
Telomere conversion in trypanosomes

Titia De Lange*, Jan M.Kooter*, Paul A.M.Michels⁺ and Piet Borst*

Section for Medical Enzymology, Laboratory of Biochemistry, University of Amsterdam, P.O.Box 60.000, 1005 GA Amsterdam, and *Department of Molecular Biology, Antoni van Leeuwenhoekhuis, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands

Received 17 October 1983; Accepted 9 November 1983

ABSTRACT

Activation of the gene coding for variant surface glycoprotein (VSG) 118 in *Trypanosoma brucei* proceeds via a duplicative transposition to a telomeric expression site. The resulting active expression-linked extra copy (ELC) is usually flanked by DNA that lacks sites for most restriction enzymes and that is thought to interfere with the cloning of the ELC as recombinant DNA in *Escherichia coli*. We have circumvented this problem by cloning an aberrant 118 ELC gene, flanked at the 3'-side by at least 1 kb DNA, that contains restriction enzyme sites. Our analysis shows that this DNA and the 3'-end of the 118 ELC gene are derived from another VSG gene (1.1006) that is permanently located at a telomeric position. We propose that the 3'-end of the 1.1006 gene and (all of) its 3' flanking sequence moved to the expression site by a telomere conversion. Such a telomere conversion can also account for the appearance of an extra copy of the 1.1006 gene detected in a sub-population of our trypanosome strain.

INTRODUCTION

The mammalian immune response to African trypanosomes, like *Trypanosoma brucei*, is directed against the main component of the parasite surface coat, the variant surface glycoprotein or VSG [1,2]. Trypanosomes can change the antigenic nature of their coat and escape immune destruction with low frequency (10^{-4} to 10^{-6} per division) by switching to the expression of another member of the large VSG gene repertoire (reviewed in ref. 3).

Two modes of VSG gene activation have been characterized: duplicative and non-duplicative. Some VSG genes are activated by a duplicative transposition of a silent basic copy (BC) of the gene to an expression site elsewhere in the genome, yielding an active expression-linked extra copy (ELC) [4-8]. The expression site is thought to activate VSG genes by donation of a promoter for VSG gene transcription and a mini-exon coding

for the common 5' 35 nucleotides of VSG messenger RNAs (mRNAs) [9-11]. ELC genes are located near a chromosome end [12] and flanked by long stretches of DNA that are not cut by the majority of the restriction enzymes tested [13]. These 'barren' regions vary in length [14,15]. Growth of trypanosomal telomeres, balanced by occasional deletions [16,17], is the main source of length variation of the 3' 'barren' region. The length variation of the 5' 'barren' region is related to the switching process [15], but the mechanism of this variation has not been clarified.

VSG gene switching by duplicative transposition probably involves a process akin to gene conversion in which the incoming gene replaces the resident ELC. The conversion process is thought to be guided by blocks of homology at the borders of the transposed segment [13]. The 3' recombination point has been determined in a number of cases by direct sequence analysis of VSG mRNAs and can occur anywhere in the last 150 base pairs of the gene, often leading to an exchange of gene ends [8,15].

The second mode of VSG gene activation does not involve gene duplication or other major rearrangements detectable in Southern blots and has only been observed for genes that are permanently located at chromosome ends [18-21]. Like the chromosome-internal BC genes, the telomeric genes are split into a mini-exon [10] and a main exon. The size of the intervening sequence is not known, but is at least 8 kb for the telomeric gene coding for VSG 221 [21].

In a previous paper we have described the structure of four ELCs coding for VSG 118 in four different trypanosome clones [15]. One of these (MITat 1.186, also referred to as 118c) was unusual in that its downstream flanking region contains several restriction enzyme sites (see Fig. 1). Hitherto all efforts to clone ELC genes plus their flanking sequences in E. coli were unsuccessful and we have attributed this to the 'barren' regions [22]. The aberrant configuration of the 118c ELC gene allowed us to circumvent this problem and to clone the ELC gene plus its 3' flanking sequences. Using this cloned DNA we have performed a detailed analysis of the recombination events in the expression site. The results suggest a novel duplicative mechanism for activation of telomeric VSG genes: telomere

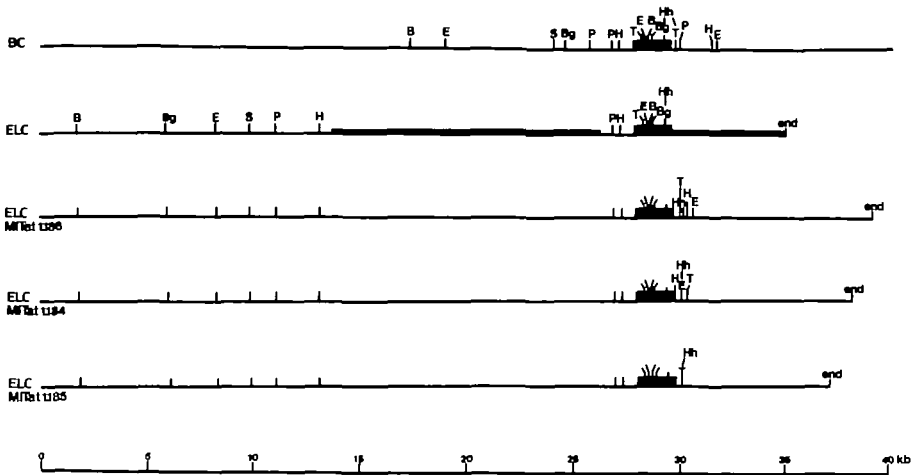


Fig. 1. Restriction maps of the BC and ELCs of the gene coding for VSG 118 showing the aberrant structure downstream of the three ELC genes analysed. The upper two maps describe the genomic environment of the 118 BC and a generalised structure of three 118 ELC genes described previously [15]. These ELC genes differ with respect to the length of the barren region (stippled) in which the transposed segment is ensconced (see text). The lower three maps are of the three aberrant 118 ELC genes described here that are distinct from the other 118 ELCs in the presence of restriction sites 3' of the gene. In all maps the coding region of the 118 gene is represented by a black box, transcription is from left to right, and 'end' indicates the end of a chromosome (see text). The HhaI and TaqI sites shown are the ones closest to the 3'-end of the genes. Abbreviations: B, BamHI; Bg, BglII; E, EcoRI; H, HindIII; Hh, HhaI; P, PvuII; S, SalI; T, TaqI.

conversion. Some of the conclusions from this work have been discussed in recent reviews [17,21,23,24].

MATERIALS AND METHODS

Trypanosomes. The isolation and characterization of cloned trypanosomes from *T. b. brucei* strain 427 have been described by Cross [1] (MITat 1.2, 1.4 and 1.5) and Michels *et al.* [15] (MITat 1.186, 1.185, 1.184, 1.192 and 1.195). *T. b. brucei* strain EATRO 1125 and *Trypanosoma evansi* SAK have been described by Frasch *et al.* [25]. *T. b. brucei* strains EATRO 3 and 795 were gifts from Dr. J.D. Barry (University of Glasgow, U.K.).

Molecular cloning. The cloning of HindIII fragments of nuclear DNA of MITat 1.186 in *E. coli* HB101 was done as described [22] with pAT 153 as cloning vehicle [26]. The library was screened with a probe for the 5'-half of the 118 VSG gene from plasmid TcV118-3 [27] that contains the complementary DNA of the VSG 118 mRNA and with a probe for the sub-telomeric element from clone TgB1.186 (see below). Plasmid DNA was isolated as described [28]. All cloned DNAs were checked for co-linearity with nuclear DNA by co-migration of appropriately digested cloned and nuclear DNA and hybridization with

the cloned insert.

Nucleotide sequence analysis. Nucleotide sequences were determined with the Maxam and Gilbert technique [29] using the modifications introduced by Cooke *et al.* [30] and Smith and Calvo [31], and with the bacteriophage M13/dideoxynucleotide technique described by Sanger *et al.* [32,33] and Messing *et al.* [34]. The sequence strategy was as follows: the 118 BC sequence from Liu *et al.* [35] was completed by the Maxam and Gilbert technique from the HincII site and the XbaI site opposite positions 200 and 221, respectively, in Fig. 4. The DNA used for this analysis was a sub-cloned fragment from a cosmid that contains the 118 BC gene area (TgBc118-29) [22]. The details of the analysis will be published elsewhere (Bernards *et al.*, in preparation). The sequence of the 1.1006 gene area was determined by both nucleotide sequence analysis techniques from the HinfI site in TgB1.1006 opposite position 216 in Fig. 4. The sequence of the 118c ELC gene area was determined by the Maxam and Gilbert technique from the HinfI and HindIII sites in Fig. 8 and by the phage M13/dideoxynucleotide technique from the HindIII, MspI, TaqI and HinfI sites indicated in Fig. 8 and from an MspI site 300 bp upstream of the 3'-end of the gene. Part of the sequence of the 3'-end of the 118c ELC gene was verified by direct sequence analysis of the 118 mRNA by reverse transcription from a 14-nucleotide primer complementary to the 3'-end of the messenger [15].

Nuclear blotting analysis. Isolation, restriction enzyme digestion, gel electrophoresis and transfer to nitrocellulose filters of nuclear DNA was done as described [8]. Gels were treated with 0.25 N HCl (40 min) prior to denaturation to improve recovery of large fragments.

Hybridization probes were isolated from appropriate digestions of plasmid DNA by separation on low-melting-temperature agarose followed by diethylaminoethyl (DEAE)-cellulose chromatography and labelling by nick translation [36]. Hybridization of ³²P-labelled probes to nitrocellulose-bound DNA was done as described [13].

BAL-31 digestion. Intact nuclear DNA from MITat 1.4 and 1.192 trypanosomes was treated with BAL-31 as described [12].

RESULTS

Isolation of the sub-telomeric element associated with the 118c ELC gene

The 118c ELC gene in MITat 1.186 is located on a 3.2-kb HindIII fragment that also contains some 600-bp 3' flanking sequences, *i.e.* the sub-telomeric element. We have isolated this HindIII fragment from a clone bank of MITat 1.186 DNA by screening with a VSG 118 cDNA probe. Fig. 2 shows the restriction map of the clone designated TgB1.186. To establish the nature and origin of the sub-telomeric element, two restriction fragments containing this region were isolated from TgB1.186 and used as hybridization probes on nuclear blots. The gene-distal probe, the 260-bp PvuIIIXHindIII fragment located 400 bp downstream of the 3'-end of the gene (Fig. 2), contains a highly repetitive element as judged from the large number of

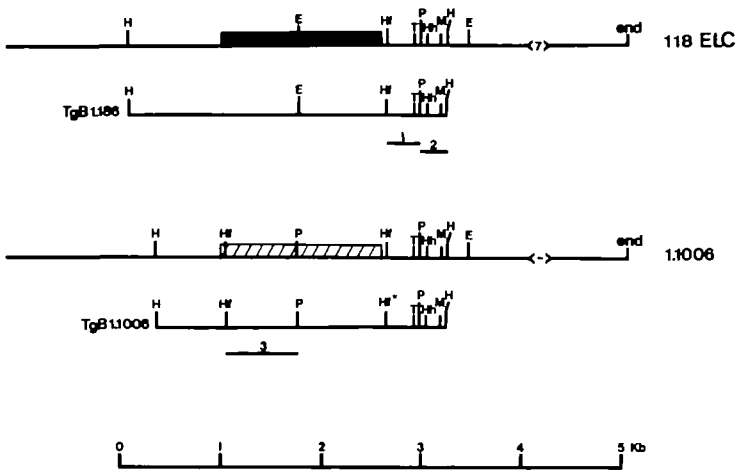


Fig. 2. Comparison of the sub-telomeric elements downstream of the 118c ELC gene and the 1.1006 gene and their cloned counterparts. The upper half shows the restriction maps of the 118c ELC gene (black box) in MITat 1.186 and the HindIII fragment from this area cloned in pAT 153 (TgB1.186). From the cloned DNA two hybridization probes (1 and 2) specific for the sub-telomeric element were isolated whose locations are indicated beneath the map of TgB1.186. With the use of probe 1, the progenitor of the sub-telomeric element was identified and isolated on a 2.9-kb HindIII fragment in clone TgB1.1006. The lower half shows the restriction map of the cloned fragment and its genomic counterpart, the 1.1006 locus that contains a tentatively identified telomeric VSG gene (see text) represented by the hatched box. Probe 3 below the map of TgB1.1006 contains the putative 5'-half of the 1.1006 gene. Deletions (—<—) were made for practical reasons; —<7>— indicates that 7 kb was removed; —<v>— indicates that the distance to the end of the chromosome downstream of the 1.1006 gene varies in different trypanosome clones. The HindIII and EcoRI sites are all shown. For other enzymes the maps are incomplete but for the sub-telomeric element. Abbreviations: Hf, HinfI; M, MspI; and see legend to Fig. 1.

HindIII fragments that are recognized under the stringent conditions used in Fig. 3 (0.1 x SSC, 65°C). The gene-proximal 330-bp HinfI x PvuII fragment, however, mainly recognizes two HindIII fragments in DNA from MITat 1.186 under the same hybridization conditions: one fragment of 3.2 kb from which the probe is derived and a 2.9-kb one (Fig. 3). The latter fragment behaves like the progenitor of the sub-telomeric element in the expression site and was, therefore, isolated from the 1.186 clone bank described above. Fig. 2 compares the restriction map of this clone (designated TgB1.1006) with that of the TgB1.186 insert derived from the 118c ELC. The 3'-halves of both maps

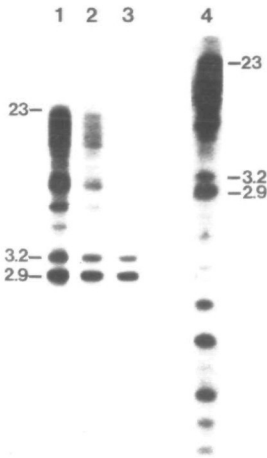


Fig. 3. The sub-telomeric element downstream of the 118c ELC has a repetitive character and is derived from a 2.9-kb HindIII fragment. Southern blots of HindIII-digested DNA from variant MITat 1.186 were hybridized to probes for the sub-telomeric element (see Fig. 2) from clone TgB1.186. Lanes 1-3 were hybridized to probe 1 and washed in 3 x SSC (lane 1), 1 x SSC (lane 2) and 0.1 x SSC (lane 3) at 65°C. Lane 4 was hybridized to probe 2 and washed in 0.1 x SSC at 65°C. Molecular weights are indicated in kb. The 3.2-kb fragment contains the 118c ELC gene; the 2.9-kb fragment contains the progenitor of the sub-telomeric element.

are identical, confirming the homology between the sub-telomeric element flanking the 118c ELC and part of clone TgB1.1006.

A complication in the analysis of the origin of the sub-telomeric element is the fact that MITat 1.186 contains two copies of the 2.9-kb HindIII fragment (see Fig. 3 and below), one of which is isolated as TgB1.1006. As the two copies of the 2.9-kb HindIII fragment have the same restriction map (see below), we assume TgB1.1006 to be representative of both. The sub-telomeric element and the 3'-end of the 118c ELC are derived from another VSG gene (MITat 1.1006)

We have determined part of the nucleotide sequence of clone TgB1.1006 starting from the HinfI site marked with an asterisk in Fig. 2. The sequence in Fig. 4 shows an open reading frame of 123 bp containing all the elements that are characteristic for 3'-ends of VSG genes (cf. ref. 35). These elements include the invariant serine and lysine residues and the Leu-Leu-Phe terminal triplet found in the C' extensions of most VSG genes [37,38]. The 3' untranslated part of the gene contains sequence motifs found in many VSG genes (underlined in Fig. 4) including the 14-nucleotide block present at the 3'-end of all VSG mRNAs [3,15]. We conclude that clone TgB1.1006 contains (part of) a VSG gene and call this gene 'MITat 1.1006' in accordance with the nomenclature proposed by Liu et al. [35] for VSG genes of T. brucei strain 427 that are found by sequence

```

118 BC   glu pro asp lys glu lys cys arg asn gly ser phe leu thr thr lys
         GAG CCT CAC AAA GAA AAG TGC CGA AAT GGT AGT TTT CTA ACA ACG AAA
118c ELC glu pro asp lys glu lys cys arg asn gly ser phe leu thr thr lys 48
         GAG CCT GAC AAA GAA AAG TGC CGA AAT GGT AGT TTT CTA ACA ACG AAA
1.1006   arg asp arg glu lys met glu thr thr asn thr thr ala ser asn sero
         AGA GAC AGA GAA AAG ATG GAA ACC ACA AAC ACC ACA GCC AGC AAT TCT

118 BC   gln phe ala leu met val ser ala ala phe val thr leu leu phe och
         CAA TCC GCC CTA ATG GTT TCT GCT GCT TTT GTT ACC TTG CTT TTT TAA
118c ELC gln phe ala leu met val ser ala ala phe val thr leu leu phe och 96
         CAA TTC GCC CTA ATG GTT TCT GCT GCT TTT GTT ACC TTG CTT TTT TAA
1.1006   phe val ile his lyso ala pro leu leu leu ala phe leuoleuopheooch
         TTT GTC ATT CAC AAG GCC CCT CTT TTG CTT GCA TTT TTG CTT TTT TAA

118 BC   crossing-over
         GTTTTTCCCCCTC--AATTTAAAAGAATTTTGGCTACTTGAAAA--CTTCTGATATATTTTAAAC
118c ELC GTTTTTCCCCCTC--AATTTAAAAGAATTTTGGCTACTTGAAAAACTTTTGATATATTTTAAAC 158
1.1006   -----TCCCCCTCCTCTTTTAAAAA---CTTIGCTACTTGAAAAACTTTTGATATATTTTAAAC
         CCTC AmTn TTGCTACT 14-mer

118 BC   ACCTTTAAATT-CTGCCGA-AAAA-TTTGAACTGT-TTTTGTAAGTTGACTG---TCTGATTAT
         **      *      *      *      *      *      *      *      *      *      *      *
118c ELC ACGT---AATTTCTGCCGATAAAAAATGAACGTGCTTTTCTAAGTTGTAAAACTTCTGATTCTN 219
1.1006   ACGT---AATTTCTGCCGATAAAAAATGAACGTGCTTTTCTAAGTTGTAAAACTTC

118 BC   CTAGAAA-----TTTTCTGGCAATTAACATTTTTTTCCTTCTTTCTTTTCTGTCT--T
         ***      *      *      *      *      *      *      *      *      *      *
118c ELC NNNAAAATTTNNTGTTTTTCTGGCGACTAAAAATTTTTTCT-CT--CCTTTTCTCGTTGAAT 280

118 BC   TTTTGTGGAAGTAGGGGTTTCATTAA--GGGGTTAGCGTTATCGTTTTTTATTTACTTFTGC
         ***** *      *      *      *      *      *      *      *      *      *
118c ELC TTTTGTGGAAGTGGCGGGGTTGTTAGGGGTTACTTAGGGGGTTACTACGGGGCTTAGG 344

```

Fig. 4. Printout of the nucleotide sequence of the 3' parts of the 118 BC gene, the 118c ELC gene and part of clone TgB1.1006. TgB1.1006 contains a VSG gene identified on the basis of conserved amino acid residues (°) and conserved sequence elements (underlined) of 3'-ends of VSG genes (cf. ref. 35 and text). Downstream of the TAA stop codon the three sequences are aligned to give maximal homology. *, Homology between the 3' flanking region of the 118 BC gene and the 118c ELC gene. The approximate point of crossing-over between the 118 BC and the 1.1006 gene is indicated. The sequence strategy is described under Materials and Methods. N, an unidentified nucleotide.

analysis and are not known to be used for VSG synthesis. Thus the sub-telomeric element downstream of the 118c ELC gene is derived from the 3' flanking sequence of another VSG gene. The sequence of the 3'-end of the VSG 118 mRNA in MITat 1.186 also differs from the 118 BC gene sequence and we have previously suggested that this new end is derived from an end exchange between the 118 gene and the preceding gene in the expression site [8,15]. Comparison of the 3'-ends of the 118 BC, 118c ELC

Downloaded from https://academic.oup.com/nar/article-abstract/11/23/8149/1057183 by guest on 25 July 2020

and 1.1006 genes in Fig. 4 shows that the newly obtained end of the 118c ELC is, indeed, derived from the 1.1006 gene. The precise point of crossing-over between the 118 BC and the 1.1006 gene cannot be determined, because the sequences are identical over 15 bp preceding the point of divergence between the 118 BC and ELC genes.

The analysis in Fig. 4 also shows that there is a high level of homology between the 3' flanking sequences of the 118 BC and 1.1006 genes. Although we have not observed a comparable homology between sequences downstream of VSG genes before (*cf.* ref. 35), extensive homology in the 3' flanking region might very well be a more general phenomenon, because a probe that contains the relevant area (the gene-proximal probe in Fig. 2) recognizes many fragments in MITat 1.186 DNA under reduced stringency of hybridization (Fig. 3).

If the 118 gene recombined with the 1.1006 gene after the formation of the 118 ELC, the heterogeneous population from which the MITat 1.186 was cloned should also contain trypanosomes carrying 'normal' 118 ELC genes, *i.e.* without a sub-telomeric element. We tested this prediction by blotting analysis of the 118 ELC genes in the heterogeneous trypanosome population from which MITat 1.186 was cloned (day 18 of the chronic infection) and preceding ones (day 15 and day 13). Detection of the TaqI fragments that contain VSG 118-specific sequences in Fig. 5 shows that the trypanosome population from which MITat 1.186 was isolated contains at least six different 118 ELC genes, in addition to the 118c ELC. In an attempt to show the presence of 'normal' ELC genes, and thus to establish the order of events in the genesis of the 118c ELC, we analysed two additional VSG 118-producing trypanosome clones from the day-18 population (MITat 1.184 and 1.185). Neither contains a 'normal' ELC. The restriction maps of the 118 ELC genes in these trypanosomes in Fig. 1 show that their sub-telomeric elements may well have arisen by deletion of the sub-telomeric element downstream of the 118c ELC. The deletion in MITat 1.184 may have occurred during the cloning procedure as its 118c ELC configuration is not detected in the day-18 population (Fig. 5). The day-18 population does, however, contain the 118 ELC gene found in

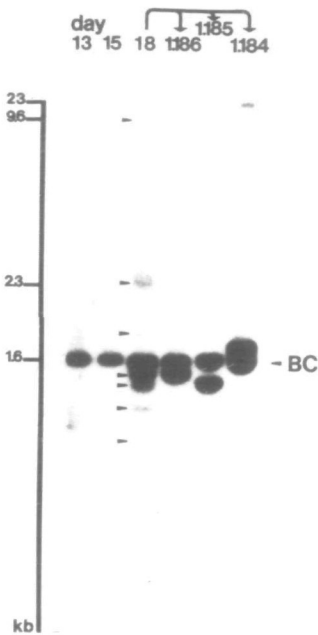


Fig. 5. Analysis of the 118 ELC genes in the chronic infection that gave rise to MITat 1.186. A Southern blot of *TaqI*-digested DNAs was hybridized to a probe for the 3'-end of VSG gene 118 isolated from a VSG 118 complementary DNA plasmid (TcV118-2; see ref. 27). The first three lanes contain DNAs from heterogeneous trypanosome populations present on day 13, 15 and 18 of a chronic infection in a rabbit. The trypanosomes were amplified in mice and rats. The last three lanes contain DNAs from three VSG 118-producing variants that were cloned from the day-18 population (MITat 1.186, 1.185 and 1.184). The *TaqI* fragment derived from the BC of the 118 gene is indicated. The arrowheads indicate the 118 ELC genes that are detected in the day-18 population.

MITat 1.185, indicating that deletions in the sub-telomeric element can also occur in vivo.

We have not found the 'normal' ELC genes in the day-18 population, that would have indicated that the 118 gene recombined with the 1.1006 gene after its duplicative transposition to the expression site. As we cannot exclude the presence of such ELCs, either, the order of events in the formation of the 118c ELC remains undetermined.

MITat 1.1006 is a telomeric gene

We have analysed the genomic environment of the MITat 1.1006 gene by nuclear blotting and observed a marked variability in the size of fragments containing the 3' flanking region. This is exemplified by the length variation of the downstream *HinfI* fragments in Fig. 6A. Such variation is exclusively seen downstream of VSG genes located near chromosome ends [3,20,39,40]. To verify this for the 1.1006 gene, intact DNA from MITat 1.4 was treated with BAL-31 nuclease for increasing time periods and then digested with *HinfI*. The *HinfI* fragments flanking the two 1.1006 genes in MITat 1.4 DNA (detected in Fig. 6B) are

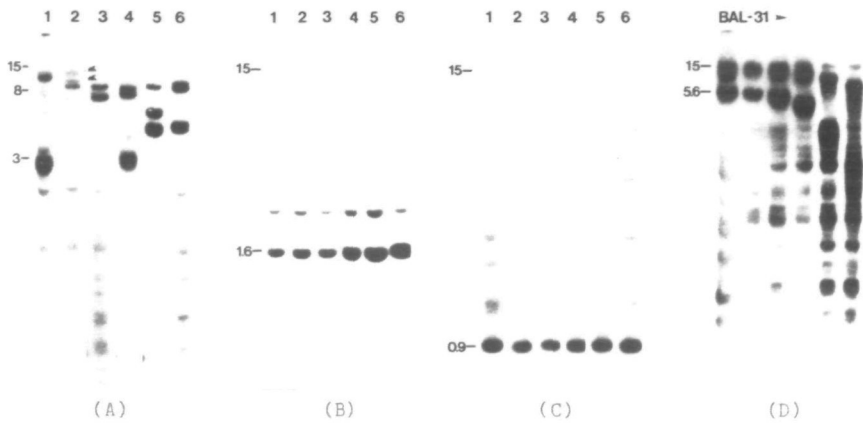


Fig. 6. Nuclear blots showing the telomeric location of the 1.1006 gene and its variable copy number. Blots of nuclear DNAs digested with *Hinf*I (A and B) and *Hinf*I×*Eco*RI (C) were hybridized to either a probe for the sub-telomeric element (probe 1 in Fig. 2) (A and C) or to a probe for the 5'-half of the 1.1006 gene (probe 3 in Fig. 2) (B). Lane 1, MITat 1.195; 2, MITat 1.186 (118c); 3, MITat 1.185; 4, MITat 1.2. Lanes 5 and 6 contain DNA from two trypanosome clones (221a-r1 and 221a-r4, resp.) that were isolated after a single relapse of clone MITat 1.2 (Bernards, A., personal communication). The intensity of the bands in A diminishes with increasing molecular weight. This is an artifact of the blotting procedure in which DNA was transferred from a 1% agarose gel without prior treatment with 0.25 N HCl. The arrowheads in A indicate two bands in lane 2 that reproduce badly on photograph. Panel D shows the BAL-31 digestion of two *Hinf*I fragments that flank the 1.1006 genes in MITat 1.4 DNA. Intact DNA was treated with BAL-31 for increasing time periods (0, 5, 10, 20, 40 and 60 min), digested with *Hinf*I, size-separated, blotted onto nitrocellulose filters and hybridized to a probe for the sub-telomeric element (probe 1 in Fig. 2).

both specifically shortened by BAL-31 exonuclease, indicating that both copies of the 1.1006 gene are located near a discontinuity in the DNA thought to be a chromosome end.

MITat 1.4, 1.195, 1.186 and 1.185 contain two copies of the 1.1006 gene whereas MITat 1.2 contains three. This third copy is probably located near a chromosome end as well, because the downstream *Hinf*I fragment varies in length in trypanosome clones derived from MITat 1.2 (Fig. 6A, lanes 5 and 6). All copies share the same restriction map (*cf.* Fig. 2) for the gene and its sub-telomeric element. Examples of the data that led to this conclusion are presented in Figs 6B and 6C, which show the invariability of the 0.9-kb *Hinf*I×*Eco*RI fragment downstream of the gene and of the 1.6-kb *Hinf*I fragment containing the gene.

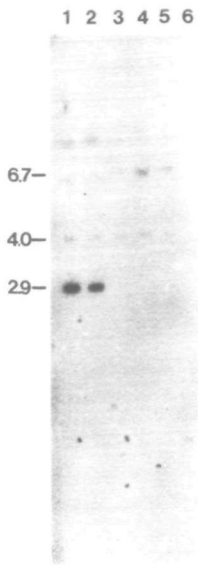


Fig. 7. Nuclear blotting analysis showing the evolutionary instability of the 1.1006 gene. Southern blots of HindIII-digested nuclear DNA from various trypanosomes were hybridized to a probe for the 5'-half of the 1.1006 gene (probe 3 in Fig. 2) and washed in 0.1 x SSC at 65°C. Lanes 1 and 2, *T. brucei* strain 427 (MITat 1.186 (118c) and 1.195, resp.); 3, *T. brucei* EATRO 1125; 4, *T. brucei* EATRO 3; 5, *T. brucei* EATRO 795; 6, *T. evansi* SAK. Molecular weights are indicated in kb.

We have previously found that the telomeric gene for VSG 221 is absent from 12 out of 13 trypanosome strains [25]. To determine whether the 1.1006 gene shows the same relative instability in trypanosome evolution, HindIII-digested DNAs from four different *T. brucei* strains and one *T. evansi* strain were probed for the presence of the 1.1006 gene. The result in Fig. 7 shows that the 1.1006 gene is present in *T. brucei* strain 427 only. Sequences related to the 1.1006 gene that are present on a 6.7-kb and on a 4-kb HindIII fragment are, however, conserved in all *T. brucei* strains.

The sequence of the sub-telomeric element and its genomic representation

The nucleotide sequence of the cloned part of the sub-telomeric element is presented in Fig. 8. It is AT-rich with a strong preference for T in the strand shown and probably does not code for protein, as the longest open reading frame is only 39 bp. The sub-telomeric element is not transcribed into stable RNA, as the gene-proximal probe (probe 1 in Fig. 2) does not detect any transcripts in total or polyadenylated RNA from MITat 1.186, whereas appropriate controls show the presence and integrity of the VSG mRNA in the same blots (not shown). The

```

14-mer                               HinfI
TCATATATTTTAAAC ACCTAATTTTC TGCCGATAAA AAATTGAACT GTGTTTTGTA AGTTGTA AAAA CTTCCTCATTC 60
NNNNAAAAAT INNTGTTTTT CTGGCGACTA AAAATTTTTT TTTCTCTCTT TTTCTGTTGA ATTTTGTGG 130
AAGTGGGGGG GGGGTCTTGA GCGCGTTACT TAGCGCGGTT AGTTAGCGCG GTTAGCTAGG ANTAGGATTA 200
GGATTAGCCT TAGGATTAGG ATTAGCGTTA CMCNGTTANG GTTAATTTTC CCTTTTTTTT AAMTCATTTT 270
TTTAATTTTC TTAATGANC CCGCGCTTTT TTAATTTTAT TTTCCGACTT TTTAAATTTT TTTTTTGT 340
TTTGANTTTC CAATTTTCAN TTTGTGCTTA GNTTTTTTAA TTAGTGATGC AGCTGCAATT TTATTATGCC 410
GACCTAACC CTCCTACCC CCGCGACGCA CCGCGGCCA CACCTCNAGT TTTTAAATT TAAAATTTAA 480
CCTTTTTTA TTCCCGACAC ATTTGTGTTT TTAATTAAT TTTTTTCTA ATTTTTTTAG TTTCCGGTTT 550
TTTCCGGGTT TGTGGTATT TTTTAAAAAT TTTCTTTAGC CATGCACCGG CTGTTTTATT ATCCGCACCC 620
TAACCCCTCT CCCCCCNCC MAACACNAGC CCGGAGAACTT
          Taq I                               PvuII
          MspI                               HindIII
    
```

Fig. 8. The nucleotide sequence of the sub-telomeric element. Printout of the nucleotide sequence downstream of the 118c ELC gene. Characteristic sequence motifs (see text) are indicated: a 21-mer pyrimidine stretch (□); 5' GGGTTA 3' elements (===); a 28-mer direct repeat (*). The sequence strategy is described under Materials and Methods. N, an unidentified nucleotide.

sequence has several unusual features. It contains many homopolymers of Ts and Gs, it contains two copies of a 28-mer sequence, a pyrimidine 21-mer, and numerous copies of the sequence 5' GGGTTA 3' and its complement CCCTAA (positions 413 and 618). Straight tandem repeats of this CCCTAA sequence have

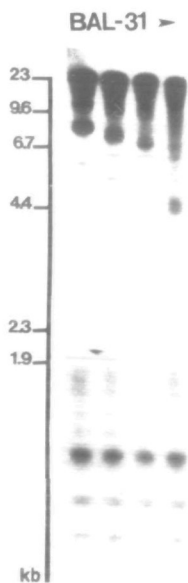


Fig. 9. BAL-31 assay indicating the telomeric location of repetitive sequences homologous to the gene-distal part of the sub-telomeric element. A Southern blot is shown of MITat 1.192 DNA treated with BAL-31 exonuclease for increasing time periods (0, 2, 7 and 20 min) and digested with BspI. The filter-bound DNA was hybridized to the PvuIIxHindIII probe from the sub-telomeric element downstream of the 118c ELC gene (probe 2 in Fig. 2) and washed in 0.1 x SSC at 65°C.

recently been found to make up the terminal part of trypanosome telomeres [24; Van der Ploeg, L.H.T., personal communication].

The highly conserved repetitive element in the gene-distal part of the sub-telomeric element (see above), situated between the PvuII site of position 390 and the HindIII site at position 657, contains a long T-rich stretch bracketed by the 28-mer direct repeats. The majority of the repeat elements homologous to this area are located close to chromosome ends. This follows from the experiment shown in Fig. 9 in which the genomic position of the repeat elements was probed for by exonuclease BAL-31 digestion of intact nuclear DNA.

DISCUSSION

Our analysis of the aberrant 118c ELC gene has revealed a novel gene conversion event between a chromosome-internal and a telomeric VSG gene: the 118 ELC gene has procured the 3'-end and the downstream area of the telomeric MITat 1.1006 gene. The recombination producing the 118c ELC probably took place during the duplicative transposition of the 118 gene to the expression site that contained an ELC of the 1.1006 gene. We do not know, however, whether the 1.1006 gene can form an ELC and we have not observed an extra copy of the gene in DNA extracted from heterogeneous trypanosome populations isolated earlier in the chronic infection (not shown). Alternatively, the 1.1006 sequences may have been added to the 118c ELC gene after its formation. This scheme predicts that the heterogeneous population from which MITat 1.186 (variant 118c) was derived should also contain trypanosomes carrying a 'normal' 118 ELC. We have not found these, but they may have been present in low concentrations.

Two observations indicate that the sub telomeric element and at least the 3'-end of the 1.1006 gene moved to the expression site by duplicative transposition. First, the 1.1006 genes, like one copy of the ILTat 1.3 gene in the ILTat repertoire [20], are located on mini-chromosomes, whereas the 118c ELC gene resides on a chromosome with an apparent molecular weight of 1800 kb that can be readily separated from mini-chromosomes by pulsed-field gradient gel electrophoresis (Van der Ploeg, L.H.T., personal communication). Second, the transposition of

1.1006 sequences to the 1800-kb chromosome must have involved a duplication as we have not observed the loss of one 1.1006 gene copy in MITat 1.186.

The duplicative transposition of the 1.1006 gene proposed here differs from the duplicative transposition of internal genes in that at least 1 kb of 3' flanking sequences and possibly the whole chromosome end is co-transposed to the expression site. Direct evidence that such a telomere conversion can occur in our trypanosome strain has recently been obtained for the 221 gene [23; Bernards, A., unpublished].

Our isolation of two trypanosome clones that express the 118 gene but have lost part of the sub-telomeric element, suggests that this element is instable downstream of the 118 ELC in the expression site. This is remarkable because we have found no indication for instability of the element in its original location (behind the MITat 1.1006 gene). It is possible that the association with a highly transcribed gene promotes internal recombination in such elements littered with internal straight repeats. Since we have not found trypanosomes in which the 1.1006 gene itself is activated, we have been unable to verify this possibility.

The lack of recombinant plasmids containing the 3' flanking sequence of ELC genes and other telomeric VSG genes had thusfar precluded analysis of this DNA. Two major points emerge from our data on the sub-telomeric element downstream of the 118c ELC and the 1.1006 genes:

1. This region harbours a highly repetitive conserved sequence that is preferentially located close to chromosome ends and that contains sequence elements present at most or all trypanosome telomeres.

2. We have found that the 118 BC gene and the 1.1006 gene have a high level of homology in their flanking sequence. The homology extends to 300 bp downstream of the 3'-end of the mature transcript and is localised in a region that is repeated in the genome. As the two genes have very different coding regions, the downstream homology may bear witness of reshuffling of parts of VSG genes possibly by (segmental) gene conversion. Analogous segmental gene conversion processes in

telomeric VSG genes have been observed by E.Pays and co-workers (personal communication). Furthermore, the observed homology shows that the homology between the incoming gene and the resident ELC need not be limited to the conserved 3'-ends of the genes, but that the 3' flanking region may play a role in the gene conversion events in the expression site.

MITat 1.2 and variants derived thereof contain an extra telomeric copy of the 1.1006 gene that could also have been generated by telomere conversion. This extra copy could represent an ELC of the 1.1006 gene that is not destroyed during switch-off. Such 'lingering' ELCs are sometimes observed after the switching-off of the 121 gene [4] and the 118 gene [23; Michels, P.A.M., personal communication]. We do not know, however, whether the appearance of an extra copy is linked to expression of the 1.1006 gene. Alternatively, telomere conversion could occur independently of VSG gene switching, but be usually missed because the product is not selected for by the host's immune system. The changes in copy number of the 1.1006 gene are probably not instrumental in VSG gene evolution. Expansion of a gene family by generation of new telomeric genes appears to be inefficient due to the high frequency at which these genes are lost. This is amply illustrated by the absence of the 1.1006 gene and two other telomeric VSG genes [25; Leegwater, P.A.J., personal communication] from other T. brucei strains.

ACKNOWLEDGEMENTS

We thank Mr A.Bernards and Mr P.A.J.Leegwater for providing part of the 118 BC sequence and sharing unpublished results; Mrs F.Fase-Fowler for isolation of DNAs; Mrs M.M.W.Van der Bijl for initial analyses of MITat 1.184 and 1.185, and Mr L.H.T.Van der Ploeg for performing the BAL-31 experiment on MITat 1.195 DNA and sharing unpublished results. We thank Dr. J.D.Barry (Department of Zoology, University of Glasgow, Glasgow G12 8QQ, U.K.) for generous gifts of strains EATRO 3 and 795. This work was supported in part by a grant to P.B. from the Foundation for Fundamental Biological Research (BION), which is subsidized by The Netherlands Organization for the Advancement of Pure Research (ZWO).

+ Present address: Research Unit for Tropical Diseases, International Institute of Cellular and Molecular Pathology, Ave Hippocrate 74, B-1200 Brussels (Belgium).

Abbreviations: BC, basic copy; bp, base pair(s); EATRO, East African Trypanosomiasis Research Organization; ELC, expression-linked (extra) copy; ILTat, ILRAD Trypanosome antigen type; kb, kilo base pair(s); MITat, Molteno Institute Trypanosome antigen type; mRNA, messenger RNA; SSC, 0.15 M NaCl, 0.015 M sodium citrate (pH 7.0); VSG, variant surface glycoprotein.

REFERENCES

- 1 Cross, G.A.M. (1975) *Parasitol.* 71, 393-417.
- 2 Cross, G.A.M. (1978) *Proc.Roy.Soc.London Ser. B* 202, 55-72.
- 3 Borst, P. and Cross, G.A.M. (1982) *Cell* 29, 291-303.
- 4 Hoeijmakers, J.H.J., Frasnch, A.C.C., Bernardis, A., Borst, P. and Cross, G.A.M. (1980) *Nature* 284, 78-80.
- 5 Borst, P., Frasnch, A.C.C., Bernardis, A., Van der Ploeg, L.H.T., Hoeijmakers, J.H.J., Arnberg, A.C. and Cross, G.A.M. (1981) *Cold Spring Harbor Symp.Quant.Biol.* 45, 935-943.
- 6 Pays, E., Van Meirvenne, N., LeRay, D. and Steinert, M. (1981) *Proc. Natl.Acad.Sci.U.S.* 78, 2673-2677.
- 7 Pays, E., Lheureux, M. and Steinert, M. (1981) *Nature* 292, 265-267.
- 8 Bernardis, A., Van der Ploeg, L.H.T., Frasnch, A.C.C., Borst, P., Boothroyd, J.C., Coleman, S. and Cross, G.A.M. (1981) *Cell* 27, 497-505.
- 9 Van der Ploeg, L.H.T., Liu, A.Y.C., Michels, P.A.M., De Lange, T., Borst, P., Majumder, H.K., Weber, H., Veeneman, G.H. and Van Boom, J. (1982) *Nucl.Acids Res.* 10, 3591-3604.
- 10 Boothroyd, J.C. and Cross, G.A.M. (1982) *Gene* 20, 281-289.
- 11 De Lange, T., Liu, A.Y.C., Van der Ploeg, L.H.T., Borst, P., Tromp, M.C. and Van Boom, J.H. (1983) *Cell* 34, 891-900.
- 12 De Lange, T. and Borst, P. (1982) *Nature* 299, 451-453.
- 13 Van der Ploeg, L.H.T., Bernardis, A., Rijsewijk, F.A.M. and Borst, P. (1982) *Nucl.Acids Res.* 10, 593-609.
- 14 Michels, P.A.M., Bernardis, A., Van der Ploeg, L.H.T. and Borst, P. (1982) *Nucl.Acids Res.* 10, 2353-2366.
- 15 Michels, P.A.M., Liu, A.Y.C., Bernardis, A., Sloof, P., Van der Bijl, M.M.W., Schinkel, A.H., Menke, H.H., Borst, P., Veeneman, G.H., Tromp, M.C. and Van Boom, J.H. (1983) *J.Mol.Biol.* 166, 537-556.
- 16 Bernardis, A., Michels, P.A.M., Lincke, C.R. and Borst, P. (1983) *Nature* 303, 592-597.
- 17 Borst, P., Bernardis, A., Van der Ploeg, L.H.T., Michels, P.A.M., Liu, A.Y.C., De Lange, T. and Sloof, P. (1984) in: *Molecular Biology of Host-Parasite Interactions* (Agabian, N. and Eisen, H., Eds), Liss, New York, in press.
- 18 Williams, R.O., Young, J.R. and Majiwa, P.A.O. (1979) *Nature* 282, 847-849.
- 19 Bernardis, A. (1982) *TIBS* 7, 253-255.
- 20 Williams, R.O., Young, J.R. and Majiwa, P.A.O. (1982) *Nature* 299, 417-421.
- 21 Borst, P., Bernardis, A., Van der Ploeg, L.H.T., Michels, P.A.M., Liu, A.Y.C., De Lange, T., Sloof, P., Veeneman, G.H., Tromp, M.C. and Van Boom, J.H. (1983) in: *Genetic Arrangement* (Chater, K.F., Cullis, C.A., Hopwood, D.A., Johnston, A.A.W.B. and Woolhouse, H.W., Eds), Croom Helm, London, pp. 207-233.
- 22 Van der Ploeg, L.H.T., Valerio, D., De Lange, T., Bernardis, A., Borst, P. and Grosveld, F.G. (1982) *Nucl.Acids Res.* 10, 5905-5923.
- 23 Borst, P., Bernardis, A., Van der Ploeg, L.H.T., Michels, P.A.M., Liu, A.Y.C., De Lange, T., Sloof, P., Schwartz, D.C. and Cantor, C.R. (1983) in: *Gene Expression, UCLA Symp. on Molecular and Cellular Biology, New Series Vol. 8* (Hamer, D. and Rosenberg, M., Eds), Liss,

- New York, in press.
- 24 Borst, P., Bernardis, A., Van der Ploeg, L.H.T., Michels, P.A.M., Liu, A.Y.C., De Lange, T. and Kooter, J.M. (1983) *Europ.J.Biochem.*, in press.
- 25 Frasch, A.C.C., Borst, P. and Van den Burg, J. (1982) *Gene* 17, 197-211.
- 26 Twigg, A.J. and Sherratt, D. (1980) *Nature* 283, 216-218.
- 27 Hoeijmakers, J.H.J., Borst, P., Van den Burg, J., Weissmann, C. and Cross, G.A.M. (1980) *Gene* 8, 391-417.
- 28 Birnboim, H.C. and Doly, J. (1979) *Nucl.Acids Res.* 7, 1513-1523.
- 29 Maxam, A.M. and Gilbert, W. (1980) in: *Methods in Enzymology* (Grossman, L. and Moldave, K., Eds), Vol. 65, Academic Press, New York, pp. 499-560.
- 30 Cooke, N.E., Coit, D., Weiner, R.I., Baxter, J.D. and Martial, J.A. (1980) *J.Biol.Chem.* 255, 6502-6510.
- 31 Smith, D.R. and Calvo, J.M. (1980) *Nucl.Acids Res.* 8, 2255-2274.
- 32 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc.Natl.Acad. Sci.U.S.* 74, 5463-5467.
- 33 Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) *J.Mol.Biol.* 143, 161-178.
- 34 Messing, J., Crea, R. and Seeburg, P.H. (1981) *Nucl.Acids Res.* 9, 309-321.
- 35 Liu, A.Y.C., Van der Ploeg, L.H.T., Rijsewijk, F.A.M. and Borst, P. (1983) *J.Mol.Biol.* 167, 57-75.
- 36 Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J.Mol. Biol.* 113, 237-251.
- 37 Rice-Ficht, A.C., Chen, K.K. and Donelson, J.E. (1981) *Nature* 294, 53-37.
- 38 Majumder, H.K., Boothroyd, J.C. and Weber, H. (1981) *Nucl.Acids Res.* 9, 4745-4753.
- 39 Donelson, J.E., Young, J.R., Dorfman, D., Majiwa, P.A.O. and Williams, R.O. (1982) *Nucl.Acids Res.* 10, 6581-6595.
- 40 Young, J.R., Donelson, J.E., Majiwa, P.A.O., Shapiro, S.Z. and Williams, R.O. (1982) *Nucl.Acids Res.* 10, 803-819.