The EMBO Journal vol.9 no.9 pp.2791 - 2801, 1990

The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*

Joost C.B.M.Zomerdijk, Marc Ouellette, Anneloor L.M.A.ten Asbroek, Rudo Kieft, Angeline M.M.Bommer, Christine E.Clayton[1] and Piet Borst

Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands and [1]The Rockefeller University, York Avenue, New York, NY 10021, USA

Communicated by P.Borst

The variant-specific surface glycoprotein (VSG) gene 221 of *Trypanosoma brucei* is transcribed as part of a 60 kb expression site (ES). We have identified the promoter controlling this multigene transcription unit by the use of 221 chromosome-enriched DNA libraries and VSG gene 221 expression site specific transcripts. The start of transcription was determined by hybridization and RNase protection analysis of nascent RNA. The 5′ ends of the major transcripts coming from the initiation region map at nucleotide sequences that do not strongly resemble rRNA transcriptional starts even though the transcripts are synthesized by an RNA polymerase highly resistant to α-amanitin. The cloned VSG gene 221 ES transcription initiation region promotes high CAT gene expression, when reintroduced by electroporation into *T.brucei*. We show that the activity of this expression site is controlled at or near transcription initiation in bloodstream trypanosomes. The 221 ES is inactivated without any sequence alteration within 1.4 kb of the transcription start site. This excludes mechanisms of promoter inactivation involving DNA rearrangements in the vicinity of the transcription start site, e.g. promoter inversion or conversion.

Key words: promoter/*Trypanosoma brucei*/transcription/transfection/variant surface glycoprotein

Introduction

African trypanosomes, such as *Trypanosoma brucei*, can evade the host immune response by altering the antigenic composition of their surface coat, which consists of a single protein species known as the variant-specific surface glycoprotein (VSG) (Cross, 1975; Vickerman, 1978). Switching from one coat to the next occurs at low frequency and is spontaneous (Lamont *et al.*, 1986). Each VSG is encoded by a separate gene and there are ~1000 VSG genes (and pseudogenes) per trypanosome genome (Capbern *et al.*, 1977; Van der Ploeg *et al.*, 1982). An expressed VSG gene is invariably localized in a telomeric expression site (ES) (De Lange and Borst, 1982). The ES consists of a single large transcription unit, that includes several expression site-associated genes (ESAGs) besides the VSG gene (Cully *et al.*, 1985; Johnson *et al.*, 1987; Kooter *et al.*, 1987). Transcription of the entire ES is resistant to α-amanitin, unlike the transcription of most of the other protein-coding genes in *T.brucei* (Kooter and Borst, 1984; Laird *et al.*, 1985; Kooter *et al.*, 1987). Hybridization studies suggest that *T.brucei* 427 may contain as many as 20 different ESs (Cully *et al.*, 1985); usually only one is operational at any time.

Antigenic switching can occur either by replacement of the VSG gene in the active ES by another previously silent or basic copy VSG gene (Hoeijmakers *et al.*, 1980; Bernards *et al.*, 1981; De Lange and Borst, 1982; Pays *et al.*, 1985), or by activation of another ES and inactivation of the formerly active one without detectable DNA rearrangements (Williams *et al.*, 1979; Buck *et al.*, 1984; Michels *et al.*, 1984; Van der Ploeg *et al.*, 1984). Gene replacement occurs by recombination or gene conversion (reviewed by Boothroyd, 1985; Donelson and Rice-Ficht, 1985; Borst, 1986; Borst and Greaves, 1987; Pays and Steinert, 1988).

How a trypanosome switches from one ES to another is not yet known. The observation that two ESs can be active simultaneously (Cornelissen *et al.*, 1985; Baltz *et al.*, 1986) suggests, however, that there is no cross-talk between ESs (i.e. not a single mobile promoter, enhancer or transcription terminator) and that each ES is activated and inactivated independently of other ESs (Cornelissen *et al.*, 1985). To test whether this stochastic (in)activation process involves DNA rearrangements near the promoter, as previously proposed (Cornelissen *et al.*, 1985; Borst and Greaves, 1987), we have cloned and analysed the promoter area of the VSG gene 221 ES in an active and in an inactive configuration.

Results

Cloning the 221 expression site promoter region

In previous studies we have characterized the 221 ES, occupied by the VSG gene 221 in trypanosome variant 221a of *T.brucei* strain 427 (Bernards *et al.*, 1984; Johnson *et al.*, 1987; Kooter *et al.*, 1987). By chromosome walking we isolated DNA clones that represent this ES and showed that the ES consists of a single 60 kb transcription unit, as illustrated in Figure 1A. DNA probes from the upstream part of this ES hybridized to at least 20 fragments in Southern blots, even at the highest stringencies. It was therefore difficult to establish that clones representing the 221 ES were actually derived from that ES. Indeed, some clones were shown not to be (Kooter *et al.*, 1987) and this is illustrated in Figure 1B. In this pulsed field gradient (PFG) gel, the 221 chromosome is marked in lane 2 by hybridization to a single-copy probe from the VSG gene 221, present only once in this trypanosome strain (Bernards *et al.*, 1984). As this ~4 Mb 221 chromosome migrates just ahead of the compression zone at the limit of resolution of the PFG gel, it is present in low amounts relative to other large chromosomes (cf. lane 1). Lanes 3 and 4 illustrate that clone pTg221.12, which contains a sequence homologous to the

promoter area of the 221 ES, is not derived from the 221 chromosome. Whereas probe a (lane 3) from this clone hybridizes to the 221 chromosome (and to most other chromosomes, except minichromosomes), probe b (lane 4) does not, even though it also recognizes the other chromosomes. Probe a of pTg221.12 is therefore suitable for screening libraries for promoter regions.

To address the question of what happens at the 221 ES promoter when it is switched off, it was necessary to obtain clones from the 221 ES chromosome itself and not one representing it, i.e. the clone pTg221.12. To this end we made a chromosome-specific library with DNA from PFG gels in which the chromosome of ~4 Mb that contains the 221 ES is relatively well separated from its neighbours (Figure 1B, lane 1). The promoter clones were identified with probe a and were counter-selected with probe b, to avoid isolating clones not originating from the 221 ES (see Figure 1A). This resulted in the set of overlapping phage λ clones used to construct the map in Figure 2. We verified the integrity of the clones by comigration of insert fragments with nuclear DNA digests (data not shown). In contrast to the promoter region of clone pTg221.12 (Figure 1A), the 221 promoter region is flanked at its 5′ end by a region of imperfect direct repeats of 50 bp (Figure 2), devoid of restriction enzyme recognition sites, except for *Rsa*I. These repeats are highly unstable in bacteriophage λ and cause poor growth of the phages, but mapping of the phage inserts of 9.5 and 12.0 kb in Figure 2, suggests that the 50 bp repeat region covers at least 10 kb. The repeats are present in many copies in the genome and, like other sequences in the promoter area, they are present in most chromosomes, except for the minichromosomes (Figure 2, probe 1). As probes of various regions along the promoter area show slightly different hybridization patterns with the chromosomes fractionated by PFGE (Figure 2, lanes 1–8), we infer that promoter sequences of different ESs are highly similar but not identical (compare for example lanes 2 and 3, Figure 2).

Evidence that the ES promoter clones are from the 221 ES

To verify whether the map in Figure 2 represents the 221 ES, rather than an ES from another chromosome contaminating the 221 chromosome preparation, we used the (semi) unique sequences associated with the ESAG-X genes (see maps in Figures 1A and 2). Although the ESAG-X genes from different chromosomes are very similar (Kooter *et al.*, 1987), we have found a sequence of 32 bp in this gene, indicated by the hatched boxes in Figure 2, that varies substantially between clones derived from different chromosomes, as illustrated in Figure 3. Our analysis of the promoter area revealed that there are three genes in the 221 ES that hybridize to an ESAG-X 5′ end probe (Figure 2, probe 6), rather than the two detected previously (Kooter *et al.*, 1987). We have named these ESAG-X, X′ and X″. Figure 3 shows that the hypervariable regions of ESAG-X, X′ and X″, as they were cloned in phage λ, are identical to that of the ESAG-X cDNA from variant 221a. We conclude that the phage λ clones obtained from chromosome 221 enriched DNA libraries are truly from the 221 ES.

Transcripts from the promoter area of the 221 ES

To map the start of transcription of the 221 ES roughly we hybridized nascent RNA from nuclear run-on assays with

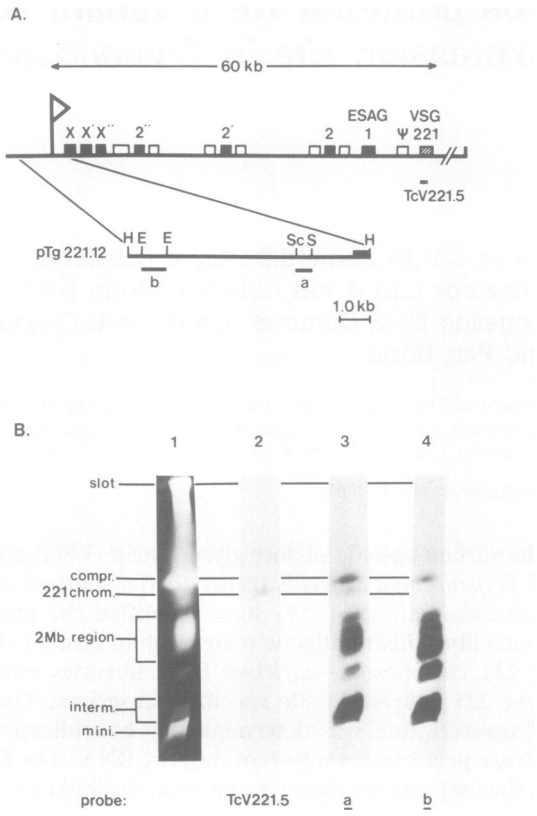

Fig. 1. Identification of the 221 ES in chromosome-sized DNA molecules of *T.brucei* 427 size-fractionated by pulsed field gradient gel electrophoresis. **A.** The genomic organization of the telomeric 221 ES and a physical map of the insert of the recombinant plasmid pTg221.12 (Kooter *et al.*, 1987). This clone represents the region far upstream in the 221 chromosome telomere, and contains the putative promoter of an ES, not however that of the 221 ES (indicated by the flag). Fragments used as probes in B are indicated below the map. Restriction sites: E, *Eco*RI; H, *Hin*dIII; S, *Sal*I; Sc, *Sca*I. **B. Lane 1** shows the ethidium bromide stained chromosomes fractionated by PFGE (see Materials and methods for conditions). The different chromosomes are indicated along the lane. The position of the 221 chromosome is indicated by 221 chrom.; the chromosomes in the size range of 2000 kb, the intermediate chromosomes and the minichromosomes are indicated by 2 Mb region, interm. and mini. respectively. Very large chromosomes are not well separated and migrate in a compression zone, indicated by compr.. **Lane 2** shows the hybridization to a VSG gene 221 probe (5′ 590 bp *Pst*I fragment of pTcV221.5, Bernards *et al.*, 1984) which marks the position of the chromosome containing the 221 ES. The hybridization of fragment a of pTg221.12 shows it to cross-hybridize with the 221 chromosome (**lane 3**) whereas fragment b does not (**lane 4**).

the cloned promoter area. The results are shown in Figure 4. Fragment 1, which overlaps with the 5′ end of ESAG-X, and fragment 7 are clearly positive in this hybridization analysis. Nascent RNA also hybridizes to fragment 8, though weakly. This reduced hybridization is not due to the relatively small size of this fragment (0.4 kb) since the 1.2 kb fragment 2 hybridizes to the same extent. Fragments 3, 4, 5 and 6 are transcriptionally silent in a run-on assay. These hybridization results were essentially the same for run-ons in the absence of α-amanitin and with [α-^{32}P]UTP instead of [α-^{32}P]GTP (data not shown). To exclude blotting and hybridization artefacts, the same blot was rehybridized with probes covering the entire promoter area. All fragments hybridized in proportion to their length and concentration

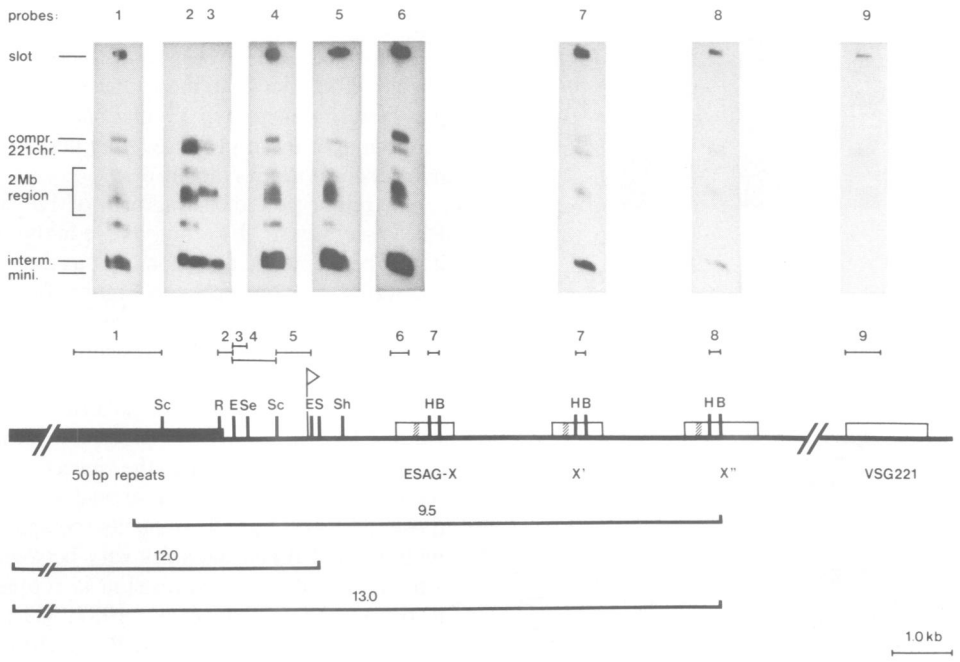

Fig. 2. Physical map of the 221 ES promoter area and repetitiveness of segments of this area within the *T.brucei* 427 genome. Physical map of the promoter area 57 kb upstream of the VSG gene 221 in the 221 ES. Boxes indicate expression site associated genes (ESAGs) and the VSG gene. The map was constructed with the phages isolated with probe a (see Figure 1B) from 221 chromosome-enriched DNA libraries of *T.brucei* variants 221a and 221aR12. The inserts (9.5 and 12.0 kb from variant 221a; 13.0 kb from 221aR12) are shown underneath the map. The hatched boxes in the ESAG-Xs represent the position of the hypervariable region within these genes. The black bar symbolizes the 50 bp direct repeats. Relevant restriction sites are indicated and the probes derived from various parts of the promoter area are numbered. Their corresponding hybridization patterns on PFG blots (as in Figure 1) are shown above the map, to illustrate the repetitiveness of this region within the *T.brucei* genome. Note the absence of hybridization of probe 8 to the compression zone. The flag (below fragment 5) marks the approximate position of the start of transcription. Restriction sites: B, *Bam*HI; E, *Eco*RV; H, *Hin*dIII; S, *Sal*I; Sc, *Sca*I; Se, *Spe*I; Sh, *Sph*I; R, *Rsa*I (every 50 bp repeat contains two *Rsa*I sites, but only one is indicated within the 50 bp repeats region).

(data not shown). We found no cross-hybridization of promoter region sequences to rRNA genes (data not shown). Hence, these data suggest that the start of transcription is close to and upstream of the *Eco*RV site within the *Sca*I−*Eco*RV fragment (fragment 8). The start is preceded by a gap in transcription that covers at least the 50 bp repeat area and the 1.4 kb between the *Eco*RV site and the 50 bp repeats. These results confirm the tentative identification of a transcription start in this area by Kooter *et al.* (1987) and Johnson *et al.* (1987), confirmed for another ES by Pays *et al.* (1989b).

Transcription initiation in the 221 ES

We mapped the start of transcription more precisely in an RNase protection assay. The probes are depicted in Figure 5. Four of these (a, b, c and d) are complementary (antisense) with respect to the direction of transcription of downstream situated ESAGs and were used to map the 5′ end of the presumed large multicistronic transcript. Since these probes do not detect abundant transcripts on RNA (Northern) blots (data not shown), we used 30 μg of total RNA from buffy coat trypanosomes to detect precursor RNA molecules. As the 221 ES transcription gradually fades out, particularly in downstream parts, upon a lengthy trypanosome isolation procedure (Kooter *et al.*, 1987), we prepared RNA immediately after bleeding the rat. RNA isolated from total blood or from buffy coat trypanosomes gave similar results. Here we present only the experiments with total RNA from buffy coat trypanosomes.

Hybridization with probes a and c results in an abundant

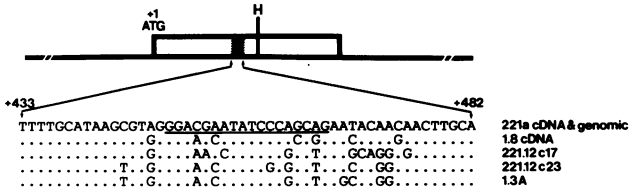

Fig. 3. The hypervariable region sequence of ESAG-X from the 221 ES is identical to the corresponding sequence of ESAG-X cDNA from *T.brucei* variant 221a. The hypervariable region is schematically indicated by a hatched box in ESAG-X and a blow up of this particular region is shown underneath. The sequence of ESAG-X″ cDNA (221a cDNA) is identical to the hypervariable region as it is present in the phages overlapping the promoter region and ESAG-X, X′ and X″ from the 221 ES in variants 221a and 221aR12 (221a genomic). Other ESAG-X hypervariable region sequences that we obtained are: the ESAG-X cDNA sequence from a *T.brucei* variant expressing VSG gene 1.8 (1.8 cDNA), from an unknown ES, and the pTg221.12-like clones isolated from a *T.brucei* 427 genomic library (221.12 c17 and c23). The sequence in ESAG-7 and 6 is also depicted (1.3 A; Pays *et al.*, 1989b). Only nucleotides which differ from the 221 ES genomic sequence are indicated. Numbers are relative to the adenine in the first translation initiation codon ATG in ESAG-X and ESAG-6 and 7 (Pays *et al.*, 1989b). Restriction site: H, *Hin*dIII.

fragment of 89 nt (Figure 5B, lanes 2, 4 and 10, 12 respectively). This protected fragment is lengthened by exactly 185 nt, the distance between the *Sal*I and *Sph*I site, when probe b is used (lanes 6 and 8). From this we conclude that the start of the corresponding transcript is located 85 nt upstream of the *Sal*I site. In addition to this abundant

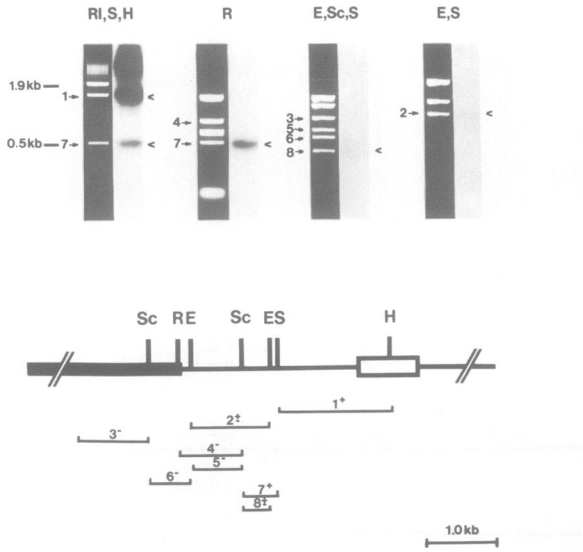

Fig. 4. Transcription of the 221 ES promoter area in isolated nuclei of *T.brucei* variant 221a. A restriction map of the 221 ES promoter area. The filled bar indicates the 50 bp repeats array, the open box the gene closest to the promoter, ESAG-X. Depicted under the map are the fragments as they are indicated at the left of each ethidium bromide stained gel lane containing different digests of subclones from the 221 ES promoter area. The left-most panel (RI, S, H digest) contains a subclone encompassing fragments 7 and 1 in the *Sma*I–*Hin*dIII sites of pGEM3. For the other digests a subclone containing fragments 3, 6, 5 and 7 was used. The 5′ end of this subclone is a *Sal*I site from the EMBL3 phage λ arm that is close to the *Bam*HI site used to clone the *Sau*3AI partially digested 221 chromosomal DNA. The *Sau*3AI site is located 1.0 kb upstream of the *Sca*I site in the 50 bp repeat region. The right-hand sides of each panel are Southern blots of these digests hybridized with nascent RNA synthesized in isolated nuclei of variant 221a in the presence of [α-^{32}P]GTP and 1 mg/ml α-amanitin. This α-amanitin concentration reduced mini-exon gene transcription by 95% (not shown). Hybridizing fragments are indicated by arrowheads at the right of these blots. The intensity of the hybridization signals are: +, strong; ±, weak; −, none. Restriction sites: E, *Eco*RV; H, *Hin*dIII; RI, *Eco*RI (polylinker site); S, *Sal*I; Sc, *Sca*I; R, *Rsa*I.

transcript, we observe a minor transcript of 127 nt with probe a (lanes 2 and 4), a 313 nt transcript with probe b (lanes 6 and 8, open triangle b) and a full-length protected fragment of 96 nt with probe c (lanes 10 and 12). This suggests that there is a transcript that overlaps with the abundant one but for an ~38 nt extension at its 5′ end. These are genuine transcripts because they are not present in the tRNA control hybridizations (lanes 3, 7, 11 and 15). Moreover, the level of the protected fragments is not strongly dependent on the hybridization temperature and the fragments are therefore not likely to result from aspecific hybridization (compare for example lanes 2 and 4). With a sense probe (e), covering the region from the *Sca*I to the *Sph*I site, no transcripts were detected (lanes 14 and 16).

The nature of the RNA polymerase that synthesizes these transcripts was examined in an RNase protection experiment in which we hybridized a non-labelled SP6 transcript (Figure 5A, probe d) with nascent labelled RNA synthesized in isolated nuclei of *T.brucei* variant 221a. Lanes 5 and 10 of Figure 5C show that the RNA molecules are heterogeneous in size, as expected. In this nascent RNA, probe d protects the same transcript of 89 nt that is also abundantly present in total RNA of *T.brucei* variant 221a (see Figure 5B). The synthesis of this transcript is completely resistant to high concentrations of α-amanitin (Figure 5C, compare lanes 1 and 2 with 6 and 7; filled triangle). At reduced hybridization stringency (55°C instead of 65°C; at 75°C no hybrids were formed, see lanes 3 and 8) additional transcripts are visible. The fact that these are synthesized by an RNA polymerase sensitive to a high concentration of α-amanitin suggests that these protected fragments result from cross-hybridization (lanes 1 and 6).

The minor protected fragment of 127 nt present in total RNA (see Figure 5B) is not visible in the nascent RNA, but it may be below the detection limit as the signal for the 89 nt transcript is much higher in Figure 5B than in Figure 5C. The nature of the RNA polymerase that synthesizes the 127 nt transcript therefore remains unknown.

We also performed a run-on analysis in combination with RNase protection with nuclei isolated from procyclic (culture form, CF) *T.brucei*, the phenotype to which trypanosomes transform upon entering the insect vector. The RNA molecules synthesized in these nuclei tend to be larger than those in bloodstream form nuclei (compare lanes 15 and 20 with 5 and 10) and the sizes vary between nuclei batches. Although VSG gene expression is repressed in procyclic trypanosomes (Kooter *et al.*, 1987; Murphy *et al.*, 1987), we detect in nascent procyclic RNA a low level of the same 89 nt transcript as was observed in bloodstream form *T.brucei* variant 221a (Figure 5C, lanes 12 and 2, respectively). This transcript is also synthesized by an RNA polymerase resistant to 1 mg/ml α-amanitin in procyclics (lane 17). In addition, we observed the same presumed cross-hybridization (lane 11) of transcripts that are synthesized by an RNA polymerase sensitive to α-amanitin (lane 16). We return to these results in the Discussion.

In conclusion, we have identified two ES promoter transcripts with their 5′ ends located around the *Eco*RV site in the 221 ES promoter area (Figures 5A and 7).

Promoter activity, as monitored by CAT assays, of the cloned VSG gene 221 ES transcription initiation region in procyclic trypanosomes

To study transcription and RNA splicing *in vivo*, Bellofatto and Cross (1989) recently developed a reproducible assay for DNA-mediated transfection in the trypanosomatid protozoan *Leptomonas seymouri*. The assay was based on expression of the *Escherichia coli* chloramphenicol acetyl-transferase (CAT) gene flanked by *Leptomonas* transcription signals.

To evaluate promoter activity of the cloned VSG gene 221 ES transcription initiation region *in vivo*, we constructed vectors in which the promoter region is fused to the CAT gene. The CAT gene in these constructs is flanked at its 5′ end by a 3′ splice site derived from either the VSG gene 118 (Liu *et al.*, 1983) or the procyclic acidic-repetitive protein (*parp*) gene A (Clayton *et al.*, 1990). These constructs (pRK4+ and pRK8+; see Figure 6 and Materials and methods) were introduced into procyclic *T.brucei* by a modified electroporation procedure (Bellofatto and Cross, 1989). Using a phase extraction assay for CAT activity (Seed and Sheen, 1988), followed by TLC, we detected 1- and 3-butyryl-[^{14}C]chloramphenicol synthesized in extracts made from the procyclics 20 h after electroporation (Figure 6). The promoterless constructs (pJZ4 and pJF6dSca), in which the CAT gene still contains the 3′ splice site fused to it, did not show detectable levels of CAT activity in this assay, implying that the promoter activity was not in plasmid

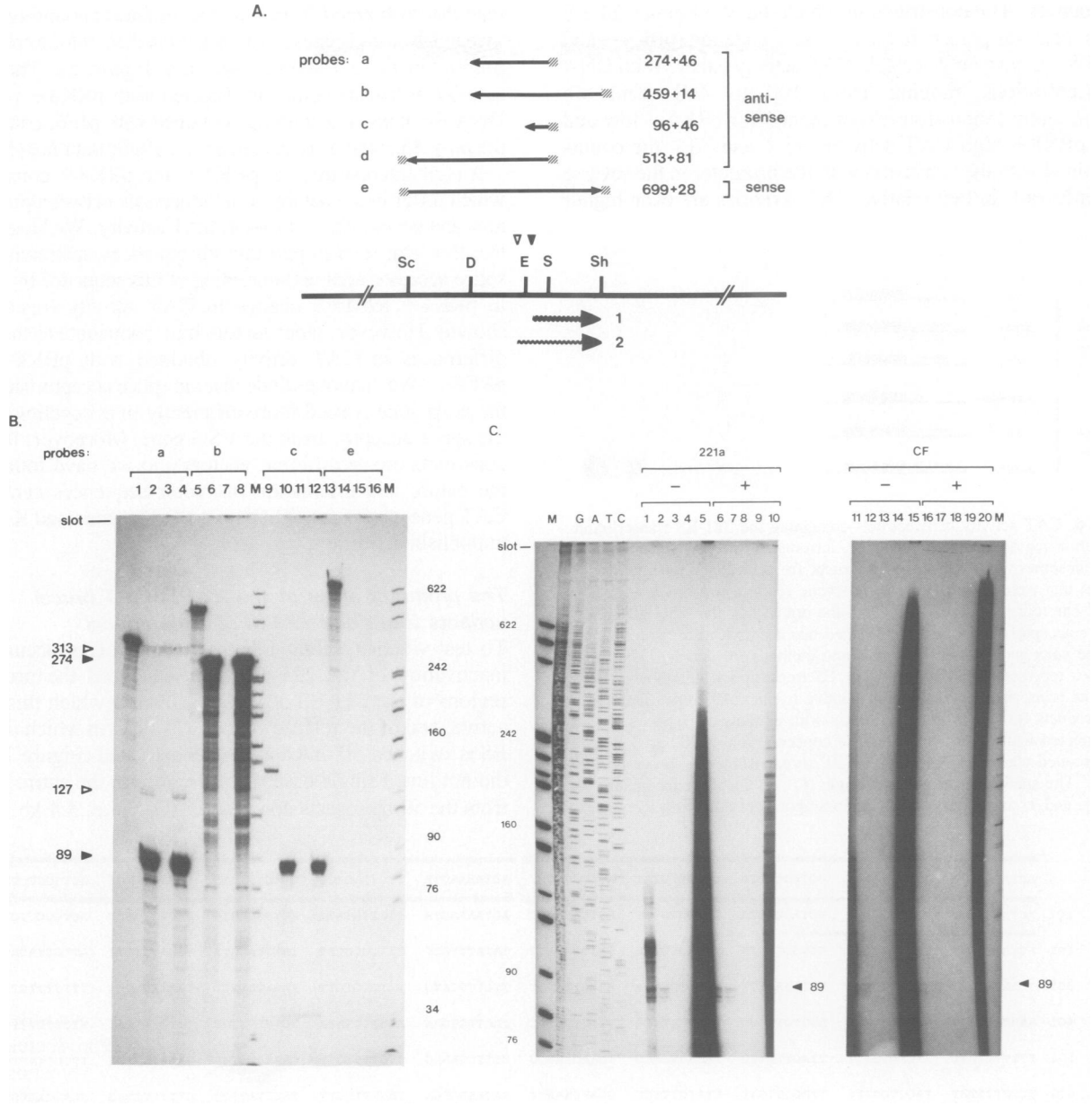

Fig. 5. RNase protection to map the start of transcription in the 221 ES. Total trypanosome RNA and nascent RNA synthesized in isolated nuclei were hybridized with various run-off transcripts synthesized *in vitro* for 16 h in 80% formamide (see Materials and methods) at the temperatures specified for each lane. The hybrids were subsequently digested with RNase A and T1, size-fractionated in a 6% polyacrylamide, 7 M urea gel and visualized by autoradiography. M (marker) is end-labelled pAT153 DNA digested with *Hpa*II. The sizes of the fragments (in bp) are indicated alongside the gel. **A.** Schematic representation of the sense and antisense transcripts used to map the start of transcription in the promoter area of the 221 ES. Above the map the SP6 and T7 transcripts synthesized *in vitro* are depicted with their length in nucleotides (nt) of trypanosome (straight line) and pGEM3 vector polylinker (hatched box) sequences. Antisense probes are: *Sal*I−*Dra*I run-off SP6 transcript of 320 nt (a); *Sph*I−*Dra*I run-off SP6 transcript of 473 nt (b); *Sal*I−*Eco*RV run-off SP6 transcript of 142 nt (c) and *Sal*I−*Sca*I run-off SP6 transcript of 594 nt (d). Sense probe is: *Sca*I−*Sph*I run-off T7 transcript of 727 nt (e). The results from the RNase protection experiments presented in (B) and (C) are summarized in the two transcripts depicted below the map and are both directed towards the VSG gene in the 221 ES (wavy arrows). Transcript 1 is formed by an RNA polymerase resistant to 1 mg/ml α-amanitin. The α-amanitin sensitivity of the less abundant transcript 2 is not known. The filled and open triangles point to the 5′ ends of transcripts 1 and 2 respectively and correspond to those depicted in (B) and (C). Restriction sites: D, *Dra*I; E, *Eco*RV; S, *Sal*I; Sc, *Sca*I; Sh, *Sph*I. **B.** RNase protection of total trypanosome RNA. 30 μg of total RNA from bloodstream form *T.brucei* variant 221a was hybridized for 16 h at 50°C (**lanes 4, 8, 12** and **16**) or at 60°C (**lanes 2, 6, 10** and **14**) with the uniformly labelled probes indicated above the lanes (probes a, b and c are antisense and probe e is sense, see A). The probes not treated with RNase are shown in **lanes 1** (a), **5** (b), **9** (c) and **13** (e). In **lanes 3, 7, 11** and **15**, 30 μg *E.coli* tRNA was hybridized (at 50°C in 80% formamide) to the probes a, b, c and e, respectively, and RNase-treated. Filled triangles point to abundant protected fragments of either probe a or b (89 and 274 nt, respectively) and the open ones to minor specific bands (127 and 313 nt, respectively). The triangles correspond to those depicted in (A). **C.** RNase protection on run-on RNA. Nascent RNA was synthesized in the presence of [α-^{32}P]UTP in isolated nuclei of bloodstream form variant 221a or culture form (CF) trypanosomes (lanes 1−10 and 11−20, respectively) and was hybridized to 0.5 μg of a non-labelled *in vitro* synthesized antisense transcript covering the promoter area (transcript d, see A) at 55°C (**lanes 1, 6, 11** and **16**), at 65°C (**lanes 2, 7, 12** and **17**) or at 75°C (**lanes 3, 8, 13** and **18**) or with *E.coli* tRNA at 65°C (**lanes 4, 9, 14** and **19**). Input is shown in **lanes 5, 10, 15** and **20**. Run-ons were performed either without α-amanitin (−; lanes 1−5 and 11−15) or with (+; lanes 6−10 and 16−20) 1 mg/ml α-amanitin. A sequence reaction was loaded in parallel to estimate the size of the protected fragments. The filled triangle marks the protected fragment of 89 nt also seen in total RNA (see B) and maps to the position indicated in A (transcript 1).

sequences. The constructs in which the VSG gene 221 ES promoter was present in the reverse orientation (pRK4− and pRK8−), gave no detectable CAT activity either. With DNA concentrations ranging from 100 to 400 μg/ml we consistently found (four experiments) for pRK4+ low and for pRK8+ high CAT activities (3× and 70× the counts obtained with the constructs with the promoter in the reverse orientation). In fact relative CAT activities are even higher than those obtained from liquid scintillation counting of the xylene fractions because of residual labelled chloramphenicol present in these fractions (see also Figure 6). The CAT activity in trypanosomes transfected with pRK8+ plasmid DNA is comparable with that obtained with pJF6, containing the *parpA* promoter (Clayton *et al.*, 1990; data not shown).

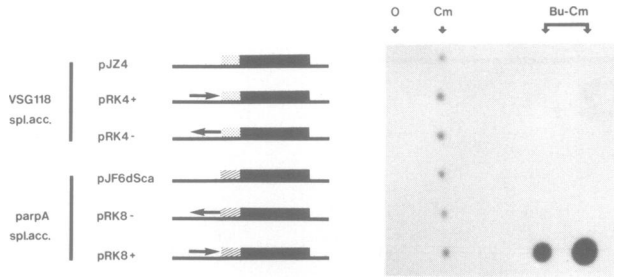

Fig. 6. CAT activity of constructs containing the 221 ES transcription initiation region and the CAT gene, introduced into procyclic trypanosomes. At the left the constructs are depicted (not drawn to scale) that were transfected into procyclic (culture form) *T.brucei*. The solid bar represents the CAT gene, the dotted box the VSG gene 118 splice acceptor region and the hatched box the splice acceptor region of the *parp* gene A (see Materials and methods for details). The large arrows represent the VSG gene 221 ES transcription initiation region and its transcriptional orientation relative to the CAT gene. The TLC of products formed in the CAT assay with extracts of trypanosomes transfected with these constructs at a concentration of 100 μg supercoiled plasmid DNA per 5×10^7 trypanosomes is shown on the right. The small arrows indicate origin (O), [^{14}C]chloramphenicol (Cm), and 1- and 3-butyryl-[^{14}C]chloramphenicol (Bu-Cm).

It is of interest that the pRK4+ and pRK8+ constructs, which differ in the nature of the sequences between promoter area and gene, differ 20-fold in CAT activity. We have found that this is at least in part due to sequences upstream of the splice acceptor area, as shortening of this sequence by 100 nt in pRK4+ led to a change in CAT activity (results not shown). However, other factors may contribute to the large differences in CAT activity obtained with pRK4+ and pRK8+. We cannot exclude that the splice acceptor site from the *parp* gene is used more efficiently in procyclics than is the splice acceptor from the VSG gene. Moreover, the two constructs are in different vectors and we have found that the nature and orientation of vector sequences can affect CAT gene expression (C.Clayton, D.Sherman and K.Fung, unpublished results).

The promoter areas of the 221 ES in T.brucei variants 221a and 221aR12 are identical

To test whether subtle DNA rearrangements occur upon inactivation of the 221 ES, we sequenced the promoter regions of the 221 ES of variant 221a, in which this ES is active, and of the relapse variant 221aR12 in which the 221 ES is switched off and another ES activated (Figure 7). We did not find a single basepair difference in the entire region from the 50 bp repeats down to ESAG-X, i.e. 3.4 kb. These

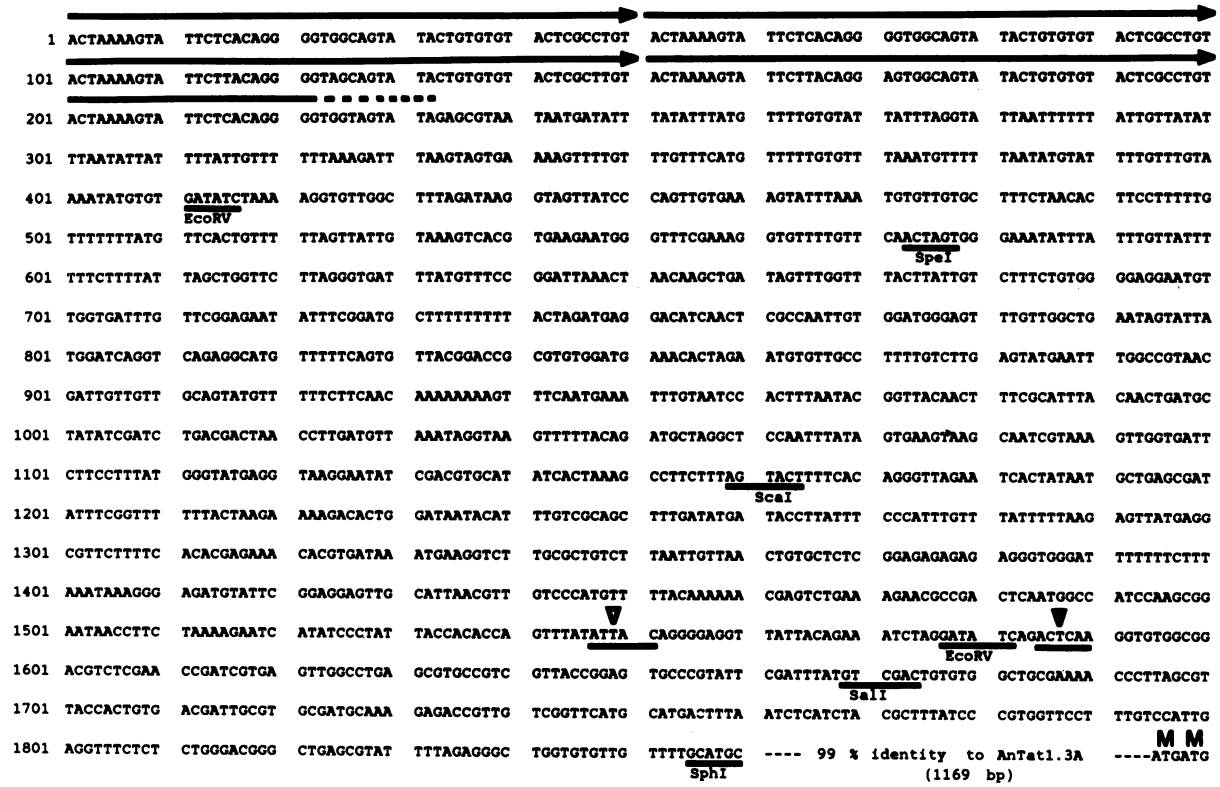

Fig. 7. Sequence of the 221 ES promoter area from *T.brucei* variants 221a and 221aR12. Position 1 is within the imperfect 50 bp direct repeat sequence, indicated by an arrow above the sequence. The filled and open triangles mark within 5 bp (underlined) the 5′ ends of transcripts 1 and 2 respectively (see Figure 5). Only relevant restriction sites are indicated. The sequence of 1169 bp between the *Sph*I site and the first methionine codon (M) of ESAG-X is not shown, and has 99% identity to the corresponding region in clone ES12 of the AnTat 1.3A ES published by Pays *et al.* (1989b). This sequence will appear in sequence databases under the accession number X17350.

data show that the shut-off of the 221 ES promoter in the switch from variant 221a to variant 221aR12 does not involve DNA rearrangements at or near the transcription start site.

The only features that stand out from the sequences around the two 5′ ends, mapped at positions 1549 and 1587 (see Figure 7) are three direct repeats at positions 1528−1533, 1546−1553 and 1561−1568. In addition, there are two imperfect inverted repeats at positions 1529−1542 together with 1550−1563, and 1584−1593 together with 1614−1623. The significance of these sequence elements is questionable.

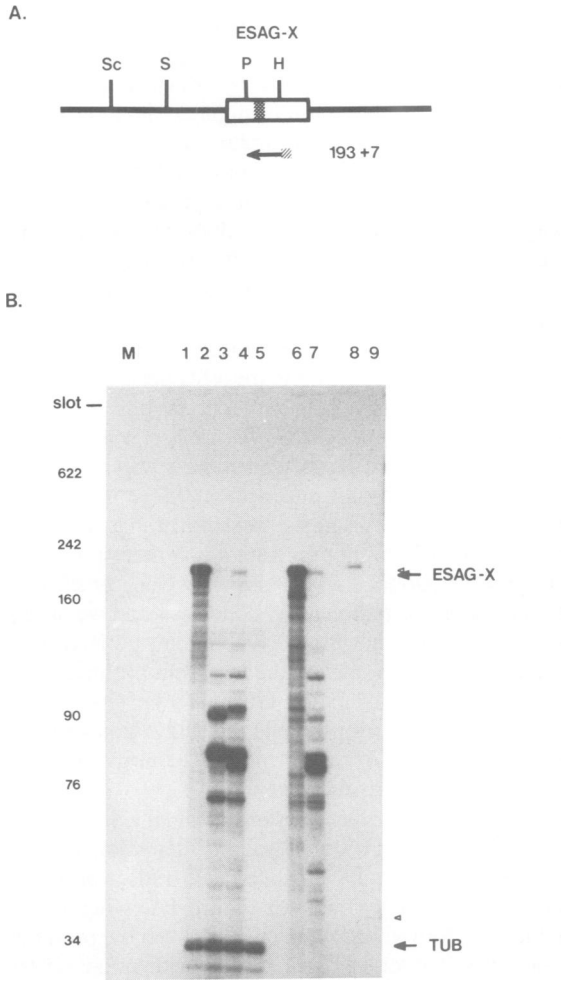

Fig. 8. ESAG-X of the 221 ES is not expressed in *T.brucei* variant 221aR12. **A.** Schematic presentation of ESAG-X to show the antisense SP6 transcript of 200 nt which includes seven polylinker nucleotides at the 5′ end (hatched box). *Pvu*II (a cluster of three sites) was used as run-off site. The transcript overlaps the hypervariable region (filled box in ESAG-X). This probe was used in the RNase protection experiment that is shown in B. Restriction sites: H, *Hin*dIII; P, *Pvu*II; S, *Sal*I; Sc, *Sca*I. **B.** 10 μg of total RNA from bloodstream form *T.brucei* variants 221a (**lane 2**), 221aR12 (**lane 3**), 118a (**lane 4**) and from culture form *T.brucei* 427 (**lane 5**) was hybridized for 16 h at 50°C in 80% formamide (see Materials and methods) with the probe overlapping the hypervariable region of ESAG-X (as described in A) and a β-tubulin 5′ end *Pst*I−*Kpn*I antisense probe. The hybrids were subsequently digested with RNase A and T1, size-fractionated in a 6% polyacrylamide, 7 M urea gel and visualized by autoradiography. **Lanes 6** and **7** contain *in vitro* synthesized sense ESAG-X transcripts of the 221 ES and another ES, respectively. **Lanes 8** and **9** show the ESAG-X and tubulin probes, respectively, not treated with RNase (and indicated by the arrowheads). In **lane 1**, 10 μg tRNA was hybridized to the probes and RNase treated. Arrows indicate full length protected fragments for ESAG-X and tubulin. The sizes of the fragments (in bp) are indicated alongside the gel.

The absence of VSG 221 mRNA in T.brucei variant 221aR12 is not due to transcription termination downstream of the 221 ES promoter

Inactivation of an ES might occur by insertion of a transcriptional terminator rather than by promoter inactivation. Although this is unlikely to be a general mechanism (see Discussion) it is not without precedent, since we found a relapse variant of 221a, named 221aR2, in which a 30 kb insert between the VSG gene 221 and the pseudogene prevented the transcription of the VSG gene but did not alter the transcription of sequences upstream of the pseudogene (Cornelissen *et al.*, 1985). A small insertion interfering with transcription would go undetected by the methods employed thus far.

To examine the presence of a terminator of transcription somewhere between the VSG gene 221 and ESAG-X in the 221 ES of variant 221aR12, we tested whether ESAG-X of the 221 ES is still expressed in this *T.brucei* variant. We used an antisense probe spanning the hypervariable region of ESAG-X (Figure 8A) to distinguish ESAG-X mRNAs from different expression sites in an RNase protection experiment. This probe is protected over its full length (193 nt) by the ESAG-X mRNA present in total RNA from variant 221a (Figure 8B, lane 2). The protected fragment can be distinguished from incomplete RNase digestion products by the removal of a 7 nt polylinker sequence from the pGEM3 vector (compare lanes 2 and 8). The ESAG-X mRNA from the 221 ES is absent in RNA from variant 221aR12, but the presence of ESAG-X messages from another ES results in specific degradation products as a consequence of mismatches in the hypervariable region (lane 3). The amounts of ESAG-X mRNAs in total RNA from the two variants do not differ as judged from the mRNA levels of the constitutively expressed β-tubulin genes (lanes 2−5 and 9). As expected there is no ESAG-X mRNA from the 221 ES in total RNA from variant 118a (lane 4), and in culture form *T.brucei* there is no detectable ESAG-X expression (lane 5). As positive and negative controls in this RNase protection experiment we used *in vitro* synthesized sense RNA from the 221 ES ESAG-X and another ESAG-X (1.8 cDNA, see Figure 3) (lanes 6 and 7, respectively). The residual full-length protection that we observe in lanes 3 and 4 is not significantly above that seen in lane 7 and is probably due to the relatively inefficient recognition by RNase A and T1 of a few single mismatches. Hence, these experiments show that in *T.brucei* variant 221aR12 the ESAG-X gene in the 221 ES is shut off. Since we did not find sequence differences upstream of ESAG-X in the 221 ES of variants 221a and 221aR12 (Figure 7), these data exclude an insertional inactivation of the 221 ES by a terminator of transcription.

Discussion

We have previously shown that variant surface glycoprotein (VSG) gene expression sites (ESs) are long multigene transcription units (Johnson *et al.*, 1987; Kooter *et al.*, 1987). Whereas most of the genes and intergenic regions of these units have been characterized in recent years (Cully *et al.*, 1985, 1986; Florent *et al.*, 1987; Kooter *et al.*, 1987; Murphy *et al.*, 1987; Shah *et al.*, 1987; Shea *et al.*, 1987; Alexandre *et al.*, 1988; Gibbs and Cross, 1988; Myler *et al.*, 1988; Pays *et al.*, 1989b; Borst *et al.*, 1990), the promoter controlling this complex transcription unit remained

to be identified. The results presented here fill this gap. UV inactivation of transcription (Johnson *et al.*, 1987), run-on transcription, RNase protection and DNA transfection experiments have led us to the identification of the promoter of such a multicistronic transcription unit.

The rapid processing of nascent transcripts of VSG ESs and other trypanosome transcription units (see Borst, 1986) has made it difficult to locate transcription starts unambiguously. We have relied on the nuclear run-on assay, optimized for trypanosome nuclei (Kooter and Borst, 1984), to obtain labelled nascent transcripts with minimal processing. When such transcripts are hybridized to DNA fragments corresponding to the 221 ES, the level of hybridization is proportional to fragment length, taking into account internal duplications (Kooter *et al.*, 1987). Moreover, inactivation of transcription with UV light shows the fade-out of ES transcription expected for a single 60 kb transcription unit (Johnson *et al.*, 1987).

Under our experimental conditions we find no nascent transcripts upstream of the flag in Figure 1A. Since the 50 bp repeat area, which starts 1.4 kb upstream of the flag and probably extends for at least 10 kb, also does not hybridize to nascent transcripts, we conclude that the flag represents the start of the ES transcription unit and not an RNA processing site. The UV inactivation of the transcription of the 221 ES provides independent support for this conclusion, as it places the transcription start approximately at the flag and not >10 kb further upstream (Johnson *et al.*, 1987; Borst *et al.*, 1990). Moreover, this transcription initiation region promotes CAT gene expression when introduced into procyclic trypanosomes. The divergence of sequences upstream of the *Spe*I site (see Figure 2, and unpublished results) in different ESs, also suggests that the transcription controlling DNA elements are confined to the highly conserved region ~1.0 kb upstream of the start sites. In RNase protection experiments we detect two starts, a major one mapping just downstream of the *Eco*RV site in Figure 4, and a minor one 38 bp upstream of the major start site. The minor transcript cannot explain the weak hybridization of the *Sca*I–*Eco*RV fragment 8 in Figure 4. This hybridization might be due either to multiple minor starts not detected in the nuclease protection assay, or, more plausibly, to some cross-hybridization by α-amanitin resistant nascent RNA from other transcription units, e.g. telomeres (Rudenko and Van der Ploeg, 1989), procyclin genes (Rudenko *et al.*, 1989) or from other segments of the 221 ES. Such cross-hybridization might not be seen in a stringent RNase protection assay.

It is even possible that processing is so rapid in the early part of this transcription unit that labelled fragments are generated that are too small to give detectable products in RNase protection, but large enough to hybridize. It seems unlikely, however, that the initiation sites deduced are, in reality, splice sites. The AG dinucleotides at positions 1552 and 1590 (see Figure 7) do not map exactly at the 5′ ends of the minor and major start sites. Moreover, the sequences in front of these AGs do not conform to the consensus 3′ splice sites in *T.brucei* (Layden and Eisen, 1988).

On the basis of these arguments, we conclude that the transcription of the 221 ES starts in the *Sca*I–*Sal*I fragment 7 in Figure 4 and probably at the positions indicated by the arrowheads in Figure 7. This conclusion has several implications for the control of VSG gene transcription: the most important implication is that a VSG gene ES in *T.brucei* can switch off without a single base pair change within 1.4 kb of the transcription start site. This eliminates switching models invoking promoter inversion or conversion (cf. Borst and Greaves, 1987). Obviously, DNA rearrangements outside the area sequenced may still be involved, although we have no evidence for such rearrangements.

We also find that the start of the VSG gene transcription unit has no strong homology with the sequence at the start of the rRNA transcription unit in trypanosomes (White *et al.*, 1986). We have scanned the 1.4 kb sequence upstream of the putative start sites of transcription and found no sequence with substantial homology to the rRNA promoter. Our results therefore provide no support for the idea that VSG genes are transcribed by RNA polymerase I. A modified RNA polymerase II might do the job (Evers *et al.*, 1989; Grondal *et al.*, 1989; Smith *et al.*, 1989). Comparison of the VSG gene promoter region with that of another protein coding gene transcribed by an α-amanitin resistant RNA polymerase, the *parp* gene, also does not reveal any obvious sequence homologies although the inverted repeat structures found in both promoter regions might be significant (this paper; Clayton *et al.*, 1990).

Our results contrast with those obtained for putative accessory promoters closer to the VSG gene in the ES, which were reported to have substantial homology with the rRNA promoter (Shea *et al.*, 1987; Alexandre *et al.*, 1988). The existence of these accessory promoters was mainly inferred from a gap in run-on transcription. More recently Pays *et al.* (1989b) have raised the possibility that the gap in the AnTat 1.3A ES might represent a polymerase re-entry site. The transcriptional gap observed by Shea *et al.* (1987) is dependent on the run-on assay conditions and we have shown by UV fade-out that the corresponding putative promoter is completely dependent on the upstream promoter identified here for its activity (Borst *et al.*, 1990; Crozatier *et al.*, 1990). It is therefore probable that these transcription gaps reflect processing or transcriptional lethargy during run-on assays, rather than true promoters.

Our results show that the shut-off of an ES in bloodstream trypanosomes must occur at transcription initiation or shortly after initiation. We have proven that transcription does not reach the ESAG-X gene (see Figure 1A) when the 221 ES is shut off. Moreover, nascent RNA from the promoter area is not grossly overrepresented in hybridization experiments (Kooter *et al.*, 1987; this paper), as would be expected if the ESs were controlled by early termination shortly after initiation. There are at least five (Liu *et al.*, 1985) and probably up to 20 ESs (Cully *et al.*, 1985) and the promoter areas of these ESs are all highly homologous (Kooter *et al.*, 1987; Pays *et al.*, 1989b; Borst *et al.*, 1990; and our unpublished results). A 5- to 20-fold overrepresentation has not been observed by Kooter *et al.* (1987) or by Pays *et al.* (1989b). Of course, the theoretical possibility that nascent transcripts are too short or too unstable to detect in run-on transcription cannot be formally excluded.

In contrast, the shut-off of the VSG gene ES that occurs in the transition from bloodstream to procyclic (insect form) trypanosomes clearly involves transcription attenuation, rather than a block in initiation. Pays *et al.* (1989a) have shown that this attenuation occurs near the ESAG-X gene and our results, showing promoter-proximal transcripts in culture form trypanosomes, confirm the conclusions of Pays

et al. We conclude therefore that trypanosomes can shut off VSG gene expression in two different ways. When a trypanosome is taken up by a tsetse fly, the active ES is shut down by early termination of RNA elongation at least 1 kb downstream of transcription initiation (Pays *et al.*, 1989a). When a bloodstream trypanosome inactivates an ES, this occurs by preventing initiation of RNA synthesis (or stopping elongation directly after initiation).

The 221 ES promoter region identified by the analysis of nascent transcripts is able to direct transcription of a CAT construct after transformation of trypanosomes. As our attempts (and those of others) to transform bloodstream trypanosomes have failed thus far, we were forced to use culture form trypanosomes, which do not produce VSG mRNA, to test our constructs. Remarkably, we find that the transient expression of the ES promoter in the procyclic trypanosomes is comparable with that of promoters that are normally highly active in procyclic trypanosomes, i.e. the *parpA* gene promoter (Clayton *et al.*, 1990) and the ribosomal promoter (White *et al.*, 1986) (data not shown). This proves that the shut-off of all but one ES promoter in procyclic trypanosomes must depend on factors outside the promoter area itself and must be caused by repression, possibly by attachment to nuclear structures (Van der Ploeg and Cornelissen, 1984), rather than by the lack of transcription machinery.

Materials and methods

Trypanosomes

The trypanosomes used belong to strain 427 of *T.brucei brucei*. Trypanosome variant 221a (MiTat 1.2a) is described by Cross (1975). Variant 221aR1 (Bernards *et al.*, 1984) is a relapse of 221a, in which the VSG gene 221 is replaced by VSG gene 1.8 in the 221 ES. Variant 221aR12 is a relapse variant of 221a in which the VSG gene 221 is no longer expressed but is retained as a lingering expression linked copy in the 221 ES. This *T.brucei* variant 221aR12 expresses the VSG gene 1.8 in another ES (unpublished results). Variants 118a (MiTat 1.5a), 118a′ and 1.8b are described by Michels *et al.* (1983, 1984). Trypanosomes were grown in Wistar rats and blood was collected from animals with a high parasitaemia by cardiac puncture. Trypanosome populations were tested for antigenic homogeneity by indirect immunofluorescence using VSG polyclonal antibodies. Buffy coat was used for RNA isolations; for PFGE analysis trypanosomes were separated from blood cells by DEAE−cellulose chromatography according to Fairlamb *et al.* (1978). Procyclic culture form trypanosomes were grown in the semi-defined medium described by Brun and Schönenberger (1979).

Isolation and blotting analysis of DNA and RNA

Nuclear DNA was isolated as described by Bernards *et al.* (1981). Total RNA was isolated from trypanosomes by lysis of cells from buffy coat in guanidium isothiocyanate followed by centrifugation over a caesium chloride cushion (MacDonald *et al.*, 1987). DNA and RNA were blotted onto nitrocellulose by standard procedures (Maniatis *et al.*, 1982), for large DNA molecules preceded by a limited acid hydrolysis as described by Wahl *et al.* (1979). Blots were hybridized with random primed ^{32}P-labelled probes (Feinberg and Vogelstein, 1983) as described by Jeffreys and Flavell (1977), with the addition of 10% (w/v) dextran sulphate (Wahl *et al.*, 1979). Blots were washed at a stringency of 0.3 × SSC, 0.1% SDS at 60°C.

Pulsed-field gradient gel electrophoresis and construction of the 221 chromosome specific libraries

Trypanosomes were prepared as described by Van der Ploeg *et al.* (1984). To optimize the separation of large chromosomes, the number of cells per plug of LMP-agarose was reduced to 5 × 10^8 per ml (Johnson and Borst, 1986). The chromosome-sized DNA molecules were separated in 0.5% GTG−agarose (FMC Corp.) gels in 1 × TBE (Maniatis *et al.*, 1982) for 24 h at 14°C, and the electric field (10 V/cm; 200 V, 170 mA) was switched every 260 s. After the run, the DNA was visualized with ethidium bromide and the region containing the 221 chromosome, either of variant 221a or variant 221aR12, was cut out and DNA was recovered from the gel slices by PFGE in dialysis tubes. A chromosome specific library in phage λ EMBL 3 (Frischauf *et al.*, 1983) was constructed. In brief, the electroeluted DNA was concentrated by *n*-butanol extraction and extensively dialysed against 10 mM Tris−HCl, pH 8.0, and 1 mM EDTA prior to partial *Sau*3AI digestion. After the digestion, the DNA was treated with calf intestinal alkaline phosphatase and size fractionated on low melting point agarose (ultra pure, BRL) gels. DNA fragments in the 10−25 kb size range were recovered by phenol extraction and subsequently ligated onto *Bam*HI digested phage λ EMBL 3 arms. Packaging, plating and screening of the banks were standard. Phage DNA was purified as described (Maniatis *et al.*, 1982).

DNA sequence analysis

DNA sequence analysis was performed using the dideoxy method (Sanger *et al.*, 1977) and modified T7 DNA polymerase (Sequenase version 2.0) according to the manufacturer's instructions (United States Biochem.). All sequences were determined on both strands. Genomic fragments from λ clones were subcloned in pGEM and pGEM-Zf vectors (Promega Biotech). Cloning and preparation of template DNA were standard. Computer analysis of the nucleotide sequence was performed using the software package of the University of Wisconsin Genetics Computer Group (UWGCG) (Devereux *et al.*, 1984). The sequence will appear in the EMBL/GenBank/DDBJ nucleotide sequence databases under the accession number X17350.

Preparation of nuclei from trypanosomes and elongation of nascent RNA in nuclear run-ons

Nuclei from bloodstream form trypanosomes were prepared by passing total blood of highly infected rats through a Stansted cell disrupter as described by Kooter *et al.* (1987). Procyclic culture form trypanosomes were grown to a density of ~5 × 10^7 cells/ml and prepared for the nuclei preparation as described (Kooter *et al.*, 1987). Approximately 1 × 10^9 nuclei were resuspended in 100 mM Tris−HCl, pH 7.5, 20% (v/v) glycerol, 0.15 mM spermine, 0.5 mM spermidine, 2 mM DTT, 2.5 mM $MgCl_2$, 4 mM $MnCl_2$, 50 mM NaCl, 10 mM KCl, 70 U RNasin (Promega Biotech.), 2 mM ATP, 1 mM CTP, 1 mM UTP, 10 μM GTP or UTP, 200 μCi [α-^{32}P]GTP or [α-^{32}P]UTP (3000 Ci/mmol), 100 μg creatine kinase and 10 mM creatine phosphate in a volume of 100 μl and incubated in the absence of α-amanitin or in the presence of 1 mg/ml α-amanitin (Sigma) for 15 min at 0°C. The nascent RNAs were elongated *in vitro* by incubating the mixture at 37°C for 5 min. The reaction was terminated by adding proteinase K (50 μg/ml) and SDS [0.2% (w/v)]. After 30 min at 37°C, the mixture was extracted with phenol/chloroform. Nucleic acids were precipitated with ethanol and redissolved in 200 μl of 10 mM Tris−HCl, pH 8.0, 5 mM $MgCl_2$ and 100 U RNase-free DNase (Boehringer Mannheim). After 15 min at 37°C the mixture was phenol/chloroform extracted and ethanol precipitated in the presence of 2 M ammonium acetate.

RNase protection and in vitro transcription

RNase protection and *in vitro* transcription were performed as described by Zinn *et al.* (1983). Ten to fifty μg of total RNA or run-on RNA from 2 × 10^8 nuclei was hybridized with RNA probes under standard conditions (80% formamide, 40 mM PIPES, pH 6.4, 400 mM NaCl and 1 mM EDTA) at the temperature indicated in the figure legends, followed by an RNase A and T1 treatment (12 μg RNase A and 3 U RNase T1, 40 min at 30°C). Protected fragments were visualized by electrophoresis in a 6% acrylamide, 7 M urea gel, followed by autoradiography.

DNA transfection of procyclic trypanosomes by means of electroporation

The method was essentially as described in Bellofatto and Cross (1989), adapted for insect form *T.brucei* by Clayton *et al.* (1990). Briefly, cells in mid-log phase (5−8 × 10^6 trypanosomes/ml) were harvested, washed and resuspended in Zimmerman Postfusion medium to a density of 5 × 10^7 cells/ml. 0.5 ml of this trypanosome suspension was transferred to a Biorad Gene Pulser cuvette (0.2 cm electrode distance), containing DNA. Two pulses of 1600 V, 25 μF capacitance were delivered by a Biorad Gene Pulser. Approximately 40% of the trypanosomes survived these electro-shocks. The trypanosomes were immediately transferred to 10 ml SDM 79 medium, containing 50 U penicillin−streptomycin (Gibco) per ml, and incubated for 20 h at 28°C. Cell extracts were prepared as described (Bellofatto and Cross, 1989). CAT activity was assayed as described (Seed and Sheen, 1988), using 28 μM *n*-butyryl CoA (Sigma) and 88 μM D-thero[dichloroacetyl-1,2-^{14}C]chloramphenicol (NEN, DuPont) at a final specific activity of 59.5 mCi/mmol. Assays were for 4 h at 37°C. After xylene and two aqueous phase extractions (Seed and Sheen, 1988), the xylene fractions were dried, resuspended in ethyl acetate and spotted onto silica plates, which were subsequently developed in 95:5 (v/v) chloroform−methanol. Alternatively,

before drying, 50 μl of the xylene fractions was mixed with 5 ml of Optifluor (Packard) and radioactivity was determined by liquid scintillation counting.

The plasmids used in the procyclic transfection experiments were constructed as follows: the 786 bp CAT gene was cloned as a *Hin*dIII–*Bam*HI fragment (from plasmid pJF6; Clayton *et al.*, 1990) into Bluescript M13+ (Stratagene), resulting in plasmid pJZ1. The splice acceptor region of the VSG gene 118 was fused via a *Nco*I site to the ATG of the CAT gene. To obtain a *Sal*I–*Nco*I fragment from 150 bp upstream of the splice acceptor site in the VSG gene 118 up to the ATG, this region was amplified from plasmid pc 118-29I (Liu *et al.*, 1983) with Taq DNA polymerase (Perkin Elmer Cetus) in a polymerase chain reaction (Saiki *et al.*, 1988) according to the manufacturer's instructions. The oligonucleotides used were a specific 5′ primer containing at the 5′ end a *Sal*I site and overlapping bases 2007–2024 and a specific 3′ primer overlapping bases 2278–2307, converting the ATG start codon of the VSG gene 118 into a *Nco*I site. The ATG start codon of the CAT gene was converted to a *Nco*I site, by PCR amplification with a primer overlapping nucleotides 4950–4979 and a 3′ primer at position 4743–4772, overlapping the *Eco*RI site in the CAT gene (pSV2CAT; Gorman, 1985). The former PCR fragment was digested with the restriction enzymes *Sal*I and *Nco*I, the latter with *Nco*I and *Eco*RI. Gel-purified PCR fragments were subsequently subcloned into pJZ1, replacing the *Sal*I–*Eco*RI sequence by the VSG gene 118 splice acceptor region fused to the ATG of the CAT gene, resulting in plasmid pJZ4. This construct was checked by sequencing. A 2.6 kb *Sal*I fragment of the VSG gene 221 ES transcription initiation region, encompassing fragments 3, 6, 5 and 7 (see Figure 4), was cloned into the *Sal*I site of pJZ4 in the correct (pRK4+) and reverse (pRK4−) orientations. pJF6dSca was a deletion derivative of pJF6 lacking the promoter region. This construct was made by ligating the polylinker blunted *Kpn*I site upstream of the *parpA* promoter–splice acceptor region to the *Sca*I site 78 bp upstream of the AG splice acceptor site of the *parpA* (at bases 1105–1106; Clayton *et al.*, 1990). Into the blunted *Sac*I polylinker site we cloned the 2.6 kb blunted *Sal*I fragment of the VSG gene 221 ES transcription initiation region (as described above) in the correct (pRK8+) and reverse (pRK8−) orientations.

Acknowledgements

We thank Janet Gommers-Ampt, Carsten Lincke, Jaap Smit and Drs Frank Baas, Peter Laird, Alfred Schinkel and Bart Swinkels for interest and advice throughout the course of this work and for critical reading of the manuscript. We thank Francesca Fase-Fowler for the supply of enzymes and reagents. This work was supported in part by a grant from the Netherlands Foundation for Chemical Research (SON), with financial aid from the Netherlands Organization for Scientific Research (NWO), and by a postdoctoral fellowship from NATO (Canada) to M.O.

References

Alexandre,S., Guyaux,M., Murphy,N.B., Coquelet,H., Pays,A., Steinert,M. and Pays,E. (1988) *Mol. Cell. Biol.*, **8**, 2367–2378.

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

Bellofatto,V. and Cross,G.A.M. (1989) *Science*, **244**, 1167–1169.

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

Bernards,A., De Lange,T., Michels,P.A.M., Liu,A.Y.C., Huisman,M.J. and Borst,P. (1984) *Cell*, **36**, 163–170.

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

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

Borst,P. and Greaves,D.R. (1987) *Science*, **235**, 658–667.

Borst,P., Zomerdijk,J., Ouellette,M., Crozatier,M. and Ten Asbroek,A. (1990) In Agabian,N. and Cerami,A. (eds), *Parasites: Molecular Biology, Drug and Vaccine Design. UCLA Symposium on Molecular and Cellular Biology, New Series*. Wiley-Liss, Inc., New York, Vol. 131, pp. 59–72.

Brun,R. and Schönenberger,M. (1979) *Acta Trop.*, **36**, 289–292.

Buck,G.A., Jacquemot,C., Baltz,T. and Eisen,H. (1984) *Gene*, **32**, 329–336.

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

Clayton,C.E., Fueri,J.P., Itzhaki,J.E., Bellofatto,V., Sherman,D.R., Wisdom,G.S., Vijayasarathy,S. and Mowatt,M.R. (1990) *Mol. Cell. Biol.*, **10**, 3036–3047.

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**, 393–418.

Crozatier,M., Van der Ploeg,L.H.T., Johnson,P.J., Gommers-Ampt,J. and Borst,P. (1990) *Mol. Biochem. Parasitol.*, in press.

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

Cully,D.F., Gibbs,C.P. and Cross,G.A.M. (1986) *Mol. Biochem. Parasitol.*, **21**, 189–197.

De Lange,T. and Borst,P. (1982) *Nature*, **299**, 451–453.

Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.

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

Evers,R., Hammer,A., Kock,J., Jess,W., Borst,P., Memet,S. and Cornelissen,A.W.C.A. (1989) *Cell*, **56**, 585–597.

Fairlamb,A.H., Weislogel,P.O., Hoeijmakers,J.H.J. and Borst,P. (1978) *J. Cell Biol.*, **76**, 293–309.

Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.

Florent,I., Baltz,T., Raibaud,A. and Eisen,H. (1987) *Gene*, **53**, 55–62.

Frischauf,A.M., Lehrach,H., Poustka,A. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827–842.

Gibbs,C.P. and Cross,G.A.M. (1988) *Mol. Biochem. Parasitol.*, **28**, 197–207.

Gorman,C. (1985) In Glover,D.M. (ed.), *DNA Cloning—A Practical Approach*. IRL Press, Oxford, Vol. 2, pp. 143–190.

Grondal,E.J.M., Evers,R., Kosubek,K. and Cornelissen,A.W.C.A. (1989) *EMBO J.*, **8**, 3383–3389.

Hoeijmakers,J.H.J., Frasch,A.C.C., Bernards,A., Borst,P. and Cross,G.A.M. (1980) *Nature*, **284**, 78–80.

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

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

Johnson,P.J., Kooter,J.M. and Borst,P. (1987) *Cell*, **51**, 273–281.

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

Kooter,J.M., Van der Spek,H.J., Wagter,R., d'Oliveira,C.E., Van der Hoeven,F., Johnson,P.J. and Borst,P. (1987) *Cell*, **51**, 261–272.

Laird,P.W., Kooter,J.M., Loosbroek,N. and Borst,P. (1985) *Nucleic Acids Res.*, **13**, 4253–4266.

Lamont,G.S., Tucker,R.S. and Cross,G.A.M. (1986) *Parasitology*, **92**, 355–367.

Layden,R.E. and Eisen,H. (1988) *Mol. Cell. Biol.*, **8**, 1352–1360.

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

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

MacDonald,R.J., Swift,G.H., Przybyla,A.E. and Chirgwin,J.M. (1987) *Methods Enzymol.*, **152**, 219–227.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Michels,P.A.M., Liu,A.Y.C., Bernards,A., Sloof,P., 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.

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

Murphy,N.B., Guyaux,M., Pays,E. and Steinert,M. (1987) In Agabian,N., Goodman,H. and Nogueiro,N. (eds), *Molecular Strategies of Parasitic Invasion*. Alan R.Liss, Inc., New York, pp. 449–469.

Myler,P.J., Aline,R.F., Scholler,J.K. and Stuart,K.D. (1988) *Mol. Biochem. Parasitol.*, **29**, 243–250.

Pays,E. and Steinert,M. (1988) *Annu. Rev. Genet.*, **22**, 107–126.

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

Pays,E., Coquelet,H., Pays,A., Tebabi,P. and Steinert,M. (1989a) *Mol. Cell. Biol.*, **9**, 4018–4021.

Pays,E., Tebabi,P., Pays,A., Coquelet,H., Revelard,P., Salmon,D. and Steinert,M. (1989b) *Cell*, **57**, 835–845.

Rudenko,G. and Van der Ploeg,L.H.T. (1989) *EMBO J.*, **8**, 2633–2638.

Rudenko,G., Bishop,D., Gottesdiener,K. and Van der Ploeg,L.H.T. (1989) *EMBO J.*, **8**, 4259–4263.

Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) *Science*, **239**, 487–491.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Seed,B. and Sheen,J.-Y. (1988) *Gene*, **67**, 271–277.

Shah,J.S., Young,J.R., Kimmel,B.E., James,K.P. and Williams,R.O. (1987) *Mol. Biochem. Parasitol.*, **24**, 163–174.

Shea,C., Lee,M.G.S. and Van der Ploeg,L.H.T. (1987) *Cell*, **50**, 603–612.

Smith,J.L., Levin,J.R., Ingles,C.J. and Agabian,N. (1989) *Cell*, **37**, 815–827.

Van der Ploeg,L.H.T. and Cornelissen,A.W.C.A. (1984) *Phil. Trans. R. Soc. Lond. B.*, **307**, 13–26.

Van der Ploeg,L.H.T., Valerio,D., De Lange,T., Bernards,A., Borst,P. and Grosveld,F.G. (1982) *Nucleic Acids Res.*, **10**, 5905–5923.
Van der Ploeg,L.H.T., Schwartz,D.C., Cantor,C.R. and Borst,P. (1984) *Cell*, **37**, 77–84.
Vickerman,K. (1978) *Nature*, **273**, 613.
Wahl,G.M., Stern,M. and Stark,P.G.R. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3683–3687.
White,T.C., Rudenko,G. and Borst,P. (1986) *Nucleic Acids Res.*, **14**, 9471–9489.
Williams,R.O., Young,J.R. and Majiwa,P.A.O. (1979) *Nature*, **282**, 847–849.
Zinn,K., DiMaio,D. and Maniatis,T. (1983) *Cell*, **34**, 865–879.

Received on April 6, 1990; revised on June 8, 1990

Note added in proof

Deletion experiments have shown that sequences upstream of position 1489 in Figure 7 are not required for the high CAT activity in Figure 6.