The EMBO Journal vol.4 no.12 pp.3299–3306, 1985

Transcription of a transposed trypanosome surface antigen gene starts upstream of the transposed segment

Titia De Lange, Jan M.Kooter, Joen Luirink and Piet Borst

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

Communicated by P.Borst

The non-telomeric variant surface glycoprotein (VSG) genes in *Trypanosoma brucei* are activated by a duplicative transposition to a telomeric expression site. We have determined the 5′ end of the transposed segment of the gene for VSG 117 and infer from comparison with similar data obtained by others that the crossover can occur at variable positions within short repeats present upstream of this gene and in the expression site. We have analysed nascent and steady state transcripts of the transposed gene and its neighbouring expression site DNA. The results indicate that transcription starts upstream of the transposed gene segment in the expression site and that transcripts are rapidly processed at specific points identified by protection of DNA-RNA hybrids against digestion by nuclease S1 or Exo VII. Hence, this gene appears to be activated by a process akin to promoter addition.
Key words: antigenic variation/duplicative transposition/VSG gene activation

Introduction

Unlike other protozoan parasites that infect mammals, the African trypanosomes do not enter the cells of their host. Instead they multiply in the bloodstream where they are fully exposed to the host's immune defences. Trypanosomes survive this hostile environment by antigenic variation (for recent reviews, see Borst *et al.*, 1983; Parsons *et al.*, 1984a; Michels, 1984; Bernards, 1985; Boothroyd, 1985).

Bloodstream trypanosomes are completely covered by a homogeneous layer of antigenic protein, the variant surface glycoprotein (VSG) (Cross, 1975, 1978). Through the sequential expression of different VSGs from a large repertoire, trypanosomes evade immune destruction (Vickerman, 1978). Most of the ~1000 surface antigen genes occur linked in the genome (Van der Ploeg *et al.*, 1982a); in addition, many VSG genes reside close to chromosome ends (Williams *et al.*, 1982; Van der Ploeg *et al.*, 1984a). The trypanosome genome contains an unknown number of telomeres in which VSG genes can be expressed. Exchange of genes in such an expression site can lead to the production of a new surface coat (Bernards *et al.*, 1981; Pays *et al.*, 1981, 1983a; Michels *et al.*, 1983). Alternate activation of different expression sites can also result in antigenic switches (Young *et al.*, 1982; Laurent *et al.*, 1984; Myler *et al.*, 1984; Van der Ploeg *et al.*, 1984a).

Non-telomeric VSG genes (called Basic Copy or BC genes) are activated by a duplicative transposition to a telomeric expression site. The duplicate gene is called Expression Linked extra Copy (ELC) and transcribed (Hoeijmakers *et al.*, 1980a; Pays *et al.*, 1981; Bernards *et al.*, 1981) (cf. Figure 1). During the genesis of a new ELC, the preceding ELC gene is destroyed, probably by a process akin to gene conversion (Pays *et al.*, 1983b; Michels *et al.*, 1983). The 3′ end of the converted gene segment resides at or close to the highly conserved 3′ untranslated trailer region (Bernards *et al.*, 1981; Michels *et al.*, 1983). The other end falls outside the coding region, some 1–2 kb upstream of the start codon (Van der Ploeg *et al.*, 1982b; Michels *et al.*, 1982, 1983). At this position, VSG genes have a variable number of short repeats (called 70-bp repeat here), that were proposed to be the sequence in which the crossover occurs (Van der Ploeg *et al.*, 1982b; Liu *et al.*, 1983). The upstream region of ELC genes, beyond the 5′ crossover, is characterized by a paucity of restriction enzyme recognition sites and by its length variation in different trypanosome clones (cf. Figure 1). Campbell *et al.* (1984a) first succeeded in cloning a small segment of this 'barren' region upstream of the 117a ELC gene and showed by sequence analysis that the 5′ crossover is in the first 70-bp repeat upstream of the 117 BC gene. Furthermore, their analysis showed that at least the first 1.5 kb of the barren region consists of 70-bp repeats. Here we extend this analysis and show that the crossover point, like the 3′ crossover point, can vary in different ELCs formed from the same BC gene.

How duplicated VSG genes are activated in the expression site has remained unclear. Transcriptional activation could be a position effect bestowed as a result of entry into the expression site, or the expression site could provide the ELC gene with DNA signals required for transcription, i.e., an activation mechanism analogous to promoter addition. Localization of the point where transcription of VSG genes is initiated should clarify this matter. Unfortunately, in trypanosomes the start of transcription cannot be found simply by a search for the DNA that codes the 5′ end of the VSG mRNA. Trypanosomes have a unique system for transcription of protein coding genes yielding mRNAs with a constant block of 35 nucleotides at the 5′ end (De Lange *et al.*, 1984a; Parsons *et al.*, 1984b). This mini-exon sequence is not encoded by protein coding genes, but provided *in trans* from a small transcript, derived from a highly repeated element (cf. Kooter *et al.*, 1984; Campbell *et al.*, 1984b; De Lange *et al.*, 1983, 1984b; Nelson *et al.*, 1983, 1984; Milhausen *et al.*, 1984; Laird *et al.*, 1985). How this mini-exon sequence becomes linked to the remainder of the mRNA is not yet clear. As a consequence of this discontinuous transcription, the sequence at the 5′ end of mRNAs is not necessarily found close to the start of transcription. We have therefore analyzed steady state and nascent RNAs from an ELC gene for VSG 117 to determine whether the transcription of this gene starts in the expression site or within the transposed gene segment.

Results

The 5′ end of the transposed gene segment

From trypanosome variant 117b we have isolated a small *Mbo*I-*Sal*I restriction fragment that spans the 5′ end of the transposed gene segment (see clone TgE 117b in Figure 1) and determined

part of its nucleotide sequence. In Figure 2 the sequence around the 5′ end of the transposed segment is compared with the BC gene sequence and with the 5′ end sequence of the transposed segment of the 117 gene in another VSG 117 expressor (variant 117a), previously determined by Campbell *et al.* (1984a). In both 117 ELC genes the crossover is within the first four 70-bp repeats upstream of the 117 BC gene and in both cases the barren region upstream of the ELC gene contains 70-bp repeats (Figure 2). The point of recombination in the duplicative transposition forming the 117a and 117b ELC genes differs. The 5′ end of the transposed segment of gene 117a is within the first 70-bp repeat whereas the first sequence divergence in the 117b ELC occurs 140 bp further upstream in an unusual half-deleted 70-bp repeat that is typical for the 117 BC gene. The presence in the expression site of this rare tell-tale 70-bp repeat (only observed once in >25 70-bp repeats examined thus far, Liu *et al.*, 1983; Campbell *et al.*, 1984a; Bernards *et al.*, 1985) indicates that the 5′ end of the transposed segment of the 117b gene indeed resides at or close to the point of sequence divergence.

Fig. 1. The BC gene and two ELC genes for VSG 117. The physical maps are from Michels *et al.* (1983). Chromosome ends are indicated with 'end'. The black boxes depict the main exon of the 117 gene, transcription is from left to right (see arrow in the BC map). The extent of the transposed gene segment is shown above the 117b ELC gene. Beneath the maps, three cloned DNAs are shown: TgB 117-1 is a genomic *Eco*RI-*Hind*III fragment from the 117 BC gene (Bernards *et al.*, 1981), TcV 117-5 is a cDNA of mRNA for VSG 117a (Bernards *et al.*, 1981; Hoeijmakers *et al.*, 1980b) and TgE 117b is an *Mbo*I-*Sal*I fragment that spans the 5′ end of the transposed segment of the ELC gene for VSG 117b. Abbreviations: E, *Eco*RI; H, *Hind*III; M, *Mbo*I or *Sau*3A; P, *Pst*I; Pv, *Pvu*II; S, *Sal*I.

Stable RNA from the 117b ELC gene

The RNA blot in Figure 3 illustrates the hybridization of stable poly(A)$^+$ RNA with DNA probes from the 117b ELC gene and upstream DNA. Previous work by Van der Ploeg *et al.* (1982c) and Pays *et al.* (1982) on other ELC genes had already demonstrated the presence of several minor polyadenylated transcripts that hybridized with probes from the co-transposed segment. An analogous set of RNAs is detected by probe B3 in 117b RNA. The apparent size of one of these transcripts is 3.7 kb (indicated by an arrow in Figure 3) and this transcript is also detected by a probe from the coding region (C1). This RNA may therefore cover the entire transposed segment, as indicated by S1 analysis

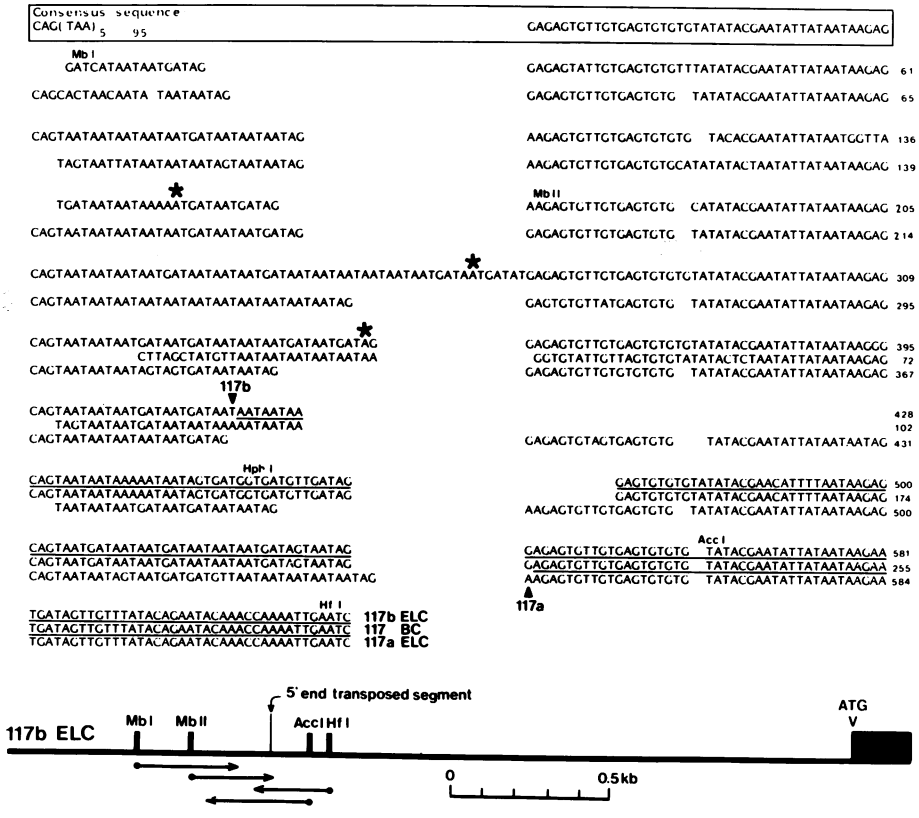

Fig. 2. Comparison of the 5′ end of the transposed segment of two ELC genes for VSG 117. The nucleotide sequence 5′ of the *Hinf*I site 1613 bp upstream of the 117 BC gene (Boothroyd and Cross, 1982), the 117a ELC gene (Campbell *et al.*, 1984a) and the 117b ELC gene are shown. The arrowheads indicate the position of sequence divergence of ELC and BC; 100% homology downstream of these positions is indicated with an unbroken line between two sequences. The sequences are aligned with the 70-bp repeat consensus (Campbell *et al.*, 1984a) (top) to show the repetitive nature of the DNA at the 5′ end of the transposed segments. Restriction sites used for sequence analysis and a *Hph*I site used in the analysis of nascent transcripts (Figure 5) are shown. The map below shows the position of the area upstream of the 117b ELC gene of which the sequence was determined following the indicated strategy. The ends (±5 bp) of the S1 and Exo VII protected fragments found in panels **A** and **B** of Figure 4 are indicated by an asterisk. Abbreviations: HfI, *Hinf*I; MbI, *Mbo*I or *Sau*3A; MbII, *Mbo*II.

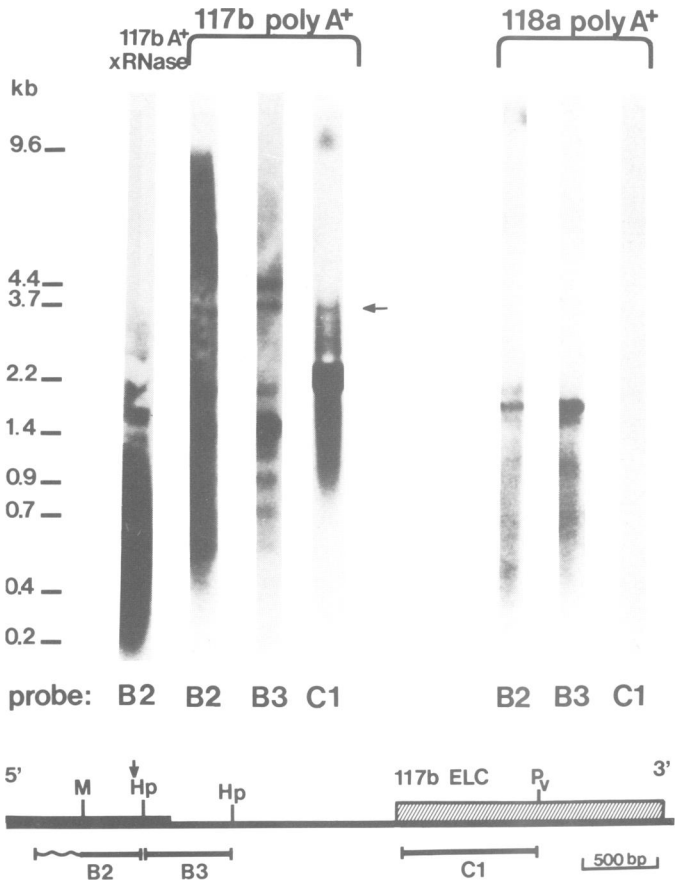

Fig. 3. Blot analysis of steady state RNA from the ELC gene for VSG 117b. Polyadenylated RNA (5 μg per lane) from variants 117b and 118a was size separated on a 1% agarose gel after glyoxal treatment and blotted onto a nitrocellulose filter. In the first lane on the left 5 μg RNase A-digested polyadenylated RNA from variant 117b was run. The filter strips were hybridized to nick-translated DNA probes from TcV 117-5 (probe C1) and TgE 117b (probes B2 and B3). The localization of the probes is shown in the lower part of the figure. Probe B2 contains vector sequences (wavy line). Post-hybridization washes were in 1 × SSC/0.1% SDS at 65°C. The right hand panel was exposed four times longer. The exposure time times specific activity of probe C1 was ~4 times lower than for the other probes. The approximate mol. wts. indicated are derived from glyoxal-treated DNA markers (λ × *Hind*III and pAT 153 × *Hinf*I). The arrow points at a minor transcript of ~3.7 kb that is detected with probes C1 and B3. Abbreviations: Hp, *Hph*I; M, *Mbo*I or *Sau*3A; Pv, *Pvu*II.

for the corresponding RNA from variant 118a (Van der Ploeg *et al.*, 1982c). Although the presence of these transcripts is compatible with the existence of a rapidly processed precursor of mature VSG mRNA, these results do not resolve whether transcription starts at the border of the transposed segment or more upstream. Probe B2 (see Figure 3) provides information on this point. This probe, which contains 70-bp repeats, detects a remarkable abundance of transcripts. This hybridization pattern is not an artefact of contaminating DNA in the RNA preparation: limited RNase A treatment reduces the size of the hybridizing RNA molecules (Figure 3), whereas DNase I had no effect (not shown). We attribute the high hybridization with this probe to the presence of processing products of transcripts of the barren region upstream of the transposed segment. Since the barren region probably mainly consists of 70-bp repeats, transcripts derived thereof should hybridize strongly with probe B2 under the reduced stringency used. The synthesis of these transcripts is dependent on VSG gene expression, because probe B2 does not detect transcripts in cultured (pro-cyclic) trypanosomes that do not make VSGs (not shown). The limited hybridization of poly(A)$^+$ RNA from variant 118a to probes B2 and B3 can be attributed to the presence of 70-bp repeats at corresponding positions upstream of the 118 and 117 ELCs. The processing products do not seem to contain the coding region of the 117 gene, because they are not detected by probe C1 (Figure 3).

The RNAs that hybridize to probe B2 are rather unstable: frequently their abundance in RNA preparations is much lower than in the RNA preparation used for the experiment in Figure 3. What factor(s) influence this variation is unknown, but similar observations have been made on transcripts from the upstream region of other ELC genes (Van der Ploeg *et al.*, 1982c). This irreproducibility may be linked to the rapid switch off of VSG gene expression that we have observed after cardiac puncture of infected rats (Kooter and Borst, 1984; Bernards *et al.*, 1985).

To demonstrate that stable transcripts exist in variant 117b that actually cross the boundary between transposed segment and upstream sequences, the DNA fragments containing this boundary were used in S1 nuclease and Exo VII protection experiments. Figure 4 shows the results. Panel A, lane 5 shows that S1 digestion yields four specific protected fragments and protection is only observed with homologous 117b RNA, not with RNA from variant 118 (panel A, lanes 6 and 7). The largest protected fragment is 620 nucleotides, which corresponds to the size of the trypanosome DNA contained in the S1 probe (probe I in Figure 4). To test whether the three smaller protected fragments are derived from preferred processing sites in the precursor RNA or reflect internal cutting of nuclease S1, the hybrids were also digested with Exo VII, a pure exonuclease (Berk and Sharp, 1978a, 1978b). Panels B and C of Figure 4 show that Exo VII gives the same protected fragments (panel B, lane 5) as nuclease S1. Protection is only observed with one strand, the complementary strand does not yield specific products (panel B, lane 15). Protection is only observed with homologous RNA and not with RNA from another variant (panel B, lanes 8 – 11). These results prove the existence of variant-specific transcripts that cross the boundary of the transposed segment in the expression site. The yield of these transcripts obtained with Exo VII is much higher, however, than with S1 nuclease. We attribute this to the difference in substrate specificity between S1 and Exo VII. S1 is prone to digest AT-rich hybrids, even if perfectly matched. Figure 2 shows that the DNA segment analysed contains long stretches of pure AT.

In an attempt to stabilize the hybrid against S1 nuclease by using a longer DNA fragment, the entire insert of plasmid TgE117b was used (probe II in Figure 4). This gave a rather broad band centered at 1350 nucleotides after Exo VII digestion (panel C, lanes 6 and 8) in agreement with the results obtained with probe I. However, probe II yields a second band of ~700 nucleotides which identifies an RNA with its 5′ end within the transposed segment as indicated in the lower part of Figure 4. This might reflect another processing site in the precursor RNA. The approximate positions of some of these sites are shown in Figure 2. The results are compatible with preferential processing of the precursor RNAs in the AU-rich areas of the 70-bp repeats.

The results of this analysis of stable RNA are compatible with a model in which transcription starts far upstream of the transposed segment and yields transcripts that are rapidly and differentially processed. The variable representation of different transcripts and the absence of very long transcripts detected by the mRNA-specific probe C1, made it desirable to test the

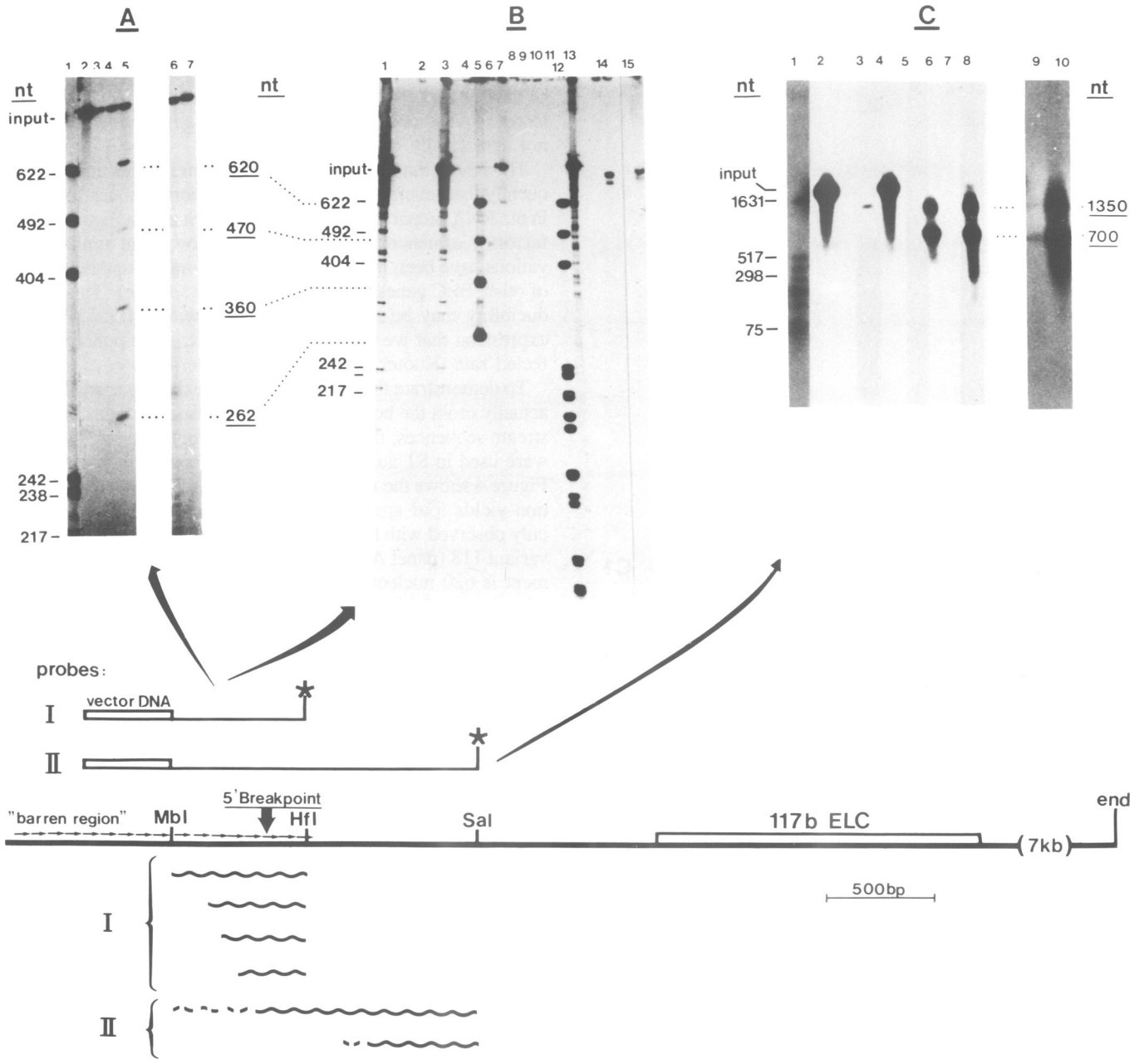

Fig. 4. Mapping of steady state RNA of the upstream region of the 117b ELC by S1 nuclease and Exo VII. In **panel A** and **B** a 1000 nucleotide single-stranded DNA fragment was used, 5′ end-labeled at the *Hinf*I site, indicated by probe I in the lower part of the figure. **Panel A: lane 1**, pAT 153 × *Msp*I 5′-end labeled marker fragments; **lane 2**, input fragment; **lane 3**, hybridized with *E. coli* tRNA at 38°C and digested with S1; **lane 4**, hybridized with variant 117b RNA at 42°C, S1 digested; **lane 5**, as **lane 4** but hybridized at 38°C. **Lanes 6** and **7**, as **lanes 4** and **5** but hybridized with non-homologous RNA from variant 118a. **Panel B: lane 1**, input fragment (and some degradation products); **lane 2**, hybridized with *E. coli* tRNA at 38°C and digested with S1; **lane 3**, hybridized with 117b RNA undigested; **lane 4**, hybridized with 117b RNA at 38°C, treated with S1 (only after a very long exposure are the protected fragments seen); **lane 5**, as **lane 4** but digested with Exo VII. **Lanes 6** and **7**, as **lanes 4** and **5**, respectively, but hybridized at 42°C. **Lanes 8–11**, as **lanes 4–7**, respectively, except that non-homologous RNA from variant Mitat 1.3 was used. **Lane 12**, 5′-end labeled pAT 153 × *Msp*I marker fragments; **lanes 13–15**, as **lanes 1, 2** and **5** but with the complementary strand. The sizes of the S1 nuclease and Exo VII protected fragments are underlined and indicated between **panels A** and **B**. In **panel C** a single-stranded probe of 2000 nucleotides was used, 5′-end labeled at the *Sal*I site (probe II); the products were loaded on an alkaline agarose gel. **Lane 1**, 5′-end labeled pAT 153 × *Hinf*I marker fragments; **lane 2**, input fragment; **lane 3**, hybridized with *E. coli* tRNA at 38°C, digested with Exo VII; **lane 4**, hybridized with 117b RNA at 38°C, undigested; **lane 5**, hybridized with 117b RNA at 42°C, digested with S1; **lane 6**, as **lane 5** but digested with Exo VII; **lanes 7** and **8**, as **lanes 5** and **6** but hybridized at 38°C. **Lanes 9** and **10** present a longer exposure of **lanes 5** and **6** to show the co-migration of the S1 and Exo VII protected fragments. The sizes of these fragments are underlined. The lower part of the figure shows the phsyical map of the 117b ELC telomere in which only the relevant restriction sites are shown; the VSG coding region is indicated by the open bar, the end denotes the end of the chromosome; the 5′ edge of the transposed segment is indicated by the arrowhead; upstream of it lies the 'barren region' probably also composed of 70-bp repeats (little arrows). Above the map the positions of the probes are indicated, the * denotes the labeled end. The wavy lines below the map schematically show the 5′ ends of the RNAs that gave rise to the protected fragments; because of the broad bands in **lanes 4** and **6** of **panel C** the 5′ ends of RNAs are interrupted. Abbreviations: HfI, *Hinf*I; MbI, *Mbo*I; SaI, *Sal*I.

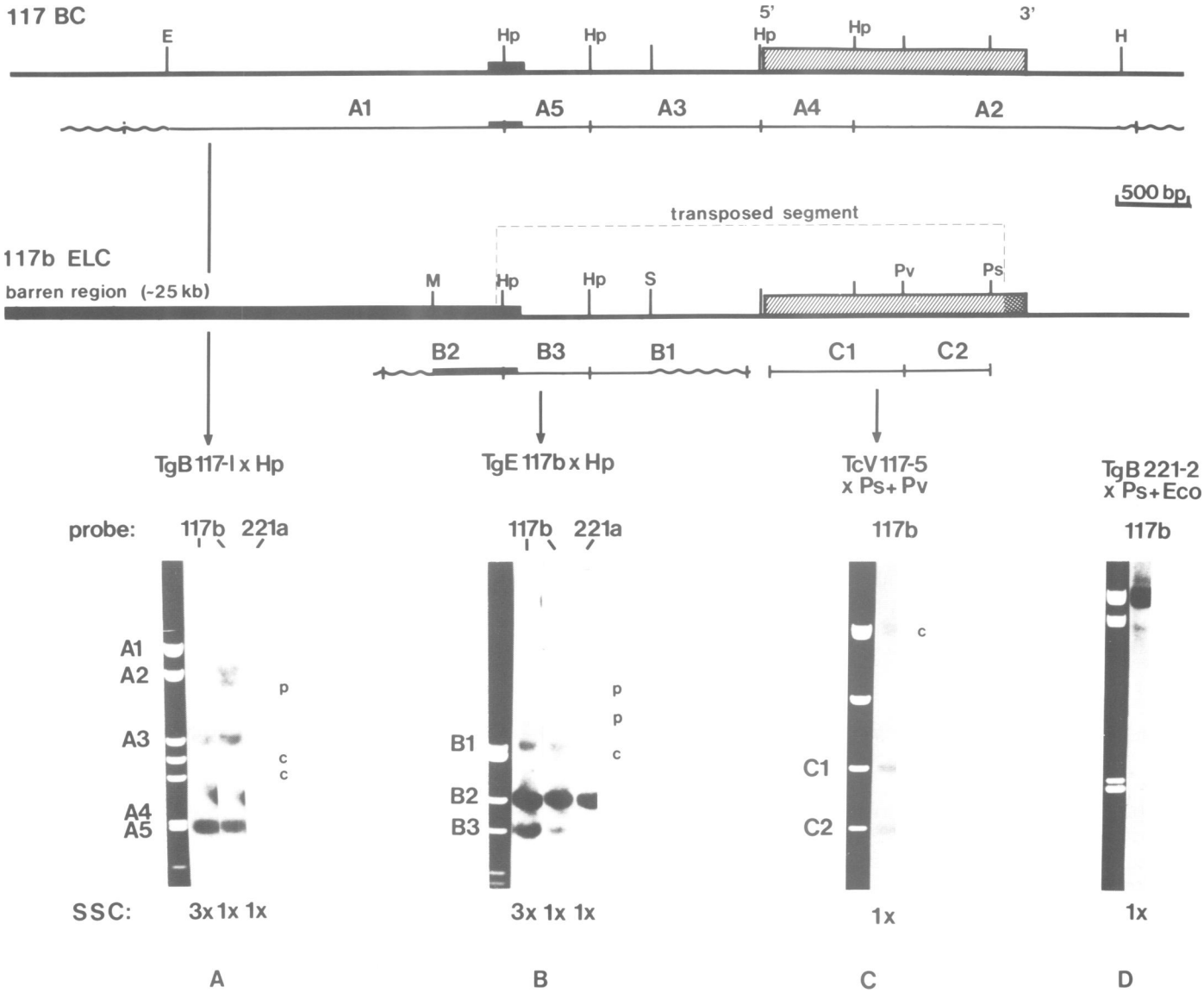

Fig. 5. Nascent RNA from the ELC gene for VSG 117b. The **top part** of the figure shows the origin of three DNA clones and restriction fragments of these clones that were used to analyse nascent RNA from the 117b ELC gene. Some restriction sites are identified in one of the two maps only. The direction of transcription is from left to right; the transposed gene segment is indicated. The thick lines indicate the presence of 70-bp repeats; the 70-bp repeat nature of the 25-kb barren region beyond the *Mbo*I site is speculative. The **lower part** shows the hybridization of labeled nascent RNA generated in isolated nuclei from variant 117b or variant 221a to blots of restricted cloned DNAs. The origin of the nascent RNA, the cloned DNA and the stringency of hybridization (at 65°C) is indicated in each panel. The relevant restriction fragments are identified by a code next to the ethidium bromide/u.v. pattern and in the top part of the figure. Some fragments contain vector sequences (wavy lines). **Panel D** shows the cross-hybridization of nascent RNA from variant 117b to a *Pst*I-*Eco*RI fragment upstream of the gene for VSG 221 (in clone TgB 221-2, see Bernards *et al.*, 1985), that contains ~50 70-bp repeats. In some lanes partial digestion products (p) occur. Cross-hybridization of nascent RNA to vector DNA (c) is occasionally seen. Abbreviations: Eco, *Eco*RI; Hp, *Hph*I; M, *Mbo*I or *Sau*3A; Ps, *Pst*I; Pv, *Pvu*II; S, *Sal*I; SSC, 0.15 M NaCl, 0.015 M Na citrate, pH 7.0.

representation of the area depicted in Figure 3 in nascent RNA.

Nascent RNA from the 117b ELC gene

Nuclei from trypanosomes expressing the 117b ELC gene or another VSG gene (221) were isolated and nascent RNAs were elongated *in vitro* in the presence of [α-^{32}P]UTP. The nascent RNAs were identified by hybridization to blots containing restriction digests of a cloned cDNA of 117a mRNA and of two cloned segments that represent the upstream sequences of the 117b ELC gene (see Figure 1). The results in Figure 5 show that the entire co-transposed segment and the adjacent upstream area of the expression site yield nascent transcripts at the same level as the gene itself or higher. Like the data obtained with stable RNA this result argues against transcription initiation close to the gene and indicates that transcription starts upstream of the transposed segment in the expression site.

There are quantitative anomalies in the data, however, that require some discussion. Transcripts from the co-transposed segment and the barren region are always over-represented, with some variation depending on the nuclei preparation. This is seen most clearly from the comparison of the 612-bp *Hph*I fragment that contains the 5′ end of the gene (fragment A4 in panel A) and the 583-bp *Hph*I fragment from the co-transposed segment (fragment A5 in panel A, fragment B3 in panel B). Whereas transcripts from the latter fragment are readily detected, the 5′ end of the gene seems much less represented. Although a lower

labeling of the 5′ end of the gene would be expected on the basis of the T-content of the non-coding strand (12% for the 5′ end of the gene and 25% for the 835-bp *Hph*I fragment from the co-transposed segment; Boothroyd *et al.*, 1981; Boothroyd and Cross, 1982), this alone cannot explain the observed labeling ratio. More likely, transcripts from the 583-bp *Hph*I fragment are over-represented in these experiments due to cross-hybridization of nascent RNAs with 70-bp repeat sequences. Indeed, the over-representation is even more prominent with the 705-bp *Hph*I fragment (fragment B2 in Figure 5) that contains six 70-bp repeats (see Figure 2), especially at reduced stringency (3 × SSC) (Figure 5, panels A and B). If the 25-kb barren region upstream of the 117b ELC largely consists of 70-bp repeats and if the whole barren region is transcribed, nascent RNA should contain an impressive amount of these sequences and at high specific activity due to the high U-content of typical 70-bp repeats containing RNAs (~30%). In support of this interpretation, nascent RNA from 117b nuclei (Figure 5, panel D), but not from nuclei of cultured (pro-cyclic) trypanosomes (not shown), hybridizes to 70-bp repeats upstream of the telomeric gene for VSG 221. Conversely, nascent RNA from trypanosomes that express VSG 221 hybridize to 70-bp repeats upstream of the 117b ELC, but not to other parts of the 117 gene (Figure 5, panels A and B). The most plausible interpretation of our 'run-on' analysis, therefore, is that transcription of the 117b ELC gene starts within the expression site, within or upstream of the barren region.

Discussion

Trypanosomes use ~10% of their genome to program antigenic variation (Van der Ploeg *et al.*, 1982a). Most of this DNA is used to encode the large repertoire of variant surface glycoproteins. In addition, there are telomeric expression sites that allow transcription of VSG genes once they become embedded in this DNA environment. We have analyzed stable and nascent RNA to determine how a translocated gene is activated in the expression site. Our results are compatible with transcription initiation upstream of the transposed segment, i.e., an activation mechanism analogous to promoter addition. In addition to a transcription initiation site, the expression site provides part of the transcription unit of the ELC gene. This part of the transcription unit has no known function and may be termed intron-like sequence until more information about the mechanism of discontinuous transcription becomes available. The processing products of this intron-like sequence are complex and very variable in abundance. In addition, they have the unusual feature of binding to oligo(dT) cellulose, probably due to the presence of a poly(A) tail, as already noted by Pays *et al.* (1982). Although nothing is known about polyadenylation in trypanosomes this result is unexpected, because it suggest that in trypanosomes poly(A) addition is not limited to the 3′ end of primary transcripts. Whether the presence of poly(A) tails on processing products is a general phenomenon in trypanosomes is not easily verified, because none of the non-VSG genes characterized thus far have introns (cf. Thomashow *et al.*, 1983; Seebeck *et al.*, 1983; Tschudi *et al.*, 1985; K. Osinga, W.Gibson and P.Borst, unpublished observations).

We briefly consider here other explanations for our results. We have not excluded that the transcripts from the expression site derive from other transcription units that immediately flank the expressed VSG gene. We consider this explanation unlikely for two reasons: first, we have demonstrated here the existence of stable transcripts crossing the boundary of the transposed segment. Second, we have recently found that transcripts of VSG genes and of upstream 70-bp repeats share an unusual resistance to 500 μg α-amanitin per ml, whereas the transcription of other protein coding genes in isolated nuclei is almost completely inhibited by 5 μg α-amanitin per ml (Kooter and Borst, 1984). This strongly indicates that the 70-bp repeat sequences upstream of the transposed segment in the expression site are part of the VSG gene transcription unit.

On the basis of the experiments presented here, the start of VSG gene transcription cannot be precisely located. The possibility even remains that transcription of VSG genes does not start at a defined site, but rather can be initiated in a region which encompasses both the 5′ barren region and the co-transposed segment. Our recent data on the transcription of the telomeric gene for surface antigen 221 argue against this explanation. The 221 gene is located permanently in a telomeric expression site with a short barren region which was readily crossed by genomic cloning (Bernards *et al.*, 1984, 1985). We have now analyzed 22 kb upstream of the start codon of this gene and find this whole region to belong to the VSG transcription unit on the basis of α-amanitin resistant transcription (Kooter and Borst, 1984; Cornelissen *et al.*, 1985; J.Kooter, R.Wagter and P.Borst, unpublished data). Furthermore, the level of transcription of the upstream area does not diminish with distance from the gene, indicating that transcription does not start randomly downstream from a defined site, but rather that the transcription unit of VSG genes is exceedingly large.

A second point that emerges from our analysis of the 117b ELC gene is the variability of crossover point at the 5′ end of the transposed segment. This variation can explain the length alterations of the barren region upstream of ELC genes (Michels *et al.*, 1983, see Figure 1). Since the barren region upstream of the ELC gene for VSG 118d is stable over >350 generations (Bernards *et al.*, 1983), it was proposed that changes in this DNA segment are linked to gene exchange in the expression site. Our data on the 5′ end of the transposed gene segment indicate that contraction and expansion of the barren region at least in part results from variation in the recombination point both in the expression site and in the incoming VSG gene segment. For instance, the telomeric gene for VSG 221 that contains as many as 50 copies of the 70-bp repeat (Bernards *et al.*, 1985) can potentially increase the length of the 5′ barren region by 3.5 kb. In agreement with the proposition that the 5′ barren region is (in part) composed of relics from previously expressed VSG genes, the sequence of the 5′ barren region in front of the 117a and 117b ELC genes, which have a completely different pedigree, is different (Figure 2). In addition, we have found a difference in the first *Mbo*I and *Dde*I sites in the barren region upstream of ELC genes in this expression site (not shown). The remarkable variability of the barren regions around ELC genes that was so baffling to early investigators of VSG genes can therefore now be satisfactorily explained by a combination of two unusual processes: the growth and contraction of telomeres (Bernards *et al.*, 1983) downstream of ELC genes, unrelated to gene switching; and the expansion and contraction of 70-bp repeats upstream of the ELC, associated with the exchange of genes in the expression site.

Materials and methods

Trypanosomes

The trypanosomes used in this study belong to strain 427 of *Trypanosoma brucei brucei*. Variant antigen types 221a and 118a have been described by Cross (1975)

and variant antigen type 117b by Michels *et al.* (1983). Parasites were grown as described by Fairlamb *et al.* (1978).

Isolation of DNA and genomic cloning

Nuclear DNA was isolated using the procedure described by Bernards *et al.* (1981) and digested with *Mbo*I and *Sal*I. The restriction fragments were ligated in *Bam*HI-*Sal*I digested, dephosphorylated pAT 153 DNA (Twigg and Sheratt, 1980). The ligate was used to transform *Escherichia coli* DHI (Hanahan, 1983) according to the protocol of Dr D.Hanahan for *E. coli* x 1776 reported by Maniatis *et al.* (1982). About 80 000 ampicillin resistant colonies were obtained. From this library six plasmids containing BC sequences and one ELC clone (TgE 117b) were isolated by screening with a 522-bp *Hinf*I fragment from the 5′ end of the transposed segment. Further manipulations with cloned DNA were carried out using standard procedures (cf. Maniatis *et al.*, 1982).

Nucleotide sequence analysis

DNA sequences were determined by the chemical degradation technique of Maxam and Gilbert (1980). Five modification reactions (G; G+A; C; C+T; A+C) were used to minimize ambiguities.

Preparation of ^{32}P-labeled single-stranded probes

DNA fragments used for mapping steady state transcripts were derived from TgE 117b, isolated from low melting point agarose and kinated using standard procedures (cf. Maniatis *et al.*, 1982). The 5′ end-labeled fragments were ethanol precipitated and resuspended in 60 μl H_2O, boiled for 2 min, followed by a quick chill on ice. DMSO was added to a final concentration of 30% and the mixture was immediately loaded onto a 0.3-mm thick polyacrylamide gel (5% acrylamide, 0.1% methylene bis-acrylamide). This was run overnight at 8 V/cm; after autoradiography the labeled DNA strands were eluted from the acrylamide by soaking the gel slices in 1 − 1.5 ml of 0.5 M ammonium acetate, 1 mM EDTA and 0.5% SDS at 37°C; the eluate was phenol extracted and ethanol precipitated.

RNA blotting

Infected rat blood was mixed with 10 volumes of ice-cold 3 M LiCl/6 M urea immediately after cardiac puncture and homogenized in a Sorval Omnimix at full speed for 1 min. RNA was precipitated for 5 h on ice and collected by centrifugation in a Sorvall SS-34 rotor at 14 000 r.p.m. at 0°C. DNase I digestion and isolation of poly(A)$^+$ RNA were as described earlier (Van der Ploeg *et al.*, 1982c). RNase A treatment of 117 RNA (5 μg) was for 5 min at room temperature with 2 ng enzyme that had been treated for 10 min at 80°C to inactivate deoxyribonuclease. RNA blots were prepared by the glyoxal method described by Thomas (1980).

S1 nuclease and Exo VII protection experiments

Total RNA (50 − 100 μg) was dissolved in 30 μl hybridization buffer as described by Berk and Sharp (1977) containing ^{32}P-5′ end-labeled single-stranded DNA and was hybridized overnight either at 42°C or 38°C. The samples were split in two and digested with S1 nuclease (Sigma) or Exo VII (BRL). The S1 digestion was done in a final volume of 300 μl (500 U/ml S1 nuclease) at 30°C for 1 h; the Exo VII digestion was started by adding 300 μl ice-cold 30 mM KCl, 10 mM Tris-HCl (pH 7.5), 10 mM EDTA containing three units Exo VII and was continued for 4 − 6 h at 30°C. Both S1 and Exo VII digestions were terminated by adding 100 μl of 4.0 M NH_4 acetate and 0.1 mM EDTA. The products were recovered by ethanol precipitation and resolved either by alkaline agarose or polyacrylamide gel electrophoresis and visualized by autoradiography using intensifying screens. The agarose gels were dried after the DNA was immobilized by soaking the gel in 10% TCA for 30 min as described by Maniatis *et al.* (1982).

Analysis of nascent RNA

All procedures used for the analysis of nascent RNA are described by Kooter *et al.* (1984). Briefly, nuclei were isolated from infected rat blood as quickly as possible using a Standsted Cell Disrupter and used either immediately or after storage at −70°C. *In vitro* elongation of nascent RNAs was allowed to proceed for 4 min in the presence of labeled UTP and terminated by the addition of DNase I or by incubation at 70°C followed by DNase I digestion. Both procedures for termination gave the same result. We have also varied the total UTP concentration between 0.5 and 10 μM and did not find qualitative differences in the results. Nascent RNAs were hybridized to blots of cloned DNA in the presence of dextran sulphate (10%) and 100 μg tRNA per ml for 18 h, followed by washing and autoradiography at −70°C.

Acknowledgements

We are indebted to Mr H.J.G.Veerman for technical assistance and to Mrs B. van Houten for preparing the manuscript. This work was supported in part 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

Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.
Berk,A.J. and Sharp,P.A. (1978a) *Proc. Natl. Acad. Sci. USA*, **75**, 1274-1278.
Berk,A.J. and Sharp,P.A. (1978b) *Cell*, **14**, 695-711.
Bernards,A., Van der Ploeg,L.H.T., Frasch,A.C.C., Borst,P., Boothroyd,J.C., Coleman,S. and Cross,G.A.M. (1981) *Cell*, **27**, 497-505.
Bernards,A., Michels,P.A.M., Lincke,C.R. and Borst,P. (1983) *Nature*, **303**, 592-597.
Bernards,A., De Lange,T., Michels,P.A.M., Huisman,M.J. and Borst,P. (1984) *Cell*, **36**, 163-170.
Bernards,A. (1985) *Biochim. Biophys. Acta*, **824**, 1-15.
Bernards,A., Kooter,J.M. and Borst,P. (1985) *Mol. Cell. Biol.*, **5**, 545-553.
Boothroyd,J.C. (1985) *Annu. Rev. Microbiol.*, in press.
Boothroyd,J.C., Paynter,C.A., Cross,G.A.M., Bernards,A. and Borst,P. (1981) *Nucleic Acids Res.*, **9**, 4735-4743.
Boothroyd,J.C. and Cross,G.A.M. (1982) *Gene*, **20**, 281-289.
Borst,P., Bernards,A., Van der Ploeg,L.H.T., Michels,P.A.M., Liu,A.Y.C., De Lange,T. and Kooter,J.M. (1983) *Eur. J. Biochem.*, **137**, 383-389.
Campbell,D.A., Van Bree,M. and Boothroyd,J.C. (1984a) *Nucleic Acids Res.*, **12**, 2759-2774.
Campbell,D.A., Thornton,D.A. and Boothroyd,J.C. (1984b) *Nature*, **311**, 350-355.
Cornelissen,A.W.C.A., Johnson,P.J., Kooter,J.M., Van der Ploeg,L.H.T. and Borst,P. (1985) *Cell*, **41**, 825-832.
Cross,G.A.M. (1975) *Parasitology*, **71**, 393-417.
Cross,G.A.M. (1978) *Proc. R. Soc. Lond. Ser. B*, **202**, 55-72.
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.
De Lange,T., Michels,P.A.M., Veerman,H.J.G., Cornelissen,A.W.C.A. and Borst,P. (1984a) *Nucleic Acids Res.*, **12**, 3777-3790.
De Lange,T., Berkvens,T.M., Veerman,H.J.G., Frasch,A.C.C., Barry,J.D. and Borst,P. (1984b) *Nucleic Acids Res.*, **12**, 4431-4443.
Fairlamb,A.H., Weislogel,P.O., Hoeijmakers,J.H.J. and Borst,P. (1978) *J. Cell. Biol.*, **76**, 293-309.
Hanahan,D. (1983) *J. Mol. Biol.*, **166**, 557-580.
Hoeijmakers,J.H.J., Frasch,A.C., Bernards,A., Borst,P. and Cross,G.A.M. (1980a) *Nature*, **284**, 78-80.
Hoeijmakers,J.H.J., Borst,P., Van den Burg,J., Weissmann,C. and Cross,G.A.M. (1980b) *Gene*, **8**, 391-417.
Kooter,J.M. and Borst,P. (1984) *Nucleic Acids Res.*, **12**, 9457-9472.
Kooter,J.M., De Lange,T. and Borst,P. (1984) *EMBO J.*, **3**, 2387-2392.
Laird,P.W., Kooter,J.M., Loosbroek,N. and Borst,P. (1985) *Nucleic Acids Res.*, **13**, 4253-4266.
Laurent,M., Pays,E., Van der Werf,A., Aerts,D., Magnus,E., Van Meirvenne,N. and Steinert,M. (1984) *Nucleic Acids Res.*, **12**, 8319-8329.
Liu,A.Y.C., Van der Ploeg,L.H.T., Rijsewijk,F.A.M. and Borst,P. (1983) *J. Mol. Biol.*, **167**, 57-75.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Michels,P.A.M. (1984) *Oxford Surveys on Eucaryotic Genes*, Vol. **1**, 145-168.
Michels,P.A.M., Bernards,A., Van der Ploeg,L.H.T. and Borst,P. (1982) *Nucleic Acids Res.*, **10**, 2353-2366.
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. and Van Boom,J.H. (1983) *J. Mol. Biol.*, **166**, 537-556.
Milhausen,M., Nelson,R.G., Sather,S., Selkirk,M. and Agabian,N. (1984) *Cell*, **38**, 721-729.
Myler,P., Nelson,R.G., Agabian,N. and Weber,H. (1984) *Nature*, **309**, 282-285.
Nelson,R.G., Parsons,M., Barr,P.J., Stuart,K., Selkirk,M. and Agabian,N. (1983) *Cell*, **34**, 901-909.
Nelson,R.G., Parsons,M., Selkirk,M., Newport,G., Barr,P.J. and Agabian,N. (1984) *Nature*, **308**, 665-667.
Parsons,M., Nelson,R.G. and Agabian,N. (1984a) *Immunol. Today*, **5**, 43-50.
Parsons,M., Nelson,R.G., Watkins,K.P. and Agabian,N. (1984b) *Cell*, **38**, 309-316.
Pays,E., Lheureux,M. and Steinert,M. (1981) *Nature*, **292**, 265-267.
Pays,E., Lheureux,M. and Steinert,M. (1982) *Nucleic Acids Res.*, **10**, 3149-3163.
Pays,E., Van Assel,S., Laurent,M., Dero,B., Michiels,F., Kronenberger,P., Matthijssens,G., Van Meirvenne,N., Le Ray,D. and Steinert,M. (1983a) *Cell*, **34**, 359-369.
Pays,E., Van Assel,I., Laurent,M., Darville,M., Vervoort,T., Van Meirvenne,N. and Steinert,M. (1983b) *Cell*, **34**, 371-381.
Seebeck,T., Whittaker,P.A., Imboden,M.A., Hardman,N. and Braun,R. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4634-4638.

Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201-5205.
Thomashow,L.S., Milhausen,M., Rutter,W.J. and Agabian,N. (1983) *Cell*, **32**, 35-43.
Tschudi,C., Young,A.S., Ruben,L., Patton,C.L. and Richards,F.F. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3998-4002.
Twigg,A.J. and Sheratt,D. (1980) *Nature*, **283**, 216-220.
Van der Ploeg,L.H.T., Valerio,D., De Lange,T., Bernards,A., Borst,P. and Grosveld,F.G. (1982a) *Nucleic Acids Res.*, **10**, 5905-5923.
Van der Ploeg,L.H.T., Bernards,A., Rijsewijk,F.A.M. and Borst,P. (1982b) *Nucleic Acids Res.*, **10**, 593-609.
Van der Ploeg,L.H.T., Liu,A.Y.C., Michels,P.A.M., De Lange,T., Borst,P., Majumber,H.K., Weber,H., Veeneman,G.H. and Van Boom,J. (1982c) *Nucleic Acids Res.*, **10**, 3591-3604.
Van der Ploeg,L.H.T., Schwartz,D.C., Cantor,C.R. and Borst,P. (1984a) *Cell*, **37**, 77-84.
Van der Ploeg,L.H.T., Cornelissen,A.W.C.A., Michels,P.A.M. and Borst,P. (1984b) *Cell*, **39**, 213-221.
Vickerman,K. (1978) *Nature*, **273**, 613-617.
Williams,R.O., Young,J.R. and Majiwa,P.A.O. (1982) *Nature*, **299**, 417-421.
Young,J.R., Donelson,J.E., Majiwa,P.A.O., Shapiro,S.Z. and Wiliams,R.O. (1982) *Nucleic Acids Res.*, **10**, 803-819.

Received on 11 January 1985; revised on 28 August 1985