Volume 14 Number 4 1986 Nucleic Acids Research

The trypanosome spliced leader small RNA gene family: stage-specific modification of one of several similar dispersed genes

Marilyn Parsons, Richard G.Nelson and Nina Agabian+

Department of Biochemistry, University of Washington, Seattle, WA 98195 and +University of California, Naval Biosciences Laboratory, 140 Warren Hall, Berkeley, CA 94720, USA

Received 30 September 1985; Accepted 22 January 1986

ABSTRACT

Diverse mRNAs of Trypanosoma brucei possess the same 5' terminal 35 nucleotides, termed the spliced leader (SL), which appears to be derived from a separate 135 nucleotide transcript. This small SL RNA is encoded within a 1.4 kb unit of DNA which is tandemly reiterated in the genome. In addition, there are at least 4 orphon elements containing SL sequences dispersed from the tandem array. Here we show that during the trypanosome life cycle one of the SL orphons undergoes a stage-specific modification that prevents cleavage of an EcoRV site and we further demonstrate that although only one orphon is modified, three of the SL orphons are flanked by very similar sequences. Each of these contains SL reiteration units including the non-transcribed spacer DNA, suggesting that they did not originate through an RNA intermediate. In addition no evidence of direct repeats at the junction of 1.4 kb and non-1.4 kb DNA was observed. Finally, a phylogenetic survey indicates that while many trypanosomatid species possess similarly organized SL-like sequences, only the SL orphons of closely related subspecies of the T. brucei - T. evansi complex share similar flanking regions.

INTRODUCTION

Eukaryotic cells possess a wide variety of reiterated gene families which show modes of genomic organization ranging from clustered to highly dispersed. Many of the clustered gene families exist in tandem arrays, with the transcribed region plus spacer DNA forming a monomer unit which is repeatedly linked head to tail. Frequently, gene families which show tandem linkage also possess some members which are dispersed (1). First described for the histone gene family of the sea urchin (1), these dispersed genetic elements are termed orphons. Orphons have since been found in gene families as diverse as the 5S and tRNA genes of Xenopus laevis (2), a major tandem repeat family from the human X chromosome (with dispersed members on autosomes) (3), and genes encoding the rRNAs of yeast (1). Both RNA and DNA mediated mechanisms of gene dispersal have been proposed to account for the

origin of these elements (4).

Among gene families organized in tandem arrays are those which encode the first 35 nucleotides of mRNAs of trypanosomatid organisms (5,6). Many, if not all, mRNAs of the protozoan parasite Trypanosoma brucei possess the same 5' terminal 35 nucleotide leader (7,8). This untranslated sequence, termed the spliced leader (SL) (or mini-exon) was originally observed on mRNAs encoding those molecules mediating antigenic variation in African trypanosomes, the variant surface glycoproteins (VSGs) (9,10). Using a synthetic probe complementary to 22 nucleotides of the SL, we demonstrated that the SL resides on a genomic 1.4 kb monomer unit (6) composed of a 135 bp transcribed region (including the SL) plus spacer DNA (11). The 135 nucleotide SL RNA transcript contains the SL at its 5' end and is hypothesized to be the donor of the SLs found on mature trypanosome mRNAs (11-13). The vast majority of the 100 to 200 1.4 kb units are directly and tandemly repeated in the genome, but a few are orphons (6). The SL sequences, either in the tandem array or in the orphons, do not appear to be linked to those genes on whose mRNAs the SL is found (6,8).

In general, the copy number and genomic location of orphon elements is highly variable (1). We have also observed that different T. brucei stocks show unique orphon patterns, indicating a high level of genetic polymorphism (6). In addition, we have previously observed different orphon patterns in Eco RV cleaved DNA isolated from two different stages of the trypanosome life cycle, the procyclic (insect midgut) stage and the mammalian bloodstream stage (6,14). These two stages differ in a number of other features, including cellular morphology (15), variant surface glycoprotein expression (16), and mitochondrial gene expression (15). Occasionally we have noted similar differences in the Eco RV orphon patterns of various bloodstream stage variant antigen type (VAT) clones descended from the same progenitor cell (6).

As described in this report, we have cloned and analyzed in detail one of the SL orphons from the IsTaR serodeme of T. brucei (17). Comparison of this orphon with the other IsTaR SL orphons indicates that, unlike orphons of other gene families, three of the four SL orphons are embedded in similar regions of genomic DNA. At least one, and probably all three of these orphons contain multiple tandem copies of the SL reiteration unit. The stage specific orphon alteration appears to result from a DNA modification in a palindromic region immediately 3' to the last 1.4 kb unit of one of the

orphons. Despite the fact that three orphons share very similar flanking regions, only one undergoes this modification.

MATERIALS AND METHODS

Antigenically homogeneous bloodstream stage trypanosomes of various variant antigen types (VATs) of the IsTaR serodeme of Trypanosoma brucei brucei were grown and isolated as described (17,18). Procyclic stage trypanosomes were propagated in vitro (14). DNA was isolated as detailed by Milhausen et al. (18), and Southern analysis was performed as described (19). Final post-hybridization washes for hybridizations employing nick-translated cloned cDNA and genomic DNA probes were 0.3 X SSC at 65°C (1X SSC is 150 mM NaCl, 15 mM Na_3 citrate). Conditions for specific hybridization with the ^{32}P-end-labelled 22 nucleotide oligomer complementary to a portion of the SL have been described (6); final post-hybridization washes were in 5X SSC at 37°C.

The largest Eco RV SL orphon (11 kb) was cloned from Eco RV cleaved DNA isolated from bloodstream stage VAT 5 cells. Eco RV fragments of 9-15 kb were excised from a gel and the presence of only the large orphon was verified by Southern analysis of the excised fragments. After addition of Eco R1 linkers, the fragments were cloned into the Eco R1 site of bacteriophage lambda L47.1 (20). Two recombinant clones containing sequences which hybridized with the 22-mer SL probe were plaque purified and DNA was prepared according to standard techniques. Subclones were prepared by ligating isolated restriction fragments into pUC 8 (21) and M13 vectors (22). DNA sequencing was performed by the dideoxy chain termination method (23). The molecular cloning of this Eco RV fragment has allowed us to more accurately estimate its size as 11 kb (it was previously estimated to be 12 kb [6, 14]).

RESULTS

Eco RV does not cleave within the 1.4 kb SL monomer, and thus leaves the large SL array intact. Our previous work demonstrated that the SL orphon pattern revealed by Eco RV digestion differed in the insect procyclic and the mammalian bloodstream stages of the life cycle (6). The SL probe detected four fragments (at 11, 7, 5.7 and 3.5 kb) in bloodstream stage DNA from cells of VAT A, but only the three smaller fragments were detected in DNA isolated from procyclic cells derived by in vitro differentiation of VAT A

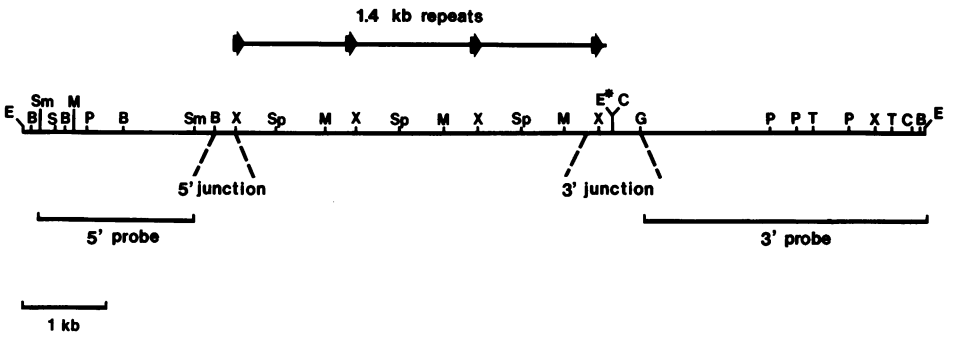


Figure 1. Map of the 11 kb Eco RV SL orphon element. The location of the 1.4 kb units, shown above the restriction map, was determined by restriction mapping and hybridization with the SL probe. The orientation of these units is based on the orientation of the sequence homologous to the small SL RNA, as determined by hybridization analyses and DNA sequencing. The locations of the SLs are denoted by heavy arrows. The regions denoted 5' junction and 3' junction were subcloned and their nucleotide sequences determined (see Figure 2). Also shown are the flanking regions which were subcloned and used as probes in Figure 3. The 5' probe was the indicated Sma I subclone while the 3' probe was the indicated Bgl II - Eco RI subclone. Restriction enzyme sites are noted. In particular, E* denotes an Eco RV site which was not seen in genomic DNA, but which was found by restriction mapping and nucleotide sequencing of the cloned DNA. Other sites include: B, Bam HI, C, Cla I; E, Eco RV; G, Bgl II; M, Mlu I, P, Pst I; S, Sal I; Sm, Sma I, Sp, Sph I, T, Sst I, X, Xmn I. The SL itself contains an Xmn I site.

cells. This finding was extended to five other pairs of bloodstream stage cells and their procyclic counterparts: in each case the procyclic DNA lacked the largest orphon (14). In order to investigate this phenomenon, and to further understand the nature of orphon elements, we have cloned the largest Eco RV orphon from bloodstream stage VAT 5 DNA.

Analysis of a Genomic Clone Containing the 11 kb Eco RV SL Orphon

Eco RV fragments of 9-15 kb were cloned into the bacteriophage lambda vector L47.1 and two recombinant clones were isolated which hybridized with the SL probe. Preliminary experiments indicated that the two cloned Eco RV fragments were identical, therefore further analysis concentrated on one clone. As shown in Figure 1, several points of interest emerged during our analysis of the cloned orphon. First, there are multiple complete 1.4 kb units within the orphon. These are tandemly linked, as they are in the large array. The orphon 1.4 kb repeat commences within an SL (at nucleotide 7 or 8) and extends 3' just over three repeat units, terminating some 20 bp

Figure 2. DNA sequence of the junctions between 1.4 kb repeat DNA and non-repeat DNA in the 11 kb Eco RV orphon. The 1.4 kb DNA sequence is italicized, with the SL underlined. Due to the tandem organization of 1.4 kb units within the orphon 49 bp of the 1.4 kb sequence is directly repeated at the junctions with non-repeat DNA. Asterisks are placed above nucleotides which differ from those in the canonical 1.4 kb repeat unit. The cryptic Eco RV site is overlined in Figure 2B. Figure 2A. The 5' junction. Figure 2B. The 3' junction.

downstream from the fourth SL sequence, within the region encoding the small SL RNA (See Figure 2). Limited sequence analysis of internal 1.4 kb regions upstream from the SL sequences indicate that they retain a high level of homology to the reiteration unit of the tandem array, showing less than 2% mismatch in the areas examined. Most of the differences were 1 or 2 bp insertions or deletions, which particularly occured in non-transcribed homopolymer tracts. Similar levels of mismatch occur between 1.4 kb elements isolated from a given cloned cell line or between different T. brucei stocks (11,13,24,25).

Sequence analysis of the junctions of 1.4 kb repeat and non-repeat DNA (noted in Figure 1) reveals no evidence of a directly repeated sequence flanking the terminal 1.4 kb units (Figure 2). Such direct repeats are

characteristically generated when mobile genetic elements are inserted into DNA, and are derived by duplication of a short sequence at the target site (26). Since the 1.4 kb unit is tandemly reiterated within the orphon, a portion of 1.4 kb sequence (as opposed to flanking DNA sequence) is necessarily directly repeated at the junctions.

At the 5' junction (as oriented from the small RNA transcription unit), the truncated SL sequence abuts a poly A tract. There is no open reading frame 5' to this area within the region examined, nor is the poly A tract preceded by the putative trypanosome polyadenylation signal AAAATTPyT (27). The poly A tract is however directly preceded by an inverted repeat of 9 bp (indicated by arrows in Figure 2A). The hypothesized polyadenylation signal is also missing from trypanosome α-tubulin and β-tubulin genes (28). DNA flanking the 3' junction is dA/dT rich and shows a high strand bias, with dA and dG predominating on one strand, and dT and dC on the other. This strand bias is also characteristic of the intergenic region separating the α-tubulin and β-tubulin genes in the tubulin tandem array (28).

Surprisingly, within the region flanking the last 1.4 kb element, both DNA sequencing (Figure 2B) and restriction enzyme analysis (not shown) demonstrated an internal Eco RV site. The Eco RV site is located 86 bp downstream from the last SL in the orphon, and 66 bp from the junction of the reiteration unit and the flanking region. The Eco RV recognition site also forms a part of several overlapping short palindromes and interrupted palindromes, one of which contains a Cla I and a Taq I site. Preliminary results indicate that this Cla I site is also refractory to cleavage.

As predicted from the DNA sequence, cleavage of the cloned orphon with Eco RV releases a 7 kb fragment which hybridizes with the SL probe (not shown). Interestingly, this fragment is the same size as the second largest orphon detected in bloodstream stage DNA and corresponds in size to the largest orphon in procyclic DNA. In genomic Southern blots hybridized with the SL probe the signal produced by the 7 kb orphon is reproducibly more intense in procyclic DNA (see below, and Figure 3A) suggesting that procyclic cells contain two copies of the 7 kb orphon. These data are compatible with the hypothesis that the bloodstream stage also contains two 7 kb Eco RV orphons, but that the 3' Eco RV site is modified in one of these such that it is not cleaved by Eco RV digestion. If this hypothesis is correct, regions flanking the 5' side of the 1.4 kb reiteration units should be shared by the 7 and 11 kb Eco RV fragments, but those 3' to the Eco RV site should be found only on the 11 kb Eco RV fragment. To test this, we subcloned segments of

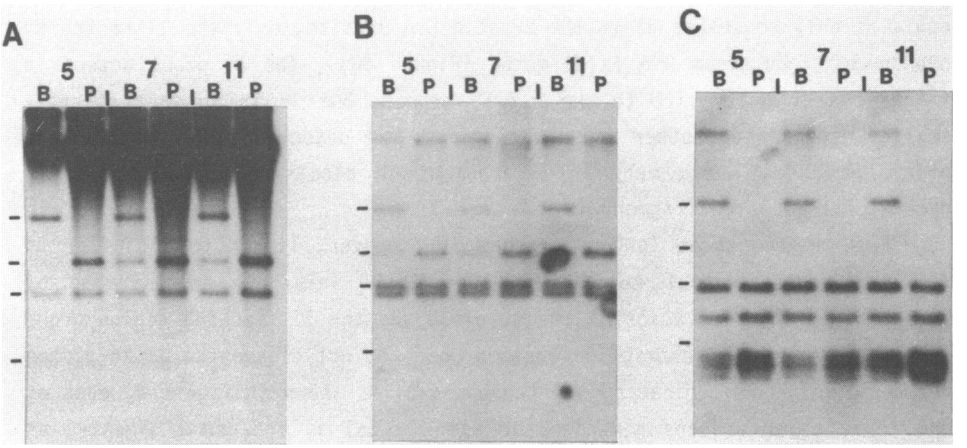


Figure 3. SL orphons share common 5' flanking sequences. Genomic DNA was isolated from bloodstream stage (B) IsTaR VATs 5, 7, and 11, and procyclic stage (P) cells derived from the same VATs. The Eco RV cleaved DNA was separated by electrophoresis and transferred to a nitrocellulose filter. The same blot was serially hybridized with the 22-mer probe complementary to the SL (Figure 3A), the 5' flanking region probe (Figure 3B), and the 3' flanking region probe (Figure 3C). These probes are described in Figure 1. Lack of signal from the previous hybridization was verified by autoradiography. Hash marks show the positions of the 11 kb, 7 kb, 5.7 kb, and 3.5 kb orphons.

DNA from the regions flanking the 1.4 kb units of the cloned orphon for use as probes in Southern hybridization analyses.

SL Orphons Share 5' and 3' Flanking Sequences.

DNA isolated from several different IsTaR VATs and the procyclic cells derived from them, was cleaved with Eco RV. Figure 3A shows the hybridization pattern obtained with the 22-mer SL probe, revealing the four orphons in bloodstream stage DNA and only three in procyclic DNA (the 3.5 kb orphon is only faintly visible in this exposure -- longer exposures render it more clearly visible but obscure the 11 kb orphon). As shown in Figure 3B, the 5' flanking region probe hybridized with both the 11 kb orphon in bloodstream stage DNA, and the 7 kb orphon in bloodstream and procyclic DNA. It also hybridized with the 5.7 kb orphon, indicating that it is related to the larger orphons. In addition, the probe detected a 5.5 kb fragment which does not contain SL sequences and a very large fragment (>25 kb) which comigrates with the SL tandem array. The 3.5 kb fragment which hybridizes weakly with the SL probe does not hybridize with this 5' probe.

When the blot was rehybridized with the 3' flanking region probe, which

contains only sequences 3' to the cryptic Eco RV site, only the 11 kb Eco RV orphon in bloodstream DNA is detected (Figure 3C). The 3' probe detects a fragment similar in size to the 5.7 kb orphon, but it is in fact slightly smaller. At least 5 other Eco RV fragments were detected in bloodstream and procyclic DNA, plus a very large fragment in bloodstream DNA only. The nature of this large fragment is unknown.

Each of the three largest orphons is, however, linked to a similar 3' flanking region distal to the Eco RV site. This was demonstrated by comparing the hybridization of the SL probe and the 3' flanking region probe to DNA cleaved with restriction enzymes which do not cleave the 11 kb orphon between regions recognized by the two probes. As shown in Figure 4, each of the three major orphons generated in Bam HI, Sal I, and Sma I digests, as detected by the SL probe, also hybridized to the 3' flanking probe. Subsequent Eco RV cleavage of these restriction digests removed the 3' flanking regions from each of the orphons but one - the 11 kb orphon (not shown). Thus, the three orphons share similar 3' flanking regions which each possess an Eco RV site, but the Eco RV site adjacent to one of the dispersed 1.4 kb reiteration units is refractory to cleavage.

To further compare the three largest orphons, we used the 5' and 3' flanking region probes, as well as the SL probe, in additional genomic Southern analyses (not shown). In every case, the data were consistent with the proposal that there are two highly homologous orphons which contain three 1.4 kb units, but that in one of these the Eco RV site is refractory to cleavage, while in the other it is not. Furthermore, there was always a fragment 1.4 kb smaller than the larger orphons, which hybridized to the 5' and 3' probes, as well as the 22-mer SL probe. These data suggest that the 5.7 kb Eco RV orphon has not three, but two 1.4 kb repeat units flanked by DNA highly homologous to that in the 11 kb and 7 kb orphons. Proof of this hypothesis awaits molecular cloning of each orphon. Furthermore, each of these three orphons also contains a Cla I site which separates the 1.4 kb units from their 3' flanking region. Double digests using the enzymes Bam HI and Cla I, and Bam HI and Eco RV, indicate that in each orphon the Cla I and Eco RV sites are in close proximity. The fourth Eco RV fragment, which hybridized weakly with the SL but not the flanking region probes, appears to be an unrelated element. In fact, it does not hybridize with a 1.3 kb clone (6) containing most of the reiteration unit except the SL (not shown) and thus represents a very different element indeed.

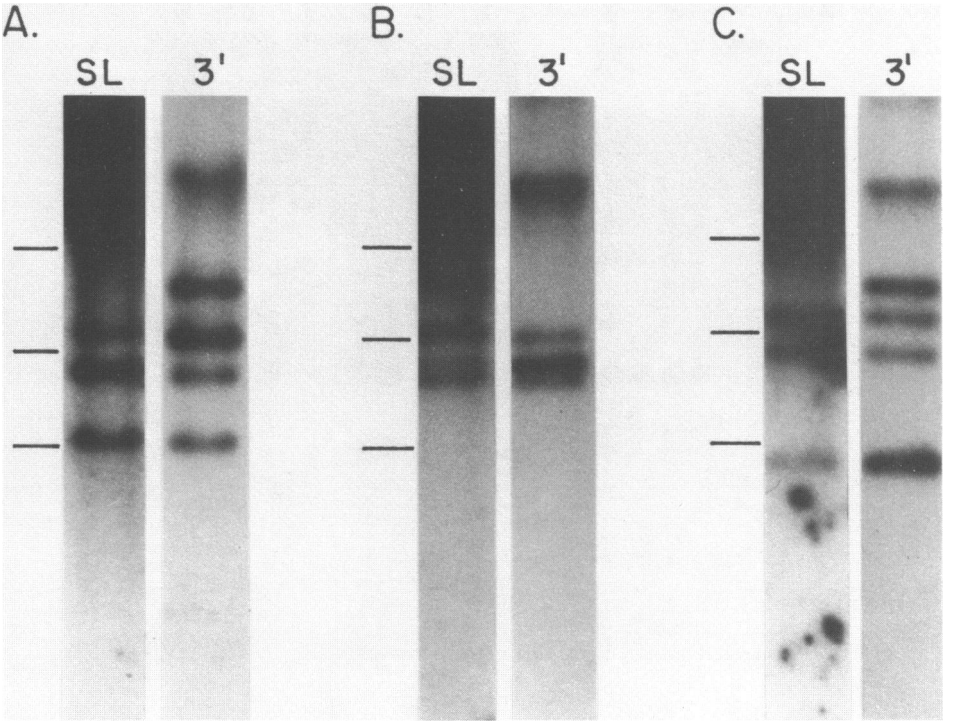
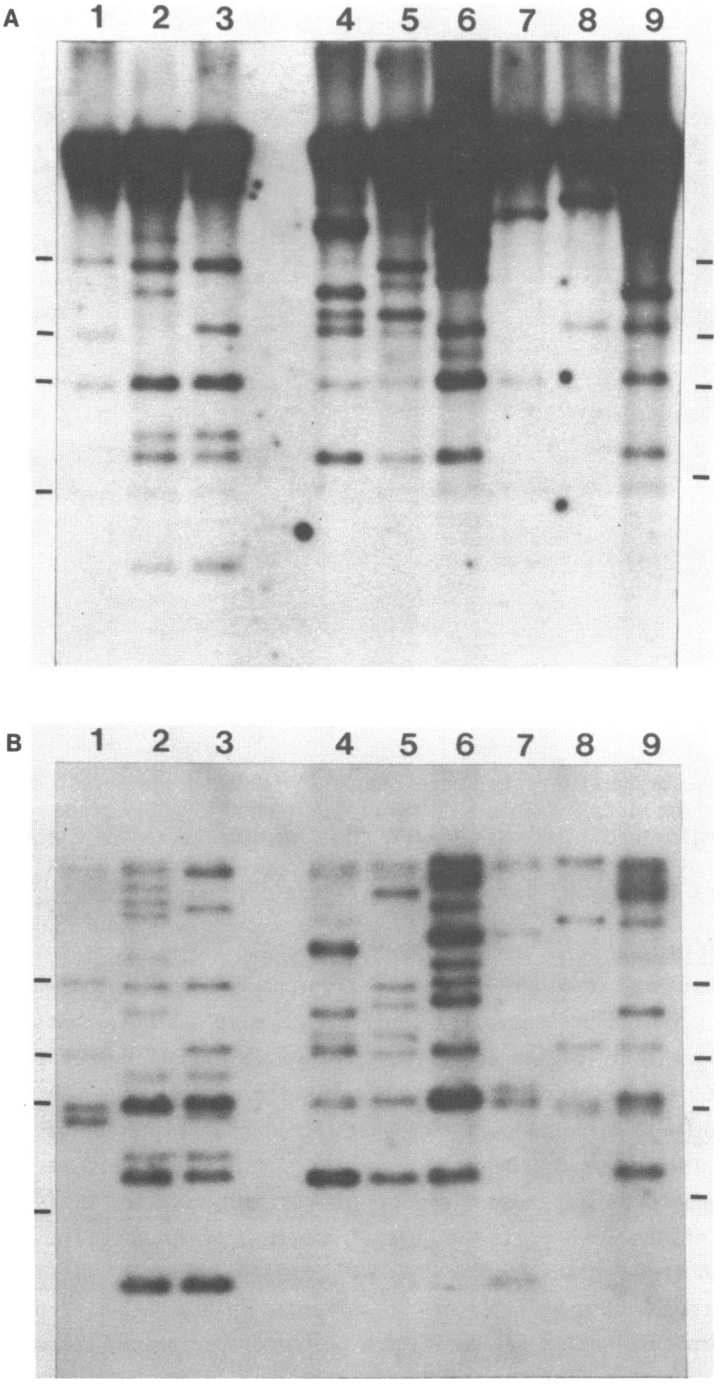


Figure 4. SL orphon elements share common 3' flanking sequences. Bloodstream stage DNA from IsTaR VAT 5 was hybridized with the SL probe or the 3' flanking region probe (see Figure 1). The hash marks indicate the position of DNA size markers of 23 kb, 9.6 kb, and 6.6 kb. Figure 4A. Sma I digested DNA. Figure 4B. Bam HI digested DNA. Figure 4C. Sal I digested DNA.

Orphon Flanking Sequences are Species Specific and Polymorphic.

The fact that the orphon flanking regions were homologous in our T. brucei stock led us to search for their presence in other trypanosome stocks and species. Hybridization analyses performed at low stringency (final wash 1X SSC at 50°C) demonstrated that the related African trypanosomes T. vivax and T. congolense do not possess sequences homologous to either of the orphon flanking regions, nor do the more divergent species T. cruzi and Leptomonas collosoma (not shown). These elements are restricted to the closely related subspecies of the T. brucei - T. evansi complex, where they are detectable even at high stringencies (0.3X SSC, 65°C). Within stocks of the T. brucei - T. evansi complex, Eco RV fragments which hybridize to the SL (Figure 5A) and to the 5' orphon flanking region (Figure 5B) often comigrate, suggesting that these too are orphons and that their flanking regions are homologous to those


A
1 2 3 4 5 6 7 8 9
B
1 2 3 4 5 6 7 8 9

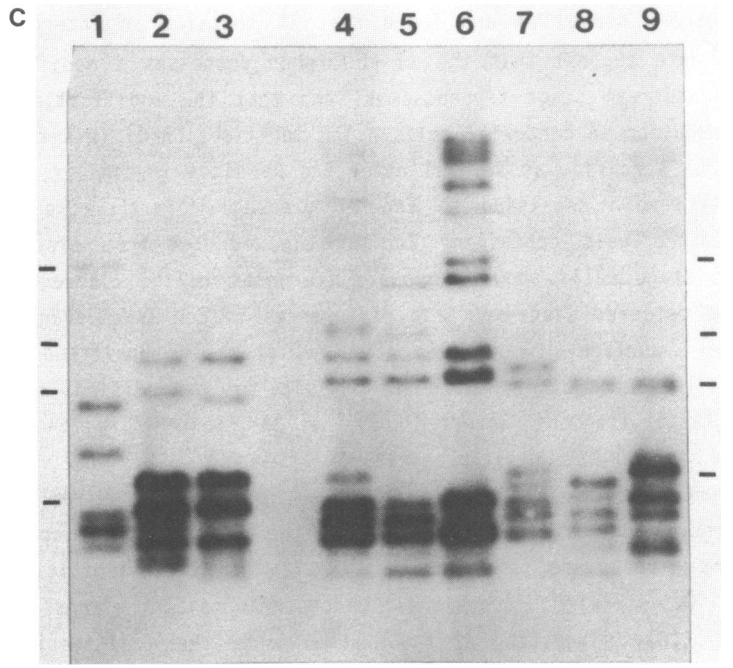


Figure 5. T. brucei subspecies retain orphon flanking regions. Eco RV digested genomic DNA from various T. brucei subspecies was transferred to a nitrocellulose filter and serially hybridized with the various orphon probes. Figure 5A. The 22-mer SL probe. Figure 5B. The 5' flanking region probe. Figure 5C. The 3' flanking region probe. (See Figure 1 for description of the probes). Lane 1-3. T. b. brucei, stocks: 164, IsTaR VAT 11 (lane 1), 1026 (lane 2), and 427 (lane 3). Lane 4-6. T.b. gambiense, stocks: U1 (lane 4), L2 (lane 5), and 1375 (lane 6). Lane 7-9. T.b. rhodesiense, stocks: 1799 (lane 7), 2002 (lane 8), and WraTaT 3 (lane 9). The hash marks indicate the positions of the IsTaR serodeme orphons at 11 kb, 7 kb, 5.7 kb, and 3.5 kb.

in the IsTaR stock. Sequences homologous to the 3' flanking region were also observed in each of these stocks (Figure 5C). In Eco RV digests, most of these hybridizing fragments did not comigrate with the SL orphons, suggesting that these orphons may also possess an Eco RV site between the dispersed 1.4 kb units and their 3' flanking regions. The level of polymorphism in the number and complexity of sequences detected by the 5' and 3' orphon flanking region probes appears to parallel that detected by the SL probe itself.

DISCUSSION

We have cloned an 11 kb Eco RV orphon containing reiterated SL sequences

from bloodstream stage DNA and found that it contains an internal Eco RV site. The data suggest that the 11 kb orphon possesses a modified Eco RV site in bloodstream stage trypanosomes, and that the modification is lost during propagation in bacterial cells. The modification of this Eco RV site appears to be specific, as several other Eco RV sites in similar regions of genomic DNA remain sensitive to Eco RV cleavage. Furthermore, ethidium bromide stained gels containing Eco RV cleaved DNA from procyclic and bloodstream stage cells show no gross differences in the cleavage pattern, nor have we detected stage-specific differences in Eco RV cleavage patterns using variant surface glycoprotein cDNAs as probes (unpublished results). However, at least one other related Eco RV site may be modified, as shown by the very large fragment detected by the 3' flanking region probe in bloodstream stage DNA, but not procyclic DNA (See Fig. 3C). Thus, it appears that in bloodstream stage cells a specific DNA modification system is active, while in procyclic forms it is not. Whether the modification we observed has any functional relevance is not known. While the stage-specificity of its expression is striking, occasional bloodstream stage clones lack the modification (6), suggesting that in some cells the modification system becomes inactive or that the inheritance of the modification pattern is not strict.

The 3' flanking region probe should detect an additional fragment derived from the unmodified 11 kb orphon in Eco RV cleaved procyclic as compared to bloodstream DNA. Although an additional fragment unique to procyclic DNA was not observed, there was an apparent increase in the hybridization intensity of one or more of a triplet of fragments of approximately appropriate size. If the 3' flanking regions of the 5.7, 7 and 11 kb orphons are identical with respect to Eco RV sites, the expected increase in the hybridization intensity would only be one and a half fold.

The nature of the proposed DNA modification is not known. Certainly, the palindromic nature of the sequences surrounding the cryptic Eco RV site affords possibilities for enzymatic recognition. In fact, it has recently been shown that *in vitro* modification of Taq I sites (methylation of dA) renders adjacent Eco RV sites refractory to cleavage (29). A Taq I site occurs within the Cla I site overlapping the Eco RV site in the 11 kb orphon. Each of the 3 major orphons appear to possess a similarly placed Cla I (and thus Taq I) site, although whether the exact position relative to the Eco RV site is maintained has not been determined.

The *in vivo* modification of chromosomal DNA sequences in African

trypanosomes has been previously reported (30 - 33). For example, in T. equiperdum an unidentified nucleoside appears to partially replace deoxycytidine residues in the telomeric regions adjacent to telomere-linked VSG genes (32). In T. brucei transcriptionally silent, telomeric VSG genes from bloodstream stage trypanosomes are partially protected from cleavage by the restriction enzymes Pst I and Pvu II (30, 31) and Sph I and Hind III (31) suggesting that such sequences may possess DNA modifications. Only a dGpC dinucleotide is shared among these four restriction enzyme cleavage sites. Telomere linked Sal I sites, which do not contain dGpC, also have been reported to be partially modified (33). However, neither transcriptionally active telomeric VSG genes from bloodstream stage cells nor transcriptionally silent telomeric VSG genes from procyclic cells are refractory to cleavage by these enzymes (31). Thus, this modification system appears to be active in the bloodstream stage, and not the procyclic stage and further appears to be specific for telomeric sequences associated with transcriptionally inactive VSG genes. Although modification of the Eco RV site of the 11 kb SL orphon manifests the same life cycle stage specificity, the orphon does not appear to be telomere-linked nor does the modification appear to be any less than complete. Furthermore neither the Eco RV, nor the overlapping Cla I and Taq I restriction enzyme cleavage sites contain a dGpC dinucleotide. Finally, examination of telomeric VSG genes containing EcoRV sites does not reveal any modifications similar to that observed in the SL orphon. Thus, the modification we describe here appears to be different from the telomeric modifications described by others.

The most well-studied orphons, aside from those presented here, are those of the histone gene family of the sea urchin (1), those of a gene cluster which encodes six tRNAs in X. laevis (2), and those of the U small nuclear (sn) RNA gene families of man (4,34,35,36). The sea urchin histone gene family is composed of reiteration units which contain coding regions for the five histone genes plus spacer DNA. The vast majority of the orphons, however, contain only one or two histone genes, along with some spacer DNA (1). The presence of nontranscribed spacer suggests that these orphons did not arise via an RNA intermediate. Sequence analysis of one orphon revealed no direct repeats at the junctions of the dispersed element with non-histone DNA (1). The tDNA gene family of X. laevis has given rise to an intriguing orphon family which contains some of the tRNA genes, but lacks others, and is itself tandemly repeated (2). Orphon pseudogenes of the U snRNA gene families were studied before the organization of the functional genes became

apparent. It is now clear that the human U1 gene family is highly clustered, if not tandemly linked on very large reiteration units (37,38). Most of the dispersed U1 snRNA pseudogenes appear to have arisen by an RNA mediated mechanism, as they contain only sequences homologous to the transcribed portion of the reiteration unit (4). Many of these are also flanked by direct repeats. In at least one case (34) a U snRNA has been shown capable of functioning *in vitro* as a self-priming template for AMV reverse transcriptase which generates 3' truncated cDNA copies very similar in structure to the observed pseudogenes. Aside from the U snRNA orphons, many of which are known to be non-functional, the transcriptional status of orphon elements is not known.

The orphons of the *T. brucei* SL family present yet another picture, with at least one, and probably three, containing multiple and faithful reiteration units. Three of the four orphons do not appear to have been generated by RNA mediated mechanisms, since they contain the non-transcribed region of the 1.4 kb unit. No direct repeats, such as those characteristically (26) but not necessarily (39) generated by transposable elements, were found in the flanking DNA at its junction with the SL reiteration units of the 11 kb orphon. The similarity of the orphon flanking regions suggests that rather than arising from independent gene dispersal events, a single dispersed cluster along with flanking sequences may have been amplified by gene duplication, transposition, or conversion-like events. Using 5' and 3' orphon flanking regions as probes we have not detected orphon transcripts in Northern analyses, although we have found evidence for DNAse 1 hypersensitive sites (unpublished results).

The phylogeny of the orphon flanking regions parallels that of the SL reiteration unit. Despite the fact that trypanosomatids as diverse as *L. collosoma*, a parasite of insects, and *T. brucei*, possess sequences homologous to the SL organized in tandem arrays and orphon elements (6), the reiteration unit in which the SL resides is species-specific (40). Thus, like the reiteration unit, the *T. brucei* orphon flanking regions hybridize with DNA from the closely related subspecies of the *T. brucei* - *T. evansi* complex, but not with DNA from other species of African trypanosome or other genera of trypanosomatid.

Given the fact that most, if not all, trypanosome mRNAs contain a 5' terminal SL, as does the small SL RNA, and the probable presence of a reverse transcription-cDNA insertion process in trypanosomes (27) it is perhaps surprising that we did not detect more dispersed SL sequences placed

in the genome by RNA-mediated mechanisms, as is common in other small RNA gene families. In fact, we detected only one dispersed SL which might have resulted from such a process, on the 3.5 kb Eco RV fragment which does not hybridize to the orphon flanking region subclones nor to a probe containing most of the reiteration unit. Whether this SL sequence represents an inserted cDNA copy of the small RNA, a cDNA copy of an mRNA, or a pseudo SL (such as that present 5' to the α-tubulin gene) remains to be elucidated. The relative absence of SL orphons generated by RNA mediated processes may result from an unusual secondary structure of the small SL RNA or it may be a consequence of a deleterious effect of such a dispersed SL.

ACKNOWLEDGMENTS

The authors thank Dr. Kenneth Stuart for helpful discussions and Drs. K. Stuart, J. Doyle, and R. Gray for the gifts of trypanosome stocks.

This work was supported by the Rockefeller Foundation's Great Neglected Diseases Program, NIH Grants AI17309 and AI21975. Nina Agabian is a Burroughs Wellcome Scholar of Molecular Parasitology.

REFERENCES

1. Childs, G., Maxson, R., Cohn, R. and Kedes, L. (1981) Cell 23, 653-663.
2. Rosenthal, D. and Doering, J. (1983). J. Biol. Chem. 259, 7402-7410.
3. Willard, H., Smith, K. and Sutherland, J. (1983) Nucl. Acids Res. 11, 2017-2033.
4. Denison, R. and Weiner, A. (1982) Molec. Cell. Biol. 2, 815-828.
5. DeLange, T., Liu, A., Van der Ploeg, L., Borst, P., Trompe, M. and Van Boom, J. (1983) Cell 34, 891-900.
6. Nelson, R.G., Parsons, M., Barr, P., Stuart, K., Selkirk, M. and Agabian, N. (1983) Cell 34, 901-909.
7. DeLange, T., Michels, P.A.M., Veerman, H.J.G., Cornelissen, A.W.C.A. and Borst, P. (1984) Nucl. Acids Res. 12, 3777-3790.
8. Parsons, M., Nelson, R.G., Watkins, K.P. and Agabian, N. (1984) Cell 38, 309-316.
9. Boothroyd, J.G. and Cross, G.A.M. (1982) Gene 20, 281-289.
10. Van der Ploeg, L., Liu, A. Michels, P., DeLange, T., Borst, P., Majumder, H., Weber, H., Veeneman, G. and Van Boom, J. (1982) Nucl. Acids Res. 10, 3591-3604.
11. Milhausen, M., Nelson, R.G., Sather, S., Selkirk, M. and Agabian, N. (1984) Cell 38:721-729.
12. Kooter, J., DeLange, T. and Borst, P. (1984) EMBO J. 3, 2387-2392.
13. Campbell, D., Thornton, D. and Boothroyd, J. (1984) Nature 311, 350-355.
14. Parsons, M., Nelson, R.G., Stuart, K. and Agabian, N. (1984) Proc. Natl. Acad. Sci. 81, 684-688.
15. Brown, R.C., Evans, D.A. and Vickerman, K. (1973) Int. J. Parasitol. 3, 691-704.

16. Barry, J.D. and Vickerman, K. (1979) Exptl. Parasitol. 48, 313-324.
17. Stuart, K., Gobright, E., Jenni, L., Milhausen, M., Thomashaw, L. and Agabian, N. (1984) Parasitology, 70, 747-754.
18. Milhausen, M., Nelson, R.G., Parsons, M., Newport, F., Stuart, K. and Agabian, N. (1983) Molec. Biochem. Parasitol. 9, 241-254.
19. Parsons, M., Nelson, R.G., Newport, G., Milhausen, M., Stuart, K. and Agabian, N. (1983) Molec. Biochem. Parasitol. 9, 225-269.
20. Loenen, W. and Brammer, W. (1980) Gene 10, 249-259.
21. Vieira, J. and Messing, J. (1982) Gene. 19, 259-268.
22. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
23. Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. 74, 5463-5467.
24. DeLange, T., Berkvens, T., Veerman, H.J.G., Carlos, A., Frasch, C., Barry, J.D. and Borst, P. (1984) Nucl. Acids Res. 12, 3777-3790.
25. Dorfman, D.M. and Donelson, J.E. (1984) Nucl. Acids Res. 12, 4907-4920.
26. Calos, M. and Miller, J. (1980) Cell 20, 579-595.
27. Hasan, G., Turner, M. and Cordingley, J. (1984) Cell 37, 333-341.
28. Sather, S. and Agabian, N. (1985). Proc. Natl. Acad. Sci. 82, 5695-5699.
29. Christ, C. and Nelson, M. (1984) BioTechniques 2, 218-222.
30. Bernards, A., van Harten-Loosbroek, N. and Borst, P. (1984) Nucleic Acids Research 12, 4153-4170.
31. Pays, E., Delauw, M.F., Laurent, M. and Steinert, M. (1984). Nucl. Acids Res. 12, 5235-5247.
32. Raibaud, A., Gaillard, C., Longacre, S., Hibner, U., Buck, G., Bernardi, G. and Eisen, H. (1983) Proc. Natl. Acad. Sci. USA. 80, 4306-4310.
33. Pays, E., Laurent, M., DeLauw, M.F. and Steinert, M. (1985) in UCLA Symposium on Molecular and Cellular Biology, Simon, M. and Hereskowitz, I. Eds., New Series Vol. 20, pp. 173-187, Alan R. Liss, Inc., New York.
34. Bernstein, L., Mount, S. and Weiner, A. (1983) Cell 32, 461-472.
35. Van Arsdell, S. and Weiner, A. (1984) Molec. Cell. Biol. 4, 492-499.
36. Denison, R., Van Arsdell, S., Bernstein, L. and Weiner, A. (1981) Proc. Natl. Acad. Sci. 78, 810-814.
37. Lund, E., Bostock, M., Robertson, S., Christie, J., Mitchen, J. and Dahlberg, J. (1983) Mol. Cell. Biol. 3, 2211-2220.
38. Manser, T. and Gesteland, R.F. (1982) Cell 29, 257-264.
39. Murphy, E. and Lofdahl, S. (1984) Nature. 307, 292-294.
40. Nelson, R.G., Parsons, M., Selkirk, M., Barr, P. and Agabian, N. (1984) Nature. 308, 665-667.