33. Kataoka, T., Niakido, T., Miyata, T., Moriwaki, K. & Honjo, T. *J. biol. Chem.* **257**, 277–285 (1982).
34. Clarke, C. *et al. Nucleic Acids Res.* **10**, 7731–7749 (1982).
35. Rice, D. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7862–7865 (1982).
36. Oi, V., Morrison, S., Herzenberg, L. & Berg, P. *Proc. natn. Acad. Sci. U.S.A.* **80**, 825–829 (1983).
37. Gillies, S., Morrison, S., Oi, V. & Tonegawa, S. *Cell* **33**, 717–728 (1983).
38. Queen, C. & Baltimore, D. *Cell* **33**, 741–748 (1983).
39. Banerji, J., Olson, L. & Schaffner, W. *Cell* **33**, 729–740 (1983).
40. Perry, R. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 1937–1941 (1980).
41. Kemp, D. J., Harris, A. W. & Adams, J. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7400–7404 (1980).
42. Walker, I. D. & Harris, A. W. *Nature* **288**, 290–293 (1980).
43. Silhavy, T. S., Benson, S. A. & Emr, S. D. *Microbiol. Rev.* **47**, 313–344 (1983).
44. Alt, F. W., Enea, V., Bothwell, A. L. M. & Baltimore, D. *Cell* **21**, 1–12 (1980).
45. Hood, L. E., Hunkapiller, T. & Kraig, E. in *Modern Cell Biology* (ed. McIntosh, J. R.) 305–328 (Liss, New York, 1983).
46. Williams, A. F. *Nature* **308**, 108–109 (1984).

47. Messing, J. & Vieira, J. *Gene* **19**, 269–276 (1982).
48. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. *J. molec. Biol.* **143**, 161–178 (1980).
49. Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. *Cell* **25**, 59–66 (1981).
50. Ollo, R., Auffray, C., Sikorav, J.-L. & Rougeon, F. *Nucleic Acids Res.* **9**, 4099–4108 (1981).
51. Givol, D., Zakut, R., Effron, K., Rechavi, G., Ram, D. & Cohen, J. B. *Nature* **292**, 426–430 (1981).
52. Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. & Baltimore, D. *Cell* **24**, 625–637 (1981).
53. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
54. Caccia, N. *et al. Cell* **37**, 1091–1099 (1984).
55. Steinmetz, M. *et al. Nature* **300**, 35–42 (1982).
56. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294–5299 (1979).
57. Kronenberg, M., Davis, M. M., Early, P. W., Hood, L. E. & Watson, J. D. *J. exp. Med.* **152**, 1745–1761 (1980).
58. Kavaler, J., Davis, M. M. & Chien, Y. *Nature* **310**, 421–424 (1984).

Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes

David A. Campbell, Deborah A. Thornton & John C. Boothroyd

Department of Medical Microbiology, D-305 Fairchild Building, Stanford University School of Medicine, Stanford, California 94305, USA

The repeated mini-exon sequence that encodes the first 35 base pairs of all variant surface antigen mRNAs of Trypanosoma brucei *directs the synthesis of a discrete 137-nucleotide transcript. It thus seems that variant surface antigen mRNAs are transcribed discontinuously, and we present two alternative models for how this might occur.*

THE ability of African trypanosomes to establish chronic infections in their mammalian hosts depends on a highly developed system of antigenic variation, whereby individual members of the parasite population change the composition of their surface coat[1–3]. This antigenic variation is controlled at the level of gene expression[4,5], each trypanosome possessing a large repertoire of genes (estimated at over 100; refs 6, 7) coding for the antigenically distinct variant surface glycoproteins (VSGs) which comprise the coat. At any one time and on any one trypanosome, only one species of VSG can be detected[8], suggesting that mutual exclusion operates between the VSG genes. Activation of VSG genes occurs by a two-step process requiring first the duplication and transposition of a silent basic copy (BC) of the gene into one of a few telomeric expression sites by a process equivalent to gene conversion, then selection of the expression site over others for transcription[9–15]. The extra copy of the gene thus produced is called the expression-linked copy (ELC). The so-called 'non-duplication-activated' VSG genes described by others probably represent genes which had already undergone the first step in activation (that is, gene conversion into an expression site) before the variants were isolated for study. Although the structure of expression sites and the sequences involved in the gene conversion event have been identified for some variants[16–19], little is known about the second step, particularly how activation occurs and how the mutual exclusion operates between the different expression sites.

Recently, a major clue to this problem has come from the finding that the 5′ 35 nucleotides of VSG mRNAs are not contiguously encoded with the protein-coding portion of the gene[13], and that this spliced leader segment is identical for different VSG mRNAs regardless of which telomeric expression site they occupy[14]. The genomic location of the 35-base pair (bp) mini-exon coding for the spliced leader has recently been reported to be a 1.35-kilobase pair (kb) segment tandemly repeated 100–200 times (as one or more clusters) at an unidentified locus in the genome but at least 30 kb upstream of the active VSG gene[20,21]. Two other important observations concerning the spliced leader are that it is detected on many RNA molecules of varying size in blot analyses using RNA from different life stage forms of *Trypanosoma brucei* (ref. 21 and M. Parsons *et al.*, personal communication) and that homologous sequences have been detected in the genomes of related species and genera[22].

Several models have been proposed to explain how a contiguous mRNA is produced from such an unusual arrangement of exons[20–23] and how the activity of different expression sites is regulated. These have included chromosome end exchange with splicing of very long transcripts, and discontinuous transcription whereby the mini-exon and ELC are flanked by their own initiation and termination sites. To determine how VSG mRNAs are produced and processed, we have studied the 1.35-kb repeats and identified their transcriptional products.

Molecular cloning of a mini-exon repeat

To enable nucleotide sequence analysis of the mini-exon repeat, recombinant plasmids containing individual repeats were generated. This was initially done using a *Pvu*II digest of the genomic DNA of *T. brucei* as this enzyme has been reported to release the mini-exon-containing portion of the repeat as a discrete band at ~720 bp[21]. Such a band was excised from an agarose gel, the DNA purified by electroelution[24] and ligated into the *Pvu*II site of the plasmid vector pBR322 (ref. 25). A sample of this same excised material was radiolabelled by nick-translation[26] and used as a probe in colony hybridization[27], thereby identifying recombinants containing repetitive DNA. Four plasmids thus identified were further characterized and three were found to contain indistinguishable inserts with at least one *Rsa*I site. As the mini-exon should include an *Rsa*I site, nucleotide sequence analysis in the vicinity of this site was performed on one of the plasmids (pMEP.1). The strategy and sequence obtained are shown in Figs 1 and 2, respectively. This demonstrated unambiguously that pMEP.1 contains a portion of the mini-exon repeat.

To obtain a recombinant plasmid containing an insert representative of the complete 1.35-kb repeat, a ligation was performed using the 1.35-kb fraction of trypanosome genomic DNA digested with *Sau*3A and the plasmid vector pAT153 (ref. 28) digested with *Bam*HI. Recombinants were screened with the
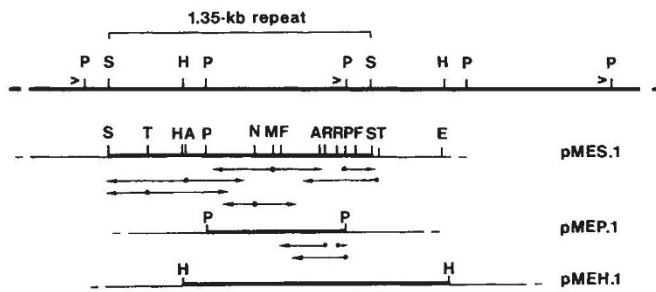
Fig. 1 Restriction maps and sequencing strategies for mini-exon-containing plasmids. The top line of the figure represents a restriction map for the genomic cluster of 1.35-kb tandem mini-exon repeats as deduced from Southern blot analysis (data not shown). The rightward-pointing arrowheads represent the approximate position of the 35-bp mini-exon sequence; beneath this are the recombinant-containing portions of this genomic region as cloned in the plasmid vectors pAT153 (pMES.1 and pMEH.1) and pBR322 (pMEP.1; see text). The restriction sites used in the nucleotide sequencing (by the chemical modification method of Maxam and Gilbert[43]) are shown for each insert. The arrows with solid circles indicate the labelled sites and the region over which unambiguous sequence information was obtained. Restriction sites: A, *Acc*I; E, *Eco*RI; F, *Hinf*I; H, *Sph*I; M, *Mlu*I; N, *Nar*I; P, *Pvu*II; R, *Rsa*I; S, *Sau*3A; T, *Taq*I.

Fig. 2 Nucleotide sequence of a complete 1.35-kb mini-exon repeat. The nucleotide sequence of the sense strand of pMES.1 is presented with numbering from the 5'-end of the 35-bp mini-exon, which is boxed. The restriction sites used in the cloning (see Fig. 1) are underlined and annotated. The *Sau*3A site (GATC) used in cloning the insert is duplicated at the beginning and end of the sequence. The thickly underlined segment represents the 17-nucleotide sequence for which a complementary oligonucleotide was synthesized and used in subsequent experiments (see Figs 3 and 5). The arrow between positions 137 and 138 indicates the 3'-terminus of mini-exon-derived RNA as described in the text.

*Pvu*II insert of pMEP.1 and after preliminary restriction mapping, one (pMES.1) was characterized further and found to contain an apparently complete 1.35-kb mini-exon repeat.

Nucleotide sequence of mini-exon repeat

Using the recombinant plasmid pMES.1 and the strategy shown in Fig. 1, the complete nucleotide sequence of one repeat unit containing the mini-exon was determined. This is presented in Fig. 2 with position +1 being the first nucleotide of the mini-

exon. The region from −270 to +61 was also sequenced from the independently generated plasmid pMEP.1 (see Fig. 1). The results from the two plasmids were identical in this region.

Although Southern blot analysis suggested that the repeat unit was defined by a single *Sau*3A site (refs 20, 21 and D.A.T. and J.C.B., unpublished results), the possibility remained that two closely adjacent *Sau*3A sites were contained within each repeat and that pMES.1, therefore, lacked a small portion of the repeat. To test this, a further recombinant plasmid containing a mini-exon repeat was generated using an *Sph*I digest of genomic DNA (as cloned in a bacteriophage λ recombinant containing several 1.35-kb repeats; D.A.C., unpublished results) and ligated to *Sph*I-digested pAT153. *Sph*I was known to cut the repeat once (from sequence analysis of pMES.1 and Southern blot analyses; D.A.C., unpublished results) and thus should give a fragment spanning the *Sau*3A site in question. Mini-exon-containing plasmids were identified as above and pMEH.1 was obtained, mapped and sequenced (see Fig. 1). From this plasmid, a contiguous sequence spanning the *Sau*3A site used to construct pMES.1 was obtained (data not shown), demonstrating that the sequence presented in Fig. 2 represents a complete mini-exon repeat.

To ensure that this sequence was not only complete but also representative, we compared its restriction map with the fragment sizes observed on Southern blots of genomic DNA cut with several enzymes, using the *Pvu*II insert of pMEP.1 as a radiolabelled probe (data not shown). The predominant bands obtained in each digest agreed in all cases with the fragment sizes predicted from the sequence; this demonstrated that the mini-exon repeats are highly conserved and that pMES.1 contains a typical member.
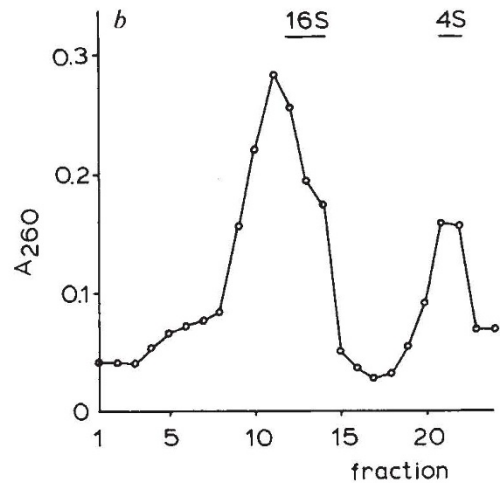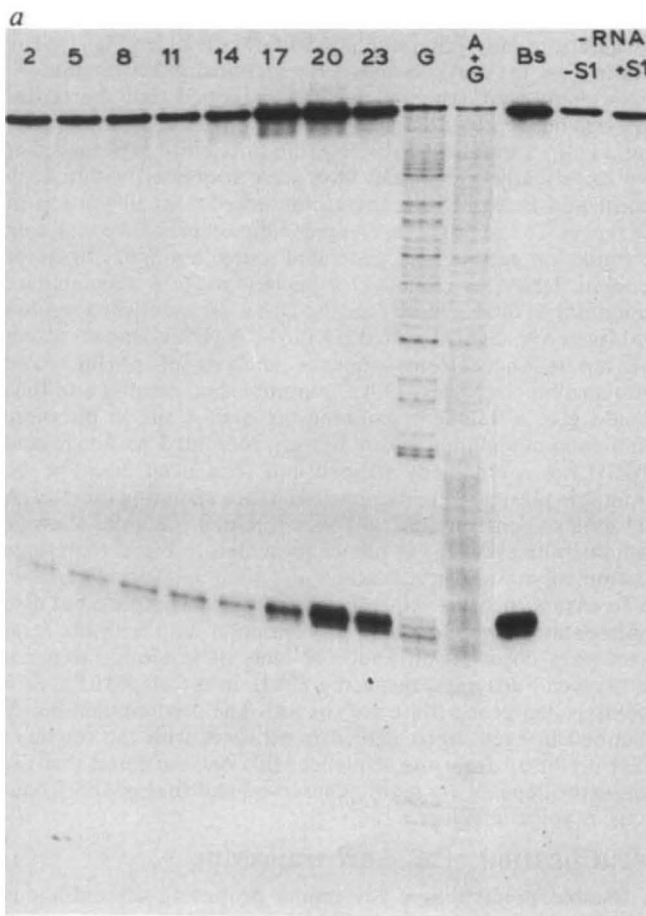
Identification of a short transcript

A testable prediction of the model proposing discontinuous transcription is that there should be detectable levels of a small RNA (that is, <1.35 kb) derived from each mini-exon repeat. One of the most sensitive methods for detecting and characterizing particular transcripts is S₁-protection[29], whereby a radiolabelled DNA probe from the region of interest is denatured and renatured in the presence of RNA; the resulting material is then digested with the single-strand-specific S₁ nuclease and the length of any protected fragments measured by gel electrophoresis and autoradiography. The DNA substrate used here was from pMES.1, extending from the [3'-^{32}P]-labelled *Pvu*II site at position +58, downstream through the *Sau*3A site (position +183), to the *Eco*RI site in the vector (see Fig. 1). This fragment was mixed with total trypanosome bloodstream-form RNA (or sucrose-gradient fractions thereof), precipitated, redissolved, denatured and renatured in conditions favouring RNA/DNA hybridization[30]. This mixture was then cooled, diluted and digested with S₁ nuclease, before electrophoresis on a 7 M urea polyacrylamide gel. The resulting autoradiogram (Fig. 3a) shows a single major end point of protection by total RNA which corresponds to nucleotide +137 in the sequence presented in Fig. 2. Minor bands immediately flanking this protected region may be indicative of slight heterogeneity at the 3' end of the RNA or nonspecific digestion by S₁ nuclease at the end of the RNA/DNA duplex.

This result could be interpreted as protection by a short RNA of 137 nucleotides derived from the mini-exon repeats and/or by a longer transcript with a discontinuity at this position in the RNA/DNA duplex. To discriminate between these two alternatives, two procedures were used. The first examined the S₁-protection by different size fractions of RNA produced by sucrose-gradient centrifugation (Fig. 3b). Maximal protection was found in fractions 19–21, which possessed RNA with a peak size slightly larger than 4S RNA (that is, ~100–200 nucleotides), with diminishing but detectable protection in the remaining fractions (Fig. 3a). This strongly suggested that the protection observed with total RNA was due, at least in part, to the presence of a small RNA. The protection observed with other fractions, particularly those containing larger RNA, may be due

a

b

Fig. 3 *a*, S$_1$ nuclease mapping of the 3'-terminus of medRNA. Lanes 2–23 are reactions performed with RNA from pooled sucrose gradient fractions, as described in *b* below. Next to these are two marker sequencing reactions[43] on the labelled probe. The lane labelled Bs contains the result of a protection experiment with 50 μg of total RNA. The two controls, which contain no added RNA, are with (+S1) or without (−S1) S$_1$-nuclease, respectively. *b*, A$_{260}$ profile of total trypanosome RNA fractionated by sucrose gradient centrifugation. The size markers are based on samples from each fraction analysed by electrophoresis and ethidium bromide staining (not shown).

Methods: *a*, A [3'-^{32}P]-end-labelled *Pvu*II digest of pMES.1 was re-cut with *Eco*RI to generate a uniquely end-labelled fragment corresponding to positions +59 to +183 plus 375 bp of the vector (see Fig. 1). S$_1$ protection was done essentially as described elsewhere[30]. Briefly, the probe was mixed with different fractions of bloodstream-form RNA, precipitated, redissolved and denatured in 80% formamide at 80 °C for 10 min. Renaturation of RNA/DNA hybrids was promoted by incubating at 45 °C for 3 h in the same buffer, before diluting with 10 vol of aqueous buffer, cooling to 30 °C and adding 35 units of S$_1$ nuclease. Digestion was for 30 min, after which the reactions were phenol-extracted, isopropanol-precipitated, resuspended in formamide loading buffer[43] and re-solved by PAGE in a 6% polyacrylamide gel with 7 M urea[44]. *b*, A 15–30% linear sucrose gradient was prepared in 0.1 M NaCl, 1 mM EDTA, 20 mM Tris-HCl (*p*H 7.5), 0.1% SDS. About 1 mg of total trypanosome RNA was overlaid with the same buffer and the samples spun in an SW-41 rotor (Beckman) at 40,000 r.p.m. at 20 °C for 9.5 h. Fractions were collected from the bottom of the tube and the A$_{260}$ measured. For the S$_1$-protection experiments, 5% of each sample was removed and pooled in groups of three (numbered by the middle fraction).

to contamination with the small RNA. Alternatively, they may be a result of protection by hybrid transcripts comprising the small RNA at their 5' ends linked to RNA from, for example, the ELC region, by the mechanisms discussed below.

The second method to detect and size transcripts from the mini-exon repeat used Northern blot analysis[31]. For this procedure, total trypanosome RNA was resolved by polyacrylamide gel electrophoresis (PAGE) in denaturing conditions, transferred to GeneScreen membrane[32] and hybridized with a [5'-^{32}P]-labelled, synthetic oligonucleotide complementary to positions +117 to +133. This 17-mer was chosen because it excluded the mini-exon itself, which was known to give intense hybridization throughout RNA blots when used as a probe[21] and because the S$_1$-protection experiments indicated that it should hybridize to mini-exon-derived transcripts. Figure 4*a* shows the ethidium bromide stain of a polyacrylamide/7 M urea gel in which the major small RNA species is readily apparent. Figure 4*b* presents the autoradiogram produced after hybridizing a GeneScreen transfer of a duplicate half of this gel with the 17-mer; only one band is detected in the track containing trypanosome RNA and no signal is observed in the track containing RNA from the unrelated parasitic protozoan, *Toxoplasma gondii*. The size of the band is ~140 nucleotides based on published values for the small ribosomal RNAs of trypanosomes[33], thus confirming that the mini-exon repeats are transcribed to yield a short, discrete RNA which we have termed mini-exon-derived RNA (medRNA). With regard to its precise size, this result is, within experimental error, consistent with medRNA being 137 nucleotides long. We cannot, however, exclude the possibility that medRNA is as many as 4–5 nucleotides longer and that S$_1$ nuclease nonspecifically digested the 3' end of the RNA/DNA hybrid. Note, however, that the nucleotides at positions 138 and 140 are cytosines which, as CG pairs, would be less susceptible to such action. A definite answer awaits 3'-sequence analysis of medRNA.

These results also suggest that the 5'-terminus of medRNA is at about position +1 in the mini-exon repeat. To confirm this, we used the synthetic oligonucleotide from the RNA blot analyses as a primer for reverse transcriptase extensions of medRNA. The result (Fig. 5) shows a major band co-migrating with a heterologous DNA marker of 135 nucleotides. Using a different heterologous marker (a *Hin*dIII digest of bacteriophage λ DNA), the size of this band was estimated as 133 nucleotides. This slight discrepancy is probably due to differences in base composition. Allowing for this variation, the extension has an approximate length of 134 ± 1 nucleotides. Given that the 5'-end of the primer represents position +133 of the mini-exon repeat, this result confirms that the 5'-end of medRNA corresponds to position +1, the same as that found for the spliced leader of mature VSG mRNAs[13,14]. The minor band in Fig. 5 at 138 ± 1 nucleotides may be an artefact of reverse transcriptase (for example, loop-back synthesis) or it may be indicative of a minor RNA 4 nucleotides longer than the major species. Such minor heterogeneity might be due to variation in the mini-exon repeat sequences and/or in the site of transcriptional initiation. No other major bands were observed in other loadings of this sample run for longer or shorter periods (data not shown).

—————————————————ARTICLES—————————————————
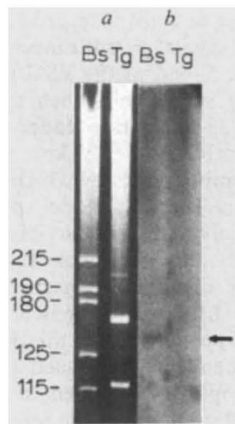
Fig. 4 Detection of the mini-exon-derived RNA. *a*, Ethidium bromide stain of electrophoretically separated trypanosome RNA. Lane Bs contains 10 μg of total RNA from bloodstream-forms of *T. brucei* (strain MITat 1.4). Lane Tg contains 10 μg of RNA from *T. gondii* as a negative control. The major bands in lane Bs are polysome-derived and their sizes in nucleotides are taken from ref. 33. The two major bands in the *T. gondii* RNA are presumably 5S and 5.8S. *b*, Autoradiograph of an RNA blot of the other half of the gel shown in *a*, containing identical samples, and having been probed with [5'-^{32}]-labelled 17-mer complementary to positions +117 to +133 of Fig. 2.

Methods: *a*, Total RNA was separated on an 8% polyacrylamide gel (1.5 mm thick) containing 7 M urea[44]. *b*, The electrophoresed RNA was electroblotted to GeneScreen membrane[32] at 4 V for 18 h at 4 °C in 25 mM Na phosphate (*p*H 6.5). The filter was prehybridized in 50% formamide, 5 × SSC (1 × SSC = 0.15 M NaCl, 0.015 M Na citrate), 1 × Denhardt's, 25 mM Na phosphate (*p*H 6.5), 0.1% SDS, 0.1% Na pyrophosphate, 100 μg ml^{-1} denatured calf thymus DNA at 30 °C for 2 h. A synthetic oligonucleotide complementary to positions +117 to +133 (see Fig. 2) was synthesized by standard means, [5'-^{32}P]-labelled by polynucleotide kinase, added to the prehybridization solution and the incubation continued with gentle shaking for 15 h at 30 °C, followed by 5 h at 22 °C. The filter was washed in four changes of 2 × SSC, 0.5% SDS for 15 min each at 22 °C. Autoradiography was at −70 °C with an intensifying screen for 6 days.

Conclusions

We have cloned and sequenced a complete copy of the 1.35-kb mini-exon repeat of *T. brucei* and identified a short RNA of 137 nucleotides as its transcriptional product. The data indicate that this repeat is a typical member of a highly conserved family—this is demonstrated further by the virtually complete agreement of the sequence presented in Fig. 2 with the partial sequence for the region from −130 to +61 recently published elsewhere[21]; the four differences (three single base changes, one insertion) are all localized to between −115 and −125. As the sequences in both cases were determined from two plasmids, the differences may be indicative of minor, but real, variation in the repeats.

In addition to the region coding for medRNA, each 1.35-kb mini-exon repeat contains about 1,200 nucleotides whose function is not clear; part of this region is presumably involved in the regulation of medRNA expression, but the remainder could be involved in some other aspect of trypanosome gene expression. Of particular note is the repetitive nature of the sequence around the *Sau*3A site, consisting of 1-, 2-, 4- and 6-nucleotide repeats (T, AT/AC, ATTT/GTTT and ACACTC, respectively). It will be interesting to determine whether these regions of the repeat are as exactly conserved as that coding for the medRNA.

The predicted sequence of medRNA can be drawn with several alternative regions of intramolecular base-pairing (data not shown). This suggests that medRNA might adopt a stable conformation with substantial secondary structure, but, as we have no data on RNase sensitivity, we cannot present a reliable structure.

The detection of medRNA raises questions about the signals directing its transcription. As noted by De Lange *et al.*[21], the
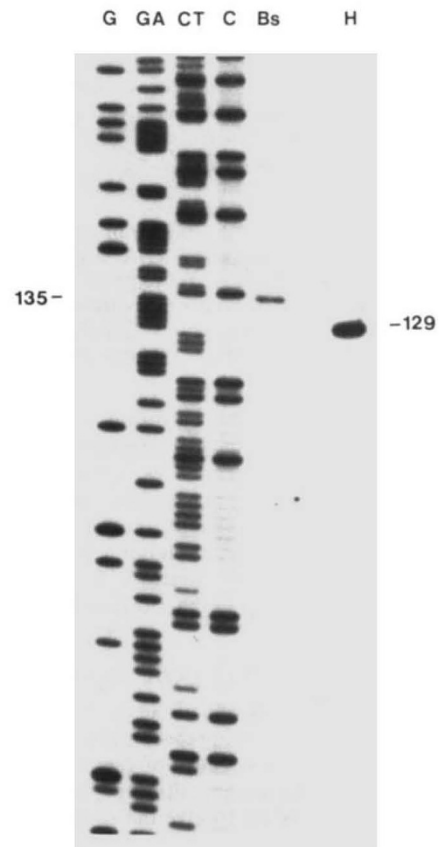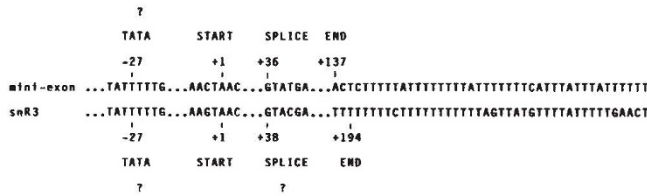

Fig. 5 Identification of the 5'-terminus of the medRNA. Polyacrylamide gel electrophoresis of the reverse transcriptase product primed off medRNA (lane Bs). Marker tracks contained standard sequencing reactions[43] of a mouse cDNA clone (provided by J. Kavaler; lanes G, A + G, C + T and T) or a [3'-^{32}P]-labelled *Hind*III digest of λ DNA (lane H).

Methods: The [5'-^{32}P]-labelled synthetic oligonucleotide described in Fig. 4 legend used as a primer in extension reactions using AMV reverse transcriptase. About 4 μg of primer were hybridized to 20 μg of total bloodstream-form RNA in 1.2 M Na phosphate buffer (*p*H 6.8), 0.5% SDS, 12.5 mM EDTA at 21 °C for 4 h. After ethanol precipitation, the reverse transcriptase reaction was done in 50 mM Tris-HCl (*p*H 8.3), 6 mM MgCl$_2$, 1 mM dithiothreitol, 60 mM KCl, 0.1 mM dATP, dCTP, dTTP and dGTP, and AMV reverse transcriptase (50 units) at 37 °C for 1 h. The sample was ethanol-precipitated, resuspended in formamide loading dye and electrophoresed on a 7 M urea, 8% polyacrylamide gel. The dried gel was autoradiographed at −70 °C for 72 h.

octanucleotide centred at position −27 upstream of the putative initiation site for transcription consists of TATTTTTG, similar to the consensus sequence generally found at this position in eukaryotic RNA polymerase II promoters (the so-called TATA box; see ref. 34). Assuming trypanosomes are usual in their types and functions of RNA polymerase and given that the medRNA ultimately forms the 5'-end of the VSG mRNA, the finding of a putative TATA box at this position is expected. Further studies on relative drug sensitivities and 5'-cap structures are needed to confirm medRNA as an RNA polymerase II transcript. The termination site cannot be compared with the general case as too few sequences of eukaryotic transcriptional terminators have been reported to enable consensus sequences to be made.

Recently, the complete sequence of a small nuclear RNA (snR3) and its gene have been reported in yeast[32,35]. This transcription unit shows remarkable similarity to the mini-exon repeat at four critical positions (Fig. 6), the remainder of the two sequences sharing no significant homologies. First, the snR3 gene also has the octanucleotide TATTTTTG centred at position −27 relative to the start site for transcription. Second, the

```
              ?
           TATA      START    SPLICE   END
           -27        +1       +36     +137
            |          |        |       |
mini-exon  ...TATTTTTG...AACTAAC...GTATGA...ACTCTTTTTATTTTTTTTTATTTTTTTCATTTATTTATTTTT
snR3       ...TATTTTTG...AAGTAAC...GTACGA...TTTTTTTTCTTTTTTTTTTTTAGTTATGTTTTATTTTTGAACT
           -27        +1       +38     +194
            |          |        |       |
           TATA      START    SPLICE   END
              ?                  ?
```

Fig. 6 Nucleotide sequence of the mini-exon repeat of *T. brucei* compared with that of the small nuclear RNA (snR3) gene of yeast (from ref. 32). The two sequences are aligned at the indicated positions with position +1 in both cases corresponding to the 5'-terminus of their respective transcripts. Question marks designate functions which have not yet been confirmed.

initiation site for transcription lies within a heptanucleotide sequence which is very similar in the two genes (six out of seven bases identical). Third, both sequences, although of different overall length (137 versus 194 nucleotides for medRNA and snR3, respectively), end near the beginning of an extremely T-rich region. Finally, the region containing the putative donor splice site of the medRNA also shares a 5- out of 6-residue homology with the yeast snR3, the potential significance of which is discussed below.

Role of medRNA in VSG gene expression

The finding of medRNA fulfils one of the predictions of a model for VSG gene expression involving discontinuous transcription. However, it is not definitive because of the multiplicity of the mini-exon repeats: it could be that, although the majority of the repeats are indeed transcribed to give the short medRNA, one repeat is used in the initiation of a very long transcript including the VSG coding exon. This is a difficult possibility to exclude because pulse-chase experiments, which have been traditionally used to demonstrate precursor/product relationships, suffer from the same shortcoming—it is impossible to demonstrate that a given mini-exon repeat is used for one purpose compared with another, indistinguishable repeat. Nevertheless, together with published observations regarding the abundance of transcripts containing mini-exon-derived sequences and the inability to detect linkage between a mini-exon sequence and the VSG coding exon, the identification of a short, discrete medRNA strongly supports a model of discontinuous transcription. In this model, the mini-exon-derived leader segment of VSG

mRNAs is transcribed as a discrete product of the mini-exon repeats and, by one of several possible mechanisms, is ultimately found attached to the 5'-end of the VSG precursor RNA. The chimaeric intervening sequence is then removed by splicing, bringing together the 35-nucleotide leader with the coding portion to give the final mRNA of ~1.7 kb.

There are two general schemes by which this can most easily be imagined to occur. The first model, presented in Fig. 7a, predicts that RNA polymerase transcribes the medRNA and then dissociates from the mini-exon repeat, possibly as a medRNA/polymerase complex; transcription then reinitiates just upstream of the ELC using medRNA as a primer. The second model (Fig. 7b) proposed that medRNA and the coding-region RNA are independently produced as discrete transcripts and then post-transcriptionally ligated to each other. In both cases, splicing ultimately removes the intervening sequence to generate the mature mRNA.

The critical difference between the two alternatives is that post-transcriptional ligation requires a functional promoter (in the sense of a site for *de novo* initiation of transcription) upstream of the ELC whereas the reinitiation model proposes a site in this same position where RNA polymerase could only act in the presence of a primer (that is, medRNA). We are now attempting to test for this distinction *in vitro*.

We have previously reported the sequence upstream of an expressed copy of a VSG gene and compared it with a silent copy of that gene[19]. We found that the upstream region of the expressed copy consisted of multiple tandem repeats of ~76 bp flanking an unusual sequence of $(TAA)_{90}$. Although three 76-bp repeats are found upstream of the silent copy, $(TAA)_{90}$ is unique to the expressed copy of this gene and is, therefore, a candidate for the (re)initiation site upstream of expressed VSG genes. If this sequence is unique and mobile, it could be the basis for the mutual exclusion observed between the different expression sites. Equally, another, as yet undetected, mobile control element might have this role.

A critical question raised by the results reported here is whether discontinuous transcription involving a small RNA is a unique property of *T. brucei*. Is it, for example, a special adaptation which has co-evolved with antigenic variation? This is argued against by the finding that the same or similar mini-exon sequences are found in many RNAs (of unknown coding function), not only from bloodstream-forms of the parasite, but also from the procyclic insect forms where VSG genes are not
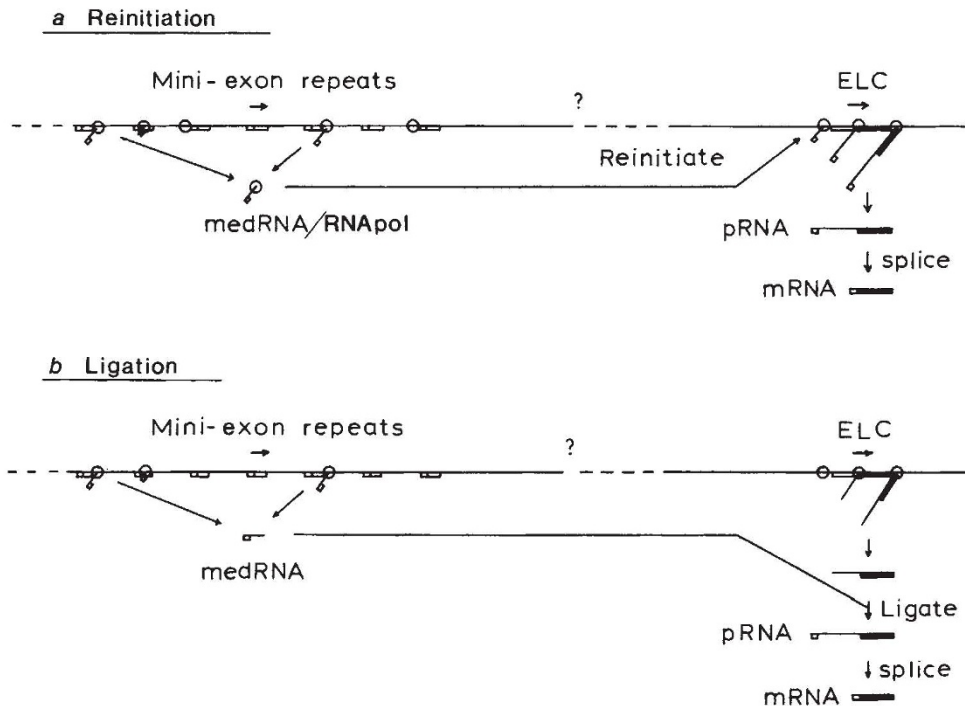
Fig. 7 Two models for discontinuous transcription of VSG mRNAs. *a*, The reinitiation model; *b*, the ligation model. The figure illustrates schematically the important features of the two models which are described in the text. The region of the mini-exon repeat which codes for the medRNA is represented by an open segmented box, the left half indicating the mini-exon coding for the 35-nucleotide spliced leader. The ELC region is also represented as a segmented box with the protein-coding exon shaded. The exon-derived portions of the precursor RNA (pRNA) are boxed in the same way. Uncertainty about whether the mini-exon repeats are linked to the ELC and if so over what distance, is indicated by a question mark over the dashed region. RNA polymerase is represented by an open circle.

expressed (ref. 21 and M. Parsons *et al.*, personal communication). In addition, the mini-exon sequence (or a conserved homologue) has been shown to be present in the genomes of several related species and genera which lack antigenic variation of the type observed for *T. brucei*[22].

Could discontinuous transcription be operating in these other cases? This question cannot, as yet, be answered, but some precedent does exist in systems which are otherwise totally unrelated. Transcription of the influenza virus is known to require a 5'-capped primer (10–15 nucleotides long) derived from cleavage of host mRNAs[36,37]. Coronavirus transcripts, on the other hand, are known to have a common 5'-leader sequence which is seemingly not joined to the coding portion of the RNA by conventional splicing, again strongly suggesting discontinuous transcription[38,39]. The fact that, in both these cases, transcription occurs by an RNA-dependent viral transcriptase may be an important distinction but they do provide a mechanistic precedent for discontinuity in primary transcripts. The observation made in Fig. 6 that a snRNA of yeast has substantial homology with the region known to be a donor splice site in medRNA, suggests that this snRNA (of unknown function) might be similarly involved in discontinuous transcription, albeit on a lesser scale (snR3 is nonessential; ref. 32). Such activity could further provide a molecular basis for the long-standing observation that heterogeneous nuclear RNAs often possess sequences at their 5'-ends which are derived from middle repetitive DNA[40–42] and that such sequences might be important in the control of developmentally regulated gene expression. The testing of this interesting hypothesis in trypanosomes may now be possible.

1. Cross, G. A. M., Holder, A. A., Allen, G. A. & Boothroyd, J. C. *Am. J. trop. Med. Hyg.* **29**(5) Suppl., 393–417 (1980).
2. Englund, P. T., Hajduk, S. L. & Narini, J. C. *A. Rev. Biochem.* **51**, 695–726 (1982).
3. Borst, P. & Cross, G. A. M. *Cell* **29**, 291–303 (1982).
4. Pays, E., Lheureux, M. & Steinert, M. *Nature* **292**, 265–267 (1981).
5. Hoeijmakers, J. H. J., Borst, P., Van den Burg, J., Weissmann, C. & Cross, G. A. M. *Gene* **8**, 391–417 (1980).
6. Van der Ploeg, L. H. T. *et al. Nucleic Acids Res.* **10**, 593–609 (1982).
7. Capbern, A., Giroud, C., Baltz, T. & Mattern, P. *Expl Parasit.* **42**, 6–13 (1977).
8. Cross, G. A. M. *Parasitology* **71**, 393–417 (1975).
9. Hoeijmakers, J. H. J., Frasch, A. C. C., Bernards, A., Borst, P. & Cross, G. A. M. *Nature* **284**, 78–80 (1980).
10. Pays, E., Van Meirvenne, N., Le Ray, D. & Steinert, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2673–2677 (1981).
11. Williams, R. O., Marcu, K. B. & Majiwa, P. A. O. *Nature* **282**, 847–849 (1979).
12. Van der Ploeg, L. H. T. *et al. Nucleic Acids Res.* **10**, 3591–3604 (1982).
13. Boothroyd, J. C. & Cross, G. A. M. *Gene* **20**, 281–289 (1982).
14. Longacre, S. *et al. Molec. cell. Biochem.* **3**, 399–409 (1983).
15. Bernards, A. *et al. Cell* **27**, 497–505 (1981).
16. Van der Ploeg, L. H. T., Bernards, A., Rijsewijk, F. A. M. & Borst, P. *Nucleic Acids Res.* **10**, 593–609 (1982).
17. Michiels, F. *et al. EMBO J.* **2**, 1185–1192 (1983).
18. Liu, A. Y. C., van der Ploeg, L. H. T., Rijsewijk, F. A. M. & Borst, P. *J. molec. Biol.* **167**, 57–75 (1983).
19. Campbell, D. A., van Bree, M. & Boothroyd, J. C. *Nucleic Acids Res.* **12**, 2759–2774 (1984).

20. Nelson, R. G. *et al. Cell* **34**, 901–909 (1983).
21. De Lange, T. *et al. Cell* **34**, 891–900 (1983).
22. Nelson, R. G. *et al. Nature* **308**, 665–667 (1984).
23. Borst, P. *et al. Eur. J. Biochem.* **137**, 383–389 (1983).
24. McDonell, M. W., Simon, M. N. & Studier, F. W. *J. molec. Biol.* **110**, 119–146 (1977).
25. Bolivar, F. *et al. Gene* **2**, 95–113 (1977).
26. Rigby, P. W. J., Diekmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
27. Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965 (1975).
28. Twigg, A. J. & Sherratt, D. *Nature* **283**, 216–218 (1980).
29. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721–732 (1977).
30. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning. A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
31. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
32. Tollervey, D., Wise, J. A. & Guthrie, C. *Cell* **35**, 753–762 (1983).
33. Cordingley, J. S. & Turner, M. J. *Molec. biochem. Parasit.* **1**, 91–96 (1980).
34. Breathnach, R. & Chambon, P. *A. Rev. Biochem.* **50**, 349–383 (1981).
35. Wise, J. A. *et al. Cell* **35**, 743–751 (1983).
36. Plotch, S. H., Bouloy, M., Ulmanen, I. & Krug, R. M. *Cell* **23**, 847–858 (1981).
37. Lamb, R. A. & Choppin, P. W. *A. Rev. Biochem.* **52**, 467–506 (1983).
38. Siddell, S., Wege, H. & ter Meulen, V. *J. gen. Virol.* **64**, 761–776 (1983).
39. Baric, R. S., Stohlman, S. A. & Lai, M. M. C. *J. Virol.* **48**, 633–640 (1983).
40. Firtel, R. A. & Lodish, H. F. *J. molec. Biol.* **79**, 295–314 (1973).
41. Zuker, C. & Lodish, H. F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5386–5390 (1981).
42. Davidson, E. H. & Posakony, J. W. *Nature* **297**, 633–635 (1982).
43. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
44. Sanger, F. & Coulson, A. R. *FEBS Lett.* **87**, 107–110 (1978).

——————————— LETTERS TO NATURE ———————————

On the origin of Triton and Pluto

William B. McKinnon

Department of Earth and Planetary Sciences, and McDonnell Center for the Space Sciences, Washington University, Saint Louis, Missouri 63130, USA

Lyttleton hypothesized long ago that Triton and Pluto originated as adjacent prograde satellites of Neptune[1]. With the presently accepted masses of Triton and Pluto–Charon[2,3], however, the momentum and energy exchange that would be required to set Triton on a retrograde trajectory is impossible. The mass of Triton has probably been seriously overestimated[4,5], but not by enough to relax this restriction. It is implausible that the present angular momentum state of Pluto–Charon has been significantly influenced by Neptune[6]. It could not acquire such angular momentum during an ejection event unless a physical collision was involved, which is quite unlikely. The simplest hypothesis is that Triton and Pluto are independent representatives of large outer Solar System planetesimals. Triton is simply captured, with potentially spectacular consequences that include runaway melting of interior ices and release to the surface of clathrated CH_4, CO and N_2 (ref. 7). Condensed remnants of this proto-atmosphere could account for features in Triton's unique spectrum[8–11].

The dynamics of Triton's orbital evolution are considerably simplified by the fact that its specific dissipation function, Q, at tidal frequencies, is much less than that of Neptune ($Q_T \ll Q_N$). Here I assume a standard solid-body Q for Triton of ~ 100 (ref. 12). A lower bound on Q_N can be derived by requiring that the outward orbital evolution of a satellite given by

$$\frac{da_s}{dt} = 3k_{2N}\sqrt{\frac{G}{m_N}}\frac{R_s^5 m_s}{Q_N a_s^{11/2}} \quad (1)$$

is not so rapid that the satellite originated at the corotation radius of Neptune 4,500 Myr ago[12] (where m_s and a_s are the satellite's mass and semimajor axis; m_N, R_N and k_{2N} are Neptune's mass, radius, and tidal-effective potential Love number of the second degree; and G is the gravitational constant). k_{2N} is estimated at 0.43, subject to uncertainties in Neptune's rotation rate and J_2 (the coefficient of the second harmonic of the gravitational potential) (see ref. 13); other parameter values are given in Table 1. If Neptune's third satellite[14] is confirmed and proves to be regular and non-commensurate, then $Q_N \gtrsim 10^4$. A lower bound on the Q of Uranus of $\sim 2 \times 10^4$ is set using Miranda[15] and a k_2 for Uranus of 0.28. The Qs of both planets should be comparable, and are probably much larger.

Accordingly, the monthly radial tide raised on Triton by Neptune dominates Triton's orbital evolution, except for orbits of very small eccentricity. This tide does not transfer angular momentum for a synchronously-rotating Triton, but in the cases of interest here, such non-synchronous spin angular momentum would be negligible compared with orbital angular momentum. The present fractional rate of change in Triton's orbital angular momentum, due to the tide raised on Neptune by Triton, is