Volume 15 Number 24 1987 Nucleic Acids Research

The structure and transcription of an element interspersed between tandem arrays of mini-exon donor RNA genes in *Trypanosoma brucei*

Mark Carrington*, Isabel Roditi and Richard O.Williams

Kernforschungszentrum, Institut fuer Genetik, Postfach 3640, 75 Karlsruhe, FRG

Received October 7, 1987; Revised and Accepted November 17, 1987

ABSTRACT

Messenger RNAs in *Trypanosoma brucei brucei* have the same 35 bases at the 5' end. These 35 bases are not encoded contiguously in the genome, but are donated from a 140 base RNA (the mini-exon donor RNA). The mini-exon donor RNA (medRNA) is encoded by 1.35 kbp genes that occur in tandem repeats. A DNA element that is associated with mini-exon donor RNA medRNA genes has been identified and characterised by restriction enzyme mapping and partial sequencing. This element (the medRNA gene associated element: MAE) varies in length between 5.5 and 7 kbp. There are between 20 and 40 copies of the element per haploid genome. In clones of genomic DNA MAEs occured between two medRNA gene arrays and on both sides of a medRNA gene array. The MAEs were always in the same orientation with respect to the medRNA genes. It is proposed that in the genome MAEs are interspersed between tandem arrays of medRNA genes. The transcription of the element has been investigated. Low levels of a 70-80 base transcript derived from a small part of MAE were detected in steady state RNA. Nuclear run off transcript studies indicated that MAEs were transcribed at high levels and that they possibly contain at least one start and stop of transcription.

INTRODUCTION.

The causal agent of sleeping sickness of cattle in large parts of Africa is the protozoan parasite *Trypanosoma brucei brucei*. This species, other trypanosomes and a range of other protozoan parasites, such as Leishmania, belong to the order Kinetoplastida.

The structure of mRNA in members of this order has an unusual feature; in any one species the majority (if not all) of the mRNAs have the same sequence at the extreme 5' end. In *T. b. brucei* (1), *T. cruzi* (2, 3) and *Leishmania enriettii* (4) this sequence is 35 bases long. The actual sequence varies between species. Homologous sequences have been identified and shown to be transcribed in a range of other Kinetoplastida (2, 3). The 35 bases is not encoded contiguously with the rest of the mRNA in the genome and is thus termed the mini-exon (5). The transcription of any one mRNA is discontinuous (6, 3), the capped mini-exon (7, 8) being donated in a *trans* splicing reaction to the mRNA precursor from the 5' end of the 140 base mini-exon donor RNA (the medRNA) (reviewed in 9) via a Y-shaped intermediate (10, 11). The medRNA is encoded within a 1.35 kbp gene (6, 2, 12) that is tandemly repeated (6). There are estimated to be approximately 200 copies of this gene per nucleus, occurring mostly in tandem arrays of 10 or more (13).

A small proportion of the medRNA genes occur in shorter tandem arrays termed orphons (14). The 3' and 5' flanking regions of one of these orphons have been characterised and are conserved between medRNA gene arrays (15). The junction between the 3' end of the orphon medRNA gene array and the surrounding DNA occured just 20 bp 3' of a mini-exon (in the sense of transcription) (15). Consequently there are at least two sequences, the orphon flanking sequence and the medRNA gene sequence, found close to the 3' end of the mini-exon. As well as being conserved between medRNA gene arrays in *T. b. brucei*, the orphon flanking sequences are conserved between trypanomastid species (15).

The DNA surrounding the medRNA gene arrays is not well characterised. As part of an investigation to characterise any genes associated with these arrays the structure and transcription of this DNA was investigated. In this report it is shown that there is a DNA element (the medRNA gene associated element: MAE) that varies between 5.5 and 7 kbp in length associated with medRNA gene arrays. MAEs are found interspersed between medRNA gene arrays. The orphon medRNA gene array flanking regions previously reported (15) are part of MAEs. Although MAEs are strongly transcribed, only low levels of steady state transcripts are found. The transcript is 70-80 bases in length and is degraded very rapidly.

METHODS.

Isolation of RNA and DNA from Trypanosomes.

RNA was isolated from *T. b. brucei* MIAG 203 bloodstream form (16) by the hot phenol-SDS method (17). DNA was isolated from the same stock as previously described. (18). DNAs prepared in the same manner from *T. b. brucei* STIB 247L (19) and *T. b. gambiense* STIB 386AA (20) were a kind gift of J. Wells (Dept. of Pathology, University of Cambridge).

Gel electrophoresis and blotting.

RNA was denatured with glyoxal prior to electrophoresis through 1.6 % agarose gels (21). After electrophoresis, gels were blotted without further treatment onto Hybond-N (Amersham) following manufacturers' instructions. Restriction enzyme digests of DNA, agarose gel electrophoresis and blotting were performed using standard conditions (22). Hybridisation of the probes was in 4x SSC, 5x Denhardts, 0.1 % SDS, 0.1 % sodium pyrophosphate, 50 µg/ml heparin solution at 65°C, except for the oligonucleotide probe when hybridisation was at 40°C. The conditions of washing are given in the figure legends.

Origin of hybridisation probes.

Two genomic Sau3A fragments containing the mini-exon (ME1.35 and ME1.5) were cloned and sequenced (see results). Fragment ME1.35 contained the medRNA gene (6, 2,12), fragment ME1.5 contained the junction of a medRNA gene with the surrounding DNA (15).

Restriction fragments derived from the two cloned Sau 3A fragments were subcloned in order to produce specific probes for the two different sequences found 3' of the mini-exon (figure 1). An RsaI-Sau3A fragment from ME1.35 was subcloned into the SmaI site of pGEM1

(Promega) to form pME-A. An RsaI-RsaI fragment from ME1.5 was subcloned into the SmaI site of pGEM1 to form pME-B (figure 1).

Probes were synthesised from pME-A and pME-B using T7 or SP6 RNA polymerase (23, 24) and are shown in figure 1. All the probes were single stranded, the senses of the probes are shown in figure 1. pME-B contained no sequence found in ME1.35. There was a small amount of cross-reactivity between pME-A and ME1.5 due to 29 bp of common sequence.

Scanning densitometry.

Hybridisation of probes to Southern blots was quantitated using a Transdyne 2955 scanning densitometer linked to a Apple IIe with integrating software.

Cloning of DNA.

To clone a specific DNA fragment, the relevant restriction enzyme digest was performed and the DNA fragment purified by electroelution after agarose gel electrophoresis (22). The fragment was then ligated into a plasmid vector cut with the appropriate restriction enzyme prior to transformation of *E. coli* . This method was used for cloning individual genomic repetetive elements and subcloning of specific fragments from larger clones.

A genomic library was constructed from a partial Sau 3A digest of *T. b. brucei* MIAG 203 DNA. The partially digested DNA was treated with calf intestinal alkaline phosphatase and then size fractionated through a sucrose gradient (22). The fractions from the gradient containing DNA of 11-22 kbp were pooled and ligated into Eco RI and Bam HI cut λEMBL3 DNA (25). The ligation was packaged *in vitro* using Gigapack packaging mixes (Vector Cloning Systems) following manufacturers' instructions. The library was plated on *E. coli* NM539 (25) for screening without amplification. The insert size in λEMBL3 is between $1.1x10^4$ and $2.2x10^4$ bp and assuming a haploid genome size of $3x10^7$ bp (26) a genome is represented by between 1400 and 2800 plaques. Colonies (27) or plaques (28) were screened by nucleic acid hybridisation.

DNA sequencing.

DNA to be sequenced was subcloned into M13mp18 or M13mp19 (29) and sequenced by dideoxy chain termination (30) using universal primers (N.E. Biolabs 1211 and 1212) or specific oligonucleotide primers (31). Additional subclones were generated by cutting the insert from the plasmid vector using restriction enzymes, treating the digest with Bal 31 nuclease (17), then recloning the shortened fragments in an M13 vector for sequencing.

Restriction enzyme mapping of DNA.

Restriction enzyme maps of clones were determined by partial restriction enzyme digestion (32). The maps were confirmed as far as possible by comparison with complete digests.

S1 Nuclease mapping of RNAs.

Single stranded DNA probes were synthesised using the prime cut probe method (33). The probes were hybridised with total RNA at 50 °C overnight then digested with S1 nuclease (34). The products were analysed using DNA sequencing gels. Total RNA extracted from the seeds of

Canavalia ensiformis (17) was used as the control RNA. The amounts of S1 nuclease used are detailed in the figure legend.

Synthesis of run-off transcripts.

Run-off transcripts were produced using the method of J. Young (I.L.R.A.D., Nairobi, personal communication). A fraction containing nuclei was isolated from 427.01 procyclic form trypanosomes. The trypanosomes were grown in 500 ml of SDM-79 medium (35) to a density of 2 x 10^7/ml, harvested by centrifugation (2000 rpm for 10 minutes in a Sorvall GSA rotor) and washed with 100 ml SDM-79 minus haemin and foetal calf serum (SDM-79*). The washed cells were collected as before and resuspended in 25 ml of ice cold SDM-79*. Immediately 25 ml of ice cold lysis buffer (see below) were added and the solutions mixed by inverting the tube 3-4 times. The lysate was centrifuged at 4000 rpm for 10 minutes in a Sorvall HB-4 rotor at 4°C, and the supernatent discarded. The pellet was resuspended in 25 ml ice cold preparation buffer (see below) and recentrifuged under the same conditions. This wash was repeated 3 more times. After the final wash the pellet was resuspended in storage buffer (see below) at 4 x 10^9 starting cells/ml.

The suspension was used immediately to produce run-off transcripts. 100 µl of this suspension were added to 100 µl transcription buffer (see below) and incubated at 37°C for 7 minutes. 2 µl of 1 mg/ml DNAse I were then added and the incubation continued at 37°C for 5 minutes, then 200 µl of 10 µg/ml proteinase K in 10mM Tris-HCl, 10mM EDTA, 2 % SDS solution pH 8 were added and the incubation continued at 37°C for a further 5 minutes. The mixture was phenol extracted and the unincorporated nucleotides removed by Sephadex G-50 chromotography. The usual incorporation was 2-5 x 10^7 cpm/10^9 starting cells (nuclei).
Solutions: Preparation buffer, 20 mM PIPES-KOH, 15 mM NaCl, 60 mM KCl, 0.5 mM EGTA, 4 mM EDTA, 14 mM mercaptoethanol, 0.15 mM spermine, 0.5 mM spermidine and 0.125 mM phenylmethylsulphonylfluoride (PMSF) pH 7.5. Lysis buffer, 1 % Nonidet NP-40 in preparation buffer. Storage buffer, 20 mM Tris-HCl, 75 mM NaCl, 0.5 mM EDTA, 0.85 mM dithiothreitol, 50 % glycerol and 0.125 mM PMSF pH 7.9. Transcription buffer, 180 mM Tris-HCl, 25 mM NaCl, 4 mM $MgCl_2$, 8 mM $MnCl_2$, 20 mM creatine phosphate, 1.5 mM dithiothreitol, 8 mM ATP, 2 mM CTP, 2 mM GTP, 1U/µl RNAsin (Amersham) pH 7.9 containing 350 µCi of 800 Ci/mmol [α^{32}P]-UTP. When α-amanitin was used the nuclei were pre-incubated for 15 minutes on ice.

RESULTS.

A 1.5 kbp Sau 3A genomic fragment contains the junction between the 3' end of an medRNA gene array and the surrounding DNA.

There is only one site for the restriction enzyme Sau 3A in the medRNA gene (6, 12, 2), consequently a Southern blot of genomic DNA probed with an oligonucleotide containing the mini-exon sequence would be expected to reveal a 1.35 kbp fragment. This was indeed observed, in addition a less intensely hybridising fragment of 1.5 kbp was also found (see below / figure 5).

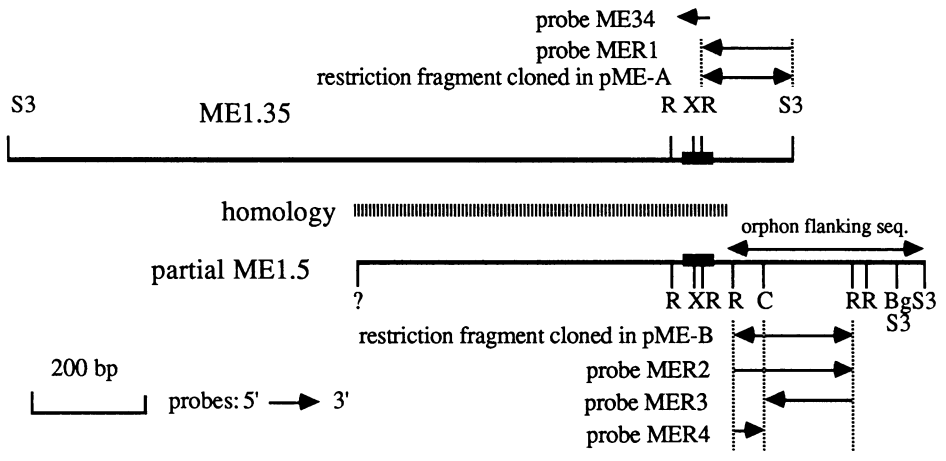


Figure 1. Two sequences occur 3' of the mini-exon. Restriction enzyme maps of, top, a clone of ME 1.35 (a medRNA gene) and, bottom, a partial clone of ME 1.5. The location of the mini-exon is indicated by the thickened line. The maps are drawn so that the direction of transcription of the mini-exon is from left to right. Bg, Bgl II; C, Cla I; R, Rsa I; S3, Sau 3A; X, Xmn I.The ? indicates the position of the lost Sau 3A site in the clones of ME 1.5. Of the 4 clones of ME 1.5 characterised 2 terminated at the Bgl II site and 2 at the Sau 3A site shown. The region common to both ME 1.35 and ME 1.5 is indicated. The locations of the probes used are shown. The probes were all single stranded, the arrow indicates the orientation of each probe 5' to 3'.

DNA corresponding to both these Sau 3A fragments was cloned and the DNA sequence determined. The DNA sequence of the 1.35 kbp fragment (ME1.35) was basically the same as previously published for the medRNA gene (6, 12, 2) with the exception of a few point mutations, insertions and deletions (data not shown). Plasmids containing DNA corresponding to the 1.5 kbp fragment (ME1.5) contained a deletion of approximately 600 bp in four independently isolated clones; the Sau 3A site at one end having been lost (figure 1).

The two DNAs (ME1.35 and the truncated ME1.5) were essentially homologous from the 5' end of the truncated ME1.5 up to 20 bp after the 3' end (in the direction of transcription) of the mini-exon (figure 1). After this point the sequences diverged totally. The sequence unique to ME1.5 was that of the 3' flanking region of an medRNA gene orphon previously reported (15), there being only minor variations from the published sequence (data not shown). Probes specific to each of the two sequences found to the 3' of the mini-exon (one derived from ME1.35 and one from ME1.5, see Materials and Methods) were used to further characterise the structure of the DNA surrounding the medRNA gene arrays.

Probe MER2 detects a sequence repeated many times in the genome that is associated with the mini-exon.

To investigate the DNA surrounding the medRNA gene arrays structures containing the flanking sequences (derived from ME1.5) were characterised. Because of the difficulty in

restriction enzyme mapping non-homogenous repetitive elements directly from genomic DNA, the problem was approached by mapping clones of genomic DNA and confirming the results, as far as possible, using Southern blots of genomic DNA.

A genomic library constructed in λEMBL3 was screened with two probes: MER2 and ME34 (an oligonucleotide 34 bases long, the reverse complemement of bases 2 to 35 of the mini-exon, see figure 1). Probe ME34 was used to detect clones containing mini-exon sequences and probe MER2 to identify those also containing 3' flanking sequences. From 21000 plaques 204 were positive with ME34, of these 181 were also positive with probe MER2. Probe MER2 hybridised with a further 9 plaques that did not hybridise with ME34. This result suggests that probe MER2 detects a sequence repeated many times in the genome since 0.9% of the plaques screened were positive for probe MER2. The result also suggests that the sequence detected by probe MER2 is usually closely associated with the mini-exon as 95% of the plaques positive for MER2 were also positive for probe ME34 which detected the mini-exon.

To obtain a wider view of the relationship between MER2 and the medRNA genes plaques were selected from two categories, first, those positive with both probes MER2 and ME34 and, second, those positive for probe ME34 but negative for probe MER2. Six plaques from the first and two plaques from the second category were used to prepare DNA that was mapped with restriction enzymes.

Probe MER2 detects part of a 5.5-7 kbp element interspersed between medRNA gene arrays.

Figure 2a shows the restriction enzyme maps of four of the eight cloned DNAs. These four represent all the types of structural relationships found. Clones 8.3, 8.4 and 8.5 were positive for both probes ME34 and MER2, clone 8.7 was positive for just probe ME34.

Figure 2b is a summary of the structures found. The probe MER2 detects one end of an 6-7 kbp medRNA gene associated element (MAE), that occurs adjacent to medRNA gene arrays. This element is assigned on the basis of 9 (out of 9) conserved restriction enzyme sites per element. The sequences homologous to MER2 were located from a cluster of three (Xmn I-Cla I-Bgl II) restriction sites mapped in ME1.5 (figure 1). The location and orientation of the medRNA genes are derived from the known restriction enzyme map (6, 12, 2). These results were confirmed by hybridisation of probes ME34, MER1 and MER2 to Southern blots of the cloned DNA and by cross hybridisation of the genomic clones after restriction enzyme digestion and Southern blotting (data not shown).

Different MAEs were always found in the same orientation (based on the order of the restriction sites in MAE) with respect to the direction of transcription of the medRNA genes. Copies of MAE occurred as an apparent tandem pair (clone 8.3), in between two medRNA gene arrays (clone 8.4), or flanking both ends of a medRNA gene array (clone 8.5). Among the 6 clones containing MAE(s) mapped 5 entire MAEs were found. Partial copies of MAEs were only found at the ends of the cloned DNA, truncated, presumably, as a result of the cloning event.

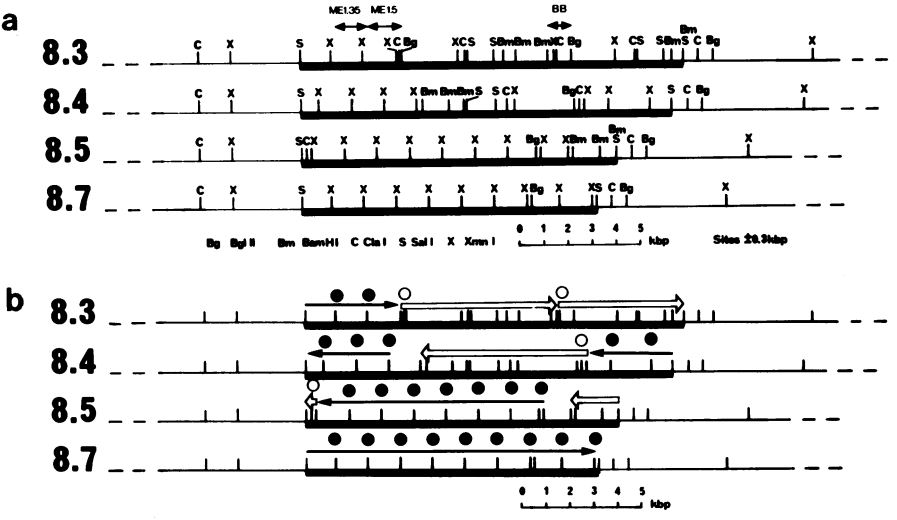


Figure 2. A medRNA gene associated element detected by restriction enzyme mapping.
(a). Restriction enzyme maps of clones of genomic DNA containing the mini-exon. Bg, Bgl II; Bm, Bam HI; C, Cla I; S, Sal I; X, Xmn I. No Eco RI or Hind III sites were found. The cloned DNA is represented by the thick line and the λ EMBL3 vector by the thin line. The Sal I site at the ends of each insert is of vector origin. Example locations of the Sau 3A fragments ME1.35 and ME1.5 are shown above clone 8.3. The Bam HI-Bgl II fragment subsequently sequenced is indicated above clone 8.3 (BB).
(b). A summary of the structural elements above each restriction enzyme map. The solid arrow indicates entire medRNA genes, the direction of the arrow shows the orientation in the direction of transcription of the medRNA. The open arrow indicates the medRNA gene associated element (MAE), the direction of the arrow indicates the orientation but is not an indication of the direction of transcription. shows the location of mini-exon coding sequences as found in ME1.35 and shows mini-exon coding sequences adjacent to sequence homologous to probe MER2. Where the restriction map is not overlined is an indication that the identity of the restriction fragment was not determined.

There was one aspect of the restriction enzyme maps of different MAEs that showed some heterogeneity. This heterogeneity was in the size of the internal Cla I fragment in MAE which varied between 2.75 and 4 kbp. Other restriction fragments such as the two small Bam HI fragments showed no size heterogeneity at all.

Two tandem copies of MAE in clone 8.3 are separated by 49 bp of a medRNA gene.

In clone 8.3 it appeared that there was a copy of the mini-exon in an MAE that was 6.5 kbp distant from the medRNA gene array (figure 2a and 2b). To investigate whether this represented a second type of gene encoding the mini-exon the BamHI-BglII fragment (fragment BB in figure 2a) that contained the mini-exon was subcloned and sequenced. This fragment also spanned the junction of two MAEs. The sequence is shown in figure 3. As predicted there was indeed mini-exon sequence in this fragment, but the mini-exon was incomplete, bases 1-6 being absent. Along

```
 Bam HI/Sau 3A
GGATCCCTCTCCCCCAATCGACCGAGTAGGTCTTCTTTTTTCGGTGGTGC      50

GGGCTCTCCCATAAGCCCGATGGAGAAAATCTCTTTCCATATAGGGC-AT      100

AAAATAATAATAAATAGATAGGATTATCCGGTCCATTAAAGACCACGTAA      150

CATGAAAAAGGTTACACTGCATGTTCCGTGAAAATCGGATGAGGTCTCGG      200
                Sau 3A
AGATCAACAAAGGGTGATCACGTTTAACTGCGGAGGTCGGGGGCA-TTAA      250

AAAAAAAAAAAAAAAAAAATTATTAGAACAGTTTCTGTACTATATTGGTA      300

TGAGAAGCTCCCAGTAGGAATTATCCGTACTTGGGGTCAATTTTCGGGAA      350

GAGAAAAAAGTAAGAAATCGCTGCATTTATGATATCGATAGGAAAGGAGG      400

GAAAACCTCAAAACCAAAAAAAGTCTTGTTTTGGGGGTTCGAACCCGGAC      450

CTCCAAAACACAAACACAATTAGGAGAGGAGTGTTGCCAGTTGGGCTATT      500

TCGACAAATCGGAGGAAAAATTGAAAATTTATGCTCAGATGAAATATACC      550

AAATCGTACACATCGAAGAGAAAAAACATAGGTGTACAAACGGTGCACAC      600

GTAGAAAAAAGTAGCGGAATTGATAGATCT                          630
                          Bgl II/Sau 3A
```

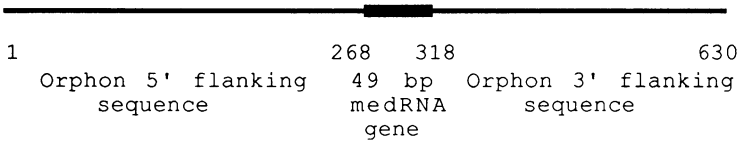


Figure 3. A partial copy of a medRNA gene occurs between two MAEs. The DNA sequence of the Bam HI-Bgl II fragment from clone 8.3 (fragment BB in figure 2) containing sequences homologous to the mini-exon. This fragment occurs 6 kbp distant from a medRNA gene array. The fragment contains bases 7 to 35 of the mini-exon, these are indicated by the box. In total there are 49 bases homologous to a medRNA gene, which are underlined. The remaining sequence is homologous to the 5' and 3' flanking regions of a medRNA gene orphon (15) as shown in the diagram below the sequence. One MAE ends at base 268 and the next starts at base 318.

with this partial mini-exon there were a further 20 bp of the medRNA gene immediately 3' of the mini-exon. A 'pseudo mini-exon' sequence which was also found to be missing bases 1-6 has been reported previously (13).

Comparison of the sequence of this Bam HI-Bgl II fragment with those of the 5' and 3' flanking regions of a medRNA gene orphon (15) revealed that the sequenced fragment was the two flanking sequences in series with just 49 bases of a med RNA gene in between (figure 3). In the sequences presented in figure 3 and reference 15 the 5' and 3' junctions between medRNA genes and surrounding DNA occurred at the same positions in the medRNA gene. In figure 3 one

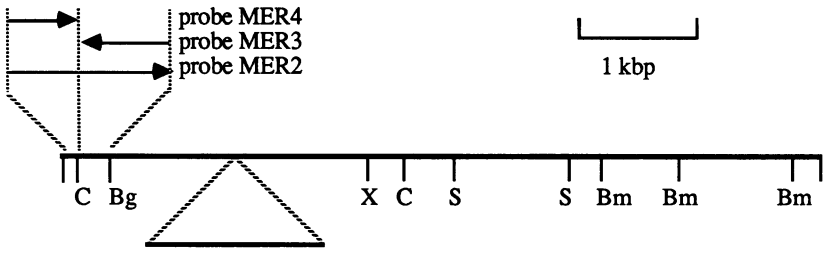


Figure 4. Restriction map of a single MAE. The insert in the internal Cla I fragment represents the variation between 2.75 and 4 kbp in length. In the orientation shown transcription of the adjacent medRNA genes would be from left to right. Bg, Bgl II; Bm, Bam HI; C, Cla I; S, Sal I; X, Xmn I. The location of sequences homologous to probes MER2, 3 and 4 are indicated.

junction occurs after base 318 where the run of adenine residues in MAE are joined to base +6 of the medRNA gene. The other junction is after base 368 where base +49 of a medRNA gene is joined to a MAE. This data concerning the ends of MAEs and the restriction enzyme data shown in figure 2a were used to draw a restriction map of a single MAE (figure 4).

The structures observed in cloned DNA are present in the genome.

Southern blots of genomic DNA were used to confirm that the results obtained from restriction enzyme maps of cloned DNA were representive of the genome. The results are shown in figure 5.

In Sau 3A digested genomic DNA probe ME34 hybridised predominantly to two fragments of 1.35 and 1.5 kbp as described above; when probe MER1 was used the hybridisation to the 1.5 kbp Sau 3A fragment was reduced. Two minor Sau 3A fragments of 0.7 and 2.7 kbp detected by both probes were probably due to slight sequence variation between individual medRNA genes causing the addition or loss of Sau 3A sites (M.C., I.R. and R.W., unpublished data).

Probe MER2 strongly hybridised to the 1.5 kbp Sau 3A fragment from which it was derived. There was weaker hybridisation to a range of smaller fragments, the most intense of these being a 400 bp fragment. This is the same size as the Sau 3A fragment from the junction of two MAE elements in the Bam HI-Bgl II fragment sequenced. The location of these Sau 3A sites is shown on figure 3. Surprisingly probe ME34 did not hybridise to this Sau 3A fragment, possibly because it contained only a partial copy of the mini-exon. The origin of the other minor fragments that showed hybridisation to probe MER2 remain unknown.

Genomic DNA was digested with Hind III or Cla I to compare the restiction fragments found in cloned MAEs with those found in genomic DNA. Probe MER2 hybridised only with high molecular weight Hind III fragments of genomic DNA, consistent with the absence of Hind III sites in any of the cloned MAE containing DNAs.

There are two Cla I sites in MAE, one is spanned by probe MER2 (figure 4). In order to simplify the interpretation of results probes MER3 and MER4 were used (figure 1).

In the restriction enzyme maps of cloned MAEs probe MER3 is homologous to part of an

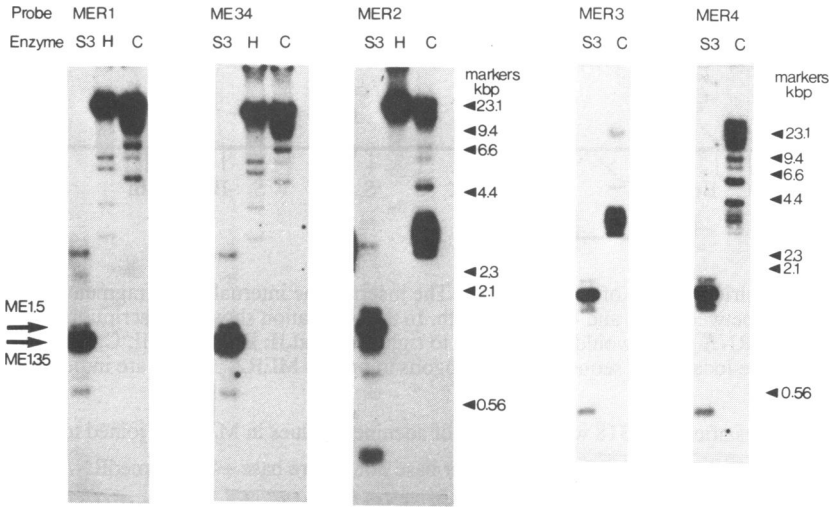


Figure 5. Restriction fragments predicted from cloned DNA are found in genomic DNA. Southern blots of genomic DNA from *T. b. brucei* MIAG203. 1 µg of DNA was loaded in each track after digestion with Sau 3A (S3) Hind III (H) or Cla I (C). The probes are detailed in figure 1. After hybridisation all blots were washed in 0.1 x SSC, 0.1 % SDS solution at 60°C for 2 hours and exposed for 8 hours (MER1 MER2 and MER3), 16 hours (ME34) or 36 hours (MER4).For each blot the next size standards are to the right are applicable.

(MAE) internal Cla I fragment that varies in size from 2.75 to 4 kbp. Probe MER4 is homologous to the extreme end of an MAE (figure 4). Consequently the size of the probe MER4 homologous genomic Cla I fragment is determined by the Cla I site in the adjacent DNA. In clones 8.4 and 8.5 it occurred next to tandem arrays of medRNA genes which contain no Cla I sites, an equivalent location in the genome would result mainly in large genomic Cla I fragments unresolved on a Southern blot. In clone 8.3 probe MER4 lies in a 4 kbp Cla I fragment that arose when there were two MAEs separated by 49 bp of a medRNA gene.

When genomic DNA was digested with Cla I and Southern blotted, probes MER3 and MER4 hybridised with a range of fragments (figure 5). Probe MER3 hybridised to a set of discrete fragments of between 2.75 and 4 kbp (more clearly visible in figure 6). This result is consistant with the size heterogeneity observed in this Cla I fragment in cloned DNA. Probe MER4 hybridised to large unresolved fragments, a 4 kbp fragment and a 5.5 kbp fragment. The first is consistent with MAE occurring next to a medRNA gene array, and the second with the Cla I fragment arising from two MAEs with 49 bp of one medRNA gene in between. The identity of the 5.5 kbp fragment remains obscure.

The results from Southern blots of genomic DNA were in agreement with those obtained by restriction enzyme mapping of cloned DNA. The Southern blot data yielded some genomic

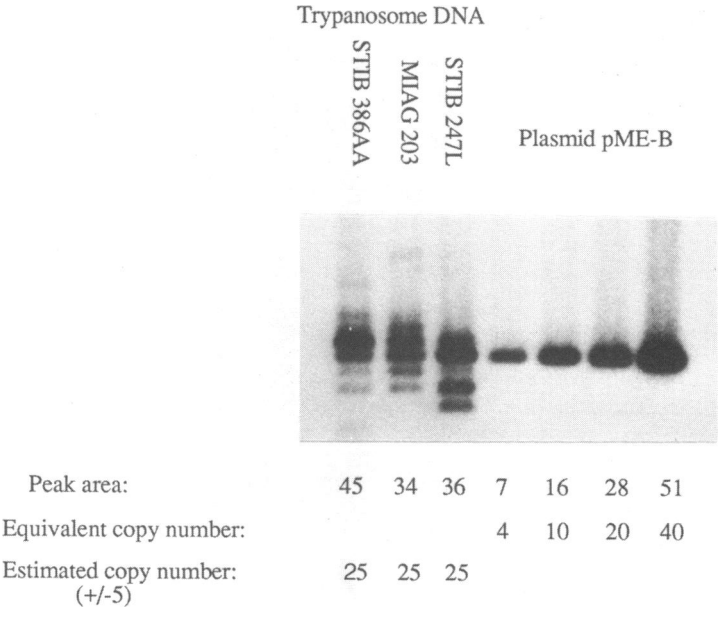


Figure 6. Estimation of MAE copy number. The hybridisation of probe MER3 to Cla I cut trypanosome DNA was quantitated by scanning densitometry and calibrated by comparison with the hybridisation to known amounts of Cla I cut pME-B (see figure 1). The Southern blot was washed with 0.1 x SSC, 0.1% SDS solution at 60°C.

restriction enzyme fragments that were not analagous to any in the cloned DNA. This was to be expected as a fraction of the copies of MAE in the genome were characterised.

Copy number of MAE.

To estimate the copy number of MAE the hybridisation of probe MER3 to Cla I digested genomic DNA from two *T. b. brucei* and one *T. b. gambiense* isolates was quantitated (figure 6). The hybridisation was measured by scanning densitometry of an autoradiograph of a Southern blot on which there also were known amounts of Cla I digested plasmid pME-B.The two *T. b. brucei* isolates (MIAG 203 and STIB 247L) were estimated to contain approximately 25 copies of MAE per haploid genome. The estimation of copy number was based on a genome of 3×10^7 bp for *T. b. brucei* (26). The *T.b. gambiense* isolate was estimated to also contain approximately 25 copies, despite the larger peak area, as the genome is approximately 70 % the size of the *T. b. brucei* genome (E. Pays, personal communication).

Transcription of MAE.

Transcription of MAE was investigated by two methods. First, Northern blotting and S1 mapping were used to identify transcripts in steady state RNA. Second, run off transcripts were used to probe cloned DNA to identify regions being transcribed. Identical results were obtained from both bloodstream form and procyclic trypanosomes.

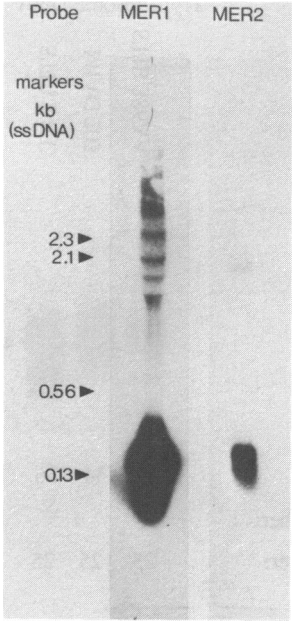


Figure 7. Probe MER2 detects transcripts in steady state RNA. Northern blot of total RNA from *T. b. brucei* MIAG203 bloodstream form trypanosomes. 10 µg of total RNA was loaded in each track. After hybridisation the blots were washed in 0.1 x SSC, 0.1 % SDS solution at 60°C for 2 hours and exposed for 16 hours.

The results of the Northern blots are shown in figure 7. Probe MER1 detected a small RNA, presumably the medRNA (6, 3) as well as processing intermediates of the attachment of the mini-exon to high molecular weight RNA (the high molecular weight smear) (10). Probe MER2 also detected a small RNA in total RNA. The transcript is derived from the sequence immediately adjacent to the associated medRNA gene array (figures 1 and 2a), however transcription is from the other strand of DNA to the medRNA gene array and consequently in the opposite direction. Size estimates of small RNAs from agarose gels are unreliable, accurate estimates of the size of the RNA detected by probe MER2 were obtained by S1 mapping (see below). Probe MER2 also hybridised faintly with an RNA of appproximately 2.0 kbp, the identity of this RNA was not investigated and remains unclear. A single stranded probe of the opposite sense to MER2 detected no transcripts in an equivalent experiment (data not shown).

The relative strength of signal on the two blots is a crude estimate of the relative abundance of the RNAs detected as both probes were of the same specific activity and were used at the same concentrations. Attempts to detect transcripts in steady state RNA using other regions of the MAE were not successful.

A more accurate estimate of the size of the transcripts was obtained by S1 mapping (figure

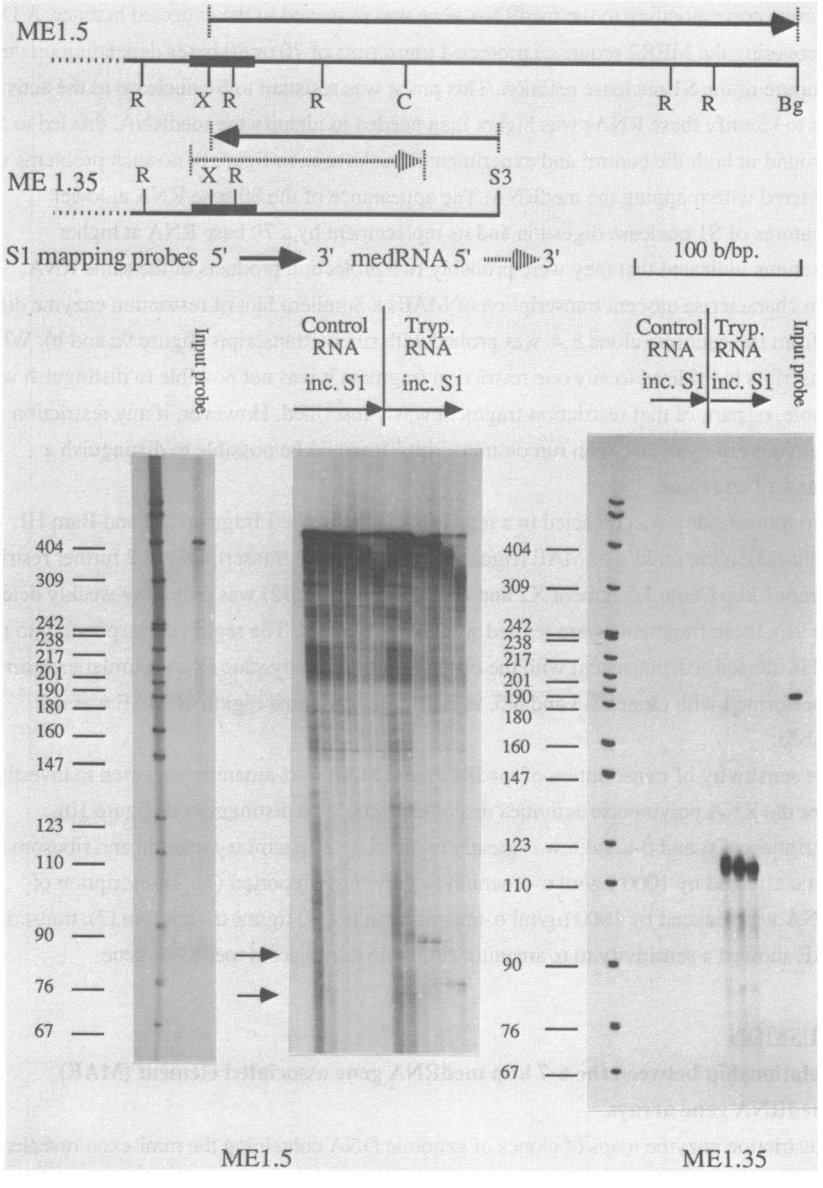


Figure 8. MAE is transcribed to produce a 70-80 base RNA. The probes used for the S1 nuclease protection are shown above and the analysis of the products below. All digestions were for 30 minutes. For the probe derived from ME1.5 six reaction conditions were used: 500 U/ml at 20°C, 1000 U/ml at 20°C, 2000 U/ml at 20°C, 2000 U/ml at 25°C, 2000 U/ml at 30°C, 2000 U/ml at 37°C and 2000 U/ml at 42°C.The autoradiograph was exposed for 60 hours. For the probe derived from ME1.35 three reactions were used: 500 U/ml at 20°C, 1000 U/ml at 20°C and 2000 U/ml at 20°C. The autoradiograph was exposed for 4 hours.

8). A probe corresponding to the medRNA gene was protected in the expected manner. A DNA probe covering the MER2 sequence protected transcripts of 70 or 80 bases depending on the temperature of the S1 nuclease reaction. This probe was resistant to S1 nuclease as the activity needed to identify these RNAs was higher than needed to identify the medRNA, this led to a high background in both the control and experimental tracks seen in figure 8, no such problems were encountered with mapping the medRNA. The appearance of the 80 base RNA at lower temperatures of S1 nuclease digestion and its replacement by a 70 base RNA at higher temperatures indicated that they were probably two protection products of the same RNA.

To characterise nascent transcription of MAEs a Southern blot of restriction enzyme digested DNA from the genomic clone 8.4 was probed with run on transcripts (figure 9a and b). When run on transcripts hybridised to any one restriction fragment it was not possible to distinguish whether the whole, or part, of that restriction fragment was transcribed. However, if any restriction fragment did not hybridise with run on transcripts, it would be possible to distinguish a non-transcribed region.

No transcription was detected in a region of 1.9 kbp (Sal I fragment S2 and Bam HI fragment B3) in the middle of MAE (figure 9b). Furthermore transcription of 2 further restriction fragments 1 kbp (Xmn I fragment X2 and Bam HI fragment B2) was only very weakly detected (figure 9b), these fragments were termed weakly transcribed. The region corresponding to probe MER2 is transcribed, consistent with the observations in steady state RNA. Similar experiments were performed with clones 8.3 and 8.5, in both cases the same region of MAE was not transcibed.

The sensitivity of transcription of medRNA and MAE to α-amanitin was used to investigate whether the RNA polymerase activities responsible could be distinguished (figure 10). Transcription of α and β tubulin was greatly reduced by 20 μg/ml α-amanitin and ribosomal RNA was little affected by 1000 μg/ml α-amanitin as previously reported (7). Transcription of medRNA was reduced by 1000 μg/ml α-amanitin but not 20 μg/ml α-amanitin (7), transcription of MAE showed a sensitivity to α-amanitin similar to the adjacent medRNA gene.

DISCUSSION

The relationship between the 6-7 kbp medRNA gene associated element (MAE) and medRNA gene arrays.

Restriction enzyme maps of clones of genomic DNA containing the mini-exon revealed a medRNA gene associated element (MAE) that varied between 5.5 and 7 kbp in length (figures 2a and 2b).

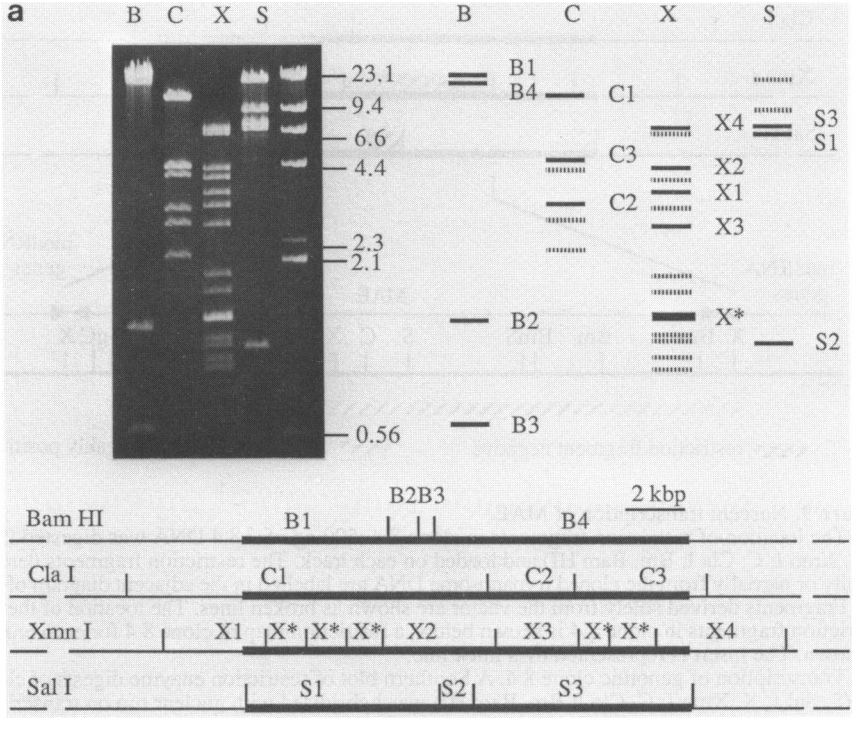
MAEs appeared, in the majority of cases, to be associated with mini-exon coding sequences. In a genomic library, 95% of clones that contained part or all of an MAE also contained the mini-exon. Only 13% of the clones that contained the mini-exon sequence were not positive for probe MER2. Such clones that were analysed contained solely medRNA gene arrays. Since probe

MER2 detected only a small part of one end of an MAE (figure 4) some of these 13% could have contained a partial copy of an MAE that would not be detected by MER2. This data strongly suggests that the medRNA gene arrays and MAEs are usually closely linked. No other medRNA gene flanking sequences were detected in clones containing the mini-exon, and, if existant, must be relatively rare.

The restriction enzyme maps of the genomic clones showed MAE elements situated between two medRNA gene arrays, as well as at both ends of an individual medRNA gene array (figure 2). This data and the estimate of 25 copies (in *T. b. brucei*) of MAE per haploid genome (figure 6) can be combined to form a model of the large scale structure of the medRNA genes. In the genome there is a series of tandem arrays of medRNA genes interspersed with MAE elements to form a super array. Whether these super arrays occur at one or more loci remains unknown. However, as there are only 200 copies of the medRNA gene per nucleus (2), the number of possible loci is limited.

The origin of medRNA gene orphons.

The detection of orphons of medRNA genes by Southern blotting would arise when restriction enzymes that cut in MAEs but not in the intervening medRNA genes were used. An



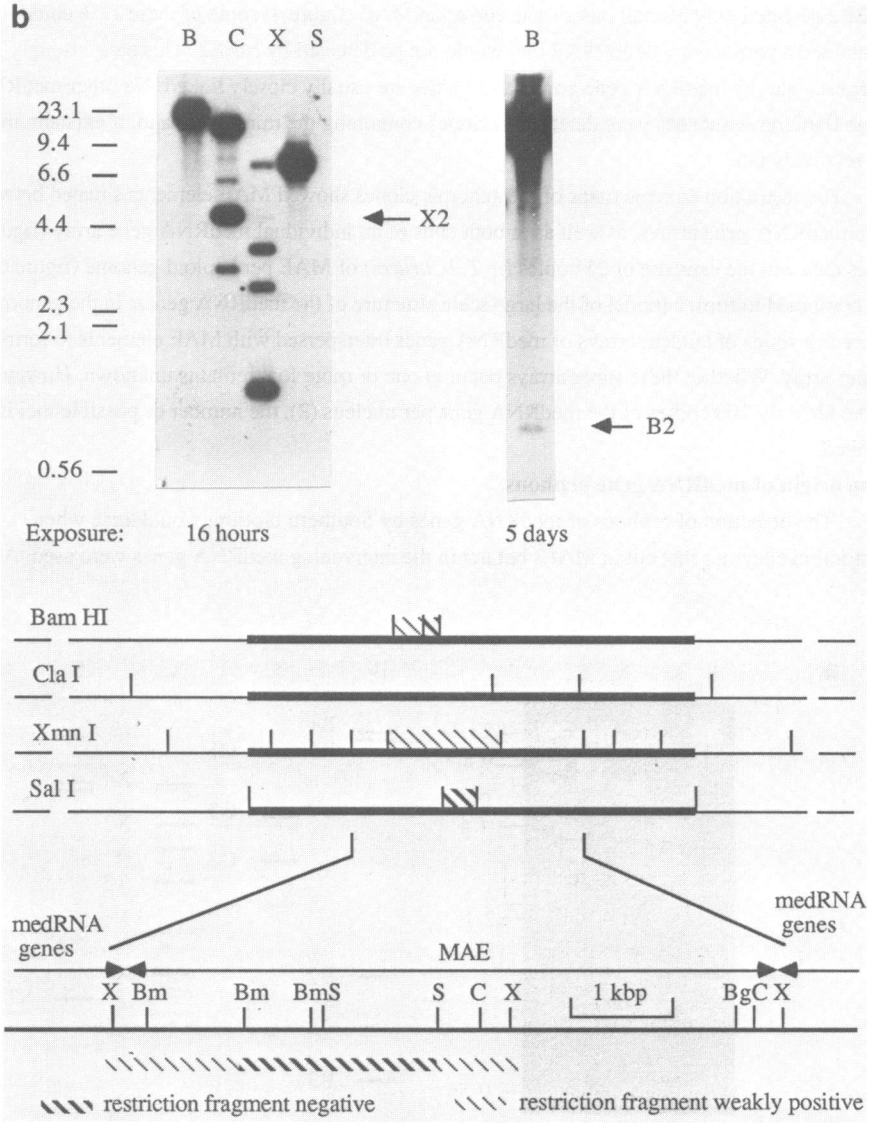


Figure 9. Nascent transcription of MAE.
(a). The location of restriction fragments in clone 8.4. 500 ng of λ8.4 DNA was digested (S, Sal I; X, Xmn I; C, Cla I; Bm, Bam HI) and loaded on each track. The restriction fragments derived wholly or partially from the cloned trypanosome DNA are labelled in the adjacent diagram of the gel. Fragments derived solely from the vector are shown as broken lines. The location of the restriction fragments in clone 8.4 is shown below, a restriction map of clone 8.4 for each enzyme is shown. The insert is represented by a thick line.
(b). Transcription of genomic clone 8.4. A Southern blot of restriction enzyme digests of clone 8.4 (S, Sal I; X, Xmn I; C, Cla I; Bm, Bam HI) was hybridised with nuclear run on transcripts.

250 ng of clone 8.4 DNA was digested and loaded on each track. After hybridisation with run on transcripts the blot was washed in 0.2 x SSC, 0.1 % SDS solution at 60°C and exposed for 16 hours. After an exposure of 5 days the Bam HI fragment of B2 was weakly positive. The non- and weakly transcribed restriction fragments are shown below on the individual enzyme restriction maps of clone 8.4 and below that in an enlarged restiction map of the mae in clone 8.4.

orphon would then result from a short medRNA gene array. This means that the orphon medRNA gene array may be separated from other medRNA gene arrays solely by a single MAE and, thus, is not necessarily at a distant locus.

This explanation is complicated by sequence variation in medRNA genes giving rise to restriction site polymorphism. To illustrate this genomic DNA was cut with Hind III and probed with ME34, MER1 and MER2 (figure 5). Hind III was used as no sites in an MAE or a medRNA gene have been reported. There are, however, 7 locations in a medRNA gene (the sequence from reference 12 was used) where a point mutation would lead to the creation of a previously absent Hind III site. Probes MER1 and ME34 detected four orphons between 3.5 and 7 kbp, which did not hybridise to probe MER2 (the immediate flanking region at one end of a medRNA gene array). The most probable explanation is that the Hind III site at one end, or both ends, of these orphons has arisen in a medRNA gene by a point mutation. Orphons could arise due to restriction sites in adjacent MAEs and/or rare restriction site polymorphisms in medRNA genes.

Contraction of medRNA gene arrays ?

Clone 8.3 contained two MAE elements in tandem with just 49 bp of a medRNA gene in between. The clone gave rise to characteristic Cla I and Sau 3A fragments that were also found in genomic DNA. The structure possibly arose after contraction of a medRNA gene array which could have occurred as the result of a crossing over event. The intensity of hybridisation of probe MER2 to the Sau 3A genomic fragment corresponding to this 49 bp of the medRNA gene (figure 5) suggest this structure occurs several times in the genome, so the contraction does not appear to have been an isolated event. If some arrays were contracting others must be expanding to maintain the number of medRNA genes. The expansion and contraction of the medRNA gene arrays could be involved in maintaining the same DNA sequence in different medRNA genes. Periodically, as part of this process, a single medRNA gene would be expanded to an array. The array would be eliminated before individuals diverged, as a result of random mutations, beyond functional constraints. The hypothesis would predict that the individual members of one array would be more similar to one another than to members of a different array, this is currently under investigation.

The transcription of MAE.

There is a marked contrast between the steady state and nascent levels of the short transcripts derived from MAEs. Allowing for the respective gene copy numbers, the rate of transcription of individual MAEs would appear to be comparable to that of an individual medRNA gene and probably, based on the sensitivity to α-amanitin, by the same RNA polymerase. However, the levels of MAE transcripts in steady state RNA, when detectable, are far lower than that of the

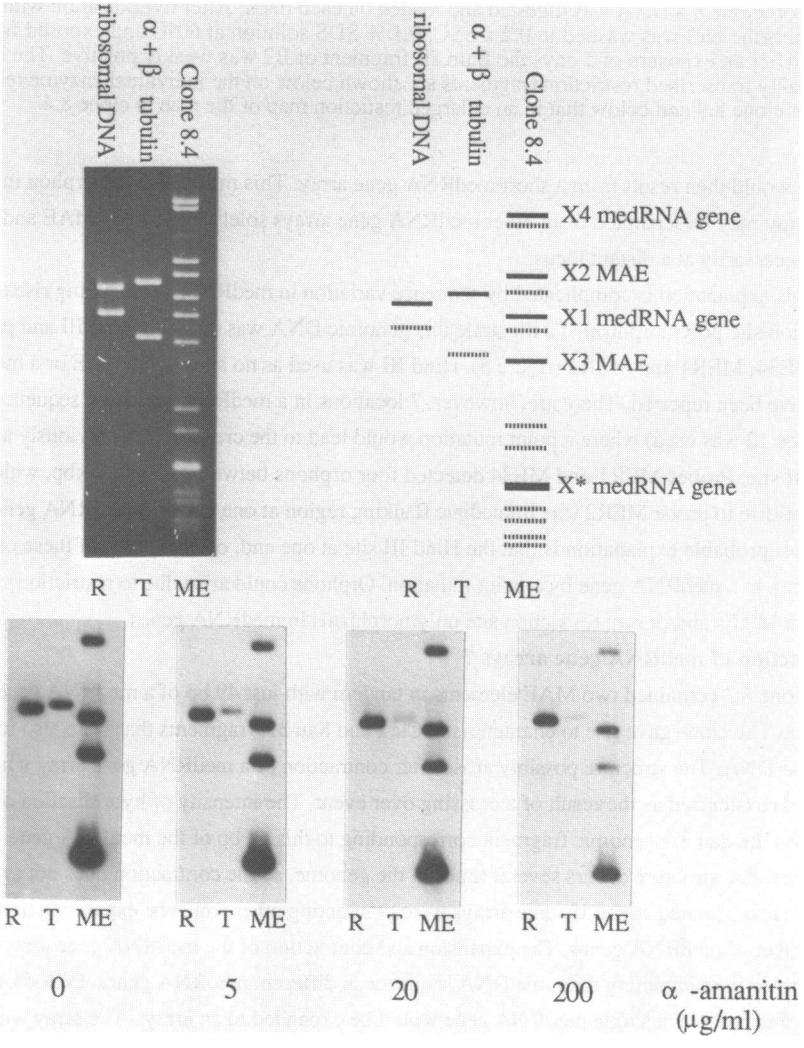


Figure 10. The sensitivity of transcription of MAE to α-amanitin. One of the gels blotted in this experiment is shown stained with ethidium bromide. The origin of the restriction fragments is shown to the right. Fragments derived solely from the vectors are shown as broken lines. The ribosomal DNA was pGH331 (37). The tubulin DNA was a clone of genomic DNA containing both the α and β tubulin genes (unpublished data). Clone 8.4 was used as a source of MAE and medRNA genes (see figure 9a). After hybridisation with run on transcripts the blots were washed in 0.2 x SSC, 0.1 % SDS solution at 60°C and exposed for 36 hours.

medRNA. The nascent transcripts from MAEs must be degraded extremely rapidly, since the half life of medRNA itself is 6 minutes (36).

There is a region of 1.9 kbp in the middle of MAE that is not transcribed. As the MER2

homologous RNA is transcribed away from this non-transcribed region there are probably promotors of and site(s) for the initiation of transcription. Furthermore there are probably also sequences where transcription terminates. The only other such sites definitely recognised in trypanosomes are in the medRNA genes (7, 36) and, more recently for varible surface antigen genes (38, 39). More detailed analysis of the relevant regions of MAE will allow definition of these sites. The function of MAEs and the transcripts derived from them remains unclear. However, the transcription of a region of MAE adjacent to medRNA genes from the opposite strand to the medRNA coding sequence, and the rapid turnover of the products are intriguing clues.

ACKNOWLEDGEMENTS.

We would like to thank Dr Gaiti Hasan for the generous gift of pGH331, and Dr Jerry Wells for the generous gift of trypanosome DNA. This work was undertaken as part of a long term EMBO fellowship held by M.C. and was funded in part by a grant from Deutsche Gesellschaft fuer Technische Zusammenarbeit and the DGTZ supplied as Special Project Funding (85.7860.1-01.100) to the IGT supplemently to the core research grant of ILRAD Nairobi.

*To whom correspondence should be addressed at: Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

REFERENCES.

1. De Lange,T., Michels, P.A.M., Veerman, H.J.G., Cornelissen, A.W.C.A., and Borst, P. (1984). Nucl. Acids Res. **12**, 3777-3790.
2. De Lange,T., Berkvens, T.M., Veerman, H.J.G., Frasch, A.C.A., Barry, J.D., and Borst, P. (1984). Nucl. Acids Res. **12**, 4431-4443.
3. Milhausen, M., Nelson, R.G., Sather, S., Selkirk, M., and Agabian, N. (1984). Cell **38**, 721-729.
4. Miller, S.I., Landfear, S.M., and Wirth, D.F. (1986). Nucl. Acids Res. **14**, 7341-7360.
5. 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. (1982). Nucl. Acids Res. **12**, 3591-3604.
6. Cambell, D.A., Thornton, D.A. and Boothroyd, J.C. (1984). Nature **311**, 350-355.
7. Laird, P.W., Kooter, J.M., Loosbroek, N. and Borst, P. (1985). Nucl. Acids Res. **13**, 4253-4266.
8. Lenardo, M.J., Dorfman, D.M. and Donelson, J.E. (1985). Mol. Cell. Biol. **9**, 2487-2490.
9. Borst, P. (1986). Ann. Rev. Biochem. **55**, 701-732.
10. Sutton, R.E. and Boothroyd, J.C. (1986). Cell **47**, 527-535.
11. Murphy, W.J., Watkins, K.P. and Agabian, N. (1986). Cell **47**, 517-525.
12. Dorfman, D.M. and Donelson, J.E. (1984). Nucl. Acids Res. **12**, 4907-4920.
13. De Lange, T., Liu, A.Y.C., Van der Ploeg, L.H.T., Borst, P., Tromp, M.C. and Van Boom, J.H. (1983). Cell **34**, 891-890.
14. Nelson, R.G., Parsons, M., Barr, P.J., Stuart, K., Selkirk, M. and Agabian, N. (1983). Cell **34**, 901-909.
15. Parsons, M., Nelson, R.G. and Agabian, N. (1986). Nucl. Acids Res. **14**, 1703-1718.
16. Miller, E.N., and Turner, M.J. (1981). Parasitology **82**, 63-80.
17. Carrington, M., Auffret, A. and Hanke, D. (1985). Nature **313**, 64-67.

18. Young, J.R., Donelson, J.E., Majiwa, P.A.O., Shapiro, S.Z. and Williams, R.O. (1982). Nucl. Acids Res. **10**, 803-819.
19. Geigy, R. and Kauffmann, M. (1973) Acta Trop. **30**, 12-48.
20. Felgner, P. *et al.* (1981) Tropen. Med. Parasit. **32,** 134-140.
21. McMaster, G.K. and Carmicheal, G.G. (1977). Proc. Natl. Acad. Sci. USA **74**, 4835-4838.
22. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbour, New York.
23. Nielsen, D.A. and Shapiro, D.J. (1986). Nucl. Acids Res. **14**, 5936 only.
24. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984). Nucl. Acids Res. **12**, 7035-7056.
25. Frischauf, A-M., Lehrach, H., Ponstka, A. and Murray, N. (1983). J. Mol. Biol. **170**, 827-842.
26. Borst, P., Fase-Fowler, F., Frasch, A.C.C., Hoijmakers, J.H.J. and Weijers, P.J. (1980). Mol. Biochem. Parasitol. **1**, 221-246.
27. Grunstein, M and Wallis, J. (1979). Meth. Enzymol. **68**, 379-388.
28. Grunstein, M. and Hogness, D. (1975) Proc. Natl. Acad. Sci. USA **72**, 3961-3965.
29. Vieira, J. and Messing, J. (1982). Gene **19**, 259-268.
30. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
31. Srauss, E.C., Kobori, J.A., Siu, G., Hood, L.E. (1986). Anal. Biochem. **154**, 353-360.
32. Rackwitz, H-R., Zehetner, G., Frischauf, A-M. and Lehrach, H. (1984). Gene **30**, 195-200.
33. Biggin, M., Farrell, P. and Barrell, B. (1984). EMBO J. **3**, 1083-1090.
34. Ward, E.S. and Ellar, D.J. (1986). J. Mol. Biol. **191**, 1-11.
35. Bruen, M. and Schoenenberger, M. (1979). Acta Tropica **36**, 289-292.
36. Laird, P.W., Zomerdijk, J.C.B.M., de Korte, D. and Borst, P. (1987). EMBO J. **6** 1055-1062.
37. Hasan, G., Turner, M.J. and Cordingley, J.S. (1982). Nucl. Acid Res. **10**, 6747-6761.
38. Shea, C., Lee M.G-S. and Van der Ploeg, L.H.T. (1987). Cell **50** 603-612.
39. Kooter, J.M., van der Spek, H.J., Wagter, R., d'Oliviera, C.E., van der Hoeven F., Johnson, P.J. and Borst, P. (1987). Cell **51** 261-272.