The EMBO Journal Vol.2 No.7 pp.1185–1192, 1983

Gene activation and re-expression of a *Trypanosoma brucei* variant surface glycoprotein

F. Michiels, G. Matthyssens*, P. Kronenberger, E. Pays[1], B. Dero[1], S. Van Assel[1], M. Darville[1], A. Cravador[1], M. Steinert[1] and R. Hamers

Laboratorium Algemene Biologie, Instituut voor Moleculaire Biologie, Vrije Universiteit Brussel, 1640 St-Genesius-Rode, and [1]Département de Biologie Moléculaire, Université Libre de Bruxelles, 1640 Rhode-St-Genèse, Belgium

Communicated by T. Rabbitts
Received on 5 April 1983; revised on 9 May 1983

The expression of the *Trypanosoma brucei* variant surface glycoprotein AnTat 1.1 proceeds by a mechanism that transfers a duplicated gene copy into a new genomic environment, the so-called expression site, where it will be expressed. We have isolated a genomic fragment containing the region spanning the expression site-transposon junction, and the 5′ half of the coding sequence. Comparing this DNA segment with its template copy (basic copy) allowed us to identify the exact breaking point and indicated a base sequence which could be involved in initiating the transposition event. Sequencing data also indicated that the co-transposed segment 5′ to the coding sequence is 430 bp in length. The extreme 5′ end of the mRNA is derived from a region in the expression site not immediately adjacent to the transposed DNA segment. This particular sequence exists in multiple copies in the genome and is common to the mRNA of all variant surface glycoproteins so far analysed.

Key words: trypanosome antigenic variation/expression-linked copy/gene cloning/nucleotide sequence

Introduction

African trypanosomes are unicellular parasites that evade the immune response of their host by varying a single surface glycoprotein (VSG) in such a way that host antibodies directed against one VSG will be unable to recognize antigenic determinants on other VSGs. The repertoire of VSGs is encoded in the genome, and recent studies of these VSGs and their corresponding genes have given us some idea about the molecular mechanisms underlying this 'antigenic variation' phenomenon. Although trypanosome species may have a restricted host range, we have concentrated on *Trypanosoma brucei* which is infective for cattle. Cloned populations of *T. brucei* have been generated in several laboratories and extensive repertoires are now available derived from a single trypanosome cloning event (Van Meirvenne *et al.*, 1975). We report here on further analyses of two serologically cross-reacting and independently obtained variants, AnTat 1.1 and AnTat 1.1 bis belonging to the same antigen repertoire: AnTaR 1. We have shown recently, that at least six gene copies are recognized by an AnTat 1.1-specific cDNA probe, one of which is present only in the variant expressing the AnTat 1.1 and AnTat 1.1 bis protein (Pays *et al.*, 1981a). This copy is derived by a duplication event from a template copy (BC or basic gene copy) present among the remaining five gene copies (AnTat 1.1 family members) (Pays *et al.*, 1981b). The duplicated copy (ELC or expression-linked copy) is inserted into a new environment where it is transcribed (Pays *et al.*, 1981c). This ELC is absent in procyclic forms derived from AnTat 1.1 trypanosomes which do not express VSG on their surface. The AnTat 1.1 bis expression is controlled by a similar duplication-transposition event. One of the questions we set out to answer was whether the AnTat 1.1 and 1.1 bis genes share the same expression site.

Activation of VSG genes involving duplication and transposition into an expression site have been reported for other *T. brucei* repertoires (Hoeijmakers *et al.*, 1980; Majiwa *et al.*, 1982). The analysed ELCs share a number of common features. (a) Between 0.4 and 2 kb upstream of the VSG coding sequence are co-transposed into the expression site. (b) The transposed segment is flanked by DNA regions lacking restriction enzyme sites, and varying in size in different variants. These barren regions have impaired the cloning of ELC copies. (c) A putative cluster of restriction sites flanking the barren region at the 3′ end has been shown to be the end of a chromosome or (large) DNA segment (De Lange and Borst, 1982; E. Pays *et al.*, in preparation; Williams *et al.*, 1982).

The expression site, therefore, seems to have special properties and we decided to clone a DNA segment containing the expression site-transposed segment boundary. The AnTat 1.1 and 1.1 bis ELC genes seemed to be particularly suitable since, unlike other ELCs, appropriate restriction enzyme sites could be found between the 'barren' 5′ region and the transposed gene segment.

Results

Cloning of ELC-derived DNA segments from variant AnTat 1.1 and AnTat 1.1 bis and of the corresponding AnTat 1.1 BC

Previous data has indicated a 2-kb fragment characteristic of the AnTat 1.1 and 1.1 bis ELC in *Pst*I digests of genomic AnTat 1.1 and 1.1 bis DNA. In an *Eco*RI digest, the ELC appears as a 2.1-kb fragment (Figure 1). In both ELC fragments, the 5′ cutting sites are located in the expression site [260 bp and 1120 bp, respectively, from the junction (J) between expression site and transposon], and the 3′ cutting sites are located within the coding region (Pays *et al.*, 1981a, 1981b; Matthyssens *et al.*, 1981). Furthermore, among the five additional AnTat 1.1 family members (Figure 1) which are present in all AnTaR 1 clones so far studied, only one shows restriction enzyme sites identical to the AnTat 1.1 ELC. This is the only copy that has the *Hind*III site characteristic for AnTat 1.1 and 1.1 bis cDNAs (indicated by the arrow in Figure 1). This gene copy gives rise to the 6.4-kb *Pst*I fragment which is therefore derived from the AnTat 1.1 BC gene segment. Further restriction mapping and cloning have indicated that the *Pst*I 9-kb family member (Figure 1) contributes to the formation of the AnTat 1.1 bis ELC gene (see Figure 1, Discussion and E. Pays *et al.*, in preparation).

The 2.1-kb *Eco*RI fragment derived from AnTat 1.1 and 1.1 bis was eluted from an agarose gel and ligated into phage λgt WES.λB (Tiemeier *et al.*, 1976). Several positive clones

*To whom reprint requests should be sent.

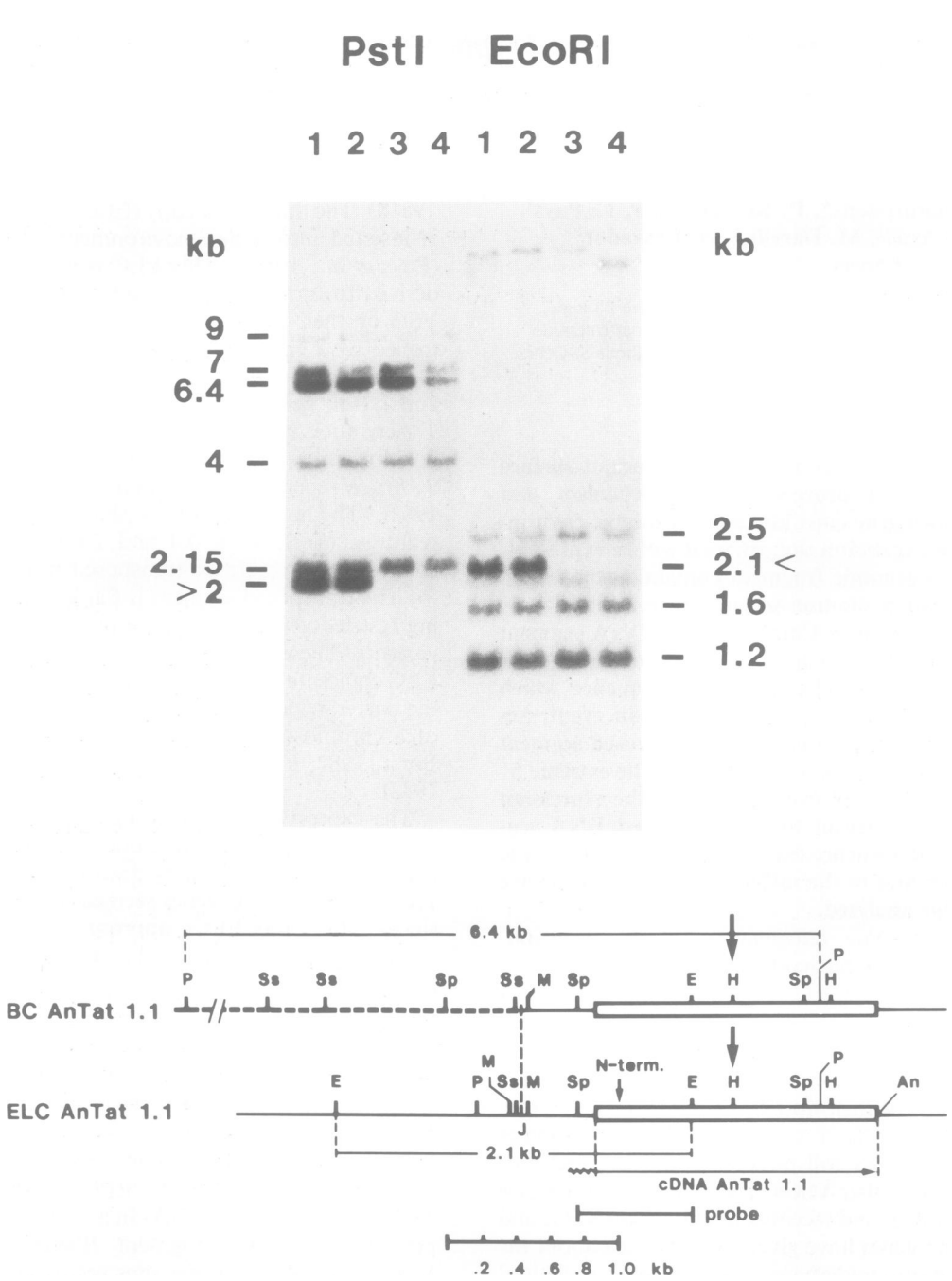


Fig. 1. Top: hybridization of the ELC-derived 650 bp *Sph*I-*Eco*RI fragment (indicated by the solid bar under the restriction map) to trypanosome DNA digested with *Pst*I and *Eco*RI. Tracks labelled **1, 2, 3** and **4** contain DNA isolated from clones expressing AnTat 1.1, AnTat 1.1 bis, AnTat 1.8 (which is not recognized by antiserum directed to AnTat 1.1 or 1.1 bis) and AnTat 1.1-derived procyclic trypanosomes (expressing no VSG on the surface), respectively. The arrowheads point to ELC-derived bands. **Bottom**: restriction enzyme maps of the AnTat 1.1 BC and ELC genes. These data are derived from Southern blotting and genomic cloning: BC genomic clones were isolated containing either the 5.7-bp *Pst*-*Eco*RI fragment or the 6.4-kb *Pst*I-*Pst*I fragment; ELC genomic clones contained the 2.1-kb *Eco*RI-*Eco*RI fragment. The open box indicates the AnTat 1.1 mRNA sequence as derived from two overlapping cDNA clones covering the whole gene; the wavy lane at the 5′ end of the cDNA tracing indicates a 153-bp segment not encoded within our derived genomic clones (see also Figure 3). An: poly(A) addition site. N-term: indicates the N-terminal amino acid position of the mature VSG as determined by protein sequencing of the isolated AnTat 1.1 VSG. The arrow above the *Hind*III site (H) indicates a conserved site found only in the *Pst*I 6.4-kb family member (BC-AnTat 1.1) and the *Pst*I 2.0-kb family member (ELC-AnTat 1.1 and 1.1 bis). J: indicates the breakpoint between expression site and transposed gene segment, as deduced from nucleotide sequence comparison between the BC and ELC genomic clones. The restriction enzyme abbreviations are as follows: P = *Pst*I; Ss = *Sst*I; M = *Msp*I; Sp = *Sph*I; E = *Eco*RI; H = *Hind*III; Rs = *Rsa*I; D = *Dde*I.

hybridizing to AnTat 1.1 cDNA (Pays *et al.*, 1980) were selected for study and one derived from AnTat 1.1 and 1.1 bis cloning was completely sequenced, with partial verifying sequencing done on others. The AnTat 1.1 and 1.1 bis ELC clones turned out to be identical. Southern blotting data using the 650-bp *Sph*I-*Eco*RI fragment containing the 5′ coding region showed the expected hybridization pattern (Figure 1). The restriction enzyme map shown (Figure 1) was identical to the map generated previously using AnTat 1.1 cDNA as probe on digests of total genomic DNA (Pays *et al.*, 1981a, 1981b).

To clone the BC of AnTat 1.1, AnTat 1.1 DNA was

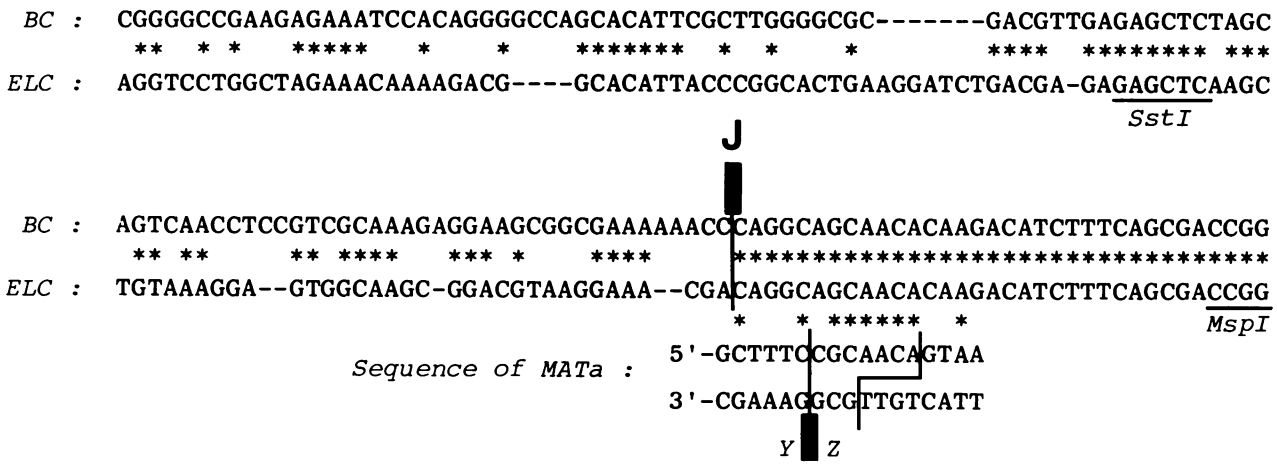


Fig. 2. Comparison of the 5′ transposon boundary as it occurs in the BC and ELC (J in Figure 1) DNA segments. Dashes indicate gaps introduced as required for alignment. The double-stranded sequence at the bottom is that of the yeast Y-Z boundary at the MATa locus, with the indication of the staggered cut generated by the yeast double strand-specific endonuclease.

digested with *Pst*I and the 6.4-kb fragment was isolated from an agarose gel. Southern blotting showed it to be free of the 7-kb fragment. The DNA was made blunt using T4 polymerase 3′ exonuclease activity followed by *Eco*RI-linker addition. Upon *Eco*RI cleavage, two fragments are generated: a 5.7-kb *Pst-Eco*RI fragment and a 0.7-kb *Eco*RI-*Pst*I fragment (see Figure 1). These were ligated into λgt WES.λB. Positive clones containing the 5.7-kb *Pst*I-RI fragment were selected and one of them was sequenced starting at the 5′ *Sph*I site down to the 3′ *Eco*RI site within the coding sequence. Among the AnTat 1.1 family members, the two *Sph*I sites flanking the 5′ transposon boundary J are found only in the 6.4-kb *Pst*I fragment. One of our AnTat 1.1 BC clones contains the *Pst*-RI 5.7-kb fragment ligated head to tail to the 3′ RI-*Pst* 0.7-kb fragment (confirmed by electron microscopy, Ch. Brack, unpublished data).

Figure 2 compares the ELC and BC sequences around the junction between the expression site and transposed segment. After aligning the sequences to allow for maximum homology, we found a 60% base sequence homology within a 60 nucleotide region upstream of the breakpoint, J. A 19-bp sequence flanking J in the ELC is composed of two 10 bp segments which are repeated 650−700 bp upstream and downstream of J (arrows in Figure 3). Furthermore, we have compared our data with eucaryotic DNA sequences known to be involved in transposition events. Comparing our data with the MAT locus in yeast *Saccharomyces cerevisiae* known to be involved in the mating type switching, revealed a homology with either MATα or MATa (Astell *et al.*, 1981) at the Y-Z boundary. The homology with MATa is indicated by asterisks in Figure 2.

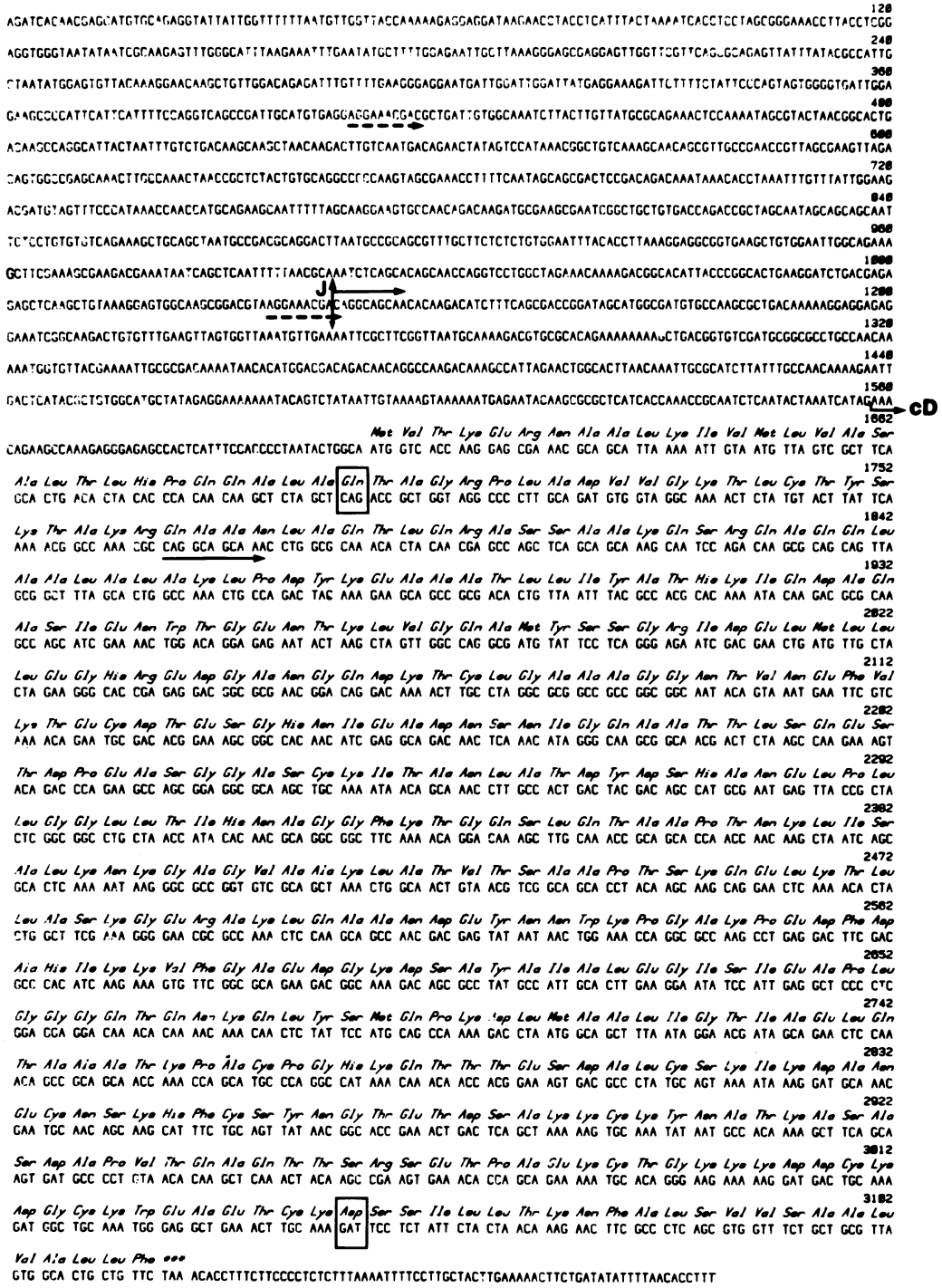
Complete amino acid sequence of the AnTat 1.1 variant surface glycoprotein

To identify the protein coding sequence within the ELC-derived fragment, we isolated two AnTat 1.1 cDNAs. The first one extends from the extreme 5′ boundary (see below) to 100 bp downstream of the *Eco*RI site, the second one extends from 140 bp upstream of the *Eco*RI site to the poly(A) (Figure 3 and Matthyssens *et al.*, 1981). The former allowed us to identify the exact 5′ boundary of the transcribed sequence in the ELC (cD in Figure 3). Since we have identified the expression site-transposon junction, J, the length of the non-transcribed co-transposed DNA segment is 432 bp.

The complete sequences of the cDNAs were determined and aligned with the BC and ELC sequences. At the overlapping positions, the sequences were found to be identical. From the only open reading frame, we were able to derive the complete amino acid sequence of the AnTat 1.1 VSG (Figure 3). Base position 1696 was found to be the first base of the triplet coding for the N-terminal amino acid, as determined by amino acid sequencing of the mature AnTat 1.1 VSG protein. The mature glycoprotein contains 450 amino acid residues, which is very similar to the 11Tat 1.3 mature glycoprotein (Rice-Ficht *et al.*, 1981) and 20 residues shorter than the VSG 117 (Allen *et al.*, 1982). At the N-terminal end, two methionine residues occur in-frame: in accord with the rule that a basic residue (e.g., Arg) precedes the hydrophobic region of the leader (Davis and Tai, 1980), we have assigned the first methionine (at position 1610) as the initiator residue. Apart from the hydrophobic character, there is no sequence homology or length conservation with other published VSG leader sequences. The remainder of the coding sequence is again very different from other VSGs except for the conserved position of the cysteine residues. As mentioned earlier, the only region of homology, therefore, is located in the C-terminal domain of the protein (Matthyssens *et al.*, 1981).

The 5′ end of VSG mRNAs is conserved

From an AnTat 1.1 cDNA bank, we selected a clone containing the extreme 5′ end by hybridization with the ELC-derived *Sph*I-*Eco*RI fragment (solid bar, Figure 1). This cDNA clone with an insert size of 800 bp, and with its 3′ end 100 bp downstream of the *Eco*RI site, had an identical sequence to the ELC up to position 1558 (Figure 3). Towards the 5′ end the cDNA contains another 154 bases which are not found either in the cloned ELC or BC segment (Figure 4). The boxed 35-bp segment was found to be identical to the 5′ end of another VSG mRNA belonging to a different antigen repertoire (Van der Ploeg *et al.*, 1982a). In the meantime, this sequence has been identified in a number of different VSG mRNAs (Boothroyd and Cross, 1982). Our data confirm therefore that the VSG mRNA is derived from a larger precursor RNA (Bernards *et al.*, 1981). The ELC sequence at the divergence point with the mRNA reads AG.A, which is different from the 'AG.G' intron-exon boundary consensus



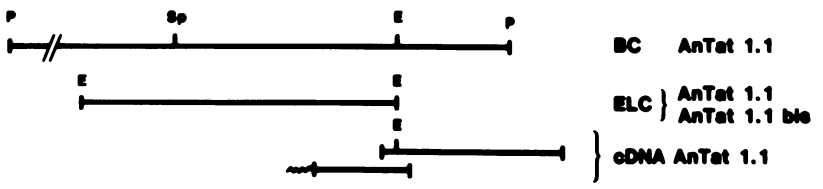


Fig. 3. Top: complete sequence of the AnTat 1.1 transposon plus 1.1-kb upstream sequence derived from the expression site. Amino acid sequence of the AnTat 1.1 VSG, starting at the methionine initiator is indicated. N-terminal and C-terminal residues of the mature protein are boxed. J: breakpoint between expression site and transposon; the two arrows indicate the 10-bp direct repeats, referred to in the text. The start of the cDNA is also indicated (cD). **Bottom**: line drawing of the genomic clones and cDNA from which the sequence data were derived. Except for the BC which was sequenced downstream of the indicated *Sph*I, all the others were sequenced to completion. The *Eco*RI site is at position 2104 in the coding sequence.

```
cDNA :  AAGCGACCGCAACTCCCTNAGTTCTAAGAACTTTCTTAATCCATATACATGGGGTA
                          Dde I  Dde I                         50

cDNA :  AATACATTTATAAGGGATAATTGCAAATAATAATAATAATAATTAAAAAAAAACAA
                                                          100

cDNA :  GACCAG|AACGCTATTATTAGAACAGTTTCTGTACTATATTG|AAACAGAAGCCAAAG
             ●●●                          RsaI    153
          *    *       *    **           **     **   **************
ELC  :  CGCGCTCATCACCAAACCGCAATCTCAAT.ACTAAATCATAGAAACAGAAGCCAAAG
                                                ●●●  1563
```

Fig. 4. Comparison of the sequence at the 5′ extremity of an AnTat 1.1 cDNA with the AnTat 1.1 ELC-derived genomic fragment. The boxed sequence is conserved between all VSGs and is referred to in the text. Putative intron-exon boundaries are indicated (●). ELC numbering is as in Figure 3. The base denoted by 'N' could not be determined.

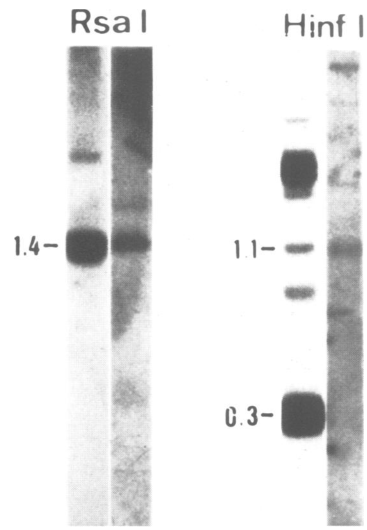


Fig. 5. Hybridization of 35-bp mini-exon and AT probe with *Rsa*I and *Hinf*I digests derived from AnTat 1.1. The left hand tracks show hybridization to mini-exon oligonucleotide labelled by kinasing (sp. act. 3 x 10^8 c.p.m./μg), the right hand tracks were hybridized to the AT probe (sp. act. 2 x 10^8 c.p.m./μg), in 6 x SSC and 10 x Denhardt's solution at 65°C. Washes were in 6 x SSC (mini-exon) and 1 x SSC (AT probe). Mini-exon probed filters were exposed for 12 h, AT-probed filters for 4 days, respectively, using tungstate intensifying screens. Numbers are in kilobases.

sequences of other eucaryotic structural genes (Breathnach and Chambon, 1981).

The 35-bp sequence was verified by direct sequencing of AnTat 1.1 and 1.1 bis mRNAs using a 15-bp synthetic oligonucleotide primer complementary to the position 1558–1612 in the ELC sequence shown in Figure 3. A strong sequencing stop occurred in all four dideoxy runs, 35 nucleotides beyond the primer. It was possible, however, to read specific though faint signals beyond this stop (data not shown). This would indicate that our cDNA was derived from the putative cloning of a precursor RNA: the boundary confirms the AG.A intron-exon junction (see also Van der Ploeg *et al.*, 1982a). One of the striking features of this cDNA is the AAT repeat ending in a stretch of nine As. Similar AAT blocks were found upstream of the 118 BC gene (Liu *et al.*, 1983) and the IlTAT 1.3 gene (copy A, Young *et al.*, 1983; J. Shah, unpublished data).

Subsequently, we tried to define the origin of the 35-bp mini-exon and the AT-rich precursor-derived sequence by using a 35-bp oligonucleotide complementary to the mini-exon, as well as the complete 5′ 118-bp precursor-derived sequence (AT probe) as probe in Southern blots of *Rsa*I and *Hinf*I digests of genomic DNAs (Figure 5). A strong 1.4-kb *Rsa*I and 0.3-kb *Hinf*I fragment hybridizes to the mini-exon probe. A copy-number experiment using the min-exon probe has indicated that ~250 copies are present per haploid genome (F. Michiels, unpublished results). The repetitive nature of the mini-exon has also been found by T. De Lange (unpublished results). We think that the mini-exons occur as a tandem repeat 1.4 kb apart, because digests with *Xmn*I (not shown) and *Rsa*I (Figure 5) which both cut the mini-exon, generate the same 1.4-kb fragments. Longer exposures, especially in the *Hinf*I digest, reveal additional hybridizing fragments, one of which (1.1 kb) also hybridizes to the AT probe (Figure 5, *Hinf*I digest, right track). The fact that the AT probe hybridizes to one or a subset of DNA fragments containing the mini-exon indicates that the AT probe is situated upstream of a mini-exon in the genome.

Discussion

VSG genes present in all trypanosomes, are selected for expression from their large repertoire, by a mechanism that involves DNA rearrangement. Some VSG genes belong to families of related sequences. One way of expressing a particular VSG gene among a repertoire is by gene duplication and transposition in an expression site (Hoeijmakers *et al.*, 1980; Pays *et al.*, 1981a; Majiwa *et al.*, 1982). Another type of rearrangement seems not to be linked directly to the expression of that VSG gene, though a similar duplication mechanism may be involved (Young *et al.*, 1983; Pays *et al.*, unpublished data). The expression of the AnTat 1.1 gene described here is clearly linked to the duplication-transposition mechanism (Pays *et al.*, 1981a). A second variant AnTat 1.1 bis, which is serologically cross-reactive to

AnTat 1.1 was derived from AnTat 1.1 *via* a third variant AnTat 1.10 (for a concise pedigree, see Figure 1 in Pays *et al.*, 1981a). Southern blotting data indicate that the genes for these three variants have been transposed into the same or a homologous expression site since the restriction enzyme pattern beyond the 5′ barren region flanking the cloned sequence is identical for AnTat 1.1, 1.1 bis and AnTat 1.10 (Pays *et al.*, in preparation). Moreover, our results clearly indicate that the 3′ 1125 bases of the expression site are identical in AnTat 1.1 and AnTat 1.1 bis. This shows, therefore, that these two VSG genes use the same expression site. Furthermore, within the transposon, the two variants are identical for the major part of their sequence. Comparison of the AnTat 1.1 and 1.1 bis cDNAs (unpublished results) shows that only the 3′ end of these two variants are different, with the divergence point between 600 and 700 nucleotides upstream of the poly(A). Hence it is not surprising that both genes are serologically cross-reactive, since the antigenic sites of VSGs are considered to be N-terminal. Taken together, it seems logical to conclude that in generating the AnTat 1.1 bis (and 1.10) variants, the original AnTat 1.1 ELC has remained in the expression site giving rise to the AnTat 1.1 bis (and 1.10) expressed gene by a recombination involving the AnTat 1.1 ELC and AnTat 1.1 bis BC, (located on the 9-kb *Pst*I fragment) leading to a replacement of the 3′ end of the AnTat 1.1 ELC. Differences at the 3′ end have also been described for two other serologically identical clones isolated independently (Michels *et al.*, 1982).

What is the transposition mechanism?

From what we have seen above, the formation of the ELC involves duplication and insertion into a recipient site. We have compared the sequence around this site with other sites involved in gene rearrangements associated with control of gene expression, e.g., phase variation in *Salmonella* (Zieg *et al.*, 1978), mating type switching in yeast (Hicks *et al.*, 1979; Nasmyth *et al.*, 1981) and antibody gene expression (Sakano *et al.*, 1979; Matthyssens and Rabbitts, 1980; Dunninck *et al.*, 1980). In Figure 2, we have indicated a homology that exists between our J region and the sequence found at the Y-Z boundary of the MAT locus in yeast: at Y, the transposition substitution reaction removes the cassette and inserts another donor cassette. A double strand-specific endonuclease recognizes, cuts and generates a 3′ overhang at the site near the MAT Y-Z junction thus initiating the switching event (Strathern *et al.*, 1982). The sequence GCAACA occurs, however, in two other locations: once in the expression site (position 567), and once in the coding sequence (position 2837). Probably, the specificity of the putative trypanosome endonuclease is conferred by a larger recognition site than the one revealed by the Y-Z homology, and/or by positional effects of the expression site (see Introduction). It should be possible to delineate DNase I-hypersensitive regions, highly accessible to enzymatic cleavage, around the BC and ELC. Studies are underway to identify trypanosome enzymes with endonucleic activity. If such enzymes can be found, site-specific cutting might identify potential expression sites in the genome.

The sequence we have analysed 5′ to J is derived from a separate duplication event of a DNA segment located at an unstable DNA terminus (Pays *et al.*, in preparation). There is no homology with known VSG coding sequences, and the longest open reading frame is 558 nucleotides long, in the same frame as the AnTat 1.1 gene. Two AnTat 1.1 specific non-VSG transcripts initiate upstream of the *Pst*I site: careful examination of the sequence within this area actually reveals two potential promoter regions showing homology to the 'CAAT' (position 513 and 690) and 'TATA'-box (position 551 and 736) consensus sequences (Breathnach and Chambon, 1981). Whether these transcripts are involved in the transposition and somatic recombination mechanisms linked to VSG expression remains to be seen. The expression site-derived *Eco*RI-*Pst*I fragment furthermore hybridizes to a large number of bands in genomic digests of DNA derived from AnTaR 1 variants (data not shown). We have therefore compared this sequence with that from the ELC of IoTat 1.3 (Donelson, unpublished data): we found a 70% base homology between region 130–410 (i.e., the sequence upstream of the 10-bp repeat, see Figure 3 of ELC AnTat 1.1) and a segment of the same length near the IoTat 1.3 J region (this region is less defined here due to multiple single base substitution between BC and ELC). This similarity is a further indication of conserved DNA segment homologies that may be used in the transposition event between different VSG genes.

The sequence 5′ to J in AnTat 1.1 is not found in a third AnTat 1.1 expressor AnTat 1.1D (E. Pays, in preparation), and seems therefore not to be required for AnTat 1.1 expression. This finding does not invalidate the significance of the homology with the MAT endonuclease recognition sequence since it is believed that this sequence is contained within a 12 nucleotide stretch downstream of the first C in the Z locus (Figure 2), and therefore located downstream of J (Kostriken *et al.*, 1983). This finding is also in agreement with the position of the putative promotor sequence which seems to be located much further upstream (see below).

Looking for the promoter

Translocation into an expression site may imply the presence of a strong promoter upstream of the expression site-transposon junction J. Since the 5′ end of the VSG mRNA is not encoded within the 1.1-kb expression site sequence we have analysed, transcripts have to originate upstream of the cloned DNA segment. The conserved nature of this part of the mRNA indicates a common promoter region for all VSG genes. This is in agreement with restriction enzyme mapping data showing homologous restriction enzyme sites at the 5′ end of the barren region upstream of the junction (J) point. We have not yet identified the origin of the 35-bp conserved DNA segment (Figure 4). Our data confirm that the initial transcript has to be spliced. The reasons for the repetitive nature of this mini-exon are less clear. Since the co-transposed DNA segment of various VSG genes differs and since restrictions may exist on the length of transcript, flexibility may be needed in transcription starts. Alternatively, the mini-exon cluster may act as a separate transposition unit (eventually on a separate mini-chromosome or circular piece of DNA) which can be transposed into the vicinity of existing VSG genes. This would require base pairing between homologous regions on both sides of the transposition unit. In this respect, it is interesting to note that a 'TAA' repeat which we identified 5′ to the 35-bp sequence, is also found 5′ to the BC of the 118 VSG gene (Liu, unpublished data), and 5′ to the I1Tat 1.3 silent copy A (Shah, unpublished data). In both cases, the TAA repeat occurs at the 5′ boundary of the DNA segment that can be duplicated and transposed.

We have presented evidence that the cDNA sequence 5′ to the mini-exon and the mini-exon itself are present on the same *Rsa*I and *Hin*fI restriction fragment. Contrary to the high

mini-exon copy number, the sequence 5′ to it is present in one copy per haploid genome only. If we assume that all AnTat 1.1 VSG RNA contains the AT-rich region identified in the cDNA, this implies that only one mini-exon is transcribed at any one time. Double digests with *Dde*I (an enzyme present in the 5′ AT region) and *Rsa*I indicate, furthermore, that these regions are not contiguous (F. Michiels, unpublished data). VSG trancripts must, therefore, be subject to a number of splicing events generating the mature mRNA. Furthermore, *Rsa*I and *Xmn*I digests (Figure 5 and unpublished) show that the mini-exon copies are 1.4 kb apart. Restriction sites in the spacer regions are loosely conserved, resulting in, e.g, 0.3-kb *Hinf*I repeat and larger bands containing fewer mini-exon copies. The AT-rich sequence is associated with one of these larger bands (1.1 kb).

Other DNA regions close to the methionine translation initiator may be of importance in regulating gene expression. The sequence starting at position 1417 matches perfectly with the first nine bases of the consensus sequence found in the 5′-flanking region of steroid hormone-responsive genes (Mulvihill *et al.*, 1982). This sequence was also found to be involved in poly(rI).poly(rC)-dependent inducibility of the interferon γ gene (Tavernier *et al.*, 1983). The distance from this sequence to the start of the cDNA as indicated in Figure 3 (141 bp), is similar to that in the γ-interferon gene (174 bp). Maybe trypanosomes can modulate their VSG transcription by binding regulator molecules (protein or RNA) to such receptors.

Finally, comparison of the DNA-derived amino acid sequence of the AnTat 1.1 glycoprotein with other published data confirm that sequence homology is confined to the C-terminal region of the molecule. cDNA sequencing has indicated the presence of a hydrophobic C-terminal extension on the precursor VSG which is lacking in the mature protein (Boothroyd *et al.*, 1980; Matthyssens *et al.*, 1981). This part of the protein is involved in several processing steps such as glycosylation and proteolytic cleavage possibly preparing the trypanosome for insertion of a new VSG. Such a phenomenon is another exciting aspect of the trypanosome antigenic variation which will need further effort to understand its genetic bases.

Materials and methods

Trypanosomes

Derivation of the trypanosome clones from *T. brucei brucei* stabilate EATRO 1125 have been described (Van Meirvenne *et al.*, 1975). This repertoire, called AnTaRI (Antwerp Trypanosome Antigenic Repertoire no. 1) consists of different cloned variants (proven by immunofuorecence to be 99% pure) indicated by the second number (AnTat 1.1, AnTat 1.3, etc.). Variant AnTat 1.1 was the first cloned member of the series. From this, new variants were obtained by neutralization-infection tests and subsequent cloning. The fully homogeneous variants used in this work were kindly provided by Dr. Van Meirvenne (Tropical Institute, Antwerp). Procyclic forms, lacking VSG on the membrane, were derived from AnTat 1.1 grown in culture, as described (Le Ray, 1975), and were kindly provided by Dr. Le Ray (Tropical Institute, Antwerp). Trypanosomes were isolated from infected rats by Percoll gradient flotation (Grab and Bwayo, 1982).

Recombinant DNA clones

Derivation of AnTat 1.1 cDNA clones has been described (Pays *et al.*, 1980). The cDNA clone extending towards the 5′ end was cloned into the *Pst*I site of pBR322 using GC-tailing.

Genomic clones were obtained from restriction fragments purified on 0.85% low melting agarose.

Direct ligation to λgt WES.λB-arms, or using linker addition, were performed as described (Maniatis *et al.*, 1978). Ligated DNA was packaged into phage particles as described by Blattner *et al.* (1977), and plaque screening using the 650-bp *Hind*II-*Hind*III fragment from AnTat 1.1 cDNA (Pays *et al.*, 1980) was performed according to Benton and Davis (1977).

Restriction enzyme mapping and sequencing of the AnTat 1.1 gene

Procedures for isolating trypanosome and plasmid DNA were as described (Pays *et al.*, 1981a) as were conditions for Southern blot analysis using specific cDNA and genomic DNA restriction fragments as probe. This allowed the construction of detailed maps. Confirmation was obtained by sequence analysis: overlapping restriction fragments were ligated into M13mp8 and M13mp9 (Messing and Vieira, 1982) either directly or by filling in and ligation into the M13 *Sma*I site. The dideoxy chain termination procedure was used in all experiments (Sanger *et al.*, 1980).

Synthesis of the oligonucleotide primer 5′-HO.CpTpTpTpGpGpCpTpTpCp TpGpTpTpTp.OH-3′

The synthesis was performed on a polystyrene solid support *via* the phosphotriester method using two protected monomers and four protected dimers. Purification was on Sephadex G50 and on reverse phase h.p.l.c. using a C18 column.

Construction of the AT probe

The cDNA clone, with the insert at the *Pst*I site of pBR322, containing the sequence 5′ to the mini-exon was digested with *Rsa*I, subjected to limited *Bal*31 exonuclease treatment, phenol extracted and digested with *Pst*I. The fragments were ligated into *Pst*I-*Sma*I cut M13mp9, and a clone containing the first 118 bp (Figure 4) was selected by sequencing. This so called 'AT probe' was released from the M13 and labeled to a specific activity of 2×10^8 c.p.m./μg.

Acknowledgements

We are indebted to T. Vervoort for supplying us with VSG protein AnTat 1.1 and to M. Lauwereys and R. Vansteelandt for determining the N-terminal amino acid sequence. We also would like to thank W. Verheulpen for computer programmes, E. Wittouck for technical assistance, A.Y.C. Liu and J. Donelson for providing us with unpublished sequence information and R.O. Williams for critically reading the manuscript. We also would like to express our appreciation to Lucy Thairo for typing the manuscript. F. Michiels thanks the I.W.O.N.L., Belgium, for a fellowship. This work was supported by ILRAD/Belgium Research Centers Agreement for Collaborative Research (Nairobi).

References

Allen,G. Gurnet,L.P. and Cross,G.A.M. (1982) *J. Mol. Biol.*, **157**, 527-546.
Astell,C.R., Ahlstrom-Jonasson,L., Smith,M., Tatchel,K., Nasmyth,K.A. and Hall,B.D., (1981) *Cell*, **27**, 15-23.
Benton,W.D. and Davis,R.W. (1977) *Science (Wash.)*, **196**, 180-182.
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.
Blattner,F.R., Williams,B.G., Blechl,A.E., Denniston-Thompson,K., Faber, H.E., Furlong,L., Grunwald,D.J., Kiefer,D.O., Moore,D.D., Schumm, J.W., Sheldon, E.L. and Smithies,O. (1977) *Science (Wash)*, **196**, 161-169.
Boothroyd,J.C., Cross,G.A.M., Hoiejmakers,J.H.J. and Borst,P. (1980) *Nature*, **288**, 624-626.
Boothroyd,J.C. and Cross,G.A.M. (1982) *Gene*, **20**, 281-289 (1982).
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Davis,B.D. and Tai,P.C. (1980) *Nature*, **283**, 433-438.
De Lange,T. and Borst,P. (1982) *Nature*, **299**, 451-453.
Dunnick,W., Rabbitts,T.H. and Milstein,C. (1980) *Nature*, **286**, 669-675.
Grab,D.J. and Bwayo,J.J. (1982) *Acta Tropica*, **39**, 363-366.
Hicks,J.B., Strathern,J.N. and Herskowitz,I. (1977) in Bukhari,A.I., Shapiro,J.A. and Adhya,S.I. (eds.), *DNA Insertion Elements, Plasmids and Episomes*, Cold Spring Harbor Laboratory Press, NY, p. 457.
Hoeijmakers,J.H.J., Frasch,A.C.C., Bernards,A., Borst,P. and Cross, G.A.M. (1980) *Nature*, **284**, 78-80.
Le Ray,D. (1975) *Ann. Soc. Belge Med. Trop.*, **55**, 129-311.
Liu,A.Y.C., Van der Ploeg,L.H.T., Rijsewijk,F.A.M. and Borst,P. (1983) *J. Mol. Biol.*, in press.
Majiwa,P.A.O., Young,J.R., Englund,P.T., Shapiro,S.Z. and Williams, R.O. (1982) *Nature*, **297**, 514-516.
Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connel,C., Quon,D., Kee Sim,G. and Efstratiadis,A. (1978) *Cell*, **15**, 687-701.
Matthyssens,G. and Rabbitts,T.H. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6561-6565.
Matthyssens,G., Michiels,F., Hamers,R., Pays,E. and Steinert,M. (1981) *Nature*, **293**, 230-233.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269-276.

Michels,P.A.M., Bernards,A., Van der Ploeg,L.H.T. and Borst,P. (1982) *Nucleic Acids Res.,* **10**, 2353-2366.
Mulvihill,E.R., Le Pennec,J. and Chambon,P. (1982) *Cell,* 621-632.
Nasmyth,K.A., Tatchell,K., Hall,B.D. Astel,C. and Smith,M. (1981) *Nature,* **289**, 244-250.
Pays,E., Delronche,M., Lheureux,M., Vervoort,T., Bloch,J., Gannon,F. and Steinert,M. (1980) *Nculeic Acids Res.,* **8**, 5965-5981.
Pays,E., Van Meirvenne,N., Le Ray,D. and Steinert,M. (1981a) *Proc. Natl. Acad. Sci. USA,* **78**, 2673-2677.
Pays,E., Lheureux,M. and Steinert,M. (1981b) *Nucleic Acids Res.,* **9**, 4225-4238.
Pays,E., Lheureux,M. and Steinert,M. (1981c) *Nature,* **292**, 365-367.
Rice-Ficht,A.C., Chen,K.K. and Donelson,J.E. (1981) *Nature,* **294**, 53-57.
Sakano,H., Huppi,K., Heinrich,G. and Tonegawa,S. (1979) *Nature,* **280**, 288-294.
Sanger,F., Coulson,A.R., Barrell,B.C., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.,* **143**, 161-178.
Strathern,J.N., Klar,A.J.S., Hicks,J.B., Abraham,J.A., Ivy,J.M., Nasmyth, K.A. and McGill,C. (1982) *Cell,* **31**, 183-192.
Tavernier,J., Gheysen,D., Dureinck,F., Van der Heyden,J. and Fiers,W. (1983) *Nature,* **301**, 634-636.
Tiemeier,D., Enquist,L. and Leder,P. (1976) *Nature,* **263**, 526-527.
Van der Ploeg,L.H.T., Liu,A.Y.C., Michels,P.A.M., De Lange,T., Borst, P., Majumder,H.K., Weber,H., Veeneman,G.H. and Van Boom,J. (1982a) *Nucleic Acids Res.,* **10**, 3591-3604.
Van Meirvenne,N, Janssens,P.C. and Magnus,E. (1975) *Ann. Soc. Belge Med. Trop.,* **55**, 1-23.
Williams,R.O., Young,J.R. and Majiwa,P.A.O. (1982) *Nature,* **299**, 417-421.
Young,J.R., Shah,J.S., Matthyssens,G. and Williams,R.O. (1983) *Cell,* in press.
Zieg,J., Hilmen,M. and Simon,M. (1978) *Cell,* **15**, 237-244.