Volume 10 Number 10 1982 Nucleic Acids Research

Structure and expression of a *Trypanosoma brucei gambiense* variant specific antigen gene

E.Pays, M.Lheureux and M.Steinert

Département de Biologie Moléculaire, Université libre de Bruxelles, 67, rue des Chevaux, 1640-B Rhode St Genèse, Belgium

Received 23 February 1982; Revised 8 April 1982; Accepted 30 April 1982

SUMMARY

The expression-linked copy of the T. b. gambiense variant specific antigen gene LiTat 1.6 is transposed in a 20 kb DNA region devoid of restriction sites, located near a chromosome end. This expression site is very similar to that of T. b. brucei variants 117, 118 (1) and AnTat 1.8. In the basic copy, the transposable element (TE) is flanked by repetitive sequences; it includes the gene copy as well as a sequence of 0.9 to 2.1 kb (probably around 1.1 kb) long, upstream from the gene. Probes derived from the 5' part of the TE specifically reveal three polyadenylated transcripts of 4.2, 1.45 and 0.85 kb, respectively, distinct from the 2.1 kb mRNA. The amount of the 4.2 kb sequence is probably less than 0.01% of total trypanosome RNA. Whereas the mRNAs coding for the three isotypic antigens AnTat 1.8 (T. b. brucei), 12.2 (T. b. rhodesiense) and 3.3 (T. evansi) are recognized by LiTat 1.6 probes extending into the 3' half of the transposed sequence, the 5' genomic probes do not hybridize with any of these RNAs. These observations suggest that the LiTat 1.6 gene could be first transcribed in a large precursor molecule. This precursor would be rapidly processed, loosing a large portion of less conserved sequence from its 5' half. Our data are compatible with a model in which the promoter would be provided by the expression site.

INTRODUCTION

The genes coding for the variant specific antigens (VSAs) in African trypanosomes are subject to sequence rearrangements (2-6). In some cases the rearrangement seems not to be linked to expression and no gene amplification is involved (2,6). In other cases a "basic" copy (BC) of the gene is duplicated, and the additional or "expression-linked" copy (ELC) is transposed to a new site (3,4,5), where it is transcribed (7). In two different models which can account for this last observation, the transcription promoter is either present in the BC and would be activated only after transposition, or it is provided by the expression site. We have previously demonstrated (8) that, in the case of AnTat 1.1 and 1.8 from T. b. brucei, the transposable element (TE) is larger than the mRNA or the full-length cDNA. A similar situation has been described for another stock of T. b. brucei (1). These observations have been extended to the three AnTat 1.8 isotypes: LiTat 1.6, AnTat 12.2 and

0305-1048/82/1010-3149$ 2.00/0

AnTat 3.3, from T. b. gambiense, T. b. rhodesiense and T. evansi, respectively (9,10). It implies that in all these cases, an extra stretch of DNA is duplicated and transposed together with the messenger sequence. If the second model for gene activation is correct, one may thus expect to find a RNA precursor larger than the mature mRNA, which includes transcripts of the cotransposed sequence.

We present here the characterization of the BC, the ELC and the expression site of the LiTat 1.6 gene. We bring evidence that the TE extends about 1 kb upstream from the gene, up to a repetitive element in the BC. Probes corresponding to the 5' region of the TE do not recognize the mature mRNA but reveal transcripts both larger and smaller than this mRNA. It is inferred that the LiTat 1.6 gene could be transcribed in a large precursor which would be rapidly processed.

MATERIAL AND METHODS

Trypanosomes. Cloned populations of variant-specific antigen types (VATs) AnTat 1.1 and 1.8 from Trypanosoma brucei brucei, AnTat 12.1 and 12.2 from T. b. rhodesiense, AnTat 3.3 from T. evansi, LiTat 1.3 and 1.6 from T. b. gambiense have been obtained from Dr. N. Van Meirvenne (Institute for Tropical Medicine Prince Leopold, Antwerp). Their growth and harvesting have been described (4,10,11).

DNA isolation was carried out as desribed (8). Total RNAs were prepared by the LiCl/urea procedure (12,13). The poly(A)+ RNAs were isolated and purified by two passages on oligo(dT)-cellulose (14). The RNAs were size fractionated by gel electrophoresis under denaturing conditions (15) and transferred onto cellulose nitrate according to Thomas (16).

Cloning of the LiTat 1.6 specific sequences. Partial Sau3A digests of DNA purified from variant LiTat 1.6 were size fractionated on 0.6% low melting point agarose gels (13), and fragments between 15 and 20 kb were cloned in the BamHI sites of λ1059 as described in (17). Screening was done with a ^{32}P-labelled probe obtained by nick-translation (18) of the 1170 bp PstI fragment of the cloned AnTat 1.8 cDNA (13). Clones harboring characteristic internal fragments of the LiTat 1.6 BC (9) were selected. In addition, we cloned two LiTat 1.6 related sequences, but were unable to clone the LiTat 1.6 ELC.

Probes. Specific parts of the LiTat 1.6 cloned sequence were isolated by at least three rounds of preparative electrophoresis on low melting point agarose, as described earlier. Fragments were ^{32}P-labelled by nick-translation

(18) and the specificity of each probe was checked by back-hybridization with a mixture of all the fragments.

Hybridization of the probes with "Southern" or "Northern" blots was performed as described (8). The RNA/DNA hybrids were usually washed in 0.6 M NaCl, 0.06 M Na citrate, 0.1% Na dodecyl sulphate at 42°C, but in some instances more stringent conditions were applied (see text). DNA/DNA hybrids were always washed in 15 mM NaCl, 1.5 mM Na citrate at 65°C.

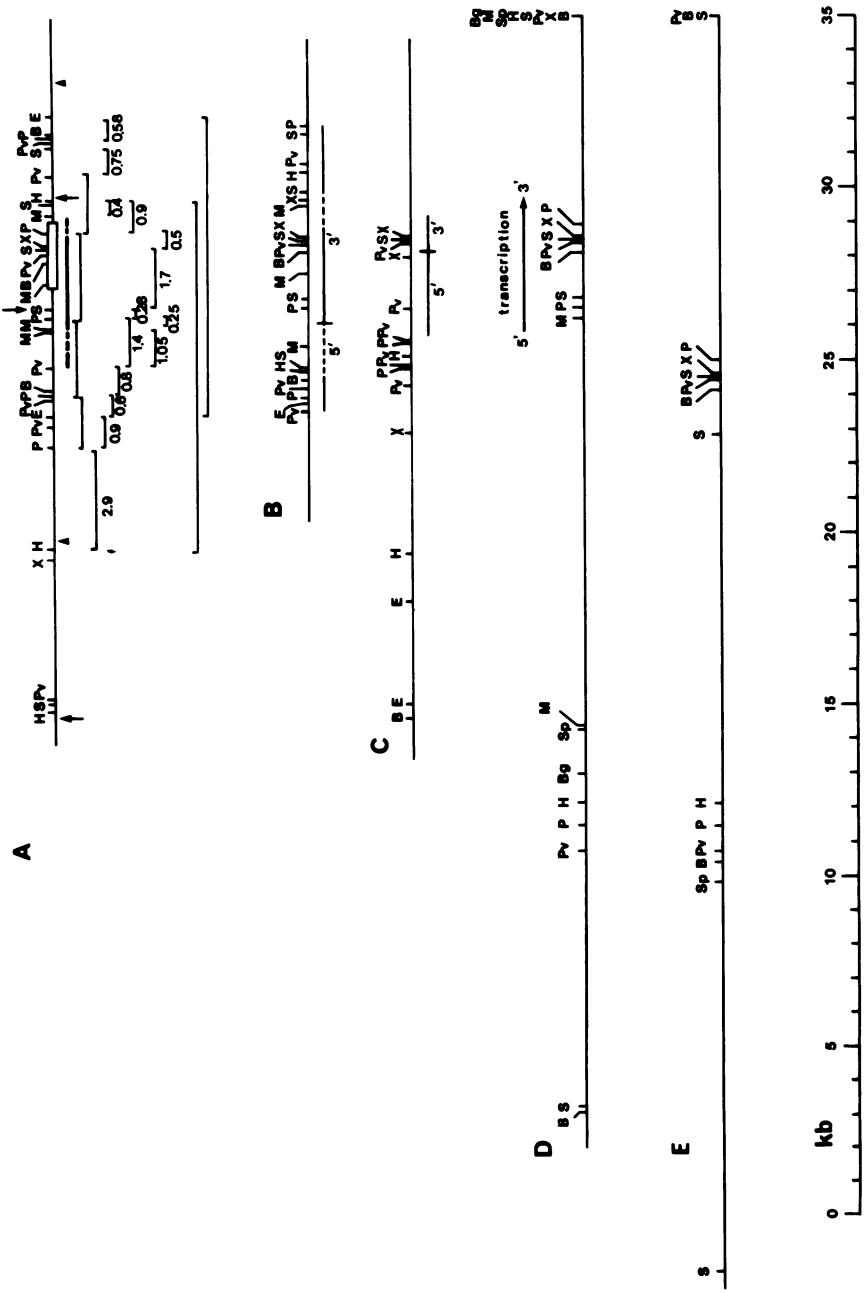
RESULTS

The BC of the LiTat 1.6 gene, as well as two related sequences, were cloned in the BamHI sites of λ1059 after partial Sau3A digestion and size-fractionation of the genome. The restriction map of the BC, included within a 19 kb genomic sequence, is presented in figure 1A. This sequence is the actual gene and not a related member of the LiTat 1.6 family: firstly it is the only one containing the 3' PstI site also found in the isotypic AnTat 1.8 cDNA (9,19); secondly, this PstI site defines the 3' limit of the 2.5 kb PstI fragment characteristic of both the BC and ELC, as demonstrated by the fact that, after hybridization, the labelling intensity of this band is doubled in genomes expressing the LiTat 1.6 gene (8,9). As mentioned previously (8,9), the expression of this gene is indeed linked to a duplication-transposition process. The doubling with gene expression of the labelling intensity of this band unequivocally implies that, besides the ELC, there is only one other LiTat 1.6 sequence, the BC, containing this 3' Pst I site.

Fragments of the gene and of its flanking sequences have been prepared from the cloned BC, as shown under map A of figure 1. These fragments have been labelled by nick-translation and used as probes to analyse the genomic DNAs of a LiTat 1.6 clone and of another cloned variant, LiTat 1.3, from the same stock. The genomic DNAs have been digested by BglI, then hybridized with different probes (fig. 2).

BglI cuts the LiTat 1.6 TE once, so that the BC and ELC of the LiTat 1.6 gene are each contained in two fragments: either 25 kb (5') and 6.9 kb (3') for the ELC, or 3.6 kb (5') and 3.75 kb (3') for the BC. Other hybridizing fragments are from related sequences that we cannot discuss at this stage of the work.

The additional copy starts to be clearly seen in 5' with the 1.4 kb PvuII-PstI probe, and only disappears with the 0.4 kb MspI-SalI probe in 3' (fig. 2). This analysis has been extended to other digestions, either cutting outside the TE (EcoRI, HindIII), or cutting once (PvuII, XorII) or twice into


A
B
C
D
E
transcription
5'
3'
kb
0
5
10
15
20
25
30
35

the TE (SalI, PstI). The data (not shown) enabled us to roughly determine the limits of the TE as summarized in figure 1. Its maximal possible length (4.2 kb) exceeds by 2.1 kb the size of the mRNA (ref.9 and see below). The minimal estimate is 3.0 kb, thus still 0.9 kb more than the mRNA.

Maps of the LiTat 1.6 ELC (fig. 1D) and of its isotypic T. b. brucei AnTat 1.8 ELC (fig. 1E) have been derived from the results of this study, as well as from additional experiments. Strikingly enough, large DNA stretches located on both sides of the ELC do not contain any site for the 13 restriction enzymes tested, indicating that they most probably consist of unusually simple sequences. Moreover, the restriction endonucleases whose site could be mapped downstream from the 3' end of the gene seem to cut the DNA at the same locus, suggesting that this locus might be the end of the chromosome.

On the other hand, a repetitive sequence is contained in the 1.4 kb PvuII-PstI fragment, at the 5' end of the TE, as indicated by the labelled smear produced by this probe. The repetitive element is more precisely located within the 1.05 kb PvuII-MspI segment, probably just outside the TE, since in EcoRI (not shown), BglI and SalI digests (fig. 3), the 1.05 kb probe shows the smear but not the ELC, whereas the 0.25 kb MspI-PstI probe reveals the ELC and not the smear. Another repetitive element has already been observed at the 3' extremity (ref. 4 and also fig. 2, probe 0.9 kb PstI-SalI). Repeated sequences are also observed in the BC clone, 2.5-3 kb outside both ends of the TE (fig. 2, probes 2.9 kb HindIII-PstI and 0.58 kb PstI-EcoRI). Using stringent washing conditions, we found no cross-hybridization between these different repeated sequences, in any of all the possible cross-reacting combinations (data not shown).

Figure 1. Restriction maps of the LiTat 1.6 BC (A), of two related sequences (B and C) and of the LiTat 1.6 and AnTat 1.8 ELCs (D and E, respectively). The extent of the TE is indicated by a thick line under (A), with uncertainties at both ends. The open box corresponds to the sequence transcribed into mRNA. Arrows and arrow heads under map A point to the limits of each of two clones used to construct the BC map. The arrowed 5' SalI site of the TE is referred to in the text. Fragments isolated to prepare the probes are shown under map (A). The fragments whose size is mentioned were ^{32}P-labelled and used as probes. B, BglI; Bg, BglII; E, EcoRI; H, HindIII; M, MspI; P, PstI; Pv, PvuII; S, SalI; Sp, SphI; X, XorII. For clarity, only the MspI sites located at the 5' and 3' ends of the TE are indicated; five MspI sites have been mapped within the messenger sequence. Lines under maps (B) and (C) indicate the regions of these clones which cross-hybridize with 5' and 3' BC probes, dashed sections meaning poor hybridization; pooled 0.6 kb EcoRI- PstI + 0.8 kb PstI-PvuII + 1.4 kb PvuII-PstI fragments have been used as 5' BC probe, and a mixture of 1.7 kb SalI-SalI + 0.5 kb SalI-PstI + 0.9 kb PstI-SalI as 3' BC probe. EcoRI sites could not be mapped with certainty in (D) and (E).

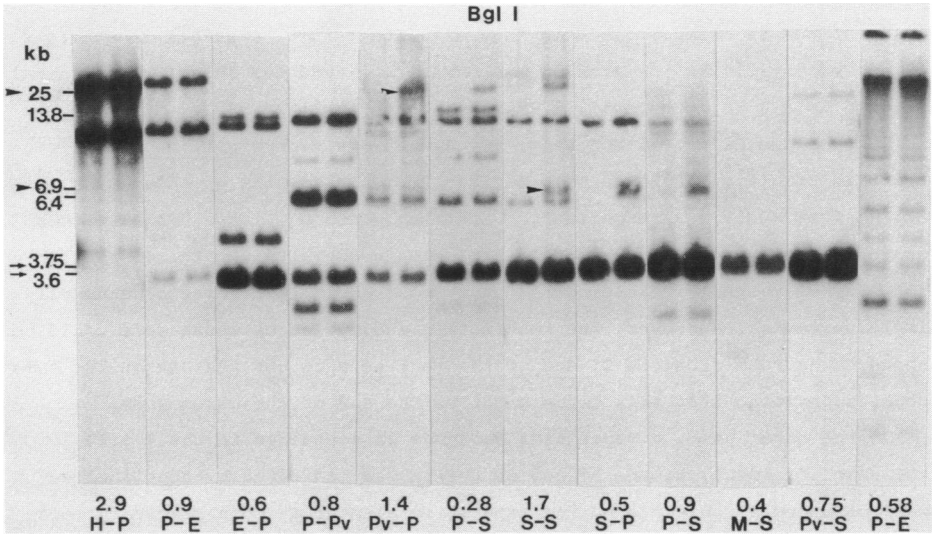


Figure 2. Hybridization pattern of LiTat 1.3 and LiTat 1.6 DNAs with probes isolated from the LiTat 1.6 BC clones. DNA was digested by BglI, then hybridized with ^{32}P-labelled probes as indicated, after "Southern" blotting on cellulose nitrate. Probes are identified by their length, in kb, and restriction cuts (same abbreviations as for fig. 1). The bands containing the BC and ELC are indicated by an arrow and an arrow head, respectively. HindIII digestion products of phage lambda and SV40 DNAs have been used as markers to determine the size of fragments.

The restriction maps of two LiTat 1.6 related sequences is presented in figure 1, B and C. One of them (fig. 1B) appears very similar to both BC and ELC. Six of the seven restriction sites mapped inside the LiTat 1.6 TE are conserved in all three cases, only the 3' PstI site being absent from the related sequence. Sites are also conserved outside the TE and cross-hybridization between BC probes and the cloned sequence extends over 6.5 kb. Interestingly enough, differences are concentrated in the regions flanking the TE. The other related sequence that we cloned (fig. 1C) differs more extensively in 5', although still cross-hybridizing with probes derived from the 5' part of the TE.

Selected fragments of the LiTat 1.6 BC clone have also been hybridized to "Northern" blots of total or poly(A)+ RNAs isolated from LiTat 1.6 trypanosomes, from their isotypes AnTat 3.3, AnTat 12.2 and AnTat 1.8, and from the heterotypes LiTat 1.3, AnTat 12.1 and AnTat 1.1. The probes derived from the TE, as expected, hybridize with specific RNA molecules in LiTat 1.6 extracts (fig. 4), whereas probes made from fragments outside the TE do not (not

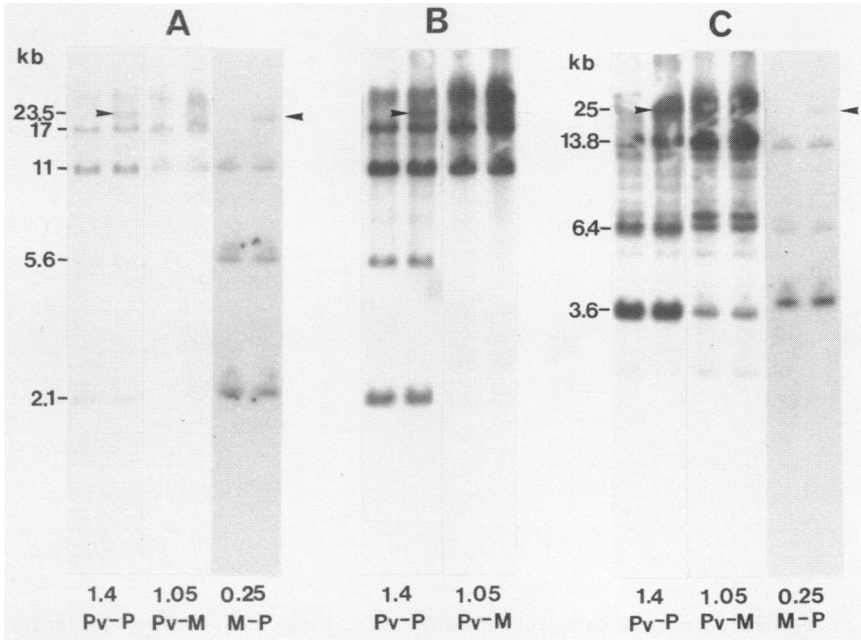


Figure 3. Hybridization pattern of LiTat 1.3 and LiTat 1.6 DNAs, respectively, with the 1.4 kb PvuII-PstI probe, or with two fragments isolated from this probe (1.05 kb PvuII-MspI and 0.25 kb MspI-PstI). Digestion was by SalI, (A and B) or BglI (C). Arrow heads point to the ELC-containing bands. In (A), exposure of the autoradiogram was short, allowing the ELC derived band to be seen in the smear of repeated sequences when using the 1.4 kb PvuII-PstI probe, but not with the 1.05 kb PvuII-MspI fragment of the probe. With the longer exposure in (B and C), the presence of repeated sequences in the 1.05 kb PvuII-MspI fragment appears more clearly. The hybrids were washed under stringent conditions (15mM NaCl, 1.5 mM Na citrate at 65°C)

shown). LiTat 1.6 probes corresponding to the gene (box in fig. 1A) also recognize the isotypic AnTat 3.3, AnTat 12.2 and AnTat 1.8 transcripts (fig. 4B), as described elsewhere for AnTat 1.8 cDNA probes (9). No LiTat 1.6-specific transcript could be detected in heterotypic clones LiTat 1.3, AnTat 12.1 and AnTat 1.1.

With probes covering a large segment downstream from the arrowed SalI site (fig. 1A) and thus extending within the LiTat 1.6 gene, the major RNA band detected is the 2.1 kb mRNA (fig. 4B), as noted previously (9). However, provided autoradiograms are overexposed, minor bands of 4.2 and 0.8 kb will come out, as illustrated in figure 4D with the 1.7 kb SalI-SalI probe. On the other hand, a probe derived from the 5' portion of the TE (1.4 kb PvuII-PstI)

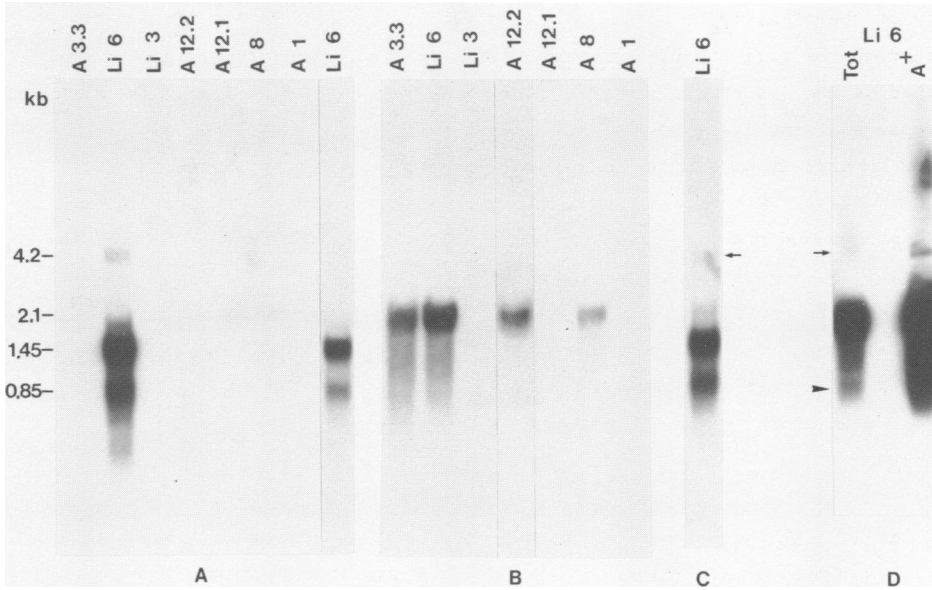


Figure 4. Hybridization patterns of AnTat 3.3, LiTat 1.6, LiTat 1.3, AnTat 12.2, AnTat 12.1, AnTat 1.8 and AnTat 1.1 poly(A)+ RNAs (10 µg) with the 1.4 kb PvuII-PstI probe (A), the 1.7 kb SalI-SalI and 0.5 kb SalI-PstI probes (B) or the 0.28 kb PstI-SalI probe (C) of the LiTat 1.6 TE. The last lane of (A) is the same as the second one, but 5-fold less exposed. Exposure for (B) was 50-fold less than for (A) and (C). The arrow in (C) points to the 4.2 kb transcript. Size markers were T. b. brucei ribosomal RNAs (2.4, 2.0 and 1.6 kb) and E. coli ribosomal RNAs (3.5 and 1.7 kb), visualized by ethidium bromide staining, as well as the 2.9 kb HindIII-PstI and 2.5 kb PstI- PstI fragments hybridized with the homologous probes (see fig.1). Comparative hybridization experiments with total and poly(A)+ LiTat 1.6 RNAs are shown in (D): the probe is the 1.7 kb SalI-SalI fragment. The arrow head and arrow point to 0.8 kb and 4.2 kb transcripts, respectively. The amounts used per slot of total and poly(A)+ RNAs are 20 µg and 2 µg, respectively.

does not reveal the 2.1 kb transcript, but only the minor components 4.2, 1.45 and 0.85 kb long (fig. 4A). Whether the 1.45 kb RNA also hybridizes with the 1.7 kb SalI-SalI probe is not clear, as the hybrid band would have been lost in the overlap of the highly labelled 2.1 kb mRNA. All these hybrids resist stringent washing conditions (0.045 M NaCl, 4.5 mM Na citrate, 0.1% Na dodecyl sulphate at 42°C; not shown). Interestingly, with our LiTat 1.6 probe these minor component are detected in LiTat 1.6 extracts only, not in the isotypes. The same pattern of minor RNA bands is obtained with another 5' probe, prepared from the 0.28 kb PstI-SalI fragment (fig. 4C)

The relative amount of RNA contained in the minor bands can be only very roughly estimated. Taking exposure times into account, the labelling intensi-

ty of the mRNA, in original autoradiographs, is about 50 fold higher than that of the 1.45 and 0.85 kb transcripts, and perhaps 500 fold higher than that of the 4.2 kb transcript. Since comparable hybridization of the mRNA is measured in at least 20 fold more total than poly(A)+ RNA, the amount of the 4.2 kb transcript can be estimated to be less than 0.01% of total RNA.

Whereas, as mentioned above, no hybridization is observed between the two 5' probes and the RNAs from the LiTat 1.6 isotypes AnTat 12.2, AnTat 3.3 and AnTat 1.8, sequences are recognized by the same probes in the genomic DNA of these isotypes (not shown). It should be stressed however that the experimental conditions were different and particularly stringent in the case of the RNA-DNA hybridization. That changes occurred between isotypes within this 5' region of the TE is evident, since the 5' PstI site of the TE is absent from the T. b. brucei DNA (fig. 1). Similarly, the PvuII and PstI sites mapped 2.3 and 3.1 kb upstream from the LiTat 1.6 BC are absent in T. evansi AnTat 3.3 DNA (not shown). These changes contrast with the conservation, in all isotypes, of the five restriction sites mapped within the limits of the AnTat 1.8 cDNA (9).

DISCUSSION

This study has revealed that the LiTat 1.6 TE is flanked by two repetitive elements. The same observation has been made by Van der Ploeg et al. (1) for other VSA genes, including VAT 117, an isotype of LiTat 1.6 in T. b. brucei. These sequences could correspond to recognition stretches involved in recombination with the expression site (1,19,20,21). In addition, we found repeated sequences about 2.5 to 3 kb outside both ends of the TE in the BC clone. The meaning of this observation is unclear.

The mapping of at least 8 different restriction sites at the same locus, 5.8 kb from the 3' end of the gene, is best interpreted in terms of this locus being the end of a chromosome. This situation is thus similar to the one described by Van der Ploeg et al. (1) in the case of the expression site for VATs 117 and 118, in T. b. brucei.

Also in accord with observations by these authors (1), the expression-linked copy of the gene is flanked by two regions devoid of most restriction sites, although in the LiTat 1.6 case, the 5' sequence without restriction sites is 12 kb long, instead of about 8 for the 117 and 118 clones. Except for this difference, it seems that the LiTat 1.6 expression site, in T. b. gambiense, is similar to the expression site occupied by the 117 and 118 ELCs in the brucei sub-species: indeed in both cases PvuII and HindIII sites are

located at the same positions, in front of the 5' site-depleted sequence, although the location of SalI and BglII sites differs in both cases. Moreover, the restriction map of the LiTat 1.6 TE appears to be very similar to that of VAT 117 (1). It thus appears that both the gene and its expression site are well conserved even among isotypes belonging to different subspecies. The expression site of AnTat 1.8, from T. b. brucei, also looks similar to that of its isotype LiTat 1.6 (fig. 1): the AnTat 1.8 TE is also located near a DNA end, within a 20 kb region depleted in many restriction sites, but the gene seems to have been transposed about 4 kb more to the left than the LiTat 1.6 ELC. Moreover, whereas a PvuII, PstI and HindIII sites appear to be conserved in front of both 5' site-depleted sequences, BglI, BglII, SphI and SalI sites differ in the two cases. Stretches of DNA devoid of restriction sites are found also at both sides of the ELC of the AnTat 1.1 gene (8 and unpublished results). However, in this case two restriction cuts (EcoRI and PstI) have been mapped in the expression site just in front of the TE (7,8). Preliminary observations (not shown) suggest that the expression site occupied by the AnTat 1.1 ELC is probably the one used for the expression of the LiTat 1.6 isotypic genes, but in a modified form.

The role of the unusual, probably very simple sequences which are bordering the ELC is still a matter of conjecture. Simple sequences are known to be present in the centromeric regions of chromosomes and are therefore probably involved in some way in the structural relations of chromatin with microtubules. In a large number protozoa, centromeres are directly associated to the nuclear envelope. That the ELC region of the genome is in close contact with the nuclear envelope and that the unusual sequences of the ELC environment are in some way involved in this attachment, are testable hypotheses. Whatever its function in the living trypanosome may be, this unusual DNA may be responsible of the repeated failures in cloning the ELC: we were unable to pick up a single ELC clone among more than 150 VSA-specific clones obtained. This has been observed also in other laboratories (R.O. Williams, P. Borst, personal communications). We recently succeeded however in the cloning of a 2.1 kb EcoRI fragment containing both the 5' portion of the AnTat 1.1 TE and its flanking sequence from the expression site. Study of this sequence is in progress.

Of the two LiTat 1.6 related sequences we have studied, one appears to be remarkably similar to the BC, along most of the TE as well as outside the TE. Differences, both in restriction sites and cross-hybridization, have been noticed only in the two regions corresponding to those flanking the TE.

Despite its overall similarity with the BC, this sequence seems not to be used for the synthesis of any LiTat 1.6-like VSA in trypanosomes stocks lacking the LiTat 1.6 BC: a close correlation has indeed been found between the presence in the genome of the LiTat 1.6 BC (but not LiTat 1.6 related sequences) and the detection of a LiTat 1.6-like VAT in the repertoire (22). It would thus appear that this sequence is never or only rarely used to synthesize a VSA of the same serological specificity as the LiTat 1.6 antigen. This could be correlated with alterations in sequences flanking the TE. We have shown that repeats are present in these regions in the BC; alteration in the repeats could prevent the specific recognition of the expression site, rendering the related sequence untransposable and thus unexpressed. It is however possible that despite its homology with the BC, this sequence could encode a VSA serologically unrelated to LiTat 1.6.

The LiTat 1.6 mRNA is about 2.1 kb long, including its poly(A) tail; the length of its transcribed part should be around 1.9 kb, considering that this sequence is similar to the AnTat 1.8 gene (9), whose full-length cDNA is 1.85 kb long (13). Since the length of the LiTat 1.6 TE is between 3.0 and 4.2 kb, a sequence of at least 0.9 to 2.1 kb is cotransposed in front of the gene. The most probable estimate of the LiTat 1.6 TE length should be just above 3 kb: indeed the additional copy cannot be seen with the 5' 1.05 kb PvuII-MspI probe, and it is revealed with the 0.25 kb MspI-PstI one (fig. 3). Therefore most of the 1.4 kb PvuII-PstI sequence, upstream from the MspI site, is not represented in the TE. That the TE length exceeds the mRNA size has also been observed for other VSA genes (1,8).

Using two genomic probes derived from the 5' section of the TE, we found no hybridization of these probes with the mature mRNA, thus confirming that a non-messenger sequence is cotransposed in 5', upstream from at least most of the messenger sequence.

With the same probes, we detected three RNAs distinct from the LiTat 1.6 mRNA. Models of LiTat 1.6 gene transcription have to take the following evidences into account:

1) The small transcripts of 1.45 and 0.85 kb are polyadenylated; they are indeed enriched in the poly(A)+ fraction (about 3%) of total RNA. Polyadenylation could result in higher stability (23,24) of the small transcripts and accounts for their easy detection. 2) Both 1.4 kb PvuII-PstI and 0.28 kb PstI-SalI probes recognize the same three RNAs: the 0.85 kb transcript corresponds thus to a sequence also present in and probably derived from the 1.45 kb RNA. 3) The strict LiTat 1.6 specificity of the three transcripts is

supported by the fact that the 0.28 kb probe does not contain a repetitive element and is thus represented in LiTat 1.6 sequences only. 4) These three transcripts are absent from the non-expressor LiTat 1.3 RNA: they are thus linked to LiTat 1.6 expression. Taken together, remarks 3 and 4 indicate that the 0.85, 1.45 and 4.2 kb transcripts are synthesized on the LiTat 1.6 TE in the expression site.

Different schemes can be proposed, producing the different polyadenylated transcripts:

1) Transcription starts in the expression site, about 4 kb upstream from the mRNA polyadenylation site, which is known from cDNA sequencing in the isotype AnTat 1.8 (19). Termination can however be variable, giving rise to the production of a 4.2 kb pre-mRNA and of smaller poly(A)+ RNAs (1.45 and 0.85 kb). In the case of μ immunoglobulin chain production, two polyadenylation sites can be alternatively used, determining whether the protein will be secreted or membrane bound (25). That polyadenylation sites can be passed by the RNA polymerase has been reported several times (25-30).

2) A 4.2 kb precursor is synthesized and processed as represented in figure 5. That the 4.2 kb RNA might be the primary transcript of the LiTat 1.6 gene is in keeping with the observation that it is recognized by the 1.7 kb SalI-SalI probe, which contains the 5' half of the messenger sequence. In this case however, it is evident that the non-messenger sequences would be excised from the 5' half of the 4.2 kb precursor and we must admit that the excised sequences can be polyadenylated. This has not been reported yet for processed intervening sequences in other eucaryotic RNA precursors, although the β-globin (30) as well as the SV40 and adenovirus mRNAs (26-29), are polyadenylated after resection of a 3' segment extending downstream from the poly(A) addition site. Moreover, polyadenylation could obey different rules in trypanosomes, since the consensus -AAUAAA- polyadenylation sequence (31) has not been found in any of the VSA genes studied so far (19,20,32,33). It is improbable that the 1.45 and 0.85 kb RNAs could encode proteins necessary for the expression of the gene: indeed they are transcribed from the 5' cotransposed sequence, which either differs or is not transcribed in the LiTat 1.6 isotypes and therefore does not seem to be essential for the expression of VSA genes. Finally, we do not exclude that a short 5' initial section of the mRNA could be transcribed from the expression site. In this case the cotransposed sequence would be intronic, at least in the ELC.

3) There is a promoter just in front of the gene in the TE, which is only activated by transposition, controlling the synthesis of the 2.1 kb mRNA.

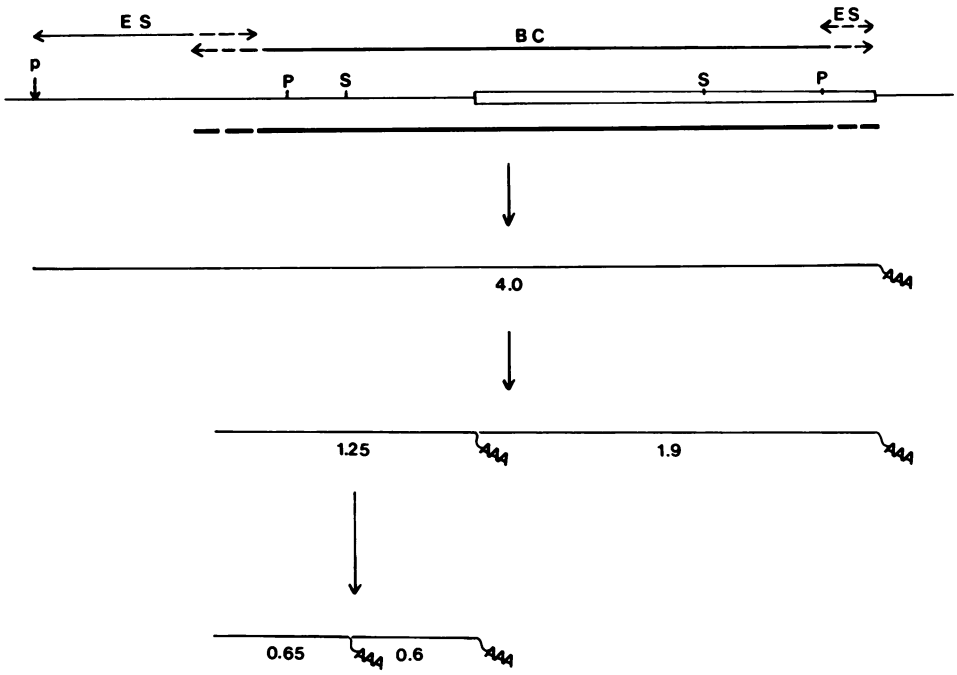


Figure 5. Diagrammatic representation of a hypothetical LiTat 1.6-specific mRNA processing. A pre-mRNA, 4.2 kb after polyadenylation, is transcribed from a promoter possibly located upstream from the transposition site. The TE should be between 2.75 and 3.15 kb long, a value in accordance with the most probable estimate (just above 3 kb). This pre-mRNA is rapidly processed into mature mRNA (2.1 kb, including the poly(A) tail) and at least two other fragments, one of which (1.45 kb, including poly(A)) is detected by TE probes. This 1.45 kb transcript is then cleaved in pieces, at least one of which (0.85 kb, including poly(A)) is detected by TE probes. The open box represents the DNA sequence covering the mRNA length; the thick line represents the TE, with uncertain ends. The size of the poly(A) tail has been estimated to be about 0.2 kb, on the observation that the AnTat 1.8 mRNA is about 2.05 kb long (8), whereas the full length of the AnTat 1.8 cDNA is approximately 1.85 kb (13). P, PstI; S, SalI; p, promoter; ES and BC, sequences from the expression site or copied from the basic copy, respectively, and engaged in transcription.

Another promoter is located in the expression site about 4 kb upstream from the last one, giving rise to the synthesis of a large 4.2 kb transcript from which the small transcripts are derived. One may speculate that the transcription of these RNAs could activate the gene promoter, although their sequence *per se* is probably not crucial for gene expression, for reasons given in scheme 2.

In any one of these models, the presence of a promoter is postulated in

the expression site. If the gene is activated by transposition of a copy beside the promoter, this model of DNA rearrangement would appear to differ from that of the mating type expression in yeast, where the equivalent of the BC (HMR or HML) is actively maintained silent by the product of a gene (SIR) (34,35).

ACKNOWLEDGEMENTS

We are grateful to Drs. N. Van Meirvenne and T. Vervoort (Institute for Tropical Medicine, Antwerp) for providing us with clone populations of trypanosomes from their collections, Dr. G. Marbaix for helpful discussions and to Dr. J. Deschamps for aid and advice on cloning in lambda vectors. This investigation received support from the Fonds de la Recherche Scientifique Médicale (FRSM, Brussels) and from the ILRAD / Belgian Research Centres Agreement for Collaborative Research (Nairobi).

Abbreviations: VSA, Variant-specific Surface Antigen; VAT, Variable Antigen Type; AnTat, Antwerp Trypanozoon antigen type; LiTat, Lille Trypanozoon antigen type; kb, kilo-base (or kilo-base pairs for double stranded DNA); BC, basic copy of the gene; ELC, expression-linked copy of the gene; TE, transposable element.

REFERENCES

1. Van der Ploeg, L.H.T., Bernards, A., Rijsewijk, F.A.M. and Borst, P. (1982) Nucleic Ac. Res. 10, 593-609.
2. Williams, R.O., Young, J.R. and Majiwa, P.A.O; (1979) Nature 282, 847-849.
3. Hoeijmakers, J.H.J., Frasch,A.C.C., Bernards, A., Borst, P. and Cross, G.A.M. (1980) Nature 284, 78-80.
4. Pays, E., Van Meirvenne, N., Le Ray, D. and Steinert, M. (1981) Proc. Natl. Acad. Sci. USA 78, 2673-2677.
5. Agabian, N., Thomashow, L., Milhausen, M. and Stuart, K. (1980) Am. J. Trop. Med. Hyg., Suppl. 29, 1043-1049.
6. Young, J.R., Donelson, J.E., Majiwa, P.A.O., Shapiro, S.Z. and Williams, R.O. (1982) Nucleic Ac. Res. 10, 803-819.
7. Pays, E., Lheureux, M. and Steinert, M. (1981) Nature 292, 265-267.
8. Pays, E., Lheureux, M. and Steinert, M. (1981) Nucleic ac. Res. 9, 4225-4238.
9. Pays, E., Lheureux, M. Vervoort, T. and Steinert, M. (1981) Molec. Biochem. Parasitol. 4, 349-357.
10. Vervoort, T., Barbet, A.F., Musoke, A.J., Magnus, E., Mpimbaza, G. and Van Meirvenne, N. (1981) Immunology 44, 223-232.
11. Magnus, E., Vervoort, T. and Van Meirvenne, N. (1981) Ann. Soc. belge Med. Trop. in press.
12. Auffray, C. and Rougeon, F. (1981) Eur. J. Biochem. 107, 303-314.
13. Pays, E., Delronche, M., Lheureux, M., Vervoort, T., Bloch, J., Gannon, F. and Steinert, M. (1980) Nucleic Ac. Res. 8, 5965-5981.
14. Arnemann, J., Heins, B. and Beato, M. (1977) Nucleic Ac. Res. 4, 4023-4036.
15. MacMaster, G.K. and Carmichael, G.G. (1977) Proc. Natl. Acad. Sci. USA 74, 4835-4838.

16. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5981.
17. Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980) Proc. Natl. Acad. Sci. USA 77, 5172-5176.
18. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
19. Matthyssens, G., Michiels, F., Hamers, R., Pays, E. and Steinert, M. (1981) Nature 293, 230-233.
20. Majumder, H.K., Boothroyd, J.C. and Weber, H. (1981) Nucleic Ac. Res. 9, 4745-4753.
21. 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, 497-505.
22. Pays, E., Dekerck, P., Van Assel, S., Le Ray, D., Eldirdiri A. Babiker, Van Meirvenne, N. and M. Steinert, (1982) submitted for publication.
23. Marbaix, G., Huez, G., Burny, A., Cleuter, Y., Hubert, E., Leclercq, M., Chantrenne, H., Soreq, H., Nudel, U. and Littauer, U.Z. (1975) Proc. Natl. Acad. Sc. USA 72, 3065-3067.
24. Huez, G., Bruck, C. and Cleuter, Y. (1981) Proc. Natl. Acad. Sci. USA 78, 908-911.
25. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) Cell 20, 313-319.
26. Nevins, J.R. and Darnell, J.E. (1978) Cell, 15, 1477-1493.
27. Ford, J. and Hsu, M.T. (1978) J. Virol. 28, 795-801.
28. Fraser, N.W., Nevins, J.R., Ziff, E. and Darnell, J.E. (1979) J. Mol. Biol. 129, 643-656.
29. Nevins, J.R., Blanchard, J.M. and Darnell, J.E. (1980) J. Mol. Biol. 144, 377-386.
30. Hofer, E. and Darnell, J.E. (1981) Cell 23, 585-593.
31. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
32. Boothroyd, J.C., Cross, G.A.M., Hoeijmakers, J.H.J. and Borst, P. (1980) Nature 288, 624-626.
33. Rice-Ficht, A.C., Chen, K.K. and Donelson, J.E. (1981) Nature 294, 53-57.
34. Klar, A.J.S., Strathern, J.N., Broach, J.R. and Hicks, J.B. (1981) Nature 289, 239-244.
35. Nasmyth, K.A., Tatchell, K., Hall, B.D., Astell, C. and Smith, M. (1981) Nature 289, 244-250.