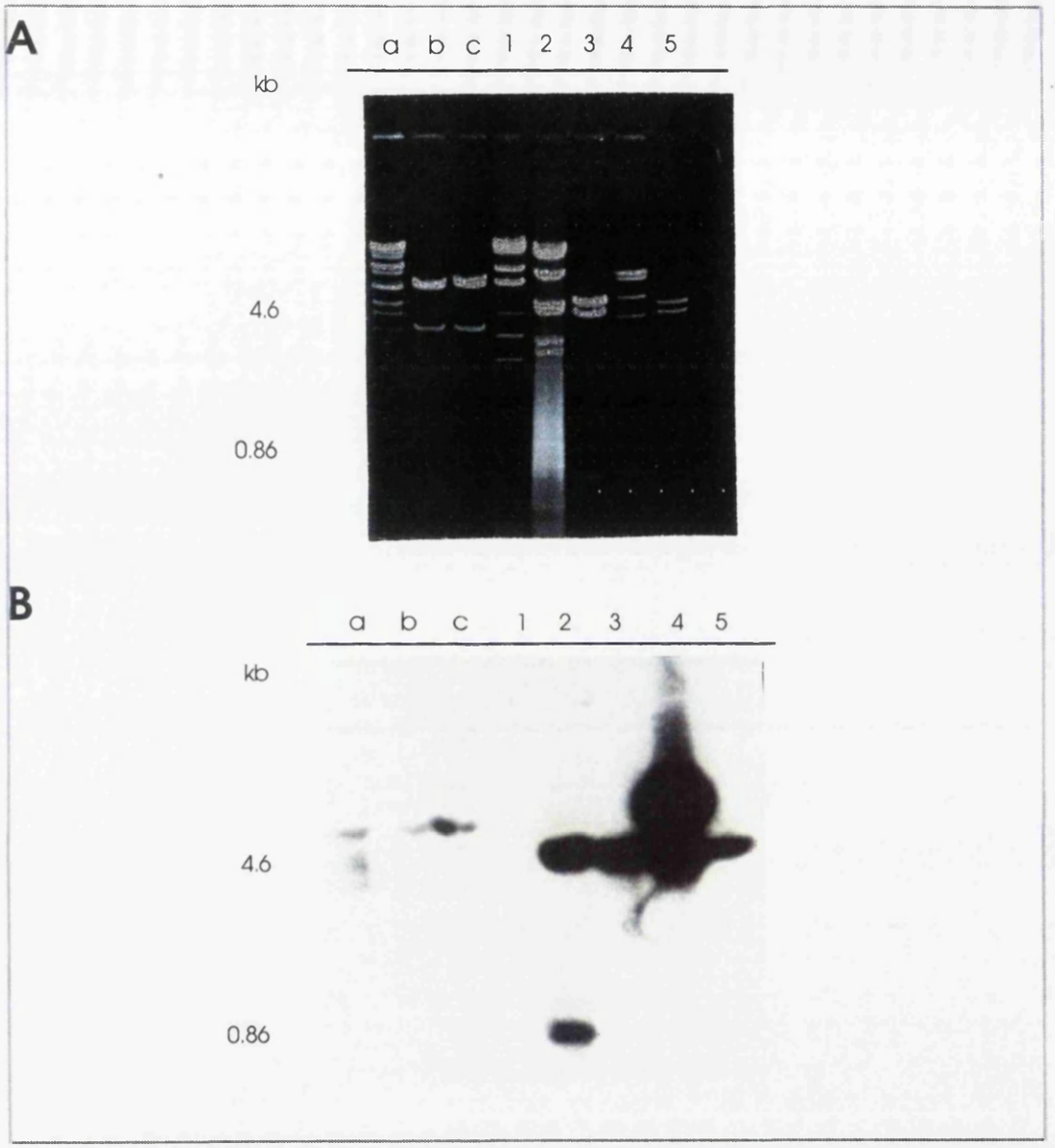


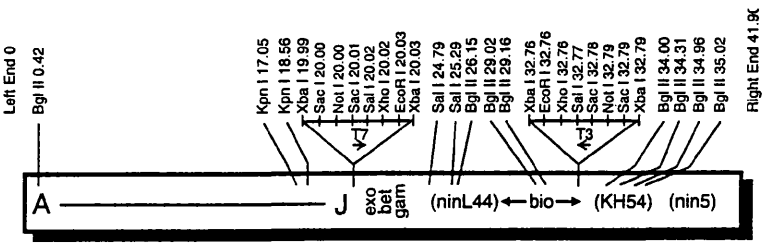
Figure 4.4 Transcriptional run-on analysis of LλYa and Yλ122. Panel A is an ethidium bromide stained gel with experimental and control clones. Panel B is an autoradiogram of the run-on hybridisation (3XSSC; 55°C) washed to 0.1XSSC at 55°C. Lanes a-c, experimental clones derived from the 1.61 metacyclic telomere by Vincent S. Graham; Lane 1, *Hin* dIII/*Not* I digest of LλYa; lane 2, *Hin* dIII/*Not* I digest of Yλ122; lane 3, *Hin* dIII digest of pMG7.1-1 (1.22 BC control); lane 4, rDNA control (pR4); lane 5, tubulin control (pT1).

The controls indicate that the elongation reaction and hybridisation have been successful. rDNA, tubulin and the 1.22 M-VSG gene all yield strong signals while there is no cross hybridisation to vector bands. The left arm containing LλYa clone appears to be transcriptionally silent while Yλ122 has two fragments that give rise to very strong signals. These two bands are interpreted as a 4.6 kb *Hin* dIII fragment and an 860 bp *Hin* dIII/*Not* I fragment. It is not yet known whether these are contiguous or represent disparate fragments within the clone showing evidence of being transcriptionally active. Clearly, more analysis is required, in particular of Yλ122.

4.8 Some mapping of the λ clones

It was decided to generate some restriction site mapping data to assist later analyses, to obtain some indication where the transcriptionally active units lay within the Yλ122 clone and to help confirm the identity of the clones. Single, double and triple restriction digest were carried out, in appropriate buffer conditions, on Yλ122 and RλYa. The resultant fragments were separated on 20 cm long agarose gels to provide maximal separation. End fragment identity and insert polarity were established by similar gels being southern blotted onto Hybond N (Amersham) membranes and hybridised with short end-specific (T3 or T7) riboprobes. The riboprobes were synthesized from *Rsa* I digested DNA from each clone and prepared as described in the Stratagene protocol to their riboprobe sythesis kit. **Figure 4.5** shows the map of the λFIXII vector used in the library construction and tabulated results of restriction digests for the two clones examined. **Figure 4.6** presents the deduced maps.

A



B

	A	B	C	D	E
1	Not I	(kb)	~20		
2	Bam HI		NO	Bam HI	
3	Bam HI/Not I		SITES	FOUND	
4	Hin dIII		>23	13	9.8
5	Hin dIII/Not I		9.8	4.2 (T7)	3 (T3)
6	Hin dIII/Bam HI		>23	13	9.8
7	Sst I		14	3.7	
8	Sst I/Bam HI		14	3.7	
9	Xba I		NO	Xba I	
10	Xba I/Bam HI		SITES	FOUND	
11	Xho I/Not I		13.5	4.8 (T7)	
12	Sst I/Hin dIII		4.2	3	
13	Xho I/Hin dIII/Not I		8	4.2	3

C

	A	B	C	D	E	F	G	H	I	J
1	Bam HI	(kb)	10.5	6.8	5.7					
2	Bam HI/Not I		6.8	5.7	1.7 (T3)	1.4 (T7)				
3	Eco RI		8	4.3	0.1					
4	Eco RI/Not I		8	4.3	3.4 (T7)	0.1(T3)				
5	Hin dIII		14	4.8	4.2	2.2	0.66	0.45		
6	Hin dIII/Not I		4.8	4.2	2.6 (T3)	2.2	0.86 (T7)	0.66	0.45	
7	Bam HI/Eco RI		>23	>9	5.7	4.3	1.5	1.2	0.37	
8	Bam HI/Eco RI/Not I		5.7	4.3	1.6	1.5	1.2	0.37	0.1	
9	Bam HI/Hin dIII		10	3.2	2.3	1.6	1.35	0.98	0.66	0.45
10	Bam HI/Hin dIII/Not I		3.2	2.3	1.6	1.35	0.98	0.75	0.66	0.45
11	Hin dIII/Eco RI		4.3	2.9	2.5	2.4	2.3	1.8	0.66	
12	Hin dIII/Eco RI/Not I		4.2	2.9	2.5	2.4	2.3	1.8	0.86	0.66

Figure 4.5 Restriction mapping data for Yλ122 and RλYa. Panel A is a restriction map of the λFIXII vector. Panel B is a Table of restriction data for RλYa. Panel C is a Table of restriction data for Yλ122. For simplicity, only fragments that have insert DNA are included in the Tables. B, Bam HI; E, Eco RI; H, Hin dIII; N, Not I; S, Sst I; X, Xba I; Xo, Xho I. T^3 and T^7 indicate end-fragment identity.

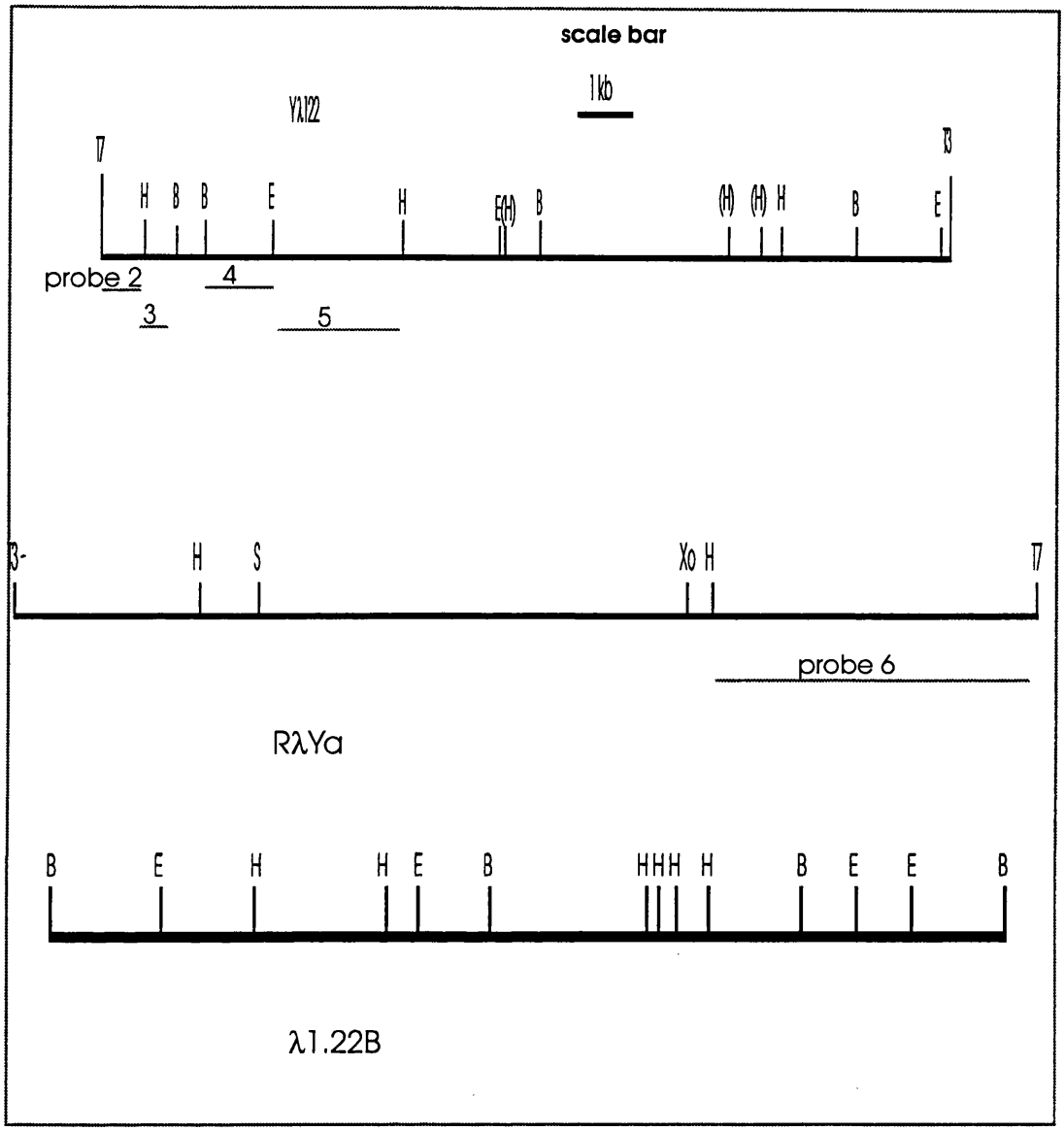


Figure 4.6 Restriction maps for Yλ122, RλYa and λ1.22B. λ1.22B is the original λ clone of the 1.22BC locus and Y1.22 is the fragment derived from this, that was used to isolate the YAC clones. B, *Bam* HI; E, *Eco* RI; H, *Hin* dIII; N, *Not* I; S, *Sst* I; X, *Xba* I; Xo, *Xho* I. The probe bars indicate probes used in the experiments documented in Figs. 4.8-4.9.

The restriction map of Yλ122 shows that the two bands that gave a positive signal in the run-on form a contiguous unit that starts at the T7 boundary of the clone. These fragments were subcloned into pBSK$^-$ for further characterisation. The remainder of the clone appears to be transcriptionally silent. On comparing the maps of Yλ1.22 and λ1.22B, the expected overlap was absent; indeed the maps differ extensively (**Figure 4.6**). These data immediately raised the spectre of a significant problem, perhaps in the original YAC cloning step. Clearly the next step was to check fragments derived from the λ clones in hybridisation experiments onto blots of trypanosome genomic DNA and karyotype profiles.

4.9 Probing the problem

There are several scenarios that can be envisioned to explain the lack of overlap between λ1.22B and Yλ122. The λ clones may not even represent trypanosome DNA. The original Ya YAC clone may be chimaeric, despite original precautions. Something may be wrong with the Y1.22 probe used to screen the YAC and genomic Ya libraries.

Genomic southerns with λ clone derived probes would help test the worst case scenario of the clones not containing trypanosome DNA. Use of those probes that appear by this southern analysis to be single copy, in karyotype analysis, would help indicate if gross chimaerism, at some level in the cloning/screening process, may be responsible for the apparent problems. **Figs. 4.7-4.9** outline the results of these experiments.

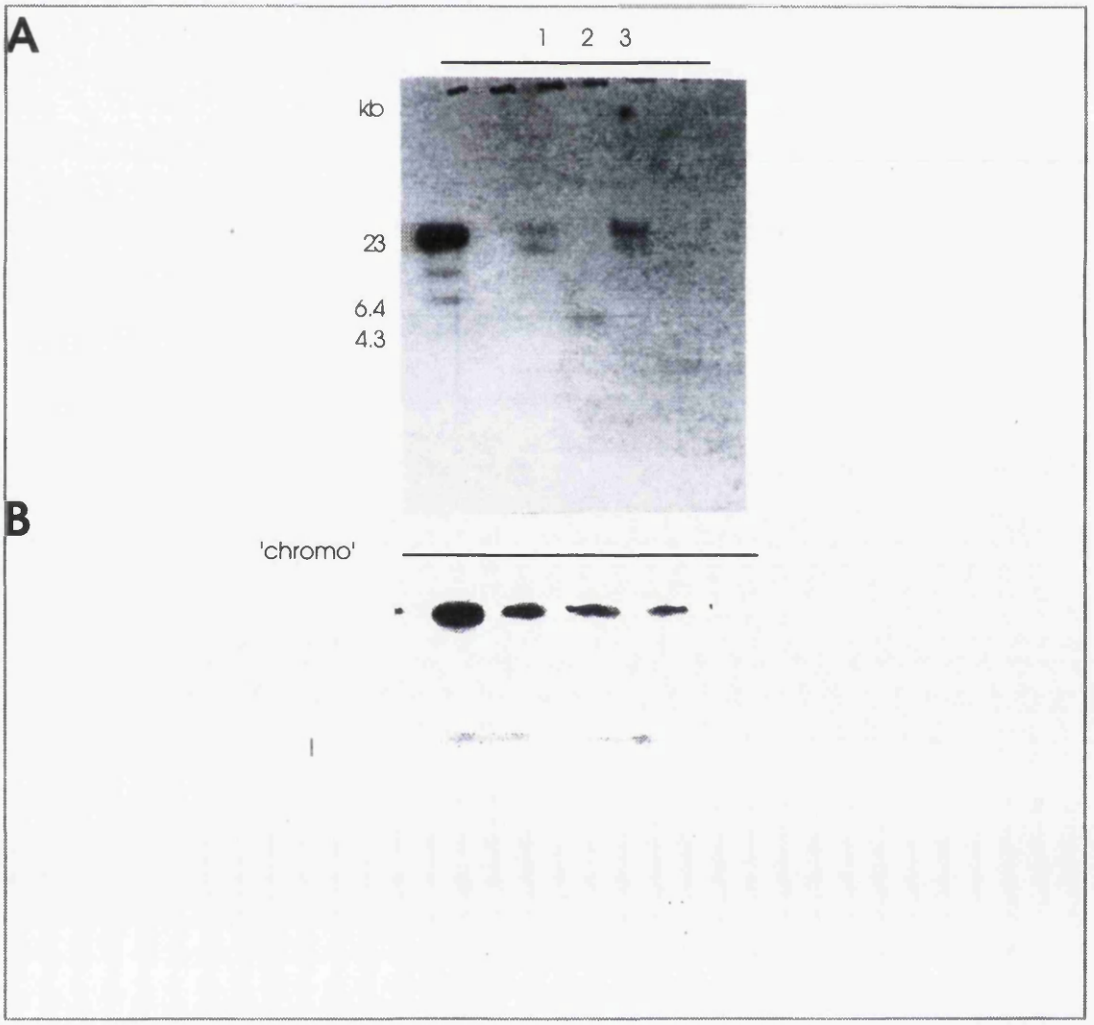


Figure 4.7 Southern and karyotype analysis of LλYa. Panel A is an autoradiogram of a southern blot of EATRO 795 DNA. Panel B is an autoradiogram of an EATRO 795 karyotype blot. Both panels were hybridised, at high stringency (h.s.), with a 3 kb *Hin* dIII/*Not* I end fragment (probe 1). Lane 1, *Bam* HI digested DNA; lane 2, *Hin* dIII digested DNA; lane 3, *Xba* I digested DNA. The karyotype gel had quadruplicate loading and was run by the parameters reported in Fig. 4.1. Probe 1 is an end fragment, the other end fragment is presumed to be composed of left arm sequences and therefore was not investigated. Along the side of the karyotype gel 'chromosome numbers' are indicated (see section 4.2).

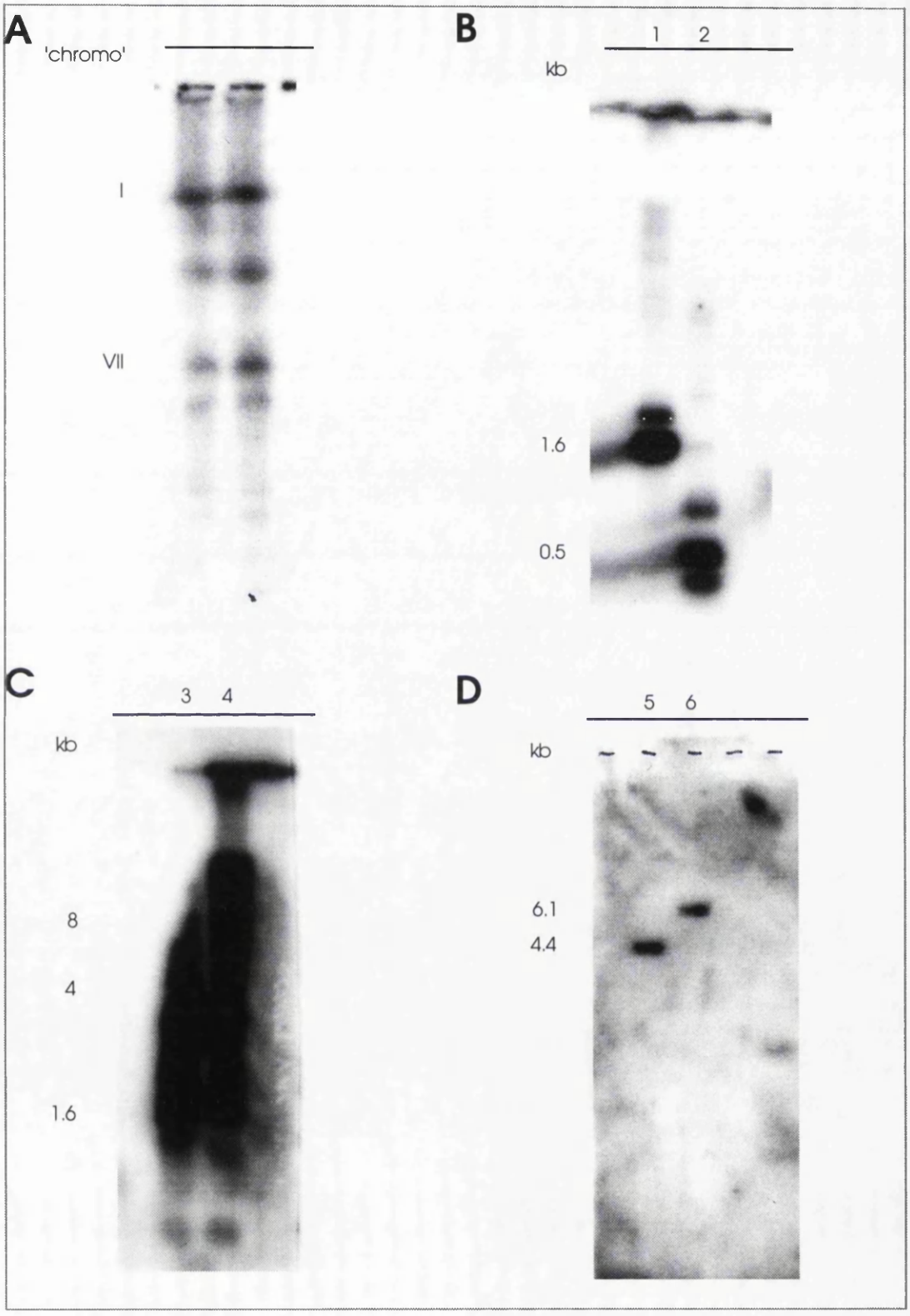


Figure 4.8 Southern and karyotype analysis of Yλ122. Panel A is an autoradiogram of an EATRO 795 karyotype blot hybridised (h.s.) with probe 2 (Fig. 4.6). Panel B is an autoradiogram of an EATRO 795 southern blot hybridised (h.s.) with probe 3 (Fig. 4.6). Panel C is an autoradiogram of an EATRO 795 southern blot hybridised (h.s.) with probe 4 (Fig. 4.6). Panel D is an autoradiogram of an EATRO 795 southern blot hybridised (h.s.) with probe 5 (Fig. 4.6). The karyotype gel had duplicate loading and was run using the parameters reported in Fig. 4.1. Lane 1, *Bam* HI digested DNA; lane 2, *Bam* HI/*Hin* dIII digested DNA; lane 3, *Bam* HI/*Eco* RI digested DNA; lane 4, *Bam* HI digested DNA; lane 5, *Eco* RI digested DNA; lane 6, *Bam* HI digested DNA. Probes 3-5 were derived from the plasmid subclone rather than directly from Yλ122.

The results of panels B,C and D are of particular interest as the probes used span the 4.6 kb *Hin* dIII fragment which was strongly positive in the transcriptional run-on experiment described in **Section 4.7**. The area covered by probes 3 and 4 is multicopy in the genome, while probe 5 appears to be of low copy number. Later transcriptional run-on analysis (**Section 6.7**) revealed that the area covered by the 860 bp *Hin* dIII/*Not* I fragment of Yλ122 and probes 3 and 4 comprise most of a contiguous unit transcribed in an apparently α-amanitin insensitive manner. Probe 5 is silent in this analysis and defines the boundary of the multicopy transcription unit. Despite the problem identified by the experiments in this section, it is of interest to determine, eventually, the identity of this transcription unit (see **section 4.13**).

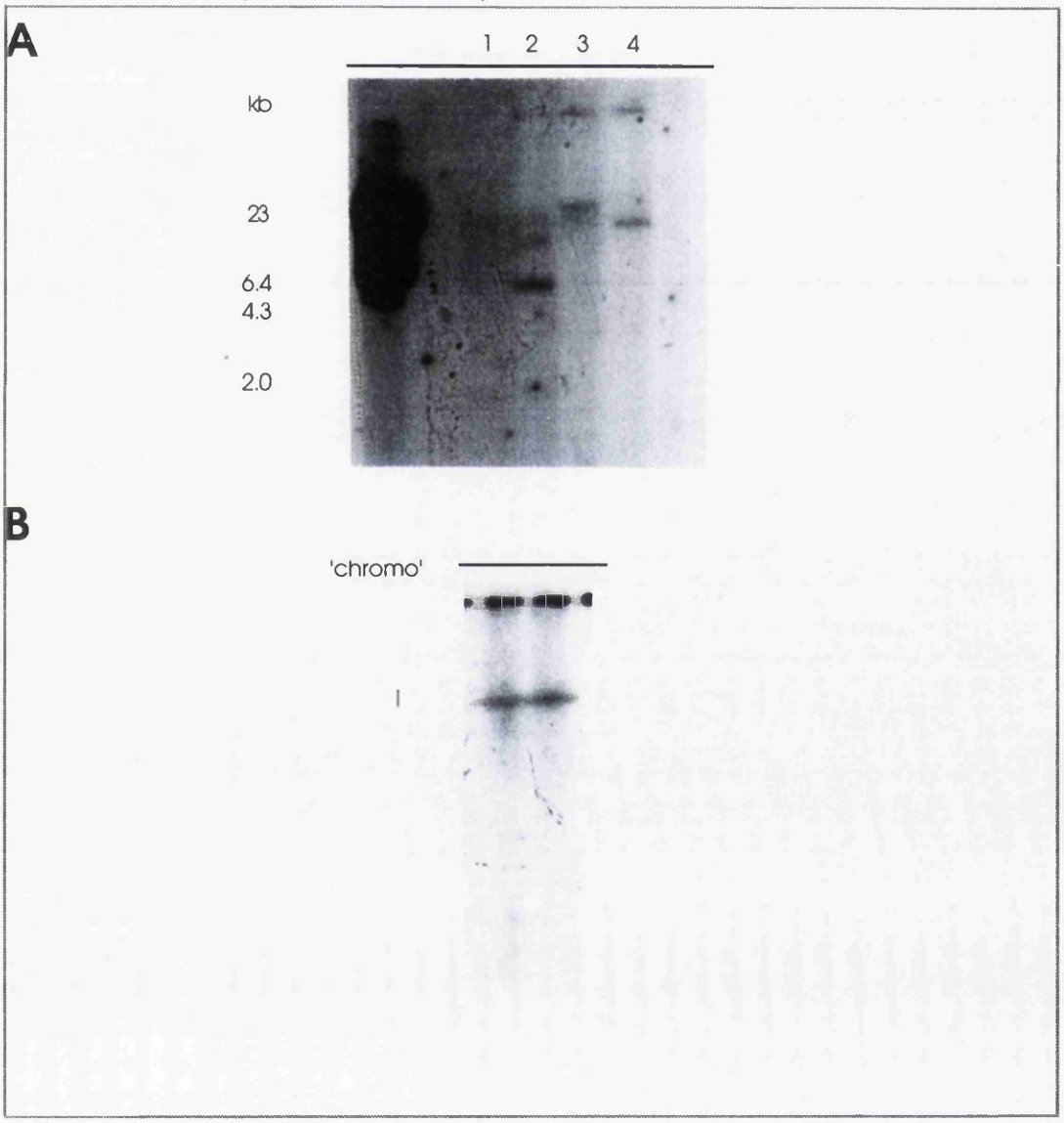


Figure 4.9 Southern and karyotype analysis of RλYa. Panel A is an autoradiogram of an EATRO 795 southern blot. Panel B is an autoradiogram of an EATRO 795 karyotype blot. Both panels were hybridised (h.s.) with probe 6 (Fig. 4.6). Lane 1, *Bam* HI digested DNA; lane 2, *Hin* dIII digested DNA; lane 3, *Sst* I digested DNA; lane 4, *Xba* I digested DNA. The karyotype gel had duplicate loading and was run using the parameters reported in Fig. 4.1. Probe 6 is an end fragment, the other end fragment (3 kb *Hin* dIII/*Not* I) failed to give a signal in a trypanosome southern blot but did hybridise (h.s.) in a Ya yeast karyotype blot, to the YAC. It is presumed that this fragment represents right arm sequences. Along the side of the karyotype gel 'chromosome numbers' are indicated (see section 4.2).

The worst case scenario of the λ clones not representing trypanosome DNA was eliminated by this analysis. However, both probes 1 and 6, which hybridise to only one band in the karyotype profiles, did so to band I as designated in **Section 4.2**. The 1.22BC was localised, in that section, to band VII. The Ya derived probes hybridise to the wrong 'chromosome'.

The result from probe 1 strongly suggested that the Ya clone was chimaeric. The result from probe 6, from the other extreme of the YAC, made this explanation seem less plausible as the two probes are separated by about 60 kb. This therefore implied a problem with the Y1.22 probe used to isolate the YACs and Yλ122.

4.10 Karyotype Analysis of the Y1.22 probe

Clone Ya was isolated by screening the YAC library with probe Y1.22 (**Section 3.7**). That this sequence is present in Ya, is demonstrated by its presence in Yλ122, a λ clone isolated from the Ya genomic library. Probes derived from the ends of Ya hybridise to the 'wrong chromosome' in a karyotypic profile. To what chromosome does probe Y1.22 hybridise?

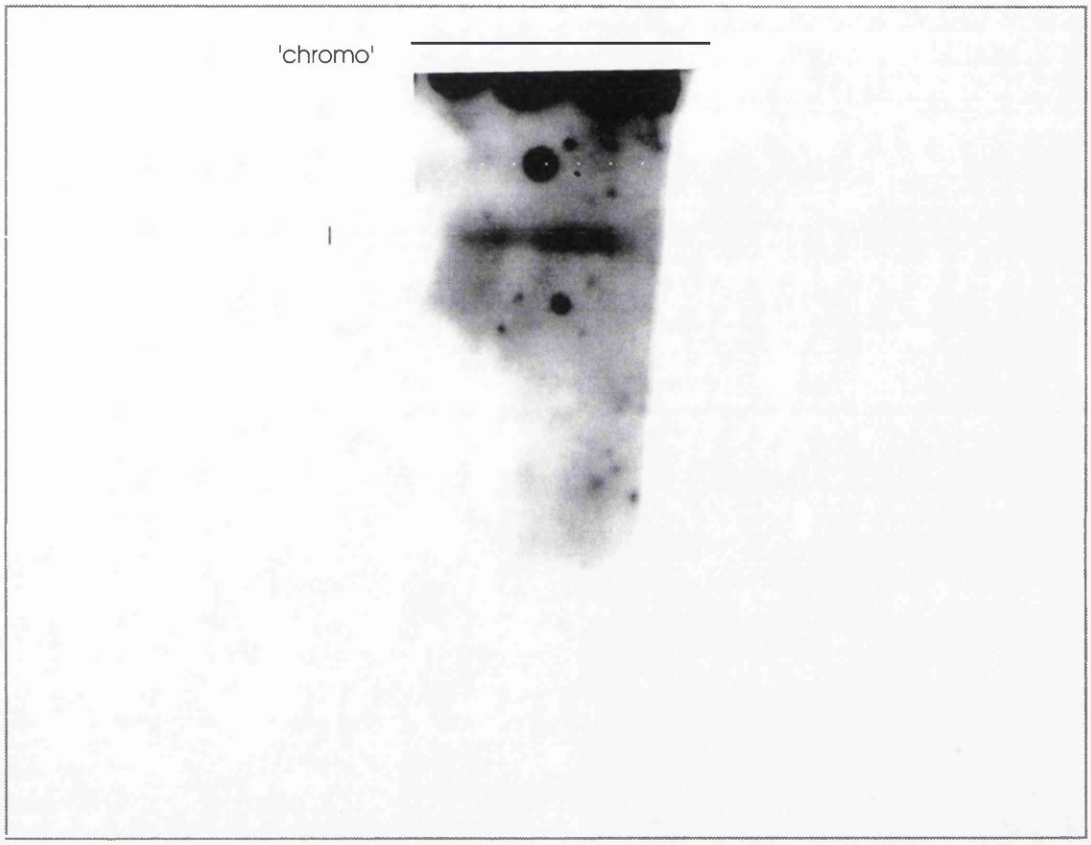


Figure 4.10 Assignment of the probe Y1.22 to a karyotypic unit. This is an autoradiogram, processed in Photoshop (Adobe), of a hybridisation (h.s.) of the remaining Y1.22 fragment onto a karyotype blot of EATRO 795 chromosome-length DNA. The karyotype gel had quadruplicate loading and was run using the parameters reported in Fig. 4.1. Along the side of the gel 'chromosome numbers' are indicated (see section 4.2)

As with probes 1 and 6 (preceding section); Y1.22 hybridises to karyotypic unit I, not to unit VII where the 1.22BC resides. Ya being from band I is a consequence of probe Y1.22 being from band I. This rules out gross chimaerism in the Ya clone as the explanation for having cloned the wrong fragment. Suspicions of YAC chimaerism are due to the fact that the YAC cloning was the most technical step in the project. Uncertainty about the quality of the YAC library persists even after gross chimaerism for the Ya clone is ruled out. It is necessary to analyze the quality of the YAC library before it can be used as a primary resource. **Chapter 5** documents the results of YAC library characterisation.

4.11 Recloning the 1.22 locus

To explain why probe Y1.22 comes from band I rather than band VII in a karyotypic profile, a test for chimaerism is required, at a different resolution. As probe Y1.22 is supposed to be derived from λ1.22B; is this clone discontinuous? Re-examining the λ1.22B clone and finding something amiss would require recloning of the 1.22 locus. Therefore, in the interests of saving time, recloning of this region was undertaken to enable restriction map comparison.

The original cloning of the 1.22 locus into λ (λ1.22B; Shiels, 1990) was very difficult. This sequence was extremely rare in an otherwise representative library. The λ1.22B clone itself has shown no evidence of instability. The telomeric location of the 1.22 M-VSG gene should impact upon the representation of this sequence in a genomic library.

Taking heed of this previous difficulty I performed some preliminary experiments to help set parameters for screening density. The approximate number of *Sau* 3AI sites in the pMG 7.1-1 insert was determined to see if this cloning site is underrepresented. The insert of this clone is 5.2 kb; in a stretch of DNA of that length one would expect approximately 20 *Sau* 3AI sites. In fact there are only 7 to 9 *Sau* 3AI sites in the pMG 7.1-1 insert. The 1.22 locus recloning was undertaken from a λDASHII (Stratagene) EATRO 795 ILTat 1.2 genomic library constructed by ligating size-selected *Sau* 3AI inserts into *Bam* HI digested λ arms. The average insert size in the library is assumed to be 18 kb. Two of every 70 *Sau* 3AI sites are cleaved to produce fragments of this size. The telomeric end is effectively fixed and therefore determines the level of representation. There are only 7 to 9 *Sau* 3AI sites between the chromosome end and any 1.22BC-specific probe that might be used to screen the library. One would expect an 18 kb fragment containing this sequence to be underrepresented 3.8 to 5-fold in the library compared to a chromosome internal sequence. According to the formula of Clarke and Carbon (1976), to obtain a 99% probability of a sequence's representation in a screen of a *T. brucei* genomic library with 18 kb inserts, 10 000 pfu need to be screened. A fourfold underrepresentation was assumed based on the paucity of *Sau* 3AI sites. A newer formula reported by Zilsel *et al.* (1992) for determining library size yields estimates two to four times greater than the classical Clarke and Carbon formula. An eightfold underrepresentation for the 1.22BC locus compared to a standard sequence was therefore assumed for screening purposes and 80 000 pfu were plated.

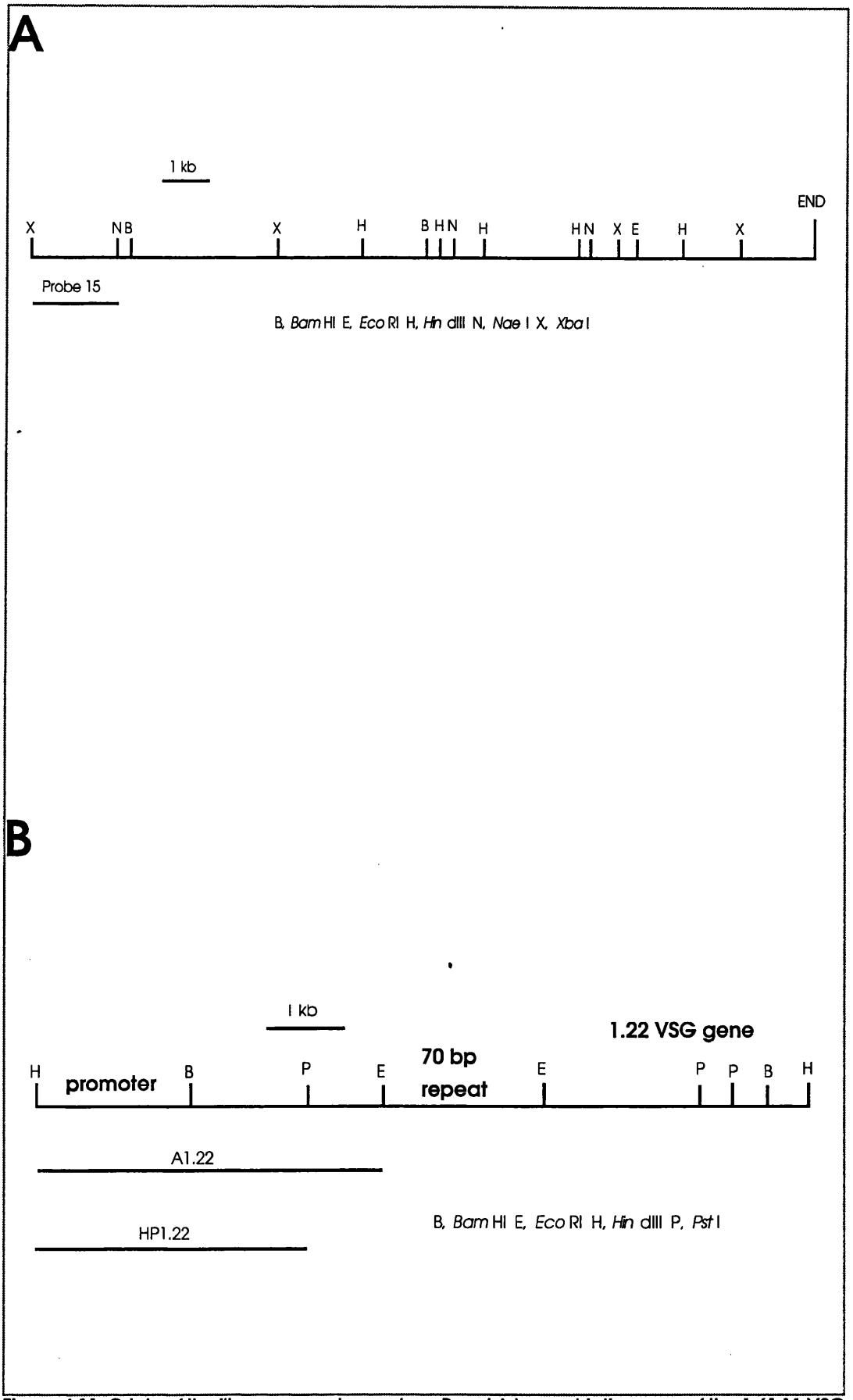


Figure 4.11 Origin of the library screening probes. Panel A is a restriction map of the 1.61 M-VSG BC locus (Matthews *et al.* 1990). Panel B is a restriction map of the insert of pMG 7.1-1.

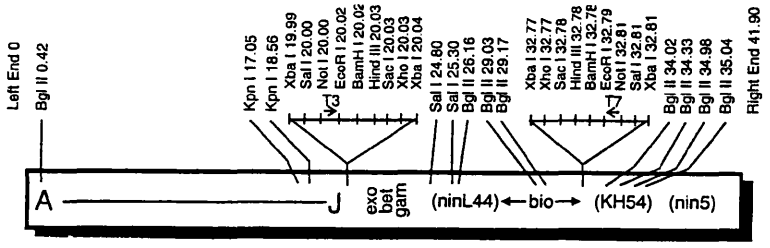
The next preliminary step was to determine the level of representation of a single copy sequence from the 1.61 M-VSG BC locus in this library. Probe 15 (see **Figure 4.11**) whose identity was first confirmed by genomic southern analysis was then used to screen the ILTat 1.2 library. 58 signals were obtained from 80 000 pfu or 1 signal/1380 pfu screened. Clearly, there are no problems with this sequence's representation in the library.

Probe HP1.22 (**Figure 4.11**), a promoter-containing *Hin* dIII/*Pst* I fragment of pMG 7.1-1 was used to screen the same 80 000 pfu as were screened with probe 15. A single positive signal was obtained. This area of the library plate was cored and the HP1.22 positive phage was plaque purified through two additional rounds of plating and screening. A λ blot of this clone (λMT1.22-2) was hybridised with HP1.22 to provide a quaternary level of confirmation.

4.12 Restriction mapping λMT1.22-2

The comparison of restriction maps has already been useful, in this chapter, for identifying problems (**Section 4.8**). In order to use this method again, restriction mapping data for λMT1.22-2 was sought. **Figure 4.12** shows the restriction map of the λDASHII vector used in the library construction and the tabulated results of restriction digests and hybridisations for λMT1.22-2. **Figure 4.13** presents the deduced restriction map.

A



B

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Bam HI	(kb)	12	7.6*	6.4	1.6	.3 (T3)						
2	Bam HI/Not I		7.6*	6.4	2.9 (T7)	1.6	0.3 (T3)						
3	Eco RI		8.6	5*	3.8 (T3)	1.1	.7 (T7)						
4	Eco RI/Bam HI		6.4	5*	1.6	1.55	1.2	1.1	.7 (T7)	.3 (T3)			
5	Hin dIII		>12	5.4	4.8	2.2*,	0.6	0.3	0.25				
6	Hin dIII/Not I		5.4	4.8	4.4 (T7)	2.2*	.75 (T3)	0.6	0.3	0.25			
7	Hin dIII/Bam HI/Not I		3.4	3.1	2.9 (T7)	2.2*	1.6	1.35	0.9	0.7	0.6	0.3	0.25
8	Hin dIII/Eco RI		4.8	2.8	2.4	2.2*	1.8	1.1	.7 (T7)	0.6	0.55	0.3	0.25
9													
10													
11													
12													

Figure 4.12 Restriction mapping data of λMT1.22-2. Panel A is a restriction map of the λDASHII vector. Panel B is a Table of restriction data for λMT1.22-2. For simplicity, only fragments that have insert DNA are included in the Table. B, *Bam* HI; E, *Eco* RI; H, *Hin* dIII; N, *Not* I. ^{T3} and ^{T7} indicate the end fragment identity. * marks fragments to which the 2.2 kb *Hin* dIII fragment hybridised. # marks fragments to which the 5.0 kb *Eco* RI fragment hybridised.

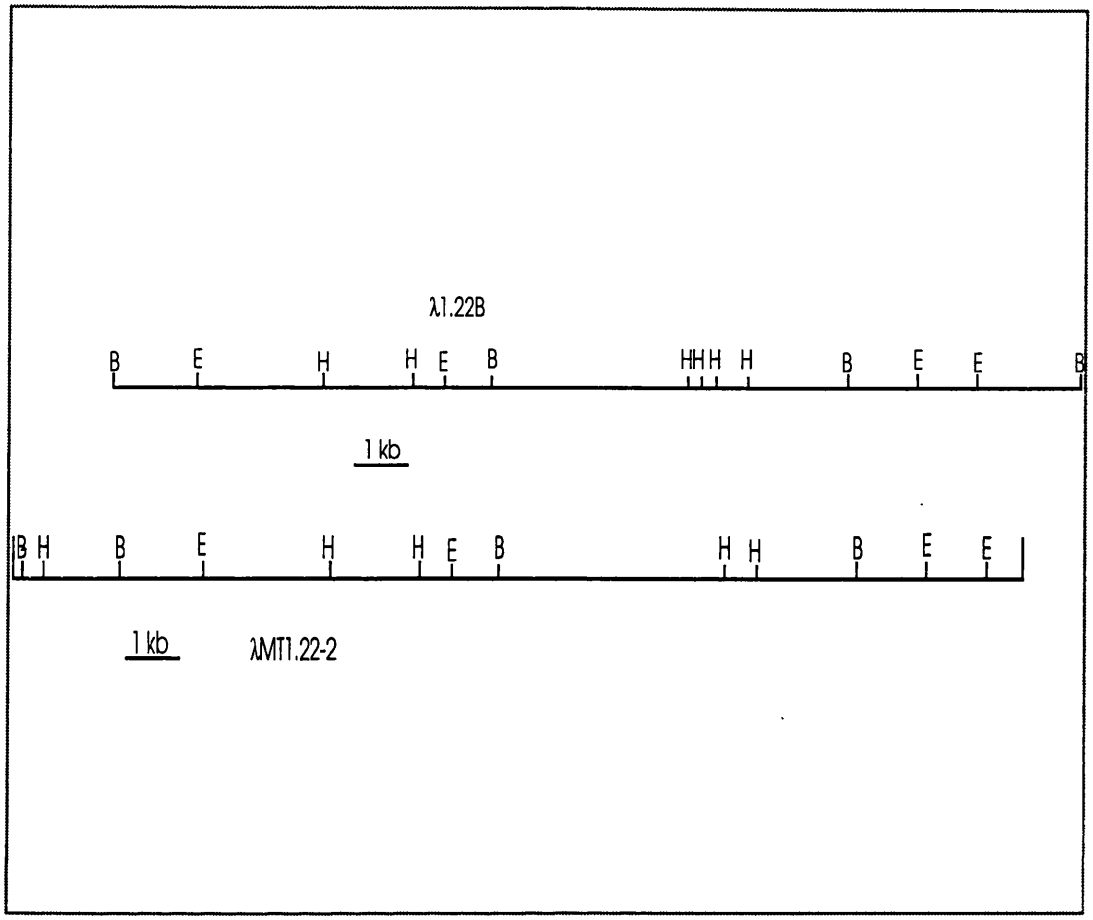


Figure 4.13 Restriction maps for λMT1.22-2 and λ1.22B. λ1.22B is the original λ clone of the 1.22BC locus and λMT1.22-2 is the result of recloning. B, *Bam* HI; E, *Eco* RI; H, *Hin* dIII; N, *Not* I.

The resolution of this mapping was unable to assign a location for the small *Hin* dIII fragments present in the 1.22BC locus. Aside from this, it is clear that these two clones overlap without any differences in the area of the overlap. This was confirmed by running the clones, digested with diagnostic enzymes, side by side, on an agarose gel. I conclude from the results of this section that λ1.22B is not chimaeric and that the probe Y1.22 is not derived from λ1.22B.

4.13 Identifying the transcription unit in Yλ122

Yλ122 was found (**section 4.7**) to harbour an active transcription unit that seemed to be highly expressed, multicopy and α-amanitin insensitive. This area was subcloned from Yλ122 into pBSK$^-$ (Stratagene) to facilitate a higher resolution characterisation. **Figure 4.14** summarizes the structure of the subclones and the nature of the transcription unit. Sequence tags were sought to determine the identity of the transcription unit. This is also indicated in **Figure 4.14**.

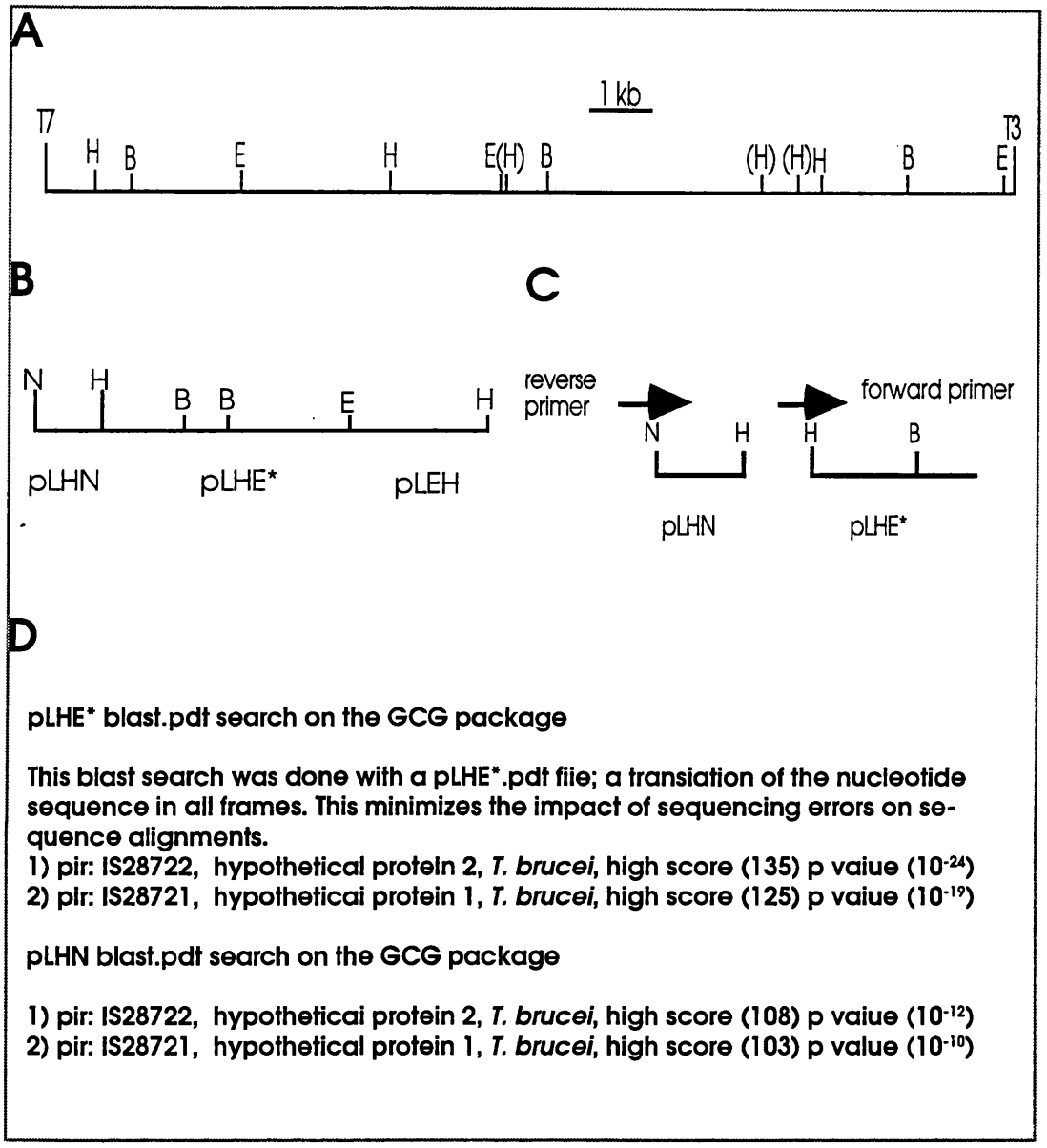


Figure 4.14 The Yλ122 transcription unit. Panel A is a schematic representation of the genomic region covered by the active transcription unit detected in Fig. 4.4. Panel B represents the 3 fragments that were subcloned into pBSK-. Panel C details the sequence tags that were obtained for the subclones derived from the Yλ122 clone, pLHN and pLHE* by using the indicated primers in dideoxy chain termination sequencing reactions (UBI Sequenase kit; manufacturer's protocol). Panel D shows the match produced by a BLAST search. The p value refers to the probability of such a match occurring by chance. B, *Bam* HI; E, *Eco* RI; H, *Hin* dIII; N, *Not* I. The *Not* I site is vector derived.

It can be concluded that this region contains a sequence related to the retroposon-like element of *T. brucei*, referred to as 'ingi' (Murphy *et al.* 1987). Ingi or TRS (Trypanosome Repeated Sequence) elements have two forms: TRS1 elements which are full length (5.2 kb) and TRS2 elements which are a truncated internal fragment of TRS1. TRS2 elements lack the antisense coding region within the ORF1 (the putative reverse transcriptase) referred to as ORF2. The sequence tags in **Figure 4.14** pick up strong homology to ORFs 1 and 2, indicating that this may be part of a full length, TRS1, element. The southern data

are consistent with this area being a TRS element. Ingi is present in the *T. brucei* genome in approximately 400 copies and the southern results in **Fig. 4.8** show that this sequence is abundantly represented in the genome. Murphy *et al.* (1987) reported that the transcription of TRS elements is moderately sensitive to inhibition by α-amanitin. The apparent lack of sensitivity documented in **Fig. 6.15** for this sequence can be explained by noting that the RNAP II tubulin control has decreased, but not down to background levels. Moderate α-amanitin sensitivity of an abundant transcription unit may not be detectable under these conditions.

4.14 Concluding remarks

This chapter has outlined the examination of the Ya YAC clone isolated with the Y1.22 probe. Due to a fragment being mislabelled or insufficiently examined, Chapters 3 and 4 have contributed nothing to the stated project aim of cloning and defining the extent of the ILTat 1.22 'metacyclic domain'. Despite this setback, this work may represent a modest achievement. The primary resource of a YAC library, the first of its kind in *T. brucei*, has been made. Experience has been gained in the analysis of YAC clones. The next chapter examines the quality of the YAC library to assess its worth as a primary resource.

Chapter 5

Characterising the YAC library

5.1 Introduction

The work in this chapter was conceived when it was believed that the insert of the Ya YAC clone was discontinuous. It was later discovered that there was no such gross cocloning artifact in this clone. I decided to proceed with the characterisation of Ya and Y1, as well as 20 randomly selected clones from the YAC library, to demonstrate the techniques involved and to assess the library's usefulness as a primary resource.

Certain issues should be addressed when characterising YAC clones in the absence of exhaustive probes to confirm the veracity of the genomic representation of the clone; stability, clonality and chimaerism. The stability of a YAC, as a complete unit, can be assessed genetically, by the colour colony assay described in **Section 3.1**. This involves determining the rate of chromosome loss, for the YAC, in the absence of auxotrophic selection. Most laboratory yeast strains, including those used for YAC work, are haploid, therefore such a colour colony assay will not be fully quantitative. A large number of white colonies (YACless) would suggest a high rate of mitotic loss of the YAC and prompt a search for a larger, more stable clone for long term manipulation. The stability of the insert, on the other hand, must be determined by karyotype analysis. The most frequent form of instability is the production of deletion derivatives from inserts with unstable, tandemly repeated sequence configurations. This shows up as multiple YACs in a karyotypic profile. Such instability is quite rare, occurring maximally at a rate of 1% in libraries from complex genomes (Vilageliu and Tyler-Smith, 1992). This problem is certain to be much rarer in the comparatively simple trypanosome genome. Multiple YAC bands in a karyotypic profile could also indicate a lack of clonality. Colony purification and reprobing with insert and vector arm probes should distinguish the trivial problems of mixed clones and a multiplicity of YACs, from the serious case of insert instability.

A quick method exists for the detection of subtle forms of chimaerism that can occur around the vector/insert junctions. **Figure 5.1** details the basis for detecting abnormalities at the vector/insert junctions.

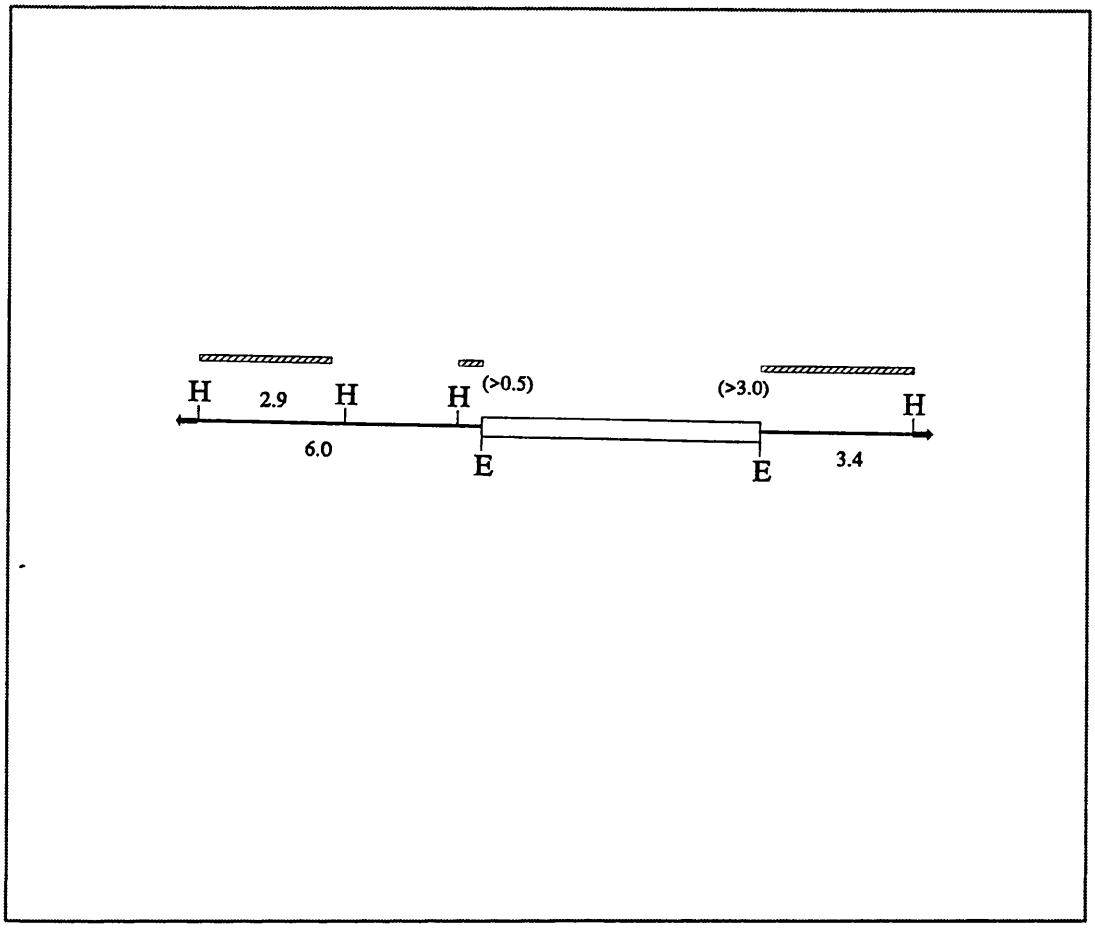


Figure 5.1 Detection of vector/insert junction abnormalities. This is a simplified restriction map of the pYAC4 vector taken from Den Dunnen *et al.* (1992). The arrowheads represent telomere tracts, the open box an insert and the hatched bars above the map indicate fragments that will hybridise to pBR322 sequences. Southern blots of *Eco* RI and *Hin* dIII digested yeast genomic DNA, when probed with pBR322 sequences, will assess the fidelity of the cloning and propagation of the insert DNA around the junctions with the vector arms. Abnormalities are detected as extra bands or as bands with altered mobilities. E, *Eco* RI; H, *Hin* dIII. The fragment sizes are in kilobase pairs (kb).

The key advantage of this method is that it is applicable to all YAC clones; no insert or trypanosome probes need be available for this analysis to proceed.

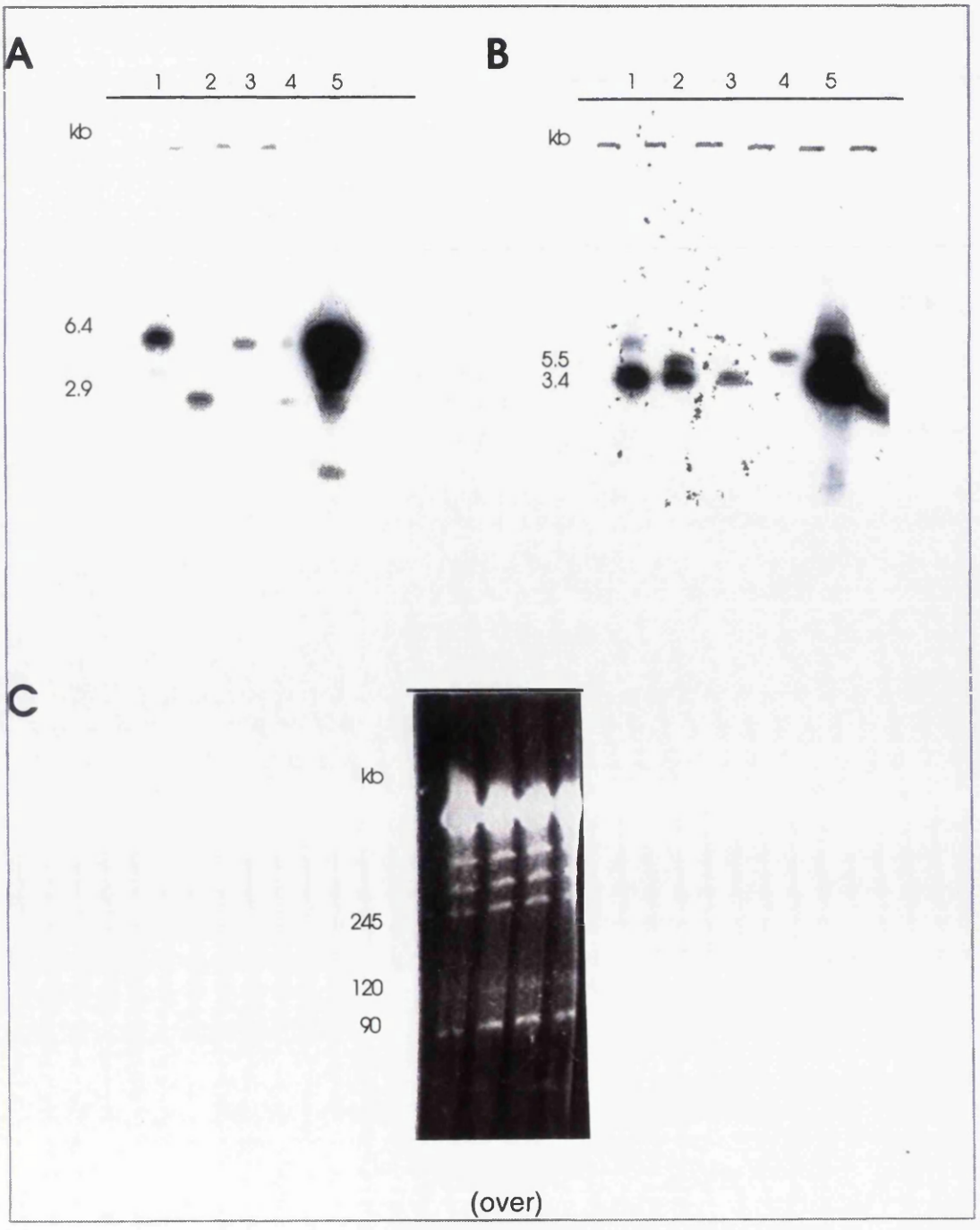
5.2 Examining Ya and Y1

Ya and Y1 were investigated to see if any unusual mitotic instability could be detected. This was done by looking for white, YACless, colonies when the clones were grown on a complex, nonselective medium; YPD. Because the YPH 252 strain used is haploid, this test cannot be quantitative. To be quantitative, it must be able to detect first division loss events (red/pink and red/white sectoring of the colony) and this requires a single copy of a suppressor in a diploid background. This test, however, should show up any obvious stability problems and predicts fewer white colonies for the larger Y1 than for Ya.

Glycerol stocks of Ya and Y1 were streaked onto plates of the selective medium, YCD, to obtain single colonies (30°C; 2 days). Single colonies were used to inoculate overnight YCD cultures. Spreading 10 µl of each of these cultures

onto YPD plates (nonselective) produced very densely populated, but not confluent, plates after incubation (30°C; 2 days). The plates were then stored at 4°C for one day to allow the red pigmentation of the colonies to develop. Inspection of the Ya plate found 16 white colonies while only 2 were found on the Y1 plate. Loss of the YACs, upon nonselective growth, was found to be rare. As predicted by the larger size of the Y1 clone, this YAC seemed more mitotically stable than Ya.

Next, Ya and Y1 were examined for abnormalities around the vector/insert boundaries by the protocol described in **Fig. 5.1**. **Figure 5.2** presents the results obtained when *Eco* RI and *Hin* dIII digested Ya and Y1 DNAs were southern blotted and hybridised with left and right arm-specific probes.



(over)

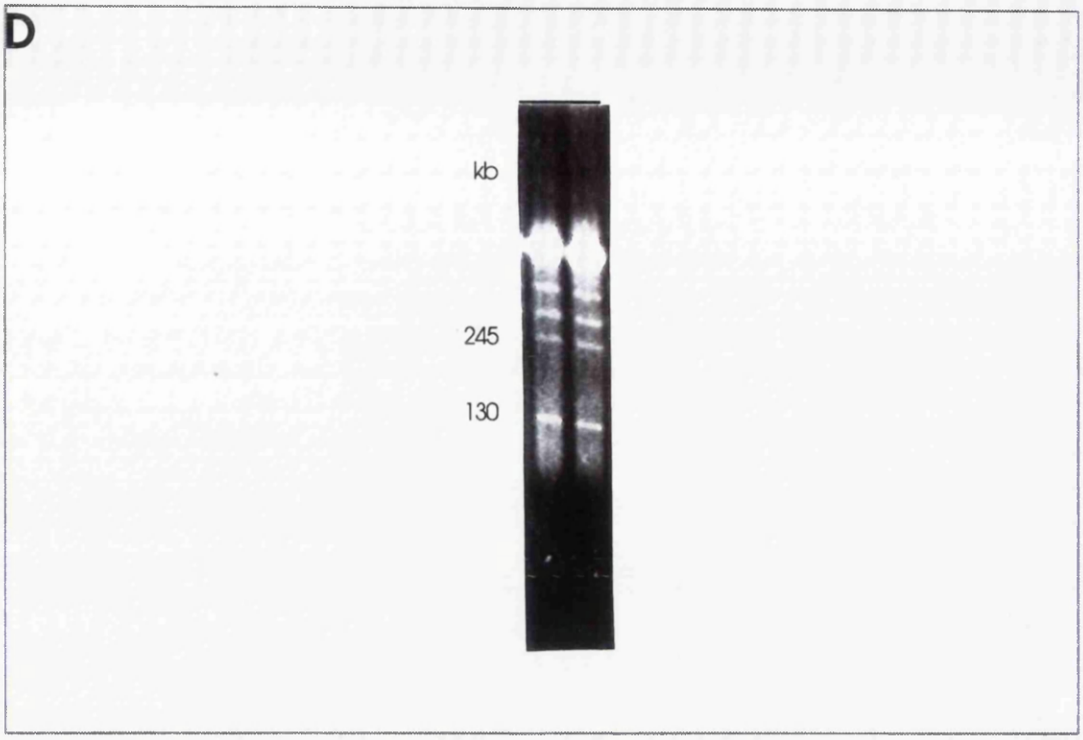


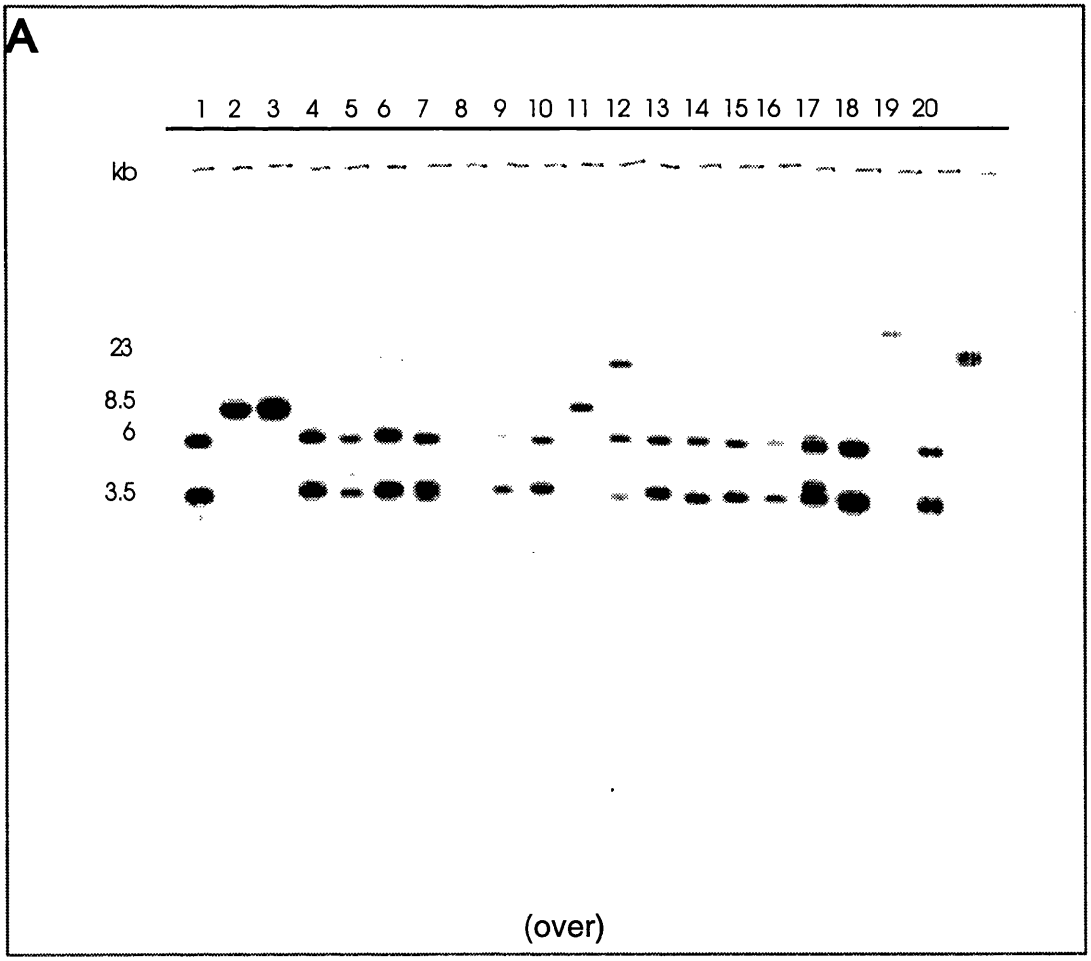
Figure 5.2 Examination of the vector/insert boundaries of Ya and Y1. Panel A is an autoradiogram of a hybridisation (h.s.) of the left arm probe (2.5 kb *Pvu* II/*Bam* HI fragment of pBR322) onto a southern blot of the yeast clones. Panel B is an autoradiogram of a hybridisation (h.s.) of the right arm probe (1.6 kb *Pvu* II/*Bam* HI fragment of pBR322) onto a southern blot of the yeast clones. Panel C is a karyotype gel, with a quadruplicate loading, of the Ya clone. Panel D is a karyotype gel, with a duplicate loading, of the Y1 clone. Lane 1, *Eco* RI digested Ya genomic DNA; lane 2, *Hin* dIII digested Ya genomic DNA; lane 3, *Eco* RI digested Y1 genomic DNA; lane 4, *Hin* dIII digested Y1 genomic DNA; lane 5, *Bam* HI/*Eco* RI digested pYAC4 vector DNA; lane 6, chromosome-length wild type yeast DNA. The karyotype gels are 1 % SeaPlaque agarose (FMC) in 1XTB0.1E running buffer run on a Biometra PFGE system. Interval: 30-3 sec. linear ramp; Angle: 120-95° linear ramp; Voltage: 10V/cm; Temp.: 13°C; Duration: 30 hrs.

A superficial examination of **Fig. 5.2** appears to indicate that there are abnormalities at the vector/insert boundaries of Ya; at both arms! It should be noted that lane 1 of panels A and B are reverse images of each other, indicating some cross contamination between the *Pvu* II/*Bam* HI fragments of pBR322 used as probes in this experiment. This is not evident in the Y1 clone, as it is underloaded with respect to the Ya lanes. The probes were gel purified; they should have been doubly gel purified to avoid this problem. Lane 2 of panels A and B show two *Hin* dIII fragments. This probing was not sensitive enough to detect the >0.5 kb fragment expected in the left arm hybridisation experiment (lanes 2 and 4; panel A). The presence of a second band here, as in the right arm experiment, represents a possible anomaly. Panel C gives an indication of where this anomaly may come from; the presence of a second YAC in the Ya clone. This probably represents a clonality problem, as this second YAC's presence has become quite pronounced in a stock that appeared, at the outset, to be clonally pure. The second YAC would give rise to *Hin* dIII fragments of different sizes than Ya, since this enzyme is used to detect clone-specific fragments. These

problems are not evident in the apparently pure, Y1 clone (panel D). Certainty is denied by the obfuscation caused by the probe and clonality problems in this experiment; but Ya and Y1 appear to be free of any significant structural abnormalities at the vector/insert boundaries.

5.3 Examination of 20 clones from the YAC library

In order to assess the quality of the YAC library, 20 clones were selected at random and subjected to karyotype and southern analysis. To avoid the problems encountered in **Section 5.2**'s characterisation, the clones were grown in overnight cultures, streaked on YCD plates and single colonies were isolated. These colonies were then used to inoculate new overnight cultures, which were divided for the preparation of genomic DNA and agarose embedded chromosome-length DNA, as described in Sherman *et al.* (1986) and **Section 2.5.3**, respectively. This should eliminate clonality problems. Southern analysis was conducted only on *Eco* RI digested DNA samples, as the interpretation of this result is unaffected by the presence of multiple YACs within a particular clone. Lastly, the experiment was conducted by pooling the left and right arm probes to derive this information simultaneously. **Figure 5.3** presents the results of this library characterisation.



(over)

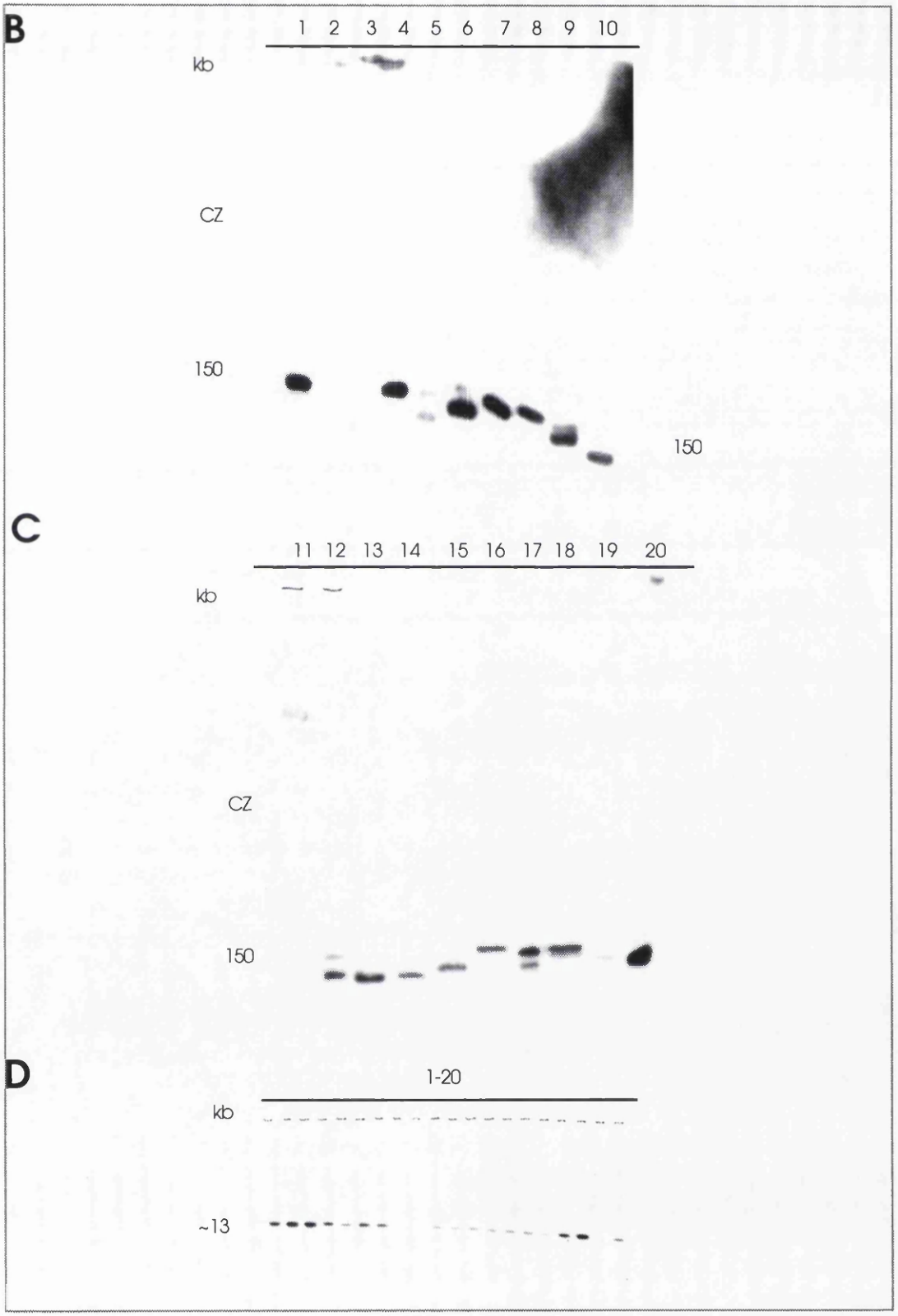


Figure 5.3 Southern and karyotype analysis of 20 YAC clones. Panel A is an autoradiogram of a hybridisation (h.s.) with the pooled left and right arm probes, onto a southern blot of *Eco* RI digested DNAs prepared from clones 1-20. Panels B and C represent autoradiograms of the same pooled probes hybridised (h.s.) onto karyotypic profiles of clones 1-20. Panel D is an autoradiogram of a control hybridisation (h.s.) of the yeast *HIS3* gene, onto the stripped filter used in panel A. The *HIS3* gene was obtained from the 1.8 kb *Bam* HI fragment of pYAC4. The lane numbers indicate the clone numbers. The karyotype gels were run as described in Fig. 5.2. CZ marks the position of the compression zone on the PFG gels.

It is clear from this characterisation that there are some problems with the YAC library; 13 of 20 clones are not typical. Clone 3 produces a diffuse hybridisation signal above the compression zone, in the karyotype blot. The southern analysis reveals a single, approximately 9 kb, fragment rather than the 6.0 and 3.4 kb bands expected. Taken together, these results strongly suggest that clone 3 represents a religated, stufferless, derivative of the pYAC4 vector. The unusual result in the PFG blot for this clone is probably due to aberrant migration of a supercoiled molecule in a pulsed electric field. The same result was observed for 'clones' from the failed telomere library (**Section 3.5**). Clone 5 is an example of a clone carrying more than one YAC. In the southern data for clone 5 an extra, anomalous band of 4.2 kb is apparent. This extra band represents an additional vector fragment, probably indicating a recombination generated double insert. These anomalies usually involve the right vector arm (Den Dunnen *et al.* 1992). Clone 5 is typical; where there is more than one YAC present in the clone, the southern data usually show evidence of a structural defect around the vector/insert boundary. Conversely, the clones that have only one YAC usually show no evidence of such defects. Some exceptions to these generalizations exist. Clones 10 and 18 have two bands in their karyotype profiles but show no evidence of abnormality in the southern blot. Clone 19 has a single YAC in its karyotypic profile but this YAC is clearly abnormal.

The aberrant bands in the southern blot result from real structural defects in the clones. Partial digestion of the DNA samples has been ruled out by the *HIS3* control hybridisation documented in panel D. Another possible form of structural defect that would be detected by the southern analysis is the ablation of one of the *Eco* RI sites that accept the insert. This would be manifested as a mobility shift, presumably significant, from either the expected 6.0 or 3.4 kb band on the southern blot. Small variations in mobility, from lane to lane, for these two bands are evident in panel A. I believe that these do not represent ablated *Eco* RI sites but rather reflect variations in telomere length and/or in the salt content between the different DNA samples. The Sherman *et al.* (1986) genomic DNA isolation protocol involves a 'rough and ready' salt precipitation that would readily cause this particular problem.

In summary, this characterisation has revealed that of the clones sampled:
1) 15% (3/20) were the products of illegitimate religation of insertless vector into an episome. The red colour of the colonies, here, indicates that the *Eco* RI site has been disrupted in an illegitimate ligation reaction (recall that the vector preparation featured a dephosphorylation reaction; **Section 3.6**).
2) of the genuine YAC containing clones, 53% (9/17) had 2 YACs present
3) of the multiple YAC containing clones, 78% (7/9) showed evidence of a double insert involving a vector arm, in at least one of the YACs
4) of the clones carrying a single YAC, 87.5% (7/8) showed no evidence of such structural aberrations

It is surprising that so many of the clones sampled are insertless or afflicted

with structural defects. The very tight correlation between these structural aberrations and the occurrence of multiple YACs within a clone suggests that, in cells that accepted more than one YAC, transformation induced the DNA repair machinery, leading to a high rate of recombination around the vector/insert junctions. This speculation is supported by the studies that have reported transformation induced instability of YACs (**Section 3.2**). This problem is probably avoidable. Using less ligation mix per aliquot of YPH 252 spheroplasts would reduce the incidence of double transformation and presumably the occurrence of structural defects.

As it stands, the library is still useful. The data obtained here emphasize that use of this type of library necessitates the initial screening of additional clones to obtain one that contains the target sequence in the desired configuration: unscrambled and unpartnered. It is strongly advised that several positives be carefully purified and subjected to a karyotype and southern analysis as described in this section, before choosing a clone for further use. Should it prove impossible to obtain an unpartnered clone, it is still possible to work on the YAC, provided it is unscrambled, by isolating it from the other YAC in the cell. Two options exist, that will appeal to different kinds of investigator, to eliminate the undesired ' partner YAC'. The geneticist would want to mate the clone to a YACless YPH 250 strain (opposite mating type from YPH 252) sporulate the diploid product and regenerate haploid lines, selecting the one with the desired YAC for further work. The molecular biologist would prefer to gel isolate the desired YAC and retransform it into the YPH 252 strain used to construct the library.

Chapter 6

Defining the ILTat 1.22 metacyclic domain

6.1 Introduction

The aim of this project is to clone and define the extent of the metacyclic domain for the 1.22 M-VSG BC locus. **Chapters 3** through **5** document the initial efforts to fulfil this objective. Due to the problems with the probe Y1.22 used to start the cloning, **Chapter 4** contributed nothing to the stated project aim. The problem discovered in that chapter was eventually traced to a 'rogue' fragment used to isolate the initial YAC clones. **Chapters 3** and **5** stand on their own as documentations of the work required to construct, establish and characterise the resource of a YAC library.

The cloning effort must be restarted. How should this be done? The option of going back into the YAC library still exists, but several factors make this option unpalatable. It is now known that a lot of work would have to be done to select a YAC clone that is suitable for further study. The results of **chapter 5** demonstrate that careful purification and subsequent southern and karyotype analysis, on several positive clones, is necessary to select a clone with the correct configuration. Only then could the process of subcloning and analysis begin. Considerable time would pass before the clones could be verified.

When the work on **chapter 5** was in progress, V.S. Graham discovered that the 1.61 BC locus, contrary to prior results, was transcriptionally silent over a shorter stretch of DNA than previously supposed. Transcription was found to commence approximately 15 kb upstream of the 1.61 M-VSG BC gene (V.S. Graham PhD. thesis in preparation).

Due to time constraints and the changed expectations regarding the size of the silent domain, contig assembly from a YAC is no longer viewed as a sensible approach to employ in order to clone the 1.22 domain. In light of the findings for the 1.61 locus, obtaining a single overlapping λ clone (overlapping with λMT1.22-2 and λ1.22B) could be all that is required to clone the 1.22 metacyclic domain. Therefore, I set about to obtain such a clone. The initial experiences of walking upstream of the 1.22 and 1.61 metacyclic telomeres had been impeded by clone instability. V.S. Graham and I experienced no such instability with our λ work on these telomeres. This is doubtless the result of the superior *E. coli* strains that are now available for λ work.

6.2 Cloning more of the 1.22 locus

Probe R5 (**Fig. 6.1**), a single copy 1.22-specific fragment, was used to screen 20 000 pfu of the ILTat 1.2 genomic λ library described in **section 4.11**. This fragment required careful double gel purification to abolish cross hybridisation, on a southern blot, to the arms of λ and to ingi sequences. Once a suitably pure R5 fragment was obtained this was used to screen the library. Several positive signals were obtained without difficulty.

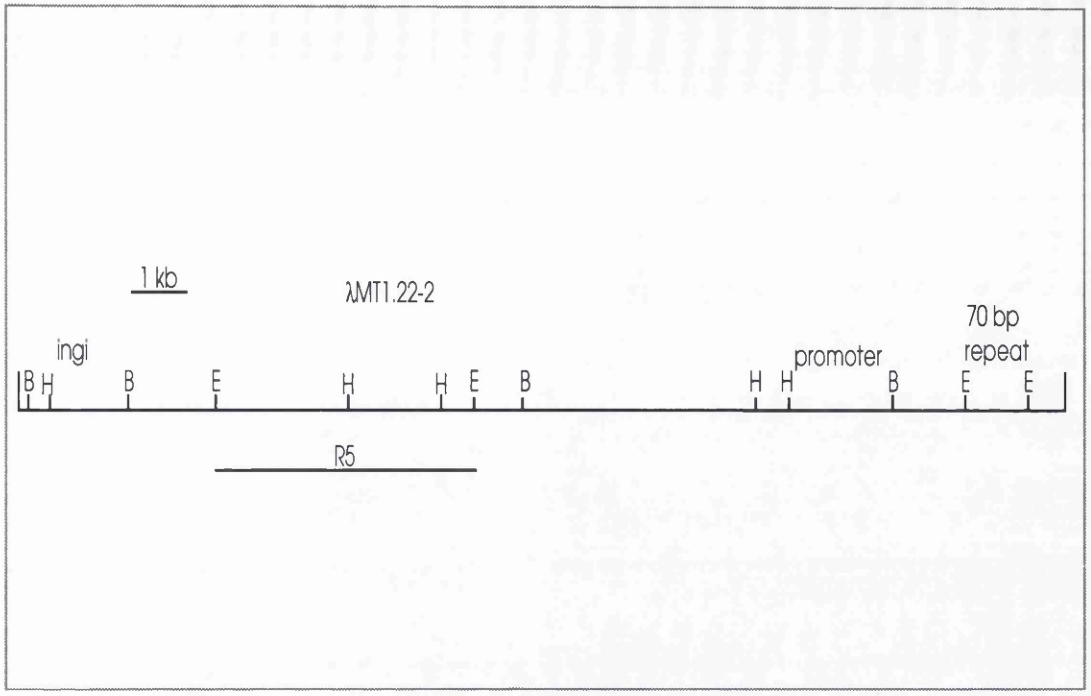


Figure 6.1 Origin of the library screening probe. R5 was derived from λMT1.22-2, the result of the recloning effort (see section 4.11).

A few of the R5 positive regions on the primary plate were cored and the putative R5 clones were plaque purified through 3 additional rounds of plating and screening (unusually, a 4° screen was required to obtain 100% purity). An extra level of confirmation was sought by fingerprinting the putative clones against λMT1.22-2 (*Bam* HI digests).

The agarose gel separation of this diagnostic digest for the new 1.22 clone (λMT1.22-3) and λMT1.22-2 is presented in **Fig. 6.2**.

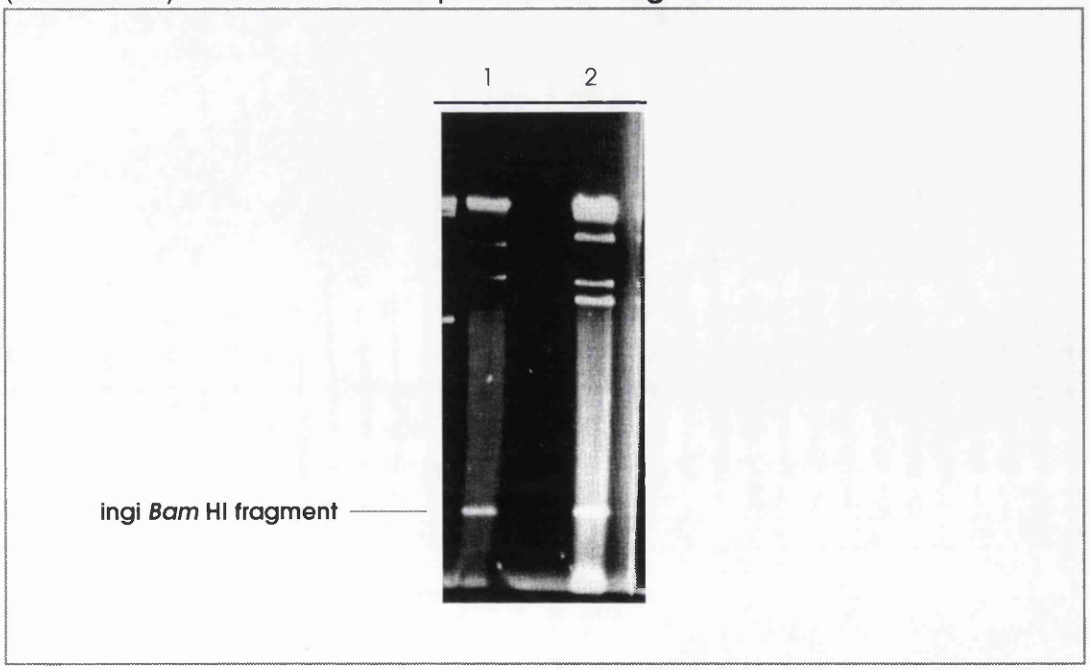


Figure 6.2 Confirming the overlap between λMT1.22-3 and λMT1.22-2. This is an ethidium bromide stained gel documenting the overlap. Lane 1, *Bam* HI digested λMT1.22-3 DNA; lane 2, *Bam* HI digested λMT1.22-2 DNA. The gel is 0.7% agarose in 1XTBE running buffer.

Unexpectedly, only two of the four clones tested had the appropriate overlap and contained new sequence. In anycase, λMT1.22-3 was found to overlap with λMT1.22-2 and appears to contain the most (about 10 kb) new sequence. λMT1.22-3 was examined further.

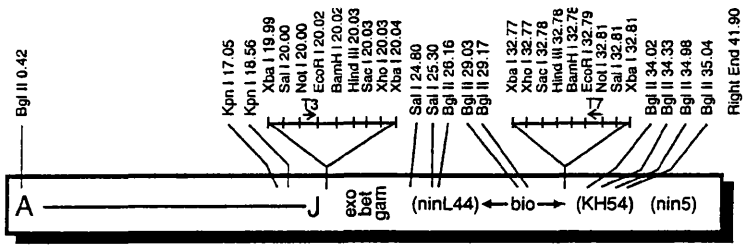
6.3 Restriction mapping λMT1.22-3

A restriction map for λMT1.22-3 is required for the complete analysis of this clone. This would aid, in the first instance, the determination of the extent of overlap with λMT1.22-2 and, hence, the amount of new sequence represented in this clone. **Figure 6.3** shows the restriction map of the λDASH II vector used in the library construction and the tabulated results of the restriction digests for λMT1.22-3.

Due to the extent of the overlap and the simplicity of the restriction patterns, no hybridisations were necessary to obtain a complete restriction map for this clone.

Figure 6.4 presents the deduced restriction map of the 1.22 locus.

A



B

	A	B	C	D	E	F	G	H	I	J	K
1	Bam HI	(kb)	>23	12	7.5	1.6					
2	Bam HI/Not I		8.8	7.5	1.6	1.3					
3	Eco RI		10	5	2.35	2.3					
4	Eco RI/Bam HI		6.4	5	2.35	1.6	1.55	1.3	1.1		
5	Hin dIII		>23	13.5	4.9	3.9	2.2	1.35			
6	Hin dIII/Not I		5.4	4.9	3.9	3	2.2	1.35			
7	Hin dIII/Bam HI/Not I		5.4	3.2	2.4	2.2	1.65	1.35	1.3	0.9	0.68
8	Hin dIII/Eco RI		3.85	2.7	2.4	2.3	2.2	2	1.7	1.35	

Figure 6.3 Restriction mapping data for λMT1.22-3. Panel A is a restriction map of the λDASH II vector. Panel B is a table of restriction data for λMT1.22-3. For simplicity, fragments that contain only vector DNA have been excluded from the table. B, *Bam* Hi; E, *Eco* RI; H, *Hin* dIII; N, *Not* I. [T3] and [T7] indicate end fragment identity.

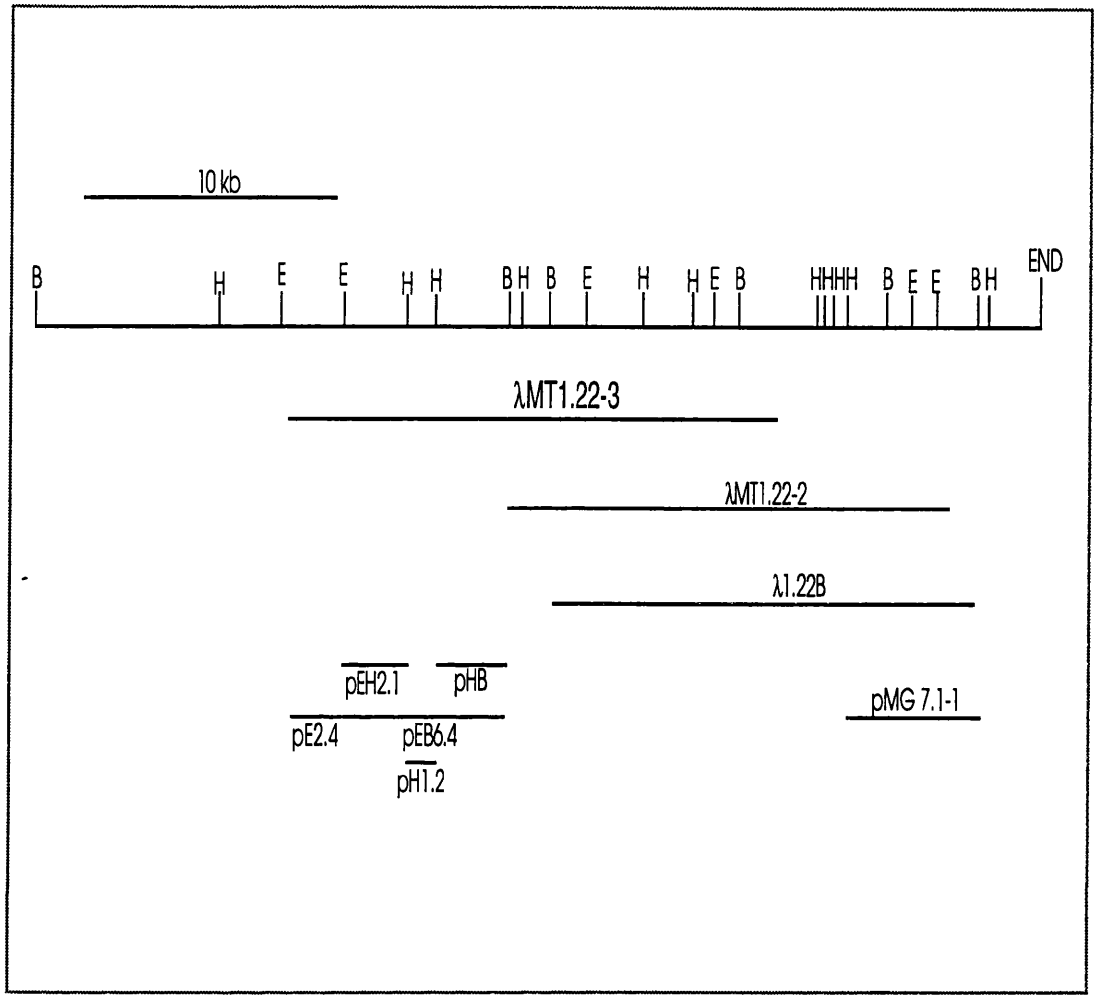


Figure 6.4 Updated restriction map of the ILTat 1.22 M-VSG BC locus. The extent of the contributing clones is indicated beneath the map. The parts of the map represented here that extend beyond this were obtained by genomic mapping. B, Bam HI; E, Eco RI; H, Hin dIII.

λMT1.22-3 contains 9.5 kb of new sequence. This should bridge the TRS element (ingi) that constitutes the upstream extreme of the λ1.22B and λMT1.22-2 clones and may complete the 1.22 metacyclic domain. This extends the cloned portion of this telomere to -27.5 kb from the start of the 1.22 M-VSG gene.

6.4 Subcloning novel 1.22 sequences/and verification of the clone inserts

Naturally, further analysis almost exclusively concentrated on the new sequence represented in λMT1.22-3. This sequence was subcloned to remove it from the previously characterised sequences in the λ. This was especially important as it had proven difficult in previous preparative gel runs of the λ clones to remove contaminating ingi sequences from the desired fragments.

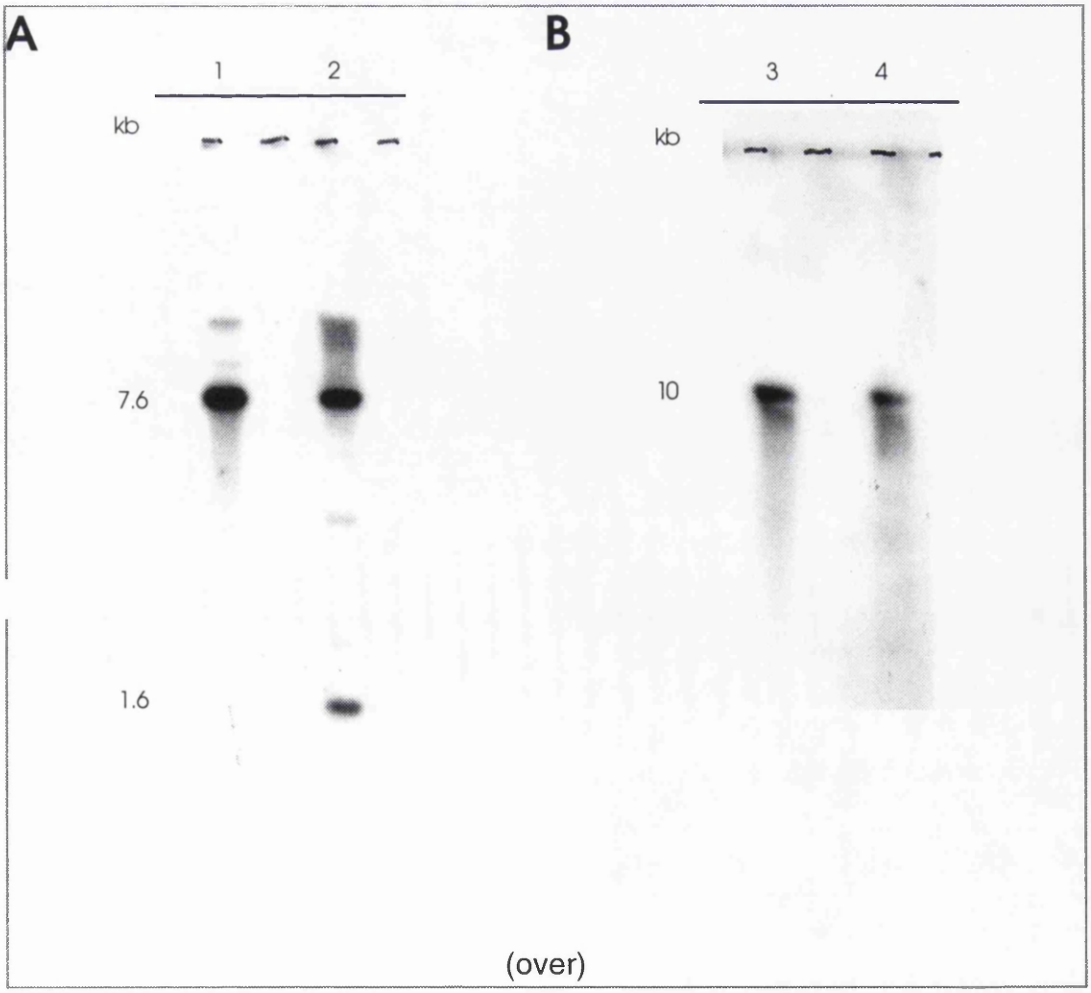
Initially, two subclones were obtained; a large *Eco* RI/*Bam* HI fragment of 6.4 kb and an *Eco* RI fragment of 2.4 kb in size. These were subcloned into pBSK$^-$ (Stratagene). The extreme upstream 2.4 kb *Eco* RI subclone (pE2.4) has one of its *Eco* RI sites derived from the polycloning site of the λDASH II vector. The large subclone, with the remainder of the sequence (pEB6.4) is expected to have

ingi sequences in the 2.9 kb *Bam* HI/*Hin* dIII fragment. Therefore, the remaining 1.35 kb *Hin* dIII and the 2.1 kb *Eco* RI/*Hin* dIII fragments were subcloned, again into pBSK⁻ to yield separate clones referred to as pH1.2 and pEH2.1, respectively. These plasmid substrates were prepared for their eventual use in transcription studies.

The location of these plasmid subcloned fragments is outlined in **Fig. 6.4**.

Next, the fidelity of all this cloning work was assessed by comparing the clones with the trypanosome genomic context from which they came. In order to include pE2.4 in this analysis it was first necessary to map a restriction site within the insert to compare with the genomic context (recall that one of the *Eco* RI sites was vector derived). pE2.4 was mapped with four enzymes that do not cut the pBSK⁻ vector: *Bgl* II, *Hpa* I, *Nco* I and *Sph* I. Of these, only *Nco* I was found to recognise a site in the insert of pE2.4. An *Eco* RI/*Nco* I digest fractioned the insert into two fragments, sized 1.6 and 0.8 kb. The 1.6 kb fragment was used in the veracity experiments and proved to be the *Eco* RI/ *Nco* I fragment represented in the genome (*i.e.* neither site was vector derived). This established the orientation of the insert and defined the position of the *Nco* I site (**Fig. 6.4**).

Figure 6.5 documents the results of the veracity experiments.



(over)

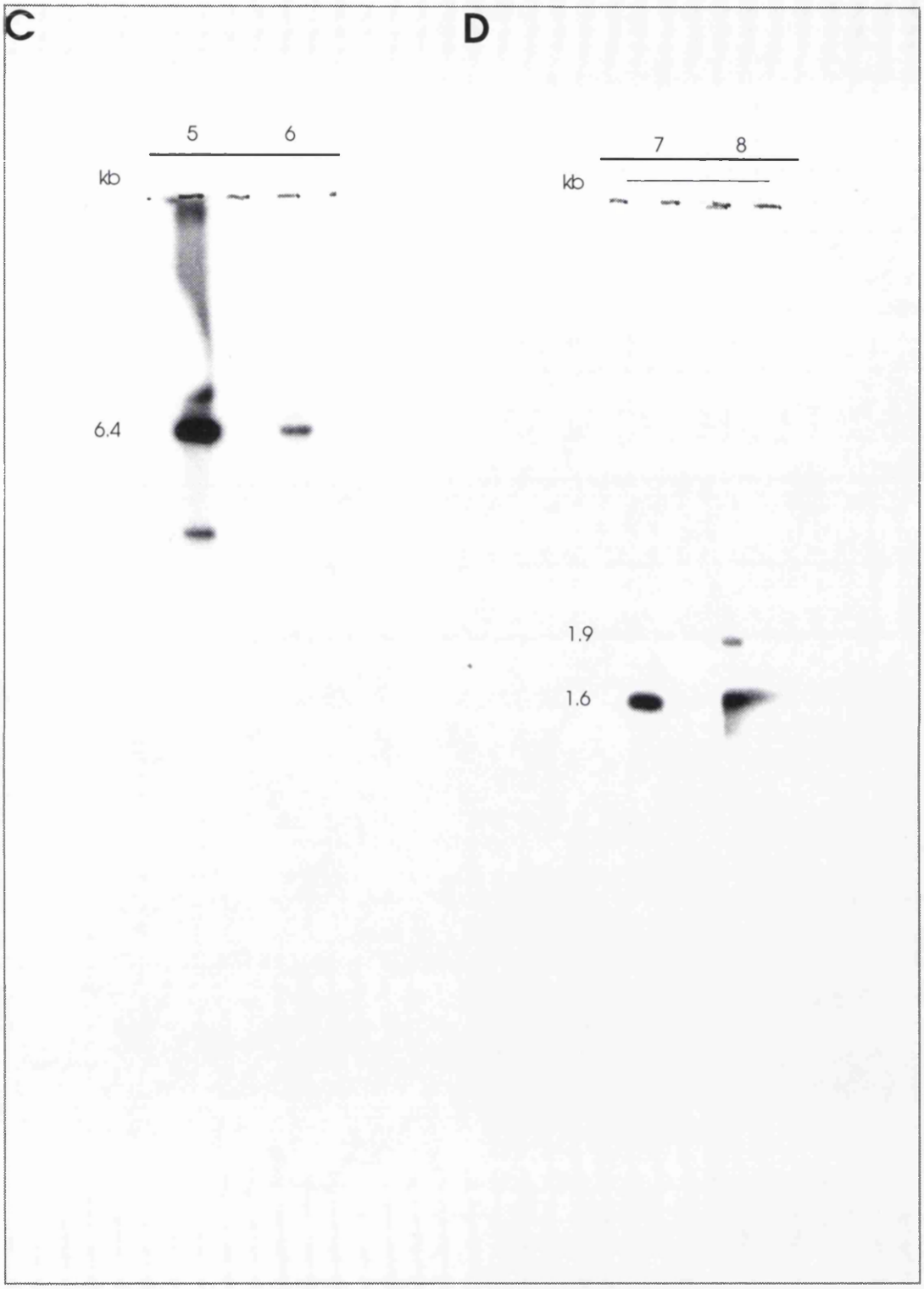


Figure 6.5 Assessing the fidelity of the 1.22 cloning. Panel A is an autoradiogram of the results of a hybridisation (h.s.) of the R5 fragment onto a southern blot of λMT1.22-3 and ILTat 1.2 DNA. Panel B is an autoradiogram of a hybridisation (h.s.) of the 1.2 kb *Hin* dIII fragment of pEB6.4 onto a southern blot of λMT1.22-3 and ILTat 1.2 DNA. Panel C is an autoradiogram of a hybridisation (h.s.) of the same probe used in panel B onto a southern blot of pEB6.4 and ILTat 1.2 DNA. Panel D is an autoradiogram of a hybridisation (h.s.) of the 1.6 kb *Eco* RI/*Nco* I fragment of pE2.4 onto a southern blot of pE2.4 and ILTat 1.2 DNA. Lane 1, *Bam* HI digested λMT1.22-3 DNA; lane 2, *Bam* HI digested ILTat 1.2 DNA; lane 3, *Eco* RI, λMT1.22-3; lane 4, *Eco* RI, ILTat 1.2 DNA; lane 5, *Bam* HI/*Eco* RI, pEB6.4; lane 6, *Bam* HI/*Eco* RI, ILTat 1.2; lane 7, *Eco* RI/*Nco* I, pE2.4; lane 8, *Eco* RI/*Nco* I, ILTat 1.2.

Panels A through C unfortunately show some evidence of contamination of the probes with ingi sequences. Nevertheless, the results of these experiments are still easily interpreted. At the resolution of this analysis, no inconsistencies are apparent. λMT1.22-3 and its plasmid subclones (pEB6.4 and pE2.4) appear to be faithful representations of the trypanosome genome. Panel D indicates that there is a polymorphism, in other copies of this sequence, with respect to the size of the *Eco* RI/*Nco* I fragment used in this analysis. This panel also gives a preliminary indication that this sequence is present in the genome in at least two copies (at high stringency).

6.5 Southern analysis of the new 1.22 sequences

The metacyclic domain has been defined in the Introduction as the portion of the telomere, at which a M-VSG gene resides, that is single copy and transcriptionally silent (excluding the VSG gene itself). The efforts to define the extent of this domain must include cloning, genomic characterisation and transcriptional mapping. The end of the metacyclic domain would be indicated by diploid (or at least multicopy) sequence that is transcriptionally active. This section details the genomic characterisation of the new 1.22 sequences represented in the subclones derived from λMT1.22-3. This kind of southern analysis yields information about copy number and linkage. The simultaneous derivation of southern data on different, carefully chosen trypanosome stocks, can contribute extra information on copy number and the degree of conservation of the sequences examined.

In this set of experiments, the newly-cloned sequences are probed onto DNA from the EATRO 795, EATRO 2340 and STIB 386 trypanosome stocks. EATRO 795 is the *T. brucei* stock from which the 1.22BC was cloned. The 1.22BC is present in one copy in this isolate. EATRO 2340 is a closely related (82% similar) *T. brucei* stock that has two copies of the 1.22BC locus. STIB 386 is a more distantly related (65% similar) stock of *T. brucei*. The rationale behind choosing these stocks is to have southern data for the 1.22BC from a diploid locus, the locus in single copy and from an unrelated stock that may be expected to lack these specific sequences. It was hoped that the STIB 386 stock would have the upstream sequences that define the limit of the metacylic domain in common with the other stocks examined. This allows a Southern comparison among DNAs in which the sequence is represented in the ratio 2:1: probable 0, which is the next best thing to having a 1.22 deletion mutant to examine (a 2:1:0 relationship).

Figure 6.6 is a dendrogram taken from Hide *et al.* (1991) that shows the relationship between the trypanosome stocks used in this analysis.

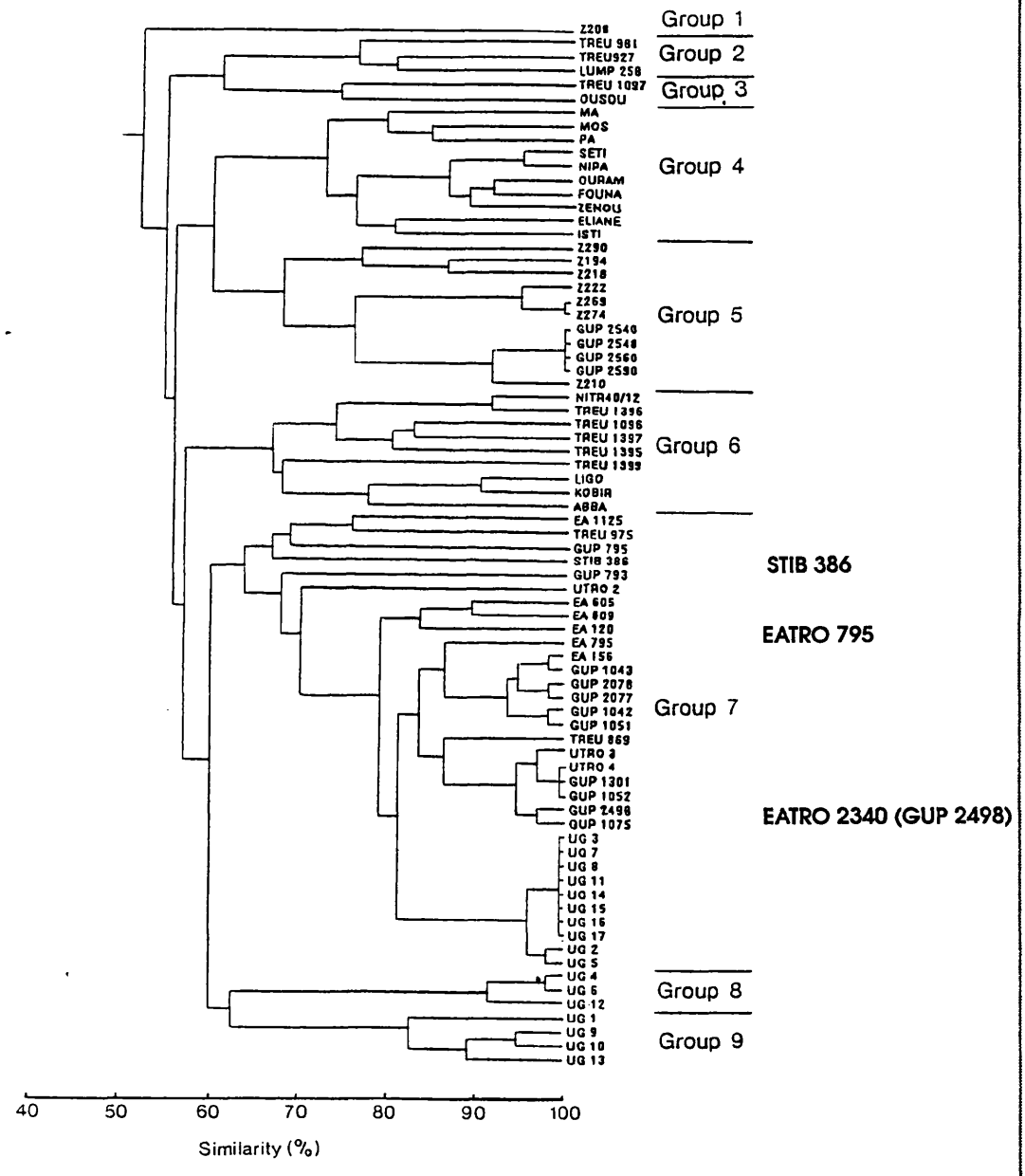


Figure 6.6 The relationship between the isolates used in southern analysis. The position of the stocks is indicated on the dendrogram (Hide *et al.* 1991).

The results of the southern analysis are presented from the telomere proximal sequences towards distal sequences.

In an effort to find an ingi-less sequence within pHB, the insert of this subclone was divided into two fragments by cleavage with *Pst* I (0.5 kb and 2.4 kb fragments resulted). Both fragments, as well as the whole *Bam* HI/*Hin* dIII insert, were used in southern hybridisation experiments. All probes produced the same result.

Figure 6.7 presents a representative example of the southern results for pHB.

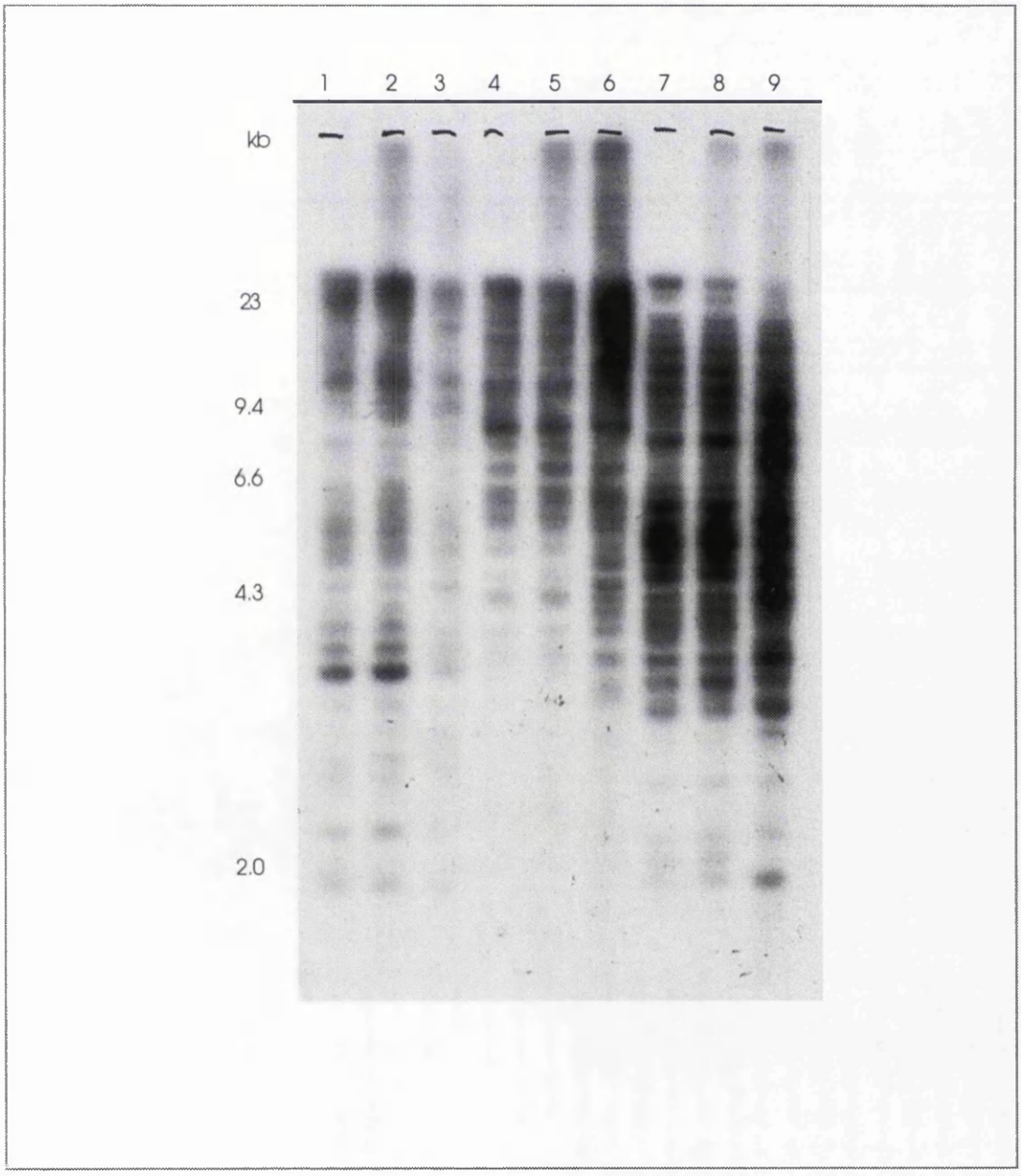


Figure 6.7 Southern data for the ingi containing pHB subclone. 16 μg of genomic DNA were digested with 40U of enzyme in 200 μl volumes (37°C; overnight) ethanol precipitated and loaded onto four 0.65% agarose gels run in 1XTBE (*i.e.* each lane represents 4 μg of DNA). This is an autoradiogram of a hybridisation (h.s.) of the 0.5 kb *Pst* I/*Bam* HI or *Hin* dIII fragment from the insert of pHB onto a genomic southern blot. Lane 1, *Bam* HI digested EATRO 2340 DNA; lane 2, *Bam* HI/EATRO 795; lane 3, *Bam* HI/STIB 386; lane 4, *Eco* RI/EATRO 2340; lane 5, *Eco* RI/EATRO 795; lane 6, *Eco* RI/STIB 386; lane 7, *Hin* dIII/EATRO 2340; lane 8, *Hin* dIII/EATRO 795; lane 9, *Hin* dIII/STIB 386.

No ingi-less fragments have been found in pHB in this survey. This southern blot does, however, serve to demonstrate how similar EATRO 2340 and 795 are to each other and how STIB 386 shows several obvious polymorphisms with respect to ingi hybridisation patterns when compared to the *T. b. rhodesiense* stocks.

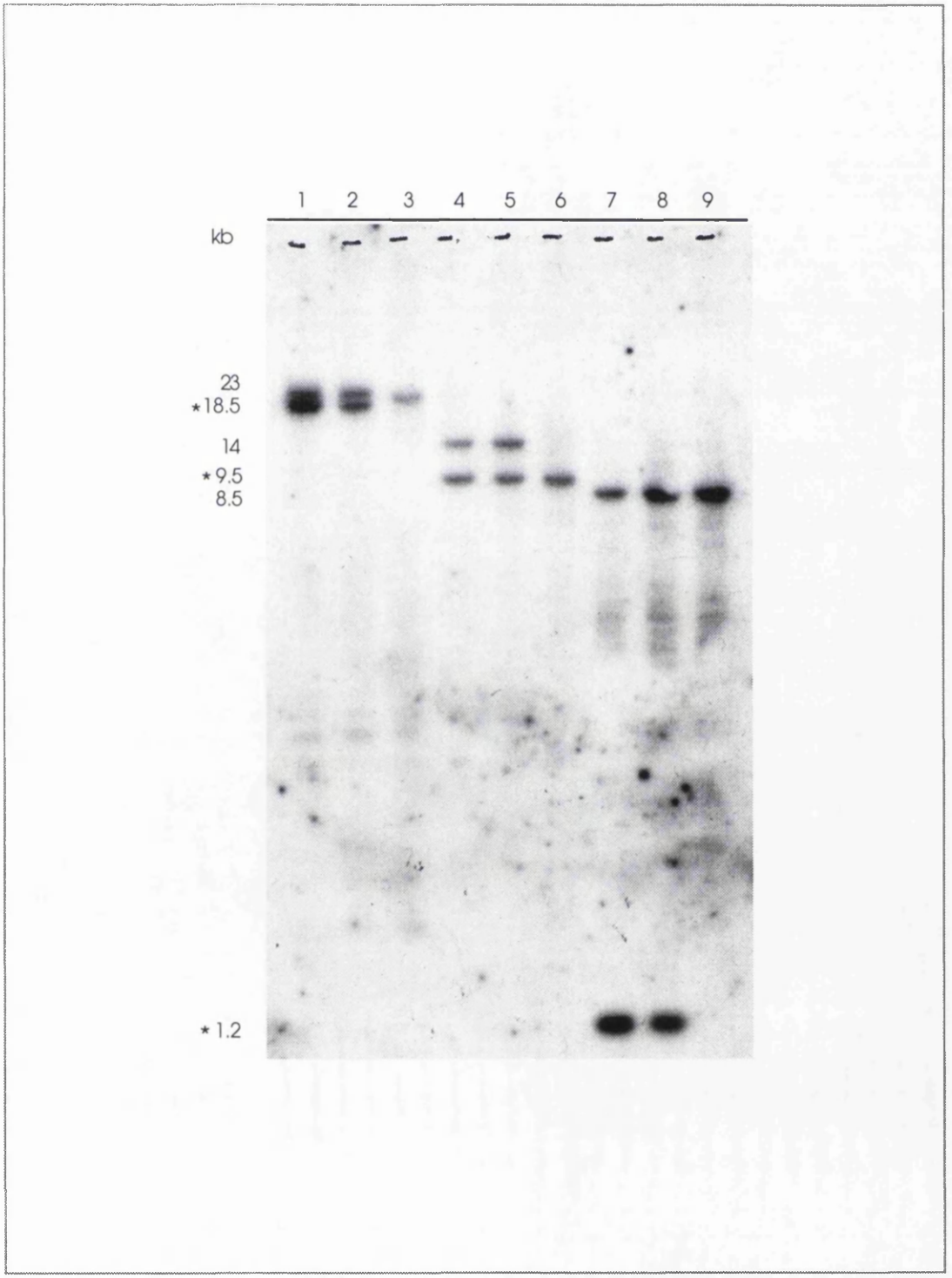


Figure 6.8 Southern data for pH1.2. See previous Figure for experimental conditions. This is an autoradiogram of a hybridisation (h.s.) of the insert of pH1.2 onto a genomic southern blot. Lane 1, *Bam* HI digested EATRO 2340 DNA; lane 2, *Bam* HI/EATRO 795; lane 3, *Bam* HI/STIB 386; lane 4, *Eco* RI/EATRO 2340; lane 5, *Eco* RI/EATRO 795; lane 6, *Eco* RI/STIB 386; lane 7, *Hin* dIII/EATRO 2340; lane 8, *Hin* dIII/EATRO 795; lane 9, *Hin* dIII/STIB 386. * indicates the presumed basic copy version of this sequence. The kb sizes at the left indicate calculated fragment sizes.

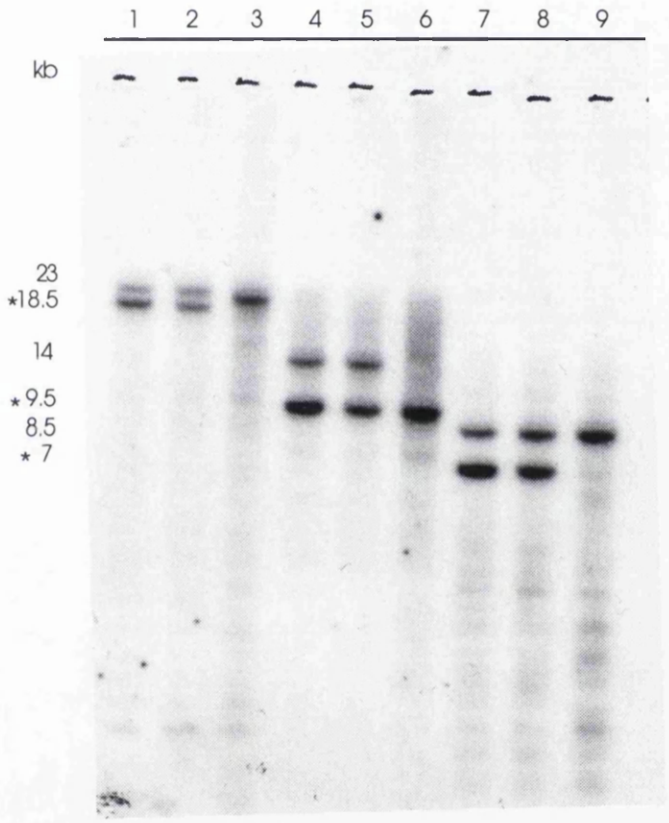


Figure 6.9 Southern analysis for pEH2.1. See Fig. 6.7 for experimental conditions. This is an autoradiogram of a hybridisation (h.s.) of the insert of pEH2.1 onto a genomic southern blot. Lane 1, *Bam* HI digested EATRO 2340 DNA; lane 2, *Bam* HI/EATRO 795; lane 3, *Bam* HI/STIB 386; lane 4, *Eco* RI/EATRO 2340; lane 5, *Eco* RI/EATRO 795; lane 6, *Eco* RI/STIB 386; lane 7, *Hin* dIII/EATRO 2340; lane 8, *Hin* dIII/EATRO 795; lane 9, *Hin* dIII/STIB 386. * indicates the presumed basic copy version of this sequence. The kb sizes at the left indicate calculated fragment sizes.

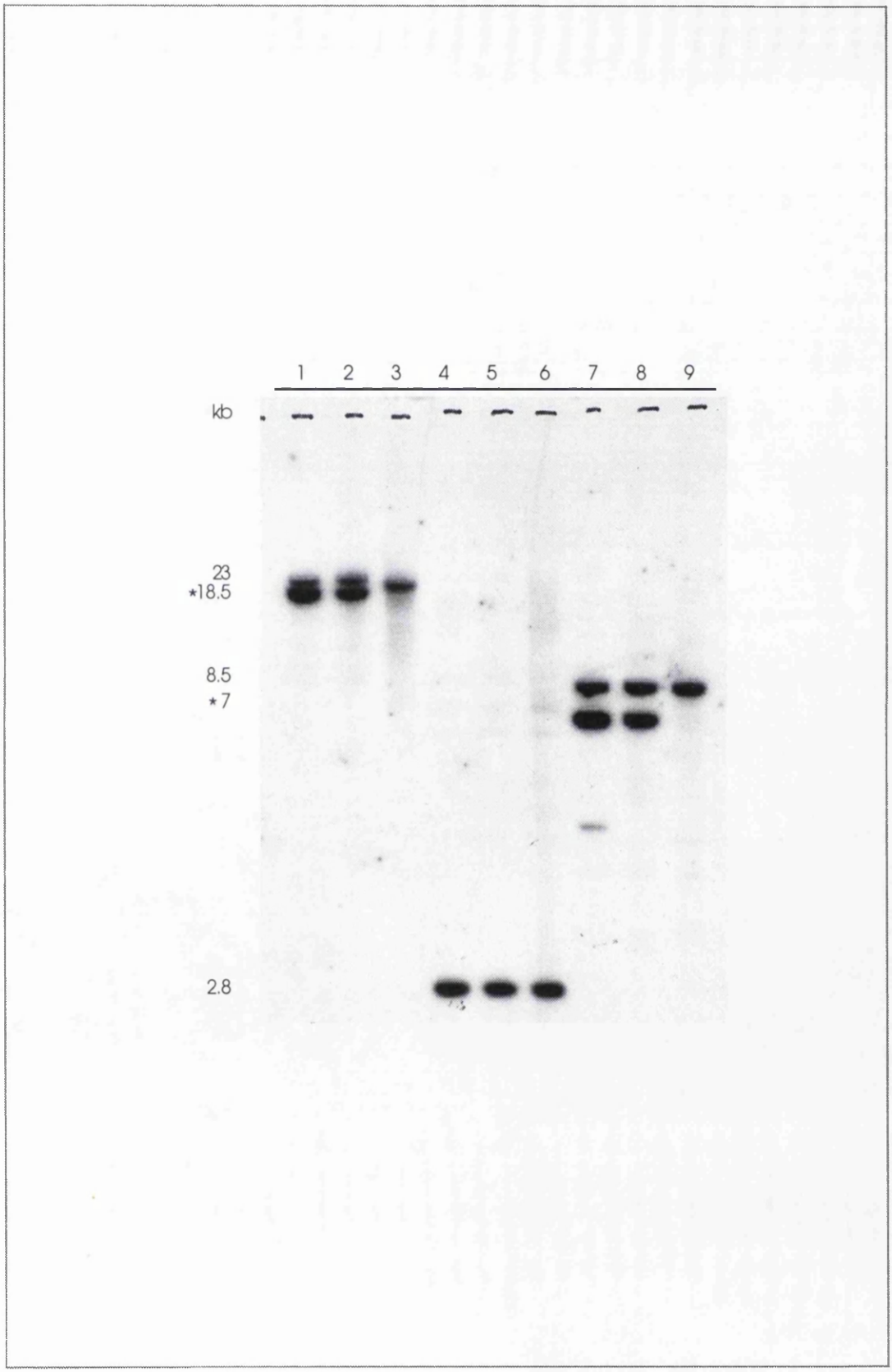


Figure 6.10 Southern analysis for pE2.4. See Fig. 6.7 for experimental conditions. This is an autoradiogram of a hybridisation (h.s.) of the 1.6 kb *Eco* RI/*Nco* I fragment of pE2.4 onto a genomic southern blot. Lane 1, *Bam* HI digested EATRO 2340 DNA; lane 2, *Bam* HI/EATRO 795; lane 3, *Bam* HI/STIB 386; lane 4, *Eco* RI/EATRO 2340; lane 5, *Eco* RI/EATRO 795; lane 6, *Eco* RI/STIB 386; lane 7, *Hin* dIII/EATRO 2340; lane 8, *Hin* dIII/EATRO 795; lane 9, *Hin* dIII/STIB 386. * indicates the presumed basic copy version of this sequence. The kb sizes at the left indicate calculated fragment sizes.

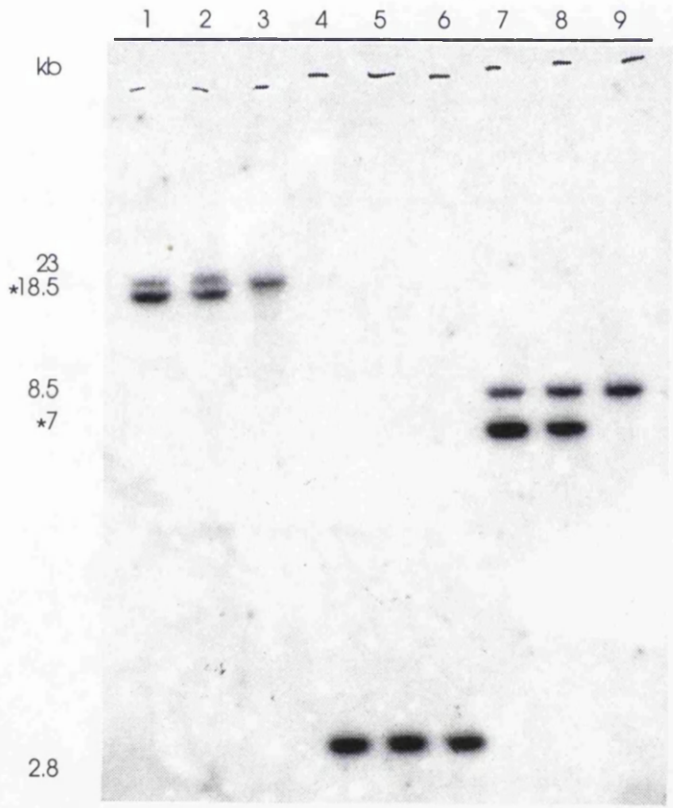


Figure 6.11 Southern data for pE2.4. See Fig. 6.7 for experimental conditions. This is an autoradiogram of a hybridisation (h.s.) of the 0.8 kb *Eco* RI/*Nco* I fragment of pE2.4 onto a genomic southern blot. Lane 1, *Bam* HI digested EATRO 2340 DNA; lane 2, *Bam* HI/EATRO 795; lane 3, *Bam* HI/STIB 386; lane 4, *Eco* RI/EATRO 2340; lane 5, *Eco* RI/EATRO 795; lane 6, *Eco* RI/STIB 386; lane 7, *Hin* dIII/EATRO 2340; lane 8, *Hin* dIII/EATRO 795; lane 9, *Hin* dIII/STIB 386. * indicates the presumed basic copy version of this sequence. The kb sizes at the left indicate calculated fragment sizes.

The results from **Figs. 6.7** through to **6.11** present a straightforward situation. The first thing to note is that the sequences, from pH1.2 to the upstream extreme of pE2.4, constitute normal genomic sequence (*i.e.* they are not obviously repetitive like ingi or pseudomonosomic like the 1.22BC locus). This fulfils the first criterion named for defining the end of the metacyclic domain: the presence of diploid or multicopy sequence. I argue that these sequences, in EATRO 795, are predominantly double copy, with one metacyclic linked copy and another copy at an undefined genomic context and location. It would be semantically inaccurate to call this sequence diploid, unless it could be proven that the two copies reside on functionally homologous chromosomes. This analysis is far from being quantitative; the assertion that this sequence is double copy comes only from the observation that in the presumptive metacyclic bands in the autoradiograms, there appears to be approximately a 2:1 ratio of signal intensity between the EATRO 2340 and 795 lanes. This 2:1 band intensity is expected for metacyclic linked sequences and has allowed me to indicate in the Figures (with an *) the bands that are likely to correspond to the metacyclic copy.

The fact that the fragments tested in **Figs. 6.8-6.11** yielded bands of the same size in the southern analysis, strongly suggests that they are linked (*i.e.* colinear with each other) in both genomic copies.

Earlier, it was asserted that the novel 1.22 sequences from pH1.2 to pE2.4 are predominantly double copy in EATRO 795. The reason I used this description is revealed by close examination of **Figs. 6.8-6.10**. The high stringency hybridisations in these experiments produced a ladder of faint bands in addition to the dominant bands on the autoradiograms. The significance of such results was unclear at the time of the experiment. It could only be speculated that these fragments might harbour a metacyclic-specific copy of some member of a multigene family. Interestingly, the most extreme upstream fragment tested (**Fig. 6.11**) failed to produce a similar series of cross hybridising bands.

The transition point between the single copy and double copy sequence associated with the 1.22BC in EATRO 795, occurs somewhere around the retroposon-like element. This copy of ingi appears to occupy the boundary between the metacyclic domain and the sequences that lie outwith it. **Fig. 6.7** documents, among other things, the failure to find a non-repetitive fragment on the upstream boundary of the ingi element in pHB. Detecting such a fragment is desirable for several reasons. Southern data from such a fragment would determine if the second copy of this sequence is ingi-less, thereby accounting for the RFLPs, as seems likely. With this information one could state, with certainty, which bands on the autoradiograms are linked to the1.22 gene. Lastly, this information would indicate whether or not the copy of this sequence in STIB 386 contains ingi as well, possibly shedding some light on the origin of the 1.22-linked sequence.

The most surprising finding in the genomic southern experiments is the conserved presence of this sequence, which seems to define the end of the

metacyclic domain, in the distantly related (65% similar) STIB 386 stock of *T. brucei*. The genomic structure of this sequence in STIB 386 is largely conserved and the signal intensity, despite the high stringency hybridisation conditions used, is comparable to that with the *T. b. rhodesiense* stocks. Lastly, the presence of the background ladder of cross hybridising bands in the STIB 386 lanes corresponds with their occurrence in the other stocks. Curiously, based on the fragment sizes, the 386 sequence resembles the presumed metacyclic version in *Eco* RI digests, like the other copy in *Hin* dIII digests and is intermediate between the two versions in *Bam* HI digests. The presence of this sequence in STIB 386 suggests that important functional sequences (a gene or genes?) may be harboured somewhere on these fragments. This strengthens the earlier speculation that a member of a multigene family may be present here.

It is possible for the project to bifurcate at this point. One could profitably pursue the question of the origin of the 1.22 metacyclic domain by continued analysis with STIB 386 and other stocks or one could continue to define the extent of the metacyclic domain and to hunt for the putative gene(s) in this area. The latter option is viewed as both more interesting and more germane to the project aim.

6.6 The 1.22 metacyclic domain in STIB 386

Section 6.5 showed that a version of the double copy sequence upstream of the 1.22BC domain exists in STIB 386. It seems possible, from the inferred genomic map, that STIB 386 may have this sequence linked with its own version of the 1.22 metacyclic domain. Therefore, I am curious to know whether STIB 386 has any of the 1.22 metacyclic domain and whether it has a copy of the 1.22 M-VSG gene itself.

Figure 6.12 is the result of the hybridisation of the R5 fragment (see **Fig. 6.1**) onto one of the Southern filters used in **section 6.5**.

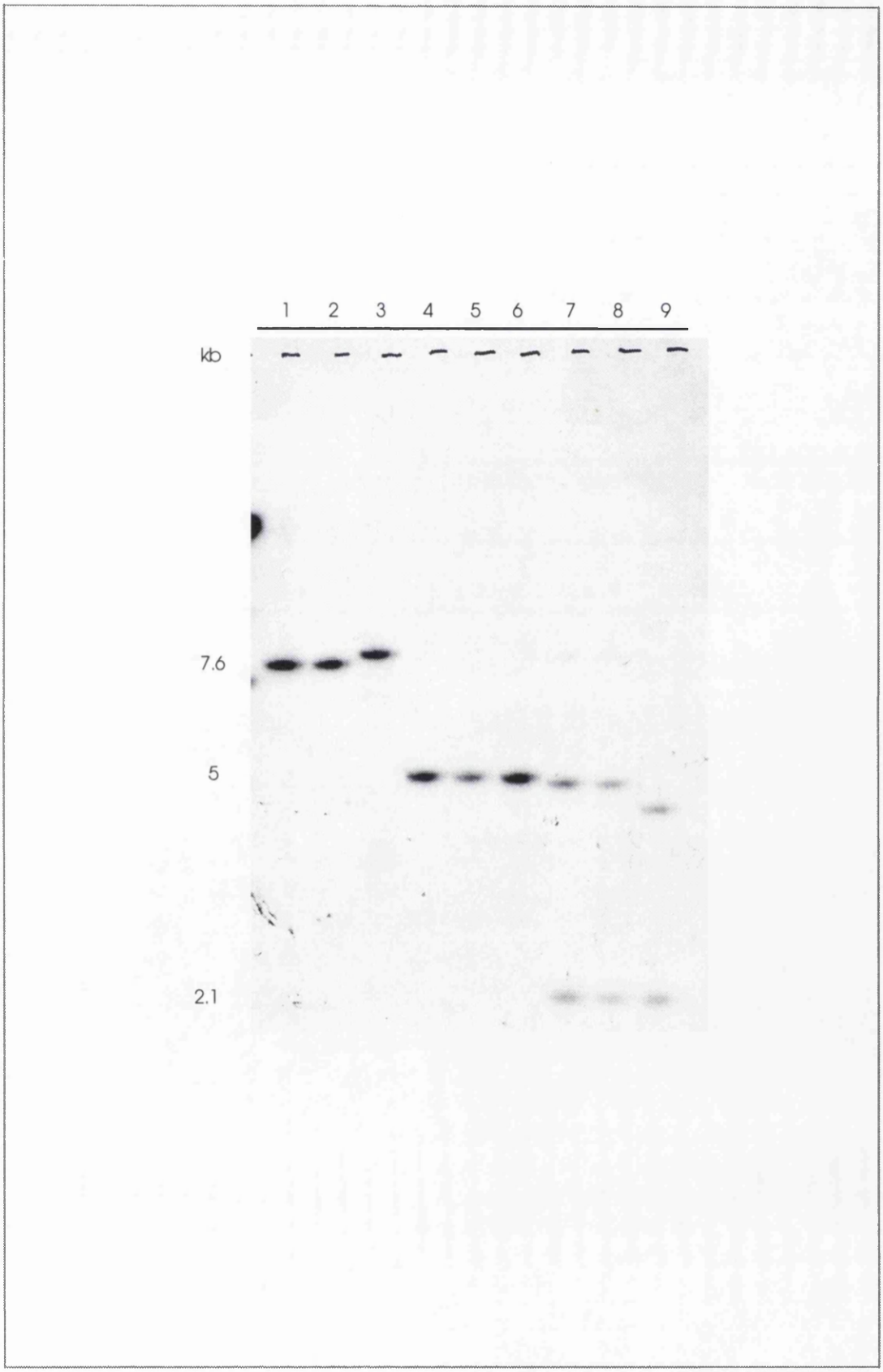


Figure 6.12 Southern data for R5. See Fig. 6.7 for experimental conditions. This is an autoradiogram of a hybridisation (h.s.) of the 5.0 kb *Eco* RI fragment of λMT1.22-2 onto a genomic southern blot. Lane 1, *Bam* HI digested EATRO 2340 DNA; lane 2, *Bam* HI/EATRO 795; lane 3, *Bam* HI/STIB 386; lane 4, *Eco* RI/EATRO 2340; lane 5, *Eco* RI/EATRO 795; lane 6, *Eco* RI/STIB 386; lane 7, *Hin* dIII/EATRO 2340; lane 8, *Hin* dIII/EATRO 795; lane 9, *Hin* dIII/STIB 386. The kb sizes at the left indicate calculated fragment sizes.

As is plainly visible from this analysis, STIB 386 has 1.22 metacyclic domain sequence and it is unexpectedly conserved; the restriction map structure is very similar and the signal intensity is comparable to that in the 2340 and 795 lanes. Though this study does not definitively prove linkage, it does suggest that STIB 386 has the 1.22 metacyclic domain linked to the newly cloned 1.22 sequences (the X sequences) since it seems probable that STIB 386 shares a copy of ingi in the same position as is found in the 1.22BC chromosome.

Fig. 6.13 is a result of the hybridisation of part of the 1.22 M-VSG gene onto one of the southern filters used in **section 6.5**.

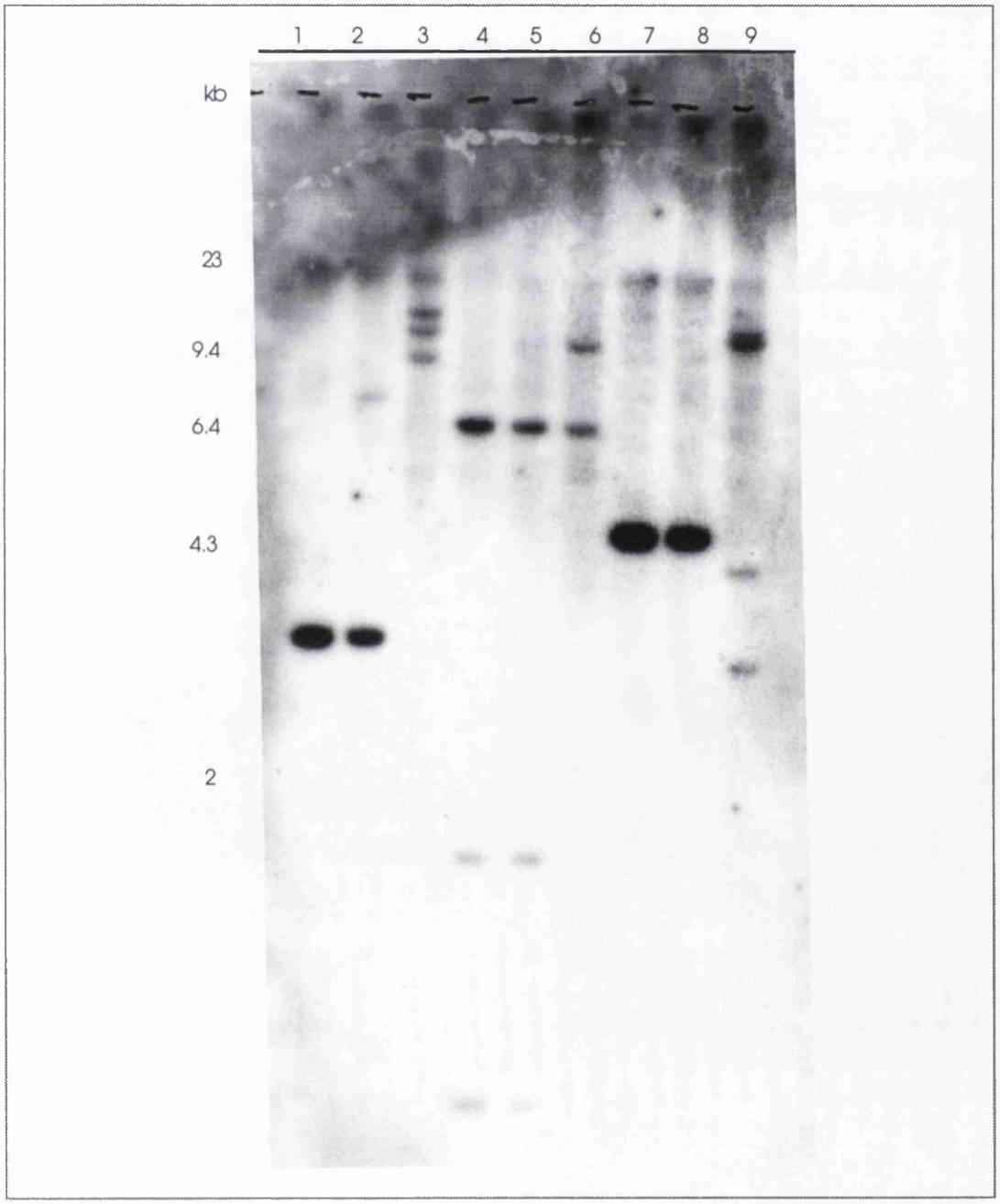


Figure 6.13 Southern analysis for the 1.22BC gene. See Fig. 6.7 for experimental conditions. This is an autoradiogram of a hybridisation (h.s.) of the 2.0 kb *Eco* RI/*Hin* dIII fragment of pMG 7.1-1 (see Fig. 4.11) onto a genomic southern blot. Lane 1, *Bam* HI digested EATRO 2340 DNA; lane 2, *Bam* HI/EATRO 795; lane 3, *Bam* HI/STIB 386; lane 4, *Eco* RI/EATRO 2340; lane 5, *Eco* RI/EATRO 795; lane 6, *Eco* RI/STIB 386; lane 7, *Hin* dIII/EATRO 2340; lane 8, *Hin* dIII/EATRO 795; lane 9, *Hin* dIII/STIB 386.

STIB 386 seems not to have an exact copy of the 1.22 M-VSG gene, though it does seem to have 1.22-like sequences (other related VSGs?). The marked decrease in signal intensity for the 386 lanes compared to the 2340 and 795 lanes indicates that STIB 386 lacks a single *bona fide* 1.22 gene. It is unknown to what the 1.22 metacyclic domain sequences are linked in STIB 386. It could be acting as a metacyclic domain to a different M-VSG gene or it could be chromosome internal and unassociated with M-VSG genes. It appears likely, based on the southern data, that STIB 386 has the same linkage of the metacyclic domain, ingi and the new sequences in λMT1.22-3, as found in the 1.22-linked copies present in EATROs 2340 and 795.

Figure 6.14 presents the inferred genomic maps of the three versions of this sequence and a speculative history of the loci in question.

Figure 6.14

The genomic maps in this figure were inferred from a combination of prior mapping data (K.R. Matthews 1990), the map of λMT1.22-3 and from the results of the southern analyses reported in **Figs. 6.8-6.12** performed with probes derived from the X sequences and the metacyclic domain. The use of a stock (EATRO 2340) in which the 1.22BC is known to be duplicated, allowed for comparison against the trypanosome stock (EATRO 795) from which the this locus was cloned and which has the 1.22BC gene in single copy. Fortunately, the duplication event seems to have encompassed a large region of the chromosome (the whole chromosome?) thus enabling the investigator to make a tentative assignment, based on signal intensities between EATRO 2340 and 795, of linkage to the metacyclic domain. Another, polymorphic, version of the X sequences was apparent in EATRO 2340 and 795 but not in the distantly related *T. brucei* STIB 386 stock. The most significant polymorphism in this 'alternate' copy arises in the area of the genomic map where ingi resides on the metacyclic telomere. The alternate copy appears to lack ingi and to be linked to a sequence unrelated to the 1.22 metacyclic domain. STIB 386 seems to have the X sequences in a conserved configuration, probably with ingi present and linked to a slightly polymorphic version of the 1.22 metacyclic domain. The analysis required to definitively prove this linkage is absent as a suitable probe flanking ingi was not found (**section 6.5**). The conservation (or even the presence) of the metacyclic domain in STIB 386 was unforeseen. This prompted an attempt, through interpretation of the genomic maps and knowledge of stock relatedness (**Fig. 6.6**), to outline a speculative history of these loci.

The reasoning was as follows:

1) The observation of central importance is that two (probably) copies of the X sequences exist in the *T. b. rhodesiense* stocks and that they are linked to different sequences at a boundary defined, in the 1.22 metacyclic telomere, by the presence of a ingi element. This can most easily be accounted for by an ancestral reciprocal recombination event between two ingi elements on non-homologues (Panel B, step 1).

2) If not already linked to its metacyclic domain before step 1, the 1.22BC gene must be acquired (Panel B, step 2). The presumed homologous copy of the metacyclic domain (provided that this sequence hasn't always been pseudomonosomic) is lost (Panel B, step 2).

3) The *Hin* dIII polymorphism within the X sequences must be generated (Panel B, step 3) and from this hypothetical ancestor, loss of the alternate X sequences (Panel B, step 5) could produce the pattern observed in STIB 386. Retention of both versions of the X sequences essentially produces the pattern expected for EATRO 795 and duplication of the 1.22BC chromosome (or a large portion thereof) could give the pattern observed for EATRO 2340 (Panel B, step 4).

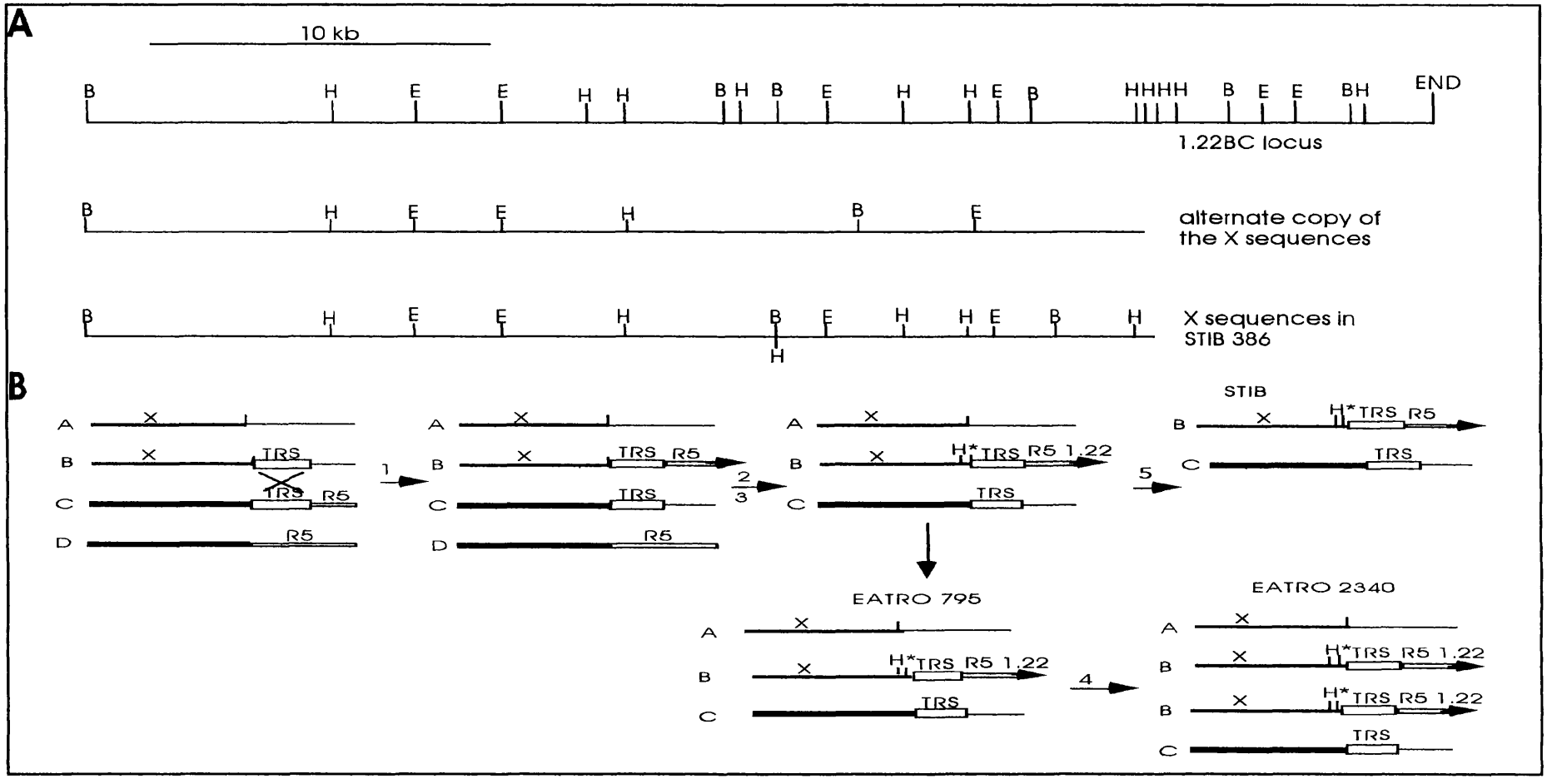


Figure 6.14 Proposed evolution of the 1.22 locus. Panel A is a representation of the genomic restriction maps for the three 'alleles' of the new 1.22 sequence. Panel B outlines a speculative history of the loci based on the genomic maps. 1) recombination between Ingi elements on non-homologous chromosomes 2) acquisition of the telomeric 1.22 gene and loss of the 'D' chromosome 3) mutation to produce (or lose) the *H site (polymorphic Hin dIII site) on the 1.22 chromosome 4) duplication of the X-TRS-R5-1.22 locus or its chromosome 5) Loss of chromosome 'A'. The restriction sites follow the same convention used all through this chapter. X, novel sequences in λMT1.22-3; TRS, the Ingi element; R5, represents the metacyclic domain; 1.22, indicates the 1.22 M-VSG gene.

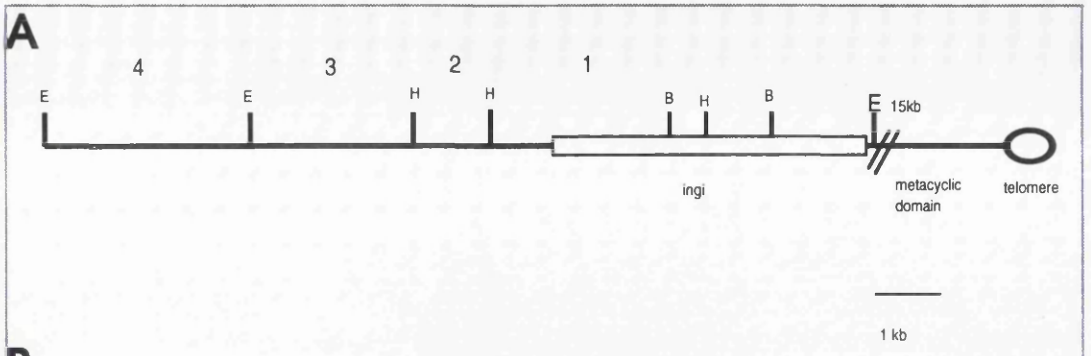
6.7 Transcriptional analysis of the novel 1.22 sequences derived from λMT1.22-3

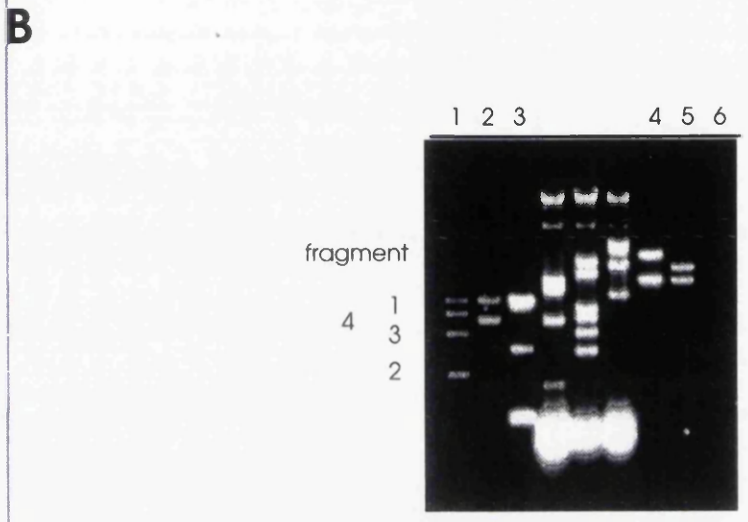
Earlier, it was noted that the upstream limit of the metacyclic domain would be defined by the presence of multicopy, transcriptionally active sequence. The TRS element upstream of the 1.22 locus was not considered to fulfil these criteria, as it is regarded as an adventitious repetitive element. The X sequences (pEB6.4 and pE2.4) that complete and extend upstream of the TRS element were found to be multicopy (probably double copy) in the analysis documented in **section 6.5**.

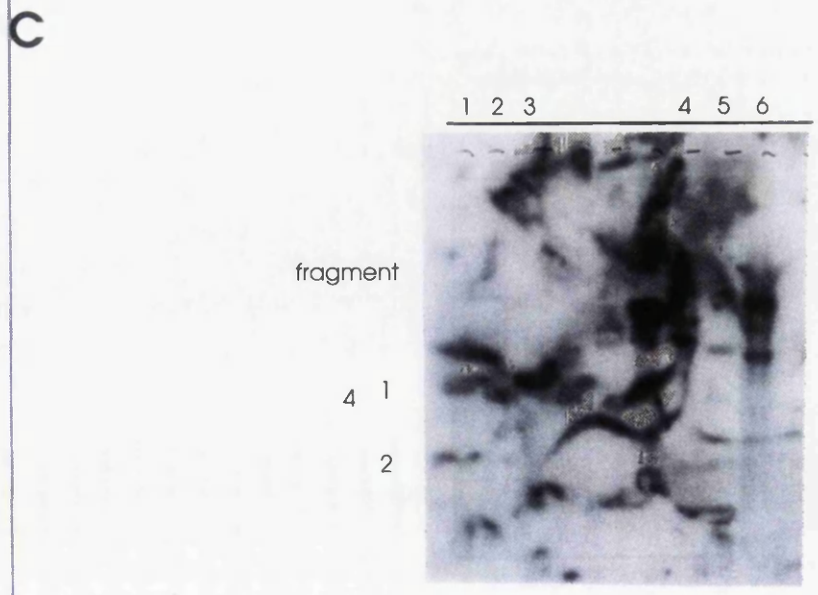
The next logical step is to examine the transcriptional status of the X sequences as a preliminary to determining whether the return to multicopy status is accompanied by the presence of coding function.

To that end, a transcriptional map for the X sequences was derived by transcriptional run-on analysis. Once again, the run-on analyses were undertaken in collaboration with V.S. Graham. Radioactive elongation reactions were performed with bloodstream nuclei from the 1.22j' clone (EATRO 795) in the presence (500 μg. ml^{-1}), and absence, of α-amanitin (37°C; 5 min). An additional, drug-free, elongation reaction was performed with EATRO 795 procyclic stage nuclei (27°C; 5 min). All nuclei used were prepared, at an earlier stage, by S.V. Graham. The RNAs were isolated by BRL's Trizol protocol and used as hybridisation probes onto filters with query and control clones.

Details of the run-on protocol can be found in **section 2.6.6**. The results of this transcriptional analysis are presented in **Fig. 6.15**.


A
4
3
2
1
E
E
H
H
B
H
B
E
15kb
ingi
metacyclic
domain
telomere
1 kb


B
1 2 3
4 5 6
fragment
4
1
3
2


C
1 2 3
4 5 6
fragment
4
1
2

(over)

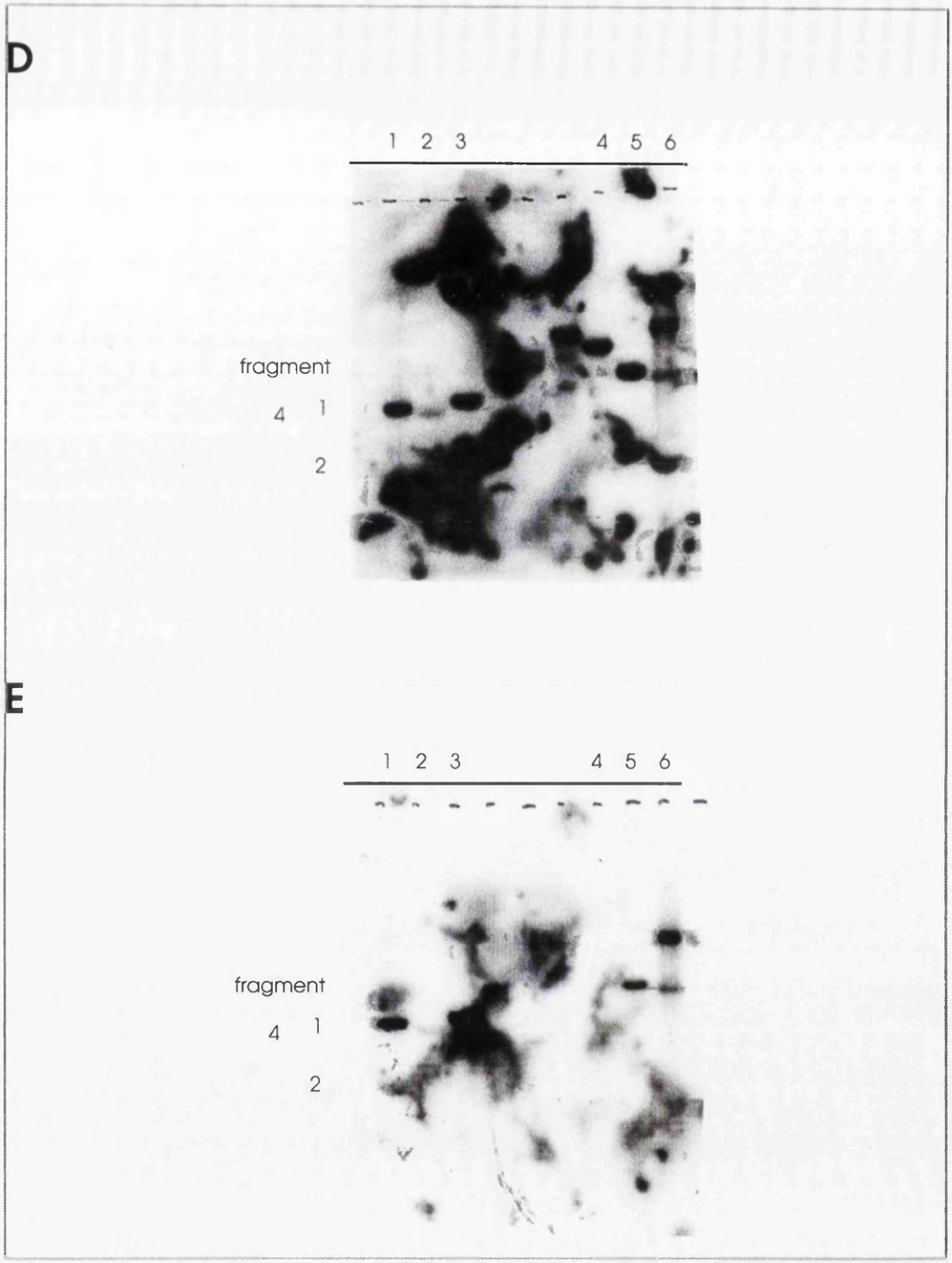


Figure 6.15 Transcriptional run-on analysis of pEB6.4, pE2.4, pLHN and pLHE*. Panel A presents the maps of the areas assayed. The numbers correspond to the fragments enumerated on the ethidium bromide stained gel. Panel B is a picture of an ethidium bromide stained gel of the substrates assayed in the run-on experiments. Quadruplicate runs were performed on a single gel; one run is shown here. Panel C is an autoradiogram of the run-on hybridisation of the α-amanitin containing bloodstream probe. Panel D is an autoradiogram of the run-on hybridisation of the drug free bloodstream probe. Panel E is an autoradiogram of the run-on hybridisation of the drug-free procyclic probe. In the α-amanitin reaction the nuclei were preincubated with 500 μg. ml^{-1} of the drug and the elongation reaction itself was conducted in drug-free buffer. All hybridisations were in 3XSSC at 55°C for 48 h and all filters were washed to medium stringency (1XSSC; 55°C). The three unnumbered lanes are clones derived from the 1.61 metacyclic telomere by Vincent S. Graham. Lane 1, *Bam* HI/*Eco* RI/*Hin* dIII digested pEB6.4; lane 2, *Eco* RI digested pE2.4; lane 3, *Hin* dIII/*Not* I digested pLHN and *Eco* RI/*Hin* dIII digested pLHE*; lane 4, *Hin* dIII digested pMG 7.1-1 (1.22 VSG gene control); lane 5, tubulin gene control (pT1); lane 6, rDNA control (pR1) underloaded and not visible by ethidium bromide staining.

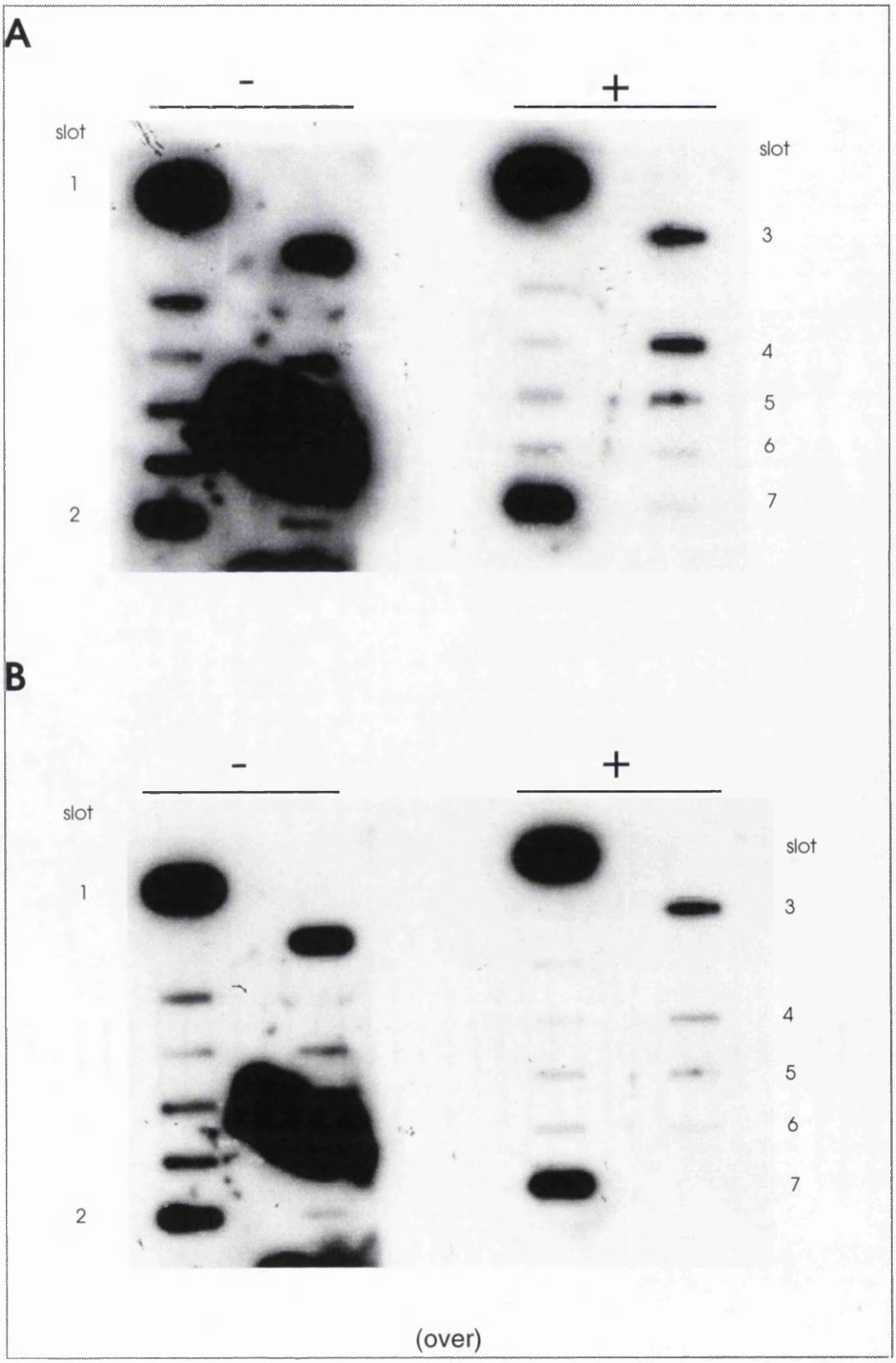
The controls indicate that the elongation reactions and hybridisations have been successful. Particular attention should be paid to the controls in Panel C (the α-amanitin containing bloodstream run-on). The tubulin control, in the presence of the drug, is greatly reduced relative to the rDNA control indicating that α-amanitin was taken up by the nuclei. This falls a little short of the ideal situation in which the transcription of tubulin is completely abolished. Lane 4 in Panel C (the 1.22 VSG control) presents a puzzling result; the transcription of the 1.22 gene appears to have a marked sensitivity to α-amanitin in this experiment. The same control behaves as expected when it gives a completely negative result in the procyclic run-on (Graham and Barry 1995). Results from this experiment must be regarded with caution due to this inconsistency.

The transcription unit from Yλ122, represented on the subclones pLHN and pLHE* (see **Fig. 4.14**), is narrowed down by this experiment, to the sequences that were shown to be repetitive in the experiments documented in **Fig. 4.8**. Due to high, patchy background on the autoradiograms, it is not possible to confirm the previous run-on result (**Fig. 4.4**) that produced a signal for pLHN's insert. This is a common occurrence in run-on analyses, since the probe is heterogeneous and unfractionated. As these subclones harbour a copy of an ingi-like sequence (**section 4.13**), it is not surprising that strong transcription is found in bloodstream and procyclic life cycle stages and that there is marked resistance to inhibition by α-amanitin observed in the Panel C experiment.

The *Bam* HI/*Hin* dIII fragment of pEB6.4, expected to harbour part of a TRS element, also produces strong signals in both bloodstream and procyclic run-on experiments. The α-amanitin sensitivity is more marked (*i.e.* is detectable) than that observed for pLHE*. This is probably due to the large relative underloading of pEB6.4 compared with pLHE*. The other two insert fragments from pEB6.4 produce very weak signals in the run-on experiments. It varies from procyclic to bloodstream experiments and from experiments in which α-amanitin is present or absent, as to whether or not a signal can be discerned from these two fragments. A judgement on the transcriptional status of these two fragments will be reserved until a more sensitive assay is employed. The results for pE2.4 are clearer. Transcription of the insert of pE2.4 is detected in all three experiments, so it can be concluded that pE2.4 sequences are being transcribed in both procyclic and bloodstream life cycle stages in EATRO 795 trypanosomes. Furthermore, this transcription is insensitive to inhibition by the drug α-amanitin (indicative of RNAP I transcription) in bloodstream trypanosomes.

A confirmation and clarification of the run-on results of **Fig. 6.15** was sought for the X sequences in a more sensitive set of transcriptional run-on experiments. The putative double copy portions of the X sequences (pH1.2, pEH2.1 and pE2.4) were slot blotted, in large quantity (5μg) onto Biodyne B nylon membranes (Pall) and used as substrates in a further set of bloodstream run-on experiments (see **section 2.6.7** for the details of the slot blotting procedure). Bloodstream nuclei from the 1.22a clone of ILTat 1.22 (EATRO 795) that were previously isolated by V.S. Graham, were used in a pair of radioactive transcriptional

elongation reactions (37°C; 5 min). One reaction was conducted after the nuclei were preincubated with a 1 mg. ml^{-1} concentration of α-amanitin. This time, the elongation reaction itself was also conducted in the presence of 1 mg. ml^{-1} α-amanitin.



(over)

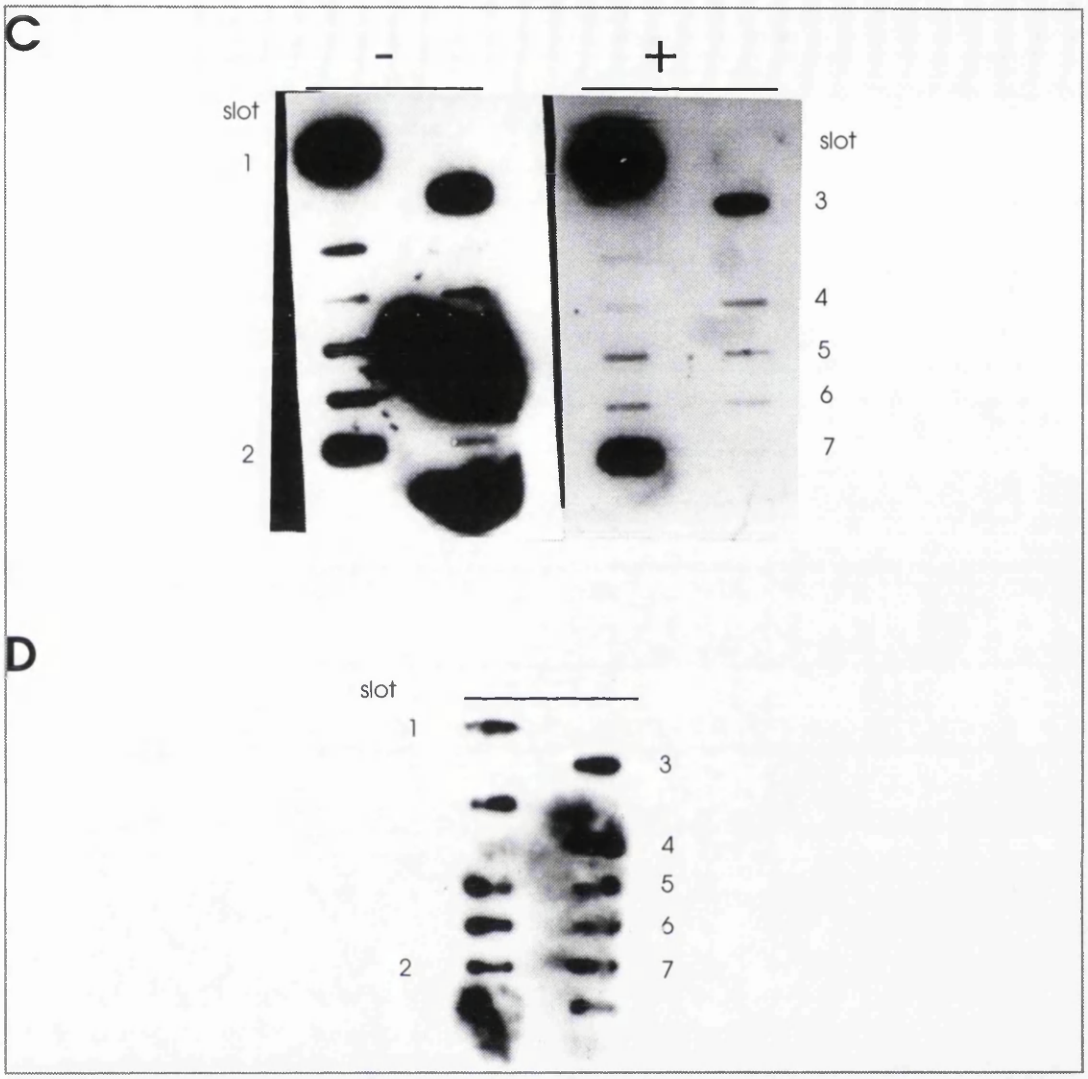


Figure 6.16 Slot blot run-on analysis. Panel A is a pair of autoradiograms of the run-on hybridisations of the 1.22a bloodstream probes, washed to medium stringency (1XSSC; 55°C). +, - indicates the presence or absence of α-amanitin in the elongation reaction. Panel B is the same pair of filters washed to moderately high stringency (0.1XSSC; 55°C). Panel C, wash at high stringency (0.1XSSC; 65°C). Panel D is an autoradiogram of a Bluescript (Stratagene) control hybridisation (6XSSC; 65°C) of the α-amanitin positive filter (*i.e.* the informative filter). This filter was reprobed after the original, experimental, signals had decayed. There is an adequate excess of target to leave the control unaffected. The hybridisations were conducted in 3XSSC at 55°C for 48 h. 1, rDNA control (0.5 μg pR4); 2, 1.22 VSG control (5 μg pMG7.1-1); 3, tubulin control (2.5 μg pT1); 4, 5 μg pE2.4; 5, 5 μg pEH2.1; 6, 5 μg pH1.2; 7, 5 μg pBluescript negative control (pBSK$^-$).

The results of **Fig. 6.16** are clearer than those of **Fig. 6.15**. The controls are also more consistent in this set of experiments. A very faint signal is detected (just above pBSK$^-$ background) for pH1.2, a stronger one for pEH2.1 and pE2.4 produces the strongest signal. The unmarked slots are samples from the 1.61 locus. Panel C reveals a result that could be very important; the signals for the X sequences seem to be more severely diminished by high stringency washing than the samples from the 1.61 locus. Perhaps, the apparently positive run-on results are being contributed by hybridisation with related transcripts derived from related sequences in the genome (the existence of which is noted in **section 6.5**). pE2.4, in particular, seems to be hybridising with heterologous tran-

scripts. This is manifested by the sharpness in signal reduction noted for this plasmid compared to the others, when the washing stringency is raised from 1XSSC/55°C to 0.1XSSC/55°C. Due to the apparent cross hybridisation with heterologous transcripts, pE2.4 is regarded as the best candidate for having a coding function.

6.7 Gene hunting in pE2.4

The transcriptional run-on experiments have not definitively established whether or not the X sequences linked to the 1.22BC are transcribed. It is not really possible to achieve this when the sequence is multicopy. The data do suggest, however, that the end of the metacyclic domain has been found in the X sequences. The genomic and transcriptional characterisation that has gone into the effort to define the extent of the metacyclic domain have pointed towards the possible existence, in the X sequences and in pE2.4 in particular, of a coding unit that is transcribed in an α-amanitin resistant manner.

To establish the identity of the putative gene in pE2.4, sequencing of the pE2.4 insert was undertaken. The easiest way to scan pE2.4 and, if necessary, to sequence it completely is to generate a nested set of deletion derivatives of the plasmid for double strand sequencing. Deletion sets from both sides of the polycloning site would help scan the entire insert quickly, as well as give substrates for deriving sequence information from both strands. The method of choice, for generating nested sets of deletions, considering an *Eco* RI insert in pBSK$^-$, is the use of exo III. pBSK$^-$ is designed so that the exo III resistant restriction sites, *Kpn* I and *Sac* I possessing 3' overhangs, are positioned one at each end of the polycloning site. Between the 'protected' sites and the *Eco* RI site used in this case for cloning, are restriction sites that are unprotected, by virtue of their 5' overhangs, and absent from the insert of pE2.4 (*Bam* HI between *Sac* I and *Eco* RI and *Hin* dIII between *Kpn* I and *Eco* RI). *Bam* HI/*Sac* I and *Hin* dIII/*Kpn* I digested pE2.4 would provide ideal substrates for the generation of exo III deletions for sequencing respectively from the T3 and T7 sides of the insert.

The double stranded Nested Deletion Kit from Pharmacia was used to generate these two sets of unidirectional deletions. The manufacturer's protocol, an adaptation of Henikoff (1984), was followed (75mM NaCl; 30°C; 10 samples taken at 3 min intervals). The resultant deletion plasmids were assigned names based on the protected restriction site used (K for *Kpn* I and S for *Sac* I) and the size of the deletion (*i.e.* 1 is the smallest deletion in the set; 5 is a bigger deletion).

Deletions for T3 sequencing are represented by plasmids S1 through to S6.

The set of deletions for T7 sequencing is made up of plasmids called K1 to K7.

6.8 Sequencing pE2.4

The so called 'forward' and 'reverse' primers of pBSK$^-$ were used in double stranded dideoxy chain termination sequencing reactions for the K and S deletion plasmids, respectively. S1, S5, S6 and K7 were sequenced using Sequenase and run on a glycerol tolerant denaturing polyacrylamide wedge gel (UBI Sequenase Kit; manufacturer's glycol/heat protocol). S5 and the remaining plasmids were Qiagen 'miniprepped' for automated termination cycle sequencing (Perkin-Elmer Cetus) and run on an autodata collection gel system (Applied Biosystems, Inc.).

The sequences derived from the 13 deletion plasmids, when assembled on GCG, form 4 contigs. The entire insert of pE2.4 is not represented without gaps in this analysis, and only small regions have been sequenced on both strands. None the less, important and revealing information has been obtained by this quick sequencing scan of the pE2.4 insert. The 0.8 kb *Eco* RI/*Nco* I fragment appears to be devoid of coding sequence, while the 1.6 kb *Eco* RI/*Nco* I fragment appears to harbour a member of the ESAG1 trypanosome gene family.

The results of the sequencing project, including work performed after I departed from the laboratory are summarized in the **appendix**.

This result makes sense of several observations noted in the previous sections. Most importantly, the positive transcriptional run-on results correlate with the presence of a gene in pE2.4. ESAG 1K (this gene) is a member of a multigene family; this neatly accounts for the faint background ladder of bands in the southern analysis reported in **Fig. 6.10** for the 1.6 kb *Eco* RI/*Nco* I fragment of pE2.4. The absence of this cross hybridising background, in the southern analysis of the 0.8 kb *Eco* RI/*Nco* I fragment of pE2.4, is explained by the probable absence of this coding region in this fragment (**Fig. 6.11**). The pattern of signal reduction observed with increased stringency washing of the slot blot run-on experiment (**Fig. 6.16**) can be attributed to transcripts from a bloodstream expression site associated copy of ESAG1 hybridising to the pE2.4 substrate. This, of course, does not rule out the possibility that ESAG 1K, itself, is expressed in bloodstream trypanosomes. The observation that transcriptional run-on signals, in the X sequences, are more prominent as the sequence assayed is more telomere distal, foreshadowed the eventual discovery that ESAG 1K lies in the reverse orientation with respect to the 1.22 M-VSG gene.

6.9 Are M-VSG expression sites remodelled B-VSG expression sites?

Superficially, the 1.22 M-VSG expression site (ES) looks like a bloodstream ES, now that a putative ESAG1 gene has been found upstream of the 1.22 gene. The possibility exists that M-VSG expression sites are remodelled or defunct bloodstream expression sites. If this is the case, it might be predicted that ESAG1 would be linked to 70 bp repeats, of which hundreds flank the VSG gene in B-ES.

It was ascertained whether or not 70 bp repeats are associated with ESAG 1K. The X sequences downstream of ESAG 1K showed evidence (**section 6.5**) of the presence of sequence that cross hybridised, faintly, with sequences present elsewhere in the genome. Are 70 bp repeats the origin of that cross hybridisation?

Figure 6.21 presents the results of an experiment that probed the X sequences for the presence of 70 bp repeats.

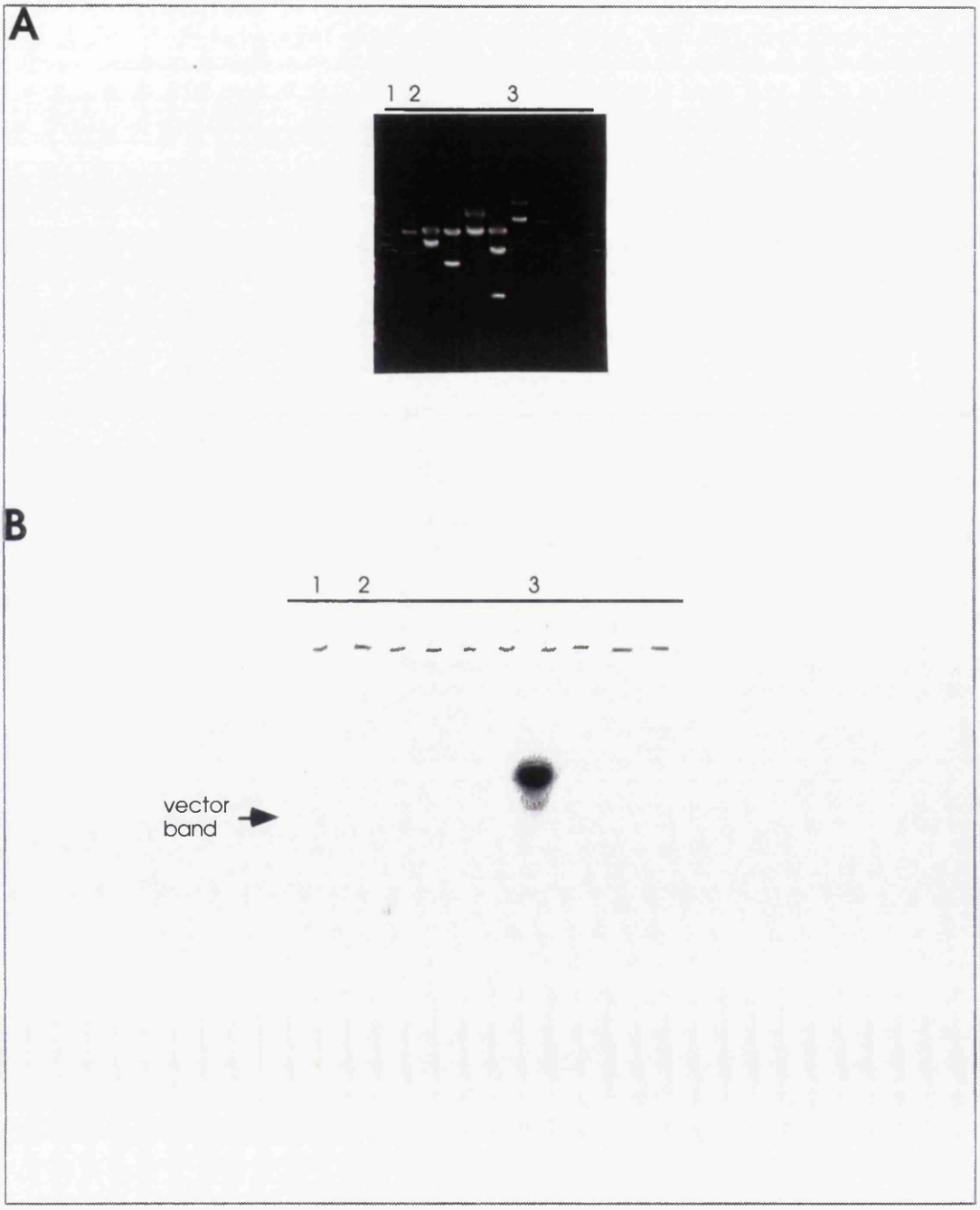


Figure 6.17 70 bp repeat probing of pE2.4, pEH2.1 and pH1.2. Panel A is a picture of the ethidium bromide stained gel that was blotted onto Hybond N for the 70 bp repeat probing experiment. Panel B is an autoradiogram of a medium stringency hybridisation (6XSSC; 50°C) of a probe containing a tract of 70 bp repeats (pTg221-1 from Bernards *et al.* 1985) onto a filter with control and experimental clones (washed at 2XSSC/0.1%SDS; 50°C). Lane 1, pE2.4/*Eco* RI; lane 2, *Eco* RI/*Hin* dIII digested pEH2.1 and *Hin* dIII digested pH1.2; lane 3, *Hin* dIII digested pMG 7.1-1 (1.22 VSG gene control with one 70 bp repeat).

The X sequences do not contain any detectable 70 bp repeat sequence, whereas the control containing just one repeat yields a strong signal. Naturally, due to the orientation of the ESAG 1K coding region, it could be argued that the search for 70 bp repeats is looking on the wrong side of this sequence. In bloodstream expression sites the 70 bp repeat sequences lie on the C-terminal side of the ESAG 1 coding region. However, there is only 1 kb cloned on the C-terminal side of ESAG 1K (this is free of detectably repetitive sequence). The orientation of ESAG 1K, itself, strongly suggests that this locus is not a remodelled bloodstream expression site. Another finding, on the 1.61BC telomere, that upstream of the 1.61 M-VSG gene, the first coding unit is not an ESAG1 but rather a possible homologue of ESAG 9, points out that the genomic environments upstream of the two metacyclic ESs studied in this laboratory do not closely resemble each other (V.S. Graham, personal communication). Taken together, these results make it seem improbable that M-VSG expression sites and their upstream domains are merely remodelled bloodstream expression sites.

6.10 Northern analysis of pE2.4

Transcriptional run-on analysis and sequencing have uncovered the existence of a possible coding unit that falls into the ESAG 1 family, within the plasmid pE2.4 (see **sections 6.7** and **6.8**). It is not possible, based on the transcriptional run-on analyses, to demonstrate that transcription is occurring from the 1.22-linked copy of ESAG 1K. Furthermore, such analysis would not be able to establish that such transcription correlates with gene expression (*i.e.* gives rise to steady state RNA). Northern analysis is required to do this. The next step is to perform northern analysis with the two fragments of pE2.4. The 0.8 kb *Eco* RI/*Nco* I non-coding fragment of pE2.4 is included in this analysis as a negative control.

Previous northern analysis of a metacyclic linked ESAG 1 gene using bloodstream stage RNA produced weak northern signals (Son *et al.* 1989). This weak signal could have been the result of cross hybridisation with an ESAG 1 expressed elsewhere or it could have reflected the existence of a short lived RNA species derived from the metacyclic VSG gene ES linked copy. In order to cover this latter possibility, total RNA was prepared from trypanosome infected rat blood as well as from chromatographically purified bloodstream trypanosomes expressing another VSG (ILTat 1.2; EATRO 795). The infected blood was mixed, immediately after exsanguination, with the guanidinium-based denaturing solution of Chomcynski and Sacchi (1986) as were the purified bloodstream trypanosomes. The RNA extractions then proceeded in tandem utilising the single step method of Chomcynski and Sacchi (1986). The quality of the extractions was assayed by non-denaturing gel electrophoresis and U.V. spectrophotometry. Both samples had large yields of undegraded RNA. A northern blot of these RNA samples was set up by running them on a denaturing agarose gel and transferring the RNA onto a Zetaprobe GT nylon membrane (Biorad) as de-

scribed in the manufacturer's protocol for GeneScreen membranes (Dupont). After fixation (120 mJ of U.V. and 1 h baking at 80°C) the RNA was visualised by methylene blue staining (Sambrook *et al.* 1989). Hybridisations were conducted at high stringency under 'nonaqueous' conditions (*i.e.* 50% formamide, 10% dextran sulphate; 5X SSPE, 42°C).

Figure 6.18 presents the results of the northern analysis of pE2.4.

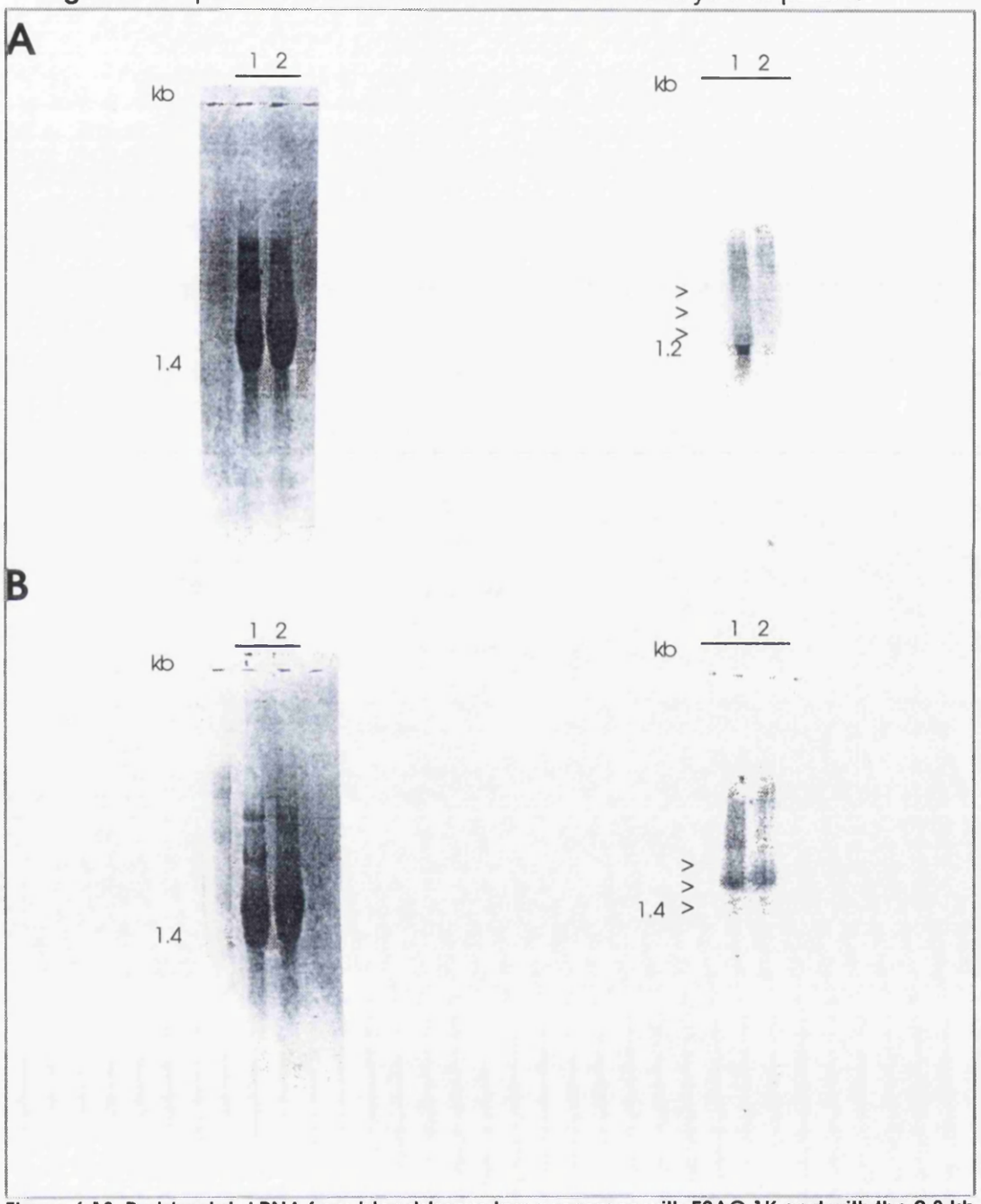


Figure 6.18 Probing total RNA from bloodstream trypanosomes with ESAG 1K and with the 0.8 kb *Eco* RI/*Nco* I fragment of pE2.4. Panel A is a photocopy of a methylene blue stained filter of RNA samples accompanied by an autoradiogram of that filter hybridised (h.s.) with the ESAG 1K containing 1.6 kb *Eco* RI/*Nco* I fragment of pE2.4. Panel B is a photocopy of another methylene blue stained filter with an autoradiogram of that filter hybridised (h.s.) with the telomere distal flank of ESAG 1K (0.8 kb *Eco* RI/*Nco* I fragment of pE2.4). Lane 1, total RNA from trypanosome infected rat blood; lane 2, total RNA extracted from purified bloodstream trypanosomes. > indicates the position of the ribosomal bands. 15μg of total RNA was loaded on each lane. The autoradiograms were documented on a phosphorimager.

Unfortunately the quality of the northerns is poor. The 0.8 kb *Eco* RI/*Nco* I fragment of pE2.4, which lies outwith the ESAG 1K coding region, yields a diffuse background on its autoradiogram when no signal is expected from this probe. The ESAG 1K probe produces the same diffuse background but no signal is discernable above this background in purified trypanosomes. However, total RNA isolated immediately after exsanguination produces a discrete and strong signal approximately 1.2 kb in size. This could represent a short lived steady state RNA produced from one or both copies of ESAG 1K. It is also possible that this signal is derived from rat RNA, as the control of rat blood RNA necessary to eliminate this possibility is absent from the experiment. It seems improbable that rat should possess such a sequence, especially not in such a highly conserved configuration, since, of the species tested, ESAG 1 sequences are found only in *T. evansi* and *T. brucei* spp. among the Trypanosomatids (Cully *et al.* 1985). These species share the property that they undergo antigenic variation. The stringency conditions used in this northern are also specific enough that the ESAG 1 family member that should be expressed from the bloodstream VSG expression site is not detected.

ESAG 1, in bloodstream VSG expression sites, is expressed as steady state RNA, at a level as much as 700 times lower than the downstream VSG gene with which it is cotranscribed (Cully *et al.* 1985). This indicates that the level of expression of ESAG 1, in bloodstream expression sites, is subject to significant posttranscriptional control. ESAG 1K, while it appears from the northern data to be subject to posttranscriptional control, is not, in any physical sense, coordinately transcribed with the 1.22 M-VSG gene as they are not physically linked in a transcription unit.

Despite the silence of bloodstream expression sites in procyclic stage trypanosomes, several genes with ESAG homology are transcribed, and yield steady-state RNAs (Pays *et al.* 1989). Therefore, copies of such ESAGs must exist outwith bloodstream expression sites. The characterisation of the parp loci revealed that they harbour members of the ESAG 6/7 and ESAG 2 gene families (Koenig-Martin *et al.* 1992; Berberof *et al.* 1991). The picture becomes even more complex when α-amanitin (in)sensitivity of transcription is considered. Pays *et al.* (1989a) discovered α-amanitin sensitive transcription of a RIME element, ESAGs 5, 4 and 2 in bloodstream and procyclic stage trypanosomes. A separate study illuminates this issue in metacyclic-derived trypanosomes (Graham and Barry 1991). All ESAGs were found to produce strong signals in transcriptional run-on experiments conducted with metacyclic-derived nuclei from trypanosome clones expressing the 1.22 M-VSG *in situ*. The strength of the signals obtained leads the authors to suggest that several copies of each ESAG are being transcribed in metacyclic-derived cells. Most relevant are the reports in that study that ESAG 1 is transcribed in procyclic and metacyclic-derived trypanosomes. The procyclic transcription of ESAG 1 is insensitive to

α-amanitin. Additionally, metacyclic-derived trypanosomes produce steady-state ESAG 1 RNA, while the procyclic stage does not. These observations raise the possibility that the products of ESAG gene family members are required in the metacyclic and procyclic stages of *T. brucei*.

The metacyclic transcription studies of the X sequences performed by S.V. Graham are intriguing. Here, the X sequences show strong α-amanitin sensitive transcription. The bloodstream slot blot run-on analyses (**Fig. 6.16**), which utilised many more nuclei and greater loadings of plasmid than the metacyclic experiment, produced very weak signals, especially for pH1.2. In the procyclic experiment, pE2.4 and pH1.2 produced signals; the intervening pEH2.1 is silent. Interestingly, the transcription of the pEH2.1 insert in **Fig. 6.15**'s bloodstream experiment appears to be α-amanitin sensitive while the transcription of the sequences on either side are not.

Any modelling of the transcriptional map of the X sequences must account for the following observations:

1) procyclic stage trypanosomes transcribe ESAG 1 sequences in an α-amanitin insensitive manner but make no steady state ESAG 1 RNA (Graham and Barry 1991)

2) procyclic stage trypanosomes appear to transcribe pE2.4 and pH1.2 but not the intervening pEH2.1

3) metacyclic-derived trypanosomes transcribe all the X sequences, including ESAG 1K, in an α-amanitin sensitive manner

4) metacyclic-derived trypanosome clones expressing the 1.22 M-VSG *in situ* produce ESAG 1 steady state RNA (Graham and Barry 1991)

5) weak transcription of the X sequences is observed in bloodstream trypanosomes; the transcription of pE2.4 and pH1.2 appears to be α-amanitin insensitive while the transcription of the intervening pEH2.1 insert seems to be sensitive to inhibition by α-amanitin

6) bloodstream trypanosomes appear to give rise to a very short lived ESAG 1K RNA species that is undetectable in purified bloodstream trypanosomes (**Fig. 6.18**)

Figure 6.19 presents a model which seeks to reconcile all the expression data enumerated above.

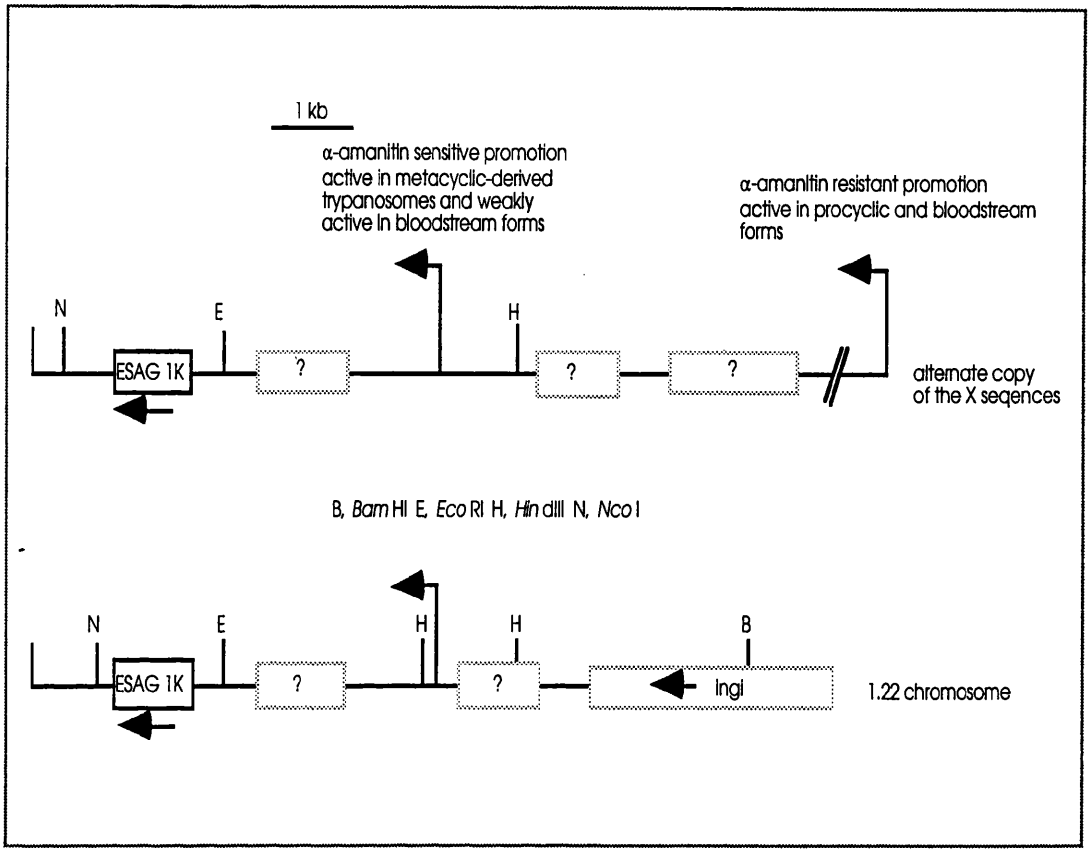


Figure 6.19 A model of the X sequences. The flags indicate putative promoters, dotted boxes indicate hypothetical genes and the arrows indicate the orientation of the coding units. Alternative scenarios are easily conceived, for instance ingi may be providing the promoter for the 1.22-linked X sequences. E, *Eco* RI; H, *Hin* dIII; N, *Nco* I.

The idea that parts of two different transcription units exist in the X sequences springs, mainly, from the observed behaviour of pEH2.1 in transcription studies. The insert of pEH2.1 gives different results than the inserts of the clones pE2.4 and pH1.2, which contain sequences that flank the insert of pEH2.1. Differences in the drug sensitivity and the life cycle stage specificity of transcription of this 'intervening' sequence could be explained by this model. It is particularly important that this model makes the prediction of where the promoter driving ESAG 1K expression lies. This proposed promoter is predicted to be represented in the insert of pH1.2, to be an RNAP II promoter and to be unable to drive expression of the X sequences in procyclic stage trypanosomes. The insert of pH1.2 is also predicted to harbour at least part of a gene occupying the end (or beginning) of another transcription unit, a unit that is complete only in the non-1.22 linked copy of the X sequences. This is speculative, based on the southern and transcriptional data, but this model does account for the observations and is easily tested. Looking for more coding regions in pH1.2 and pEH2.1 is likely to be fruitful. Such a 'gene hunting' exercise was only undertaken for pE2.4 as the bloodstream transcriptional run-on results for the rest of the X sequences looked unpromising. The metacyclic transcriptional run-on data obtained by S.V. Graham indicates greater complexity. It is clear, from those data, that there is strong α-amanitin sensitive transcription of all the X sequences in metacyclic-derived trypanosomes. The southern data for pEH2.1 and pH1.2 (**section 6.5**) show a faint ladder of cross hybridising sequences, again suggesting, as for pE2.4, that these clones may harbour copies of genes that are part of multigene families. ESAGs are obvious candidates. A quick test of the model would be afforded by a pair of hybridisation experiments using the inserts of pEH2.1 and pH1.2 onto southern blotted ESAG-containing plasmid clones. Medium stringency hybridisations followed by successively higher stringency washes (with subsequent exposures) would assay all the ESAGs simultaneously and rapidly. A more thorough test would be to sequence the inserts of pEH2.1 and pH1.2. It might also be worthwhile, eventually, to test the insert of pH1.2 and its various fragments for the ability to drive the expression of a reporter construct in transient transfection experiments.

Once again, it should be stressed that it is not possible to infer from the available evidence that the 1.22-linked copy of the X sequences is transcriptionally active. The positive transcriptional and northern signals could be coming from the alternate copy of the X sequences or even from the multiple, weakly similar sequences that are apparent in the southern data. The designation,ESAG 1K, has only been used as a convenient shorthand to mark the existence of a putative member of the ESAG 1 gene family; completion of the sequencing project could reveal the 1.22-linked copy of this sequence to be an pseudogene. Obviously, any serious efforts to test the model presented in this section should wait until the sequencing project is completed.

6.11 Postscript

During the time it has taken to write this thesis, work on the 'X sequences', including ESAG 1K, has been continued by others in the laboratory. This work has consisted of further transcriptional run-ons and ESAG 1K sequencing. Sheila V. Graham performed transcriptional run-on analysis of the X sequences utilising nuclei from metacyclic-derived trypanosomes. These are bloodstream trypanosomes harvested from rat blood ~5 days after tsetse bite. These metacyclic-derived trypanosomes still express VSG from metacyclic expression sites; the bloodstream VSG repertoire is not activated until ~day 6 of infection. M-VSGs are presumed to be expressed in metacyclic-derived trypanosomes as they are in actual metacyclic forms; the hope being that the X sequences are as well. Since it is known that ESAG 1 gene family member(s) are expressed in metacyclic-derived trypanosomes (Graham and Barry 1991), the key to run-on analysis indicating that ESAG 1K, itself, is transcribed is an examination of its flanks. Therefore, this set of run-ons included an examination of the C-terminal flank of ESAG 1K, *viz.*, the 800 bp *Eco* RI/*Nco* I fragment of pE2.4. This fragment falls outwith the ESAG 1K coding region and lacks any signs of harbouring multicopy sequence (**Fig. 6.11**). A positive signal for this fragment would be indicative of ESAG 1K transcription. This fragment, indeed all the X sequences (pH1.2, pEH2.1 and pE2.4), yielded a strong signal. Furthermore, all transcription detected, except that of the 800 bp *Eco* RI/*Nco* I fragment of pE2.4, showed marked sensitivity to α-amanitin. The α-amanitin sensitivity of transcription for the ESAG 1K C-terminal flank was, for unknown reasons, less marked.

The discord between my bloodstream transcriptional run-on results with those performed by S.V. Graham utilising nuclei isolated from metacyclic-derived trypanosomes suggested complexity in the X sequences. This prompted the formulation of the model presented in **Fig. 6.19**. My failure to get convincing data, with respect to the α-amanitin sensitivity of X sequence transcription, prompted an attempt by M. Fotheringham to repeat these experiments. A clear result was obtained. All the X sequences were found to be strongly transcribed in bloodstream nuclei and all this transcription, including that of the ESAG 1K C-terminal flank, was sensitive to inhibition by α-amanitin. Of particualr note was the discovery of the transcriptional silence of a ~500 bp *Apa* I/*Hin* dIII fragment of pHB. This fragment is predicted, by its map position, to lie outwith the ingi element and is immediately adjacent to pH1.2, where the X sequences become transcriptionally active. All this greatly clarifies the situation. The elaborate scheme of two transcription units with different polymerase requirements, interrupted by ingi in the 1.22BC chromosome is no longer tenable. Instead, it seems

that ingi is located just distal to a possible RNA polymerase II promoter. The small gap in transcription between ingi and the X sequences rules out the possibility that the ingi element is supplying a promoter for the X sequences present on the 1.22BC chromosome.

Finally, the continuation of the ESAG 1K sequencing project has identified an unambiguous stop codon in the middle of the 'coding sequence'. Aside from this the sequence is not greatly divergent from other ESAG 1 sequences, implying that the nonsense mutation is relatively recent. The presence of a nonsense mutation in the ESAG 1K linked to the 1.22BC gene suggests that there is no requirement for the coordinated (with the 1.22BC gene) production, from this gene copy, of a wild-type protein product. This is pertinent as the persistent linkage of ESAG-like sequences with M-ES could be viewed as evidence that their coordinate expression is required in metacyclic trypanosomes. The presence of this nonsense mutation in ESAG 1K suggests otherwise.

6.12 Perspectives

Summary

The so-called 'metacyclic domain' for the ILTat 1.22 M-VSG gene has been defined. This domain of silent, single copy sequence linked to the 1.22BC gene extends 16.5 kb upstream of the start of the ILTat 1.22 VSG gene, is interrupted by an apparently full length ingi element (5.2 kb) lying in the chromosome in the reverse orientation, with respect to the 1.22BC gene, and extends less than 1 kb beyond this element. Therefore, the metacyclic domain ends ~21 kb upstream of the start of the 1.22BC gene. Beyond this metacyclic domain, part of a transcription unit, probably transcribed by RNA polymerase II, has been found. This unit also lies in the reverse orientation with respect to the 1.22BC gene. One member of this transcription unit is known to be a member of the ESAG 1 gene family and is herein designated as ESAG 1K. The copy of ESAG 1K (it is probably double copy) linked to the 1.22BC gene and its metacyclic domain is interrupted 'mid-gene' by a nonsense mutation.

Thoughts about the metacyclic domain

Metacyclic domains are pseudomonosomic sequences located at chromosome ends. The two M-VSG genes studied in this laboratory (ILTat 1.61 and ILTat 1.22) both have metacyclic domains linked to them. If this situation is generalisable to all telomeres harbouring M-VSG genes, then as many as 27 chromosome termini in the *T. brucei* karyotype are pseudomonosomic. This could have important implications for the genetic behaviour of these chromosomes. *T. brucei* is known to undergo some form of nonobligatory genetic exchange in the tsetse fly (Tait 1980); models requiring meiosis are favoured. In

many organisms, telomeric and/or subtelomeric sequences are crucial for the homologue recognition step (*i.e.* the initiation) of proper chromosome pairing in early meiotic prophase [Liu *et al.* 1995 (yeast); McKim *et al.* 1993 (*C. elegans*); Curtis *et al.* 1995 (wheat)]. Any trypanosome meiotic process would have to be unusually tolerant of differences in chromosome length and terminal sequence composition in the chromosome pairing stage for meiosis to occur. One indicator that crossovers are indeed occurring would be the observation of an exchange of a metacyclic domain from one homologue to the other (provided that the homologues can be resolved).

Models of gene expression

Models of chromatin pairing/association to strengthen the silencing of B-ES in bloodstream trypanosomes, analogous to what happens in *Drosophila* (*e.g.* Laurenti *et al.* 1995) could be envisioned. Meanwhile the active B-ES may be sequestered to a position in the nucleus where an immobile VSG-specific 'transcription factory' exists (for a discussion of transcription foci and chromosome structure see Cook 1994). In this model the extremes of the VSG-ES would be bound to a transcription factory, with the intervening sequences defining a chromatin loop that slides past the fixed transcription factory. This chromatin loop defines a discrete supercoiled domain of the chromosome. This kind of model could account for the torsional stress observed in active VSG gene chromatin and perhaps also for the apparent memory of VAT expression through cyclical transmission (see **section 1.7.2**).

The pseudomonosomic nature of metacyclic domains probably rules out any silencing mechanisms requiring chromatin pairing. The silence of metacyclic domains imply that if they have a function it is an ultrastructural function. Thinking about VSG transcription in the context of the model in which chromosome structure is defined by immobile transcription factories located in various foci within the nucleus suggests a role for the metacyclic domain. The metacyclic domain may be required for the separation of the next transcription unit on the chromosome (the RNAP II transcription unit that starts in the X sequences) from the VSG-specific transcription factory. The presence of a long stretch of single copy, silent sequence in an area that would normally be populated by repeats and genes in a B-ES seems enigmatic if one works with the assumptions, implicit in most transcription work, of unordered nuclei and mobile transcription complexes. If the nuclei are ordered and the transcriptional machinery of the nucleus is distributed into immobile foci, then conserved, seemingly functionless non-coding regions of DNA no longer seem so enigmatic.

Futures

Besides sequencing the remainder of the X sequences from the transcriptional gap to ESAG 1K, the metacyclic domain is to be sequenced. This will help clarify whether, for this area, non-expressed equates with non-coding. *A priori* assumptions that the metacyclic domain is non-coding are unjustified. Perhaps the most exciting work now underway is the search for a RNA polymerase II promoter within the pHB transcriptional gap and within the insert of pH1.2. This entails the subcloning of fragments from this area, in both orientations, into reporter plasmids. Transient transfection of trypanosomes, followed by chloramphenicol acetyl transferase (CAT) or luciferase assays should identify any promising promoter candidates.

The designation of the single copy silent sequences linked to the 1.22BC gene as the 1.22 'metacyclic domain' has implicitly reified this DNA as a unitary functional entity. This designation is only intended, at present, to aid the investigators in the framing of experimental questions. It remains to be demonstrated whether the implied reification is receivable.

The most pressing issue raised by the analysis of the 1.22 M-VSG gene promoter (Graham and Barry 1995) is whether a position effect is operative at this telomere which accounts for the unique transcriptional control of this locus. The proximity of the metacyclic VSG promoter to the telomere makes this a much more probable candidate for control by telomere position effect, than the bloodstream VSG promoters located ~60 kb upstream of the chromosome end. A central hallmark of position effects is the promoter-independence of the silencing. Therefore, a promoter replacement experiment is in order to determine if a position effect is acting at this telomere. Replacement of the endogenous promoter with a constitutively active promoter that experiences no down regulation through the life cycle would be the most stringent test. The ribosomal promoter, despite possibly being transcribed by a different class of polymerase, is the best candidate available. The requirement for the insertion of the ribosomal promoter within a construct harbouring selection/reporter cassettes will place the experimental promoter further away from the end of the chromosome than the metacyclic promoter in the unmanipulated chromosome. Therefore, a parallel control replacement of the 1.22 metacyclic promoter, within the same construct, must be achieved to establish that the increased distance from the telomere isn't causing any of the differences that may be observed in the experimental promoter replacement.

The most definitive determination of whether the position effect, if one is indeed found, is a telomere position effect is, of course, afforded by deletion of the

telomere. Because of the time it takes to select the products of a targeted deletion and the rapidity with which chromosomes with broken ends are lost, this experiment may not be feasible in the native chromosome. However, a new set of molecular tools are being developed that may, eventually, make this kind of experiment possible: Trypanosome Artificial Chromosomes (TACs) (Lee *et al.* 1995). These are linear artificial chromosome vectors that include, selection cassettes, reporter cassettes, trypanosome subtelomeric and telomeric repeats, in their composition. *In vitro* assembly of TAC constructs with the 1.22 metacyclic promoter and the ribosomal promoter adjacent to a broken TAC end would be relatively straightforward. After transfection, performance of the requisite CAT or luciferase assays may be practicable in the period before the TAC is lost.

All this is pertinent, since a telomere position effect at metacyclic telomeres could account for the existence of metacyclic domains. These domains may represent the minimum required buffer size (~15 kb for the ILTat 1.61 telomere) for a non-metacyclic VSG promoter to be active in all life cycle stages. The proposal is that there may be a requirement, for the expression of the next transcription unit on the chromosome, for some non-essential DNA to act as a buffer against a silencing effect. 'Non-essential' because the sequence of this buffer seems unimportant. This is inferred from the fact that the two metacyclic telomeres studied in this laboratory have completely unrelated sequences constituting their metacyclic domains.

If a position effect **is** found to be influencing the 1.22BC locus, telomeric in origin or not, the next step is to integrate the 1.22 metacyclic promoter with selection and reporter cassettes into various positions in the metacyclic domain. This should establish how far from the telomere this promoter must be for it to be released from repression or if the silencing effect is propagated by chromatin repression spreading from within the metacyclic domain.

Lastly, I would eventually want to delete the metacyclic domain to determine if and/or how its absence affected the control of expression of the 1.22BC locus and/or the X sequences. This would have to be done twice with the selection cassette being inserted in both orientations to determine if there is any readthrough from the selection and/or reporter cassettes. A complete lack of phenotype in this experiment would suggest that the 1.22 metacyclic domain, rather than impacting upon the expression pattern of the 1.22BC locus or the X sequences, is merely a contingent historical fact of the evolution of these loci.

Appendix

ESAG 1K In a peptide pileup with other translated ESAG 1 gene sequences from the GenBank database.

gene designation	accession number	source
esagf	M31615	*T.b. gambiense* TREU 1257
esag1a	M21052	*T. b.* LVH75 (MVAT4)
esaga	M11451	*T. b.* 427 (117a)
esagca	M31612	*T. b. gambiense* TREU 1285
esagd	M31613	*T. b. gambiense* TREU 1285
esage	M31614	*T.b. gambiense* TREU 1257
esagb	M11452	*T. b.* 427 (221a)
esagcb	M20873	*T. b.* EATRO 1125
esag1k		*T. b. rhodesiense* EATRO 795

```
 esagf                                                          MKASIVG
esag1a                                                          MNLVTVE
esag1k                                                          MKVMVPR
 esaga                                                          MKVTIVE
esagca                                                               E
 esagd                                                          MRVVILE
 esage
 esagb                                                          MKVEIVE
esagcb                                                            ESYGS

esag1b  .......... .........C RLVDDFY.KN NLNESVCHLK CLSDALSRLY
 esagf  LVILEFFVMC IGGAEDEGPC PLVDDHF.KN NLRKSVCYLK CISDALNKLY
esag1a  LVALLFFVTC IDGQEE.NKC MSVTDYKGDG PLSEIVCHVS CLSDALNKLY
esag1k  WVFLSTFLNC VDGEDEVRRC TSIDDYKGIN .LSHSVCYLS CLSDALDRLY
 esaga  LVVWLFSVNF FVVVAEESRW TLVDDHYGKN .LHESVCYLR CLSNALNKLY
esagca  LVVLLFFVNC TGGSAPASRC TLVDDYYGKN .LHESVCYLR CLSNALNKLY
 esagd  WPTLILLI.. .VGAYGDNDH TLVADYEGDA PLSETVCYLR CLSDALNKLY
 esage  .......... .......... .....FK*VR MLSQLS..FR CIK*IIR*WR
 esagb  LVVLLFSVTC VDAWLQGADC TRVADHKEHA PVTEAVCYLR CLSDALNKLY
esagcb  GLVFVIFVTC VDAWRRVRIA LVWVITKSTH L*LEAVCHLR CLSDALNKLY

esag1b  GDGQKRLFVN EEVY.ASASR IFDDMEGKTG ESTKYLSVIG GVTEGEQDKL
 esagf  TDGEQKFLVN EEVY.ANASR ILDDMEGKTG ESTKYLSAIS GVIDGEYGKL
esag1a  TDGERKMLVN EEVY.ANASR ILDDMEGKTG ESTEYLSVIS GVMESKHDKL
esag1k  TNGEKKLFVN EEVY.ANASR ILDDMEGKTG ESTKYLSVIS GVMEGKHDKL
 esaga  SEGERRLFVN EEVY.ANASR ILDDMEGKTG ESTTYLSVVS SEMGGENNKL
esagca  SEGEKKLLAT EEVY.ANASL ILDDMEGRAG ESSTYLSIIS GVMEEQTDRL
 esagd  TDGEKKLLLN EEVY.ANVSR ILDSMEAKTG ESVKYLSVIS GVMGGENNKL
 esage  EEA...VCE* GSVC.KCISY IG*YGKQNR* KHNIFEYY*Q VVMVEKNNRL
 esagb  SEGEKKLLVT EEVY.ANASL ILDDMEGRAG ESSTYLSVIR GVMEEQTDRL
esagcb  SEGENKLACD RGKCTANASR ILDDMEGRTV QVLNI*VLLE V*WKGDNDKL

esag1b  EKLISYGNEM GNLVAKVGGL FAEVNESVSA VRKEIPDALI KANKYYTAVA
 esagf  EKLIFYGNTI GDLVDKVGGL FSEVNESVRA VREVLPSALI KVNKYYTAVA
esag1a  EKLISYGNEM GNLVAKAGGL FSEVNESVRA VRKEIPGALI KVNKYYTAIA
esag1k  EKLISYGNEM *DLVAKVSGL FSRVNESVRA VRKEIPEALV KANKYYTAIV
 esaga  EKLISYGNAM GDLVAKVGGL FAEVNESVRA VREEIPSALI RANKYYTAIA
esagca  EKLISYGNAM GDLVAKAGGL FAEVNESVRA VRKKIPDALM RANKYYTAIA
 esagd  EKLISYGNTM GDLVAKVGGL FAEVNESVRA VRKEILDALI KANKYYTSIA
 esage  EKLISHGNAM GDLVAKAGGL FAEVNESVRA VRKEIPGALI KVNKYYTAIA
 esagb  EKLISYGNKM GNLVAKAGGL FAALEDSLKE VRKEIPGALI KTNKYYTSVA
esagcb  EKLISCGNKM EIF*RRQVDY LLHRRTRLKE VRKEIPRCPH KNE*ILYVCR

esag1b  EITRTVWDDV MAME.KGGAG ECEDRTFKKA RALPTACAKQ MCALRD.SMS
 esagf  EITRTVWDDL KAMK.ERGTG ECEDQITREV VVLPATCADH TCPLRD.RVS
esag1a  EITRTVWSDV AVVN.HGDEA KCTEQKFRKV GELQTKCVDH TCPLSD.DVN
esag1k  EIARTVWDDV KAVS.NDEGH ECEKGTPKGV KEFSVTCSNQ TCSLSN.GVS
 esaga  EITRTVWDDV NRPL.QQDKA TCGDQKVTGV GELKTECGAH TCPLSD.GVN
esagca  EVVRTVWDDV .KAV.ERGKH ECKDQKFDGV KGFECTCGDN ACPLKD.GVS
 esagd  EIVRTVWDDV KAIP.TSIDP KCENGEFDGV KEFEVKCGDS SCPLRN.GVS
 esage  EVTRTVWDDV KAV..ESGKH ECKDQEFRGV KEFEVTCGDN ACPLRD.GVS
 esagb  EIVRTVWEDV GEILWKETEA KCGSQKVEGV GEIQTECGAH TCPFADNGVA
esagcb  *DY*DCLGGC RRDTVRRPRR PPLLQHQQG* RRW*VR.... .NRVWSFIRA

esag1b  KDAVQKYKDG CLAVTV..QN GSVSECLN.. .KPRDNAYKN GAVKNSGDVL
 esagf  KDTIQKYKDG CLAVTV..QS GSVSGCFN.. .RPRDNIYKN GVIESSDDVL
esag1a  EGTLQKYKDG CLDINV..AG GSVSECLN.. .LPRNKLYKS GGRTIPREAI
esag1k  ESTPQKYRGG CLQIDVLTGS GSVSECFN.. .LPRGKLYGH GGVNFSNDVL
 esaga  ESALQKYKGG CLEINVM..S GSVSECFN.. .LPRNKLYRS VALSSSHGFL
esagca  ESALQKYKGG CLEINVM..S GSVSECLN.. .LPRNNLYKN GAVNESKEGT
 esagd  EDALKYYKRG RIEVNVL..N GFVSRCLN.. .LQRKNLYKN GAEKHLSEVL
 esage  EGALKHYKNG LLEINVM..S GSVSKCLN.. .LPRNNLYKS GAINSSHGLL
 esagb  ASAVDKYKGH CLYVG...RN SYLRHCFN.. .LPRGRLYRH GPVNTLGDAL
esagcb  RLHLKWTPQP SINIRETVCT *AATATLENA TTSQE*FVRA WPCKYFGDAL

esag1b  IWKKNGRAAT FFQLTVEVEN IFGPLIVPFA SGQPPSVLLE IMSNITSFYS
 esagf  ..KWHDQGAT VFQLTVKVED IFGPLIAPFS AWQPPSVLLT KMTNITYLYS
esag1a  DWHDHRDEVT YFQLKLQVKS TFTPLIAPFA PGQPPSALLA LMSNITSLYS
esag1k  KWPQNYANR. .FQITVEVQK VFDPLIALFA SGLPPSVLLA VAVNVTSFYS
 esaga  KWYQ..DEAK RFQLGLRVKN IFGPLIASFG VGQPPSVLAE MINNITSLQS
esagca  DWYG..NGNT YFQLKLHVHS MFSPLSISLG AGQPTSALLV MMSNITSLQS
 esagd  KWPQ..DDAT FFQLKLEVQS MFGPLIVSFA AGRTTSALLE MMENITFLRS
 esage  KWHQ..DSAE MFQLELRVKN IFNPLIASFA AGQPPSVLAE MMANITSLYF
 esagb  EWEENWSDYM NFELTVKVQK IFGPLIASFD VGIAPSTLAE MINNITSLQS
esagcb  EWEENWSGYL GLNFG.KSRK IFGPLIASFV VGIAPSTLAE MINNITSLQS

esag1b  QFNKVYSNFT SLLLDTNTTD NLNSTGSII*
 esagf  HFNKVQSKFT SLLLDTNLIV NVNSSNSTI*
esag1a  HFNKVHNNFT SLLVDTNRTD NLSSTHYTI*
esag1k  RFKGVHSNFT SVLIDTNLTY NVNSTNSTI*
 esaga  RFNEVHSNFT SILLADNLTA DVDNTDSTI*
esagca  RFNEVHSNFT SILLAAKLKA EVDNTDSTI*
 esagd  RFNEIHSDFT SLLLNPNITD NVNSTDSTI*
 esage  RFNEVHRNFT SLLVNPNLTD TGDGIDSIF*
 esagb  RFNEVHSNFT SILFTTKLKT EVYNTDSTI*
esagcb  RFNEVHSNFT SLLLTTKLKT EVYNTDSTI*
```

References

Ahn, B-Y. and D.M. Livingston (1986). Mitotic gene conversion lengths, coconversion patterns and the incidence of reciprocal recombination in a *Saccharomyces cerevisiae* plasmid system. *Mol. Cell. Biol.* **6(11)**: 3685-93.

Alarcon, C.M., Son, H.J., Hall, T. and J.E. Donelson (1994). A monocistronic transcript for a trypanosome variant surface glycoprotein. *Mol. Cell. Biol.* **14(8)**: 5579-91.

Albertsen, H.M., Abdarrahim, H., Cann, H.M., Dausset, J., Le Paslier, D. and D. Cohen (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl. Acad. Sci.* USA **87**: 4256-60.

Barry, J.D. (1986). Antigenic variation during *Trypanosoma vivax* infections of different host species. *Parasitology* **92**: 51-65.

Barry, J.D., Hajduk, S.L., and K. Vickerman (1979). Detection of multiple variable antigen types in metacyclic populations of *Trypanosoma brucei. Trans. R. Soc. Tropical Medicine and Hygiene.* **73**: 205-8.

Barry, J.D., Graham, S.V., Matthews, K.R., Shiels, P.G. and O.A. Shonekan (1990). Stage-specific mechanisms for activation and expression of variant surface glycoprotein genes in *Trypanosoma brucei. Biochem. Soc. Trans.* **18**: 708-10.

Bentley, D.R. (1992). The analysis of YAC clones: in Techniques for the Analysis of Complex Genomes edited by R. Anand. (Academic Press Ltd.) Harcourt Brace Jovanovich Publishers, London, pp 113-135.

Berberof, M., Pays, A. and E. Pays (1991). A similar gene is shared by both the variant surface glycoprotein and procyclin gene transcription units of *Trypanosoma brucei. Mol. Cell. Biol.* **11(3)**: 1473-79.

Berberof, M., Vanhamme, L., Tebabi, P., Pays, A., Jefferies, D., Welburn, S. and E. Pays (1995). The 3'-terminal region of the mRNA for VSG and procyclin can confer stage specificity to gene expression in *Trypanosoma brucei. EMBO J* **14(12)**: 2925-34.

Bernards, A., Van Harten-Loosbroek, N. and P. Borst (1984). Modification of telomeric DNA in *Trypanosoma brucei:* a role in antigenic variation? *Nucleic Acids Res.* **12**: 4153-70.

Bernards, A., Kooter, J.M. and P. Borst (1985). Structure and transcription of a telomeric surface antigen gene of *Trypanosoma brucei. Mol. Cell. Biol.* **5**: 545-53.

Bernards, A., Kooter, J.M., Michels, P.A.M., Moberts, R.M.P. and P. Borst (1986). 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**: 313-22.

Birnboim, H.C. and J. Doly (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**: 1515-23.

Blum, M.L., Down, J.A., Gurnett, A.M., Carrington, M., Turner, M.J. and D.C. Wiley (1993). A structural motif in the variant surface glycoproteins of *Trypanosoma brucei. Nature* **362**: 603-609.

Borst, P., Bernards, A., Van der Ploeg, L.H.T., Michels, P.A.M., Liu, A.Y.C., DeLange, T., Sloof, P., Schwartz, D.C. and C.R. Cantor (1983). The role of mini-chromosomes and gene translocation in the expression and evolution of VSG genes. In Gene Expression, UCLA Symp. on Molecular and Cellular Biology, New Series, 8, D. Hamer and M. Rosenberg, eds. (New York: Liss), pp. 413-35.

Bringaud, F. and T. Baltz (1993). Differential regulation of two distinct families of glucose transporter genes in *Trypanosoma brucei. Mol. Cell. Biol.* **13(2)**: 1146-54.

Burke, D.T., Carle, G.F. and M.V. Olson (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**: 806-12.

Cardosa de Almeida, M.L. and M.J. Turner (1983). The membrane form of variant surface glycoproteins of *Trypanosoma brucei. Nature* **302**: 349-52.

Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D. and M. Turner (1991). Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues. *J. Mol. Biol.* **221**: 823-35.

Celnicker, S.E. and J.L. Campbell (1982). Yeast DNA replication *in vitro*: initiation and elongation events mimic *in vivo* processes. *Cell* **31**: 201-13.

Chomcynski, P. and N. Sacchi (1986). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* **162**: 156-59.

Chung, H-M., Lee, M.G-S. and L.H.T. Van der Ploeg (1992). RNA polymerase I-mediated protein-coding gene expression in *Trypanosoma brucei. Parasitol. Today* **8(12)**: 414-18.

Clarke, L. and J. Carbon (1976). A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. *Cell* **9**: 91.

Clarke, L. and J. Carbon (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* **287**: 504-509.

Clarke, L. and J. Carbon (1983). Genomic substitutions of centromeres in *Saccharomyces cerevisiae. Nature* **305**: 23-28.

Connelly, C., McCormick, M.K., Shero, J. and P. Hieter (1991). Polyamines eliminate an extreme size bias against transformation of large yeast artificial chromosome DNA. *Genomics* **10**: 10-16.

Cook, P.R. (1994). RNA polymerase: structural determinant of the chromatin loop and the chromosome. *Bioessays* **16(6)**: 425-30.

Cornelissen, A.W.C.A., Bakkeren, G., Barry, J.D., Michels, P.A.M. and P. Borst (1985 a). Characteristics of trypanosome variant antigen genes active in the tsetse fly. *Nucleic Acid Res.* **13**: 4661-76.

Cornelissen, A.W.C.A., Michels, P.A.M., Borst, P., Spanjer, W., Versluys-Broers, J.A.M., Van der Meer, C., Farzaneh, F. and S. Shall (1985 b). Effect of 3-aminobenzamide on antigenic variation of *Trypanosoma brucei. Biochem. Pharmacol.* **34(23)**: 4151-56.

Cornelissen, A.W.C.A., Johnson, P.J., Kooter, J.M., Van der Ploeg, L.H.T. and P. Borst (1985 c). Two simultaneously active VSG gene transcription units in a single *Trypanosoma brucei variant. Cell* **41**: 825-32.

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

Cornelissen, A.W.C.A., Evers, R., Grondal, E., Hammer, A., Jess, W. *et al.* (1991). Transcription and RNA polymerases in *Trypanosoma brucei. Nova. Acta. Leopold.* **279**: 215-16.

Coulson, A., Waterson, R., Kiff, J., Sulston, J. and Y. Kohara (1988). Genome linking with yeast artificial chromosomes. *Nature* **335**: 184-86.

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

Cross, G.A.M. (1984). Structure of the variant glycoproteins and surface coat of *Trypanosoma brucei. Phil. Trans. R. Soc. Lond. B* **307**: 3-12.

Cross, S.H., Allshire, R.C., McKay, S.J., McGill, N.I. and H.J. Cooke (1989). Cloning of human telomeres by complementation in yeast. *Nature* **338**: 771-76.

Crowe, J.S., Barry, J.D., Luckins, A.G., Ross, C.A. and K. Vickerman (1983). All metacyclic variable antigen types of *Trypanosoma congolense* identified using monoclonal-antibodies. *Nature* **306**: 389-91.

Cully, D.F., Ip, H.S. and G.A.M. Cross (1985). Coordinate transcription of variant surface glycoproteins and an expression site associated gene family in *Trypanosoma brucei. Cell* **42**: 173-82.

Curtis, C.A., Lukaszewski, A.J. and M. Chrzastek (1991). Metaphase I pairing of deficient chromosomes and genetic mapping of deficiency breakpoints in common wheat. *Genome* **34**: 553-60.

Den Dunnen, J.T., Steensma, H.Y., Grootscholten, P.M. and G.J.B. Van Ommen (1992). 'Reconstruction of megabase-sized genomic regions using overlapping yeast artificial chromosomes.' in Techniques for the Analysis of Complex Genomes. Ed. R. Anand. (Academic Press, Ltd.) Harcourt Brace Jovanovich, Publishers London, pp 197-214.

Dobson, M.J. and W.R.A. Brown (1992). 'Cloning human telomeres in yeast artificial chromosomes.' in Techniques for the Analysis of Complex Genomes. Ed. R. Anand. (Academic Press, Ltd.) Harcourt Brace Jovanovich, Publishers London, pp. 81-92.

Doolittle, R.F. (1986). Of URFS and ORFS. Oxford University Press.

Esser, K.M., Schoenbechler, M.J. and J.B. Gingrich (1982). *Trypanosoma rhodesiense* blood forms express all antigenic specificities relevant to protection against metacyclic (insect form) challenge. *J. Immunol.* **129**: 1715-18.

Fangman, W.L., Hice, R.H. and E. Chlebowicz-Sledziewska (1983). ARS replication during the yeast S phase. *Cell* **32**: 831-36.

Felgner, P., Brinkmann, V., Zillman, V., Mehlitz, D. and S. Abuishira (1981). Epidemiological studies on the animal reservoir of *Gambiense* Sleeping Sickness: parasitological and immunodiagnostic examination of the human population. *Tropomed. Parasit.* **32(3)**: 134-40.

Ferguson, M.A.J., Homans, S.W., Dwek, R.A. and T.W. Rademacher (1988). Glycosyl-phosphatidylinositol moiety that anchors *Trypanosoma brucei* variant surface glycoprotein to the membrane. *Science* **239**: 753-59.

Fincham, J.R.S. and P. Oliver (1989). Initiation of recombination. *Nature* **338**: 14-15.

Gommers-Ampt, J.H., Van Leeuwen, F., de Beer, A.L.J., Vliegenthart, J.F.G., Dizdaroglu, M., Kowalak, J.A., Crain, P.F. and P. Borst (1993). β-D-glucosyl-hydroxymethyluracil: A novel modified base present in the DNA of the parasite protozoan *T. brucei. Cell* **75**: 1129-36.

Gottesdiener, K., Chung, H-M., Brown, S.D., Lee, M.G-S. and L.H.T. Van der Ploeg (1991). Charaterisation of VSG gene expression site promoters and promoter-associated DNA rearrangement events. *Mol. Cell. Biol.* **11**: 2467-80.

Graham, S.V., Matthews, K.R., Shiels, P.G. and J.D. Barry (1990). Distinct, developmental stage-specific activation mechanisms of trypanosome VSG genes. *Parasitology* **101**: 361-67.

Graham, S.V. and J.D. Barry (1991). Expression site-associated genes transcribed independently of variant surface glycoprotein genes in *Trypanosoma brucei. Mol. Biochem. Parasitol.* **47**: 31-42.

Graham, S.V. and J.D. Barry (1995). Transcriptional regulation of metacyclic variant surface glycoprotein gene expression during the life cycle of *Trypanosoma brucei. Mol. Cell. Biol.* **15(11)**: 5945-56.

Greaves, D.R. and P. Borst (1987). *Trypanosoma brucei* variant-specific glycoprotein gene chromatin is sensitive to single-strand-specific endonuclease digestion. *J. Mol. Biol.* **197**: 471-83.

Haber, J.E. (1992). Mating-type switching in *Saccharomyces cerevisiae. Trends Genet.***8(12)**: 446-52.

Hajduk, S.L. and K. Vickerman (1981 a). Absence of detectable alteration in the kinetoplast DNA of a *Trypanosoma brucei* clone following loss of ability to infect the insect vector (*Glossina morsitans*). *Mol. Biochem. Parasitol.* **4**: 17-28.

Hajduk, S.L. and K. Vickerman (1981 b). Antigenic variation in cyclically transmitted *Trypanosoma brucei.* Variable antigen composition of the first parasitaemia in mice bitten by trypanosome-infected *Glossina morsitans. Parasitology* **83**: 609-21.

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

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**: 351.

Hide, G., Buchanan, N., Welburn, S., Maudlin, I., Barry, J.D. and A. Tait (1991). *Trypanosoma brucei rhodesiense*: characterisation of stocks from Zambia, Kenya and Uganda using repetitive DNA probes. *Exp. Parasitol.***72**: 430-39.

Hieter, P., Mann, C., Snyder, M. and R.W. Davis (1985). Mitotic stability of yeast chromosomes: A colony color assay that measures nondisjunction and chromosome loss. *Cell* **40**: 381-92.

Hoffman, J.F-X., Laroche, T., Brand, A.H. and S.M. Gasser (1989). RAP-1 factor is necessary for DNA loop formation *in vitro* at the silent mating type locus HML. *Cell* **57**: 725-37.

Hoheisel, J.D., Nizetic, D. and H. Lehrach (1989). Control of partial digestion combining the enzymes *dam* methylase and *Mbo* I. *Nucleic Acids Res.* **17**: 9571.

Horn, D. and G.A.M. Cross (1995). A developmentally regulated position effect at a telomeric locus in *Trypanosoma brucei. Cell* **83**: 555-61.

Jackson, J.A. and G.R. Fink (1981). Gene conversion between duplicated genetic elements in yeast. *Nature* **292**: 306-11.

Jefferies, D., Tebabi, P. and E. Pays (1991). Transient activity assays of the *Trypanosoma brucei* variant surface glycoprotein gene promoter: control of gene expression at the posttranscriptional level. *Mol. Cell. Biol.* **11(1)**: 338-43.

Kamper, S.M. and A.F. Barbet (1992). Surface epitope variation via mosaic gene formation is potential key to long-term survival of *Trypanosoma brucei. Mol. Biochem. Parasitol.* **53**: 33-44.

Koenig-Martin, E., Yamage, M. and I. Roditi (1992). *Trypanosoma brucei*: a procyclin-associated gene encodes a polypeptide related to ESAG 6 and 7 proteins. *Mol. Biochem. Parasitol.* **55**: 135-46.

Kooter, J.M., van der Spek, H.J., Wagter, R., d'Oliveira, C.E., van der Hoeven, F., Johnson, P.J. and P. Borst (1987). The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in *T. brucei. Cell* **51**: 261-72.

Kooter, J.M. and P. Borst (1984). Alpha-amanitin insensitive transcription of variant sruface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. *Nucleic Acids Res.* **12**: 9457-72.

Koshland, D., Kent, J.C. and L.H. Hartwell (1985). Genetic analysis of the mitotic transmission of minichromosomes. *Cell* **40**: 393-403.

Lamont, G.S., Tucker, R.S. and G.A.M. Cross (1986). Analysis of antigen switching rates in *Trypanosoma brucei. Parasitology* **92**: 355-68.

Landegent, J., Jansen in de Wal, N., Dirks, R. and F. Baao (1987). Use of whole cosmid cloned genomic sequences for chromosomal localisation by non-radioactive *in situ* hybridisation. *Hum. Genet.* **77**: 366-70.

Laurent, M., Pays, E., Van der Werf, A., Aerts, D., Magnus, E., Van Meirvenne, N. and M. Steinert (1984). Translocation alters the activation rate of a trypanosome surface antigen gene. *Nucleic Acids Res.* **12(22)**: 8319-28.

Laurenti, P., Graba, Y., Rosset, R. and J. Pradel (1995). Genetic and molecular analysis of terminal deletions of chromosome 3R of *Drosophila melanogaster. Gene* **154**: 177-81.

Lee, M. G-S., Yaping, E. and N. Axelrod (1995). Construction of trypanosome artificial chromosomes. *Nucleic Acids Res.* **23(23)**: 4893-99.