INTERNATIONAL REVIEW OF CYTOLOGY, VOL. 99

The Molecular Biology of Antigenic Variation in Trypanosomes: Gene Rearrangements and Discontinuous Transcription

TITIA DE LANGE[1]

Division of Molecular Biology, Antoni van Leeuwenhoekhuis, The Netherlands Cancer Institute, Amsterdam, The Netherlands

I. Introduction

Protozoan parasites of mammals have developed various strategems to cope with their host's defense system (reviewed by Bloom, 1979). First, some invaders have managed to infiltrate the ultimate hiding place: the inside of a mammalian cell. Examples are *Leishmania,* surviving in mac-

[1] Present address: Department of Microbiology and Immunology, University of California, San Francisco, California.

rophages, and *Trypanosoma cruzi,* which shelters in various cell types. Second, many parasites have found ways to impair the immune system. A third strategy, cunningly employed by some trypanosomes, matches the immense collection of immunoglobulins with an equally impressive assortment of sequentially expressed antigens. The molecular biology of trypanosome antigenic variation will be discussed in detail here; other aspects of the biochemistry of trypanosomes have been discussed elsewhere (cf. Vickerman and Barry, 1982; Englund *et al.,* 1982; Opperdoes, 1984).

The bloodstream-dwelling African trypanosomes are covered by a homogeneous layer of coat protein, variant surface glycoprotein, (VSG),[2] of which antigenically different types can be synthesized during a chronic infection (cf. Cross, 1978; Vickerman, 1978). Trypanosomes have reserved ~10% of their genome to program antigenic variation; every trypanosome may have as many as 1000 genes that code for VSG (Van der Ploeg *et al.,* 1982a, 1984a).

The occasional switches in the type of coat protein sythesized are a consequence of the sequential activation of different VSG genes (cf. Borst and Cross, 1982). Among the mechanisms for gene activation, gene rearrangements feature prominently. Often, surface-antigen genes are activated by duplicative transposition to a dominant telomeric expression site, where they may procure a transcriptional promoter (cf. Bernards *et al.,* 1981; De Lange *et al.,* 1985). In addition, some telomeric VSG genes are activated without any apparent rearrangements or transposition to the dominant expression site (cf. Van der Ploeg *et al.,* 1984a). Hence, there must be more than one locus where surface-antigen genes can be expressed. In spite of the presence of multiple expression sites, only one gene is expressed at a time. The molecular basis of this exclusion is unknown.

The study of transcripts for VSG has uncovered a novel mode of mRNA synthesis that appears to be a general characteristic of the family of Trypanosomatidae. The first 35 nucleotides of many, if not all, mRNAs are the same and not encoded contiguously with the rest of the transcript (cf. De Lange *et al.,* 1984b; Parsons *et al.,* 1984a). The common leader sequence is provided in trans from highly repetitive genes that are probably not linked to protein-coding genes. These mini-exon genes are transcribed into a small precursor RNA which is thought to donate its 5′ end

[2] Abbreviations: BC, basic copy; bp, base pair(s); cDNA, complementary DNA; ELC, expression-linked extra copy; kb, kilobase pair(s); L-ELC, lingering expression-linked extra copy; Mb, megabase pair(s); mf, membrane form; mRNA, messenger RNA; PFG, pulsed filed gradient; VSG, variant surface glycoprotein.

sequence to mRNAs either by bimolecular splicing or by serving as a primer in the transcription of structural genes.

II. Antigenic Variation in African Trypanosomes

A. Trypanosome Biology

All trypanosomes and their allies discussed here belong to the family of the Trypanosomatidae [see Fig. 1 for classification and Lumsden and Evans (1976, 1979) for a comprehensive treatise on trypanosome biology]. They are uniflagellated, parasitic protozoa with a DNA-containing kinetoplast. The trypanosomes that show antigenic variation belong to the section Salivaria of the genus *Trypanosoma.*

The life cycle of *Trypanosoma brucei,* a representative of the Salivarian trypanosomes, is shown in Fig. 2. The infection of a mammal usually results from the injection of a few hundred or thousand trypanosomes by an infected tsetse fly. The trypanosomes multiply in a chancre formed at the site of the fly bite, from which they drain into lymph vessels and enter the circulatory system.

In the mammal, trypanosomes hide behind a compact layer of glycoprotein that covers the parasite body and flagellum. This surface coat protects against the nonspecific host defense and, moreover, presents antigens that readily elicit a humoral immune response, eventually removing the trypanosomes from the circulation (Vickerman, 1969, 1978). A few trypanosomes, however, survive the massacre because they have switched to another antigenically distinct coat protein and initiate a new parasitemic wave. Antigenic switching occurs at an apparent frequency of 10^{-4}–10^{-6} per cell division (Doyle, 1977) and is indepenent of the immune response: *in vitro* cultured bloodstream trypanosomes are also capable of

Phylum	Protozoa					
Class	Mastigophora					
Order	Kinetoplastida					
Family	Trypanosomatidae					
Genera	Leptomonas	Leishmania	Trypanosoma		Crithidia	Phytomonas	Herpetomonas
Sections		Stercoraria	Salivaria			
Subgenera		Schizotrypanum	Duttonella	Trypanozoon	Nannomonas	
Species		T. cruzi	T. vivax	T. equiperdum T. evansi T. brucei	T. congolense	

FIG. 1. Classification of some trypanosomes (after Lumsden and Evans, 1976).

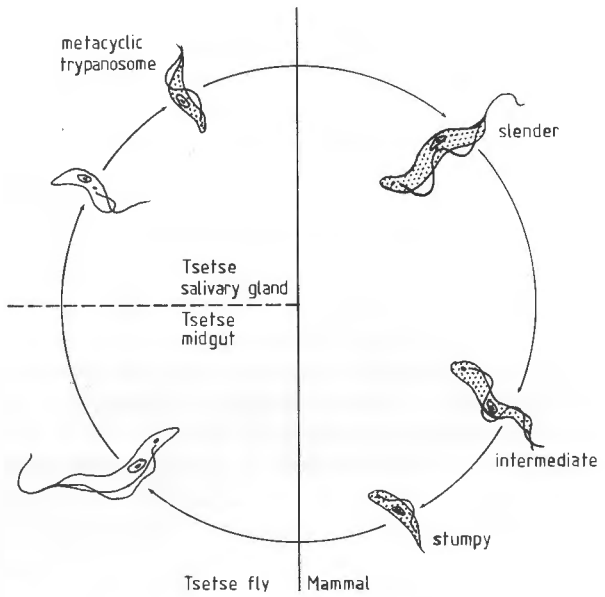

FIG. 2. The life cycle of *T. brucei*. The shaded trypanosomes carry a surface coat (see Section II,A for explanation).

coat exchange (Doyle *et al.,* 1980). Due to the large number of different coats available [e.g., more than 100 for *Trypanosoma equiperdum* (Capbern *et al.,* 1977)] and the economical use of this repertoire, trypanosomes can maintain a lengthy chronic infection with on average one parasitemic wave each week. The human disease associated with the infection is called sleeping sickness. The causative agents are *Trypanosoma brucei rhodesiense* and *Trypanosoma brucei gambiense*. *Trypanosoma brucei brucei* is not man infective due to lysis in human serum (Rifkin, 1978) and is, therefore, rather popular with biochemists. This subspecies, along with *Trypanosoma vivax* and *Trypanosoma congolense,* gives rise to diseases in cattle and other domestic mammals.

In nature, bloodstream trypanosomes are pleomorphic (see Fig. 2). Long slender forms, which divide by binary fission, are accompanied by nondividing short stumpy forms thought to be preadapted to life in the tsetse fly (Vickerman and Barry, 1982). Rodent-adapted virulent laboratory strains are usually monomorphic and often not transmissible by tsetse flies. Here, the fly is substituted by a syringe.

Upon entering the fly as part of its bloodmeal, the trypanosomes lose their surface coat and transform to the procyclic stage. A similar transformation occurs *in vitro* when bloodstream trypanosomes are incubated in

the appropriate media at 25°C (Brown *et al.,* 1973; Overath *et al.,* 1983). These cultured procyclic trypanosomes are not infective to mammals: in the absence of their surface coat, trypanosomes are lysed by serum. The fly stage of the life cycle is completed in a few weeks. After a period of differentiation in the midgut, trypanosomes migrate to the salivary glands where they reacquire the surface coat and become infective. These metacyclic trypanosomes (see Fig. 2) show a limited heterogeneity with respect to their surface antigens (Barry *et al.,* 1979): often reexpression of the same set of probably less than 20 metacyclic coats occurs after cyclic transmission of a trypanosome strain (Crowe *et al.,* 1983). Variations in the metacyclic coat repertoire are, however, not rare (Barry *et al.,* 1983). Therefore, the prospects of vaccination against metacyclic trypanosomes are rather dim.

The occurrence of trypanosomiasis in tropical Africa is more or less restricted to the distribution of the insect vector, i.e., the tsetse belt. Salivarian trypanosome species found outside this area (e.g., in North Africa, South America, and Asia) have lost the need for transmission by the tsetse fly; e.g., *Trypanosoma evansi,* which is pathogenic to camels and horses and is mechanically transmitted by bloodsucking flies and infected vampire bats, and *T. equiperdum,* which is venereally transmitted between horses.

Trypanosomes are relatively easy organisms for molecular biologists to study. First, trypanosomes can be cloned and maintained in small laboratory rodents, such as mice and rats, or stored under liquid nitrogen (Van Meirvenne *et al.,* 1975a,b). The study of antigenic variation especially profits from the availability of fairly efficient cloning procedures. Due to the low switching frequency and the high growth rate (generation time about 6 hours), large batches of antigenically identical trypanosomes can be analyzed. Second, the genome of trypanosomes is small, about 4×10^4 kb per haploid genome for *T. brucei,* with ~70% single-copy sequences (Borst *et al.,* 1980, 1982). On the other hand, the absence of *in vitro* and *in vivo* genetics that can be manipulated (no DNA-mediated transformation system, no mutants, and no test-tube sex) hampers the study of trypanosome molecular biology.

Several lines of evidence indicate that trypanosomes are diploid organisms. Studies on isoenzyme patterns (Gibson *et al.,* 1980; Tait, 1980) and on the DNA content of the nucleus (Borst *et al.,* 1982) are compatible with this idea. Recently, more information on the ploidy of *T. brucei* has come from the analysis of restriction-site polymorphisms in and around three genes that code for glycolytic enzymes (Gibson *et al.,* 1985). Like the isoenzyme patterns, the polymorphisms reveal the presence of allelic copies of each gene, which appear to segregate independently. This sup-

ports the suggestion that *T. brucei* has a diploid genome and indulges in genetic exchange. Preliminary experiments suggest that recombination of genomes may occur in the fly (L. Jenni, personal communication).

B. The Variant Surface Glycoprotein

The surface coat of trypanosomes consists of a uniform layer of about 10^7 closely packed variant surface glycoprotein (VSG) molecules (Cross, 1975, 1978). Antigenically distinct *T. brucei* clones each yield a unique VSG of about 60,000 Da that differs from other VSGs in isoelectric point, peptide map, amino acid sequence, and carbohydrate content (Cross, 1975, 1978). The protein can be subdivided into an exposed N-terminal domain and a membrane-oriented C-terminal part, which is glycosylated. The N-terminal amino acid sequence is variable and probably determines the antigenic characteristics of the VSG (Bridgen *et al.*, 1976; Miller *et al.*, 1984). In contrast, the last 110 amino acids at the C-terminus show some conservation (Rice-Ficht *et al.*, 1981; Majumder *et al.*, 1981; Matthyssens *et al.*, 1981; Boothroyd *et al.*, 1981). This may reflect a limitation on the structure nearest the plasma membrane, imposed by the necessity to interact with other VSGs and/or by the membrane proximity.

In spite of the presence of sugar side chains, the C-terminal part is not immunogenic *in situ*, probably because the antigenic determinants are shielded by the rest of the protein. Indeed, one trypanosome variant does not bind concanavalin A unless the N-terminal domain is removed by trypsin (Cross and Johnson, 1976). Carbohydrate makes up 7–17% (w/w) of the soluble VSG that is released during mild detergent lysis and consists of mannose, galactose, and glucosamine residues (no sialic acid) (Cross, 1975, 1977; Johnson and Cross, 1977). The high-mannose oligosaccharides are added in a tunicamycin-sensitive manner using the Asn-X-Ser/Thr recognition site (Rovis and Dube, 1981).

The soluble VSG, which was initially analyzed, has recently turned out to be incomplete. The complete membrane form (mf) VSG can be obtained by boiling trypanosomes in detergent, thereby inhibiting a membrane-bound enzyme that under milder conditions clips off the membrane anchorage site (Cardoso de Almeida and Turner, 1983). Although the physiological role of this enzyme(s) has not been studied, it is tempting to speculate about a "stripte-ase" function for rapid coat shedding when trypanosomes enter the fly, or a role in coat turnover. In mf VSG the C-terminal aspartic acid or serine residue is covalently linked through an ethanolamine residue (Holder, 1983) to a phosphoglycolipid, which contains myristic acid (Ferguson and Cross, 1984; Ferguson *et al.*, 1985). This saturated fatty acid is thought to anchor VSGs in the plasma membrane. Acylation has also been reported for the VSGs of *T. equiperdum* (Duvillier *et al.*, 1983). The addition of the lipid-containing carbohydrate group happens early in the synthesis of the coat protein after removal of a short C-terminal extension peptide and the N-terminal signal peptide (Boothroyd *et al.*, 1980, 1981; Boothroyd and Cross, 1982; McConnel *et al.*, 1981, 1983).

III. The Structure and Organization of VSG Genes

A. Gene Structure

VSG mRNA is ~2 kb, polyadenylated, and very abundant (Williams *et al.*, 1978; Hoeijmakers *et al.*, 1980). Not surprisingly, therefore, large parts of VSG mRNA sequences could readily be obtained from cDNA libraries (Williams *et al.*, 1979; Hoeijmakers *et al.*, 1980; Pays *et al.*, 1980). The use of these cDNAs as VSG-specific probes in the analysis of DNA and RNA showed that each VSG is encoded by a separate gene and that each VSG gene is transcribed only when the corresponding VSG is synthesized; i.e., VSG synthesis is regulated at the level of transcription (Hoeijmakers *et al.*, 1980).

To date the structure of over 30 VSG genes has been investigated in detail, and (part of) the nucleotide sequence of many of these has been determined (cf. Majumder *et al.*, 1981; Boothroyd *et al.*, 1980, 1981; Matthyssens *et al.*, 1981; Rice-Ficht *et al.*, 1981, 1982; Bernards *et al.*, 1981; Boothroyd and Cross, 1982; Donelson *et al.*, 1982; Liu *et al.*, 1983; Young *et al.*, 1983; Pays *et al.*, 1983b–d; De Lange *et al.*, 1983a; Campbell *et al.*, 1984a; Bernards *et al.*, 1985). The scheme in Fig. 3 summarizes their most salient general features. As may be expected for a gene in a lower eukaryote, VSG-coding regions lack intervening sequences that commonly interrupt protein-coding genes in higher eukaryotes. At both extremes VSG genes encode untranslated RNA segments. The leader sequence is variable in size and sequence (Boothroyd and Cross, 1982; Liu *et al.*, 1983; Dorfman and Donelson, 1984); in contrast, the 3′-untranslated trailer (UT) is highly conserved with notable motifs like the 14-mer (5′-UGAUAUAUUUUAAC-3′), which is present within 15 nucleotides upstream of the poly(A) tail in nearly every VSG mRNA of *T. brucei* (Rice-Ficht *et al.*, 1981; Borst and Cross, 1982; Michels *et al.*, 1983; Liu *et al.*, 1983). Using the 14-mer sequence as priming site for reverse transcriptase catalyzed cDNA synthesis, the 3′ end sequence of a score of VSG mRNAs has been determined directly on the RNA template, thus omitting the laborious cDNA cloning step (Michels *et al.*, 1983; Merrit *et*

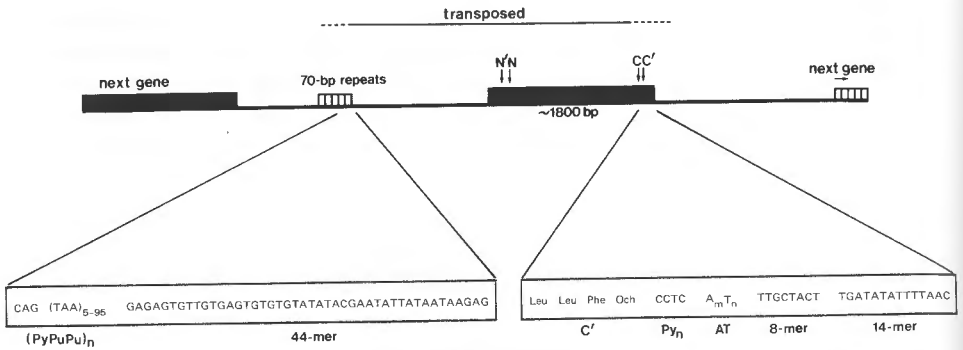

FIG. 3. The structure of chromosome-internal VSG genes. The black box depicts the region that codes for the mature VSG (between N and C), the N-terminal signal peptide (N'), the C-terminal extension (C'), and the untranslated termini of VSG mRNA. Two VSG gene-specific repeat sequences are shown in the blocks below the scheme: the left-hand block shows the consensus sequence of the 70-bp repeats upstream of the gene, the right-hand block shows the conserved features at the 3' end of VSG genes. The extent of the gene segment that is transposed upon activation is indicated, as is the close linkage between VSG genes.

al., 1983). From this analysis and data obtained with cloned cDNAs it is evident that the homology at the 3' end of VSG genes extends into the C-terminus of the mature protein (Rice-Ficht *et al.*, 1981; Borst and Cross, 1982). A homology gradient from 5' to 3' in the gene was already predicted on the basis of hybridization patterns of nuclear DNAs with cDNA probes (Borst *et al.*, 1981). Whereas 5'-half gene probes usually detect one or two restriction fragments in nuclear DNA, 3'-half gene probes often hybridize to a multitude of bands, so much so that they have proved useless for mapping analyses in some cases. The unique 5' halves of VSG genes coincide with a functional domain: the N-terminus of the protein that carries the variable antigenic determinants. Likewise, the conservation of the 3' ends of VSG genes probably in part reflects a structural constraint on the amino acid sequence of the membrane-oriented C-terminus. In addition, the conserved sequence at the 3' ends plays a role in the activation VSG genes (see Section IV,A).

Contrary to what is seen for housekeeping genes (see Section II,A), the copy numbers of many individual VSG genes are not in agreement with the presumed diploidy of trypanosomes: genes for surface antigens are present at one, two, or three copies per nucleus (Frasch *et al.*, 1982; De Lange *et al.*, 1983a). How trypanosomes manage a partly haploid genome is not clear (see Gibson *et al.*, 1985, for discussion of this problem).

B. Genomic Context

VSG genes are found in two radically different genomic contexts: at chromosome-internal sites and at chromosome ends (telomeres) (Williams *et al.*, 1982; De Lange and Borst, 1982). Until recently, the contribution of telomeric VSG genes in the gene repertoire could not be fully appreciated; no reliable information on chromosome number had been obtained because trypanosomes do not have discrete condensed chromatin bodies during any phase of the cell cycle. The application of the pulsed field-gradient (PFG) gel electrophoresis technique, by which chromosome-sized DNA molecules can be separated, has recently uncovered the interesting structural details of the trypanosome genome (Schwartz and Cantor, 1984; Van der Ploeg *et al.*, 1984a). A representative PFG gel pattern of *T. brucei* chromosomes is shown in Fig. 4. Roughly four fractions are separated (I) large chromosomes that hardly move away from the origin, (II) several chromosomes that migrate in the megabase-pair size range and cannot be accurately sized due to a compression effect, (III) several intermediately sized chromosomes, and (IV) an impressive number of small (50–150 kb) chromosomes (Van der Ploeg *et al.*, 1984a; A. Bernards, personal communication). Obviously, there is ample room for VSG genes at telomeres. Indeed, many (if not all) small chromosomes may carry VSG genes at their ends, since probes for the conserved 3' ends of VSG genes hybridize strongly to this DNA fraction (Van der Ploeg *et al.*, 1984a).

The majority of VSG genes, however, has a chromosome-internal position. Unlike telomeric VSG genes, chromosome-internal genes and their surrounding sequences are usually well represented in genomic libraries made in *Escherichia coli*. Extensive analysis of cosmid clones containing the gene for VSG 118 in *T. brucei* strain 427 has shown that this gene is closely linked to other VSG genes, detected through the conserved 3' end sequence. Less than 3 kb separates the 118 gene from its nearest neighbors, and a similar VSG gene clustering was found around the gene for VSG 117 (Van der Ploeg *et al.*, 1982a).

Another point to emerge from the analysis of cloned genes is the presence of a second conserved feature of VSG genes, the 70-bp repeats (Van der Ploeg *et al.*, 1982a; Liu *et al.*, 1983; Campbell *et al.*, 1984a). This conserved sequence is found a few kilobase pairs upstream of the start codon and is made up of direct imperfect repeats, usually 70–80 bp long. Their consensus sequence (Campbell *et al.*, 1984a) is given in Fig. 3; the length variation is mainly due to the number of PyPuPu (typically TAA) motifs that precede the more conserved 44-bp 3' half of the repeats. Using

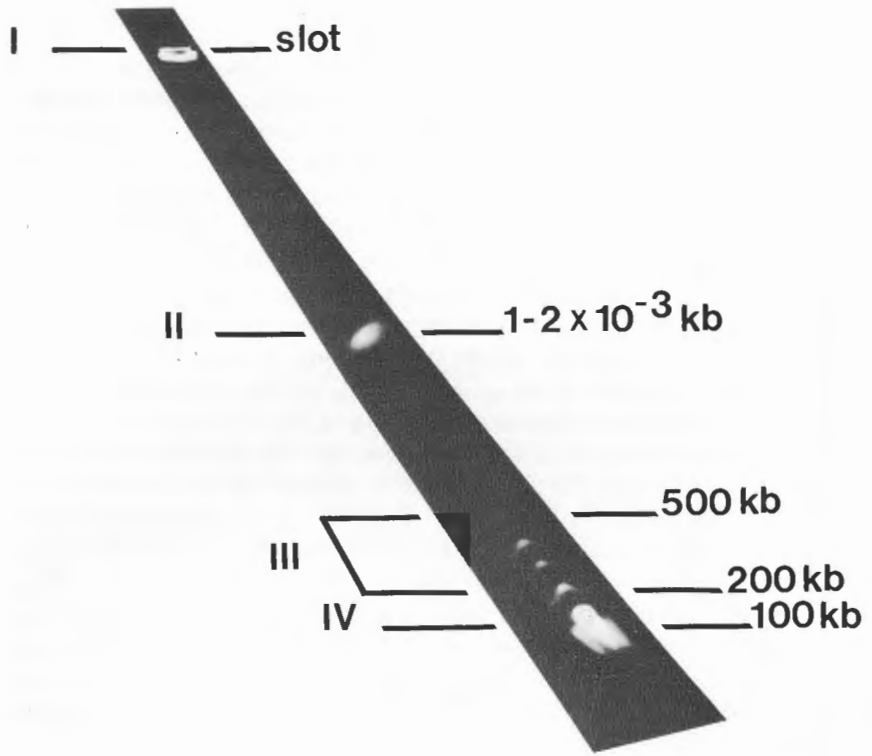

FIG. 4. Chromosomes of *T. brucei*, separated by pulsed field-gradient gel electrophoresis, stained with ethidium, and photographed over ultraviolet light. Four classes of chromosomes are visualized, I, large chromosomes that hardly leave the slot; II, a few chromosomes in the megabase-pair size range, which cannot be sized accurately due to a compression in this area; III, intermediate-sized chromosomes; IV, minichromosomes. The tentative molecular weights are derived from comigration with phage λ (50-kb) oligomers (A. Bernards, personal communication). For details on the electrophoresis technique, see Schwartz and Cantor (1984) and Van der Ploeg *et al.* (1984a).

this sequence to detect VSG genes in a cosmid library, Van der Ploeg *et al.* (1982a) estimated that 9% of the colonies contain VSG gene clusters, together representing an estimated total of 10^3 potential chromosome-internal VSG genes in the genome of *T. brucei*.

The only telomeric VSG gene that has been cloned together with a large stretch upstream codes for VSG 221. Like the chromosome-internal genes, the area upstream of gene 221 contains 70-bp repeats (Bernards *et al.*, 1984a, 1985) (see Fig. 5). A difference between telomeric and chromosome-internal VSG genes may be the number of 70-bp repeats: whereas the genes for VSG 117 and 118 have, respectively, four and five copies,

the telomeric 221 gene has about 50. Although this has not been verified for other telomeric VSG genes, indirect evidence suggests this difference may be the rule. First, Van der Ploeg *et al.* (1982a) have shown that satellite-like long stretches of 70-bp repeats are present in the genome. Second, restriction maps of telomeric VSG genes invariably show a region of up to 40 kb that is scarce in restriction enzyme cutting sites (cf. Young *et al.*, 1983a; Pays *et al.*, 1983b; Laurent *et al.*, 1983; Bernards *et al.*, 1984a). These "barren" regions could very well contain large arrays of 70-bp repeats because these repeats lack sites for most restriction endonucleases (with notable exceptions such as *Acc*I, *Dde*I, *Hph*I, and *Mbo*I).

C. Telomere Structure and Instability

Originally some VSG genes were suspected to reside close to a duplex discontinuity in the DNA on the basis of their position upstream of a site where all tested restriction endonucleases appeared to cut (Van der Ploeg *et al.*, 1982b; Michels *et al.*, 1982). This supposition was verified by showing that the 3′-flanking sequence of these genes can be progressively shortened by digestion of intact nuclear DNA with exonuclease *Bal*31 (Williams *et al.*, 1982; De Lange and Borst, 1982).

Currently, there are additional criteria that set telomeric VSG genes apart from their chromosome-internal counterparts. Trypanosome telomeres have proved to be unusually unstable DNA segments that frequently show extensive length alterations. As a result, the distance between the gene and the end of the chromosome varies in different trypanosome clones, and now this feature is usually the first indication for the telomeric location of a newly analyzed VSG gene.

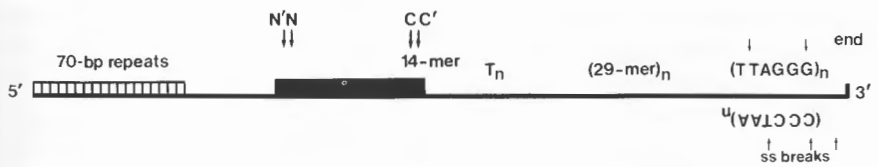

FIG. 5. Common sequence elements around telomeric VSG genes. A speculative scheme based on data from Van der Ploeg *et al.* (1984b), Blackburn and Challoner (1984), Bernards *et al.* (1985), and De Lange *et al.* (1983a). The repeats indicated are the 70-bp repeats (see Fig. 3 for consensus sequence), a T-rich element (T_n), a stretch of direct repeats of 29 bp [$(29\text{-mer})_n$] with 5′-TTAGGG-3′ motifs, and a tandem array of 5′-TTAGGG-3′ hexamers (note that this region reads CCCTAA from telomere to centromere). The arrows symbolize the presence of single-stranded (ss) breaks in trypanosome telomeres. (See Fig. 3 for explanation of the structure of the gene and Sections III,B and C for further details.)

At least two processes are responsible for this curious phenomenon. First, telomeres in trypanosomes gradually increase in length by 6–7 bp per generation (28 bp per day) (Bernards *et al.*, 1983; Pays *et al.*, 1983a; Van der Ploeg *et al.*, 1984b). Second, this growth is balanced by occasional deletions that become apparent when trypanosomes are carried through cloning procedures or heat-shock treatment (Bernards *et al.*, 1983). Telomeres harboring actively transcribed VSG genes may be slightly different, as their growth rate is significantly higher (36–40 bp per day) and they appear more prone to deletions (Pays *et al.*, 1983a). A model explaining the growth of trypanosome telomeres has been presented by Van der Ploeg *et al.* (1984b).

The precise structure of the chromosome end flanking telomeric VSG genes is not known because no such structure has been cloned in its entirety. Nonetheless, a fairly accurate picture, schematically drawn in Fig. 5, can be extracted from the separate pieces of information now available. Van der Ploeg *et al.* (1984b) succeeded in cloning a DNA segment that is, in all likelihood, derived from a telomere. The most striking feature of this DNA is the presence of over 50 tandemly repeated hexamers, which read 5′-CCCTAA-3′ from telomere to centromere. As nuclear DNA fragments that contain 5′-$(CCCTAA)_n$-3′ behave like chromosome ends and are abundant, it is probable that this sequence is the common denominator of trypanosome telomeres; the sequence at the very end remains unknown, however. The telomeric sequence 5′-$(CCCTAA)_n$-3′ is common to all trypanosomatid genera and concurs nicely with the consensus sequence for protozoan telomeres, 5′-$[CCCP_y(A_n)]_n$-3′ (Van der Ploeg *et al.*, 1984b,c; Blackburn and Challoner, 1984). Furthermore, trypanosome telomeres contain single-stranded breaks in both strands, an additional feature shared with telomeres of other protozoa (Blackburn and Challoner, 1984). Two other more or less telomere-specific repeats are found more internal in the chromosome analyzed by Van der Ploeg *et al.*, a 29-bp 5′-$(CCCTAA)_n$-3′-derived repeat and a T-rich element (dT_n stretches are encountered, when reading 5′ to 3′ from centromere to telomere, as shown in Fig. 5). The latter had been found previously downstream of a telomeric VSG gene, showing the linkage between this telomeric repeat and VSG genes at chromosome ends (De Lange *et al.*, 1983a).

IV. Activation of VSG Genes

VSG genes can be activated in various ways. On the basis of observations at the DNA level, three different pathways can be discerned (see Fig. 6). Gene switching can be achieved by an exchange of genes in an expression site (scheme A), by a switch from one expression site to another (scheme B), and by the duplication of a telomeric VSG gene and its surroundings (scheme C). The three pathways are discussed separately for reasons of clarity.

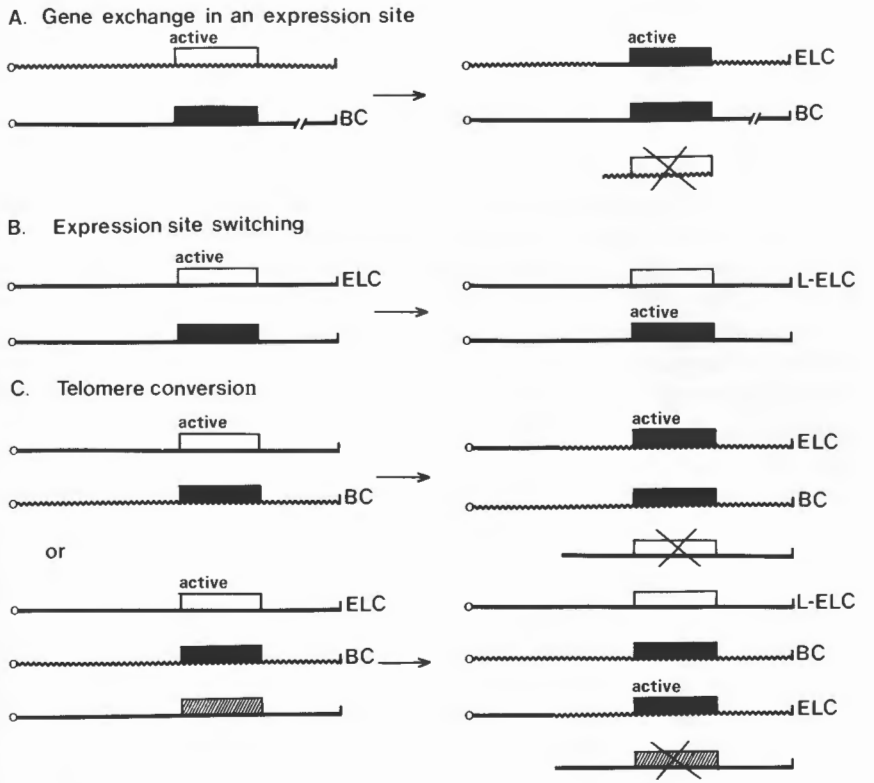

FIG. 6. Three modes of VSG gene activation. Schemes A, B, and C are discussed separately in Sections IV,A, B, and C, respectively. BC, Basic copy; ELC, expression-linked extra copy; L-ELC, lingering expression-linked extra copy. A gene with a cross symbolizes that this gene copy is destroyed during the switching event.

A. Exchange of VSG Genes in a Telomeric Expression Site

Chromosome-internal VSG are invariably activated by a duplicative transposition to a telomeric expression site. The expression-linked extra copy (ELC) of the gene thus formed is transcribed (Pays *et al.*, 1981a; Bernards *et al.*, 1981); the original basic copy (BC) of the gene is silent.

The transcription of ELC genes starts in the expression site (Bernards *et al.*, 1985; De Lange *et al.*, 1985), i.e., the duplicative transposition activates VSG genes by addition of a promoter (see Section VI,D). Here, the mechanism of duplicative transposition is discussed.

1. *The Structure of the Expression Site*

We have studied four separately generated ELC genes coding for VSG 118 and two for VSG 117 (Michels *et al.*, 1983). Their genomic environment, drawn in Fig. 7, is rather similar and characterized by three typical features. First, the ELC genes are proximal to a telomere, Second, upstream of the gene there are long barren regions of variable length. Third, 5′ of the barren region there is a constant set of restriction enzyme cutting sites, suggesting that all six ELCs occupy the same expression site. In support, the 118 and 117 ELC genes invariably reside on a chromosome that migrates in the mega-base-pair range (fraction II in Fig. 4), whereas the corresponding BC genes are housed in larger chromosomes that do not enter the gel (Van der Ploeg *et al.*, 1984a).

Some of these characteristics of the dominant acceptor site for ELC genes are typical for *T. brucei* strain 427. Other strains have different expression sites. For example, the expression site in *T. brucei* strain 1125 has a different restriction map and resides on a small (about 300-kb) chromosome (Pays *et al.*, 1983b; M. Guyaux, E. Pays, A. W. C. A. Cornelissen, and P. Borst, unpublished). However, all expression sites are telomeric and contain a barren region 5′ of the ELC gene.

2. *The Transposed Gene Segment*

The duplicative transposition of chromosome-internal VSG genes is probably a gene conversion of a previously expressed ELC by the incoming gene (Pays *et al.*, 1981b, 1983c; Borst and Cross, 1982; Michels *et al.*, 1982, 1983). The data on the boundaries of the transposed segment suggest as much, although other possibilities (see Borst, 1983) have not been rigorously excluded. The 3′ end of the transposed gene segment has been determined in many cases, often by direct sequence analysis of the 3′ end of VSG mRNA using the conserved 14-mer (see Fig. 3) as priming site. Michels *et al.* (1983) inferred that the recombination can occur anywhere in and around the conserved 3′ end of the gene (see Fig. 3), from 16 bp into the mature C-terminus (140 bp upstream of the 14-mer) to downstream of the 14-mer. For instance, the point of sequence divergence between the 118 BC gene and its ELCs has been found from 42 bp upstream of the stop codon (86 bp upstream of the 14-mer) to 237 bp downstream of the 14-mer (Michels *et al.*, 1983; T. De Lange and M. Timmers, unpublished).

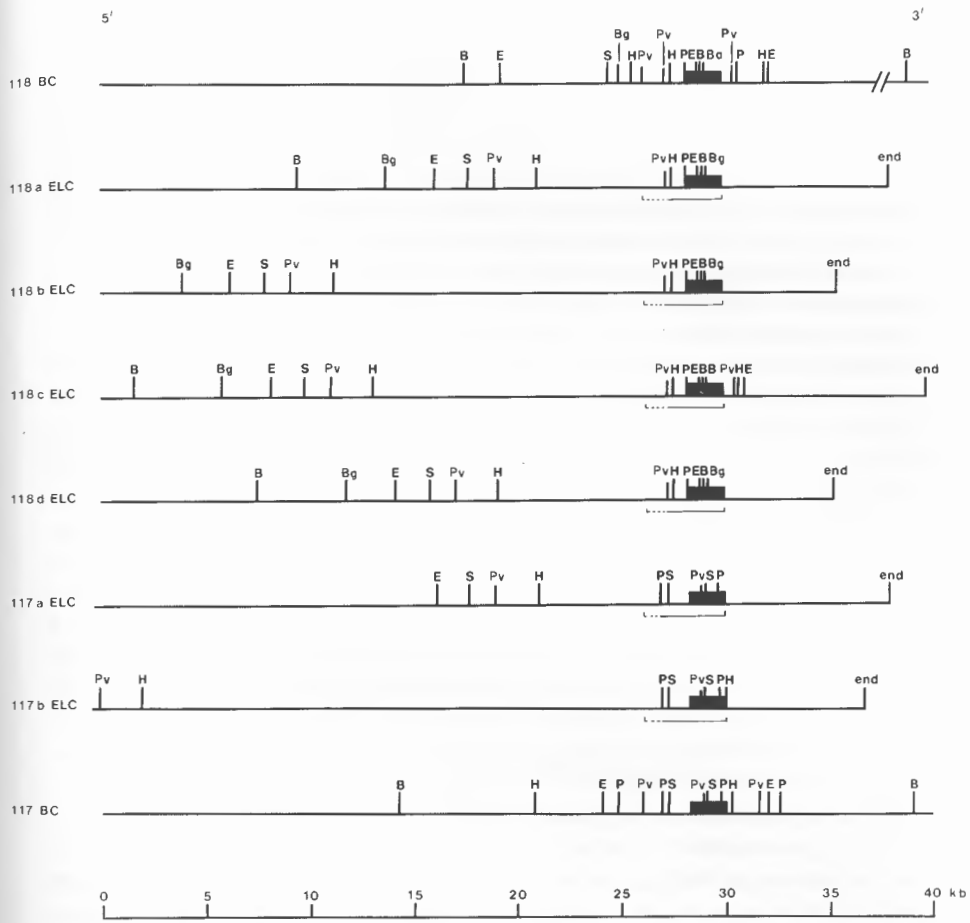

FIG. 7. Restriction maps of the basic copy (BC) genes for VSG 117 and 118 and six expression-linked extra copy (ELC) genes in six trypanosome clones that produce VSG 117 or 118 (see Michels *et al.*, 1983, and Section IV,A). The transposed gene segments are indicated beneath the maps. "End" denotes the end of the chromosome. B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; Pv, *Pvu*II; S, *Sal*I.

A second region of homology that may play a role in the recombination in the expression site is located downsteam of some VSG genes (see also Section V,A). Initially, the limited data on 3′-flanking regions indicated that each VSG gene has a unique sequence downstream. However, the analysis of additional genes now suggests that there are several classes of VSG genes with substantial sequence homology downstream. For example, the sequence downstream of the 118 BC gene is reiterated as judged

from the number of restriction fragments that are detected by a probe from this region under relaxed hybridization conditions (P. Leegwater, and A. Bernards, personal communication). By DNA sequence analysis we have found that one of these homologous sequences indeed flanks the 3′ end of a VSG gene, called 1.1006 (De Lange *et al.*, 1983a). Another example is the 3′-flanking sequence of the BC gene for VSG 117 that is homologous to the sequence downstream of a pseudo-VSG gene (1.1010), but not to the 118 BC gene area (Bernards *et al.*, 1985).

The 5′ end of the transposed gene segment resides some 1–2 kb upstream of the start codon in the region that contains the 70-bp repeats (see Fig. 3) (Van der Ploeg *et al.*, 1982b; Michels *et al.*, 1983; Liu *et al.*, 1983). For this reason we proposed that these repeats are instrumental in the gene activation process (Van der Ploeg *et al.*, 1982b; Liu *et al.*, 1983). Verification of this proposal has been delayed, largely due to the fact that the 5′ barren region that should contain the 5′ breakpoint has repeatedly eluded cloning in *E. coli*. Recently, however, Boothroyd and co-workers succeeded in cloning a small segment of the 5′ barren region upstream of the 117a ELC gene and determined the point where the sequence diverges from the BC gene (Campbell *et al.*, 1984a). As predicted, the 5′ breakpoint falls in a 70-bp repeat, the first in front of the BC gene. We have similarly determined the 5′ breakpoint in the 117b ELC gene (De Lange *et al.*, 1985). In this case, the first sequence divergence is in the unusual third repeat that lacks the 44-bp 3′ repeat half. The presence of this rare telltale repeat in the ELC strongly suggests that this sequence is cotransposed and therefore that the 5′ breakpoints are different for these independent activations of the same VSG gene.

Furthermore, the ELC clones allow a glimpse of the structure of the 5′ barren region. In both cases the cloned expression site segments are made up of 70-bp repeats. Whether other repeats inhabit the barren region remains to be seen. The information now available suffices to explain the length alterations of this DNA segment (see Fig. 7). Provided that the 70-bp repeats in which the crossover takes place can be selected randomly both in the expression site and in the incoming gene, as suggested by our results with the 117 gene, contraction and expansion can be understood as a consequence of gene replacement (Campbell *et al.*, 1984a; De Lange *et al.*, 1985). A switch to a gene carrying many 70-bp repeats (e.g., gene 221, see Section III,C) could increase the length of the barren region considerably and any switch could shorten it.

In the transposition of at least some telomeric VSG genes to the dominant expression site, the 5′ breakpoint of the transposed segment is probably also located within the 70-bp repeats a few kilobase pairs upstream of the start codon (Bernards *et al.*, 1984a). In contrast, the 3′ end of the

transposed segment in the two cases examined in detail resides at least 0.8 kb downstream of the gene (De Lange *et al.*, 1983a; Bernards *et al.*, 1984a). It is quite possible that here the complete telomere is transposed to the expression site by a telomere conversion. For lack of landmarks (e.g., restriction sites) in the last part of telomeres, this proposal is not easily verified.

The rules for gene conversion sketched here, although seemingly loose enough, are not always obeyed. A notable exception is an ELC of gene 1.1 in *T. brucei* strain 1125 that is accompanied by a second (earlier?) transposed element at the 5′ side, appropriately referred to as companion (Pays *et al.*, 1983b). The boundary between the companion and the 1.1 ELC does not carry 70-bp repeats. Possibly these repeats are present more upstream at the companion–expression site junction. Another aberrant gene conversion event occurred during switch-off of the 1.10 ELC gene in the same strain (Pays *et al.*, 1983d). Instead of being completely replaced by an unrelated VSG gene, the 1.10 ELC underwent a partial gene conversion by the related 1.1 gene. This gene conversion involved the 5′ end of the coding region, but not the cotransposed segment, and ended more than 630 bp upstream of the 3′ end, i.e., much further into the protein-coding region than observed in the six switches of the 117 and 118 genes in strain 427. I note that this aberration may be a consequence of the high-sequence homology of the interacting genes.

B. Switching of Expression Sites

Obviously, the exchange of genes in a single expression site as discussed above would elegantly explain the mutually exclusive expression of VSG genes. However, this attractive model does not apply in its simplest form.

Early observations by the group of Williams and others indicated that telomeric genes can be activated without duplication or any other DNA rearrangement within the range of view of restriction maps (Williams *et al.*, 1980, 1982; Young *et al.*, 1982, 1983a,b; Majiwa *et al.*, 1982; Donelson *et al.*, 1982; Penncavage *et al.*, 1983; Laurent *et al.*, 1984; Bernards *et al.*, 1984a; Myler *et al.*, 1984). These data could only be reconciled with the single expression site hypothesis by invoking a reciprocal translocation of the expression site telomere and the telomere with the activated gene. If this recombination would occur far upstream or within highly homologous regions, it could have gone unnoticed, especially because the chromosomal location of these genes had not been studied.

This hypothesis was recently disproved by a study on the gene that codes for VSG 221. When not expressed, this gene is present at one copy

per nucleus on a chromosome that remains close to the slot in PFG gels (Bernards *et al.,* 1984a; Van der Ploeg *et al.,* 1984a). When the gene is activated, it either follows the duplicative transposition route to the dominant expression site on a chromosomes of 1–2 × 10^3 kb (in fraction II in Fig. 4), or no significant change occurs within the gene or its surroundings. The latter situation is observed in trypanosome clone 221a, which expresses VSG 221. In this clone, the active 221 gene apparently occupies exactly the same telomere as the inactive gene, i.e., the gene is activated *in situ* without concomitant duplication (see Fig. 6B). The detailed restriction map reaching up to 45 kb 5′ of the coding region indicates that no gross rearrangement has accompanied the activation of the gene, but does not exclude that the gene is activated in the dominant expression site: a reciprocal translocation with the expression site telomere could fully account for the results. This was ruled out by Van der Ploeg *et al.* (1984a), who showed that the active single copy of the 221 gene in 221a trypanosomes still resides on a large chromosome that hardly leaves the slot of PFG gels, implying that the genome of *T. brucei* strain 427 contains at least two separate loci where VSG genes can be expressed. The elucidation of the activation mechanism of the 221 gene in its own telomere awaits the location of the start of transcription of this gene (see Section VI,D).

There is much additional circumstantial evidence in favor of multiple expression sites. For instance, in some cases the switch-off of an ELC is not accompanied by its destruction by the next gene in line. Rather, the old ELC persists in an inactive form, called a lingering ELC (see Fig. 6) (Michels *et al.,* 1984; Buck *et al.,* 1984; Laurent *et al.,* 1984). As far as can be judged from restriction analysis and PFG gel electrophoresis, the lingering ELC remains in the dominant expression site. Obviously, a gene in another expression site has been activated in these trypanosomes.

The most challenging question raised by the use of multiple expression sites concerns the coordination of the activity of these telomeres. The simplest models retain the idea that only one expression site is active at a time. These models invoke a unique mobile element that confers telomere activity in cis (Borst *et al.,* 1983) or a unique site at the nuclear matrix where a telomere must be bound to be active (Van der Ploeg and Cornelissen, 1984). Other models do not incorporate a unique activating entity, but instead invoke a stochastic activation mechanism. For instance, in analogy with the phase variation system of *Salmonella* (reviewed by Silverman and Simon, 1983), every expression site could contain an invertible element that activates the telomere in one orientation. Bernards *et al.* (1984b) have argued that a stochastic on and off switch of expression sites could be mediated by DNA modification. All but one expression site would be repressed due to modification, and every expression site would

have a low probability of losing its modification during replication. The attraction of this model resides in the fact that there is indirect experimental evidence for modified nucleotides in inactive telomeres (Bernards *et al.,* 1984b; Pays *et al.,* 1984). The restriction enzymes *Pst*I and *Pvu*II are incapable of cutting some sites in inactive telomeric VSG genes to completion. Most likely this is due to a partial modification of the recognition site. The level of modification is highest proximal to the telomere and strictly linked to the absence of gene expression.

In stochastic activation models, two types of aberrant trypanosomes are predicted, "zebras" that have two active expression sites and "naked" trypanosomes that have none. Neither have been observed. This is as expected because coatless trypanosomes are lysed in serum, and trypanosomes that do not cease production of the previous coat will probably be removed by the immune response (cf. Bernards *et al.,* 1984b). However, we have derived an exceptional trypanosome clone from variant 221 in which two expression sites appear to be simultaneously active (A. W. C. A. Cornelissen, P. Johnson, and P. Borst, unpublished). In this variant, the previously expressed 221 gene has been inactivated by an insertion of 30-kb DNA at a site about 3 kb upstream of start codon. The insertion prevents expression of the 221 gene from the VSG gene promoter far upstream (see Section VI,D), probably by providing a termination signal, since the region upstream of the insertion is transcribed, whereas the downstream area is silent. Hence, in this variant the expression site for VSG 221 is still active, although it yields no VSG mRNA. As the surface antigen of this variant (VSG 1.1) is produced from a VSG gene on another telomere, it appears that two expression sites are simultaneously active in this trypanosome clone. This argues against a mobile promoter or any other unique entity controlling VSG gene transcription and in favor of a stochastic mechanism of expression site activation/inactivation.

C. Telomere Conversion

Telomeric VSG genes are often switched on without concomitant duplication; in some cases, however, the activation is accompanied by the duplication of the gene and a large area upstream (see Fig. 6C) (Pays *et al.,* 1983d; Michels *et al.,* 1984; Laurent *et al.,* 1984; Van der Ploeg *et al.,* 1984d). Usually, the extent of the duplication has remained undetermined because it exceeds the limits of the available restriction maps (about 30 kb). The duplication of one telomere most likely occurs at the expense of another, i.e., by an extensive telomere conversion instead of by duplication of the whole chromosome.

Three observations argue in favor of telomere conversion. First, the genesis of a new active copy of the telomere that harbors the gene for VSG 1.8 in our stock was followed using PFG analysis of the chromosomal location of this gene. The ELC is found on a new chromosome of about 425 kb, whose appearance accompanies the loss of a smaller chromosome (about 375 kb) (Van der Ploeg *et al.,* 1984d). The simplest explanation for this set of events is a telomere conversion of the 375-kb chromosome by the 1.8 telomere, thereby changing its migration rate in PFG gels considerably. The second observation supporting telomere conversion was made on the antigenic switch of a trypanosome from clone 221a to expression of the telomeric gene for VSG 224 (J. M. Kooter, R. Moberts, A. Bernards, and P. Borst, unpublished). The 224 telomere in this case is duplicated over an area of less than 45 kb, as judged from the analysis of large restriction fragments on PFG gels. At the same time the previously active telomeric 221 gene is lost, together with at least 15 kb of upstream sequence. Although rigorous proof is still wanting, it is very likely that the 224 telomere has converted the 221 telomere during the switch. In fact, this rather drastic way of switching off a telomeric VSG gene seems to be a hallmark of gene 221. In five out of six inactivation events, the gene and its upstream region have been lost (Bernards *et al.,* 1984a; Liu *et al.,* 1984). Finally, Pays and co-workers (1983d) have reported a switch from expression of the gene for VSG 1.1C to expression of the telomeric gene for VSG 1.3B. In this switch the (telomeric) 1.1C gene is lost and the 1.3B gene together with at least 30 kb of upstream sequences are duplicated. Again, a conversion of the 1.1C telomere by the 1.3B telomere is the simplest explanation for these observations.

V. Changes in VSG Repertoire

A. The Order of VSG Gene Expression

Early serological data argue that the expression sequence of VSGs is nonrandom: some VSGs are predominant early and others usually appear later in a chronic infection (Van Meirvenne *et al.,* 1975a,b; Capbern *et al.,* 1977; Kosinski, 1980; Miller and Turner, 1981). More recent analyses support the hypothesis that the predominant early VSGs are encoded by genes with a permanent telomeric location and that these genes are switched on preferentially (Laurent *et al.,* 1983; Michels *et al.,* 1984; Myler *et al.,* 1984).

This is clearly seen in single relapse experiments. For instance, when trypanosome clone 118a (expressing VSG 118a) is treated with antibodies

against VSG 118, the few surviving trypanosomes that had apparently switched prior to the experiment were found to express either the telomeric gene for VSG 1.8 or the telomeric gene for VSG 1.1 (Michels *et al.,* 1984). An extreme case is the telomeric VSG 1 gene in *T. equiperdum* which is preferentially switched on in single relapse experiments and always dominates the early phase of chronic infections (Capbern *et al.,* 1977; Longacre *et al.,* 1983; Raibaud *et al.,* 1983). In contrast, chromosome-internal VSG genes are expressed late. Examples are the genes for VSG 118 and 117 that are activated reproducibly in the third week of chronic infections in rabbits (Michels *et al.,* 1983). Michels *et al.* (1984) elegantly showed that the genomic environment of the 118 gene determines the frequency with which it is activated. When the gene has a chromosome-internal position, it is never switched on in single relapse experiments. In contrast, when the 118 ELC is still present as a (telomeric) lingering ELC, it is switched on with high preference in such experiments. Similar results have been obtained by Laurent *et al.* (1984) with *T. brucei* strain 1125.

It is not clear whether there is an order of activation of the chromosome-internal genes later in infection. Capbern *et al.* (1977) found that in different chronic infections certain VSGs repeatedly come up at the same time, suggesting some order in the expression of VSGs. It is not excluded, however, that this order is the consequence of factors [growth rate, competition, host factors (see Kosinski, 1980)] other than a fixed sequence in the activation of VSG genes. Aside from these uncertainties, it could be argued that extensive homology downstream of VSG genes (see Section IV,A,2) may influence the sequence of gene conversion events in the expression site. This hypothesis has not been tested extensively yet, mainly because the sequences downstream of ELC genes are usually not present in genomic libraries in *E. coli.* The single exception to this rule is the downstream area of the 118c ELC that aberrantly has several restriction enzyme cutting sites (see Fig. 7). Analysis of this DNA indicated that it is derived from another (telomeric) VSG gene (1.1006) and, moreover, that the 3′-flanking sequences of the 1.1006 gene and the 118 BC gene are rather alike (De Lange *et al.,* 1983a). In this case, then, the recombination between these two genes may have been aided by the nucleotide sequence homology downstream of the 14-mer. More data are needed to assess whether such homologies indeed influence the order of VSG gene expression later in chronic infections.

Of course, an expression sequence based on nucleotide sequence homology would lead to many retrograde switches, yielding trypanosomes that probably would be destroyed by the host. However, this may be no intolerable burden for the trypanosome population. In view of the prefer-

red switch-on of telomeric genes, one may expect that late in a chronic infection trypanosomes often switch to a coat protein that has already been used.

B. Evolution of VSG Genes

Given the high frequency of gene conversion at telomeres and the strong selection for such changes, it is not surprising that telomeric VSG genes are particularly unstable. This instability is illustrated by a survey of different trypanosome isolates for the presence of three telomeric genes (221, 1.1006, and 1.1005). These genes are only present in strain 427 (13, 6, and 5 strains analyzed, respectively) (Frasch *et al.*, 1982; De Lange *et al.*, 1983a; Borst *et al.*, 1984). It has been argued that the enormous number of small chromosomes in *T. brucei* has evolved to increase the number of early telomeric genes that can be changed rapidly (Van der Ploeg *et al.*, 1984a). However, the presence of small chromosomes is not essential for antigenic variation: *T. equiperdum* and *T. b. gambiense* manage without them (Van der Ploeg *et al.*, 1984c; W. Gibson and P. Borst, unpublished).

Some chromosome-internal VSG genes also show a remarkable instability. We have compared about 40 kb surrounding the 118 BC gene in strains 1125 and 427 and found a deletion of one neighboring VSG gene in strain 1125 and a quadruplication of another in strain 427 (Borst *et al.*, 1984; W. Gibson, A. Bernards, and L. H. T. Van der Ploeg, unpublished). The high frequency of recombination in and around chromosome-internal VSG genes is also indicated by the loss of the 118 gene in 5 out of 13 trypanosome strains (Frasch *et al.*, 1982). Another chromosome-internal gene, the 117 gene, is much more stable (Pays *et al.*, 1982; Frasch *et al.*, 1982).

VI. Discontinuous Transcription

The analysis of transcripts coding for VSGs in *T. brucei* has revealed a novel mode of mRNA synthesis that may be a general feature of kinetoplastid flagellates. The first 35 nucleotides of many (all?) mRNAs are identical and come from highly reiterated genes that are not linked to structural genes. Transcription therefore appears to be a discontinuous process.

A. Many Trypanosomal mRNAs Share a Common 5′ Terminal Sequence

The common 5′ end sequence of trypanosomal mRNAs was first found on the transcripts coding for VSG 118 in *T. brucei*. As noted, the protein-coding part of VSG BC genes does not contain introns. Still, VSG mRNAs are transcribed from two separate exons. The first observation that pointed in this direction was reported by Bernards *et al.* (1981), who characterized a VSG cDNA that contains 9 bp at the 5′ end that are not encoded by the corresponding BC gene. Subsequent sequence analysis of the 5′ end of the mRNA for VSG 118 and the transposed 118 gene segment showed that the first 35 nucleotides of the messenger are absent from the transposed segment. Van der Ploeg *et al.* (1982c) therefore proposed that these 35 nucleotides are encoded by a mini-exon in the expression site. In apparent agreement with this supposition, Boothroyd and Cross (1982) found the same mini-exon sequence at the 5′ end of the transcript of the 117a ELC gene, which resides in the same expression site. A search for the mini-exon in this expression site yielded negative results: as judged from restriction analysis and quantitative hybridization experiments, the mini-exon is not present within 10 kb upstream of the 118d ELC gene (De Lange *et al.*, 1983b). In contrast, many mini-exons are present elsewhere in the genome. A 22-nucleotide probe for the mini-exon was found to hybridize to large arrays of a tandemly linked 1.35-kb repeat (De Lange *et al.*, 1983b; Nelson *et al.*, 1983). Sequence analysis showed that the complete mini-exon is present in these repeats, flanked by the sequence TTG/GTAGT that resembles the consensus sequence (CAG/GTPuAGT (see Mount, 1982) for splice donor sites in protein-coding genes (De Lange *et al.*, 1983b). We speculated that the expression site may contain one of these arrays of mini-exons and that this reiteration of initiation sites somehow would enhance the transcription of VSG genes. More recent evidence argues strongly against this model, however.

From the analysis of transcripts that do not code for surface antigens, it has become clear that the mini-exon sequence is not exclusive to the VSG messenger. Although VSG mRNA is very abundant, it accounts for at most 10% of the polyadenylated transcripts that hybridize to a 22-mer probe for the mini-exon (De Lange *et al.*, 1984b). The other RNAs are heterogeneous in size, also have the mini-exon sequence at a 5′-terminal position, but are not encoded in toto by the mini-exon repeat (De Lange *et al.*, 1984b). We have investigated 13 of these transcripts; another six were analyzed by Parsons *et al.* (1984a). Although the function of none of these RNAs is known, they appear to be part of the pool of messenger RNAs:

they are discretely sized transcripts of variable length and abundance that are encoded by single or low copy number genes whose expression is in some cases dependent on the differentiation stage of the trypanosomes. The simplest interpretation of these results is that the bulk of mRNAs in *T. brucei* has the same mini-exon-derived 5′ end. We note, however, that the existence of a class of protein-coding genes that yield mRNAs without the common 5′ end sequence cannot yet be excluded. The transcripts that were analyzed were selected for the presence of the mini-exon sequence. A systematic analysis of randomly picked mRNAs has not been carried out. Apart from VSG mRNAs, only two transcripts of known function thus far were available for study. These code for the α- and β-tubulins, and both have the mini-exon sequence, as shown by sandwich hybridization (T. De Lange, unpublished).

B. Evidence for Discontinuous Transcription

On theoretical grounds it seems unlikely that each structural gene has its own mini-exon upstream. Assuming that more than 1000 genes yield a transcript with the common 5′ end sequence, there are simply too few mini-exons (about 200 per nucleus, De Lange *et al.*, 1983b) to serve each gene. Even if mini-exons are used by considerably fewer genes, the strong clustering of the repeats that contain mini-exons would probably demand that several genes share one repeat array. Therefore, we and others raised the possibility that mini-exons and structural genes are not linked, i.e., that transcription is a discontinuous process (cf. Borst *et al.*, 1983; Campbell *et al.*, 1984a; De Lange *et al.*, 1984b; Parsons *et al.*, 1984a). Indeed, Parsons *et al.* (1984a) did not find mini-exons within 8 kb upstream of four genes that yield a transcript with a mini-exon-derived 5′ end. Another case in point is the genomic environment of the α- and β-tubulin genes. These two genes are closely linked on a 3.7-kb element that is repeated about 10 times per haploid genome in one cluster (Thomashow *et al.*, 1983; Seebeck *et al.*, 1983; T. De Lange, unpublished). The mini-exon is neither present on the 3.7-kb element nor within 13.5 kb upstream of the cluster (T. De Lange, unpublished).

A second line of evidence evolved from the analysis of the transcription of mini-exon repeats. Both Kooter *et al.* (1984) and Campbell *et al.* (1984b) have shown that these repeats are transcribed into an RNA of about 140 nucleotides that starts with the mini-exon. Kooter *et al.* also showed that this molecule is not a processing product: the rest of the mini-exon repeat is transcribed 750-fold less in isolated nuclei. Preliminary

data indicate that the mini-exon repeats and VSG genes are not transcribed by the same polymerase (Kooter and Borst, 1984). Whereas transcription of mini-exon genes in isolated nuclei is 80% inhibited by the addition of 200 μg of α-amanitin per milliliter, VSG gene transcription is not affected by α-amanitin at all. Unlike other protein-coding genes whose transcription is blocked by 20 μg of α-amanitin per milliliter, genes for surface antigens still yield nascent RNA in the presence of 1000 μg of α-amanitin per milliliter. Whether this remarkable α-amanitin resistance also exists *in vivo* remains to be seen. Why VSG genes are transcribed by a (partly) different polymerase than other protein-coding genes is also not known.

A third line of evidence in favor of discontinuous transcription has recently come from the analysis of an expressed VSG gene in *T. brucei* strain 1125 (M. Guyeaux, E. Pays, A. W. C. A. Cornelissen, and P. Borst, unpublished). In these trypanosomes the ELC gene for VSG 1.1A is located on a chromosome of about 300 kb that does not contain a mini-exon. Nevertheless, the mRNA for VSG 1.1A has the mini-exon sequence. Thus, unless the mini-exon in this case is split and therefore undetectable with the available probes, transcription of this ELC gene is a discontinuous process. A similar observation was made by Van der Ploeg *et al.* (1984d).

Currently, two models for discontinuous transcription prevail (De Lange *et al.*, 1984a; Parsons *et al.*, 1984a; Campbell *et al.*, 1984b). One of these proposes that the transcripts of mini-exons and structural genes are synthesized independently and that the mini-exon sequence is transferred posttranscriptionally from the 140-nucleotide transcript onto mRNAs by a splicing event that involved two rather than one molecule. The other model says that (part of) the 140-nucleotide mini-exon transcript functions as a primer for the initiation of transcription in front of structural genes. A conventional splicing step would subsequently link the two parts of the mature mRNA.

The more fundamental question of why discontinuous transcription has evolved has also not been answered satisfactorily. Is the trypanosome RNA polymerase II crippled in that it needs a primer to start transcription or in that it does not start at a specific site and the spliced on mini-exon has to create a defined 5′ end? Does the mini-exon sequence have a function in the RNA, e.g., in stability, transport, or translation? If so, why is this sequence not encoded in each gene? Perhaps the clustering of mini-exon genes in tandem arrays helps to keep this sequence homogeneous. Elucidation of the mechanism of discontinuous transcription may shed some light on these questions.

C. Structure of Mini-Exon Genes

The repeat elements that carry mini-exons have been investigated in several trypanosome species. In general, the elements are small (1.35 kb or less) and tandemly linked in large clusters: tandem arrays are found in all *Trypanozoon* species and in *T. vivax, T. congolense,* and *T. cruzi* (De Lange *et al.*, 1983b, 1984b; Nelson *et al.*, 1983, 1984). A probe for the mini-exon of *T. brucei* furthermore detects putatively repeated elements in other trypanosomatid genera, such as *Crithidia, Leptomonas, Herpetomonas, Phytomonas,* and *Leishmania,* suggesting that discontinuous transcription is a general characteristic of trypanosomatids and may be used by other kinetoplastid flagellates (Nelson *et al.*, 1984; De Lange *et al.*, 1984b). The number of mini-exon repeat clusters in *T. brucei* is probably not much larger than 20 and many repeats are linked to more than 10 others (De Lange *et al.*, 1983b, 1984b). In addition, a few elements are linked in smaller clusters and others occur singly dispersed (Nelson *et al.*, 1983; Parsons *et al.*, 1984b).

The repeat units from *T. brucei, T. vivax,* and *T. cruzi* have been cloned and subjected to nucleotide sequence analysis (De Lange *et al.*, 1984b). From the comparison of these repeats, it has become clear that very little is conserved outside the mini-exon and its direct vicinity. The most prominent conserved features are (1) a 10-mer around the first nucleotide of the mini-exon, (2) the 3′ part of the mini-exon and the putative 5′ splice donor site, and (3) a T-rich stretch downstream (see Fig. 8). The T-rich stretch may play a role in defining the 3′ end of the mini-exon transcripts because the 3′ end of the mini-exon repeat transcript in *T. brucei* maps just in front of it (De Lange *et al.*, 1984b; Kooter *et al.*, 1984; Campbell *et al.*, 1984b). The significance and function of the other two conserved blocks are unclear. The conservation within the mini-exon could reflect a role in the RNA or in transcription of the genes. The same holds for the apparent conservation at the 3′ border of the mini-exon. It is tempting to speculate that this sequence functions in the splicing event that links mini-exons to the rest of messenger RNAs. This idea is supported by the fact that VSG BC genes are flanked by a sequence that remotely resembles the 3′ splice acceptor site in protein-coding genes in other eukaryotes, e.g., the AG rule for intron ends is obeyed (Van der Ploeg *et al.*, 1982c; Boothroyd and Cross, 1982; Liu *et al.*, 1983), but other possibilities are not ruled out. Information on which polymerase transcribes mini-exon genes and which nucleotides are necessary for transcription should clarify this matter.

Thus far, conflicting results as to the identity of the polymerase making the mini-exon transcript have been obtained. The effect of α-amantin on the transcription of these genes in isolated nuclei, i.e., inhibition between 20 and 200 μg/ml (Kooter and Borst, 1984), suggests that a type III polymerase transcribes these genes (cf. Chambon, 1975). However, recent experiments suggest that the 140-nucleotide mini-exon transcript contains a cap structure (P. Laird and P. Borst, unpublished). This is atypical for a RNA polymerase III product, and trypanosomes apparently obey this rule, since their 5 S RNA does not contain a cap. Finally, the conserved sequence blocks in the mini-exon repeats do not resemble the consensus sequence for RNA polymerase III promoters (RGYNNRRYNGG- 30–60 bp -GATTCRANNC; see Ciliberto *et al.*, 1983). Rather, the presence of a conserved 10-mer around the initiation site is reminiscent of an RNA polymerase I transcription unit (Bach *et al.*, 1981; Financzek *et al.*, 1982; Verbeet *et al.*, 1984).

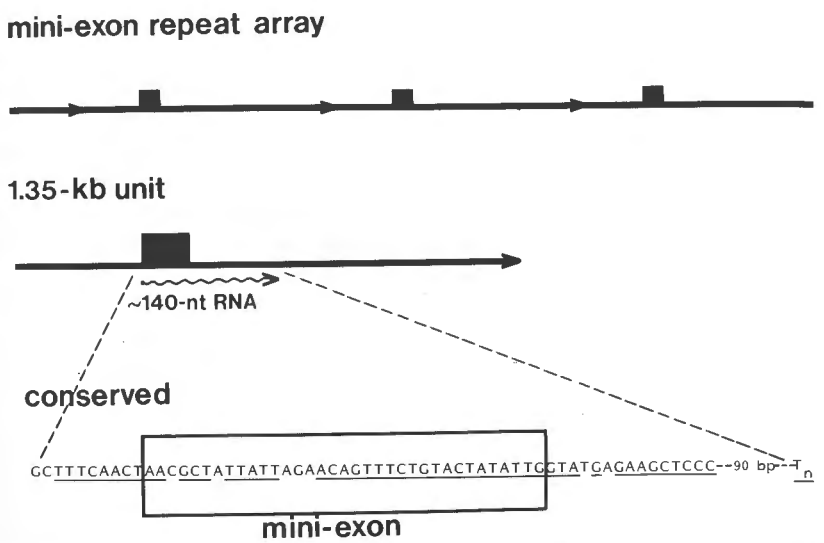

FIG. 8. The structure of mini-exon genes in *T. brucei.* Part of a tandem array of mini-exon repeats is drawn schematically, and below (part of) one of the 1.35-kb repeat units is shown with the 140-nucleotide transcript that is encoded by mini-exon genes. The lower part shows the sequence of the mini-exon and its immediate flanking sequence in *T. brucei.* The conserved elements (also found in the mini-exon repeats of *T. vivax* and *T. cruzi*) are underlined. The position of the T-rich stretch (T_n) varies in different trypanosome species. (See also Section VI,C.)

D. Transcription of VSG Genes

Notwithstanding the paucity of information on the discontinuous transcription process, it is anticipated that structural genes have their own promoter for transcription, i.e., a site from which the (re?)initiation of transcription is directed. A search for this entity is especially interesting in the case of VSG genes, for it could potentially reveal the mechanism of

VSG gene activation and provide insight into the control of expression sites. Early observations on the transcripts of two ELC genes showed that the cotransposed segment yields discrete, steady-state RNAs, albeit at very low levels and with an unknown relation to the mature VSG mRNA (Pays *et al.*, 1982, 1983b; Van der Ploeg *et al.*, 1982c). Similar transcripts derive from the upstream area of the telomeric gene for VSG 221 (Bernards *et al.*, 1985). Analysis of steady-state RNA appeared an inadequate way for finding the start of transcription, however, because very long primary transcripts could escape detection due to rapid degradation.

Therefore, we turned to the analysis of labeled nascent RNAs synthesized in isolated nuclei. Two genes were studied in this fashion: the ELC of the nontelomeric gene for VSG 117 in the dominant expression site of trypanosome strain 427 and the *in situ* activated telomeric gene for VSG 221 (De Lange *et al.*, 1985; Bernards *et al.*, 1985). From the 117 ELC gene, nascent RNAs were synthesized that hybridize to the complete transposed gene segment. Furthermore, an impressive amount of nascent RNAs are produced that have 70-bp repeat sequences, suggesting that a large part of the 5′ barren region is transcribed. We inferred that transcription of this ELC gene starts in the expression site; i.e., the transposition of nontelomeric VSG genes activates these genes by a process akin to promoter addition. In support of this interpretation, we have found that transcription of 70-bp repeats shows the resistance to α-amanitin, characteristic for VSG gene transcription units (Kooter and Borst, 1984). Where transcription starts and how the mini-exon sequence is juxtaposed onto the transcript may become clear when sequences beyond the 5′ barren region become available as cloned recombinant DNA.

Likewise, the region as far as 22 kb upstream of the 221 gene, including sequences beyond the short 5′ barren region, are transcribed at approximately the same rate as the coding region by an RNA polymerase that is insensitive to α-amanitin *in vitro* (Bernards *et al.*, 1985; Kooter and Borst, 1984; J. M. Kooter, R. Wagter, and P. Borst, unpublished). This indicates that the initiation of transcription occurs at least 22 kb upstream of the coding region. Since the promoter of this VSG gene has not been located yet, it is not clear how this gene and others like it are activated in their own telomere.

VII. Outlook

The study of the genetics of antigenic variation started in 1979 with the isolation of the first cDNAs of VSG mRNAs. Now, some 5 years later, the field has come of age: it has turned out to be a rewarding model system for sophisticated gene switching events. No other genome has yet yielded such an overwhelming number of DNA rearrangements, linked to development and evolution, that are worthwhile to study in detail. Furthermore, the study of VSG genes has allowed a closer look at the structure and behavior of chromosome ends. And finally, the unique, discontinuous transcription system of trypanosome protein-coding genes has given a glimpse of unexpected possibilities in genome usage.

There are still a number of questions one would like to see answered. What is the nature of the activation process of telomeric VSG genes and how is the activity of different expression sites regulated? What is the precise mechanism of gene conversion in the expression site and the frequent extensive conversion of telomeres? How do the ends of trypanosome chromosomes grow and why is there a difference in growth rate and stability of actively transcribed telomeres? What is the mechanism of discontinuous transcription and why has it evolved?

These questions can be answered partly by techniques now available, but many aspects will need the development of *in vitro* or (preferably) *in vivo* systems in which manipulated DNA segments can be tested for their biological activity. It is high time that trypanosome molecular biology lost these last growing pains.

Acknowledgments

I am grateful to Drs. P. Borst, A. Bernards, K. Osinga, W. Gibson, A. Cornelissen, P. Johnson, L. Van der Ploeg, M. Guyaux, E. Pays, J. Kooter, and P. Laird for communication of unpublished data and to Drs. P. Borst, A. Bernards, K. Osinga, W. Gibson, P. Johnson, and G. A. M. Cross for valuable comments on the manuscript. This work was supported in part by a grant from the Foundation for Fundamental Biological Research (BION) which is subsidized by the Netherlands Organization for the Advancement of Pure Research (ZWO).

References

Bach, R., Allet, B., and Crippa, M. (1981). *Nucleic Acids Res.* **9,** 5311.

Barry, J. D., Hadjuk, S. L., and Vickerman, K. (1979). *Trans. R. Soc. Trop. Med. Hyg.* **73,** 205.

Barry, J. D., Crowe, J. S., and Vickerman, K. (1983). *Nature (London)* **306,** 699.

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

Bernards, A., Michels, P. A. M., Lincke, C. R., and Borst, P. (1983). *Nature (London)* **303,** 592.

Bernards, A., De Lange, T., Michels, P. A. M., Huisman, M. J., and Borst, P. (1984a). *Cell* **36,** 163.
Bernards, A., Van Harten-Loosbroek, N., and Borst, P. (1984b). *Nucleic Acids Res.* **12,** 4153.
Bernards, A., Kooter, J. M., and Borst, P. (1985). *Mol. Cell. Biol.,* in press.
Blackburn, E. H., and Challoner, P. B. (1984). *Cell* **36,** 447.
Bloom, B. R. (1979). *Nature (London)* **279,** 21.
Boothroyd, J. C., and Cross, G. A. M. (1982). *Gene* **20,** 281.
Boothroyd, J. C., Cross, G. A. M., Hoeijmakers, J. H. J., and Borst, P. (1980). *Nature (London)* **288,** 624.
Boothroyd, J. C., Paynter, C. A., Cross, G. A. M., Bernards, A., and Borst, P. (1981). *Nucleic Acids Res.* **9,** 4735.
Borst, P. (1983). *In* "Mobile Genetic Elements" (J. A. Shapiro, ed.), p. 621. Academic Press, New York.
Borst, P., and Cross, G. A. M. (1982). *Cell* **29,** 291.
Borst, P., and Cross, G. A. M. (1982). *Cell* **29,** 291.
Borst, P., Fase-Fowler, F., Frasch, A. C. C., Hoeijmakers, J. H. J., and Weijers, P. J. (1980). *Mol. Biochem. Parasitol.* **1,** 221.
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). *Cold Spring Harbor Symp. Quant. Biol.* **45,** 935.
Borst, P., Van der Ploeg, M., Van Hoek, J. F. M., Tas, J., and James J. (1982). *Mol. Biochem. Parasitol.* **6,** 13.
Borst, P., Bernards, A., Van der Ploeg, L. H. T., Michels, P. A. M., Liu, A. Y. C., De Lange, T., and Kooter, J. M. (1983). *Eur. J. Biochem.* **137,** 383.
Borst, P., Bernards, A., Van der Ploeg, L. H. T., Michels, P. A. M., Liu, A. Y. C., De Lange, T., and Sloof, P. (1984). *In* "Molecular Biology of Host-Parasite Interactions" (N. Agabian and H. Eisen, eds.), p. 205. Liss, New York.
Bridgen, P. J., Cross, G. A. M., and Bridgen, J. (1976). *Nature (London)* **263,** 613.
Brown, R. C., Evans, D. A., and Vickerman, K. (1973). *Int. J. Parasitol.* **3,** 691.
Buck, G. A., Longacre, S., Raibaud, A., Hibner, U., Giroud, C., Baltz, T., Baltz, P., and Eisen, H. (1984). *Nature (London)* **307,** 563.
Campbell, D. A., Van Bree, M., and Boothroyd, J. C. (1984a). *Nucleic Acids Res.* **12,** 2759.
Campbell, D. A., Thornton, D. A., and Boothroyd, J. C. (1984b). *Nature (London)* **311,** 350.
Capbern, A., Giroud, C., Baltz, T., and Mattern, P. (1977). *Exp. Parasitol.* **42,** 6.
Cardoso de Almeida, M. L., and Turner, M. (1983). *Nature (London)* **302,** 349.
Chambon, P. (1975). *Annu. Rev. Biochem.* **44,** 613.
Ciliberto, G., Castagnoli, L., and Cortese, R. (1983). *Curr. Top. Dev. Biol.* **18,** 59.
Cross, G. A. M. (1975). *Parasitology* **71,** 393.
Cross, G. A. M. (1977). *Ann. Soc. Belge Med. Trop.* **57,** 389.
Cross, G. A. M. (1978). *Proc. R. Soc. London Ser. B* **202,** 55.
Cross, G. A. M., and Johnson, J. G. (1976). *In* "Biochemistry of Parasites and Host-Parasite Relationships" (H. van den Bossche, ed.), p. 413. Elsevier, Amsterdam.
Crowe, J. S., Barry, J. D., Luckins, A. G., Ross, C. A., and Vickerman, K. (1983). *Nature (London)* **306,** 389.
De Lange, T., and Borst, P. (1982). *Nature (London)* **299,** 451.
De Lange, T., Kooter, J. M., Michels, P. A. M., and Borst, P. (1983a). *Nucleic Acids Res.* **11,** 8149.
De Lange, T., Liu, A. Y. C., Van der Ploeg, L. H. T., Borst, P., Tromp, M. C., and Van Boom, J. H. (1983b). *Cell* **34,** 891.

De Lange, T., Michels, P. A. M., Veerman, H. J. G., Cornelissen, A. W. C. A., and Borst, P. (1984a). *Nucleic Acids Res.* **12,** 3777.
De Lange, T., Berkvens, T. M., Veerman, H. J. G., Frasch, A. C. C., Barry, J. D., and Borst, P. (1984b). *Nucleic Acids Res.* **12,** 4431.
De Lange, T., Luirink, J., Kooter, J. M., and Borst, P. (1985). *EMBO J.,* in press.
Donelson, J. E., Young, J. R., Dorfman, D. M., Majiwa, P. A. O., and Williams, R. O. (1982). *Nucleic Acids Res.* **10,** 6581.
Dorfman, D. M., and Donelson, J. E. (1984). *Nucleic Acids Res.* **12,** 4907.
Doyle, J. J. (1977). *In* "Immunity of Blood Parasites of Animals and Man" (L. H. Miller, J. A. Pino, and J. McKelvey, Jr., eds.), p. 31. Plenum, New York.
Doyle, J. J., Hirumi, H., Hirumi, K., Lupton, E. N., and Cross, G. A. M. (1980). *Parasitology* **80,** 359.
Duvillier, G., Nouvelot, A., Richet, C., Baltz, T., and Degand, P. (1983). *Biochem. Biophys. Res. Commun.* **114,** 119.
Englund, P. T., Hadjuk, S. L., and Marini, J. C. (1982). *Annu. Rev. Biochem.* **51,** 695.
Ferguson, M. A. J., and Cross, G. A. M. (1984). *J. Mol. Chem.* **259,** 3011.
Ferguson, M. A. J., Haldar, K., and Cross, G. A. M. (1985). *J. Biol. Chem.,* in press.
Financzek, I., Mizumoto, K., and Muramatsu, M. (1982). *Gene* **18,** 115.
Frasch, A. C. C., Borst, P., and Van den Burg, J. (1982). *Gene* **17,** 197.
Gibson, W. C., Marshall, T. F. de C., and Godfrey, D. G. (1980). *Adv. Parasitol.* **18,** 175.
Gibson, W. C., Osinga, K. A., Michels, P. A. M., and Borst, P. (1985). *Mol. Biochem. Parasitol.,* in press.
Hoeijmakers, J. H. J., Borst, P., Van den Burg, J., Weissmann, C., and Cross, G. A. M. (1980). *Gene* **8,** 391.
Holder, A. A. (1983). *Biochem. J.* **209,** 261.
Johnson, J. G., and Cross, G. A. M. (1976). *J. Protozool.* **24,** 587.
Kooter, J. M., and Borst, P. (1984). *Nucleic Acids Res.* **12,** 9457.
Kooter, J. M., De Lange, T., and Borst, P. (1984). *EMBO J.* **3,** 2387.
Kosinski, R. J. (1980). *Parasitology* **80,** 343.
Laurent, M., Pays, E., Magnus, E., Van Meirvenne, N., Matthijssens, G., Williams, R. O., and Steinert, M. (1983). *Nature (London)* **302,** 263.
Laurent, M., Pays, E., Van der Werf, A., Aerts, D., Magnus, E., Van Meirvenne, M., and Steinert, M. (1984). *Nucleic Acids Res.* **12,** 8319.
Liu, A. Y. C., Van der Ploeg, L. H. T., Rijsewijk, F. A. M., and Borst, P. (1983). *J. Mol. Biol.* **167,** 57.
Liu, A. Y. C., Michels, P. A. M., Bernards, A., and Borst, P. (1984). *J. Mol. Biol.* **182,** 383.
Longacre, S., Hibner, U., Raibaud, A., Eisen, H., Baltz, T., Giroud, C., and Baltz, D. (1983). *Mol. Cell. Biol.* **3,** 399.
Lumsden, W. H. R., and Evans, P. A., eds. (1976). "Biol Kinetoplastida," Vol. 1. Academic Press, London.
Lumsden, W. H. R., and Evans, P. A., eds. (1979). "Biol Kinetoplastida," Vol. 2. Academic Press, London.
McConnel, J., Gurnett, A. M., Cordingley, J. S., Walker, J. E., and Turner, M. J. (1981). *Mol. Biochem. Parasitol.* **4,** 226.
McConnel, J., Turner, M. J., and Rovis, L. (1983). *Mol. Biochem. Parasitol.* **8,** 119.
Majiwa, P. A. O., Young, J. R., Englund, P. T., Shapiro, S. Z., and Williams R. O. (1982). *Nature (London)* **297,** 514.
Majumder, H. K., Boothroyd, J. C., and Weber, H. (1981). *Nucleic Acids Res.* **9,** 4745.

Matthyssens, G., Michiels, F., Hamers, R., Pays, E., and Steinert, M. (1981). *Nature (London)* **293,** 230.
Merrit, S. C., Tschudi, C., Konigsberg, W. H., and Richards, F. F. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 1536.
Michels, P. A. M., Bernards, A., Van der Ploeg, L. H. T., and Borst, P. (1982). *Nucleic Acids Res.* **10,** 2353.
Michels, P. A. M., Liu, A. Y. C., Bernards, A., Sloof, P., Van der Bijl, M. M. W., Schinkel, A. H., Menke, H. H., Borst, P., Veeneman, G. H., Tromp, M. C., and Van Boom, J. H. (1983). *J. Mol. Biol.* **166,** 537.
Michels, P. A. M., Van der Ploeg, L. H. T., Liu, A. Y. C., and Borst, P. (1984). *EMBO J.* **3,** 1345.
Miller, E. N., and Turner, M. J. (1981). *Parasitology* **82,** 63.
Miller, E. N., Allan, L. M., and Turner, M. J. (1984). *Mol. Biochem. Parasitol.*, in press.
Mount, S. M. (1982). *Nucleic Acids Res.* **10,** 459.
Myler, P., Nelson, R. G., Agabian, N., and Weber, H. (1984). *Nature (London)* **309,** 282.
Nelson, R. G., Parsons, M., Barr, P. J., Stuart, K., Selkirk, M., and Agabian, N. (1983). *Cell* **34,** 901.
Nelson, R. G., Parsons, M., Selkirk, M., Newport, G., Barr, P. J., and Agabian, N. (1984). *Nature (London)* **308,** 665.
Opperdoes, F. R. (1984). *Br. Med. Bull.* **41,** 130.
Overath, P., Czichos, J., Stock, U., and Nonnengaesser, C. (1983). *EMBO J.* **2,** 1721.
Parsons, M., Nelson, R. G., Watkins, K. P., and Agabian, N. (1984a). *Cell* **38,** 309.
Parsons, M., Nelson, R. G., Stuart, K., and Agabian, N. (1984b). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 684.
Pays, E., Delronche, M., Lheureux, M., Vervoort, T., Bloch, J., Gannon, F., and Steinert, M. (1980). *Nucleic Acids Res.* **8,** 5965.
Pays, E., Lheureux, M., and Steinert, M. (1981a). *Nature (London)* **292,** 265.
Pays, E., Van Meirvenne, N., LeRay, D., and Steinert, M. (1981b). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 2673.
Pays, E., Lheureux, M., Vervoort, T., and Steinert, M. (1981c). *Mol. Biochem. Parasitol.* **4,** 349.
Pays, E., Lheureux, M., and Steinert, M. (1982). *Nucleic Acids Res.* **10,** 3149.
Pays, E., Laurent, M., Delinte, K., Van Meirvenne, N., and Steinert, M. (1983a). *Nucleic Acids Res.* **11,** 8137.
Pays, E., Van Assel, S., Laurent, M., Dero, B., Michiels, F., Kronenberger, P., Matthyssens, G., Van Meirvenne, N., Le Ray, D., and Steinert, M. (1983b). *Cell* **34,** 359.
Pays, E., Van Assel, I., Laurent, M., Darville, M., Vervoort, T., Van Meirvenne, N., and Steinert, M. (1983c). *Cell* **34,** 371.
Pays, E., Delauw, M. F., Van Assel, S., Laurent, M., Vervoort, T., Van Meirvenne, N., and Steinert, M. (1983d). *Cell* **35,** 721.
Pays, E., Delauw, M. F., Laurent, M., and Steinert, M. (1984). *Nucleic Acids Res.* **12,** 5235.
Penncavage, N. A., Julius, M. A., and Marcu, K. B. (1983). *Nucleic Acids Res.* **11,** 8343.
Raibaud, A., Gaillard, C., Longacre, S., Hibner, U., Buck, G., Bernardi, G., and Eisen, H. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 4306.
Rice-Ficht, A. C., Chen, K. K., and Donelson, J. E. (1981). *Nature (London)* **294,** 53.
Rice-Ficht, A. C., Chen, K. K., and Donelson, J. E. (1982). *Nature (London)* **298,** 676.
Rifkin, M. R. (1978). *Proc. Natl. Acad. Sci. U.S.A.* **75,** 3450.
Rovis, L., and Dube, D. K. (1981). *Mol. Biochem. Parasitol.* **4,** 77.
Schwartz, D. C., and Cantor, C. R. (1984). *Cell* **37,** 67.

Seebeck, T., Whittaker, P. A., Imboden, M. A., Hardman, N., and Braun, R. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 4634.
Silverman, M., and Simon, M. (1983). *In* "Mobile Genetic Elements" (J. A. Shapiro, ed.), p. 537. Academic Press, New York.
Tait, A. (1980). *Nature (London)* **287,** 536.
Thomashow, L. S., Milhausen, M., Rutter, W. J., and Agabian, N. (1983). *Cell* **32,** 35.
Van der Ploeg, L. H. T., and Cornelissen, A. W. C. A. (1984). *Philos. Trans. R. Soc. London Ser. B* **307,** 13.
Van der Ploeg, L. H. T., Valerio, D., De Lange, T., Bernards, A., Borst, P., and Grosveld, F. G. (1982a). *Nucleic Acids Res.* **10,** 5905.
Van der Ploeg, L. H. T., Bernards, A., Rijsewijk, F. A. M., and Borst, P. (1982b). *Nucleic Acids Res.* **10,** 593.
Van der Ploeg, L. H. T., Liu, A. Y. C., Michels, P. A. M., De Lange, T., Borst, P. Manjumder, H. K., Weber, H., Veenemen, G. H., and Van Boom, J. (1982c). *Nucleic Acids Res.* **10,** 3591.
Van der Ploeg, L. H. T., Schwartz, D. C., Cantor, C. R., and Borst, P. (1984a). *Cell* **37,** 77.
Van der Ploeg, L. H. T., Liu, A. Y. C., and Borst, P. (1984b). *Cell* **36,** 459.
Van der Ploeg, L. H. T., Cornelissen, A. W. C. A., Barry, J. D., and Borst, P. (1984c). *EMBO J.* **3,** 3109.
Van der Ploeg, L. H. T., Cornelissen, A. W. C. A., Michels, P. A. M., and Borst, P. (1984d). *Cell* **39,** 213.
Van Meirvenne, N., Janssens, P. G., and Magnus, E. (1975a). *Ann. Soc. Belge Med. Trop.* **55,** 1.
Van Meirvenne, N., Janssens, P. G., Magnus, E., Lumsden, W. H. R., and Herbert, W. J. (1975b). *Ann. Soc. Belge Med. Trop.* **55,** 25.
Verbeet, M. P., Klootwijk, J., Van Heerikhuizen, H., Fontijn, R. D., Vreugdenhil, E., and Planta, R. J. (1984). *Nucleic Acids Res.* **12,** 1137.
Vickerman, K. (1969). *J. Cell Sci.* **5,** 163.
Vickerman, K. (1978). *Nature (London)* **273,** 613.
Vickerman, K., and Barry, J. D. (1982). *In* "Immunology of Parasitic Infections" (S. Cohen and K. S. Warren, eds.), p. 204. Blackwell, Oxford.
Williams, R. O., Marcu, K. B., Young, J. R., Rovis, L., and Williams, S. C. (1978). *Nucleic Acids Res.* **5,** 3171.
Williams, R. O., Young, J. R., and Majiwa, P. A. O. (1979). *Nature (London)* **282,** 847.
Williams, R. O., Young, J. R., Majiwa, P. A. O., Doyle, J. J., and Shapiro, S. Z. (1980). *Am. J. Trop. Med. Hyg.* **29,** 1037.
Williams, R. O., Young, J. R., and Majiwa, P. A. O. (1982). *Nature (London)* **299,** 417.
Young, J. R., Donelson, J. E., Majiwa, P. A. O., Shapiro, S. Z., and Williams, R. O. (1982). *Nucleic Acids Res.* **10,** 803.
Young, J. R., Shah, J. S., Matthyssens, G., and Williams, R. O. (1983a). *Cell* **32,** 1149.
Young, J. R., Miller, E. N., Williams, R. O., and Turner, M. J. (1983b). *Nature (London)* **306,** 196.