J. Mol. Biol. (1986) **190**, 1–10

Rapid Change of the Repertoire of Variant Surface Glycoprotein Genes in Trypanosomes by Gene Duplication and Deletion

A. Bernards[a], L. H. T. Van der Ploeg[b], W. C. Gibson P. Leegwater[c], F. Eijgenraam[d], T. De Lange[e], P. Weijers[f] J. Calafat and P. Borst

Division of Molecular Biology
The Netherlands Cancer Institute
Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

(Received 28 January 1986)

To study the evolution of the variant surface glycoprotein (VSG) repertoire of trypanosomes we have analysed the DNA region surrounding the VSG 118 gene in different trypanosome strains. We find a remarkable degree of variation in this area. Downstream from the 118 gene a $5{\cdot}7 \times 10^3$ base-pair DNA segment containing a potential VSG gene has been quadruplicated in strain 427 of *Trypanosoma brucei*, but not in most other strains analysed. The VSG 1.1000 gene, located immediately upstream from the 118 gene in one trypanosome strain, has been cleanly deleted in another. Our results are most easily explained by multiple unequal cross-overs between sister chromatids and are the first indication that sister chromatid exchange occurs in trypanosomes.

1. Introduction

African trypanosomes, such as *Trypanosoma brucei*, evade the immune response of their mammalian host by the sequential production of different variant surface glycoprotein, which form a dense coat around the parasite and are the major parasite antigens recognized by the host (Vickerman, 1969, 1978; Cross, 1975, 1978). Each trypanosome makes a single type of VSG† at a time, but during a chronic infection a large repertoire of VSGs is sequentially produced. In *Trypanosoma equiperdum*, for example, 101 different VSGs have been identified by immunological means (Capbern *et al.*, 1977), and *T. brucei* may have an equally large arsenal of VSG genes at its disposal (Van der Ploeg *et al.*, 1982*a*). Studies on the mechanism of trypanosome antigenic variation have shown that trypanosomes encode each VSG in a separate gene and that activation of VSG genes often involves their duplicative transposition to one of several expression sites, located at the ends of chromosomes (reviewed by Englund *et al.*, 1982; Borst & Cross, 1982; Borst *et al.*, 1983; Parsons *et al.*, 1984; Boothroyd, 1986; Donelson & Rice-Ficht, 1985; Michels, 1984; Bernards, 1985; De Lange, 1986; Pays, 1985; Borst, 1986). The integration of VSG genes in expression sites appears to be mediated by two sets of loosely conserved sequence elements up- and downstream from VSG coding regions. Using probes that represent these conserved sequences Van der Ploeg and co-workers have shown that potential VSG genes are closely linked in *T. brucei*, and estimated that this species

[a] Present address: Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142, U.S.A.

[b] Present address: College of Physicians and Surgeons, Columbia University, Department of Human Genetics and Development, 701 West 168th Street, New York, NY 10032, U.S.A.

[c] Present address: Department of Physiological Chemistry, University of Utrecht, Vondellaan 24A, 3521 GG Utrecht, The Netherlands.

[d] Present address: Hubrecht Laboratorium, International Embryological Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.

[e] Present address: Department of Microbiology and Immunology, School of Medicine, University of California at San Francisco, San Francisco, CA 94143, U.S.A.

[f] Present address: Laboratory of Biochemistry, University of Amsterdam, B.C.P. Jansen Institute, Plantage Muidergracht 12, 1018 TV Amsterdam, The Netherlands.

† Abbreviations used: VSG, variant surface glycoprotein; kb, 10^3 base-pairs; MITat, Molteno Institute *Trypanozoon* antigen type; bp, base-pair(s).

0022-2836/86/130001-10 $03.00/0

contains in excess of 1000 VSG genes and pseudogenes per nucleus (Van der Ploeg *et al.*, 1982*b*).

The large VSG gene repertoire evolves rapidly. Strains of *T. brucei*, with minor differences in their housekeeping genes, as judged by enzyme and restriction fragment length polymorphisms (see Gibson *et al.*, 1985*b*), may show large differences in the VSG genes expressed and in the actual VSG gene repertoire (Frasch *et al.*, 1982; Pays *et al.*, 1983; Massamba & Williams, 1984; Paindavoine *et al.*, 1986). In part these differences can be attributed to a rapid turnover of VSG genes in telomeres, caused by partial or complete gene conversion in which chromosome-internal or other telomeric VSG genes act as donors (see the reviews quoted). Substantial differences are also found, however, in the complement of chromosome-internal genes, for which no such gene conversion has been detected. To obtain more information on the evolution of these chromosome-internal VSG genes, we have scrutinized the 30 kb of DNA surrounding the silent basic copy gene for VSG 118 in two related strains of *T. brucei*, strain 427 studied in this laboratory and strain EATRO 1125, studied by Pays, Steinert and their co-workers. Both strains were isolated in Uganda. Their maxi-circle (mitochondrial) DNA sequences differ by only 2% (Borst *et al.*, 1981), whereas differences of up to 8% between other *T. brucei* strains have been observed (Gibson *et al.*, 1985*a*). Their nuclear DNAs differ only marginally in the size distribution of chromosome-sized DNA molecules (Gibson & Borst, 1986) and in the restriction enzyme cleavage sites of the genes for three glycolytic enzymes (Gibson *et al.*, 1985*b*). Nevertheless, we found major differences in the VSG genes surrounding the VSG 118 gene in these strains. Our results indicate that asymmetric sister chromatid exchange and non-duplicative transfer of genes could play a major role in changing the repertoire of chromosome-internal VSG genes.

2. Materials and Methods

(a) *Trypanosomes*

The origins of *T. brucei* strains 427 (Cross, 1975), EATRO 1125 (Frasch *et al.*, 1982), EATRO 795, EATRO 3, and of *T. evansi* SAK (Le Ray *et al.*, 1977), have been described. *T. rhodesiense* TDRN 34 has been described by Gibson *et al.* (1980). Trypanosomes were grown in rats and purified from infected blood by DEAE column chromatography (Lanham & Godfrey, 1970).

(b) *Recombinant DNA*

The construction of cosmid libraries of *T. brucei* strain 427 DNA has been described by Van der Ploeg *et al.* (1982*b*). Genomic libraries of *T. brucei* strain 1125 were made: (1) by inserting a partial *Pst*I digest into plasmid pUR222 and transforming *Escherichia coli* HB101 with the recombinant plasmids; and (2) by inserting a complete *Pst*I digest into plasmid pAT153 and transforming *E. coli* DH1; the resulting transformants were plated onto nitrocellulose filters and screened with labelled probes (Maniatis *et al.*, 1982). Positive bacterial colonies were picked, rescreened and grown in large batches to provide plasmids for mapping and sequence analysis. Plasmids were isolated by the alkaline lysis procedure of Birnboim & Doly (1979).

(c) *Sequence analysis*

DNA sequences were determined as described by Maxam & Gilbert (1980) or by Sanger *et al.* (1977). In the chemical sequencing protocol 5 base-specific modification reactions (A, A+G, C+T, C, A > C) were used to minimize ambiguities. To sequence the 3′ breakpoint in the MITat 1.1000/$(TAA)_{95}$ substitution, a 2·25 kb *Hin*dIII fragment from clone p1125-I was subcloned into pAT153, linearized with either *Bam*HI or *Eco*RI and digested with *Bal*31. Following *Hin*dIII digestion, fragments shortened by 750 bp from the *Bam*HI site or 1900 bp from the *Eco*RI site were subcloned into M13 and sequenced by the method of Sanger *et al.* (1977).

(d) *Gene mapping*

Trypanosome DNA was prepared following a published procedure (Van der Ploeg *et al.*, 1982*b*), cut with restriction enzymes and electrophoresed in agarose gels (1·5 μg/lane). The size-separated DNA was transferred to nitrocellulose (Southern, 1975) and hybridized with nick-translated probes (Rigby *et al.*, 1977) as described (Jeffreys & Flavell, 1977). Prior to exposure to X-ray films the filters were washed with 3 × SSC (SSC is 0·15 M-NaCl, 0·015 M-sodium citrate, pH 7·0) 0·1% (w/v) sodium dodecyl sulphate for several hours at 65°C, and then at higher stringency as indicated in the Figure legends, to remove a specifically bound probe.

(e) *DNA–DNA heteroduplex analysis, electron microscopy and DNA length measurements*

The DNA species used were pc118-29I and p1125-I, in a 1 : 1 concentration of 2·4 μg/μl. After 4 min heating at 90°C in 10 mM-Tris · HCl (pH 8·5), 1 mM-NaEDTA, the solution was immediately cooled on ice. The samples were diluted 10-fold in 100 mM-Tris · HCl, 10 mM-EDTA, 100 mM-NaCl, 60% (v/v) formamide (final pH 8·5) and renatured during 75 min at 30°C. After addition of cytochrome *c* at a concentration of 0·01% and PM2 (double-stranded) and φX (single-stranded) DNA as internal length marker molecules, the DNA was spread on a hypophase of 10% formamide in water according to the procedure described by Davis *et al.* (1971). The DNA–cytochrome *c* film was picked up from the hypophase on 200-mesh copper grids coated with parlodion film and stained with uranyl acetate according to Davis & Davidson (1968).

Electron micrographs were made with a Philips EM300 microscope at a magnification of 7600×. The magnification was calibrated with a grating replica (Fulham 2160 lines/mm). Length measurements were made with a digitizer (Summagraphics, Fairfield, CT) directed by a calculator (HP9825) and stored on tape.

3. Results

(a) *Deletion of the MITat 1.1000 gene in strain 1125*

All VSG genes analysed so far share two sets of loosely conserved sequences up- and downstream from the coding region. At 1 to 2 kb upstream from

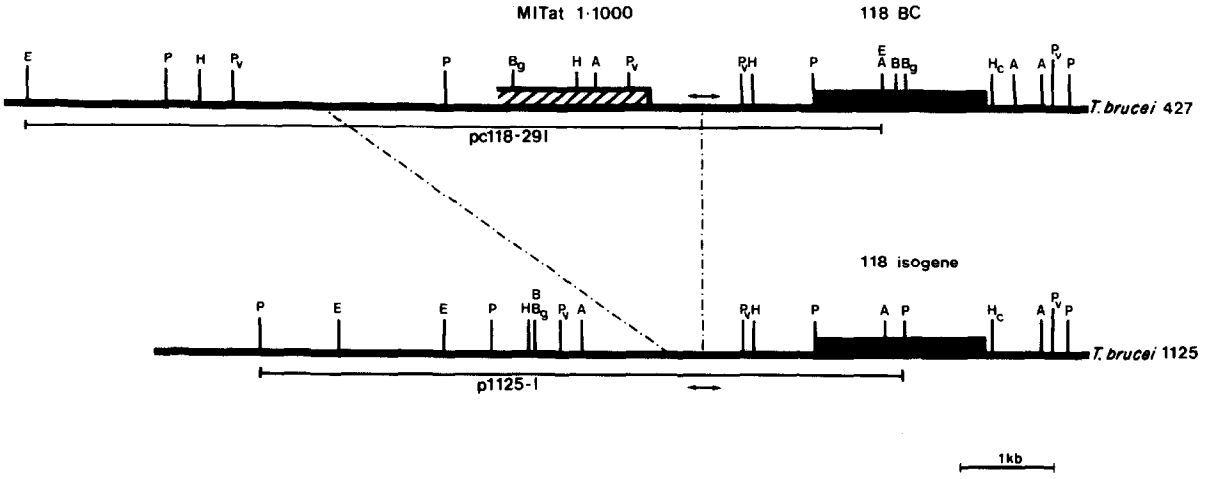

Figure 1. Comparison of the flanking regions of the VSG 118 gene in *T. brucei* strains 427 and 1125. The coding region of the VSG 118 gene is shown as a filled box; the putative VSG gene 1.1000, identified by sequence analysis, is shown as a hatched box. Broken lines show the extent of a deleted fragment containing the 1.1000 gene in strain 1125. Double-headed arrows indicate a 3′ breakpoint region that was sequenced in both strains 427 and 1125. The positions of clones pc118-29I and p1125-I are indicated below the maps. A, *Ava*II; B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; Hc, *Hin*cII; P, *Pst*I; Pv, *Pvu*II.

the start codon of a VSG gene, one characteristically finds a region consisting of a few to hundreds of tandemly repeated DNA segments of approximately 70 bp in length (Van der Ploeg *et al.*, 1982*a*; Liu *et al.*, 1983; Campbell *et al.*, 1984; Bernards *et al.*, 1985; Aline *et al.*, 1985). Downstream, VSG genes share a second region of sequence homology in the segment encoding the 3′ untranslated part of the VSG message (Majumder *et al.*, 1981; Matthyssens *et al.*, 1981; Rice-Ficht *et al.*, 1981; Borst & Cross, 1982; Michels *et al.*, 1983; Liu *et al.*, 1983). Both types of conserved sequence elements may be used to track down VSG genes in the trypanosome genome (cf. Van der Ploeg *et al.*, 1982*b*). By this means, we previously identified a putative VSG gene (named the MITat 1.1000 gene) immediately upstream from the VSG 118 basic copy gene in *T. brucei* 427 (Liu *et al.*, 1983). A map of this region is presented in Figure 1. The 5′ border of the 1.1000 gene has not been investigated; we also do not know whether this gene is functional.

A comparison of the region upstream from the VSG 118 gene in several trypanosome strains previously revealed a number of alterations in the restriction map (Frasch *et al.*, 1982). To analyse these changes in more detail, we have cloned part of the VSG 118 upstream area in *T. brucei* strain 1125, and compared it to a clone of the same upstream region in strain 427 (Fig. 1). A 3·6 kb DNA fragment containing the MITat 1.1000 gene in strain 427 appears to be deleted in strain 1125, and this was confirmed by heteroduplex analysis using clone pc118-29I from strain 427 and clone p1125-I from strain 1125 (see Fig. 1 for position of clones). The electron micrograph in Figure 2 does not only show that part of the VSG 118 upstream region is deleted in strain 1125, but that it has been replaced by a new 0·4 kb insertion. To analyse the nature of this insertion, we sequenced across the 3′ edge of the deletion/substitution. The sequence obtained is compared with that previously determined for strain 427 in Figure 3. Going from 3′ to 5′, the two sequences match fairly well up to the 70 bp repeats (indicated by the 3 boxed motifs) upstream from the 118 gene. In the first repeat substantial divergence starts and becomes complete in the second repeat, due to the presence of a $(TAA)_n$ stretch in strain 1125. This stretch contains about 95 TAA units and it must represent part, if not all, of the 0·4 kb unpaired segment in the heteroduplex, since it is not present at this location or elsewhere in the 70 bp repeat array upstream from the 118 gene in strain 427 (Liu *et al.*, 1983). Beyond the $(TAA)_{95}$ stretch, the two sequences regain homology, as shown by Figure 2. There is an additional small mismatch 1·7 kb upstream from the major mismatch; this has not been investigated further.

We have shown that some VSG genes have been more stable during evolution than others, being present in most of the *T. brucei* strains analysed (Frasch *et al.*, 1982). To test the status of the MITat 1.1000 gene, we probed restriction digests of nuclear DNA of a number of trypanosome strains with a DNA fragment specific for this gene. As shown in Figure 4(a), the 1.1000 gene is present in all five strains tested except 1125, and we therefore suggest that its absence in this strain is more likely to be due to a deletion than its presence in other strains is due to an insertion.

(b) *Quadruplication of a putative VSG gene in strain 427*

The MITat 1.1000 and the VSG 118 basic copy are not the only VSG genes in this region of the genome of strain 427. Downstream from the VSG 118 gene, Van der Ploeg *et al.* (1982*b*) located

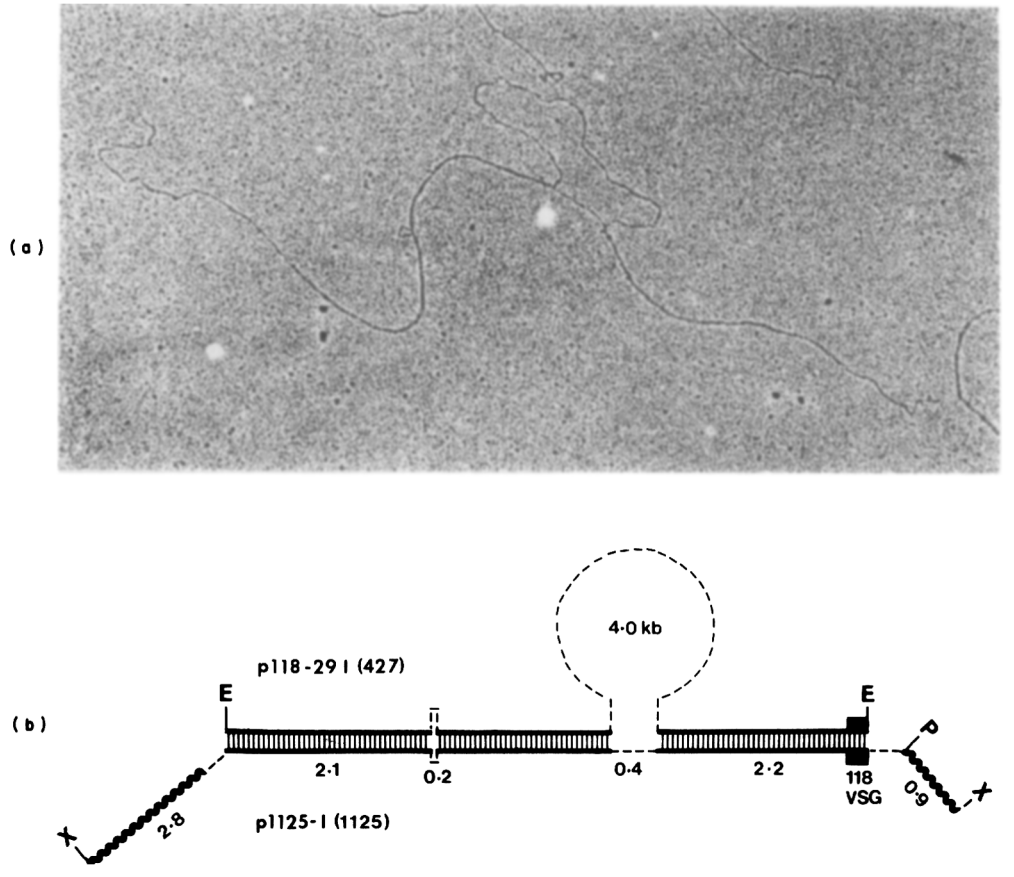

Figure 2. (a) Electron micrograph of an heteroduplex formed from clones pc118-29I of *T. brucei* 427 and p1125-I of *T. brucei* 1125 (see the maps in Fig. 1). Renaturation was in 60% formamide. Magnification, 59,000 ×. See Materials and Methods for experimental details. (b) Schematic interpretation of the heteroduplex molecule shown in (a) based on measurements of 35 molecules. The hatched area is duplex DNA; broken lines represent single-stranded *T. brucei* DNA and wavy lines single-stranded plasmid DNA. The position of the VSG 118 gene is indicated by the filled block. The numbers represent the average size of the segment in kb. The standard deviation in these measurements was 0·2 kb. See Materials and Methods for details. E, *Eco*RI; P, *Pst*I; X, *Xmn*I.

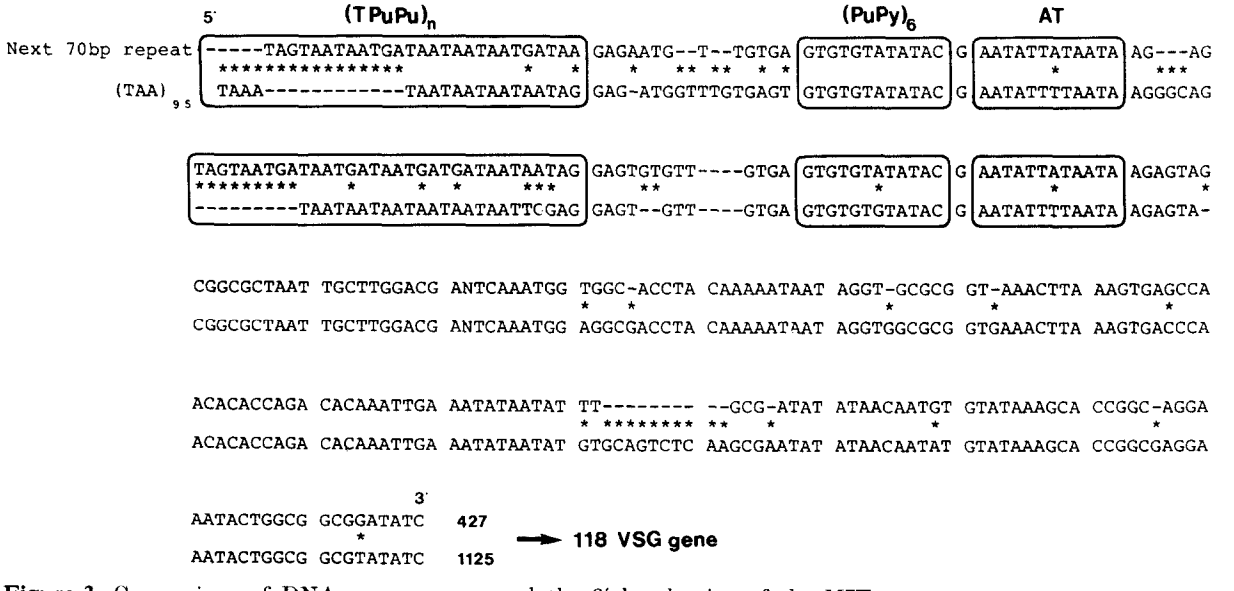

Figure 3. Comparison of DNA sequences around the 3′ breakpoint of the MITat 1.1000/$(TAA)_{95}$ substitution in *T. brucei* strains 427 and 1125. The 2 sequences have been aligned for maximum homology. Mismatches are indicated by asterisks. The 3 elements comprising a 70 bp repeat are boxed. The 2 sequences are largely homologous upstream from the VSG 118 basic copy gene, but diverge completely at the 5′ end of the second 70 bp repeat. The regions sequenced are indicated by double-headed arrows in Fig. 1. Sequence hyphens have been omitted for clarity.

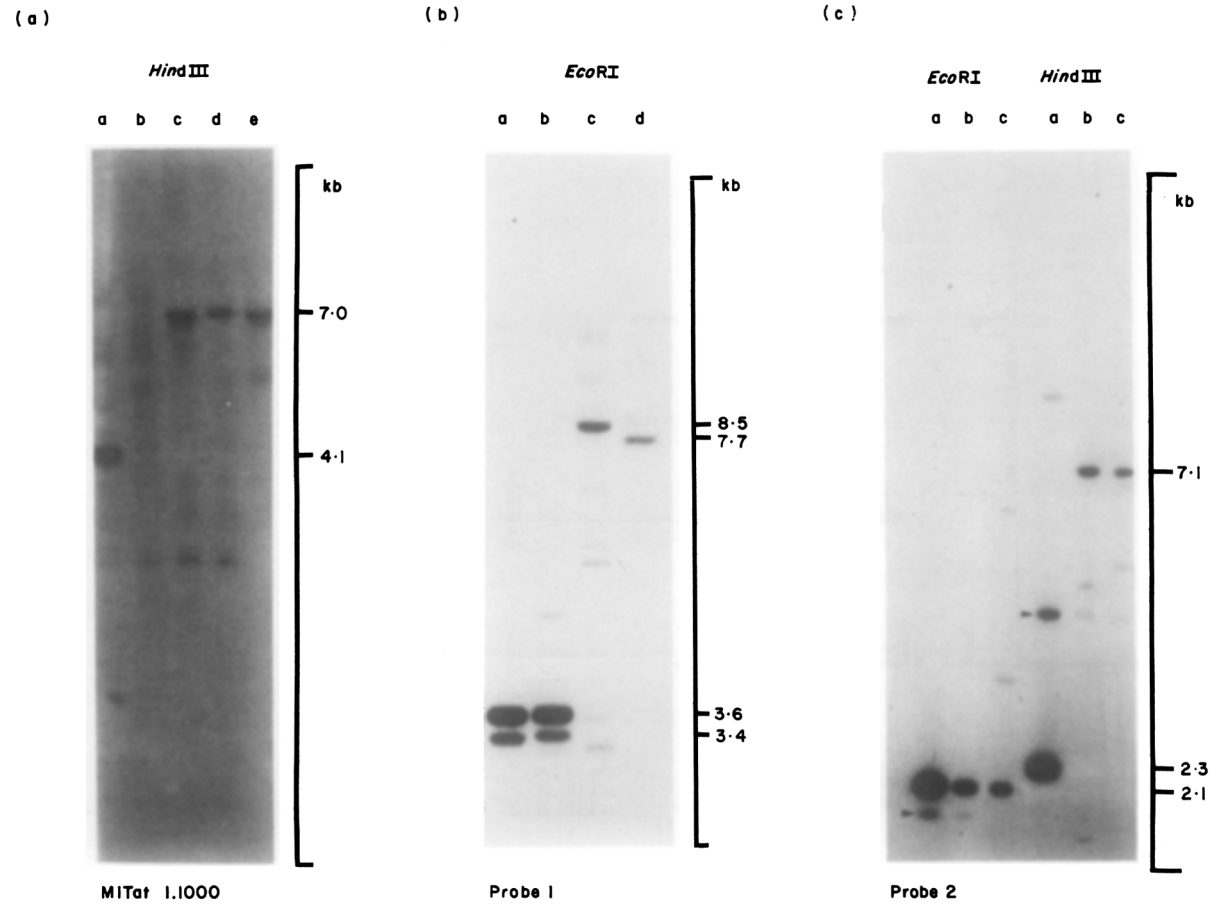

Figure 4. (a) Presence of the MITat 1.1000 gene in various trypanosome strains. Nuclear DNA was digested with *Hin*dIII and size-fractionated by electrophoresis through 0·5% (w/v) agarose. After Southern blotting, the DNA was hybridized with a 2 kb *Pst*I–*Pvu*II fragment from the MITat 1.1000 gene; post-hybridizational washes were to 3 × SSC at 65°C. Lane a, *T. brucei* 427; lane b, *T. brucei* 1125; lane c, *T. brucei* EATRO 3; lane d, *T. brucei* EATRO 795; lane e, *T. evansi* SAK. *T. brucei* 1125 is the only strain lacking the 1.1000 gene. A background of cross-hybridization bands is visible in all lanes. (b) Southern blot analysis of DNA from several trypanosome strains using a probe downstream from the VSG 118 gene (probe 1 see Fig. 5(c)). Electrophoresis was through 0·7% agarose and washing of filters after hybridization to 0·1 × SSC at 65°C. Lane a, *T. brucei* 427; lane b, TDRN 34; lane c, 1125; lane d, *T. evansi* SAK. (c) As (b) using a probe for the 1.1001 gene (probe 2 see Fig. 5(c)). Lane a, *T. brucei* 427; lane b, 1125; lane c, *T. evansi* SAK. Arrowheads indicate fragments that do not come from the VSG 118 gene region.

several restriction fragments that hybridize with probes either for the 70 bp repeats or the 3′-conserved end of VSG genes, and concluded that potential VSG genes are tightly packed in this area. We have investigated this further by mapping the area in more detail.

In these experiments we used the previously isolated cosmid clone TgBc 118-33 (Van der Ploeg *et al.*, 1982*b*), which contains the VSG 118 coding region and approximately 10 kb of upstream and 30 kb of downstream-flanking DNA (Fig. 5(a)). As shown, the 3·4 kb *Eco*RI fragment containing the 3′ end of the VSG 118 gene is flanked by several tandemly repeated 2·1 kb and 3·6 kb *Eco*RI paired fragments. The 2·1 kb fragments hybridize with a 70 bp repeat probe, the 3·6 kb fragments with a 3′ VSG cDNA probe. This indicates that the VSG 118 gene is flanked downstream by several tandemly repeated copies of a putative VSG gene. The preliminary analysis by Van der Ploeg *et al.* (1982*b*) had indicated the presence of three VSG genes in this area, but the partial restriction digests shown in Figure 5(b), clearly demonstrate that there are actually four 2·1 kb and three 3·6 kb *Eco*RI fragments contained in cosmid TgBc 118-33. These four putative VSG genes were provisionally named the MITat 1.1001 to 1.1004 genes (Fig. 5(a)). Genes 1.1001 to 1.1004 are very similar, as each enzyme that cuts in the 2·1 or 3·6 kb segments was found to do so at an identical position in every copy. We have also isolated a number of *Eco*RI subclones of cosmid TgBc 118-33, containing either 2·1 or 3·6 kb segments. The maps of these subclones are all identical (Fig. 5(c)).

Since the restriction sites mapped cover 110 bp, the sequence differences between individual 2·1 (fragment b in Fig. 5(c)) and 3·6 kb (fragment c in Fig. 5(c)) fragments must be less than 1%, and we have confirmed this by partial sequence analysis (see below). Upon inspection of the subclone maps in Figure 5(c), it becomes clear that the region immediately downstream from the 118 basic copy is

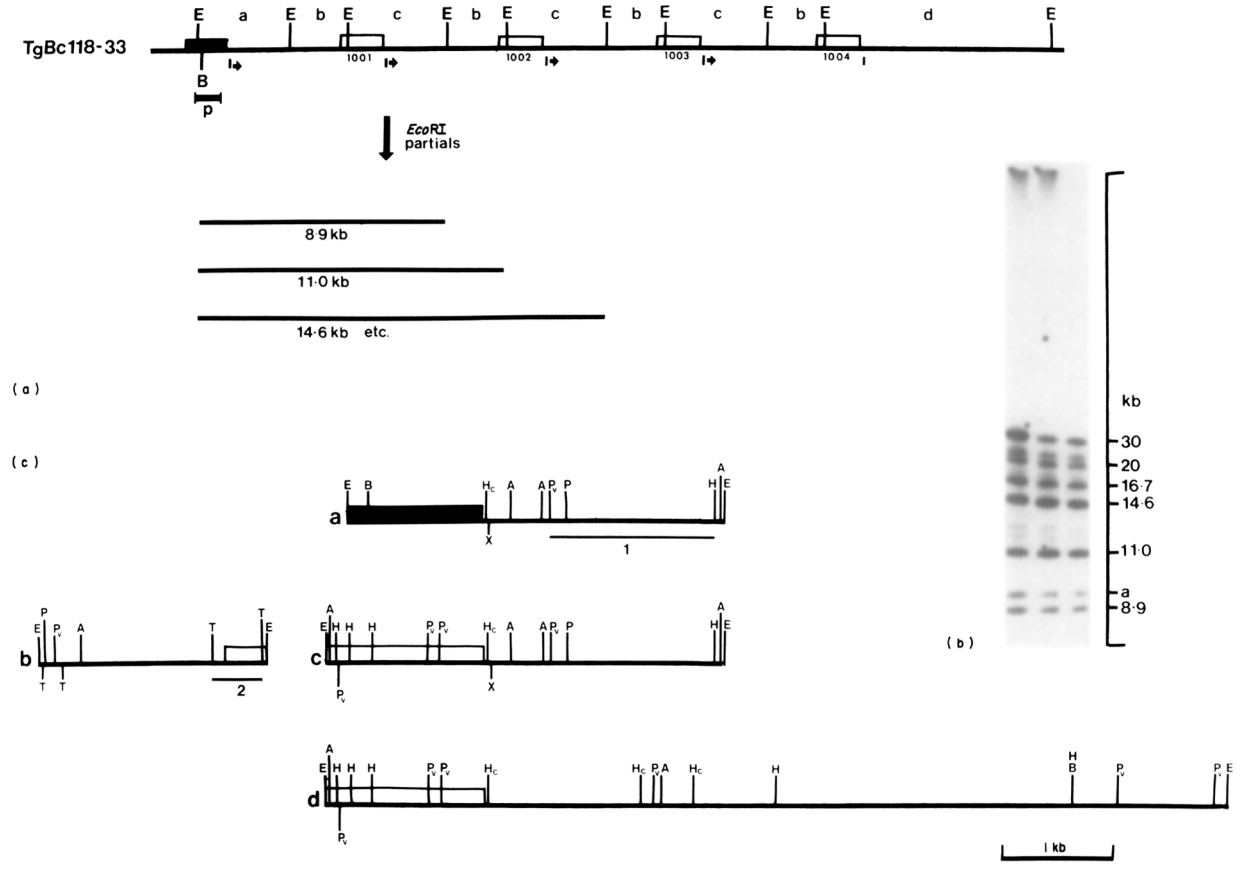

Figure 5. (a) Physical map of the VSG 118 BC gene (filled box) and downstream region in cosmid clone TgBc 118-33 (Van der Ploeg *et al.*, 1982*b*). Putative VSG genes MITat 1.1001 to 1.1004 are indicated by open boxes. The 5′ border of each 5·7 kb repeat unit is indicated by an arrow. E, *Eco*RI. 2·1 kb b fragments hybridize with a 70 bp repeat probe (1 kb *Pvu*II fragment upstream from 118 BC in strain 427: see Fig. 1) and the 3·4 kb a, 3·6 kb c and 8 kb d fragments hybridize with a probe representing the 3′ end of the VSG 117 gene (280 bp *Pst*I from TcV 117.8; Bernards *et al.* (1981)). (b) Southern blot analysis of cosmid TgBc 118-33, showing quadruplication of the gene downstream from the VSG 118 gene. A complete *Bam*HI digest of cosmid TgBc 118-33 was incubated for different times with *Eco*RI. The samples were fractionated through 0·5% agarose and analysed by Southern blotting using probe p (see (a)); post-hybridization washes were to 0·1 × SSC at 65°C. The products seen are located as shown below the map in (a); the smaller fragments have been run off the gel to achieve separation of the larger ones. The band indicated by a is irrelevant and contains the 5′ part of the VSG 118 gene, since the probe overlaps this fragment by 200 bp. (c) Physical maps of *Eco*RI fragments a, b, c, d shown in (a). See Fig. 2 for abbreviations of restriction enzymes, and additionally T, *Taq*I; X, *Xba*I.

very similar to part of the 3·6 kb *Eco*RI fragment map. This was confirmed by hybridizing Southern blots of *Eco*RI digests of strain 427 DNA with a probe representing this area. In this experiment the probe hybridized to both the 3·4 and the 3·6 kb fragments, even at conditions of high stringency (Fig. 4(b)). The tandem repetition therefore starts in or immediately beyond the 3′ end of the 118 gene, and each repeat is 5·7 kb long, ending near a position just beyond the next VSG gene, as indicated by the arrows in Figure 5(a). To verify this, we sequenced the 3′ end of the 118 gene and of several of the putative VSG genes downstream; the sequences of VSG 118 and the other genes except for MITat 1.1004 were obtained by the Maxam & Gilbert (1980) technique from the labelled *Xba*I site; for MITat 1.1004, the *Hin*cII site immediately downstream from the gene was used (Fig. 5(c)). The sequences, lined up in Figure 6, indeed show many sequence elements characteristic of VSG gene 3′ ends, such as the 14-mer T-G-A-T-A-T-A-T-T-T-T-A-A-C, which is present with minor alterations in the 3′ untranslated region of all VSG genes analysed so far (Majumder *et al.*, 1981; Matthyssens *et al.*, 1981; Rice-Ficht *et al.*, 1981; Borst & Cross, 1982; Michels *et al.*, 1983; Liu *et al.*, 1983), and several other conserved sequence elements. The sequence comparison in Figure 6, also shows that, starting from a position just upstream from the conserved 14-mer, the regions downstream from the 118 gene and the MITat 1.1001 to 1.1004 genes, are almost identical. The greatest divergence is found in the 3′ end of the MITat 1.1004 gene. This situation may have arisen by multiple unequal cross-overs, as illustrated in Figure 7. This scheme assumes that the 118 gene was originally linked to a single copy of the MITat 1.1001 gene, and that the tandem quadruplication arose later. We therefore tested whether the original situation or some of the postulated intermediate steps can be found in other

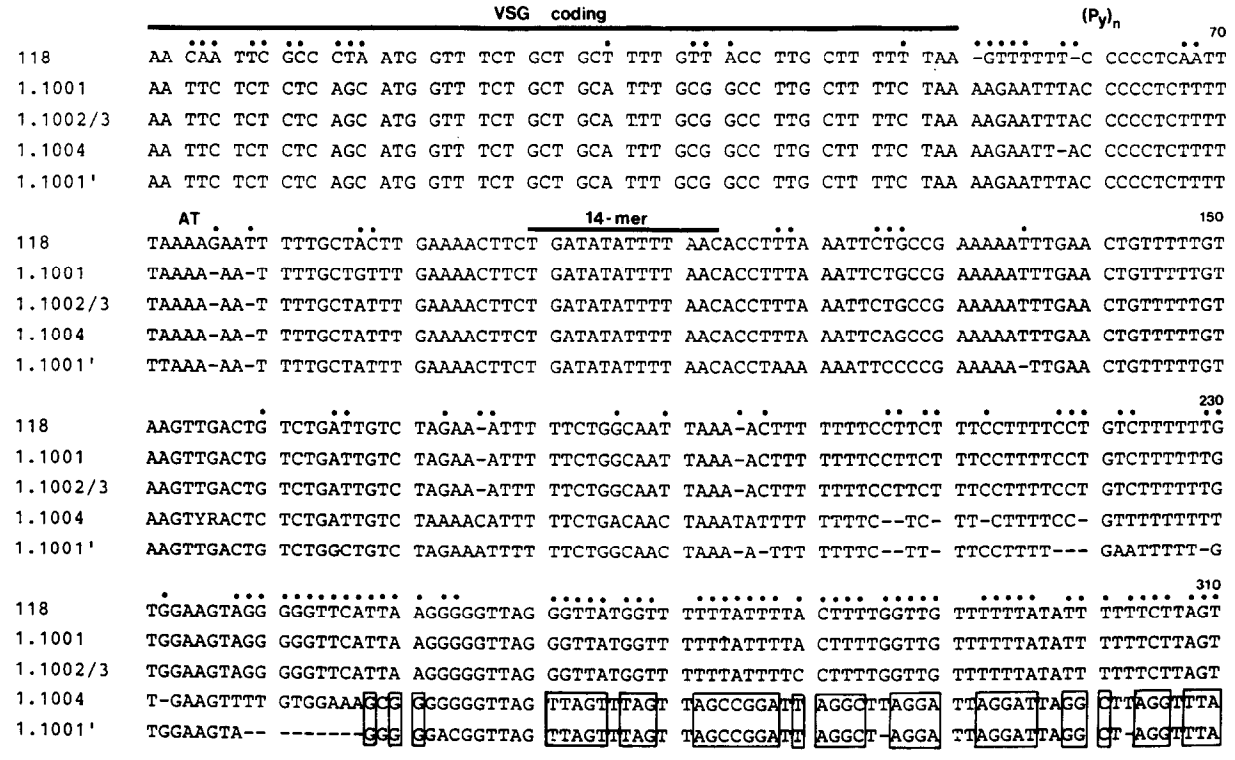

Figure 6. Sequence comparison of the 3′ ends of the VSG genes 118, 1.1001, 1.1002/3 and 1.1004 of *T. brucei* 427 and 1.1001 (designated 1.1001′) of *T. brucei* 1125. The 1.1001 sequence was determined using a subcloned fragment of cosmid TgBc 118-18. This cosmid contains the 1.1001 gene, but not the other 3 copies of the quadruplicated gene array (Van der Ploeg *et al.*, 1982*b*). The sequence designated 1.1002/3 was obtained from several individual 3·6 kb *Eco*RI subclones from cosmid TgBc 118-33 (fragment c, Fig. 5). The 1.1004 sequence was determined from fragment d (Fig. 5). Mismatches are indicated by dots above the columns. Three sequence motifs of VSG gene 3′ ends are indicated above the sequence (14-mer, A+T-rich region, pyrimidine stretch). Specific homologies between 1.1004 and 1.1001′ genes are boxed. Sequence hyphens have been omitted for clarity.

trypanosome stocks. As shown in Figure 4(b) and (c), the 118 gene is indeed linked to just a single copy of the MITat 1.1001 gene in strains *T. brucei* 1125 and *T. evansi* SAK. Restriction maps of the relevant areas are presented in Figure 8. In

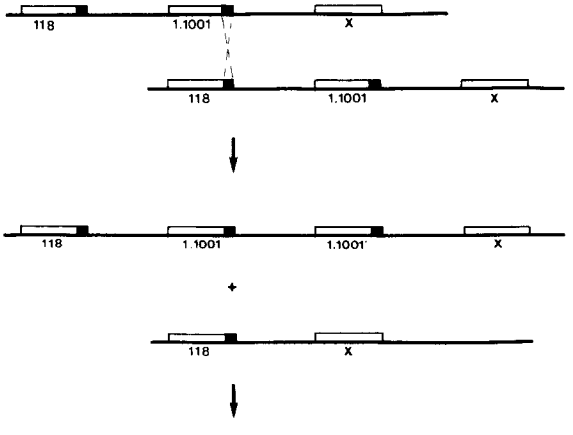

Figure 7. Schematic drawing showing how unequal cross-overs between the 3′ ends of adjacent VSG genes may have given rise to the tandem repetition of the MITat 1.1001 gene. VSG genes are indicated by boxes, the conserved 3′ ends are filled. The initial cross-over between the 118 gene on one chromatid and the 1.1001 gene on another must have occurred just upstream from the conserved 14-mer sequence in the 3′ ends (see Fig. 6).

T. rhodesiense TDRN 34 the map is identical to that of *T. brucei* 427 (Fig. 4(b)). However, no strains representing intermediate stages in the quadruplication were found.

A further prediction of the scheme in Figure 7 is that the areas downstream from the MITat 1.1004 gene in strain 427 and beyond the end of the 1.1001 gene in strain 1125 are related. We tested this prediction by cloning a 6 kb *Pst*I fragment containing the 3′ end of the 1.1001 gene in strain 1125 and sequencing around the *Hin*cII site (see Fig. 8). The sequence included in Figure 6, shows that the 3′ ends of the 1.1004 gene in strain 427 and the 1.1001 gene in strain 1125 are indeed closely related.

4. Discussion

Our analysis has demonstrated a remarkable degree of variation in the DNA area surrounding the VSG 118 gene in different trypanosome strains. Downstream we find a recent quadruplication of the VSG 1.1001 gene in strain 427; this quadruplication is absent in most other strains analysed. Upstream we find the VSG 1.1000 gene in strain 427, which is missing in strain 1125, but not in three other strains analysed. We do not think that the area around the VSG 118 gene is exceptional. The gene was chosen at random, because we happened to isolate a cDNA

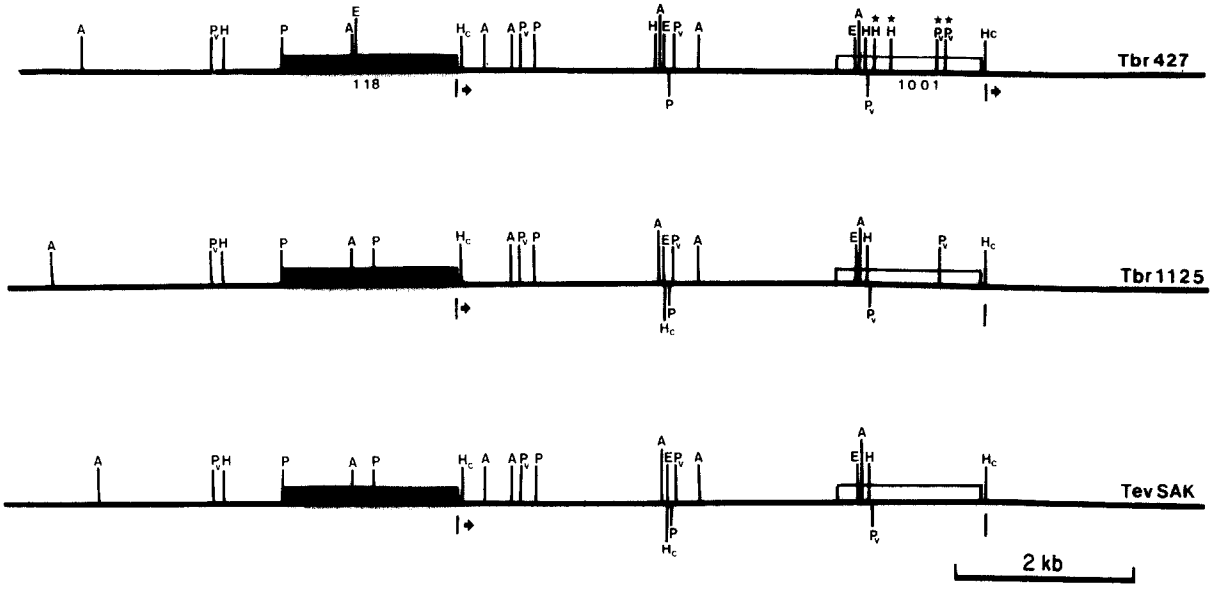

Figure 8. Physical maps of the genomic area downstream from the VSG 118 gene in *T. brucei* 427 (Tbr 427), *T. brucei* 1125 (Tbr 1125) and *T. evansi* SAK (Tev SAK). Arrows indicate the 5′ border of a 5·7 kb segment. The presence of restriction sites marked with a star has not been verified in Tbr 1125 and Tev SAK. See Fig. 2 for abbreviations of restriction enzymes.

probe for it years ago (Hoeijmakers *et al.*, 1980). Gross variability has also been observed for other VSG genes and their surroundings, as shown by the size variation of restriction fragments hybridizing with a gene-specific probe in DNA from different *T. brucei* strains and the absence of certain genes in some strains (Frasch *et al.*, 1982; Pays *et al.*, 1983; Massamba & Williams, 1984; Paindavoine *et al.*, 1986). We therefore expect that the duplications and deletions observed here will be representative of chromosome-internal VSG genes in general.

The quadruplication of VSG gene 1.1001 has all the hallmarks of a repeated asymmetric sister chromatid exchange. Although it is now clear that bloodstream trypanosomes are diploid for housekeeping genes (Gibson *et al.*, 1985*b*), no evidence for exchange of DNA between homologous chromosomes has been forthcoming. Our data are therefore the first indication that sister chromatid exchange may occur in trypanosomes. Alignment of chromosomes containing VSG genes must pose special problems. The sequences shared by all VSG genes, present around the coding regions, can be expected to be a source of frequent chromosome misalignments. The rapid evolution of VSG genes will add to the logistic problems, because homologous positions in sister chromatids will often be occupied by different VSG genes.

The interpretation of the deletion of the 1.1000 gene in strain 1125 is less straightforward. Such deletions are the unavoidable counterpart of asymmetric sister chromatid exchanges that lead to gene duplication, but in this case a $(TAA)_{95}$ stretch replaces the gene lost. Such a long $(TAA)_n$ stretch was first observed in the long array of 70 bp repeats in front of the expressed telomeric copy of the VSG 117 gene (Campbell *et al.*, 1984). After our experiments were completed, long $(TAA)_n$ stretches were also found by Aline *et al.* (1985) in front of three chromosome-internal VSG genes. Aline *et al.* also report that there are some 30 long $(TAA)_n$ stretches in the *T. brucei* genome, but whether these are mostly chromosome-internal or telomeric has not been established. It is possible that the loss of gene 1.1000 and the gain of the $(TAA)_{95}$ stretch were unconnected events, since 70 bp repeats (which contain $(TAA)_n$ stretches) are sites of frequent recombinations/gene conversions. A simpler interpretation is, however, that the loss of the 1.1000 gene and the gain of the $(TAA)_{95}$ stretch occurred as one event. Such exchanges could occur during duplicative transposition of the 1.1000 or a neighbouring gene to a telomeric expression site that contains a long $(TAA)_n$ stretch.

Three factors are now known to contribute to the rapid change of VSG gene repertoires. First, the type of gene duplications documented here should provide raw material for new VSG genes. Secondly, duplicative transposition of genes to telomeric locations leads to a rapid turnover of genes in telomeres; and thirdly, partial conversion of telomeric genes rapidly adds new genes to the repertoire (see Pays, 1985; Pays *et al.*, 1985). Whether such telomeric genes can become more than a fleeting addition to the repertoire remains to be seen, because no mechanism for the return of telomeric genes to a chromosome-internal position has been documented. In addition, genetic exchange, if occurring on a sizeable scale, could pool the diversification generated by individual trypanosome strains and add to the evolution of the repertoires. Although these factors can largely account for the large differences in VSG gene repertoire observed, three points require further

study. A major question is whether any active mutagenic mechanism is used by trypanosomes to speed up the evolution of (part of) their VSG gene repertoire. Such a mechanism was postulated by Frasch *et al.* (1982) to explain their finding that the VSG 221 gene was present only in strain 427 and not in 12 other trypanosome strains analysed. We now know that the 221 gene is telomeric (Bernards, 1982; Bernards *et al.*, 1985), which explains why it is unstable. It remains puzzling, however, that no 221-related gene has turned up in any of the 13 strains analysed thus far under hybridization conditions of moderate stringency (3 × SSC at 65°C). We have seen this also for two other telomeric genes, 1.1005 and 1.1006 (De Lange *et al.*, 1983). The possibility that a subfraction of VSG genes evolves more rapidly, for instance because of an increase in DNA replication or recombination errors near telomeres, therefore remains open. The possible contribution of patch-wise segmental gene conversion to the evolution of chromosome-internal genes, also requires further analysis. There is even the remote possibility that VSG genes from other trypanosome species might be able to enter the repertoire of *T. brucei via* abortive matings or virus infection.

A second finding that remains unexplained is the high stability of some VSG genes in the repertoire. The most intensively studied example is the VSG gene known as the 117 gene in strain 427 and the AnTat 1.8 gene in strain 1125. The variant antigen type corresponding to this gene was detected by immunological methods in nearly all *T. brucei* strains analysed (see Pays *et al.*, 1981), and the presence of the gene has been verified in many strains both by Frasch *et al.* (1982) and Pays and co-workers (Pays *et al.*, 1981, 1983; Paindavoine *et al.*, 1986). In contrast, most other genes studied have a more limited distribution. It is conceivable that some chromosomal locations may be more shielded against change than others, for instance because a gene is surrounded by housekeeping genes and therefore less liable to be the victim of asymmetric cross-overs. The 117 gene is indeed located in the large chromosome fraction (Van der Ploeg *et al.*, 1984), which also hybridizes to probes for several housekeeping genes (Gibson *et al.*, 1985*b*); but so is the VSG 118 gene (Van der Ploeg *et al.*, 1984), which is absent in about half of the trypanosome strains analysed (Frasch *et al.*, 1982). Moreover, the 117 gene is surrounded by other VSG genes, as judged by hybridization experiments with the conserved up- and downstream sequence elements as probes (Van der Ploeg *et al.*, 1982*b*). It is possible therefore that the 117 (AnTat 1.8) gene is not shielded against change, but selected for because of the superior quality of the 117 coat. This remains to be tested.

Finally, it is still not clear how the dispersed families of related chromosome-internal genes arise. The quadruplication described here generates gene families in which the family members are linked, but close linkage has not been found in the VSG gene families studied thus far (Van der Ploeg *et al.*, 1982*b*). Gene conversion of chromosome-internal genes by other VSG genes or displaced expression-linked copies could be responsible for this, as we have suggested (Van der Ploeg *et al.*, 1982*b*), but evidence for such a process is lacking. A systematic analysis of VSG gene arrays in related trypanosome strains, as initiated here, may provide information on this point.

We thank Marijke Gruter for her contribution to some of these experiments and Marleen Van den Berg for assistance with electron microscope work. We are also indebted to Helga A. Woudt and Bartie K. Van Houten for preparing the manuscript. This work was supported by a grant from The Wellcome Trust (to W.C.G.) and by a grant from the Foundation for Fundamental Biological Research (BION), which is subsidized by the Netherlands Organization for the Advancement of Pure Research (ZWO).

References

Aline, R., Macdonald, G., Brown, E., Allison, J., Myler, P., Rothwell, V. & Stuart, K. (1985). *Nucl. Acids Res.* **13**, 3161–3177.

Bernards, A. (1982). *Trends Biochem. Sci.* **7**, 253–255.

Bernards, A. (1985). *Biochim. Biophys. Acta,* **824**, 1–15.

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

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

Boothroyd, J. C. (1986). *Annu. Rev. Microbiol.* **39**, 475–502.

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

Borst, P. & Cross, G. A. M. (1982). *Cell,* **29**, 291–303.

Borst, P., Fase-Fowler, F. & Gibson, W. C. (1981). *Mol. Biochem. Parasitol.* **3**, 117–131.

Borst, P., Bernards, A., Van der Ploeg, L. H. T., Michels, P. A. M., Liu, A. Y. C., De Lange, T. & Kooter, J. M. (1983). *Eur. J. Biochem.* **137**, 383–389.

Campbell, D. A., Ven 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.

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

Cross, G. A. M. (1978). *Proc. Roy. Soc. ser. B,* **202**, 55–72.

Davis, R. W. & Davidson, N. (1968). *Proc. Nat. Acad. Sci., U.S.A.* **60**, 243–250.

Davis, R. W., Simon, M. & Davidson, N. (1971). *Methods Enzymol.* **21**, 413–428.

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

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

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

Englund, P. T., Hajduk, S. L. & Marini, J. C. (1982). *Annu. Rev. Biochem.* **51**, 695–726.

Frasch, A. C. C., Borst, P. & Van den Burg, J. (1982). *Gene,* **17**, 197–211.

Gibson, W. C. & Borst, P. (1986). *Mol. Biochem. Parasitol.* **18**, 127–140.

Gibson, W. C., Marshall, T. F. de C. & Godfrey, D. G. (1980). *Advan. Parasitol.* **18**, 175–246.

Gibson, W. C., Borst, P. & Fase-Fowler, F. (1985*a*). *Mol. Biochem. Parasitol.* **15**, 21–36.

Gibson, W. C., Osinga, K. A., Michels, P. A. M. & Borst, P. (1985*b*). *Mol. Biochem. Parasitol.* **16**, 231–242.

Hoeijmakers, J. H. J., Borst, P., Van den Burg, J., Weissmann, C. & Cross, G. A. M. (1980). *Gene*, **8**, 391–417.

Jeffreys, A. J. & Flavell, R. A. (1977). *Cell*, **12**, 429–439.

Lanham, S. M. & Godfrey, D. G. (1970). *Exp. Parasitol.* **28**, 521–534.

Le Ray, D., Barry, J. D., Easton, C. & Vickerman, K. (1977). *Ann. Soc. Belge Med. Trop.* **57**, 369–381.

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

Majumder, H. K., Boothroyd, J. C. & Weber, H. (1981). *Nucl. Acids Res.* **9**, 4745–4753.

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

Massamba, N. N. & Williams, R. O. (1984). *Parasitology*, **88**, 55–65.

Matthyssens, G., Michiels, F., Hamers, R., Pays, E. & Steinert, M. (1981). *Nature (London)*, **293**, 230–233.

Maxam, A. M. & Gilbert, W. (1980). *Methods Enzymol.* **65**, 499–560.

Michels, P. A. M. (1984). In *Oxford Surveys on Eucaryotic Genes*, vol. 1, pp. 145–168.

Michels, P. A. M., Liu, A. Y. C., Bernards, A., Sloof, P., Van der Bijl, M. M. W., 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.

Paindavoine, P., Pays, E., Laurent, M., Geltmeyer, Y., Le Ray, D., Mehlitz, D. & Steinert, M. (1986). *Parasitology*, **92**, 31–50.

Parsons, M., Nelson, R. G. & Agabian, N. (1984). *Immunol. Today*, **5**, 43–50.

Pays, E. (1985). *Progr. Nucl. Acid Res. Mol. Biol.* **32**, 1–26.

Pays, E., Lheureux, M., Vervoort, T. & Steinert, M. (1981). *Mol. Biochem. Parasitol.* **4**, 349–357.

Pays, E., Dekerck, P., Van Assel, S., Babiker, E. E., Le Ray, D., Van Meirvenne, N. & Steinert, M. (1983). *Mol. Biochem. Parasitol.* **7**, 63–74.

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. (1985). *Cell*, **42**, 821–829.

Rice-Ficht, A. C., Chen, K. K. & Donelson, J. E. (1981). *Nature (London)*, **294**, 53–59.

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

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463–5467.

Southern, E. M. (1975). *J. Mol. Biol.* **98**, 503–517.

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

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

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

Vickerman, K. (1969). *J. Cell Sci.*, **5**, 163–193.

Vickerman, K. (1978). *Nature (London)*, **273**, 613.

Edited by P. Chambon