MOLECULAR AND CELLULAR BIOLOGY, June 1988, p. 2367–2378
0270-7306/88/062367-12$02.00/0

Vol. 8, No. 6

Putative Genes of a Variant-Specific Antigen Gene Transcription Unit in *Trypanosoma brucei*

SYLVIE ALEXANDRE, MICHEL GUYAUX, NOEL B. MURPHY, HÉLÈNE COQUELET, ANNETTE PAYS, MAURICE STEINERT, AND ETIENNE PAYS*

Department of Molecular Biology, University of Brussels, 67, rue des Chevaux, B1640 Rhode St Genèse, Belgium

Received 14 December 1987/Accepted 2 March 1988

In a 7-kilobase (kb) sequence upstream from the 5′ barren region, the *Trypanosoma brucei* AnTat 1.3A expression site carries two putative genes, named ESAG 2 and ESAG 3 for expression site-associated genes, as well as a copy of ESAG 1 (D. F. Cully, H. S. Ip, and G. A. M. Cross, Cell 42:173–182, 1985). At least 3 kb of this expression site exhibits a high degree of homology with the silent telomere carrying the AnTat 1.3A basic copy, whose ESAG 1 is interrupted by stop codons. Like the antigen gene, the region containing the ESAGs is transcribed only in the bloodstream forms, although transcription of 5′ barren- and ESAG 2-related sequences also occurs in cultured procyclics. Analysis of steady-state and nascent transcripts suggests a continuous transcription of the whole expression site by an RNA polymerase resistant to α-amanitin, possibly initiating at a polymerase I-like promoter located about 17 kb upstream from the antigen gene. This polymerase seems prone to becoming inactivated upon incubation of the trypanosomes at low temperature. The putative protein encoded by ESAG 3 may carry a hydrophobic signal peptide, suggesting interaction with a membrane.

African trypanosomes evade the immune response of their mammalian host by repeatedly changing their antigenic specificity. The latter is determined by the selective expression of only one variant-specific glycoprotein (VSG) gene from a repertoire of several hundred specific sequences (for recent reviews, see references 3, 4, 9 and 30). The VSG gene is transcribed in a telomeric expression site, which, like other trypanosome telomeres, is characterized by the presence of long arrays of imperfect 76-base-pair (bp) repeats about 1.5 kilobases (kb) upstream from the gene, the 5′ barren region. The expression site may also contain at least one other gene (6, 7) and, possibly, pseudogenes (2, 28). The length of the antigen gene transcription unit seems to be variable. Several observations suggest that it may be very long (4), but recent data strongly suggest that at least in some instances, the transcription promoter might be relatively close to the gene (28).

In an attempt to characterize the DNA sequences required for the controlled expression of the antigen gene transcription unit, we cloned sequences upstream from the 5′ barren region of the expression site for the gene encoding one of the predominant surface antigens of the *Trypanosoma brucei* AnTAR 1 repertoire, AnTat 1.3A (17, 19). The expressed gene is a copy of the unique telomeric AnTat 1.3A basic-copy gene. It has been inserted by gene conversion in a telomere of a 200-kb chromosome, which is frequently used as an antigen gene expression site in the EATRO 1125 *T. brucei* stock (8).

We also cloned the corresponding sequences from the telomere harboring the AnTat 1.3A basic copy (19). This telomere is in a chromosome of the megabase class (12).

We report here the characterization and transcription properties of these sequences. Our results suggest that the expression of several genes is coordinated with that of the antigen gene. Their transcription seems to start at a polymerase I (pol I)-like promoter located about 17 kb upstream from the antigen gene.

* Corresponding author.

MATERIALS AND METHODS

The trypanosome clones AnTat 1.3A and 1.3B have been characterized elsewhere (17, 22; see Fig. 1 for an abridged pedigree). Trypanosomes were separated from blood cells by DEAE chromatography in phosphate-buffered glucose (16). Before the extraction of nuclei for in vitro transcription assays, the trypanosomes were suspended in the medium eluting through the DEAE, which contains blood components diluted about sevenfold in phosphate-buffered glucose. In this medium the trypanosomes can survive at room temperature for at least 3 days, but were used immediately for further analysis. For most transcription experiments, the isolation of trypanosomes, including the DEAE chromatography and pelleting of the cells, was performed at 37 to 38°C. The whole isolation procedure was routinely performed in less than 30 min.

The methods for DNA and RNA isolation are described by Pays et al. (23). The procedures used for cloning the expression site-associated sequences of the AnTat 1.3A variant, as well as sequences from the telomere harboring the AnTat 1.3A basic copy, have been described previously (19). In brief, sequences preceding the 5′ barren region of the AnTat 1.3A expression site have been cloned in the *Sph*I site of pBR322 by a jumping procedure to avoid an unclonable fragment in the 5′ barren region. This procedure involved ligation of partial *Sph*I digests of genomic AnTat 1.3A DNA with a pBR322 molecule from which the *Hin*dIII site had been deleted. The whole 5′ barren region and contiguous sequences were eliminated by *Hin*dIII digestion, and then the construct was resealed by ligation after extensive dilution. A probe from the AnTat 1.3A cloned cDNA allowed the selection of clones with integrated sequences from both the AnTat 1.3A gene and a region located 14 kb upstream. A similar procedure was applied to clone the telomere which contained the AnTat 1.3A basic copy, but from which the full 5′ barren region and upstream sequences could be isolated as a single clone (19). Following the initial cloning, further sequences from the expression site were picked up from pUC18 libraries containing *Hin*dIII and *Sph*I fragments of the 200-kb expressor chromosome, which was isolated by

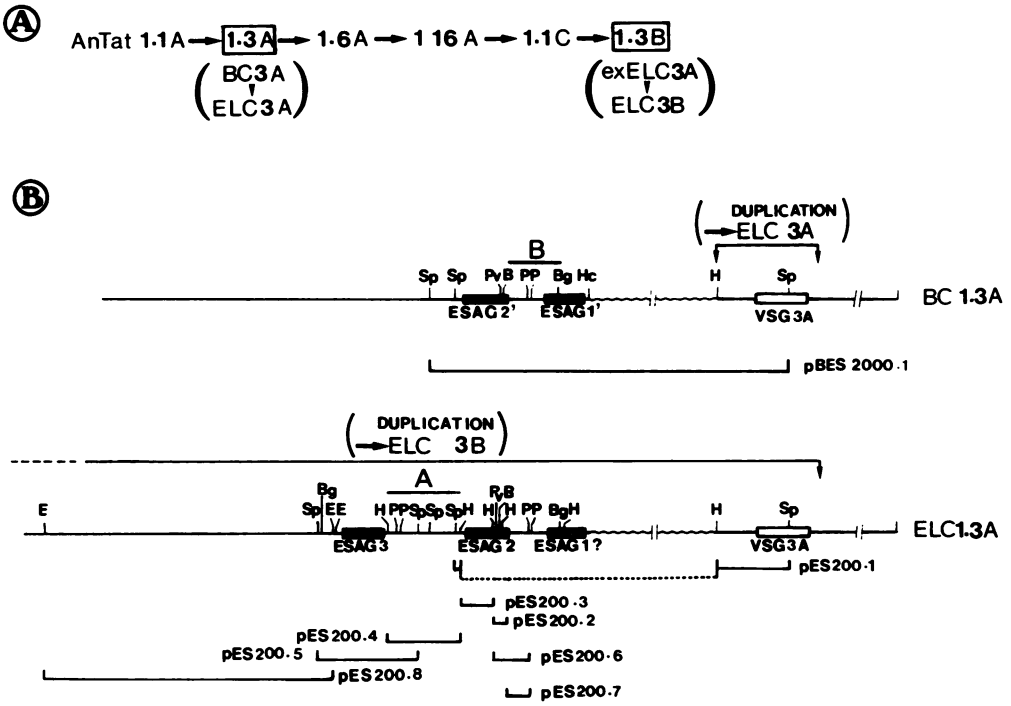


FIG. 1. Origin of the trypanosome variants and genomic DNA clones. (A) Pedigree of the *T. brucei* clones AnTat 1.3A and AnTat 1.3B. Clones of successive variant antigenic types were set up from heterotypes arising from the diversifying clone of the previous variant population. Previous analyses (17, 22) have shown that in variant 3A the active gene is a copy (ELC) of the AnTat 1.3A basic copy (BC), whereas in variant 3B the active gene is in a large duplicate of the lingering AnTat 1.3A ELC (exELC), conserved from the AnTat 1.3A clone. (B) Extent of the unique clone from the AnTat 1.3A basic copy-containing telomere (pBES 2000.1 [top]) and of eight clones from the AnTat 1.3A expression site (pES 200.1 to pES 200.8 [bottom]). With the exception of pES 200.1, pES200.8, and PBES 2000.1, which are pBR322 recombinants, all other clones are in pUC18 (19; see Materials and Methods). The pES 200.1 clone is discontinuous, as discussed in the text. Symbols: · · · ·, extent of the sequence deleted in the cloning procedure; ▭, AnTat 1.3A coding region; ▨, ■, other genes as discussed in the text; ∿∿∿, 5′ barren region. Fragments labeled A and B were used as probes for hybridization with the Southern blots shown in Fig. 2. The extent of the duplication-transpositions which generated the AnTat 1.3A and AnTat 1.3B ELCs is shown above each map. Abbreviations for restriction sites: B, *Bgl*I; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; Hc, *Hin*dII; P, *Pst*I; Pv, *Pvu*II; Sp, *Sph*I; T, *Taq*I. Only the *Taq*I and *Hin*dII sites closest to the 5′ barren region of pBES 2000.1 are indicated.

pulse-field gradient gel electrophoresis (27). Overlapping fragments have been identified by DNA sequencing. The alignment of nonoverlapping clones has been deduced from a comparison with the corresponding regions of the basic copy clone (see Results). Figure 1 shows the extent and location of the different clones obtained from both the expression site and the basic copy-containing telomere.

Specific probes were prepared from the cloned regions by subcloning restriction fragments in derivatives of the M13 bacteriophage (34).

The nucleotide sequence of DNA fragments was determined on both strands by the method of Sanger et al. (26). The method for Southern and Northern (RNA) blot hybridization have been described previously (23). S1 protection and primer extension experiments were performed as described by Gilmour (11) and Colman (5), respectively. Run-on transcription assays were conducted as described by Murphy et al. (19).

RESULTS

Specificity of the cloned sequences. The procedure used to clone sequences from the AnTat 1.3A expression site involved the deletion of a region refractory to cloning in plasmid vectors (19). Two categories of clones were selected by hybridization with an AnTat 1.3A cDNA probe. According to their restriction maps, these clones could be ascribed either to the AnTat 1.3A expression site or to the basic copy-containing telomere (19). Using the first cloned expression site sequence as a probe, we obtained several other clones, spanning 6 kb, from *Hin*dIII and *Sph*I banks of the isolated 200-kb expressor chromosome (Fig. 1). The origin of the sequences attributed to the expression site could be further verified by hybridization with AnTat 1.3A and 1.3B genomic DNAs. Indeed, in the AnTat 1.3B clone, a large region from the AnTat 1.3A expression site is duplicated (22), so that fragments of that region are twice as abundant in Southern blots of genomic AnTat 1.3B DNA compared with AnTat 1.3A DNA. Figure 2 shows typical results of such an analysis. Different probes from the cloned region reveal genomic fragments expected from the restriction map of the clones, and these fragments appear more abundant in AnTat 1.3B (arrowheads in Fig. 2). These results confirm that the selected clones are specific to the AnTat 1.3A expression site.

Nucleotide sequence of the cloned area. Figure 3 shows the nucleotide sequence of the 6-kb region cloned from the AnTat 1.3A expression site. This sequence contains two large open reading frames (ORFs) of 1,104 and 1,263 bp (positions 603 to 1706 and 3974 to 5236, respectively) in the same orientation as the VSG gene. Starting at the first codon

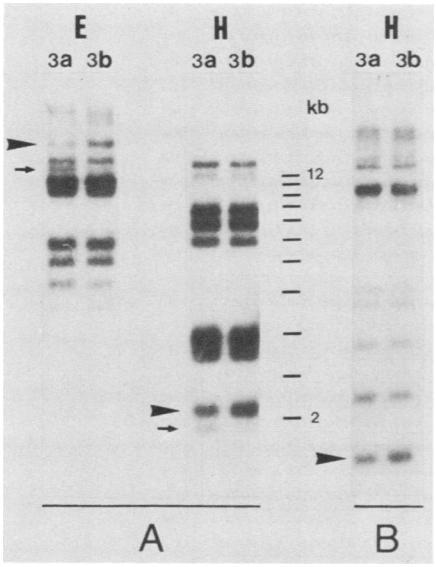


FIG. 2. The cloned region from AnTat 1.3A is duplicated in AnTat 1.3B, as expected for sequences from the expression site. The genomic DNA from variants AnTat 1.3A and AnTat 1.3B was digested by *Eco*RI or *Hin*dIII and hybridized to probes A and B as indicated (see Fig. 1 for the extent and location of the probes). Post-hybridization washes were at high stringency (0.1× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate] at 65°C). The arrowheads point to fragments duplicated in AnTat 1.3B DNA. This duplication is expected for sequences of the VSG expression site (22). The sizes of these fragments are as expected from the restriction map of the cloned sequences (Fig. 1). Fragments marked with a small arrow are lost in AnTat 1.3B DNA, as expected for sequences replaced by the AnTat 1.3A additional copy.

for methionine, these ORFs may be translated into polypeptides of 367 and 403 amino acids, respectively. These main ORFs are separated by a sequence carrying two shorter ORFs (positions 1863 to 2108 and 2512 to 3019). The first has a methionine codon at the beginning of its sequence; it would encode only a short polypeptide (78 amino acids). The second, which is not in the same frame as the first, does not exhibit a methionine codon in its 5′ region. According to their transcription characteristics (see below), the ORFs starting at positions 603 and 3974 have been named ESAG 3 and ESAG 2, respectively, for expression site-associated genes (7).

The second half of ESAG 3 (underlined in Fig. 3), exhibits 80% nucleotide homology with the cotransposed sequence of the AnTat 1.1-expressed gene copy (companion sequence [25]).

Segments of the sequence cloned from the telomere carrying the AnTat 1.3A basic copy have also been sequenced and compared with those of the expression site (Fig. 4). It appears that the corresponding regions from the two telomeres are highly homologous. In particular, a sequence similar to ESAG 2 exists in the basic-copy telomere.

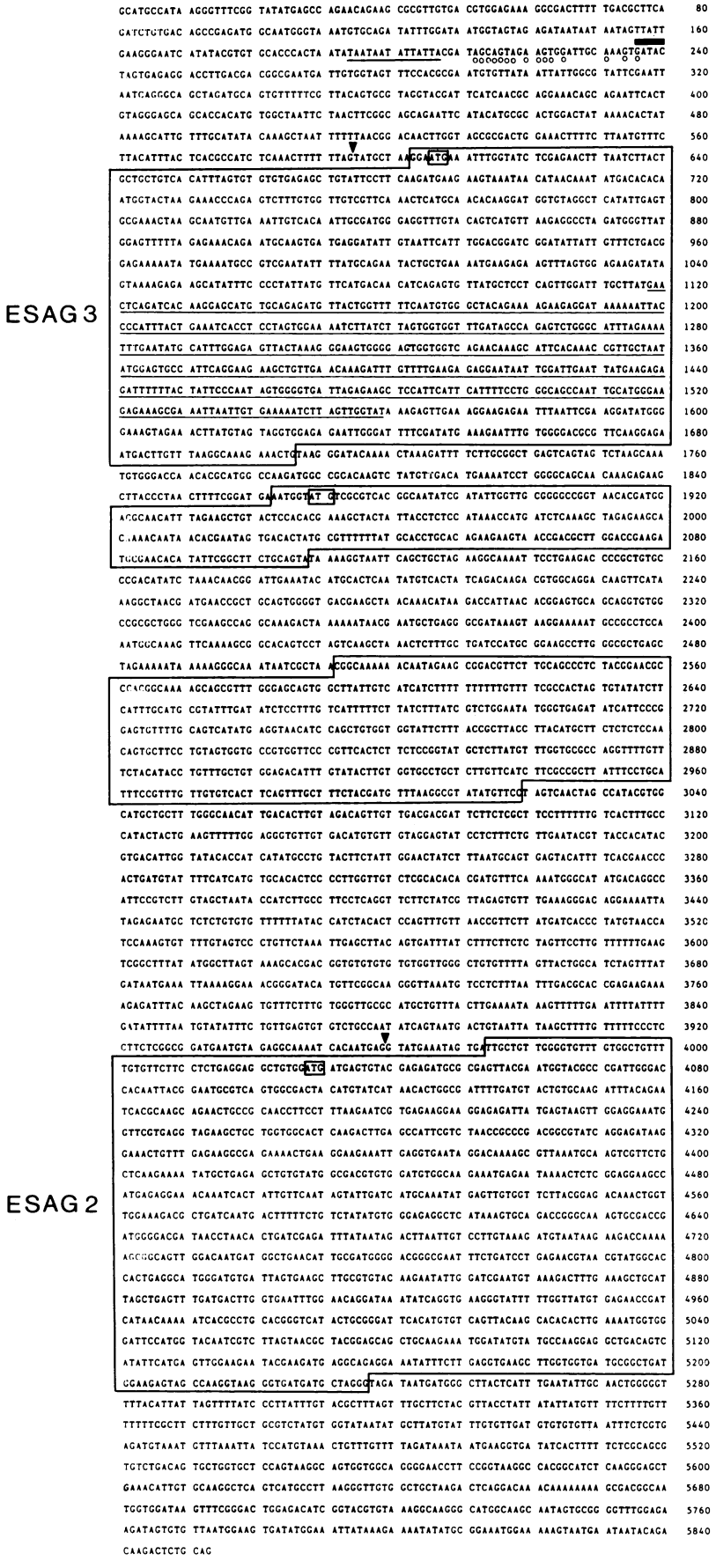
The region immediately upstream of the 5′ barren region, which could not be cloned from the expression site, was cloned from the basic-copy telomere. A comparison of this sequence with that published by Cully et al. (7) shows homology to ESAG 1 (Fig. 4). However, the ESAG 1-related sequence from the basic-copy telomere is interrupted by stop codons. With a probe extending to the ESAG 1-related sequence from the basic-copy telomere (the ESAG 1 copy is characterized by a *Bgl*II site), hybridization to Southern blots of genomic DNA shows that an ESAG 1 copy also exists in the expression site, at the same location with respect to both ESAG 2 and the 5′ barren region (Fig. 1 and 2 and data not shown).

Mapping of the steady-state transcripts. Cloned sequences from both the expression site and the basic-copy telomere were used as hybridization probes on poly(A)$^+$ RNA from bloodstream and procyclic forms of *T. brucei*. Figure 5 shows the locations of these probes, as well as a scheme summarizing the results. The data are presented in Fig. 6. All probes clearly recognize steady-state transcripts, except probe IX, which hybridizes only weakly to the blots. Six or seven main transcripts, distinct from the VSG mRNA, are observed. These transcripts cover virtually the whole cloned region, except the barren region. Moreover, the presence of several larger transcripts, some of which seem common to different probes (for instance, the 2- and 3-kb transcripts with probes VI to VIII), suggests that the main RNAs are processed from larger precursors. In addition, it can be seen that although most of the probes hybridize only to RNA from bloodstream forms, probe V (ESAG 2) also hybridizes to transcripts from procyclic forms.

Additional data about the origin of the transcripts were obtained by two approaches: S1 protection and primer extension. Cloned DNA fragments from the AnTat 1.3A expression site can be protected from digestion by S1 nuclease by hybridization with AnTat 1.3A RNA, but not AnTat 1.6C or procyclic RNA (Fig. 7). The same extent of protection is achieved after hybridization with AnTat 1.1B RNA, which is synthesized in the same expression site as AnTat 1.3A (12; results not shown). These data confirm that the cloned sequences are specific to the AnTat 1.3A expression site. In addition, measurement of the extent of their protection allows the approximate mapping of the 5′ extremity of the transcripts. Fragments 1 and 4 (Fig. 5) are protected over 340 and 850 nucleotides, respectively (Fig. 7; data not shown), mapping the beginning of ESAG 3 and ESAG 2 RNAs at, respectively, 10 and 70 nucleotides upstream from the translation initiation codon. Fragment 1 shows an additional protection over about 700 bp, which could be due to S1 hypersensitivity in an AT palindrome at positions 190 to 210 (Fig. 3). These results were confirmed by primer extension analysis (Fig. 8). After hybridization with synthetic oligonucleotide primers specific for each ESAG, reverse transcription of AnTat 1.3A and 1.6C RNAs gave rise to the synthesis of cDNA molecules of a size expected from S1 protection analysis; these cDNA molecules were not observed in experiments with RNA from procyclic trypanosomes (Fig. 8 and scheme in Fig. 5). In some instances and on long exposure, the cDNAs synthesized on total Antat 1.3A RNA from a primer starting at position 611 show, in addition to the cDNA ending at the 5′ extremity of the mature ESAG 3 mRNA (50 nucleotides), a discrete molecule of 370 nucleotides which might map the transcription start around position 240 (Fig. 3).

Both S1 protection and primer extension experiments reveal stable transcripts, apparently beginning at several loci, in the region between ESAG 3 and ESAG 2 (Fig. 5E, thin arrows). On the other hand, no conspicuous cDNA was synthesized following primer extension from position 101 (Fig. 8), suggesting that a region located about 0.5 kb upstream of ESAG 3 is not transcribed in stable RNAs.

Presence of spliced leader sequence on ESAG transcripts. On close examination of the data in Fig. 7 and 8, we found that the 5′ limit of ESAG 3 and ESAG 2 transcripts differs by about 40 nucleotides, according to whether the estimates are


ESAG3
ESAG2

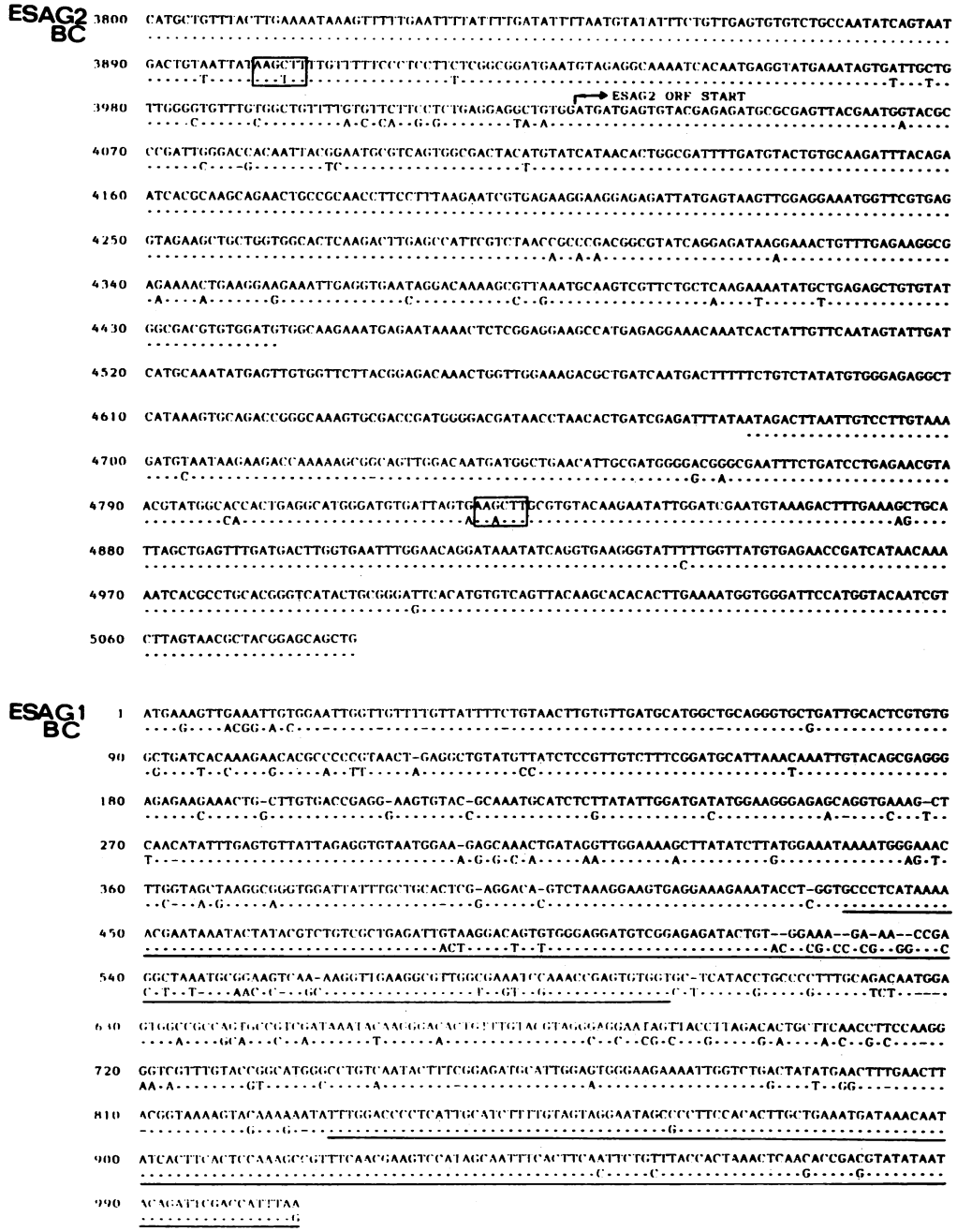


FIG. 4. Comparison of sequences from the telomere containing the AnTat 1.3A basic copy with those of different VSG expression sites. Sequenced areas of the AnTat 1.3A basic copy-containing telomere (BC) are aligned with those from the AnTat 1.3A expression site (ESAG 2) and the published sequence of ESAG 1 (7). Identical nucleotides are indicated by dots, and deletions (−) are introduced to maximize homology. The two boxed *Hin*dIII sites are present in the expression site, but absent from the basic copy-containing telomere. These sites mark boundaries between different clones from the expression site (Fig. 1). The comparison with the sequence of the clone from the basic-copy telomere allows us to align these clones. The longest ORFs of the ESAG 1 copy from the BC telomere are underlined.

FIG. 3. Nucleotide sequence of the cloned region from the AnTat 1.3A expression site. The ORFs, as well as their putative translation initiation codon, are boxed. The 3′ splice sites of the main steady-state RNAs are indicated by arrowheads. A bar above the sequence at positions 236 to 240 indicates where transcription seems to start. A 14-bp AT palindrome, located about 35 nucleotides upstream from the transcription start, is underlined. Dots under some nucleotides indicate homology with the promoter sequence for rDNA transcription (see Fig. 11). The underlined region in ESAG 3 is 80% homologous to the Antat 1.1 companion sequence (25).

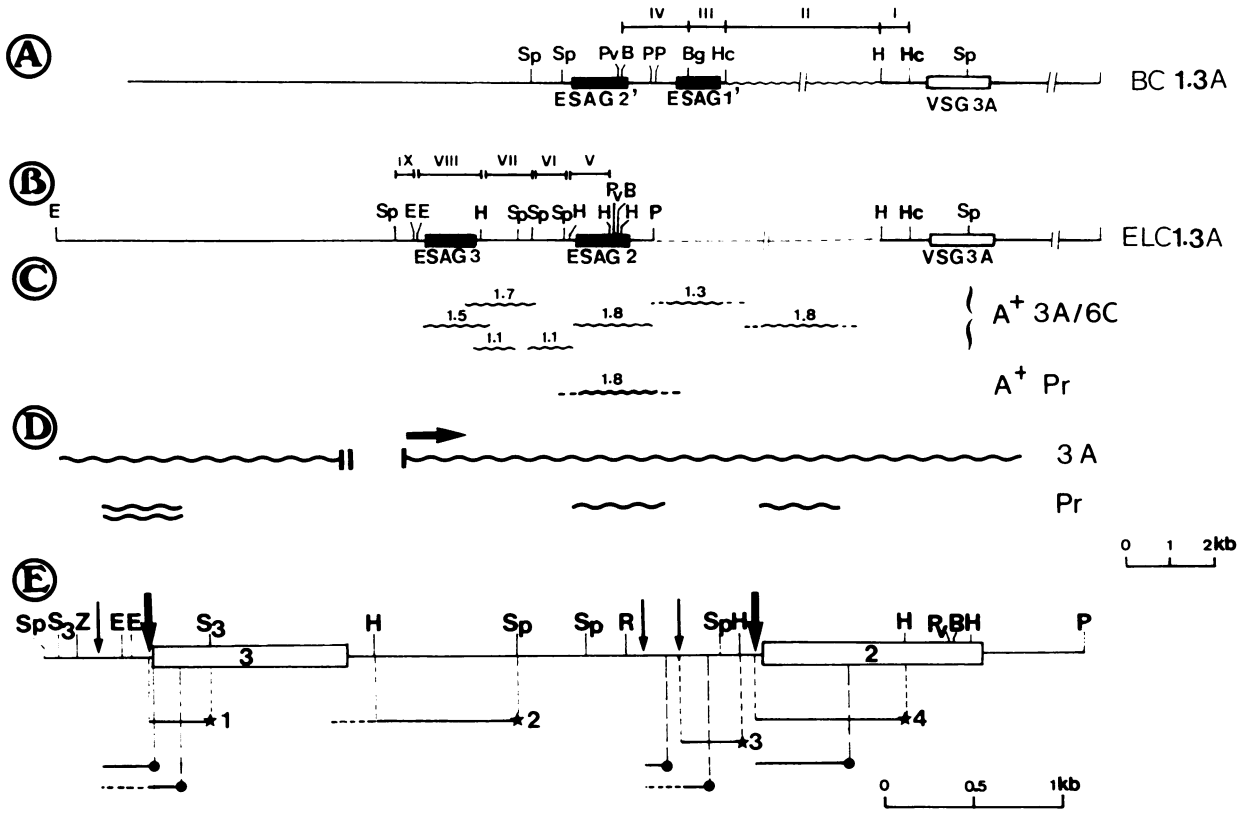


FIG. 5. Scheme summarizing the transcriptional properties of the AnTat 1.3A expression site. Restriction maps of the telomeres harboring the AnTat 1.3A basic copy and expression site are presented in lines A and B, respectively. Symbols: ▭, VSG gene; ▬, ESAGs as indicated; ∿∿, 5′ barren region; - - - -, extent of the uncloned sequence from the AnTat 1.3A expression site. Above each map is shown the extent of the probes used for hybridization with Northern blots (see Fig. 6). Line C shows an interpretation of the patterns of steady-state transcripts observed by Northern blot hybridizations, and line D summarizes the results of the in vitro transcription assays. The arrow indicates the direction of transcription. The scheme in line E summarizes the results of the S1 protection and primer extension experiments. The lines labeled with stars, 1 to 4, correspond to the S1-protected regions of the 0.85-kb *Sau*3A, 0.8-kb *Hin*dIII-*Sph*I, 0.6-kb *Rsa*I-*Hin*dIII, and 0.9-kb *Hin*dIII fragments, respectively. The lines labeled with dots correspond to the cDNA molecules observed following primer extension on RNA; the dots show the locations of the primers. The thick arrows show the main 3′ splice sites of the ESAG mRNAs, and the thin arrows show discrete cDNA starts in the primer extension experiments. The abbreviations are the same as for Fig. 1, except for the following: A^+, poly(A)$^+$ RNA; Pr, procyclic; BC, basic copy; ELC, expression-linked copy; R, *Rsa*I; S3, *Sau*3A; Z, *Ssp*I. Only the relevant restriction sites are shown.

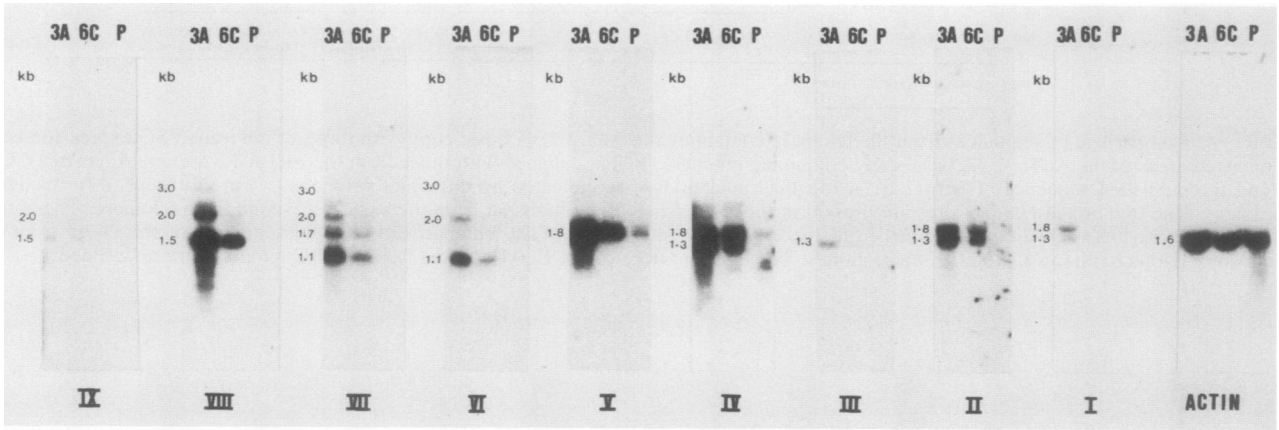


FIG. 6. Steady-state transcripts of the AnTat 1.3A expression site. Northern blots of poly(A)$^+$ RNAs from variants AnTat 1.3A and AnTat 1.6C, as well as from AnTat 1.1D-derived procyclic forms, have been hybridized with various probes from the AnTat 1.3A expression site or from the *T. brucei* actin gene, as indicated. All exposures were for about the same time, except for the blots hybridized with probes IV, V, and IX, which are fivefold overexposed. See Fig. 5 for the extent and location of probes I to IX and Ben Amar et al. (1) for the characterization of the actin probe.

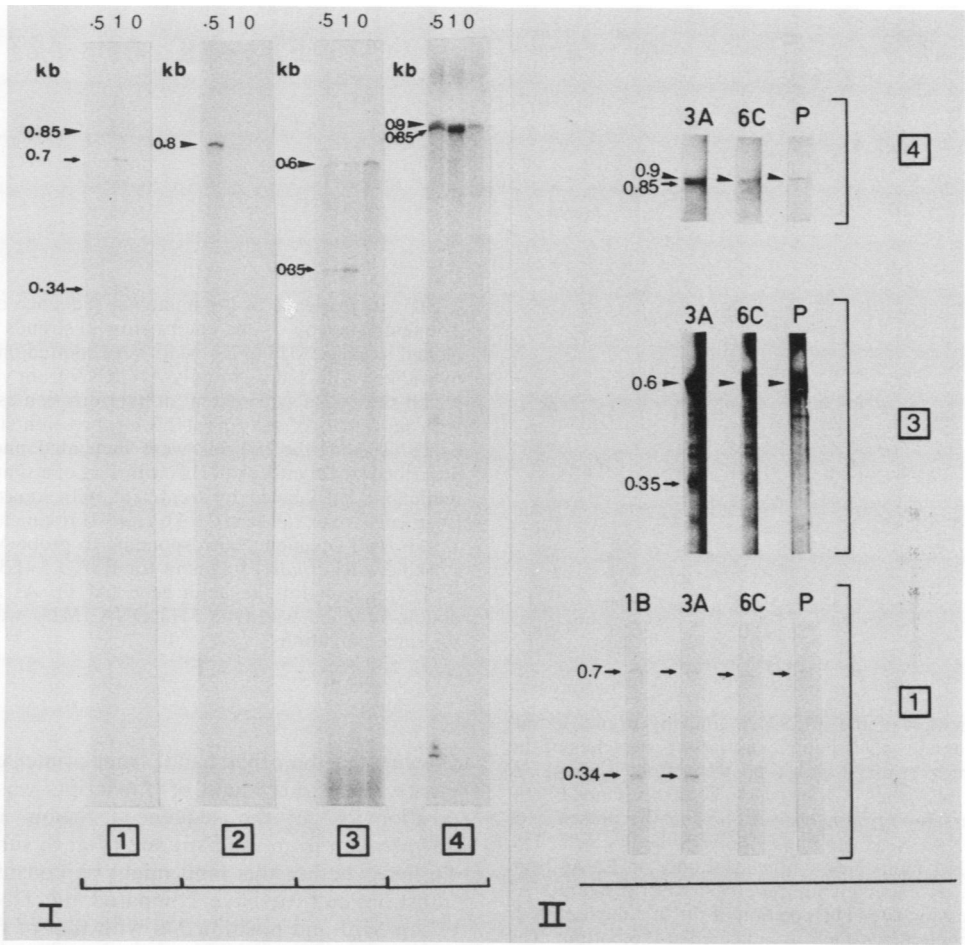


FIG. 7. Determination of the mRNA 5′ limit by S1 protection. (I) The ^{32}P-labeled 0.85-kb *Sau*3A, 0.8-kb *Hin*dIII-*Sph*I, 0.6-kb *Rsa*I-*Hin*dIII, or 0.9-kb *Hin*dIII fragment from the expression site (lanes 1, 2, 3, and 4, respectively; see Fig. 5) was hybridized with AnTat 1.3A poly(A)$^+$ RNA (0.5, 1, or 0 μg, as indicated above the lanes), treated with 72 U of S1 nuclease, and electrophoresed in a 6% polyacrylamide gel. The size was estimated with *Msp*I-digested pBR322 DNA as the marker, ^{32}P labeled by treatment with kinase. Bands indicated by small arrows correspond to the sequences protected by hybridization, whereas those indicated by arrowheads are full-sized fragments. (II) Fragments 1, 3, and 4 were hybridized with 1 μg of poly(A)$^+$ RNA from AnTat 1.1B, AnTat 1.3A, AnTat 1.6C, or procyclic forms, as indicated, and then treated with 72 U of S1 nuclease. Subsequent analysis was performed as described for panel I.

made by S1 mapping or by primer extension. This suggests that these mRNAs are capped by a 39-nucleotide spliced leader (or miniexon) sequence. This was indeed verified by hybridization of a synthetic probe of the miniexon with hybrids between cloned ESAG 2 and ESAG 3 sequences and poly(A)$^+$ RNA (Fig. 9), which indicated that the transcripts of the two ESAGs carry the miniexon. The differences in hybridization levels, which is also observed in Northern blots, could be due to different proportions of the specific transcripts.

Transcription in vitro. To estimate the extent of transcription along the cloned region of the AnTat 1.3A expression site, we allowed nascent RNA molecules to elongate in isolated AnTat 1.3A nuclei in the presence of [^{32}P]GTP and used them as probes on Southern blots of the cloned sequences. The results (Fig. 10) suggest that nearly all the cloned section, both upstream and downstream of the 5′ barren region, is transcribed. All the restriction fragments hybridize at comparable levels to the RNA probes, except those specific to ESAG 2 and the 5′ barren region, which hybridize much more strongly, and fragments upstream from the *Eco*RI sites, which hybridize only weakly (Fig. 10B, digest H+E+Sp, band 4; digest E+Pv, band 6) or not at all (Fig. 10B, digest H+Z, band 5). In the presence of 1 mg of α-amanitin per ml, most hybridization signals are only slightly reduced (about 30%), exactly as observed with transcription of ribosomal genes (Fig. 10B, fragments I). However, transcription of both ESAG 2 and the 5′ barren regions appears strongly inhibited by the drug. Unexpectedly, hybridization to ESAG 2 and 5′ barren region-specific fragments can also be detected with transcripts from procyclic nuclei. These results suggest that besides a variant-specific transcription of the cloned region of the expression site by an RNA polymerase resistant to α-amanitin, sequences related to ESAG 2 and the 5′ barren region are constitutively transcribed by an RNA polymerase sensitive to the drug.

For all experiments described above, the trypanosomes were isolated from blood components by DEAE chromatography at 37 to 38°C (see Materials and Methods). When trypanosomes are isolated at room temperature, the hybridization of nascent RNA to sequences of the antigen gene expression site appears selectively reduced (Table 1).

Putative pol I-like promoter 17 kb upstream from the

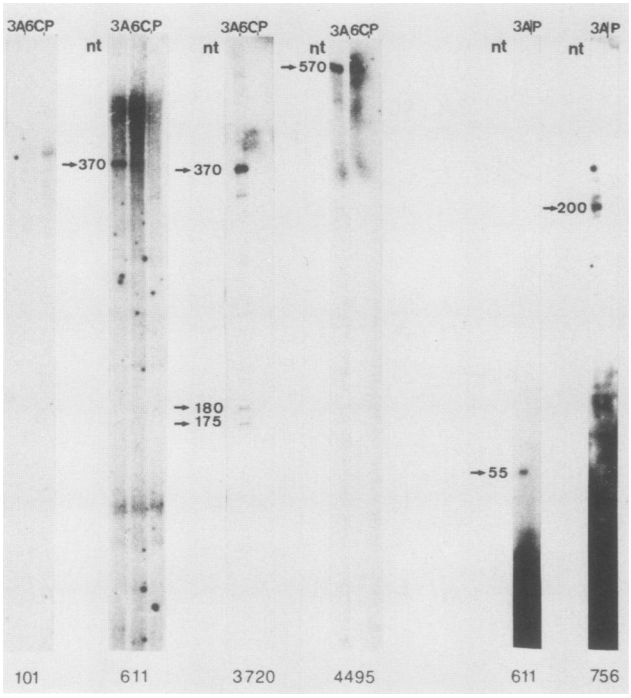


FIG. 8. Determination of the mRNA 5′ limit by primer extension. Synthetic oligonucleotides complementary to stretches of the expression site were hybridized with total RNA from AnTat 1.3A (lanes 3A), AnTat 1.6C (lanes 6C), or procyclic forms (lanes P) and then used as primers for reverse transcription in the presence of [^{32}P]dCTP. The cDNAs were run on 6% polyacrylamide gels. The arrows point to cDNA molecules synthesized only on RNAs from bloodstream forms, as expected for reverse transcripts of RNAs from the VSG expression site. The position of the first nucleotide 5′ to each primer (20 nucleotides each) is indicated at the bottom. The apparent cDNA in the 101 extension of procyclic RNA is probably an artifact, since it has not been reproducibly observed. The cDNAs primed by 611 extension were run for different periods and exposed for different times to see the two discrete stops at 55 and 370 nucleotides, respectively.

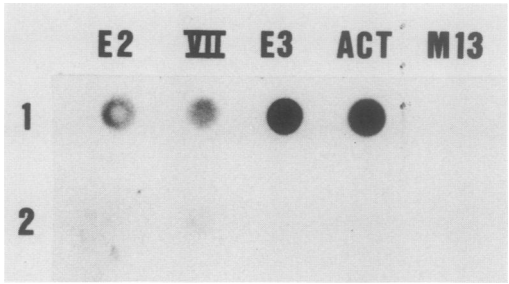


FIG. 9. Presence of the miniexon sequence on ESAG mRNAs. The template (row 1) and coding (row 2) strands of DNA fragments cloned in phage M13 (4 μg each) were spotted on nitrocellulose and hybridized with 150 μg of poly(A)$^+$ RNA from variant AnTat 1.1B, which shows the same extent of hybridization as AnTat 1.3A RNA with sequences of the AnTat 1.3A expression site (see text). After being washed, the hybrids were incubated under the conditions described by Guyaux et al. (12) with 1 μg of a synthetic 35-mer of the miniexon, ^{32}P labeled by treatment with kinase. The hybridized fragments are as follows: 0.9-kb *Hin*dIII fragment for ESAG 2 (E2); 1.3-kb *Rsa*I fragment, corresponding to probe VII of Fig. 5 (VII); 1.4-kb *Eco*RI-*Hin*dIII fragment for ESAG 3 (E3); and 1.9-kb *Sal*I-*Bgl*II fragment for the actin gene (ACT). As a control, 4 μg of double-stranded wild-type M13 DNA (M13) was incubated under the same conditions.

antigen gene. A stretch of 200 bp at the 5′ extremity of the sequenced region appears not to be transcribed, since Northern blot analyses with probes restricted to this region do not show any transcript (not shown). Moreover, DNA fragments located upstream from the *Ssp*I site at position 201 do not hybridize with in vitro-labeled nascent RNA, whereas all other fragments hybridize at comparable levels (Fig. 10; data not shown). The untranscribed region seems to extend at least 550 bp upstream from this *Ssp*I site (Fig. 10B; data not shown). On the other hand, from primer extension experiments, the 5′-most limit of transcripts was mapped around position 240 of the sequenced region, i.e., only 40 bp downstream from the *Ssp*I site (Fig. 8; summary in Fig. 5). To find whether this limit might be considered the start of transcription, we have compared the DNA sequence between *Ssp*I and position 240 with that of the promoters for rDNA and other antigen gene transcription unit (Fig. 11). A significant level of homology can be found among the three sequences, strongly suggesting the presence of a pol I-like promoter. No other region from the sequenced fraction of the AnTat 1.3A expression site was found to contain a similar element.

Amino acid sequence of ESAG translation products. The translation sequences of the two main ORFs (ESAG 2 and ESAG 3) are presented in Fig. 12. These polypeptides have been compared with sequences in the GenBank and EMBL databases, but no significant homologies could be found. The only remarkable feature is the presence, at the beginning of the ESAG 3 polypeptide, of a hydrophobic segment with a putative cleavage point, which can be interpreted as a signal peptide (boxed in Fig. 12).

FIG. 10. Extent of transcription in vitro. Restriction digests from the cloned regions of the basic copy-containing telomere (A) and expression site (B) were hybridized with ^{32}P-labeled RNA synthesized in isolated nuclei from AnTat 1.3A bloodstream forms (lanes B) or procyclic forms (lanes P). Transcription of bloodstream nuclei was conducted in the presence of 1 mg of α-amanitin per ml (lanes 1) or without α-amanitin (lanes 0), whereas procyclic nuclei were incubated without α-amanitin. The DNA fragments, shown on the left in each panel after being stained with ethidium bromide, may be identified by numbers in the maps below. ▬, DNA from the plasmid vectors. Encircled numbers refer to fragments containing the ESAG 2 gene, and fragments marked by a star are thought to contain untranscribed DNA. Fragments labeled 1′ in panel A are from a minor population of the recombinant plasmids, in which a fraction of the 5′ barren region has been deleted. Fragments labeled 4′ in panel A are due to an incomplete cut at the *Hin*dII site bracketed in the map. The results of four independent experiments are shown, accounting for the variability of hybridization level in different digests. For *Hin*dIII-*Eco*RI-*Sph*I digests, the results of two experiments are presented, to better show the relative extent of hybridization of band 4. As a control, panel B shows the hybridization pattern of an *Eco*RI-*Hin*dIII digest of a lambda 1059 clone containing an rDNA transcription unit (Pays, unpublished results). Fragments I contain rDNA. Fragment II contains a RIME transposon, which is transcribed by a pol II-like enzyme, much more in bloodstream forms than in procyclic forms (20). The two other hybridizing fragments span both rDNA and vector sequences. Abbreviations for restriction sites: B, *Bgl*I; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; Hc, *Hin*dII; P, *Pst*I; Pv, *Pvu*II; Sp, *Sph*I; Z, *Ssp*I.

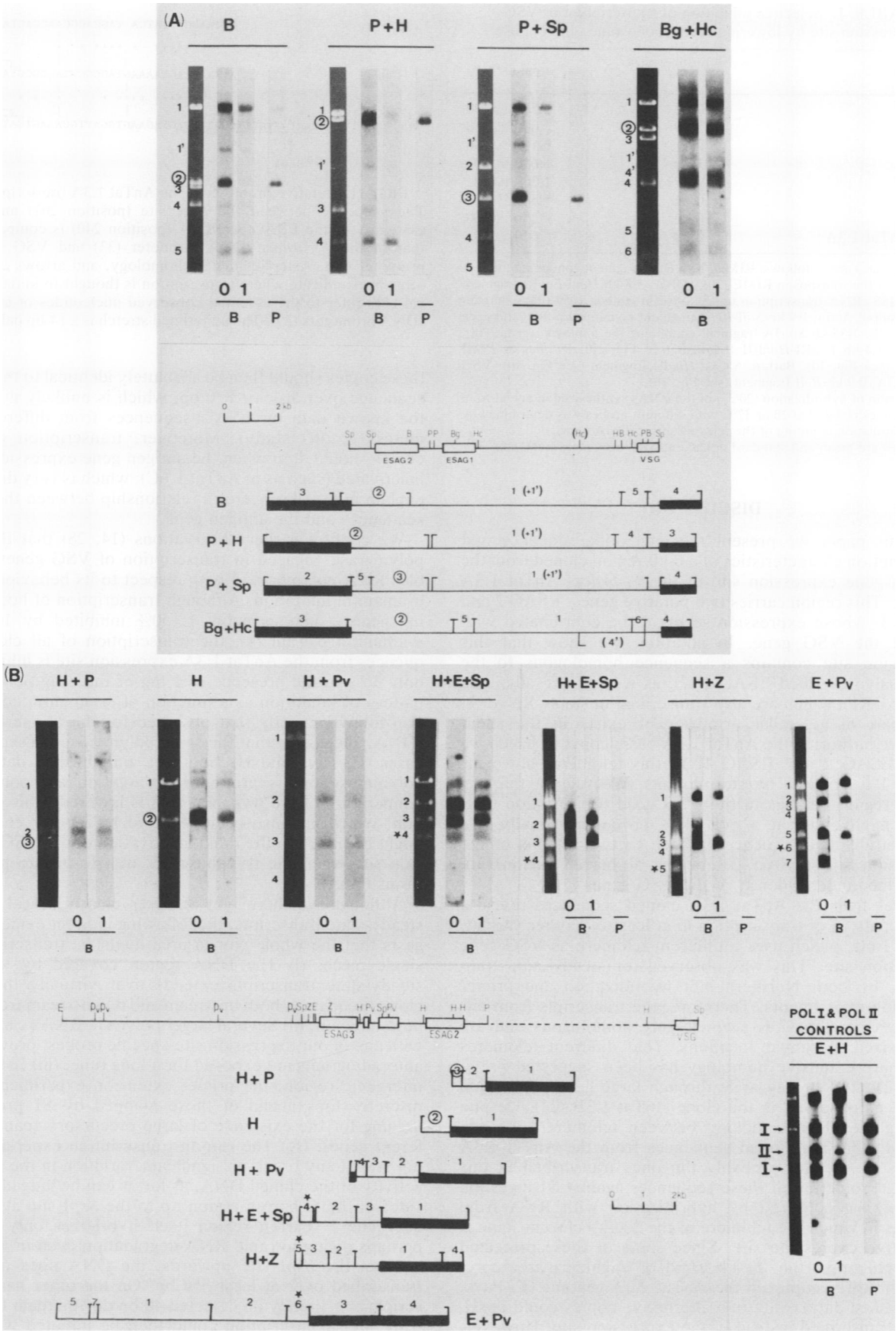

(A)
B
P+H
P+Sp
Bg+Hc
0 1
B
P
ESAG2
ESAG1
VSG
B
P+H
P+Sp
Bg+Hc
(B)
H+P
H
H+Pv
H+E+Sp
H+E+Sp
H+Z
E+Pv
ESAG3
ESAG2
VSG
H+P
H
H+Pv
H+E+Sp
H+Z
E+Pv
POLI & POL II
CONTROLS
E+H
I→
II→
I→

TABLE 1. Selective inhibition of transcription of VSG expression site by lowering the tryponosome temperature

Probe[a]	Hybridization ratio (20°C/37°C)[b]	cpm (37°C)[c]
RIME	0.87	1,500 ± 100
rDNA	1.09	2,000 ± 250
Actin	0.91	1,100 ± 100
SL	1.29	1,650 ± 50
ESAG 3	0.67	1,200 ± 60
Pv	0.62	560 ± 30
Barren	0.42	250 ± 100
VSG AnTat 1.3A	0.38	650 ± 50

[a] The probes are as follows: RIME, 1-kb *Rsa*I fragment containing the B + A halves of the transposon RIME (20); rDNA, 5.8-kb *Hin*d-*Eco*RI fragment from an 18S rRNA transcription unit, cloned in lambda 1059 (Pays, unpublished results); Actin, 1.9-kb *Sal*I-*Bgl*II fragment containing a full actin gene copy (1); SL, 1.35-kb *Sau*3A fragment containing the miniexon repeat (18); ESAG 3, 1.4-kb *Eco*RI-*Hin*dIII fragment (see Fig. 10); Pv, 0.6-kb *Pvu*II fragment (see Fig. 10); Barren, 5.4-kb *Hin*dII fragment (see Fig. 10); VSG AnTat 1.3A, 0.9-kb *Bgl*I fragment (see Fig. 10).

[b] The ratio of hybridization (20/37) of the RNAs synthesized in nuclei from trypanosomes isolated at 20 or 37°C with different probes was determined by liquid scintillation counting of the relevant nitrocellulose area.

[c] Values are mean and standard deviation from three measurements.

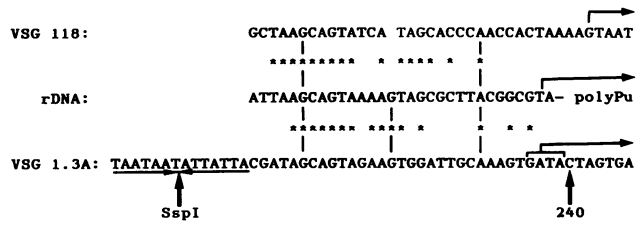


FIG. 11. Putative promoter of the AnTat 1.3A transcription unit. The sequence between the *Ssp*I site (position 201) and the 5′ extremity of the ESAG 3 mRNA (position 240) is compared with that of the *T. brucei* rDNA promoter (33) and VSG 118 gene promoter (28). Asterisks denote homology, and arrows above the sequences indicate where transcription is thought to initiate. Vertical bars refer to the essential conserved nucleotides of eucaryotic rDNA promoters (29). The underlined stretch is a 14-bp palindrome.

DISCUSSION

In this paper we present the nucleotide sequence and transcription characteristics of a 6-kb region cloned from the antigen gene expression site of the *T. brucei* AnTat 1.3A variant. This region carries two putative genes, ESAG 2 and ESAG 3, whose expression seems to be coordinated with that of the VSG gene. In addition, we show that this expression site contains a sequence homologous to the previously described ESAG 1 (7), as well as two adjoining shorter ORFs which are also transcribed in stable RNA. At least part of a similar arrangement exists in the silent telomere harboring the AnTat 1.3A basic copy, in which we found ESAG 2 and ESAG 1. In this telomere, however, ESAG 1 is crippled by stop codons. We suggest that the ESAG region belongs to the VSG gene transcription unit, which might start at a pol I-like promoter. Finally, we demonstrate that sequences related to two regions of the expression site (ESAG 2 and the 5′ barren region) are transcribed independently of the VSG gene.

Probes from the AnTat 1.3A cloned sequences reveal a similar pattern of transcription in at least one other variant, AnTat 1.6C, which uses a different telomere as VSG gene expression site. This was observed for steady-state transcripts, by both Northern blot hybridization and primer extension experiments. Therefore, the transcripts from different expression sites seem largely homologous and are synthesized at similar locations. That different telomeres may share extensive homology has been suggested previously (21, 24). It may arise through large gene conversion events, as observed in the clone AnTat 1.3B (22). Despite this high level of homology between telomeres, we can conclude that the cloned sequences from the AnTat 1.3A expression site are probably the ones transcribed in this variant. Protection of these sequences against S1 digestion can only be achieved by hybridization with RNA from variants in which the telomere of the 200-kb chromosome is the active expression site. Since some of these protected fragments (e.g., the 0.9-kb *Hin*dIII fragment) are represented in three copies in the AnTat 1.3A genome (E. Pays, unpublished data), only two alternative copies could possibly be transcribed instead of the expression site. However, these copies should then be absolutely identical to the cloned sequence over at least 850 bp, which is unlikely in view of the known data on ESAG sequences from different telomeres (7; this study). Moreover, transcription of these copies should stop when the antigen gene expression site is inactivated (such as in AnTat 1.6C), which is very difficult to explain in the absence of a relationship between the ESAG sequences and the antigen gene.

We confirm earlier observations (14, 28) that the RNA polymerase engaged in transcription of VSG genes resembles RNA polymerase I with respect to its behavior toward α-amanitin inhibition. Although transcription of housekeeping genes, such as actin, is 50% inhibited by 10 μg of α-amanitin per ml (1), the transcription of all cloned sequences from the AnTat 1.3A expression site is inhibited by only 30% in the presence of 1 mg of the drug per ml. The pattern of inhibition as a function of α-amanitin concentration follows exactly that observed for the transcription of rDNA, suggesting that the same polymerase is used in both cases (E. Pays and H. Coquelet, unpublished data). This polymerase seems prone to inactivation as soon as the temperature of the trypanosomes is lowered. This observation, which has also been reported by Kooter et al. (15), might be linked to the rapid inactivation of the VSG expression site when the trypanosomes transform into procyclic forms (10).

Although Northern blot analyses clearly reveal discrete steady-state transcripts, the following indirect evidence suggests that the whole cloned area might be transcribed in a single piece. (i) The DNA region covered by the main steady-state transcripts extends over virtually the whole cloned sequence, both upstream and downstream from the 5′ barren region. (ii) Several larger poly(A)$^+$ RNAs can be seen with any of our expression site-specific probes, provided the autoradiograms are exposed for a long time. (iii) In the main intergenic region, the primer extension experiments show discrete stops ahead of those mapped by S1 protection, arguing for the existence of large precursors spanning different genes. (iv) The run-on transcription experiments do not reveal any break or significant variation in the template activity of the cloned DNA, as far as can be judged on both sides of the 5′ barren region up to the *Ssp*I site at position 201. The 5′ barren region itself hybridizes only weakly, perhaps owing to rapid RNA degradation even in vitro.

From the *Ssp*I site upwards, the DNA does not seem transcribed over at least 550 bp. On the other hand, transcripts can already be detected 40 bp downstream from this *Ssp*I site. Transcription could then be initiated within the

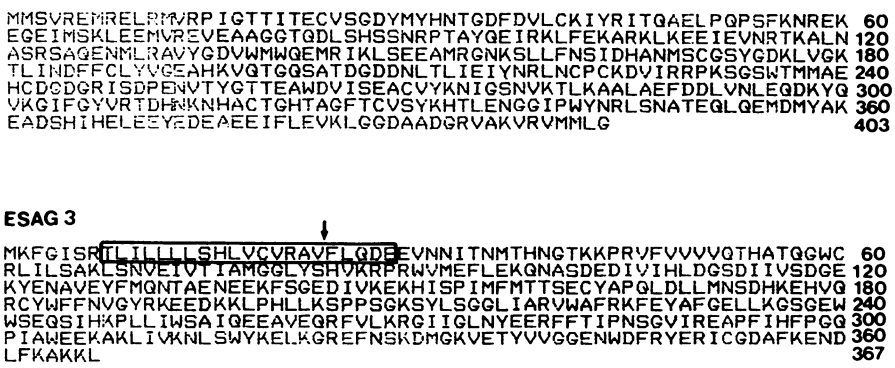


FIG. 12. Amino acid sequence of ESAG 2 and ESAG 3. The putative signal peptide of ESAG 3 is boxed, with the predicted site of cleavage (32) indicated by an arrow.

40-bp sequence following the *Ssp*I site. Analysis of this sequence has revealed the presence of a putative pol I-like promoter, which carries the consensus essential nucleotides at the correct location with respect to the 5′ limit of the transcribed domain. This observation suggests that transcription of the AnTat 1.3A expression site starts at that locus. The possibility that the promoter for transcription of the VSG gene is far upstream from the gene was suggested previously (4, 31) and is strongly supported by recent results (13, 15, 28). However, the two available estimates of the size of different transcription units in independent expression sites are widely divergent: 60 kb (13, 15) and 6 kb (28). Our estimate, around 18.5 kb, is between these two values. Further comparison between clone 221.4 of Kooter et al. (15) and our clone pES200.8 reveals identical restriction maps for all mapped sites, except for the presence of an additional *Bgl*II site near the 3′ end of our clone (Fig. 5 and 10; E. Pays, unpublished data). This difference might be significant, since 221.4 has been reported to be fully transcribed, whereas pES200.8 contains a 0.5- to 1-kb untranscribed region which precisely carries the additional *Bgl*II site. In addition, like Shea et al. (28), we have noted that sequences upstream from the putative promoter are also transcribed by a pol I-like RNA polymerase. Such a transcription occurs over at least 15 kb upstream from the untranscribed stretch (Pays, unpublished data). Other pol I-like promoters, whose activity would be under the same control as that of the VSG gene, might thus be present upstream.

Data from both in vitro transcription and Northern blot analyses show that ESAG 2-related sequences are transcribed independently of the VSG expression site. This transcription, which occurs in procyclic as well as in bloodstream forms, involves a polymerase sensitive to α-amanitin. Similarly, sequences related to the 5′ barren region appear transcribed by a pol II-like enzyme in isolated nuclei from both bloodstream and procyclic forms. However, steady-state transcripts of the 5′ barren region cannot be detected in procyclic forms, suggesting that 76-bp repeats, of which 5′ barren regions are made, can be transcribed independently from the expression site, but that the products are unstable.

The existence of at least two putative ESAG genes within the VSG gene transcription unit rests on the presence of two potentially translatable ORFs, transcribed in two stable poly(A)$^+$ RNAs. These RNAs may direct the synthesis of two polypeptides of similar sizes (403 and 367 amino acids for ESAG 2 and ESAG 3, respectively). The actual existence of these polypeptides in trypanosome bloodstream forms remains to be demonstrated, but is supported by the identification of the ESAG 1 protein in cytoplasmic extracts (6). We endeavored to get an insight into their possible function by comparing their sequences, as well as the sequence of their genes, with those available in data banks. The only significant homology found was between the ESAG 3 gene and the AnTat 1.1 "companion" sequence (25). The meaning of this observation is obscure, although a similar homology with the companion has also been found in a region close to the promoter in another VSG gene transcription unit (27). A clue about the possible function of the ESAG 3 protein might be the presence of a putative signal peptide at its N terminus, suggesting interaction with membrane elements, possibly the endoplasmic reticulum.

Along these lines, it might be relevant that the ESAG 1 copy from the telomere carrying the AnTat 1.3A basic copy is most probably a pseudogene, being interrupted by several stop codons in the three reading frames. Whether this peculiarity is linked to the fact that this telomere has never been found active as a VSG gene expression site remains to be seen, but the idea finds some support in the observation that minichromosome telomeres, which do not carry ESAG-specific sequences (19), have never been shown to be used for VSG gene transcription. This further suggests that ESAG proteins are essential whenever a VSG gene is expressed.

Two additional short ORFs were found between ESAG 2 and ESAG 3. Only the first one appears translatable as such. These adjoining ORFs appear transcribed in discrete steady-state RNAs, but it is unlikely that they might encode a functional polypeptide through RNA splicing. Probes specific to the region between the two ORFs recognize the same set of RNAs as probes from the ORFs do, suggesting that the sequence between the ORFs is not an intron (Pays, unpublished data). In any case, it is worth pointing out how densely the cloned region from the VSG expression site is packed with genes; three different coding sequences are clustered in only 6.5 kb.

ACKNOWLEDGMENTS

We thank D. Jefferies for comments on the manuscript; E. Vancutsem for help in computer analysis; and J. Kooter, Netherlands Cancer Institute, Amsterdam, The Netherlands, for having drawn our attention to the homology between ESAG 3 and the AnTat 1.1 companion.

This work has been supported by research contracts between the University of Brussels and the Commission of the European Communities (TDS-M-023B) and between the University and Solvay & Cie. Brussels, Belgium. It was also funded by the African Trypanosomiasis Component of the United Nations Development Program-/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases and by the Agreement for Collaborative Research on African Trypanosomiasis between ILRAD (Nairobi) and Belgian Research Centres. S.A. and H.Q. benefited from I.R.S.I.A. (Belgium) predoctoral training grants.

LITERATURE CITED

1. **Ben Amar, M. F., A. Pays, P. Tebabi, B. Dero, T. Seebeck, M. Steinert, and E. Pays.** 1988. Structure and transcription of the actin gene of *Trypanosoma brucei*. Mol. Cell. Biol. **8**:2166–2176.
2. **Bernards, A., J. M. Kooter, and P. Borst.** 1985. Structure and transcription of a telomeric surface antigen gene of *Trypanosoma brucei*. Mol. Cell. Biol. **5**:545–553.
3. **Boothroyd, J. C.** 1985. Antigenic variation in African trypanosomes. Annu. Rev. Microbiol. **39**:475–502.
4. **Borst, P.** 1986. Discontinuous transcription and antigenic variation in trypanosomes. Annu. Rev. Biochem. **55**:701–732.
5. **Colman, A.** 1984. Expression of exogenous DNA in Xenopus oocytes, p. 49–70. *In* B. D. Hames and S. J. Higgins (ed.), Transcription and translation: a practical approach. Practical Approach Series. IRL Press, Oxford.
6. **Cully, D. F., C. P. Gibbs, and G. A. M. Cross.** 1986. Identification of proteins encoded by variant surface glycoprotein expression site-associated genes in *Trypanosoma brucei*. Mol. Biochem. Parasitol. **21**:189–197.
7. **Cully, D. F., H. S. Ip, and G. A. M. Cross.** 1985. Coordinate transcription of variant surface glycoprotein genes and an expression site associated gene family in *Trypanosoma brucei*. Cell **42**:173–182.
8. **Delauw, M. F., M. Laurent, P. Paindavoine, D. Aerts, E. Pays, D. Le Ray, and M. Steinert.** 1987. Characterization of genes coding for two major metacyclic surface antigens in *Trypanosoma brucei*. Mol. Biochem. Parasitol. **23**:9–17.
9. **Donelson, J. E., and A. C. Rice-Ficht.** 1985. Molecular biology of trypanosome antigenic variation. Microbiol. Rev. **49**:107–125.
10. **Ehlers, B., J. Czichos, and P. Overath.** 1987. RNA turnover in *Trypanosoma brucei*. Mol. Cell. Biol. **7**:1242–1249.
11. **Gilmour, R. S.** 1984. Transcription of chromatin, p. 131–152. *In* B. D. Hames and S. J. Higgins (ed.), Transcription and translation: a practical approach. Practical Approach Series. IRL Press, Oxford.
12. **Guyaux, M., A. W. C. A. Cornelissen, E. Pays, M. Steinert, and P. Borst.** 1985. *Trypanosoma brucei*: a surface antigen mRNA is discontinuously transcribed from two distinct chromosomes. EMBO J. **4**:995–998.
13. **Johnson, P. J., J. M. Kooter, and P. Borst.** 1987. Inactivation of transcription by UV irradiation of *Trypanosoma brucei* provides evidence for a multicistronic transcription unit that includes a variant surface glycoprotein gene. Cell **51**:273–281.
14. **Kooter, J. M., and P. Borst.** 1984. Alpha-amanitin insensitive transcription of variant surface glycoprotein genes provides further evidence for discontinuous transcription in trypanosomes. Nucleic Acids Res. **12**:9457–9472.
15. **Kooter, J. M., H. J. van der Spek, R. Wagter, C. E. d'Oliveira, F. van der Hoeven, P. J. Johnson, and P. Borst.** 1987. The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in *Trypanosoma brucei*. Cell **51**:261–272.
16. **Lanham, S. M.** 1968. Separation of trypanosomes from the blood of infected rats and mice by anion-exchangers. Nature (London) **218**:1273–1274.
17. **Laurent, M., E. Pays, N. Van Meirvenne, G. Matthyssens, R. O. Williams, and M. Steinert.** 1983. DNA rearrangements linked to expression of a predominant surface antigen gene of trypanosomes. Nature (London) **302**:263–266.
18. **Michiels, F., S. Muyldermans, R. Hamers, and G. Matthyssens.** 1985. Putative regulatory sequences for the transcription of mini-exons in *Trypanosoma brucei* as revealed by S1 sensitivity. Gene **36**:263–270.
19. **Murphy, N. B., M. Guyaux, E. Pays, and M. Steinert.** 1987. Analysis of VSG expression site sequences in *T. brucei*. UCLA Symp. Mol. Cell. Biol. New Ser. **42**:449–469.
20. **Murphy, N. B., A. Pays, P. Tebabi, H. Coquelet, M. Guyaux, M. Steinert, and E. Pays.** 1987. *Trypanosoma brucei* repeated element with unusual structural and transcriptional properties. J. Mol. Biol. **195**:855–871.
21. **Myler, P. J., J. Allison, N. Agabian, and K. Stuart.** 1984. Antigenic variation in African trypanosomes by gene replacement or activation of alternate telomeres. Cell **39**:203–211.
22. **Pays, E., M. F. Delauw, S. Van Assel, M. Laurent, T. Vervoort, N. Van Meirvenne, and M. Steinert.** 1983. Modifications of a *Trypanosoma brucei* antigen gene repertoire by different DNA recombinational mechanisms. Cell **35**:721–731.
23. **Pays, E., M. Delronche, M. Lheureux, T. Vervoort, J. Bloch, F. Gannon, and M. Steinert.** 1980. Cloning and characterization of DNA sequences complementary to messenger ribonucleic acids coding for the synthesis of two surface antigens of *Trypanosoma brucei*. Nucleic Acids Res. **8**:5965–5981.
24. **Pays, E., M. Laurent, K. Delinte, N. Van Meirvenne, and M. Steinert.** 1983. Differential size variations between transcriptionally active and inactive telomeres of *Trypanosoma brucei*. Nucleic Acids Res. **11**:8137–8147.
25. **Pays, E., S. Van Assel, M. Laurent, B. Dero, F. Michiels, P. Kronenberger, G. Matheyssens, N. Van Meirvenne, D. Le Ray, and M. Steinert.** 1983. At least two transposed sequences are associated in the expression site of a surface antigen gene in different trypanosome clones. Cell **34**:359–369.
26. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA-sequencing. J. Mol. Biol. **143**:161–178.
27. **Schwartz, D. C., and C. R. Cantor.** 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell **37**:67–75.
28. **Shea, C., M. G. S. Lee, and L. H. T. Van der Ploeg.** 1987. VSG gene 118 is transcribed from a cotransposed pol-I like promoter. Cell **50**:603–612.
29. **Sollner-Webb, B., and J. Tower.** 1986. Transcription of cloned eukaryotic ribosomal genes. Annu. Rev. Biochem. **55**:801–830.
30. **Steinert, M., and E. Pays.** 1986. Selective expression of surface antigen genes in African trypanosomes. Parasitol. Today **2**: 15–19.
31. **Van der Ploeg, L. H. T.** 1986. Discontinuous transcription and splicing in trypanosomes. Cell **47**:479–480.
32. **Von Heijne, G.** 1983. Patterns of amino acids near signal sequence cleavage sites. Eur. J. Biochem. **133**:17–21.
33. **White, T. C., G. Rudenko, and P. Borst.** 1986. Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28 S rRNAs. Nucleic Acids Res. **14**:9471–9489.
34. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. Gene **33**:103–119.