Coincident Multiple Activations of the Same Surface Antigen Gene in *Trypanosoma brucei*

H. Th. Marc Timmers†, Titia de Lange‡, Jan M. Kooter and Piet Borst

The Netherlands Cancer Institute, Plesmanlaan 121
1066 CX Amsterdam, The Netherlands

Trypanosomes with a coat of variant surface glycoprotein (VSG) 118, consistently appear around day 20 when a rabbit is infected with *Trypanosoma brucei* strain 427. There is a single chromosome-internal gene for VSG 118 and this is activated by duplicative transposition to a telomeric expression site. We show here that the expression-linked extra copy of VSG gene 118 in a day 18 population of a chronic infection is heterogeneous, and we infer that the population is not monoclonal but is the result of multiple independent activations of the 118 gene. We show that the heterogeneity of expression-linked extra copies is also present in other trypanosome populations expressing chromosome-internal VSG genes. We present a model for the timing of VSG gene activation during chronic infection that emphasizes two features: the relative activation and inactivation frequencies of different expression sites, and the degree of homology of the sequences flanking VSG genes with expression sites.

1. Introduction

The African trypanosomes, exemplified by *Trypanosoma brucei*, are unicellular protozoa that can multiply for long periods in the bloodstream of mammalian hosts. The parasite survives the host immune defence by repeatedly changing its surface coat, the only structure in living trypanosomes exposed to host antibodies (Cross, 1984; Turner *et al.*, 1985). Switches of coat composition occur at a rate of 10^{-7} to 10^{-6} per trypanosome division (Cornelissen *et al.*, 1985*a*; Lamont *et al.*, 1986) and are apparently independent of external stimuli, such as antibodies (Myler *et al.*, 1985).

The surface coat of *T. brucei* is composed mainly of one protein species, the variant-specific surface glycoprotein (VSG§; Cross, 1984; Turner *et al.*, 1985). Every trypanosome may have as many as 1000 genes that code for VSGs. Many of these genes reside near the chromosome ends (telomeres) of numerous small chromosomes, but most genes are chromosome-internal and linked in clusters (see

Boothroyd, 1985; Borst, 1986; de Lange, 1986; Donelson & Rice-Ficht, 1985).

Three routes have been described for the activation of a silent VSG gene.

(1) A duplicative transposition of a gene to an active telomeric expression site, where it displaces the resident VSG gene (see Boothroyd, 1985; Borst, 1986; de Lange, 1986; Donelson & Rice-Ficht, 1985).

(2) A reciprocal recombination between two telomeres resulting in the transfer of an inactive telomeric VSG gene to an active expression site in exchange for the resident VSG gene (Pays *et al.*, 1985*a*).

(3) Activation of an inactive expression site (see Boothroyd, 1985; Borst, 1986; de Lange, 1986; Donelson & Rice-Ficht, 1985).

Strong indirect evidence indicates that duplicative transposition occurs by gene conversion, mediated by short blocks of homology that flank the transposed sequence of chromosome-internal VSG genes. Longer blocks of homology exist between telomeres and these may facilitate the transposition of telomeric VSG genes and the reciprocal translocation between telomeres. How inactive expression sites are activated without transposition is still not known (see Boothroyd, 1985; Borst, 1986; de Lange, 1986; Donelson & Rice-Ficht, 1985; Pays, 1985*a,b*).

During a chronic trypanosome infection in mammals, different antigenic variants appear in a

---

† Present address: Department of Medical Biochemistry, State University of Leiden, P.O. Box 9503, 2300 RA Leiden, The Netherlands.

‡ Present address: Department of Microbiology and Immunology HSE 405, School of Medicine, University of California, San Francisco, CA 94143, U.S.A.

§ Abbreviations used: VSG, variant-specific surface glycoprotein; BC, basic copy; ELC, expression-linked extra copy; bp, base-pair(s); kb, 10^3 bases or base-pairs.

loosely programmed order (Capbern et al., 1977). Indeed, such a temporal order is essential for antigenic variation to work at all. Several factors are known to affect this order. Firstly, some trypanosome variants outgrow others (Myler et al., 1985; Seed et al., 1984). Computer modelling has shown that this factor can explain some, but not all, of the temporal order (Kosinski, 1980). Secondly, genes expressed very early in the mammalian infection are invariably located at chromosome ends (Aline et al., 1985b; Liu et al., 1985) and late genes can become early when they are transferred to a telomeric position (Laurent et al., 1984; Michels et al., 1984). Thirdly, the degree of sequence homology between the silent VSG gene and its surroundings, with the expression site, may determine its chance of entering the site (Laurent et al., 1984; Roth et al., 1986). Indeed, Eisen and co-workers have shown that two VSG genes that lack the common elements found at the 3' end of most VSG genes, appear late only during a chronic T. equiperdum infection in rabbits (Roth et al., 1986). Finally, L.H.T. Van der Ploeg (personal communication) has raised the possibility that certain VSGs are not compatible within the same coat. During switching, the trypanosome temporarily makes a mixed coat, in which new and old VSG molecules seem intimately associated (Esser & Schoenbechler, 1985; Michels et al., 1983). If this association were to lead to a non-functional coat, the switch would be unsuccessful. A given trypanosome variant might therefore only be able to switch to a limited number of other variants.

In previous papers we have presented the analysis of several independently isolated variants of T. brucei strain 427 that express the gene for VSG 118 (Michels et al., 1983; de Lange et al., 1983). Strain 427 has a single silent basic copy (BC) gene for VSG 118 in a chromosome-internal position (Michels et al., 1983). Activation of the gene in all variants analysed thus far resulted in an expression-linked extra copy (ELC) gene located in the dominant telomeric expression site of strain 427. During chronic infections in rabbits, variant 118 appears reproducibly around days 18 to 20 (Michels et al., 1983). During the analysis of two additional 118 trypanosome variants, we noted that the expression-linked extra copies present in a single parasitaemia peak were heterogeneous (de Lange et al., 1983). Further analysis of this unexpected phenomenon has shed new light on the timing of VSG gene expression in a chronic infection and on sequences involved in the duplicative transposition of VSG genes to expression sites.

2. Materials and Methods

(a) Trypanosomes

The heterogeneous trypanosome populations from T. b. brucei strain 427 have been described by Michels et al. (1983). The isolation and characterization of cloned trypanosomes have been described by Cross (1975) (MITat 1·2 (221a)), by Michels et al. (1983) (MITat 1·86

(118c), 1·192 (117b) and 1·195 (118d)) and by de Lange et al. (1983) (MITat 1·184 and 1·185).

(b) Molecular cloning

The cloning of BamHI–TaqI fragments of nuclear DNA of MITat 1·184 and 1·185 in Escherichia coli HB101 was performed as described (Van der Ploeg et al., 1982c) with pAT153 as vector. The libraries were screened with a ^{32}P-labelled 950 bp PstI–EcoRI probe for the 3' half of the VSG 118 gene from plasmid TcV118-2 (Michels et al., 1983). Several double-positive clones were identified and one from each library was selected. pTgE 1·84-I contained the 1·47 kb BamHI–TaqI ELC fragment (see Fig. 1) and a 200 bp TaqI fragment of unknown origin. The ELC fragment was recloned in the same vector to obtain pTgE 1·184-II without the contaminating 200 bp fragment. pTgE 1·185-I contained the 1·27 kb 1·185 ELC fragment (see Fig. 1). Plasmid DNA was isolated as described by Birnboim & Doly (1979). The cloned DNAs were checked for colinearity by comigration with nuclear DNA and subsequent blotting and hybridization with the probe used in the library screening.

(c) Nuclear blotting analysis

Isolation of nuclear DNA, restriction enzyme digestion and gel electrophoresis was as described by Bernards et al. (1981). Pulsed-field gradient gel electrophoresis was performed as described (Van der Ploeg et al., 1984a). The gels were subjected to electrophoresis in 20 cm × 20 cm tanks for 19 h with a pulse frequency of 35 s. Gels were incubated for 15 min in 0·25 M-HCl prior to denaturation. Hybridization probes were isolated from appropriate digestions of plasmid DNA by separation in low-melting temperature agarose gels followed by DEAE-cellulose chromatography and labelling by nick-translation (Rigby et al., 1977). Hybridization of ^{32}P-labelled probes to nitrocellulose-bound DNA was as described by Bernards et al. (1981).

(d) Bal31 digestion

Intact nuclear DNA from MITat 1·184 and 1·185 trypanosomes was treated with exonuclease Bal31 (Biolabs) for various times as described (de Lange & Borst, 1982); 0·2 (1·184) and 0·8 unit (1·185) Bal31/μg DNA was used.

(e) Nucleotide sequence analysis

DNA sequences were determined by the Maxam–Gilbert technique (Maniatis et al., 1982). The sequence strategy is indicated in Fig. 1. The 118 BC sequence from Bernards et al. (1986a) was extended from +190 to +274 relative to the 14-mer.

3. Results

Figure 1 shows partial physical maps of the iso-ELCs of three variants (1·184, 1·185 and 1·186) expressing VSG gene 118, a chromosome-internal VSG gene invariably activated by duplicative transposition to a telomeric expression site (Michels et al., 1983). The three variants were isolated on the same day from an infected rabbit (see Fig. 2). Initially, variants 1·184 and 1·185 appeared to have arisen from 1·186 by deletions in the DNA segment
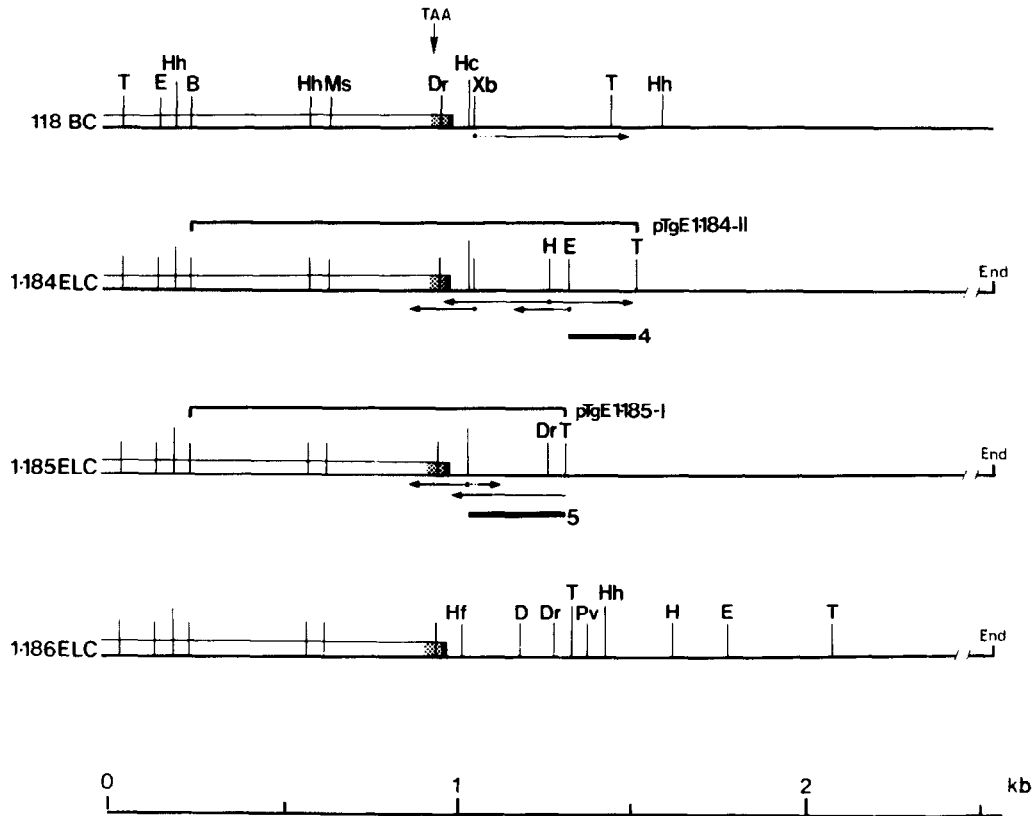
Figure 1. Physical maps of the 3' region of the VSG 118 BC gene and of the 1·184, 1·185 and 1·186 ELC genes. The cloned segments of pTgE 1·184-II and pTgE 1·185-I are overlined. Sequence strategies are indicated by arrows below the maps; large dots mark the labelling sites. The labelling site for the 1·185 sequence from the TaqI site is located in the vector and is not indicated. Common restriction enzyme cutting sites are only indicated in the 118 BC map. Unique sites are named separately. Hybridization probes 4 and 5 are indicated by a filled bar (TAA denotes the stop codon and ends at the tip of the chromosome). Abbreviations: B, *Bam*HI; D, *Dde*I; Dr, *Dra*I; E, *Eco*RI;. H, *Hind*III; Hc, *Hinc*II; Hf, *Hinf*I; Hh. *Hha*I; Ms, *Msp*I; Pv, *Pvu*II; T, *Taq*I; Xb, *Xba*I.

flanking the ELC (de Lange *et al.*, 1983). However, this interpretation is not supported by the detailed restriction maps in Figure 1. To rule out complex rearrangements, we cloned the *Bam*HI–*Taq*I ELC fragments indicated in Figure 1 and determined the

Figure 2. Pedigree of heterogeneous populations and cloned variants of *T. b. brucei* strain 427. The chronic infection is started with the variant on top (Michels *et al.*, 1983; de Lange *et al.*, 1983). The days at which trypanosomes were isolated are indicated by boxes. Variants cloned from these populations are indicated by arrows.

sequence of the areas 3' to the ELC. The results, presented in Figure 3 show clearly that the 3'-flanking sequences of the iso-ELCs differ substantially. Complete homology with the 118 BC gene is lost at position −47 in 1·185, −4 in 1·186 and as far downstream as +252 in 1·184. This makes it unlikely that 1·184 and 1·185 could have arisen from 1·186, or 1·184 from 1·185. It is also unlikely that variant 1·186 has arisen from one of the other two, as several restriction enzymes cutting downstream from the 1·186 ELC, do not cut downstream from the ELCs of variants 1·184 (*Pvu*II and *Msp*I) and 1·185 (*Pvu*II, *Eco*RI, *Hind*III and *Msp*I) (data not shown).

Further to exclude substantial homology between the sub-telomeric regions of the ELCs of the three variants, sub-telomeric probes were hybridized to chromosome-sized DNA molecules separated by pulsed-field gradient gel electrophoresis. Previous work has shown that the BC gene of VSG 118 resides in a chromosome that sticks in the slot, and that the VSG 118 ELCs reside in one of the chromosomes in the compression zone (Van der Ploeg *et al.*, 1984*a*; Johnson & Borst, 1986). Figure 4 shows that this also holds for the three variants studied here. The main point of the experiment is, however, that the sub-telomeric
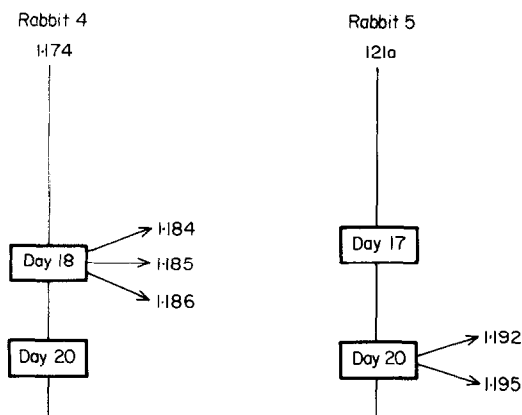
```
                                                                          Stop        -38         -28
118               TTC GCC CTA ATG GTT TCT GCT GCT TTT GTT ACC TTG CTT TTT TAA    GTTTTTTCCC  CCTCAATTTA
1·184 ELC         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...    ..........  ..........
1·185 ELC         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...    ⁻.........  ..........
1·186 ELC         ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...    ..........  ..........
                                                                                     ↑1·185
                       -18          -8                   14-mer                24          34          44
118     BC        AAAGAATTTT   TGCTACTTGA   AAA-CTTC  TGATATATTTTAAC  ACCTTTAAAT  TCTGCCGAAA  AATTTGAACT
1·184 ELC         ..........   ..........   ...⁻....  ..............  ..........  ..........  ..........
1·185 ELC         ..........   ..........   ...⁻....  ..............  ....G...G.  .A.CTGA...  .T.GA.CTG.
1·186 ELC         ..........   ..........   ...A...T  ..............  ..G.AATTTC  .GCCGAT...  ..A.......
                                               ↑1·186
                       54          64          74          84          94          104         114
118     BC        GTTTTTGTAA   GTTGACTGTC  TGATTGTCTA  GAAATTTTTC  TGGCAATTAA  AACTTTTTTT  CCTTCTTTCC
1·184 ELC         ..........   ..........  ..........  ..........  ..........  ..........  ..........

                       124         134         144         154         164         174         184
118     BC        TTTTCCTGTC   TTTTTTGTGG  AAGTAGGGGG  TTCATTAAGG  GGGTTAGGGT  TATGGTTTTT  TATTTTACTT
1·184 ELC         ..........   ..........  ..........  ..........  ..........  ..........  ..........

                       194         204         214         224         234         244         254
118     BC        TTGGTTGTTT   TTTATATTTT  TTCTTAGTAA  TGCAGCGGTT  GTTTTTGTTG  ATGTGGACCC  TAACCCTTAT
1·184 ELC         ..........   ..........  ..........  ..........  ..........  ..........  .......CC.
                                                                                               1·184↑
                       264         274
118     BC        TTGAAAAATA   AACACCTTTC
1·184 ELC         CCCCCCCCGC   GCGC..CACA
```

Figure 3. Sequence comparison of the 3′ end and flanking region of the 118 BC, 1·184, 1·185 and 1·186 ELC genes. The 118 BC and 1·186 ELC sequences are from Bernards *et al.* (1986*a*) and de Lange *et al.* (1983), respectively. Numbering of the bases is relative to the first base of the 14-mer. Bases in the coding region are given in triplets, according to the reading frame. "Stop" marks the stop codon. Dots indicate identical bases and dashes indicate gaps introduced to maximize homology. Arrows mark the position where the ELC sequence deviates from the 118 BC sequence and thus the divergence points.

probes from the 1·184 and 1·185 ELCs only hybridize to homologous chromosomes and not to chromosomes carrying the other ELCs (Figs 1 and 4).

We conclude from these results that the three 118 iso-ELCs are associated with different 3′ sequences. We have shown (de Lange *et al.*, 1983) that there are multiple sub-telomeric copies of the sequence

flanking the 1·186 ELC in the *T. brucei* genome, and that one of these had entered the expression site in variant 1·186 by duplicative transposition (telomere conversion). Figure 4 shows that there are also multiple copies of the sequences associated with the ELCs 1·184 and 1·185 in *T. brucei*, and Figure 5 shows that most (1·184) or all (1·185) of these sequences are preferentially digested by exonuclease

Figure 4. Chromosomal localization of sub-telomeric regions. The trypanosome variants, from which the chromosomes were prepared, are indicated above the slots. Chromosome-sized DNA molecules were separated by pulsed-field gradient gel electrophoresis and transferred to nitrocellulose. The filter was successively hybridized to a 950 bp probe for the 3′ half of the VSG 118 gene, probe 4 and probe 5 (see Fig. 1), as indicated below the panels. Probes were melted off between the successive hybridizations in 0·1 M-NaOH for 5 min. Hybridization stringency was 0·1 × SSC at 65°C (left panel) and 0·3 × SSC at 65°C (middle and right panels) (SSC is 0·15 M-NaCl, 0·015 M-sodium citrate, pH 7·0). Molecular weights are as determined by Bernards *et al.* (1986*b*).
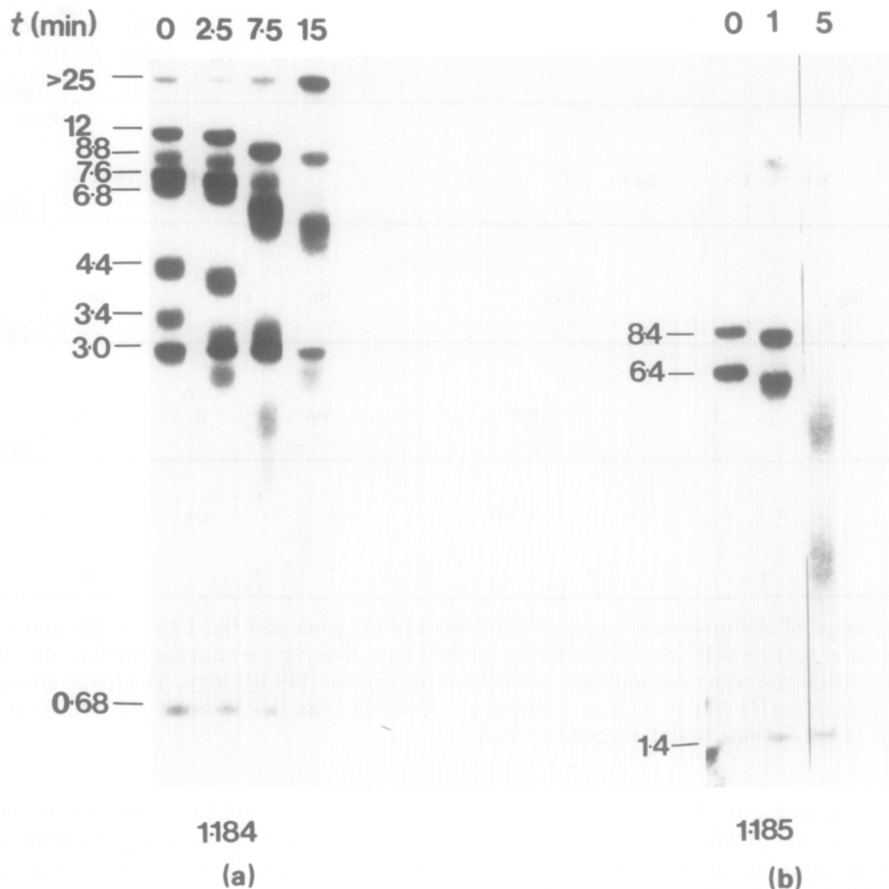
Figure 5. Preferential sensitivity to exonuclease *Bal*31 of DNA fragments recognized by the sub-telomeric segments of the 1·184 and 1·185 ELC. (a) The hybridization to probe 4, and (b) to probe 5. The hybridization stringency was 0·3 × SSC at 65 °C. Each lane contained 4 μg of nuclear DNA from variant MITat 1·184 (a) or MITat 1·185 (b). Prior to *Ava*II digestion the DNA was treated with exonuclease *Bal*31 for the times indicated. In (b) *Bal*31 activity/μg nuclear DNA was 4 times higher than in (a). Fragment sizes (in kb) are indicated on the left. The 7·6 kb (a) and the 6·4 kb fragment (b) carried the 1·184 and the 1·185 ELC telomere, respectively.

*Bal*31, indicating their sub-telomeric position. This supports a scenario in which the sub-telomeric sequences associated with the ELCs of all three variants entered the expression site by telomere conversion. The crucial question then is whether the 118 gene came into the expression site before or after the sub-telomeric sequences arrived. To analyse this, we have made a detailed map of the sequences upstream from the three ELCs. Figure 6 shows that restriction sites up to 1·1 kb in front of the BC gene are unaltered in the ELC, suggesting that this 1·1 kb is co-transposed to the expression site, as seen in previous transpositions of the 118 gene (Michels *et al.*, 1983). Beyond this common segment, however, the physical maps of the three ELCs diverge. This provides strong support for the idea that the three iso-ELCs arose from independent transpositions of the 118 gene (see Discussion).

Whereas the 1·185 and 1·186 ELCs display the same configuration of far-upstream *Bgl*III, *Eco*RI, *Taq*I, *Pst*I and *Hind*III sites characteristic of the dominant expression site of this *T. brucei* strain (Michels *et al.*, 1983), the 1·184 ELC has a completely different set of sites. Nevertheless, this

ELC is also located in the chromosome that contains the dominant expression site (Johnson & Borst, 1986, and unpublished results). This re-emphasizes that the configuration of upstream restriction sites is not a reliable criterion to identify expression sites, a point that we have stressed before (Van der Ploeg *et al.*, 1984*a*; Borst, 1986). Of course, we cannot exclude the remote possibility that the 1·184 ELC is in the same chromosome as the dominant site, but not in the same telomere.

To test whether the three variants analysed here were the only variants expressing the 118 gene at day 18 in rabbit 4, we have analysed the total uncloned populations after amplification (and therefore possible selection) in mice and rats (see Michels *et al.*, 1983). Figure 7(a) shows that this heterogeneous collection of trypanosomes contained bands corresponding to the iso-ELCs of variants 1·184 (weak), 1·185 and 1·186, but also two to three other weak bands. Only the single BC band of the VSG 118 gene is visible in the day 20 trypanosome populations isolated from rabbit 4 (Fig. 7). We have already shown in a previous paper (de Lange *et al.*, 1983) that no extra bands were present in the day 15 population either. This strongly suggests
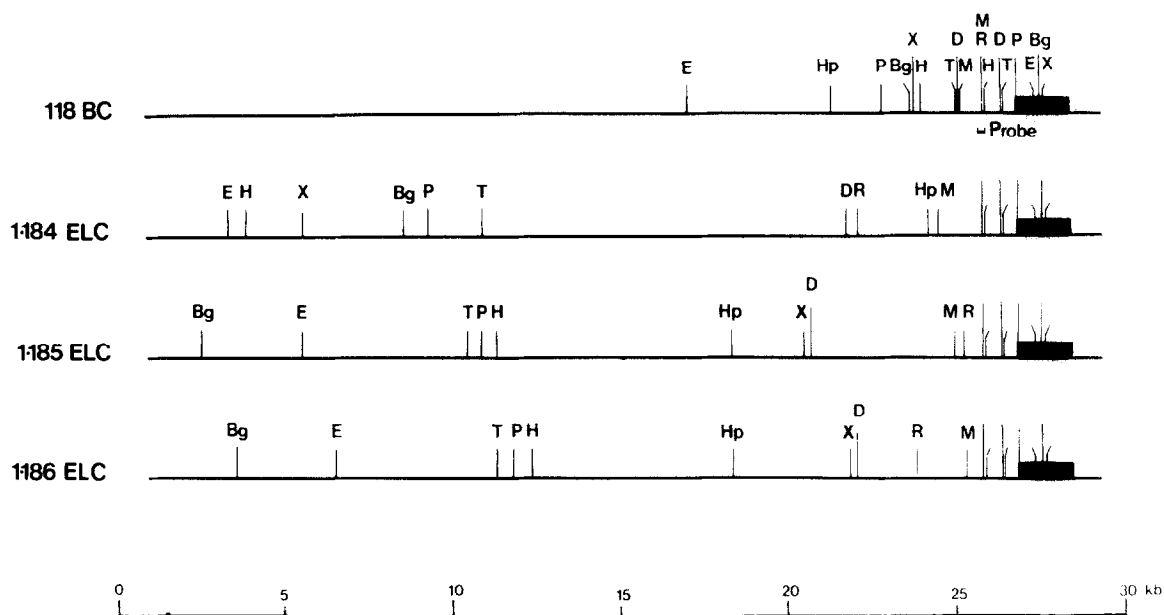
Figure 6. Physical maps of the upstream region of the VSG 118 BC gene and the 1·184, 1·185 and 1·186 ELC genes. The portion of the gene in mature mRNA is indicated by a filled box. Restriction enzyme cutting sites that are common in all maps, and thus within the transposition unit, are named only in the 118 BC map. The maps are constructed using as probe a 205 bp *Pvu*II–*Hin*dIII fragment from plasmid pc118-29III (Van der Ploeg *et al.*, 1982b), as indicated below the 118 BC map. For abbreviations, see the legend to Fig. 1.

that all additional hybridizing bands in the day 18 VSG 118 population are attributable to iso-ELC bands, and that this population is made up of at least five to six variants that have independently switched on the VSG 118 gene. The alternative explanation that restriction sites are only partially cleaved, because of partial modification, is highly unlikely. No modification has been observed in *T. brucei* DNA for any sites cleaved by *Dde*I or *Rsa*I and *Taq*I, that also gave multiple bands (data not shown).

To test whether multiplicity of variants expressing the 118 VSG gene was a peculiarity of the day 18 population from rabbit 4, we have also checked heterogeneity of putative ELC bands in another rabbit (rabbit 5 in Fig. 2). The DNA was hybridized with probes for the transposition unit of VSG gene 118 or of VSG gene 117, another chromosome-internal VSG gene activated by duplicative transposition (see Hoeijmakers *et al.*, 1980; Van der Ploeg *et al.*, 1982a; Michels *et al.*, 1983). The results in Figure 7(b) show that the day 20 trypanosome population from rabbit 5 contains three extra bands hybridizing with the 118 cDNA probe that remained unaltered at high restriction enzyme concentrations. One of these is enriched in a cloned trypanosome, MITat 1·192 or 118d, isolated previously (Michels *et al.*, 1983). None of the extra bands is visible in another trypanosome clone, isolated from this population, that expresses the VSG 117 gene. Analogous results are obtained with the 5′ end 117-transposition unit probe (Fig. 7(c)). The extra bands in the day 20 populations remained unaltered at high restriction enzyme concentrations (data not shown). We conclude that the heterogeneity of ELC bands is not just a

peculiarity of VSG 118 variants or the trypanosome populations isolated on day 18 from rabbit 4, but a common feature in VSG gene activation.

4. Discussion

(a) *Multiple interchromosomal transposition events are responsible for the heterogeneity of the trypanosome population that expresses VSG gene 118*

We show in this paper that the populations of trypanosome variants that express a chromosome-internal VSG gene are heterogeneous in the ELCs that they contain. Similar results have been obtained independently in another chronic infection with this trypanosome strain by G.-S. Lee & L.H.T. Van der Ploeg (personal communication) and some ELC heterogeneity may also have been observed in another strain by Pays *et al.* (1983).

Our results show that the three 118 variants analysed here cannot have arisen from each other by *intra*-chromosomal deletions or rearrangements. Hence, they must have originated either from three independent transpositions of the same VSG gene into different surroundings, or from a single transposition followed by alteration of flanking sequences in some of the progeny by transposition of sequences from other telomeres. We do not favour the latter alternative for three reasons. Firstly, it requires that both flanks of the ELC were invaded in the variants analysed. Independent invasion of upstream and downstream flanks would be highly unlikely at a switching rate of 10^{-6}; a simultaneous invasion by another telomere would not be expected to leave the 118 gene unscathed. Secondly, it requires that invasion occurred shortly
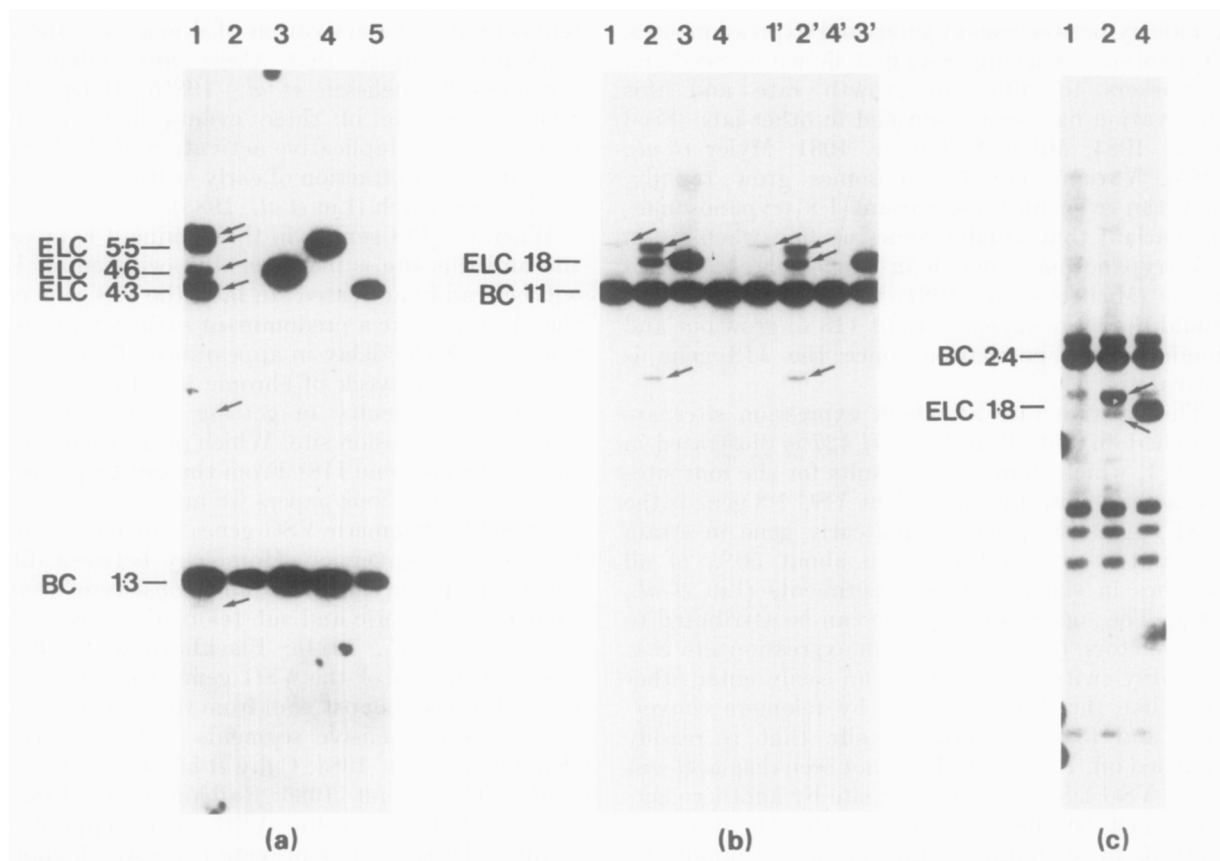
Figure 7. Multiple ELCs for VSG 118 and VSG 117 are present in heterogeneous populations. Autoradiographs of blots containing DNA from heterogeneous populations and their cloned variants (see Fig. 2). (a) Lanes contain 3 μg of nuclear DNA digested with *Dde*I from rabbit 4 day 18 (lane 1), rabbit 4 day 20 (lane 2), MITat 1·184 (lane 3), MITat 1·185 (lane 4) and MITat 1·186 (lane 5). The blot was hybridized to a 205 bp *Pvu*II–*Hind*III probe (see Fig. 6). (b) and (c) Lanes contain 3 μg of nuclear DNA digested with *Eco*RI (b) or *Mbo*I (c) from rabbit 5 day 17 (lane 1), rabbit 5 day 20 (lane 2), MITat 1·192 (118d, lane 3) and MITat 1·195 (117b, lane 4). The blot was hybridized to the 550 bp *Pst*I–*Eco*RI probe from pTcV 118-2 (Michels *et al.*, 1983) containing the 5′ half of the VSG 118 gene (b) or to the 579 bp *Hph*I probe from pTgE 117b-I (de Lange *et al.*, 1985) containing the 5′ breakpoint of the 117b transposition unit (c). This latter probe shows extensive cross-hybridization, because it contains 70 bp repeats. The hybridization stringency was 0·1 × SSC at 65 °C. Fragment sizes (in kb) are indicated on the left. Arrows denote ELCs present in the heterogeneous populations. To exclude that the extra bands were partial digestion products, DNA samples were split after digestion and half were redigested with the same enzyme. DNAs in lanes 1 to 4 were digested once, in lanes 1′ to 4′ twice.

after the VSG 118 gene had entered the expression site, since the sequences around an ELC are stable once the variant is established (Michels *et al.*, 1983; Bernards *et al.*, 1983). Finally, it requires that telomere conversion must have run up to at least 20 kb upstream from the VSG gene without dislodging the gene itself, to explain the differences in far-upstream sites between the 1·184 ELC and the other two ELCs. Hence, we prefer the interpretation that variants 1·184, 1·185 and 1·186 result from independent VSG gene 118 transpositions into different surroundings. A more detailed dissection of the genealogical relationships between variants is unfortunately complicated by the limitations of the biological system. Chronic trypanosome infections can be studied only in animals like rabbits that are relatively resistant against trypanosomiasis. Resistance implies low parasitaemia, necessitating amplification of trypanosome populations in mouse or rat for biochemical analysis. Hence, the final trypanosome

population studied is a far cry from that present in the rabbit, and precise genealogical relationships between trypanosome variants can not normally be studied in this system. These can be analysed in single relapse experiments in which a few trypanosomes from a homogeneous population are introduced in a mouse or rat and allowed to switch once. Single relapse experiments give only information about the initial phase of infection, in which activation of the 118 BC gene does not occur. By combination of the data from chronic infections and single relapse experiments, however, a speculative scenario may be constructed for the timing of VSG gene 118.

(b) *Factors that determine the timing of VSG gene expression during a chronic infection*

Three factors are known (or thought) to affect timing: trypanosome growth rate, the relative rate at which expression sites switch on and off, and the

homology between silent genes and expression sites. Trypanosome variants were first shown by Seed and co-workers to differ in growth rate and this observation has been confirmed in other labs (Seed et al., 1984; Miller & Turner, 1981; Myler et al., 1985). Variant 118 trypanosomes grow rapidly, they can even outgrow variant 1·8 trypanosomes, the variant that usually comes up first when strain 427 trypanosomes switch in single relapse experiments (Michels et al., 1984; Liu et al., 1985). One would therefore expect variant 118 to grow out and dominate the population, once the 118 gene is activated.

The relative rate at which expression sites are switched on and off in T. brucei 427 is illustrated in Table 1, which summarizes results for the four sites characterized in this strain. The VSG 1·8 gene is the most aggressive predominant early gene in strain 427 and it is switched on in about 50% of all switches in single relapse experiments (Liu et al., 1985). The success of this gene can be attributed to three factors: (1) it resides in an expression site that is readily switched on; (2) it can easily enter other sites, like the dominant site, by telomere conversion; and (3) it occupies a site that is readily switched off. In fact, we have not seen displacement of the VSG 1·8 gene in the 1·8 site by another gene. This must be the main reason why the 1·8 gene succeeds in remaining in this site and retaining its predominant early position. The dominant site is readily switched on and off as well (Michels et al., 1984), but genes are also frequently displaced from this site by other genes. The dominant site is the only one that we have seen active after the first weeks of a chronic infection (Liu et al., 1985). The 221 site, in contrast, was activated only on a single occasion. This site is not easily switched off either; the 221 gene is usually displaced by gene conversion (Bernards et al., 1985; Liu et al., 1985, and unpublished results). Antigenic variation by telomere switching requires the inactivation of the old telomere and the activation of a new one. Indirect evidence suggests that these are independent processes (Cornelissen et al., 1985b; Baltz et al., 1986). The rate of these events must be high relative to the duplicative activation of VSG genes, as a substantial fraction of early switches are due to a telomere switch (Liu et al., 1985).

When the 118 gene is in the dominant expression site and this site is inactive, this switched-off ELC will be readily activated. In fact, the 118 gene copy then behaves like a predominant early VSG gene. It follows that the delay in appearance of variant 118 until the third week of chronic infections must be due to its difficulty in getting (stably) into the dominant expression site. Which genes then occupy the site before gene 118? From the results presented here and in previous papers we infer that the site is occupied by telomeric VSG genes that have entered by telomere conversion. Homology between different telomeres is extensive and consists of several elements: telomeric and sub-telomeric repeats (Van der Ploeg et al., 1984b; Blackburn & Challoner, 1984); segments of the VSG genes themselves and the 70 bp repeats upstream from the VSG genes (see below); and extensive segments further upstream (Kooter & Borst, 1984; Cully et al., 1985; Bernards et al., 1986b; Borst, 1986; Raibaud et al., 1986) as far as 50 kb 5' to the VSG gene (unpublished results). These should provide extensive homology for telomeric genes to enter expression sites by gene conversion or reciprocal recombination. Analysis of single relapse experiments has shown that very early variants arise either by switching on new expression sites or by telomere conversion of active sites, never by duplicative transposition of chromosome-internal genes (see Liu et al., 1985). We therefore conclude that the transposition rate of chromosome-internal genes, like VSG gene 118, is too low to compete with alternative expression site activation and telomere conversion. Although this explains why variant 118 does not appear in the second week, it is not obvious why it should appear around day 20. There are some 250 telomeres in T. brucei 427 (Van der Ploeg et al., 1984b), many (if not all) carrying VSG genes. Together they might keep the dominant expression site occupied for over 100 weeks. Four factors might prevent this. Firstly, some telomeres may carry defective VSG genes (see Bernards et al., 1985) or genes for coats of slow-growing variants. Secondly, it is unlikely that all telomeres have extensive homology with expression site telomeres, as expression site probes only hybridize to about 20 to 30 bands in DNA blots (Cully et al., 1985). Thirdly, the strain 427 trypanosome populations in chronically infected rabbits are rather heterogeneous in the early weeks of infection, leading to inefficient use of the repertoire. Finally, the duplicative transposition of the 118 gene may be less efficient than telomere conversion, but as the rate is not zero, the chance of producing one is only a matter of cumulative population size after starting an infection.

Chromosome-internal genes are thought to use

Table 1

VSG gene expression sites in strain 427 of
T. brucei†

Expression site	Frequency of site: activation	inactivation	Frequency of gene displacement in single relapses	Use later in chronic infections
Dominant site (118)	+ + +	+ +	+ +	+ + +
1·8 site‡	+ + + +	+ + +	?	
1·1 site ‡	+ +	?	?	
221 site‡	+	+ §	+ + +	

† Based on data from Liu et al. (1985), Cornelissen et al. (1985b) and Johnson & Borst (1986).

‡ Expression sites are called after the VSG gene that was first found in it.

§ Inactivation by promoter switch-off has only recently been observed for this site (Cornelissen, unpublished results). In addition Cornelissen et al. (1985b) have reported one variant in which this site was silenced by a 30 kb DNA insertion between the promoter and gene 221.

short blocks of sequence homology with expression sites to enter these sites by duplicative transposition (Van der Ploeg *et al.*, 1982*a*; Laurent *et al.*, 1984). It is therefore plausible that the degree of homology could co-determine timing (Laurent *et al.*, 1984). At the 5′ edge these repetitive sequences consist of imperfect 70 bp repeats found about 1·5 kb in front of VSG gene 118 (Liu *et al.*, 1983) and many other chromosome-internal VSG genes (Van der Ploeg *et al.*, 1982*a,c*; Aline *et al.*, 1985*a*). Long arrays of 70 bp repeats are present in front of VSG genes in expression sites (Campbell *et al.*, 1984; Bernards *et al.*, 1985) and the divergence point between BC and ELC lies somewhere in these repeats (Campbell *et al.*, 1984; de Lange *et al.*, 1985). There are indications that the number and sequence of 70 bp repeats varies (Van der Ploeg *et al.*, 1982*c*; Aline *et al.*, 1985*a*) and thus could co-determine the avidity of the BC gene for expression sites.

At the 3′ edge of the transposed segment, VSG genes share short sequence blocks in the area corresponding to the 3′ untranslated part of the mRNA (see Michels *et al.*, 1983). Figure 8 illustrates that the homology between the BC gene and the expression site extends even beyond the end of the gene. It is unlikely, however, that this downstream area plays a decisive role in timing, as there is hardly any homology between these areas of VSG genes 118 and 117 (Fig. 8), genes that are both activated around day 20 of a chronic infection (Michels *et al.*, 1983). The involvement in timing of the sequence blocks in the area corresponding to the 3′ untranslated part of the mRNA is indicated, however, by recent results with *T. equiperdum*, a trypanosome species closely related to *T. brucei*. Roth *et al.* (1986) found that two BC genes activated late in infections lack the 3′ end of the gene, and suggested that lack of homology with the resident gene limits the chance for these incomplete genes of entering expression sites. Such genes have to wait until a preceding gene in an expression site provides the homology required for transposition (Laurent *et al.*, 1984; Roth *et al.*, 1986). Of course, maximal homology is provided by identical or

closely related genes in the expression site and, indeed, Pays and co-workers (Pays, 1986*a*; Pays *et al.*, 1985*b*) have reported striking examples of segmental gene conversions involving members of the same gene family. How the trypanosome avoids the risk of switching on the same gene or its family member again and again is a matter of speculation. Relatively short blocks of homology might be sufficient for efficient gene conversion, or the ability to switch expression sites may prevent complete take-over by a large dominant gene family.

In conclusion, we think that the following factors determine the appearance of 118 variants around day 20. Initially, the infection would be dominated by the expression site that contains the 1·8 gene. After the first week, another expression site would usually take over, e.g. the dominant site that appears to be preferentially used in later weeks. At first, this site would be occupied mainly by telomeric genes with a high degree of homology with this site in the far upstream area. As these are expended, chromosome-internal genes with flanking sequences with a high homology with the type of telomeric genes that can enter the dominant site would be able to occupy the site. Of course, only genes that give rise to variants with high growth rate, like the 118 gene, would be successful. In the first weeks of a chronic infection in rabbits, the trypanosome populations are not antigenically homogeneous (Michels *et al.*, 1983, and unpublished results). We assume therefore that several different telomeric genes occupy the dominant expression site at that time, each with different flanking regions. These are the flanking regions that end up in the different 118 variants that arise from independent transposition events.

How many potentially activatable expression sites are present in one trypanosome is still unknown. We consider it likely that the number is closer to 10 than to 100, and that these 10 are more often off than on. In this way the fraction of trypanosomes that express more than one VSG gene would remain low enough to allow long-term survival in mammals.
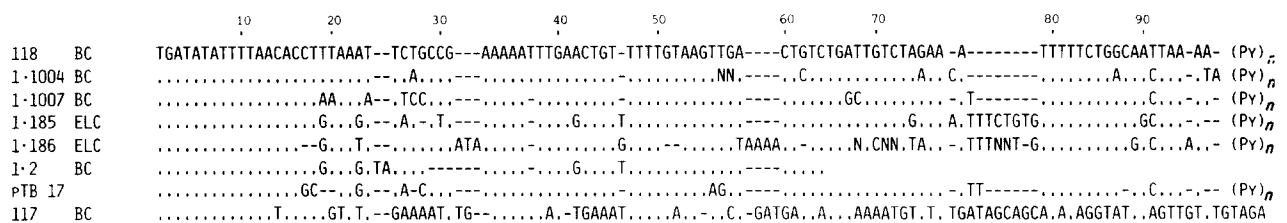

Figure 8. Sequence comparison of 3′-flanking regions of VSG genes, downstream from the 14-mer. Short homology blocks upstream from the 14-mer have been reported by Michels *et al.* (1983). Sequences of VSG genes are aligned on the basis of their homology with the 118 BC sequence. Only nucleotides differing from the 118 BC are denoted. Dots mark identical nucleotides and gaps are introduced to obtain maximal homology. Numbering is relative to the first nucleotide of the 14-mer in the 118 BC sequence. N marks unidentified nucleotides. (Py)$_n$ indicates a pyrimidine track, which is a common feature for the 118-like genes (Bernards *et al.*, 1986*a*). The 117 BC sequence has little homology with the 118 BC sequence in this region. This indicates that homology in this region is not a common feature for genes activated relatively early by duplicative transposition, but that it is restricted to a subclass of VSG genes. The 118 BC and 1·1004 BC sequences are from Bernards *et al.* (1986*a*); the 1·186 ELC sequence is from de Lange *et al.* (1983); the pTB17 sequence is from Blackburn & Challoner (1984); the 1·2 BC sequence is from J. Donelson (personal communication); and the 117 BC sequence is from Bernards *et al.* (1981).

References

Aline, R., MacDonald, G., Brown, E., Allison, J., Myler, P., Rothwell, V. & Stuart, K. (1985a). *Nucl. Acids Res.* **13**, 1361–1377.

Aline, R. F., Scholler, J. K., Nelson, R. G., Agabian, N. & Stuart, K. (1985b). *Mol. Biochem. Parasitol.* **17**, 311–320.

Baltz, T., Giroud, C., Baltz, D., Roth, C., Raibaud, A. & Eisen, H. (1986). *Nature (London)*, **319**, 602–604.

Bernards, A., Van der Ploeg, L. H. T., Frasch, A. C. C., Borst, P., Boothroyd, J. C., Coleman, S. & Cross, G. A. M. (1981). *Cell*, **27**, 497–505.

Bernards, A., Michels, P. A. M., Linke, C. R. & Borst, P. (1983). *Nature (London)*, **303**, 592–597.

Bernards, A., Kooter, J. M. & Borst, P. (1985). *Mol. Cell. Biol.* **5**, 545 553.

Bernards, A., Van der Ploeg, L. H. T., Gibson, W. C., Leegwater, P., Eijgenraam, F., de Lange, T., Weijers, P., Calafat, J. & Borst, P. (1986a). *J. Mol. Biol.* **190**, 1–10.

Bernards, A., Kooter, J. M., Michels, P. A. M., Moberts, R. M. P. & Borst, P. (1986b). *Gene*, **42**, 313–323.

Birnboim, H. & Doly, J. (1979). *Nucl. Acids Res.* **7**, 1513–1523.

Blackburn, E. H. & Challoner, P. B. (1984). *Cell*, **36**, 447–457.

Boothroyd, J. C. (1985). *Annu. Rev. Microbiol.* **39**, 475–502.

Borst, P. (1986). *Annu. Rev. Biochem.* **55**, 701–732.

Campbell, D. A., Van Bree, M. P. & Boothroyd, J. C. (1984). *Nucl. Acids Res.* **12**, 2759–2774.

Capbern, A., Giroud, C., Baltz, T. & Mattern, P. (1977). *Exp. Parasitol.* **42**, 6–13.

Cornelissen, A. W. C. A., Michels, P. A. M., Borst, P., Spanjer, W., Versluys-Broers, J. A. M., Van der Meer, C., Farzaneh, F. & Shall, S. (1985a). *Biochem. Pharmacol.* **34**, 4151–4156.

Cornelissen, A. W. C. A., Johnson, P. J., Kooter, J. M., Van der Ploeg, L. H. T. & Borst, P. (1985b). *Cell*, **41**, 825–832.

Cross, G. A. M. (1975). *J. Parasitol.* **71**, 393–417.

Cross, G. A. M. (1984). *Phil. Trans. Roy. Soc. ser. B*, **307**, 3–12.

Cully, D. F., Ip, H. S. & Cross, G. A. M. (1985). *Cell*, **42**, 173–182.

de Lange, T. (1986). *Int. Rev. Cytol.* **99**, 85–117.

de Lange, T. & Borst, P. (1982) *Nature (London)*, **299**, 451–453.

de Lange, T., Kooter, J. M., Michels, P. A. M. & Borst, P. (1983). *Nucl. Acids Res.* **11**, 8149–8165.

de Lange, T., Kooter, J. M., Luirink, J. & Borst, P. (1985). *EMBO J.* **4**, 3299–3306.

Donelson, J. E. & Rice-Ficht, A. C. (1985). *Microbiol. Rev.* **49**, 107–125.

Esser, K. M. & Schoenbechler, M. J. (1985). *Science*, **229**, 190–193.

Hoeijmakers, J. H. J., Frasch, A. C. C., Bernards, A.,

Borst, P. & Cross, G. A. M. (1980). *Nature (London)*, **284**, 78–80.

Johnson, P. J. & Borst, P. (1986). *Gene*, **43**, 213–220.

Kooter, J. M. & Borst, P. (1984). *Nucl. Acids Res.* **12**, 9457–9472.

Kosinski, R. J. (1980). *Parasitology*, **80**, 343 357.

Lamont, G. S., Tucker, R. S. & Cross, G. A. M. (1986). *Parasitology*, **92**, 355–368.

Laurent, M., Pays, E., Van der Werf, A., Aerts, D., Magnus, E., van Meirvenne, N. & Steinert, M. (1984). *Nucl. Acids Res.* **12**, 8319–8329.

Liu, A. Y. C, Van der Ploeg, L. H. T., Rijsewijk, F. A. M. & Borst, P. (1983). *J. Mol. Biol.* **167**, 57–75.

Liu, A. Y. C., Michels, P. A. M., Bernards, A. & Borst, P. (1985). *J. Mol. Biol.* **182**, 383–396.

Longacre, S. & Eisen, H. (1986). *EMBO J.* **5**, 1057–1063.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). In *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Michels, P. A. M., Liu, A. Y. C., Bernards, A., Sloof, P., Bijl, M. M. W., Van der Schinkel, A. H., Menke, H. H., Borst, P., Veeneman, G. H., Tromp, M. C. & Van Boom, J. H. (1983). *J. Mol. Biol.* **166**, 537–556.

Michels, P. A. M., Van der Ploeg, L. H. T., Liu, A. Y. C. & Borst, P. (1984). *EMBO J.* **3**, 1345–1351.

Miller, E. N. & Turner, M. J. (1981). *Parasitology*, **82**, 63–80.

Myler, P. J., Allen, A. L., Agabian, N. & Stuart, K. (1985). *Infect. Immun.* **47**, 684–690.

Pays, E. (1986a) *Progr. Nucl. Acids Res. Mol. Biol.* **32**, 1–26.

Pays, E. (1986b) *Trends Genet.* **2**, 21–26.

Pays, E., Van Assel, S., Laurent, M., Dero, B., Michiels, F., Kronenberger, P., Matthysens, G., Van Meirvenne, N., Le Ray, D. & Steinert, M. (1983). *Cell*, **34**, 359–369.

Pays, E., Guyaux, M., Aerts, D., Van Meirvenne, N. & Steinert, M. (1985a). *Nature (London)* **316**, 562–564.

Pays, E., Houard, S., Pays, A., Van Assel, S., Dupont, F., Aerts, D., Huet-Duvillier, G., Gomes, V., Richet, C., Degand, P., Van Meirvenne, N. & Steinert, M. (1985b). *Cell*, **42**, 821–829.

Raibaud, A., Buck, G., Baltz, T. & Eisen, H. (1986). *Mol. Cell. Biol.* **6**, 2950 2956.

Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977). *J. Mol. Biol.* **113**, 237 251.

Roth, C. W., Longacre, S., Raibaud, A., Baltz, T. & Eisen, H. (1986). *EMBO J.* **5**, 1065 1070.

Seed, J. R., Edwards, R. & Sechelski, J. (1984). *J. Protozool.* **31**, 48 53.

Turner, M. J., Cardoso Del Almeida, M. L., Gurnett, A. M., Raper, J. & Ward, J. (1985). *Curr. Top. Microbiol. Immunol.* **117**, 23–55.

Van der Ploeg, L. H. T., Bernards, A., Rijsewijk, F. A. M. & Borst, P. (1982a). *Nucl. Acids Res.* **10**, 593–609.

Van der Ploeg, L. H. T., Liu, A. Y. C., Michels, P. A. M., de Lange, T., Borst, P., Majumder, H. K., Weber, H., Veeneman, G. H. & Van Boom, J. (1982b). *Nucl. Acids Res.* **10**, 3591 3604.

Van der Ploeg, L. H. T., Valerio, D., de Lange, T., Bernards, A., Borst, P. & Grosveld, F. G. (1982c). *Nucl. Acids Res.* **10**, 5905–5923.

Van der Ploeg, L. H. T., Schwartz, D. C., Cantor, C. R. & Borst, P. (1984a). *Cell*, **37**, 77–84.

Van der Ploeg, L. H. T., Liu, A. Y. C. & Borst, P. (1984b). *Cell*, **36**, 459–468.