Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Cell, Vol. 38, 309-316, August 1984, Copyright © 1984 by MIT 0092-8674/84/080309-08 $02.00/0

Trypanosome mRNAs Share a Common 5′ Spliced Leader Sequence

Marilyn Parsons, Richard G. Nelson, Kenneth P. Watkins, and Nina Agabian
University of Washington
Department of Biochemistry SJ-70
Seattle, Washington 98195

Summary

A 5′-terminal leader sequence of 35 nucleotides was found to be present on multiple trypanosome RNAs. Based on its representation in cDNA libraries, we estimate that many, if not all, trypanosome mRNAs contain this leader. This same leader was originally identified on mRNAs encoding the molecules responsible for antigenic variation, variant surface glycoproteins. Studies of selected cDNAs containing this leader sequence revealed that leader-containing transcripts can be stage-specific, stage-regulated, or constitutive. They can be abundant or rare, and transcribed from single or multigene families. No linkage between the genomic leader sequences and the structural gene exons was observed. Possible mechanisms by which the leader sequences are added to trypanosome mRNAs are discussed.

Introduction

The surface of pathogenic African trypanosomes is covered with a densely packed coat composed of a single protein species, the variant surface glycoprotein (VSG). The VSG gene repertoire of the trypanosome contains some 300 to 1000 genes (Van der Ploeg et al., 1982a), yet only one VSG gene at a time is expressed. Switching transcription from one VSG gene to another leads to antigenic variation, allowing the parasite to evade the immune response of its mammalian host. The result is the relapsing parasitemia of African sleeping sickness. Some of the molecular and biological aspects of antigenic variation have been recently reviewed (Parsons et al., 1984a).

The molecular mechanism that assures the transcriptional activation of a single VSG gene remains unknown. Comparisons of the genomic contexts of VSG genes have revealed that certain VSG genes undergo duplication when activated (Hoeijmakers et al., 1980a; Pays et al., 1981a; Longacre et al., 1983; Parsons et al., 1983). The new expression-linked copy is located at a different site in the genome and is transcribed (Pays et al., 1981b; Bernards et al., 1981). Other VSG genes appear to be activated in situ, as they undergo no detectable alteration in genomic organization when expressed (Young et al., 1982). Transcriptionally active VSG genes invariably reside in regions of the genome relatively devoid of restriction enzyme recognition sites (the "barren" regions) (Van der Ploeg et al., 1982b) and close to what appears to be a telomere (DeLange and Borst, 1982; Williams et al., 1982). These data suggest that some aspects of the genomic location of VSG genes may be important for their activation. However, analyses of restriction enzyme sites located 5′ to the barren regions indicates that there are several sites in the genome from which VSG genes can be transcribed (Longacre et al., 1983; Myler et al., submitted; Allison et al., submitted). Furthermore, while a VSG gene must reside in one of these sites in order to be transcribed, the opposite is not true; occupation of a particular site does not guarantee transcription (Buck et al., 1984; Allison et al., submitted).

Whether transcribed from duplication- or nonduplication-activated genes or from genes residing in similar or distinct genomic locations, all VSG mRNAs share the same 35 nucleotide sequence at their 5′ terminus (Van der Ploeg et al., 1982c; Boothroyd and Cross, 1982). This untranslated sequence is not encoded by DNA contiguous to the structural gene, and therefore has been termed the spliced leader (SL). Sequences encoding the SL are repeated 100–200 times in the genome of Trypanosoma brucei, and each resides in a 1.4 kb unit monomer (DeLange et al., 1983; Nelson et al., 1983). The vast majority of these units are directly and tandemly repeated to form a large array(s). However, a few 1.4 kb units, with their resident SL sequences, are dispersed from the tandem array (Nelson et al., 1983; Parsons et al., 1984b) and are termed orphons according to the nomenclature of Childs et al. (1981). Surprisingly, neither the large array nor the orphons are detectably linked (i.e., within 50 kb) to active VSG genes. Nevertheless, the tandem array of SL reiteration units has been proposed to mark the 5′ boundary of the VSG expression site (DeLange et al., 1983). Immediately 5′ to the SL in the 1.4 kb repeat unit are sequences resembling eucaryotic RNA polymerase II promoters (DeLange et al., 1983); according to the multiple promoter hypothesis above, these function in the initiation of VSG gene transcription. Only that VSG gene placed downstream from the array through DNA rearrangement would be transcribed, thus only a single VSG gene at a time would be expressed.

Certain clues, however, indicate that the function of the SL repeat is more complex. First, the SL is transcribed not only by the mammalian bloodstream stage of the parasite, which expresses VSG and undergoes antigenic variation, but also by procyclic culture forms (analogous to the insect midgut stage of the trypanosome life cycle), which do not express VSG (DeLange et al., 1983; Parsons et al., 1984b). Second, sequences homologous to the SL are highly reiterated in the genomes of certain trypanosomatids that do not undergo antigenic variation, such as Trypanosoma cruzi, the intracellular parasite that causes Chagas disease, and Leptomonas collosoma, a parasite of insects (Nelson et al., 1984). These findings suggest that although the SL is used by African trypanosomes for VSG expression, it is an ancient sequence that may fulfill other more fundamental functions in gene expression in these organisms. This would predict the presence of the SL on other trypanosome transcripts as well as on VSG mRNA. To test this hypothesis we have screened cDNA libraries made from T. brucei bloodstream or procyclic RNA for clones hybrid-

izing to a synthetic probe complementary to 22 nucleotides of the 35 nucleotide SL. Recombinant clones which did not encode VSG, but which contained the SL, were detected in both libraries. The expression of the corresponding genes is regulated differently from VSG genes; they are transcribed during both bloodstream and procyclic stages. As with VSG genes however, the SL is not encoded contiguously to these various structural genes, and is apparently derived from a transcript originating elsewhere in the genome. Our data suggest that rather than providing for transcriptional exclusivity in VSG gene expression, the SL plays a universal role in trypanosome gene expression.

Results

Isolation of cDNA Clones Containing the SL

Only bloodstream-stage trypanosomes synthesize VSG mRNA (Agabian et al., 1980; Overath et al., 1983; Parsons et al., 1984b), and then, of the thousand-odd VSG genes, each variant antigen type (VAT) transcribes only the VSG gene that is ultimately expressed as its surface coat (Hoeijmakers et al., 1980b; Pays et al., 1980). Thus, in Figure 1A, when total RNA is fractionated by agarose gel electrophoresis, transferred to nitrocellulose membranes, and hybridized to a ^{32}P-labeled probe (Northern analysis), a prominent species is detected by a cDNA encoding VSG 3 in RNA isolated from bloodstream-stage cells of VAT 3 (lane B). This RNA, of about 1.9 kb, and a putative precursor of about 4 kb (seen only upon overexposure of the autoradiogram), are not observed in procyclic RNA (lane P) nor in bloodstream-stage RNA isolated from other VATs (not shown). As previously reported (Parsons et al., 1984b), when the same RNAs are hybridized to a probe complementary to 22 nucleotides of the 35 nucleotide SL (see Figure 2) a very different pattern is obtained (Figure 1B). Here, a smear of RNAs ranging in size from 0.5 to over 6 kb is detected in both procyclic (lane P) and bloodstream stage (lane B). In particular, a 1.9 kb species is seen in VAT 3 RNA (lane B). Ribosomal RNAs (their position indicated by the hash marks) do not hybridize to the synthetic probe. The SL-containing RNAs are not simply transcripts of the genomic 1.4 kb repeat unit in which the SL resides, as rehybridization of the same blot to a clone containing the 1.4 kb genomic repeat reveals no hybridizing RNA (not shown). Since many of the RNAs revealed by the SL probe were smaller than the mature VSG mRNA, it seemed unlikely that they represented VSG mRNA processing intermediates. We therefore hypothesized that the smear of RNAs hybridizing to the synthetic probe represents transcripts derived from other genes that also use the SL or an SL-like sequence.

As a first step in testing this hypothesis, we used the SL probe to screen two cDNA libraries, one made from RNA isolated from VAT 5 and one made from RNA isolated from procyclic cells derived from VAT 5. In the procyclic library of 1000 clones, 24 clones were found that hybridized with the synthetic probe. Of approximately 600 recombinant clones in the VAT 5 library, 25 VSG-encoding cDNAs were detected, but only one of these contained 5′-terminal sequences as defined by hybridization to the SL probe. Seventeen other clones also hybridized with the SL probe. By analogy with VSG transcripts, the SL should be located at the 5′ terminus of transcripts giving rise to these clones. If only 5% of our cDNAs contain 5′-terminal sequences (as is the case for the VSG clones), then the numbers cited above indicate that approximately half of the clones in the library are derived from transcripts containing the SL. Thus, as predicted from the Northern analyses, SL-containing transcripts are abundant in stages of the life cycle when VSG is (bloodstream) or is not (procyclic) expressed. Most of the SL-containing cDNAs were represented only once in the libraries, implying that the complexity of such transcripts is high.

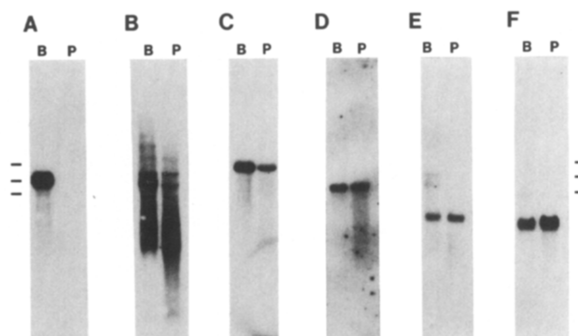

Figure 1. Northern Analyses

Total bloodstream VAT 3 RNA (B) and procyclic RNA (P) were fractionated on adjacent lanes (in quadruplicate) of a single gel. Four blots were prepared and hybridized to the probes listed below. When blots were re-used, we removed the previous probe with the appropriate denaturation conditions, verified by autoradiography. The probes used were as follows: (A) VSG 3 cDNA; (B) 22-mer SL; (C) pSLc1 cDNA; (D) pSLc2 cDNA; (E) pSLc4 cDNA; (F) pSLC3 cDNA.

Nucleotide Sequence Analysis of SL cDNAs

To determine whether the hybridizing sequences contained in these cDNAs were identical with the SL found on VSG mRNAs (Figure 2) or were divergent SL-like sequences, and to search for possible protein-coding sequences, the hybridizing regions of four cDNAs (two from the bloodstream-stage library, and two from the procyclic library) were cloned into M13 bacteriophage and their nucleotide sequences were determined. As shown in Figure 2, each clone contained a complete or partial SL. In the case of pSLc1 and pSLc3, the SL was truncated; the cDNA clone contained 23 and 26 bp, respectively, of the 35 nucleotide SL sequence. We interpret these to be incomplete copies of their corresponding transcripts. Since the synthetic probe is complementary to the 3′ portion of the SL, and since the cDNA and probe sequences match perfectly, these clones were easily detected by the SL probe. Clone pSLc4 contained a complete SL and, 5′ to that, 3 nucleotides that correspond to those immediately 5′ to the SL in the genomic repeat unit. These data suggest that sequences 5′ to the SL may be transcribed, as has

```
                           Xmn 1
                              ↓
SL          AACGCTATTA TTAGAACAGT TTCTGTACTA TATTG

22-mer               T AATCTTGTCA AAGACATGAT A

pSLc1                  AGAACAGT TTCTGTACTA TATTGTATCA AATAATAAGA GAATTAACTT

            TGTAGATAAA GAAAGCAATA AAGCATCA ATG AGC GGA AAG GAA GTT GGA GGT
                                           met ser gly lys glu val gly gly

pSLc2      *AACGCTATTA TTAGAACAGT TTCTGTACTA TATTGGTTCG CTTTAACTTG CCAGTACGCT

            TGTGAAGCGG TT

pSLc3               CA TTAGAACAGT TTCTGTACTA TATTGACTAC CTTCTCGTTA GTGTAACAAG

            TCCTCTGCAG TG ATG ATG TTC GGT CGC CCC GCC GTC CCC CAG GCA ACC TGG
                          met met phe gly arg pro ala val pro gln ala thr trp

            GAA GAG AAG TAT TTT TAT CAA AAA CTT CAC CAT CTT TTC GAC CAT GCT
            glu glu lys tyr phe tyr gln lys leu his his leu phe asp his ala

            GCT GAT TGG TTC GTA ACG AAG GTT AAC TGG TGG ATG CCG TCT ATC GGT
            ala asp trp phe val thr lys val asn trp trp met pro ser ile gly

            GCC GGG ATG GTG CTC AGT CTC T
            ala gly met val leu ser leu

pSLc4   ACT AACGCTATTA TTAGAACAGT TTCTGTACTA TATTGTGCCA CTAGCGAAGG GGGCGAAGGA

            GACCGAAGAG GAGAGGGGTT AATAATTTGT GTAACTATTA CCTGTA ATG TTG CGT CTC
                                                               met leu arg leu

            TGC CGT GTG TCA CTG CGT GTC CAG TCA CAC CAG AAG AAG CGC GCA CAG CAC
            cys arg val ser leu arg val gln ser his gln lys lys arg ala gln his

            CCC AAC GCC GGC ACA CGG TTT GGA CGT GTG TAC AAT CGC GGT TTC ATT CGG
            pro asn ala gly thr arg phe gly arg val tyr asn arg gly phe ile arg

            TAC GGC TTC GGT GGT TTC GGC AT
            tyr gly phe gly gly phe gly
```

Figure 2. Nucleotide Sequence Analysis

The sequence of the 35 nucleotide SL and the sequence of the 22 nucleotide probe complementary to the SL are shown at the top. Below are shown partial sequences of several cDNAs that hybridize to the SL probe (the SL portion is underlined). Putative reading frames are translated into amino acid sequences. The asterisk (*) indicates that another 92 bp of sequence was found 5′ to the SL (see text).

been proposed by Boothroyd and Cross (1982). The fourth clone, pSLc2, contained a complete and perfect SL. An additional 93 bp (indicated by an asterisk in Figure 2) are found 5′ to the SL in this clone. Subsequent hybridization analysis suggested that this 93 bp sequence is derived from sequences within the structural gene, implying that its appearance 5′ to the SL resulted from an artifact of cDNA cloning. Michiels et al. (1983) have described a VSG cDNA containing sequences 5′ to the SL.

The start codons of VSG mRNAs are located some 30 to 80 nucleotides downstream from the SL. We further analyzed these clones for the presence of a start codon and the existence of a possible reading frame. Our results indicate that, like the VSG leaders, these leaders are composite, containing the 5′ SL followed by additional noncoding sequences. For example, the first ATG triplet is encountered 53 bp downstream from the SL in pSLc1. In pSLc3, a start codon is found 37 bp downstream from the SL, followed by a reading frame extending at least 159 nucleotides. pSLc4 contains a reading frame that starts 72 nucleotides downstream from the SL, extending at least 135 nucleotides. This region encodes a 45 amino acid polypeptide rich in basic amino acids. Thus there is no reason to suspect that these cDNAs are not derived from bona fide mRNAs since they contain protein-coding reading frames.

Expression of Genes that Use the SL

The one well-studied class of genes that employs the SL, VSG genes, shows stage-specific expression. It was therefore interesting to determine whether stage-specific expression was a general feature of genes that use the SL. The transcripts that correspond to several of the SL-containing cDNAs were examined by Northern analysis of RNA isolated from the two easily obtained stages of the trypanosome life cycle: bloodstream and procyclic stages. The RNA blots shown in Figures 1A and 1B were rehybridized with SL cDNA probes corresponding to single-copy genes (see below). In each case one major RNA species was detected, ranging in size from 1.3 to 2.3 kb, depending on the cDNA used as probe (Figure 1, C–F). The two clones from the procyclic library, pSLc3 (Figure 1F) and pSLc4 (Figure 1E), also detect additional higher molecular weight RNAs in both bloodstream and procyclic samples. For example, pSLc4 hybridizes primarily to a 1.4 kb RNA, but also detects 1.75 and 2.0 kb RNAs. (These are more clearly visible on longer autoradiographic exposures.) Unlike VSG genes, all of these genes are tran-

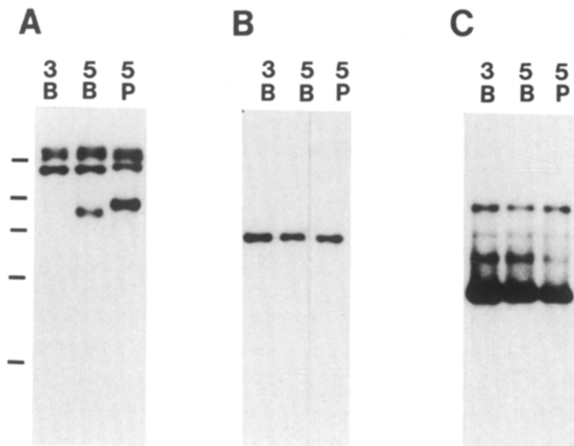

Figure 3. Genomic Southerns

DNA isolated from trypanosomes of VAT 3 (3B), VAT 5 (5B), or procyclic cells derived from VAT 5 (5P) was cleaved with Eco RV. Two micrograms of each sample was electrophoresed in triplicate on an agarose gel and transferred to nitrocellulose paper and hybridized with nick-translated SL cDNAs. The hash marks on the left indicate the position of DNA markers at 23 kb, 9.6 kb, 6.6 kb, 4.3 kb, and 2.3 kb. Probes were: A, VSG 5 DNA; B, pSLc2; C, pSLc5.

scribed during both bloodstream and procyclic phases of the trypanosome life cycle (compare lanes P and B). One of the cDNAs isolated from the bloodstream-stage library, pSLc1, detects a transcript that appears more abundant in the bloodstream than in procyclic stage (Figure 1C), while pSLc3, isolated from the procyclic library, detects a transcript that is more abundant in procyclic RNA (Figure 1F). Since the SL cDNA probes had specific activities similar to that of the VSG cDNA probe, but required autoradiographic exposures some five to ten times longer, it seems that the abundance of each of these transcripts is quite low compared to that of VSG gene transcripts (Figure 1A).

Genomic Organization of Genes That Use the SL

Genomic Southern analysis using SL cDNA probes reveals several classes of genes whose transcripts contain the SL. These experiments were performed under conditions too stringent for stable hybridization of the 35 nucleotide SL with its corresponding genomic sequences. Figure 3 shows Eco RV-cleaved genomic DNA from bloodstream-stage VATs 3 (3B) and 5 (5B) and from the procyclic population obtained by in vitro differentiation of VAT 5 (5P). When hybridized with a VSG 5 cDNA probe, the hallmark properties of many VSG genes are visible (Figure 3A): the presence of a multigene family, an expression-linked copy (the extra gene copy in VAT 5 as compared to VAT 3), and the retention of the expression-linked copy by procyclic cells (Parsons et al., 1984b). In addition, the expression-linked copy, as well as one other member of the VSG 5 gene family, resides on restriction fragments that show variations in size. Such variation is characteristic of VSG genes that are located adjacent to a telomere and is correlated with cell growth or division (Bernards et al., 1983).

In contrast, in this and many other restriction digests, pSLc2 hybridizes to a single genomic sequence (Figure 3B), indicating that pSLc2 detects a single-copy gene. These data also suggest that there are no large introns within the portion of the gene detected by the cDNA probe (an intron between the SL and the structural gene would not be detected, since the SL sequence does not hybridize to its corresponding genomic sequences under these conditions). Unlike VSG genes, this gene shows no variation between VATs or life-cycle stages. Of six SL-containing cDNA clones used as probes in Southern analyses, four appeared to detect single-copy, nonvarying genes. The remaining two cDNAs detected nonvarying multigene families such as that shown in Figure 3C. Since all trypanosome telomeric sequences thus far studied show variation in length over time, it appears that, unlike many VSG genes, these other genes that employ the SL do not reside adjacent to a telomere. Again, unlike active VSG genes, the other genes that employ the SL are not situated in DNA devoid of restriction sites (see also Figure 5). None of the SL cDNAs detected sequences that comigrate with any genomic 1.4 kb SL repeat units, suggesting they are not closely linked to the tandem array or the major SL orphons (not shown).

Cloning of Genes Encoding SL cDNAs

Although no evidence of close linkage between genomic SL sequences and the genes detected by SL cDNAs was observed in Southern analyses, it might be difficult to detect a single SL dispersed from the tandem array under the conditions employed. We therefore cloned genomic sequences corresponding to four of the SL-containing cDNAs, previously determined to detect single-copy genes. VAT 5 DNA was partially digested with Bam HI or Mbo I, fragments ranging from 12 to 20 kb were cloned into bacteriophage λ 1059, and recombinants that hybridized with the SL cDNA probes were isolated. None of these clones hybridized with more than one cDNA, indicating that the genes examined are not closely linked. None of the genomic clones hybridized with the SL, or with the genomic 1.4 kb SL repeat unit even under relaxed conditions (3× SSC, 50°C). For example, in Figure 4, cDNA clone pSLc1 (lane C), and genomic DNA clone λ(SL)g1 (lane G) were digested with Cla I and hybridized either to a 5′ 250 bp subclone of pSLc1 (Figure 4A) or to the SL probe (Figure 4B). Both the cDNA and the genomic clones hybridize with the 5′ 250 bp fragment, but only the cDNA hybridizes with the SL.

The absence of the SL in these recombinants could be the result of cloning only the 3′ portion of the gene. To rule out this possibility, representative genomic clones of each of the four genes were analyzed by restriction enzyme mapping. In each case, clones were found that contained several (3–12) kilobases of DNA upstream from the structural gene (Figure 5). The genomic clone discussed above and shown in Figure 4 contained 8 kb 5′ to the structural gene. The restriction sites present in each cDNA, up to but not including the Xmn I site contained in

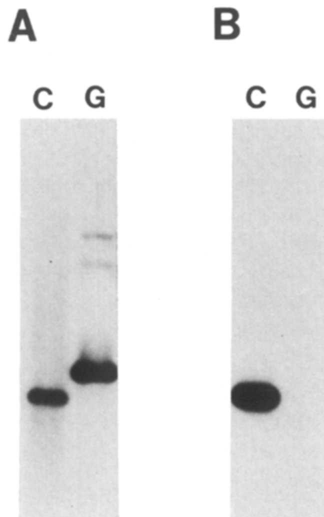

Figure 4. Hybridization of the SL Probe to Genomic and cDNA Clones

Genomic clones detected by pSLc1 were isolated. Genomic clone λ(SL)g1-3 (lane G) and cDNA clone pSLc1 (lane C) were cleaved with Cla I. Cla I has recognition sites within both of the vectors, as well as within the cloned structural gene (see Figure 5A). (A) Hybridization with a 250 bp 5′ fragment of pSLc1. (B) Hybridization with the SL probe.

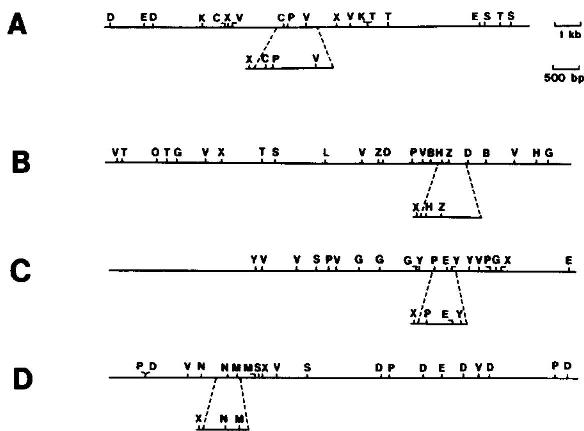

Figure 5. The Genomic Environment of Selected Genes That Use the SL

In each case, the top line shows the restriction map derived from overlapping genomic clones, while the lower line shows the map of the corresponding cDNAs. In some cases only the restriction sites closest to the gene are shown. Note that the Xmn I site (X) present in the SL was not observed in the corresponding position in the genomic clones. Restriction enzyme sites are marked as follows: Bgl I, Z; Bgl II, G; Bam HI, B; Cla I, C; Eco RI, E; Hinc II, H; Hind III, D; Kpn I, K; Mlu I, M; Nco I, N; Pst I, P; Pvu II, V; Sal I, L; Sma I, S; Sph I, Y; Sst I, T; Xho I, O; Xmn I, X.

(A) pSLc1 and its genomic environment; (B) pSLc2 and its genomic environment; (C) pSLc3 and its genomic environment; (D) pSLc4 and its genomic environment.

the SL (see Figures 2 and 5), were also present in the homologous genomic clones. For example, the Hinc II site, located 37 bp from the 3′ end of the SL in pSLc2, and the Pst I site, located 28 bp from the SL in pSLc3, are each found in the corresponding λ(SL)g2 and λ(SL)g3 genomic clones. In both cases, these sites are in the 5′ untranslated regions of the molecule. Thus the 3′ portion of the leader sequence is derived from sequences abutting the structural gene, while the 5′ portion containing the SL is transcribed from elsewhere in the genome.

Discussion

The experiments described here demonstrate that the 35 nucleotide SL is not unique to VSG mRNAs but is also present on many other trypanosome RNAs. Estimates based on the frequency of SL-containing clones in cDNA libraries suggest that a substantial fraction of cytoplasmic mRNAs may carry this 5′ sequence. Sequences common to many different cellular mRNAs have been described in other eucaryotic systems, but they are not 5′-terminal. For example, a 3′-terminal "suffix" sequence is found on approximately 2% of all Drosophila mRNAs (Tchurikov et al., 1982). Similarly the 3′ "Set 1" sequence is highly represented on mouse mRNAs transcribed during the first half of embryogenesis (Murphy et al., 1983). Another sequence is found on the intron regions of many RNA species in the rat—the "ID" sequences (Milner et al., 1984). This sequence appears to be a marker for brain-specific transcripts (Sutcliffe et al., 1984).

Each of these elements are repeated in their respective genomes, but unlike the tandemly repeated SL sequences, they are dispersed. They are encoded adjacent to or within the structural genes on whose transcripts they are found, again contrasting with the trypanosome SL, which is not detectably linked to the structural genes. The differences between the SL and other repetitive elements found on multiple transcripts, in particular regarding their locations in the genome and on the transcripts, indicate that different molecular mechanisms are responsible for their ubiquity.

Perhaps more similar to the trypanosome SL is the case of the coronavirus leader. Coronavirus RNAs are actually a series of nested transcripts that are 3′ coterminal and extend various distances 5′. Each RNA, however, shares the same 5′-terminal leader sequence (Spaan et al., 1983).

What mechanism accounts for the presence of the SL on trypanosome RNAs? Figure 6 depicts several alternative arrangements of SL and structural gene exons, and their resulting primary transcripts. Sequences just upstream from the SL resemble concensus sequences of eucaryotic RNA polymerase II promoters. The large tandem array of SL repeat units has therefore been hypothesized to provide multiple promoters for frequent initiation of VSG gene transcription (DeLange et al., 1983). However, the regulation of the transcription of the other genes whose transcripts contain the SL differs from that of VSG genes; those studied here are transcribed during both bloodstream and procyclic stages, and appear to be much less abundant than VSG mRNAs. Could all of these genes, with such different characteristics, be using the same promoter (see Figure 6A)? Recall that as with VSG genes, genomic sequences encoding the SL were not found within 3–12 kb of the structural genes encoding the SL-containing RNAs. Nor do any of these genes appear to be linked to one another. Thus if there is a single promoter used by all

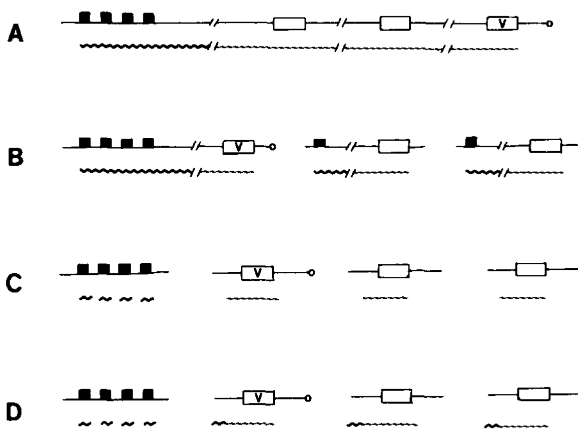

Figure 6. Models of SL and Structural Gene Transcription

In each case, the upper line depicts the genomic configuration and the lower line shows the primary transcription products. The large boxes indicate structural genes, with that labeled "V" being the active VSG gene. As indicated by the line terminating in a circle, the active VSG gene lies adjacent to a telomere. The small shaded boxes are the genomic SL sequences, either in the tandem array or in the orphons.

(A) One continuous transcription unit for all genes that use the SL. (B) Separate transcription units for each structural gene that uses the SL. (C) Discontinuous transcription, subsequent RNA joining. (D) Discontinuous transcription, SL RNA used as primer.

these genes, the primary transcription unit must be hundreds of kilobases long, provide for stage-specific and abundant transcription of the distal VSG gene (transcribed VSG genes are adjacent to a telomere), and moderate transcription of many other genes throughout the trypanosome life cycle.

Another possibility is that each of these genes has its own SL promoter, located far upstream (Figure 6B). Although each of the SL repeat units appear very similar, we have detected a low level of sequence heterogeneity between repeat units (unpublished results), which might be hypothesized to allow for differences in the regulation of transcription. For example, the developmentally regulated VSG gene could be colinear with (albeit distant from) the major array of repeat units and the three to four SL orphons of our T. brucei stock could serve as promoters for other SL genes. However, there are many more than four non-VSG genes that employ the SL sequence. Even if there were several unidentified orphons or several tandem arrays, there do not appear to be sufficient unlinked SL transcription units for each gene to have its own. A combination of the two hypotheses, with several large transcription units, one colinear with developmentally regulated genes, and one colinear with constitutively expressed genes, could also be proposed.

We believe it more likely that the SL repeat units are not linked to the structural genes, and that transcription is discontinuous. Two different hypotheses could be proposed. First, that the SL and structural gene sequences could be transcribed from separate DNA molecules, with the two RNA molecules then joined by intermolecular splicing or ligation (Figure 6C). Alternatively, the SL transcript could aid in the initiation of structural gene transcription (Figure 6D) as host RNA fragments serve as primers for transcription of influenza virus genes (Plotch et al., 1981). Interestingly, this reaction does not require homology between primer and virus sequences. The coronavirus leader sequence has also been postulated to function as a primer in discontinuous transcription (Spaan et al., 1983). With regard to these latter hypotheses, one would predict the presence of small SL-containing RNAs. We have recently characterized an RNA species of approximately 135 bp that hybridizes to the SL probe (Milhausen et al., submitted).

It is clear that the role of the SL is not limited to VSG gene expression in African trypanosomes—it is found in other trypanosomatids that do not undergo antigenic variation (Nelson et al., 1984), and on RNAs derived from genes whose pattern of expression differs dramatically from that of VSG genes. Although as yet no structural genes, aside from those encoding VSGs or tubulins, have been studied in trypanosomes, identification of the function of the genes employing the SL may serve to clarify its role in trypanosomatid gene expression. Alternatively, the SL may be present on virtually all trypanosome mRNAs, resulting from a requisite role in the transcription or RNA-processing machinery of the cell. The possibility of a unique process critical to gene expression in these pathogenic organisms may provide a target for rational chemotherapy of trypanosomiasis.

Experimental Procedures

Trypanosomes

The IsTaR serodeme of Trypanosoma brucei brucei was employed for all studies (Stuart et al., 1984). Cells of VAT 5 were converted to procyclic forms by in vitro culture (Hanas et al., 1975). DNA for Southern analyses and RNA for construction of cDNA libraries were isolated from VAT 5 cells and from procyclic cells cultivated for at least 2 months (Parsons et al., 1983; Milhausen et al., 1983). Total RNA isolated from VAT 3 cells or procyclic cells derived from VAT 5 was a gift of Dr. Jean Feagin.

cDNA and Genomic Clones

The construction of cDNA libraries in pBr322 has been described (Parsons et al., 1983). The generation of a library of genomic DNA (from VAT 1.5 cells) in bacteriophage λ 1059 is described by Aline et al. (submitted). SL-containing cDNAs were detected by hybridization with ^{32}P-labeled synthetic probe complementary to 22 nucleotides of the 35 nucleotide SL (see Figure 2) using hybridization conditions previously determined to detect specific SL sequences in the genome (see below).

Those cDNAs used for studies reported here were pTbSLc1-1(B), pTbSLc2-1(B), pTbSLc3-1(P), pTbSLc4-1(P), and pTbSLc5-1(P) and for convenience are designated in the text as pSLc1 through pSLc5, respectively. pSLc1 and pSLc2 were isolated from the bloodstream-stage cDNA library while pSLc3, pSLc4, and pSLc5 were isolated from the procyclic-stage library. Genomic clones corresponding to pSLc1 are designated λ(SL)g1-n (where n is the clone number); those corresponding to pSLc2 are designated λ(SL)g2-n, etc. Thirty-four λ(SL)g1 clones, four λ(SL)g2 clones, ten λ(SL)g3 clones, and ten λ(SL)g4 clones were isolated.

The portions of selected cDNAs that hybridized to the synthetic probe were subcloned into M13 vectors mp 8, mp 9, mp 18, or mp 19. The sequence was determined by the dideoxy chain termination method (Sanger et al., 1977).

Hybridization Analyses

Total RNA was fractionated by electrophoresis in 1.4% agarose-formaldehyde gels and transferred to nitrocellulose membranes (Milhausen et al., 1983). Hybridizations were in 5× SSPE (1× SSPE is 180 mM NaCl, 10 mM

NaH_2PO_4, 1 mM EDTA, pH 7.4), 0.2% sarkosyl, and 200 μg/ml denatured salmon testis DNA for 16–30 hr at 65°C (for cDNA probes) or 37°C (for the 22-mer probe). The final stringency of post-hybridization washes was 0.3× SSC at 65°C for cDNA probes and 5× SSC at 37°C for the SL probe (1× SSC is 150 mM NaCl, 15 mM NaCitrate, pH 7.0). Southern analyses were performed as previously described (Parsons et al., 1983). The final stringencies for post-hybridization washes were the same as those described above, except for VSG cDNA hybridization, which employed a 0.1× SSC, 65°C final wash. cDNAs were labeled with ^{32}P by nick translation to a specific activity of approximately 10^8 cpm/μg. The 22-mer, a gift of Dr. Phillip Barr (Chiron Corporation), was labeled at its 5′ end to a specific activity of approximately 5×10^6 cpm per pmole as previously described (Nelson et al., 1983).

Acknowledgments

The authors would like to thank Dr. Kenneth Stuart for providing the VAT 5 genomic library, Susan Sather for helpful discussions, and Cyndy Gartside for secretarial assistance. This work was supported in part by grants from the Rockefeller Great Neglected Disease Program and from the National Institutes of Health (AI 17309). N. A. is a Burroughs-Wellcome Scholar in Molecular Parasitology.

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 May 9, 1984

References

Agabian, N., Thomashow, L., Milhausen, M., and Stuart, K. (1980). Structural analysis of variant and invariant genes in trypanosomes. Am. J. Trop. Med. Hyg. *29* (5) Suppl. 1043–1049.

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 duplication-transposition leading to an altered 3′ end. Cell *27*, 497–505.

Bernards, A., Michels, P., Lincke, C., and Borst, P. (1983). Growth of chromosome ends in multiplying trypanosomes. Nature *303*, 592–597.

Boothroyd, J., and Cross, G. (1982). Transcripts coding for variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5′ end. Gene *20*, 281–289.

Buck, G., Longacre, S., Raibaud, A., Hibner, U., Giroud, C., Baltz, T., Baltz, D., and Eisen, H. (1984). Stability of expression-linked surface antigen gene in *Trypanosoma equiperdum*. Nature *307*, 563–566.

Childs, G., Maxson, R., Cohn, R. H., and Kedes, L. (1981). Orphons: dispersed genetic elements derived from tandem repetitive genes of eucaryotes. Cell *23*, 651–663.

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

DeLange, T. Liu, A. Y. C., Van der Ploeg, L. H. T., Borst, P., Trompe, M. C., and van Boom, J. H. (1983). Tandem repetition of the 5′ mini-exon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription? Cell *34*, 891–900.

Hanas, J., Linden, G., and Stuart, K. (1975). Mitochondrial and cytoplasmic ribosomes and their activity in blood and culture forms of *Trypanosoma brucei*. J. Cell Biol. *65*, 103–111.

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.

Longacre, S., Hibner, U., Raibaud, A., Eisen, H., Baltz, T., Giroud, C., and Baltz, T. (1983). DNA rearrangements and antigenic variation in *Trypanosoma equiperdum:* multiple expression-linked sites in independent isolates of trypanosomes expressing the same antigen. Mol. Cell. Biol. *3*, 399–409.

Michiels, F., Matthyssens, G., Kronenburger, P., Pays, E., Dero, B., Van Assel, S., Darville, M., Cravador, A., Steinert, M., and Hamers, R. (1983). Gene activation and re-expression of a *Trypanosoma brucei* variant surface glycoprotein. EMBO J. *2*, 1185–1192.

Milhausen, M., Nelson, R., Parsons, M., Newport, G., Stuart, K., and Agabian, N. (1983). Molecular characterization of initial variants from the IsTat 1 serodeme of *Trypanosoma brucei*. Mol. Biochem. Parasitol. *9*, 241–254.

Milner, R., Bloom, F., Lai, C., Lerner, R., and Sutcliffe, G. (1984). Bran-specific genes have identifier sequences in their introns. Proc. Nat. Acad. Sci. USA *81*, 713–717.

Murphy, D., Brickell, P. M., Latchman, D. S., Willison, K., and Rigby, P. W. J. (1983). Transcripts regulated during normal embryonic development and oncogenic transformation share a repetitive element. Cell *35*, 865–871.

Nelson, R. G., Parsons, M., Barr, P. J., Stuart, K., Selkirk, M., and Agabian, N. (1983). Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphons in Trypanosoma brucei. Cell *34*, 901–909.

Nelson, R., Parsons, M., Selkirk, M., Barr, P., and Agabian, N. (1984). Sequences homologous to variant antigen mRNA spliced leader in Trypanosomatidae which do not undergo antigenic variation. Nature *308*, 665–667.

Overath, P., Czichos, J., Stock, U., and Nonnengasser, C. (1983). Repression of glycoprotein synthesis and release of surface coat during transformation of *Trypanosoma brucei*. EMBO J. *2*, 1721–1729.

Parsons, M., Nelson, R. G., Newport, G., Milhausen, M., Stuart, K., and Agabian, N. (1983). Genomic organization of *Trypanosoma brucei* variant antigen gene families in sequential parasitemias. Mol. Biochem. Parasitol. *9*, 255–269.

Parsons, M., Nelson, R., and Agabian, N. (1984a). Antigenic variation in african trypanosomes: DNA rearrangements program immune evasion. Immunol. Today *5*, 43–50.

Parsons, M., Nelson, R., Stuart, K., and Agabian, N. (1984b). Variant antigen genes of *Trypanosoma brucei:* genomic alteration of a spliced leader orphon and retention of expression linked copies during differentiation. Proc. Nat. Acad. Sci. USA *81*, 684–688.

Pays, E., Delronche, M., Lheureux, M., Vervoort, T., Bloch, J., Gannon, F., and Steinert, M. (1980). Cloning and characterization of DNA sequences complementary to messenger ribonucleic acids coding for the synthesis of two surface antigens of *Trypanosoma brucei*. Nucl. Acids Res. *8*, 5965–5981.

Pays, E., Van Meirvenne, N., LeRay, 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.

Plotch, S. J., Bouloy, M., Ulmanen, I., and Krug, R. M. (1981). A unique cap (m^7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell *23*, 847–858.

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

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B., and Siddell, S. (1983). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. EMBO J. *2*, 1839–1844.

Stuart, K., Gobright, E., Jenni, L., Milhausen, M., Thomashow, L., and Agabian, N. (1984). The IsTat serodeme of *T. brucei:* development of a new *T. b. brucei* serodeme and characterization of early variants. Parasitology, in press.

Sutcliffe, J., Milner, R., Gottesfeld, J., and Lerner, R. (1984). Identifier sequences are transcribed specifically in brain. Nature *308*, 237–241.

Tchurikov, N., Naumova, A., Zelentsova, E., and Georgiev, G. (1982). A cloned unique gene of *Drosophila melanogaster* containing a repetitive 3′

exon whose sequence is present at the 3′ ends of many different mRNAs. Cell *28*, 365–373.

Van der Ploeg, L. H. T., Valerio, D., DeLange, T., Bernards, A., Borst, P., and Grosveld, F. G. (1982a). An analysis of the 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*, 5903–5923.

Van der Ploeg, L., Bernards, A., Rijsewijk, F., 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., Liu, A., Michels, P., DeLange, T., Borst, P., Majumder, H., Weber, H., Veeneman, G., 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.

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

Young, J., Donelson, J., Majiwa, P., Shapiro, S., and Williams, R. (1982). Analysis of genomic rearrangements associated with two variable antigen genes of *Trypanosoma brucei*. Nucl. Acids Res. *10*, 803–819.