Cell, Vol. 34, 891–900, October 1983, Copyright © 1983 by MIT 0092-8674/83/100891-10 $02.00/0

Tandem Repetition of the 5′ Mini-exon of Variant Surface Glycoprotein Genes: A Multiple Promoter for VSG Gene Transcription?

Titia De Lange,*† Alvin Y. C. Liu,*†
Lex H. T. Van der Ploeg,† Piet Borst,*†
Marijke C. Tromp,‡ and Jacques H. Van Boom‡
*Section for Medical Enzymology
Laboratory of Biochemistry
University of Amsterdam
P.O. Box 60.000
1005 GA Amsterdam, The Netherlands
‡Organic Chemistry Laboratory
State University Leiden
Gorlaeus Laboratories
P.O. Box 9502
2300 RA Leiden, The Netherlands

Summary

Activation of some variant surface glycoprotein (VSG) genes involves a duplicative transposition to an expression site, which completes the gene by addition of a mini-exon coding for the 5′ 35 nucleotides of VSG mRNAs. Using a 22 nucleotide probe we have found some 200 copies of the mini-exon on a tandemly arranged 1.35 kb repetitive element. This repeat is highly conserved in three trypanosome species. The mini-exon on the repeat is flanked by a 5′ splice site that resembles the consensus sequence. We have not found a single mini-exon within 10 kb of the transposed VSG gene exon in the expression site. We propose a model in which the arrays of mini-exon repeats function as a repetitive promoter for efficient transcription of VSG genes.

Introduction

Bloodstream-dwelling trypanosomes, like Trypanosoma brucei, contain a protective coat composed of a single, highly immunogenic protein species: the variant surface glycoprotein (VSG). VSGs are encoded by a large gene family (Borst et al., 1981; Bernards et al., 1981; Van der Ploeg et al., 1982a), and by switching from the expression of one VSG gene to the next trypanosomes can change their antigenic identity and evade the immune response of their mammalian host (for reviews see Borst and Cross, 1982; Englund et al., 1982).

VSG genes can be categorized as telomeric or nontelomeric according to their position in the genome. The nontelomeric genes—like the 117 and 118 genes in our T. brucei strain (427)—are activated by the duplication and transposition of a silent basic copy (BC) of the gene into an expression site, yielding an expression-linked extra copy (ELC) that is transcribed (Hoeijmakers et al., 1980a; Bernards et al., 1981; Pays et al., 1981a, 1981b). The expression sites of six independent trypanosome clones expressing either the 117 or the 118 gene look similar and have two striking features (see Figure 1): a telomeric location (De Lange and Borst, 1982); and long DNA segments (up to 30 kb), flanking the transposed VSG gene segment, that lack recognition sites for 17 restriction enzymes, including tetranucleotide recognizing enzymes (Van der Ploeg et al., 1982b; Michels et al., 1982, 1983). These 5′ and 3′ "barren" regions vary in size in different trypanosome clones (Michels et al., 1982, 1983). They could contain short repetitive sequences or modified nucleotides; we have not been able to verify this, because these regions have eluded all cloning efforts (Van der Ploeg et al., 1982a).

We have presented evidence that the duplication–transposition activates VSG genes by promoter addition. Sequence analysis of the 5′ end of the 118 VSG mRNA revealed 35 nucleotides that are not encoded within the transposed VSG gene segment (Van der Ploeg et al., 1982c; Liu et al., 1983). Hence VSG gene transcription was inferred to start outside the transposed segment, from a promoter in the expression site upstream of a predicted 35 nucleotide mini-exon (Van der Ploeg et al., 1982c). As predicted, the 117 VSG gene transcript that originates from the same or a similar expression site also contains the mini-exon sequence (Boothroyd and Cross, 1982; Michels et al., 1983).

Telomeric VSG genes, like the 221 gene in the 427 strain of T. brucei, can be activated without duplication or expression-linked rearrangements detectable in nuclear blots (Williams et al., 1979, 1981, 1982; Borst and Cross, 1982; Bernards, 1982). Nevertheless, the 221 VSG mRNA contains the 35 nucleotides encoded by the mini-exon spliced onto the main body of the message (Boothroyd and Cross, 1982; A. Bernards, unpublished data). This mini-exon sequence is not found within 8 kb of the main exon of this gene (A. Bernards and T. De Lange, unpublished data). We have proposed that the telomeric genes can be translocated to the expression site by a chromosome end exchange, which places them downstream of the promoter/mini-exon area from where they are transcribed (Borst et al., 1983a). However, in situ activation of telomeric genes that have their own upstream mini-exon or activation of these genes by insertion of a mobile promoter/mini-exon element could not be ruled out for lack of information on the genomic organization of mini-exons.

To determine the mini-exon copy number and location as well as the characteristics of the sequences surrounding the mini-exons we have synthesized a specific probe for this sequence and have studied the genomic organization by nuclear blotting and cosmid clone analyses. The results indicate that the mini-exon is part of a conserved, tandemly repeated element. We discuss the possibility that these repeat arrays function as a strong promoter for VSG gene transcription at a limited number of sites in the genome.

Results

Mini-exon Probes Hybridize to a Repetitive Element

We have synthesized a 22 nucleotide oligomer comple-

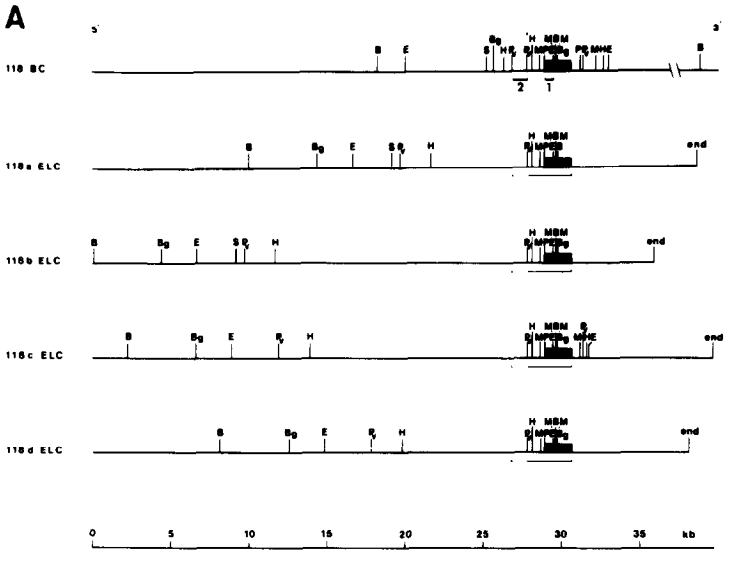
† Present address: Antoni van Leeuwenhoekhuis, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

mentary to positions 13–35 of all VSG mRNAs (see Figure 1B) and have used this oligomer as hybridization probe on Southern blots of trypanosome DNA to locate the mini-exon(s). The results in Figure 2 show that the 22 nucleotides from the VSG mini-exon are present many times in the trypanosomal genome, since the probe hybridizes to multiple Pvu II, Sal I, Bgl II, and Bam HI fragments of trypanosomal DNA. In DNA digested with Eco RI, Hind III, or Kpn I, however, only one band corresponding to high molecular weight DNA (~30 kb) is visible. These large fragments must represent large repeat clusters lacking Eco RI, Hind III, and Kpn I sites. The repetitive pattern in 22-mer hybridizations could be the consequence of an unfortunate choice of probe leading to homology with a repetitive element with no further relation to the mini-exon; however, this is not the case. Another probe—5′ complementary DNA—made by reverse transcription of the terminal 59 nucleotides of the 118 VSG mRNA (see Figure 1B) contains the mini-exon in toto and recognizes the same prominent 0.7 kb fragment as the 22-mer in Pvu II-digested DNA (Figure 2B). There are minor differences between the two mini-exon probes in their hybridization to a 0.4 kb Pvu II fragment and to the Pvu II fragments in the 4–6 kb range. Sequence analysis suggests that the 5′ complementary DNA mini-exon probe fails to detect these fragments because they contain pseudo mini-exons (see below).

The restriction fragments detected by mini-exon probes do not vary in size in different trypanosome clones (including cultured trypanosomes). Therefore, they cannot originate from the mapped part of the expression site that contains the variable 5′ barren region. This is exemplified by the comigration of Bam HI fragments from the DNA of trypanosome variants 118a and 118d (Figure 2C), which differ in the size of their flanking sequence. Rehybridization of the filter with a 118 5′ gene probe shows that the fragments detected with the 22-mer do not comigrate with the Bam HI fragments containing the ELCs of the 118 gene. This does not exclude, however, the presence of a single mini-exon within the mapped part of the expression site, because single mini-exons are not detected in this experiment (see below).

Cloning and Sequence Analysis of Mini-exon Repeats

We have cloned parts of the repetitive area identified by mini-exon probes in a cosmid vector. Screening of a library containing Mbo I partial digestion products of 118a nuclear DNA (Van der Ploeg et al., 1982a) with the 5′ complementary DNA mini-exon probe (Figure 1B) yielded six cosmid clones. All clones contain multiple copies of a 0.7 kb Pvu II fragment that hybridizes with the 22-mer (Figure 3) and comigrates with the prominent 0.7 kb Pvu II fragment in nuclear DNA (see Figure 5C). Many other Pvu II fragments appear to be repeated within the inserts and often are found in more than two clones. The repetitive nature of the



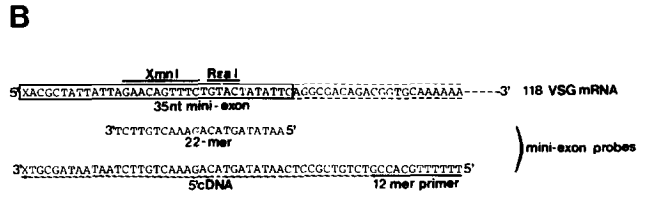


Figure 1. Physical Maps of the BC and ELCs of the 118 VSG Gene and Sequence of the 5′ End of the 118 VSG mRNA

(A) Physical maps of the 118 BC gene and 118 ELC genes in four independent trypanosome clones expressing VSG 118 (Michels et al., 1983). The black boxes represent the main exon of the genes. "End" represents the end of the chromosome. The underlined regions in the ELC maps indicate the size of the transposed 118 BC gene segment. Underneath the BC gene map the position of two probes (1 and 2) used in the experiments in Figures 2 and 8 is shown. (B) The sequence of the 5′ end of the 118 VSG mRNA (Liu et al., 1983) showing the 35 nucleotide mini-exon and 22 nucleotides encoded by the main exon. The sequences of the two mini-exon probes (22-mer and 5′ complementary DNA) are also shown. Abbreviations: B, Bam HI; Bg, Bgl II; E, Eco RI; H, Hind III; M, Msp I; P, Pst I; Pv, Pvu II; S, Sal I.

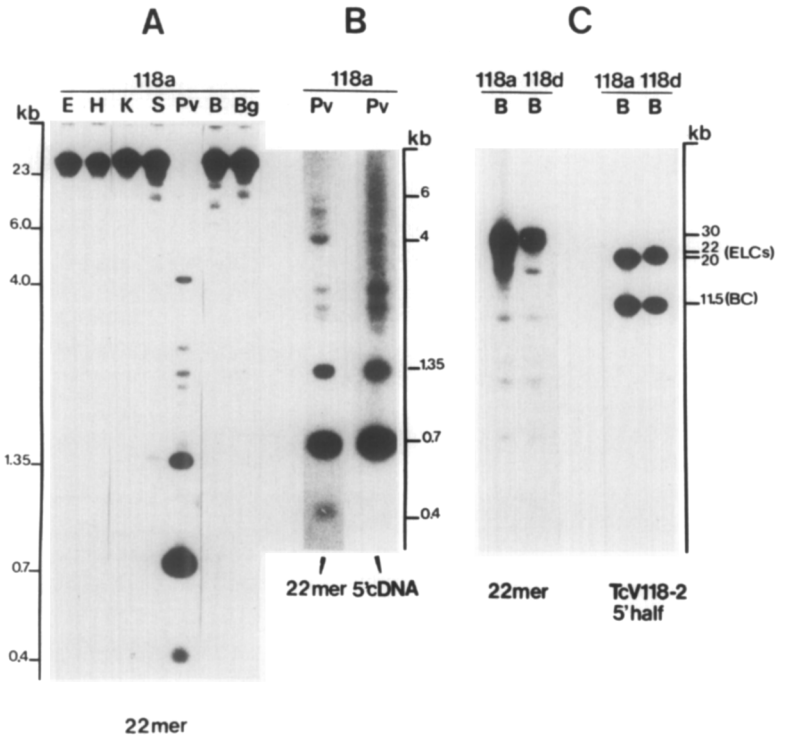


Figure 2. Nuclear Blots Showing the Repetitive Hybridization Pattern of Mini-exon Probes

(A) Variant 118a DNA was digested with the restriction enzymes indicated, size-fractionated, transferred to nitrocellulose filters, and hybridized to 22-mer mini-exon probe. (B) Two identical nitrocellulose filters containing size-fractionated 118a DNA digested with Pvu II were hybridized to the 22-mer (left) and 5′ complementary DNA (right) mini-exon probes. (C) DNA from variants 118a and 118d was digested with Bam HI, size-fractionated on a 0.5% agarose gel, and transferred to nitrocellulose filters. The filter was hybridized to the 22-mer mini-exon probe (left). After autoradiography the 22-mer signal was removed and the filter rehybridized to a probe for the 5′ half of the 118 gene from the 118 VSG complementary DNA clone TcV118-2 (probe 1, Figure 1A). Positions of the 118 BC and the 118a and 118d ELCs are indicated. Abbreviations as in Figure 1.

cloned area has precluded the construction of an overall map. It is even likely that some clones do not overlap, even though they contain comigrating Pvu II fragments. We have, however, constructed a map of one clone, cPR1 (Figure 3B), which clearly establishes the repetitive nature of the insert and also substantiates that mini-exon repeats are closely linked.

We have analyzed the 0.7 kb Pvu II fragments from cPR1 and cPR20 in more detail. cPR1 contains four copies of this fragment; one was subcloned in M13mp7 and sequenced by the dideoxy-chain-termination procedure. cPR20 contains approximately eight copies, which were sequenced as one restriction fragment by the chemical degradation procedure. The sequences are identical and show the presence of a complete copy of the mini-exon 21 bp from the Pvu II site (Figure 4). Determination of the complete sequence of the 0.7 kb Pvu II fragment (T. De Lange, unpublished data) has shown that this is the only mini-exon on this fragment. The mini-exon ends in the sequence TTG/GTATGA, which fits the consensus sequence of 5′ splice sites derived by Mount (1982) (see Figure 4). The 3′ splice sites of three VSG genes are also listed in Figure 4 and show that trypanosomes follow the AG rule for intron ends. Further homology to the 3′ splice site consensus sequence is limited, and only the 3′ splice site of the 118 VSG gene fits well.

We have searched the sequence upstream of the mini-exons for homology with canonical RNA polymerase II promoter sequences of higher eucaryotes. There is no extensive homology at analogous positions (Figure 4). However, the sequence 5′-GTAAAATATG-3′ from -67 to -58 might be equivalent to the TATA box sequence, because putative TATA boxes in front of yeast genes are also often found farther upstream (from -35 to -180; see Sentenac and Hall, 1983) than in higher eucaryotes. Further analysis of the promoter characteristics of the mini-exon repeat sequence will require information on other trypanosomal protein coding genes and the development of a trypanosome transformation system.

Besides complete copies of the mini-exon we have found pseudo (ψ) mini-exons. One was found by sequence analysis of the 5.3 kb Pvu II fragment from cPR1. Another is present on the 0.4 kb Pvu II fragment in cPR20. Both ψ mini-exons lack 6 bp at the 5′ side. From this breakpoint the sequences diverge completely (not shown). The presence of ψ mini-exons, which contain the 22-mer sequence but lack the 5′ part of the mini-exon, was already suggested by the failure to detect a minority of the Pvu II fragments, recognized by the 22-mer probe at 30°C, with the 5′ complementary DNA mini-exon probe at 45°C (see Figure 2B).

Structure and Number of the Mini-exon Repeats

The physical map of cPR1 (Figure 3) shows that the 0.7 kb Pvu II fragment is flanked by a 0.65 kb Pvu II fragment. All other cPR clones contain multiple 0.7 kb and 0.65 kb Pvu II fragments in a 1:1 stoichiometry (see Figure 3). The two fragments apparently belong to one repeat unit of about 1.35 kb. We have mapped the Msp I, Mbo I, and Taq I sites in the repeat unit by Southern blot analysis of nuclear DNA (data not shown). The map in Figure 5A shows that Mbo I and Taq I cut the 1.35 kb unit once. Both enzymes yield one major fragment of 1.35 kb in nuclear DNA (shown for Mbo I in Figure 5B), which can be

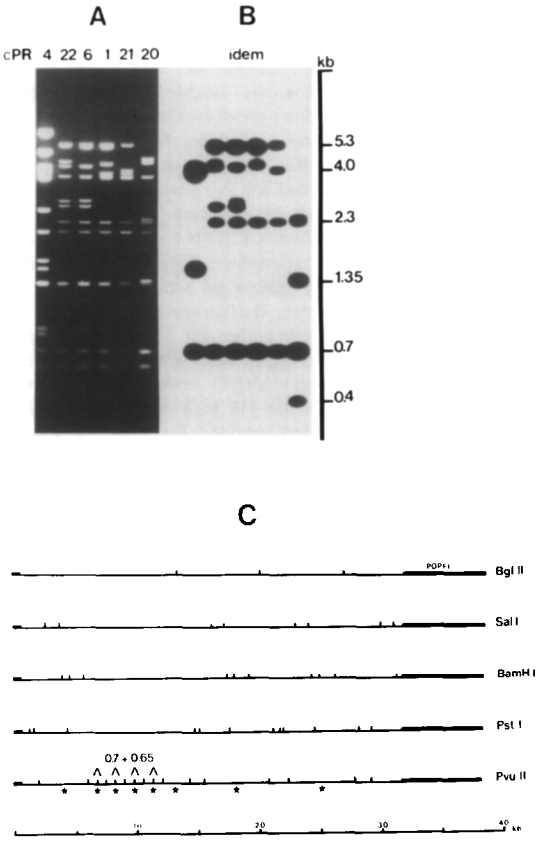


Figure 3. Cosmid Clones Containing Mini-exon Repeats

(A) Ethidium-bromide/UV staining pattern of a 0.7% agarose gel containing size-fractionated DNA of six cosmid clones digested with Pvu II. (B) The DNA from the gel shown in (A) was transferred to nitrocellulose filters and hybridized to the 22-mer mini-exon probe. (C) Physical map of cosmid clone cPR1. The map was constructed by single, double, and partial digestion analyses of clone cPR1 and isolated restriction enzyme fragments of this clone. The orientation of the four 0.7 kb and 0.65 kb Pvu II fragment pairs was not determined. The insert lacks sites for Eco RI, Hind III, and Kpn I. The sites in the cosmid vector (POPFI; Grosveld et al., 1981) are not shown. The asterisks indicate Pvu II fragments that hybridize to the 22-mer mini-exon probe.

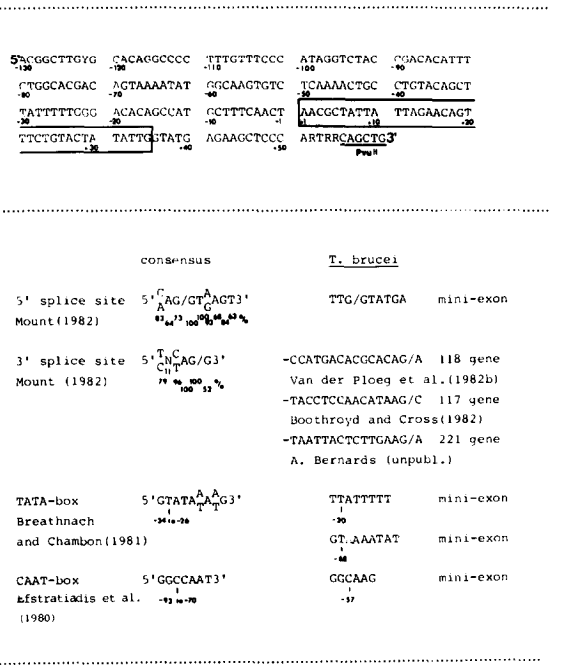


Figure 4. Flanking Sequences of the Mini-exon

The top half of the figure shows the sequences flanking the mini-exon (boxed) on the 0.7 kb Pvu II fragment from cPR1, which is identical with the sequence of the 0.7 kb Pvu II fragments of cPR20. The two sequences were determined by dideoxy and chemical degradation methods, respectively (see Experimental Procedures). In the bottom half the 5′ and 3′ splice site consensus sequences are compared with the sequences flanking VSG gene introns and the canonical promoter sequences in higher eucaryotes are compared with sequence blocks from the mini-exon prelude.

explained by tandem linkage of the majority of the repeats. Mbo I partial digestion analysis (Figure 5B) indeed yields 1.35 kb multimers. High resolution electrophoresis of the large Mbo I partials shows linkage of at least 15 repeat units. We have not obtained such long tandem arrays in clone. This may be the consequence of the cloning procedure, in which the DNA was size-selected on sucrose gradients after partial Mbo I digestion (see Van der Ploeg et al., 1982a), as cloning of Mbo I partials without size selection does yield clones containing up to ten tandemly linked repeat units (T. De Lange, unpublished data).

Long stretches of tandemly linked mini-exon repeats should yield large fragments that hybridize to the 22-mer in nuclear DNA digested with restriction enzymes that do not cut within the repeat. This is in fact what we find with Bam HI, Sal I, Bgl II, Eco RI, Kpn I, and Hind III in Figure 2.

We have determined the approximate copy number of the mini-exon repeats by quantitative Southern blotting. The hybridization intensity of the 0.7 kb Pvu II fragment in nuclear DNA was compared to the intensity of the same fragment from cPR1. From the experiment shown in Figure 5C and two other independent ones we calculate the copy number of the mini-exon repeat to be about 200 per nucleus (for calculation see legend to Figure 5).

The Mini-exon Repeats Are Highly Conserved

The flanking sequence of some 200 mini-exon copies in the genome of our T. brucei strain (427) is rather homogeneous over an area of 1.35 kb including two Pvu II sites, two Msp I sites, one Taq I site, and one Mbo I site. Such homogeneity could reflect either the recent amplification of the repeat unit or an essential role of multiple, identical repeat units in the genome. To test this we have analyzed Pvu II–digested DNA from four T. brucei strains, Trypanosoma rhodesiense, and two additional trypanosome species—T. equiperdum and T. evansi. These trypanosomes all belong to the subgenus Trypanozoon and undergo antigenic variation. Figure 6 shows that these strains all contain the mini-exon repeats. Variation occurs in the number and intensity of the faint larger repeat bands, but not in the major hybridizing 0.7 kb Pvu II band (the intensity of which varies with the amount of DNA per lane).

We also analyzed two trypanosome species from different subgenera: Trypanosoma vivax (Duttonella) and the biologically different Trypanosoma cruzi (Schizotrypanum). T. cruzi does not undergo antigenic variation; T. vivax

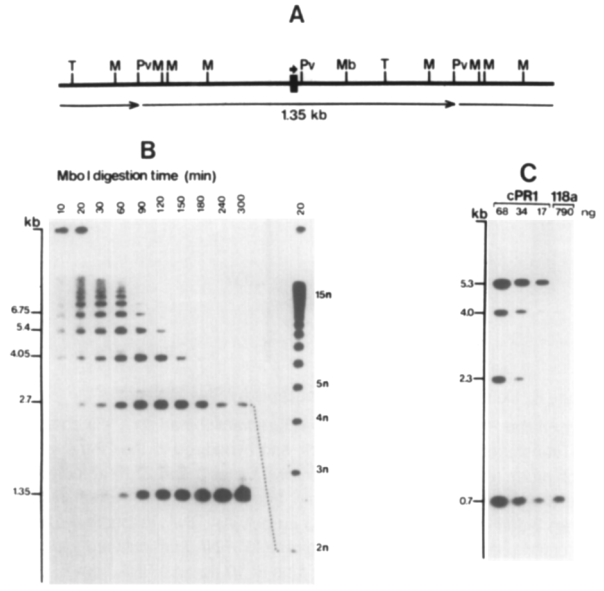


Figure 5. Structure and Copy Number of the Mini-exon Repeat

(A) Physical map of the 1.35 kb mini-exon repeat. The map was derived by blot analysis of single and double digests of nuclear DNA and hybridization with the 22-mer. The black box indicates the mini-exon. Abbreviations: M, Msp I; Mb, Mbo I; Pv, Pvu II; T, Taq I. The second and third Msp I site upstream of the mini-exon were found by sequence analysis of the 0.7 kb Pvu II fragment. (B) Mbo I partial digestion of mini-exon repeat arrays. DNA (1.5 μg) from variant 118a was incubated with Mbo I (0.5 U) for the times indicated, size-fractionated on a 0.7% agarose gel, transferred to a nitrocellulose filter, and hybridized to the 22-mer mini-exon probe. In a sample of the DNA digested for 20 min and run separately on a 0.5% agarose gel to resolve large products, partials containing up to 15 linked units (15 n) can be detected. (C) Determination of the copy number of the mini-exon repeat. The autoradiogram shows a blot containing the indicated amounts of size-fractionated nuclear (variant 118a) and cloned (cPR1) DNA digested with Pvu II and hybridized to the 22-mer mini-exon probe. Scanning of the autoradiogram showed that the intensity of the 0.7 kb Pvu II fragment in 790 ng nuclear DNA corresponds to the intensity of the same fragment in 25 ng cPR1 DNA. As the diploid genome size of T. brucei is about 74 × 10^3 kb (Borst et al., 1982), i.e. 1.85 × 10^3 the size of cPR1, and cPR1 contains four copies of the 0.7 kb Pvu II fragment, it follows that T. brucei contains 25/790 × 1.85 × 10^3 × 4 = 234 copies per diploid genome.

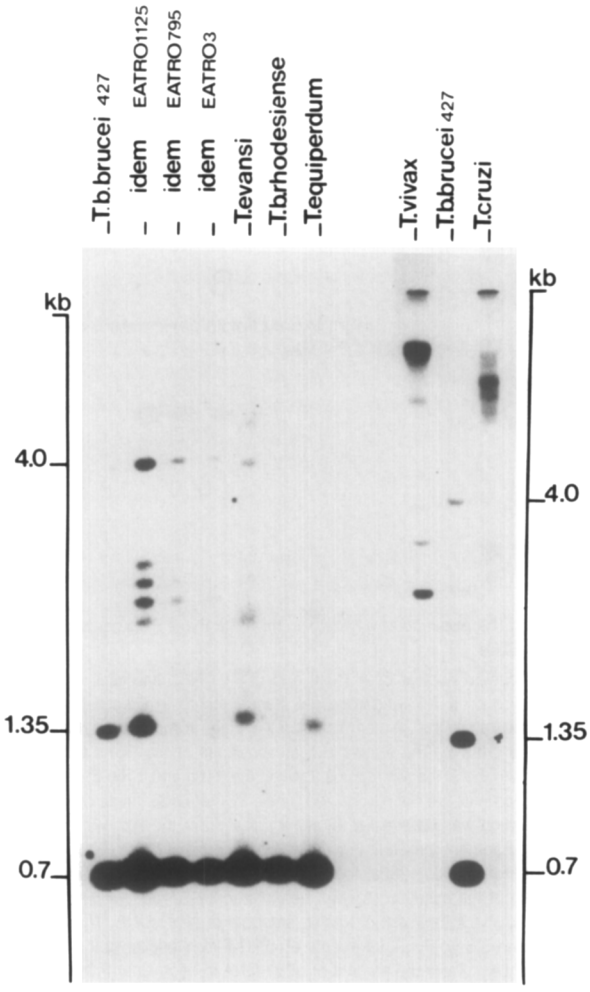


Figure 6. Nuclear Blots Showing That the Mini-exon Repeat Is Conserved in Trypanosomal Evolution

DNA from the trypanosome strains indicated was digested with Pvu II, size-fractionated on a 0.7% agarose gel, transferred to nitrocellulose filters, and hybridized to the 22-mer mini-exon probe.

does. Figure 6 shows that DNA from both species shows some homology to the 22-mer. The repeat unit, if any, must differ, however, from the 1.35 kb repeat of the members of the subgenus Trypanozoon as the 0.7 kb Pvu II fragment is absent. The hybridization with T. vivax DNA is particularly relevant, because none of our other T. brucei probes (including the conserved 3′ half of VSG genes and the 177 bp satellite sequence) hybridizes to T. vivax DNA at low stringency (T. De Lange and J. D. Barry, unpublished data). From this we infer that the mini-exon repeats are a functional part of the genome rather than an accidental satellite-like sequence.

The Nearest Mini-exon in the Expression Site Is More Than 1.5 kb Upstream of the 5′ Barren Region

The mini-exon sequence contains an Rsa I site and an Xmn I site (see Figure 1B) that can be used to determine the nearest possible location of the mini-exon in the expression site. The first Rsa I site in the expression site lies upstream of the 5′ barren region, just in front of the Hind III site (Van der Ploeg et al., 1982c). We have mapped the nearest Xmn I site 200 bp 3′ of the Pvu II site upstream of the barren region (not shown). Although this suggests that the nearest mini-exon in the expression site is located at or upstream of the Xmn I site, we cannot exclude the possibility that the barren region contains Rsa I and Xmn I sites masked by modified bases. The absence of the mini-exon in the Pvu II fragment containing the 5′ barren region was verified by quantitative Southern blotting with cPR1 DNA as standard. Under conditions that allow detection of less than 0.5 mini-exon/diploid genome, the probe does not light up the Pvu II fragment that contains the 5′-flanking region of the ELCs of VSG gene 118 in DNA from variants 118a and 118d (Figure 7). From these results we conclude that the nearest mini-exon is upstream of the Pvu II site—i.e., at least 1.5 kb upstream of the barren region and at least 20 kb upstream of the main body of the gene in variant 118b.

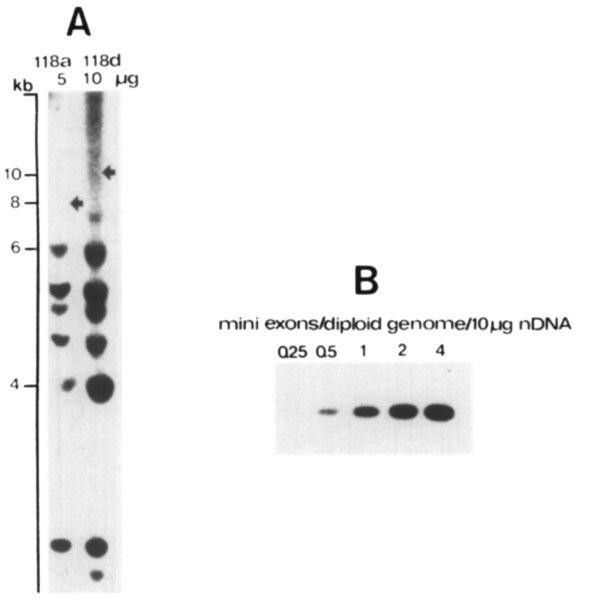


Figure 7. The Absence of a Mini-exon from the Pvu II Fragments Flanking the 118 ELCs

Pvu II–digested nuclear DNA (test) and cPR1 DNA (standard) was size-fractionated on a 0.5% agarose gel, transferred to a nitrocellulose filter, and hybridized to the 22-mer mini-exon. (A) DNA from variant 118a (5 μg) and 118d (10 μg). The arrows indicate positions of the Pvu II fragments that contain the 5′ barren regions flanking the 118a and 118d ELC genes. These fragments are not recognized by the 22-mer under these conditions. Their position was determined by rehybridization of the filter with the 970 bp Pvu II fragment (probe 2, Figure 1A) from a cosmid clone (TgBc118-29; Van der Ploeg et al., 1982a) that contains the 118 BC gene area. (B) The sensitivity of the assay in (A) was determined by 22-mer hybridization to the 0.7 kb Pvu II fragment of known amounts of cPR1 DNA. The 2.3 kb, 4.0 kb, and 5.3 kb Pvu II fragments of cPR1 that hybridize to the 22 mer are not shown. From the amount of cPR1 DNA in each lane (0.34–5.4 ng) and the parameters given in the legend to Figure 5C, the 0.7 kb Pvu II fragment was calculated to hybridize with an intensity equivalent to 0.25–4 mini-exon copies per diploid genome in 10 μg DNA.

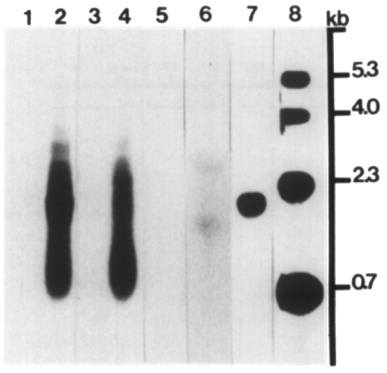


Figure 8. Detection of Mini-exon Transcripts by RNA Blotting

Glyoxylated RNA (or DNA, lane 8) was size-fractionated on a 1% agarose gel, transferred to nitrocellulose filters, and hybridized to the 22-mer mini-exon probe (except for lane 7, see below). Lanes 1 and 2, 10 μg poly(A)$^-$ and 3 μg poly(A)$^+$ RNA, respectively, from variant 118a trypanosomes. Lanes 3 and 4, 10 μg poly(A)$^-$ and 3 μg poly(A)$^+$ RNA, respectively, from cultured trypanosomes. Lane 5, 3 μg poly(A)$^+$ RNA from cultured trypanosomes treated with 2 μg RNAase A; lane 6, 10 μg total RNA from E. coli. Lane 7 shows a rehybridization of lane 2 with a 118 VSG complementary DNA probe (the 5′ half gene subprobe from plasmid TcV118-2, see probe 1, Figure 1A). Lane 8 contains 0.2 μg cPR1 DNA cut with Pvu II. Lanes 5 and 6 were exposed five times longer.

Mini-exon Transcripts

From the data presented in the preceding section it follows that the complete 118 VSG gene is very large, at least 12 kb in its smallest version (118a). The primary transcript should, therefore, be no less than 12 kb. We have probed nitrocellulose blots of size-fractionated RNA from variant 118a with the 22-mer to verify the presence of such precursors. As expected, the 22-mer hybridizes to the mature VSG transcript in 118a poly(A)$^+$ RNA (Figure 8, lane 2). In addition to this RNA species we find a broad smear of hybridizing RNAs ranging from 0.8 to 6 kb, but no 12 kb species. Appropriate controls included in Figure 8 exclude the possibility that this smear results from the presence of DNA (lane 5), nonspecific degradation of RNA (lane 7), or aspecific binding of the probe to RNA (lane 6).

As an additional control the RNA from culture form T. brucei was tested. Since these trypanosomes do not have a surface coat, transcription of VSG genes was not expected. Lane 4 of Figure 8 shows, however, that the culture form trypanosomes yield nearly the same size distribution of RNA hybridizing to the 22-mer as do blood-stream trypanosomes. Only the VSG mRNA band is missing. We have verified the absence of VSG gene transcripts in total and poly(A)$^+$ RNA from cultured trypanosomes by reverse transcription/dideoxy-chain-termination sequence analysis using a 14-mer primer complementary to the 3′ end of VSG mRNAs (Michels et al., 1983). RNA from cultured trypanosomes did not yield 14-mer primed cDNA, whereas in mixing experiments with RNA from variant 221a the expected 221a sequence was obtained. Hence the coatless cultured trypanosomes contain RNA that hybridizes with the 22-mer, but little or no mature VSG mRNA.

Discussion

We have shown that the VSG gene mini-exon is part of a 1.35 kb repetitive element, present some 200 times in the T. brucei genome. The elements are largely organized in long arrays of head-to-tail linked units. What is their function? One possibility is that they are unrelated to VSG gene expression and that a single mini-exon is used to start VSG gene transcription. Although we cannot exclude this, we favor the idea that a tandem array of mini-exon repeats is present upstream of the ELC in the expression site and that each of these mini-exons can act as start for RNA synthesis, as depicted in Figure 9. We propose that this arrangement has evolved to allow a very high rate of transcription of the VSG gene that resides in the expression site.

Although there are no experimental data on the rate of synthesis of VSG mRNA, it must be high. VSG makes up 10% of the total trypanosome protein (Cross, 1978); VSG mRNA is a major component of the poly(A)$^+$ RNA (Williams et al., 1978); and the stability of this messenger must be

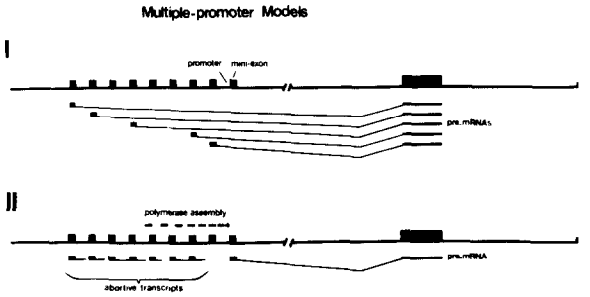


Figure 9. Multiple Promoter Models for VSG Gene transcription

Schematic representation of two alternative models in which mini-exon repeat arrays function as a repetitive promoter area for VSG genes inserted in the expression site. For explanation see Discussion.

low to allow efficient switching of the VSG synthesized. Many genes that code for abundant gene product (ribosomal RNA, dihydrofolate reductase, chorion proteins) are constitutively or temporally amplified (for review see Schimke, 1983). Trypanosomes are no exception to this rule, since they contain a 15-fold tandem amplification of the genes coding for α- and β-tubulin (Thomashow et al., 1983). However, amplification of VSG genes would not be simple. A 10-fold amplification of the 10^3 VSG BC genes (see Van der Ploeg et al., 1982a) would double the genome size. A 10-fold amplification each time that a new VSG gene duplicate is slotted into the expression site would require a logistic tour de force. A strong promoter might be the best way to meet the high demand for VSG gene products.

Two versions of the multiple-promoter model in Figure 9 are compatible with our data. One version is that RNA polymerases start at each repeated mini-exon and transcribe all downstream sequences into long precursors, which would be spliced to yield a mature message that contains the first transcribed mini-exon and the main body of the gene. We have not detected long primary transcripts in Northern blots, but as these transcripts should be over 10 kb, they might be processed during transcription or lost during isolation. The spliced out sequences could be represented by the heterogeneous transcripts we detected with the 22-mer. The recovery of spliced out RNAs in oligo(dT)-selected RNA is unexpected, but might be due to dT homopolymeric tracts in the template.

The other version is that the mini-exon repeats merely act as entry sites for polymerases in order to assemble them at the farthest downstream mini-exon from which the synthesis of the primary VSG transcript starts. During the travel downstream polymerases would start transcription at every mini-exon, yielding the abortive heterogeneous transcripts detected with the 22-mer. Polymerases could either slide on after termination or dissociate from the template and reassociate with high probability in the same area. A similar enhancing effect through polymerase assembly has been proposed for the repeated initiation sites in the spacers between Xenopus laevis ribosomal RNA genes (Moss, 1983).

The multiple-promoter model predicts physical linkage between a mini-exon repeat array and the expression site. None of the cosmid clones we have isolated overlaps the mapped part of the expression site (T. De Lange, unpublished results), possibly because of selection against cosmids containing long repeat arrays during the cloning procedure (see above). We are currently trying to demonstrate linkage by other cloning procedures and chromosome walking. A test of the second prediction of the multiple-promoter model—the enhancing effect of promoter redundancy on transcription—awaits the development of a suitable in vivo transcription system in which mutagenized templates can be tested.

Although our experiments show that the mini-exons are strongly clustered in units of at least 15 repeats, it is possible that there are several such clusters and that some of these are unrelated to mRNA synthesis or linked to genes other than VSG genes. Linkage to other genes could explain the abundance of heterogeneous mini-exon transcripts found in culture form trypanosomes, which synthesize no VSGs and contain no detectable VSG mRNA. Alternatively, the heterogeneous mini-exon transcripts could originate from the redundant VSG gene promoter, the lack of VSG gene transcripts being due to a block in the transcription of sequences downstream of the promoter, incomplete or altered RNA processing, or simply lack of physical linkage between the promoter region and a VSG gene. The latter possibility is rendered less likely by the recent finding (Parsons et al., 1983) that bloodstream trypanosomes lose their VSG, but not the ELC specifying this VSG, when brought into culture.

Our data also bear on the question whether the telomeric VSG genes that are switched on without duplication (Bernards, 1982; Williams et al., 1982; Borst et al., 1983a, 1983b) are activated in situ or by translocation. The number of telomeric genes is still unknown. We have identified three in our trypanosome strain (J. M. Kooter, T. De Lange, P. A. J. Leegwater, and A. Bernards, unpublished observations), but this may be only the tip of an iceberg. In collaboration with Schwartz and Cantor we have recently found that trypanosomes contain at least 100 telomeres, and most of these in mini-chromosomes that hybridize under low stringency with probes recognizing large families of VSG genes (Borst et al., 1983c). Potentially there are, therefore, some 100 telomeric VSG genes. Could all these be activated in situ, either by insertion of a mobile promoter–mini-exon module or by activation of such a module in situ? The clustering of mini-exon repeats that we have seen argues against the latter alternative. We find 200 mini-exon repeats clustered in blocks of at least 15. There are, therefore, not enough mini-exon blocks for each of the 100 telomeric genes. Furthermore, hybridization with mini-exon probes shows that no or very few (<10) mini-exons are located on mini-chromosomes (Borst et al., 1983c; L. H. T. Van der Ploeg and T. De Lange, unpublished data). Consequently, these genes must either be served by mobile mini-exon repeats or move by translocation into the expression site. We have not observed any

difference in mini-exon repeats in DNA from variants that do and those that do not express a telomeric VSG gene, but if the mobile mini-exon block were small, it could have escaped detection in blots. More direct tests are required to determine how telomeric VSG genes are activated.

Experimental Procedures

Trypanosomes

Trypanosoma brucei brucei strain 427 variants 118a and 221a were obtained from Dr. G. A. M. Cross (The Rockefeller University, New York). Variants 118b, 118c, and 118d are independently isolated trypanosome clones expressing VSG 118 from the same strain (Michels et al., 1982, 1983). T. brucei strain 427 (procyclic) cultured trypanosomes were obtained from Dr. R. Brun (Swiss Tropical Institute, Basel, Switzerland). T. brucei EATRO 795 (ILTat 1.22) and EATRO 3 (ETat 1.2) and T. vivax Swain (GUDat 1.1) were obtained from Dr. J. D. Barry (University of Glasgow, Scotland). T. cruzi Tulahuen epimastigotes were obtained from Dr. A. C. C. Frasch (University of Buenos Aires, Argentina). The origins of T. brucei EATRO 1125, T. rhodesiense TDRN34, and T. equiperdum ATCC 30019 have been described (Frasch et al., 1982).

DNA Isolation

Bloodstream trypanosomes were grown in rats and isolated as described previously (Michels et al., 1982). Cultured trypanosomes were grown as described by Hoeijmakers et al. (1981). T. cruzi epimastigotes were grown as described by Leon et al. (1980). Trypanosomal DNA was isolated by standard procedures (see Bernards et al., 1981). Recombinant DNA was isolated by the alkaline lysis procedure of Birnboim and Doly (1979).

RNA Isolation

RNA was isolated from purified trypanosomes by the LiCl–urea method (Auffray and Rougeon, 1980), and contaminating DNA was removed by DNAase treatment as described (Van der Ploeg et al., 1982c). Poly(A)$^+$ and poly(A)$^-$ fractions were separated by oligo(dT)-cellulose chromatography as described (Hoeijmakers et al., 1980b).

Blotting

Restriction endonuclease digestion, electrophoresis, and transfer of DNA to nitrocellulose filters were performed as described (Bernards et al., 1981). RNA (Northern) blots were prepared according to the glyoxal procedure (Thomas, 1980). DNA- and RNA-containing nitrocellulose filters were prehybridized as described (Bernards et al., 1981).

Hybridization Probes

Probes 1 and 2 in Figure 1A were obtained from the 118 VSG cDNA plasmid TcV118-2 (Hoeijmakers et al., 1980b) and 118 BC gene area cosmid clone TcBc118-29 (Van der Ploeg et al., 1982a), respectively, by isolation of the appropriate restriction fragments from low-melting-temperature agarose gels (Wieslander, 1979) and labeling by nick translation (Rigby et al., 1977). The 22-mer mini-exon probe was synthesized by the procedure of Van der Marel et al. (1982) and labeled at the 5′ end by polynucleotide kinase (BioLabs) with γ-^{32}P-ATP (NEN) (Maxam and Gilbert, 1980). The 5′ complementary DNA mini-exon probe was synthesized by reverse transcription of approximately 100 μg total RNA from variant 118a with the 12-mer primer and α-^{32}P-dCTP and -dATP using the reaction conditions described by Van der Ploeg et al. (1982c). The 59 nucleotide 5′ complementary DNA product was isolated from 8% polyacrylamide–7 M urea gels by the procedure of Maxam and Gilbert (1980).

Hybridization Conditions

The 22-mer and 5′ cDNA mini-exon probes were hybridized to nitrocellulose filters at 30°C and 45°C, respectively, for 20 hr in 6× NET hybridization buffer (0.9 M NaCl; 6 mM EDTA; 90 mM Tris-HCl, pH 7.5; 5× Denhardt; 0.5% sodium dodecylsulfate; 10% dextran sulfate) as described by Wallace et al. (1981) with 50 μg salmon-sperm DNA per milliliter and 0.1% Na-pyrophosphate. Posthybridization washes were performed at the hybridization temperature in 3× SSC, 0.1% sodium dodecyl sulfate (SSC = 0.15 M NaCl; 0.015 M Na-citrate, pH 7.0). Hybridization with probes 1 and 2 (Figure 1A) was carried out as described (Bernards et al., 1981) with final posthybridization washes in 0.1× SSC, 0.1% sodium dodecylsulfate at 65°C.

Library Screening

A library of Mbo I partial fragments from variant 118a DNA inserted in the cosmid vector POPFI (Grosveld et al., 1981) described previously (Van der Ploeg et al., 1982a) was screened with the 5′ complementary DNA mini-exon probe using the hybridization conditions described above. Positive colonies were picked and rescreened as described (Grosveld et al., 1981), and DNA was isolated from cultures inoculated with a single positive colony.

Sequence Analysis by the Dideoxy Method

The 0.7 kb Pvu II fragment (0.1 μg) from cPR1 was ligated to 0.2 μg Hind II–cut, dephosphorylated M13mp7 (Messing et al., 1981) with phage T4 DNA ligase (BioLabs), and the ligation products were used to transform E. coli JM103. Single-stranded template DNA for the dideoxy sequence method of Sanger et al. (1977) was prepared from recombinant M13 plaques (Sanger et al., 1980), and cDNA synthesis started with a 18-mer synthetic primer, complementary to the M13 sequence flanking the insert.

Sequence Analysis by Chemical Degradation

The 0.7 kb and 0.4 kb Pvu II fragments of cPR20 and a 320 bp Msp I fragment from the 5.3 kb Pvu II fragment of cPR1 were sequenced in both directions by the procedure of Maxam and Gilbert (1980) using 5′-labeled separated strands. Five chemical reactions (G, G + A, A>C, C + T, and T) were used.

Acknowledgments

We thank Mr. A. Bernards for communication of unpublished results, Dr. J. D. Barry (University of Glasgow, Scotland) for his help in the T. vivax experiments, Mrs. F. Fase-Fowler for preparation of trypanosomal DNAs, and Mrs. M. A. Ten Bosch-Tilanus for preparation of the manuscript. This work was supported in part by a grant to P. B. from the Foundation for Fundamental Biological Research (BION), which is subsidized by The Netherlands Organization for the Advancement of Pure Research (ZWO). A. Y. C. L. is the recipient of a fellowship from the Damon Runyon–Walter Winchell Cancer Fund (DRG-504).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 25, 1983; revised August 3, 1983

References

Auffray, C., and Rougeon, F. (1980). Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. Eur. J. Biochem. *107*, 303–314.

Bernards, A. (1982). Transposable genes for surface glycoproteins in trypanosomes. TIBS *7*, 253–255.

Bernards, A., Van der Ploeg, L. H. T., Frasch, A. C. C., Borst, P., Boothroyd, J. C., Coleman, S., and Cross, G. A. M. (1981). Activation of trypanosome surface glycoprotein genes involves a gene duplication–transposition leading to an altered 3′ end. Cell *27*, 497–505.

Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. *7*, 1513–1523.

Boothroyd, J. C., and Cross, G. A. M. (1982). Transcripts coding for different variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5′-end. Gene *20*, 279–287.

Borst, P., and Cross, G. A. M. (1982). Molecular basis for trypanosome antigenic variation. Cell *29*, 291–303.

Borst, P., Frasch, A. C. C., Bernards, A., Van der Ploeg, L. H. T., Hoeijmakers, J. H. J., Arnberg, A. C., and Cross, G. A. M. (1981). DNA rearrangements involving the genes for variant antigens in *Trypanosoma brucei*. Cold Spring Harbor Symp. Quant. Biol. *45*, 935–943.

Borst, P., Van der Ploeg, M., Van Hoek, J. F. M., Tas, J., and James, J. (1982). On the DNA content and ploidy of trypanosomes. Mol. Biochem. Parasitol. *6,* 13–23.

Borst, P., Bernards, A., Van der Ploeg, L. H. T., Michels, P. A. M., Liu, A. Y. C., De Lange, T., Sloof, P., Veeneman, G. H., Tromp, M. C., and Van Boom, J. H. (1983a). DNA rearrangements controlling the expression of genes for variant surface antigens in trypanosomes. In Genetic Arrangement, K. S. Chater, C. A. Cullis, D. A. Hopwood, A. A. W. B. Johnston, and H. W. Woolhouse, eds. (London: Croom Helm), pp. 207–233.

Borst, P., Bernards, A., Van der Ploeg, L. H. T., Michels, P. A. M., Liu, A. Y. C., De Lange, T., and Sloof, P. (1983b). Gene rearrangements controlling the expression of surface antigen genes in trypanosomes. In Proc. Cetus-UCLA Meeting on Molecular Biology of Host–Parasite Interactions, N. Agabian and H. Eisen, eds. Park City, Utah, in press.

Borst, P., Bernards, A., Van der Ploeg, L. H. T., Michels, P. A. M., Liu, A. Y. C., De Lange, T., Sloof, P., Schwartz, D., and Cantor, C. R. (1983c). The role of mini-chromosomes and gene translocation in the expression and evolution of VSG genes. In Gene Expression, UCLA Symp. on Molecular and Cellular Biology, New Series, *8,* D. Hamer and M. Rosenberg, eds. (New York: Liss), in press.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. *50,* 349–383.

Cross, G. A. M. (1978). Antigenic variation in trypanosomes. Proc. Roy. Soc. London Series B *202,* 55–72.

De Lange, T., and Borst, P. (1982). Genomic environment of the expression-linked extra copies of genes for surface antigens of *Trypanosoma brucei* resembles the end of a chromosome. Nature *299,* 451–453.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., De Riel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., and Proudfoot, N. J. (1980). The structure and evolution of the human β-globin gene family. Cell *21,* 653–668.

Englund, P. T., Hajduk, S. L., and Marini, J. C. (1982). The molecular biology of trypanosomes. Ann. Rev. Biochem. *51,* 695–726.

Frasch, A. C. C., Borst, P., and Van den Burg, J. (1982). Rapid evolution of genes coding for variant surface glycoproteins in trypanosomes. Gene *17,* 197–211.

Grosveld, F. G., Dahl, H. H. M., De Boer, E., and Flavell, R. A. (1981). Isolation of β-globin-related genes from a human cosmid library. Gene *13,* 227–237.

Hoeijmakers, J. H. J., Frasch, A. C. C., Bernards, A., Borst, P., and Cross, G. A. M. (1980a). Novel expression-linked copies of the genes for variant surface antigens in trypanosomes. Nature *284,* 78–80.

Hoeijmakers, J. H. J., Borst, P., Van den Burg, J., Weissmann, C., and Cross, G. A. M. (1980b). The isolation of plasmids containing DNA complementary to messenger RNA for variant surface glycoproteins of *Trypanosoma brucei.* Gene *8,* 391–417.

Hoeijmakers, J. H. J., Snijders, A., Janssen, J. W. G., and Borst, P. (1981). Transcription of kinetoplast DNA in *Trypanosoma brucei* bloodstream and culture forms. Plasmid *5,* 329–350.

Leon, W., Frasch, A. C. C., Hoeijmakers, J. H. J., Fase-Fowler, P., Borst, P., Brunel, F., and Davison, J. (1980) Maxi-circles and mini-circles in kinetoplast DNA from *Trypanosoma cruzi.* Biochim. Biophys. Acta *607,* 221–231.

Liu, A. Y., C., Van der Ploeg, L. H. T., Rijsewijk, F. A. M., and Borst, P. (1983). The transposition unit of VSG gene 118 of *Trypanosoma brucei:* presence of repeated elements at its border and absence of promoter-associated sequences. J. Mol. Biol., *167,* 57–75.

Maxam, A. M., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzymol. *65,* 499–560.

Messing, J., Crea, R., and Seeburg, P. H. (1981) A system for shotgun DNA sequencing. Nucl. Acids Res. *9,* 309–321.

Michels, P. A. M., Bernards, A., Van der Ploeg, L. H. T., and Borst, P. (1982). Characterization of the expression-linked copies of variant surface glycoprotein 118 in two independently isolated clones of *Trypanosoma brucei.* Nucl. Acids Res. *10,* 2353–2366.

Michels, P. A. M., Liu, A. Y. C., Bernards, A., Sloof, P., Van der Bijl, M. M. W., Schinkel, A. H., Menke, H. H., Borst, P., Veeneman, G. H., Tromp, M. C., and Van Boom, J. H. (1983). Activation of the genes for variant surface glycoproteins 117 and 118 in *Trypanosoma brucei.* J. Mol. Biol., *166,* 537–556.

Moss, T. (1983). A transcriptional function for the repetitive ribosomal spacer in *Xenopus laevis.* Nature *302,* 223–228.

Mount, S. M. (1982). A catalogue of splice junction sequences. Nucl. Acids Res. *10,* 459–472.

Parsons, M., Nelson, R., and Agabian, N. (1983). Genomic organization of variant surface glycoprotein genes in *T. bruci* procyclic culture forms. J. Cell. Biochem. *Suppl. 7A,* 22.

Pays, E., Van Meirvenne, N., Le Ray, D., and Steinert, M. (1981a). Gene duplication and transposition linked to antigenic variation in *Trypanosoma brucei.* Proc. Nat. Acad. Sci. USA *78,* 2673–2677.

Pays, E., Lheureux, M., and Steinert, M. (1981b). The expression-linked copy of surface antigen gene in *Trypanosoma* is probably the one transcribed. Nature *292,* 265–267.

Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. *113,* 237–251.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci. USA *74,* 5463–5467.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. *143,* 161–178.

Schimke, R. T., ed. (1983). Gene Amplification. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Sentenac, A., and Hall, B. (1983). Yeast nuclear RNA polymerases and their role in transcription. In The Molecular Biology of the Yeast Saccharomyces, *2,* J. N. Strathern, E. Jones, and J. Broach, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 561–606.

Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Nat. Acad. Sci. USA 77, 5201–5205.

Thomashow, L. S., Milhausen, M., Rutter, W. J., and Agabian, N. (1983). Tubulin genes are tandemly linked and clustered in the genome of Trypanosoma brucei. Cell *32,* 35–43.

Van der Marel, G. A., Marugg, J. E., De Vroom, E., Wille, G., Tromp, M., Van Boeckel, C. A. A., and Van Boom, J. H. (1982). Phosphotriester synthesis of DNA fragments on cellulose and polystyrene solid supports. Rec. Trav. Chim. Pays-Bas *101,* 239–241.

Van der Ploeg, L. H. T., Valerio, D., De Lange, T., Bernards, A., Borst, P., and Grosveld, F. G. (1982a). An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. Nucl. Acids Res. *10,* 5905–5923.

Van der Ploeg, L. H. T., Bernards, A., Rijsewijk, F. A. M., and Borst, P. (1982b). Characterization of the DNA duplication–transposition that controls the expression of two genes for variant surface glycoproteins in *Trypanosoma brucei.* Nucl. Acids Res. *10,* 593–609.

Van der Ploeg, L. H. T., Liu, A. Y. C., Michels, P. A. M., De Lange, T., Borst, P., Majumder, H. K., Weber, H., Veeneman, G. H., and Van Boom, J. (1982c). RNA splicing is required to make the messenger RNA for a variant surface antigen in trypanosomes. Nucl. Acids Res. *10,* 3591–3604.

Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H., and Itakura, K. (1981). The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequence to rabbit β-globin DNA. Nucl. Acids Res. *9,* 879–894.

Wieslander, L. (1979). A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gelling temperature agarose gels. Anal. Biochem. *98,* 305–309.

Williams, R. O., Marcu, K. B., Young, J. R., Rovis, L., and Williams, S. C. (1978). A characterization of mRNA activities and their sequence complexities in *Trypanosoma brucei:* partial purification and properties of the VSSA mRNA. Nucl. Acids Res. *5,* 3171–3182.

Williams, R. O., Young, J. R., and Majiwa, P. A. O. (1979). Genomic

rearrangements correlated with antigenic variation in *Trypanosoma brucei*. Nature *282*, 847–849.

Williams, R. O., Young, J. R., Majiwa, P. A. O., Doyle, J. J., and Shapiro, S. Z. (1981). Contextural genomic rearrangements of variable-antigen genes in *Trypanosoma brucei*. Cold Spring Harbor Symp. Quant. Biol. *45*, 945–949.

Williams, R. O., Young, J. R., and Majiwa, P. A. O. (1982). Genomic environment of *T. brucei* VSG genes: presence of a mini-chromosome. Nature *299*, 417–421.