The genes encoding fructose bisphosphate aldolase in *Trypanosoma brucei* are interspersed with unrelated genes

S.Vijayasarathy[+], Isabelle Ernest[1], Jane E.Itzhaki[§], David Sherman, Michael R.Mowatt[φ], Paul A.M.Michels[1]* and Christine E. Clayton*
The Rockefeller University, 1230 York Ave, New York, NY 10021, USA and [1]International Institute of Cellular and Molecular Pathology, 74/39 avenue Hippocrate, 74 B-1200 Brussels, Belgium

Received December 22, 1989; Revised and Accepted March 20 1990

ABSTRACT

The fructose bisphosphate aldolase genes of *Trypanosoma brucei* are interspersed with unrelated genes whose transcript levels show no developmental modulation. Transcription appears approximately constant across the entire locus, suggesting that aldolase mRNA abundance is regulated post-transcriptionally.

INTRODUCTION

Trypanosoma brucei is a parasite of the bloodstream and tissue fluids of mammals. Trypanosomes are transmitted from one host to the next by tsetse flies (genus Glossina). While in the mammal, they derive all their energy from substrate-level phosphorylation during glycolysis: most of the enzymes responsible are compartmentalised in peroxisome-like microbodies called 'glycosomes' (1). Adaptation to the tsetse fly environment involves extensive mitochondrial elaboration including the building of a network of cristae and the appearance of citric acid cycle enzymes and cytochromes; at the same time, the rate of glycolysis is thirty-fold reduced (2). We are interested in the details of this developmental regulation; in particular, in the regulation of expression of fructose bisphosphate aldolase, a glycosomal enzyme that is thirty times more abundant in bloodstream trypanosomes than in the 'procyclic' insect forms (3, 4). The level of stable aldolase mRNA is at least six times higher in bloodstream forms than procyclic forms (5).

Each diploid trypanosome has four aldolase genes arranged as allelic tandem repeats (5, 6). This type of genomic organization is very common in trypanosomatids (7): other examples include the genes encoding glyceraldehydephosphate dehydrogenase (8), phosphoglycerate kinase (9), and the procyclic acidic repetitive proteins (10), while those encoding tubulin and the heat shock proteins (reviewed in (7)) and calmodulin (11) occur in more extensive arrays. Evidence has been presented that transcription of many of these genes is polycistronic. The phosphoglycerate kinase genes are particularly interesting because developmental regulation of their RNAs appears to be effected post-transcriptionally (12). Meanwhile, no trypanosome RNA polymerase II promoter or transcription start site has yet been definitively identified. This is partly because all trypanosome mRNAs are processed by a *trans*-splicing reaction in which the original 5'-end of the transcript is replaced by a thirty-nine nucleotide 'mini-exon' sequence (reviewed in (7, 13)), and also because until very recently no reliable DNA transformation methods have been available.

In this paper we show that the aldolase genes are interspersed with unrelated genes whose transcript levels show no developmental modulation. Transcription appears constant across the entire locus, suggesting that mRNA abundance is regulated post-transcriptionally.

MATERIALS AND METHODS

Trypanosomes

Strain 427 parasites were cultured or grown in rats (5, 14).

DNA cloning and characterization

DNA cloning in the EMBL3, pAT153, pUC12, pTZ (LKB-Pharmacia) and Gemini (Promega Biotec) vectors, restriction mapping, Northern blot analysis and chain-termination sequencing were done using standard methods as previously described (5, 6, 14). Most of the sequencing was done with double-stranded DNA using Sequenase (United States Biochemical corporation) or Klenow fragment of E. coli polymerase I. cDNA clones were selected from existing lambda gt10 libraries (5, 10). Nucleic acid sequences were analysed with DNA Strider (Macintosh) and Pustell, ARP and other packages on the Rockefeller University UNIX system.

Primer extension and polymerase chain reaction

Primers were annealed to RNA and extended using AMV reverse transcriptase in the presence of 25 μg/ml Actinomycin D and

* To whom correspondence should be addressed
Present addresses: [+]Department of Pharmacology, Stanford University Medical Center, Stanford, CA 94305, USA, [§]Wolfson College, Oxford OX2 6UD, UK and [φ]Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA

alpha-[^{32}P]-dCTP. Products were analysed on an 8% polyacrylamide sequencing gel with a sequencing ladder as markers. A separate primer-extension was digested with RNase, extracted with phenol-chloroform, ethanol precipitated and the product amplified by the polymerase chain reaction using Taq polymerase (15) and a Techne thermal cycler. Primers used were a variety of sequences complementary to the 5′ end of RNAs, and a sense-strand primer corresponding to the last 35 nucleotides of the mini-exon of *Leptomonas seymouri* (which is 79% homologous to the *T. brucei* mini-exon) (16). The annealing temperature was 53°C. Products were analysed on gels containing 3.5% NuSeive (FMC) and 1% normal agarose in tris-borate buffer. Oligonucleotides were made by the Rockefeller University nucleic acid synthesis facility.

Transcription in isolated nuclei

Nuclei were isolated from trypanosomes using a Stansted Cell Disruptor and stored at −70°C as described by Kooter *et al.* (17). Bloodstream trypanosome nuclei were made directly from uncooled infected rat blood; cultured procyclic trypanosomes were harvested at densities of $0.5-1.0\times10^7$ parasites/ml. To analyse transcription, nuclei from 10^9 parasites were thawed, and incubated in 200μl transcription buffer after 0−2 washes in transcription buffer. The transcription buffer contained 100mM Tris pH 7.8−8.0, 50mM NaCl, 0−120mM KCl, 25% glycerol, 1.2−2.5mM DTT, 2mM $MgCl_2$, 2mM $MnCl_2$, 4mM ATP, and 1mM CTP. In some experiments, 5μM GTP and 5μM UTP were supplemented by 250μCi each of alpha [^{32}P]-UTP and -GTP (3000Ci/mMol, Amersham International). Better results (including those shown) were obtained by including in the reaction 1mM UTP and 500μCi of alpha [^{32}P]-GTP at about 2000 Ci/mMol (obtained by mixing 250 μCi each of 400Ci/mMol and 3000 Ci/mMol). Mixtures were pre-incubated on ice for 10−20min in the presence or absence of alpha-amanitin (Serva) and extension allowed to proceed at 27−37°C for 5−10 min. Reactions were terminated by digestion with RNAase-free DNAase (Promega) and proteinase K, and the products purified by two sequential spin-columns, sometimes followed by ethanol precipitation. Reactions yielded $1-9\times10^6$ cpm of acid-precipitable RNA per 10^9 nuclei.

Cloned DNAs were linearised by restriction digestion, denatured with 0.3M NaOH (incubated at 65°C for 10min) and 5−10μg applied to nitrocellulose (Schleicher & Schull), Duralon or Duralose (Stratagene) membranes in 1M ammonium chloride. (Use of 6×SSC gave very poorly reproducuble results.) DNA was immobilised by baking or uv-cross-linking using a Stratalinker (Stratagene). Radioactive RNA was hybridised to the filters at a concentration of 10^6-10^7 cpm/ml in 5×SSC, 50 mM sodium phosphate pH 7.0, 5×Denharts solution, 50% formamide, 0.25mM sodium pyrophosphate, 50 μg/ml wheat germ tRNA, 250 μg/ml sonicated salmon-sperm DNA, 100 μg/ml heparin for 12−40hr after 2−12h pre-incubation at 42°C. Filters were washed in 0.2−0.5×SSC, 0.1% SDS at 65°C after digestion at 37°C for 30−45min with RNAse A at 100−500μg/ml in 2×SSC. Films were pre-flashed before autoradiography at −70°C. The hybridization was quantitated by scintillation counting of cut-up membranes and by laser densitometry.

RESULTS

Transcription between and upstream of the aldolase genes

A region of about 20kb containing the aldolase tandem repeat was cloned from *T.b. brucei* 427 using lambda and plasmid vectors. A partial restriction map of part of this region is shown in Figure 1. The aldolase genes are arranged in a direct tandem repeat covering almost 8kb. The end of the second repeat is at about 11.5kb on the map just before the beginning of open reading frame D.

Various clones from the region shown in Figure 1 were hybridized to Northern blots of bloodstream and procyclic total or polyadenylated RNA. Three stable transcripts homologous to these areas were identified; none shows significant developmental regulation (Figure 2). The clones I1 and I2 (from between the aldolase genes) hybridize to a polyadenylated transcript of about 900bases; using a strand-specific labelled I2 RNA probe the direction of transcription was determined to be the same as that of the aldolase genes. (The additional higher molecular weight bands are caused by non-specific hybridization of the riboprobe as they are not seen using DNA probes.) Clones U1 and U2 hybridised to two polyadenylated RNAs of 800bases (Ua) and 1kb (Ub) transcribed in the same direction as aldolase (Figure 2, panel U1). To confirm the origin of these transcripts, cDNAs were cloned. One clone each corresponding to the Ua and I transcripts were obtained from a bloodstream 427 cDNA library (5); a Ua and a Ub clone were obtained from a cDNA library constructed from procyclic forms of T. brucei TREU 667 (10). The abundance of these clones in a bloodstream cDNA library

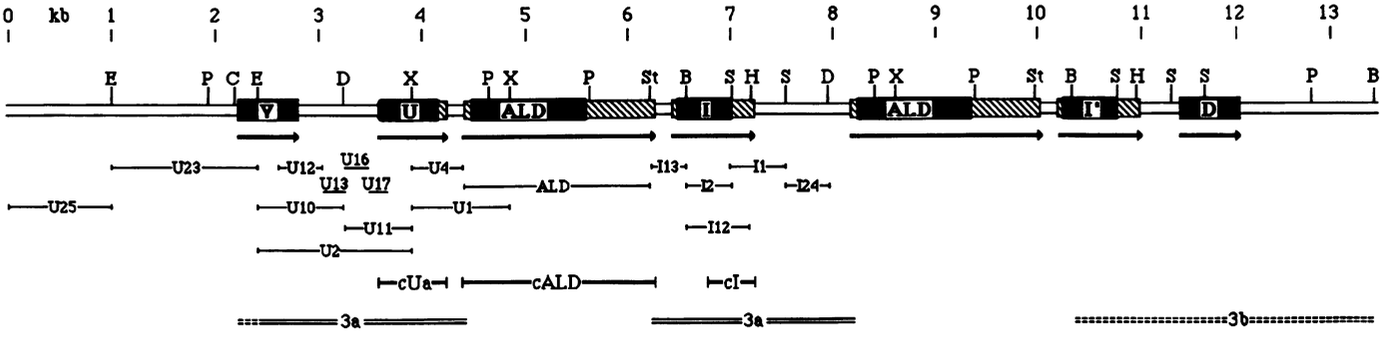

Figure 1. Map of aldolase locus. Open reading frames are represented by solid bars and untranslated regions of transcripts are shaded and the direction of translation indicated. Where known, for U, I and ALD, the direction of transcription corresponds to this. The positions of clones are below the map. Restriction sites are as follows: B-BamHI; C-HincII (only one shown out of several); D-DraI; E-EcoRI; H- HindIII; P-PstI; S-SalI; St-StuI; X-XhoI. The DraI site at 3.3kb is absent in the B allele. The direction of transcription is indicated by the arrows; the extents of cDNAs (cUa, cALD and cI) are indicated. Regions whose sequences are shown in Figure 3 are indicated.

was much lower than that of aldolase, as was expected from the signal seen by blot hybridization; only one clone of each was found in each library-screening attempt, for both bloodstream and procyclic libraries, so the frequency cannot be estimated.

Sequence analysis

The sequence of the aldolase coding region and the corresponding cDNA has been presented previously (5, 6). We sequenced the upstream 2.2kb (from the HincII site at 2.1kb on the map) to the first aldolase AUG, and the intergenic region from the PstI site at position 5.3kb to the second aldolase AUG (Figure 3), as well as the 427 intergenic cDNA and the TREU 667 Ua and Ub cDNAs (figure4). Partial sequence was also obtained from the downstream region to a BamHI site at 13.4kb (Figure 3b); also from the strain 427 Ua cDNA. There are two aldolase alleles in *T. brucei* , one of which (the A allele) has two Pst sites in each aldolase gene. Most sequence was from this allele; the other (B allele) has lost a Pst site in the upstream aldolase gene. The sequence of the published full-length aldolase cDNA contains both Pst sites but we also have a cDNA with only one. The TREU667 strain has one allele in which the upstream aldolase gene has an extra PstI site, caused by the substitution of an A for the G at position 839 in the cDNA sequence (5) as modified by Marchand *et al.* (6); this does not change the protein sequence.

The upstream segment has two long open reading frames. Data on these and others found in the aldolase locus are summarized in Table 1. The first one (V, nt 73−769) codes for a protein of 232 amino acids (25,573 Da). This protein would have a hydrophilic, somewhat basic character (19 arginines, 27 lysines, 10 aspartate and 15 glutamate). The predicted amino acid sequence has a highly hydrophobic stretch of 17 residues (Pro_{175}-Leu_{194}) suggesting that this open reading frame encodes a membrane-associated protein. However, no corresponding transcript was detected by blot hybridization with U2 and U10 DNA probes so we do not know if a stable RNA is made.

A second open reading frame (U in figure 1) coding for a protein of 190 amino acids is found between nt 1461 and 2031. The predicted 22,037Da protein is hydrophilic and very basic, having 25 arginines and 14 lysines as opposed to 10 aspartate and 10 glutamate residues. The TREU667 Ua cDNA starts at nt 1458 and continues to the polyadenylation site at a string of four As (2099−2102). Its sequence is identical to the 427 genomic sequence; the 3′-untranslated region is 66nt. Partial sequence of the 427 Ua cDNA is identical but starts just after the initiation codon.

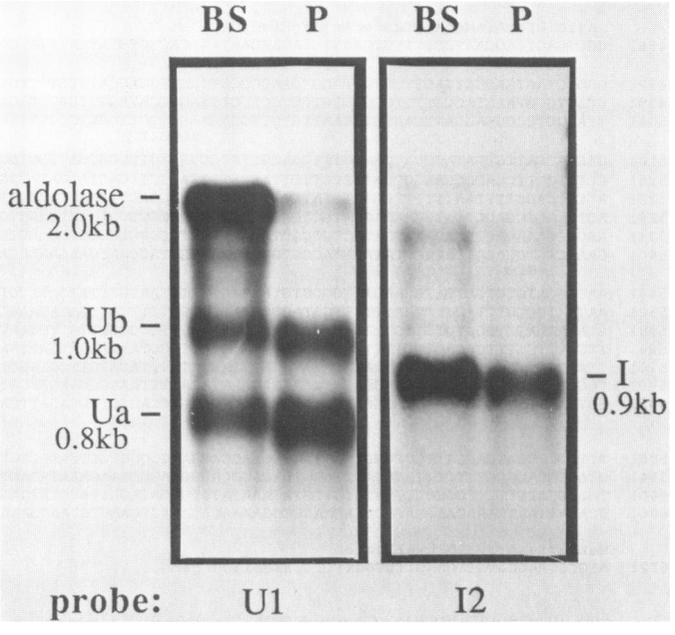

Figure 2. Blot hybridization analysis of polyadenylated RNA (10 μg/lane) from bloodstream (B) and procyclic (P) trypanosomes hybridized with antisense riboprobes from clones U1 and I2 (See Figure 1).

The TREU 667 Ub cDNA is 978 nt long (figure 4). The coding region shows only silent 3rd base changes (relative to Ua) but the 24nt at the 5′- end and the 3′-untranslated 328 nt are completely different. Probing of RNA and genomic blots with the Ub cDNA and with a subcloned 3′- fragment that is specific for Ub (generated by cleavage of the cDNA with MluI) showed that the Ub RNA is transvcribed from elsewhere in the genome of both 427 and TREU667 strains (not shown). This is interesting as the levels of Ub and Ua RNAs are similar, suggesting that they may be controlled in the same way.

The region between the aldolase genes contains a single open reading frame, 'I' in figure 1, (from nt 4343−4898) of 185 amino acids (21,727Da), again a hydrophilic, rather basic protein with 18 Arg, 18 Lys, 9 Asp and 23 Glu residues. A strain 427 cDNA covered the C-terminal half of this region from nt 4735−5122 corresponding to the 900nt intergenic transcript. A partial genomic sequence from TREU 667 DNA revealed a 38nt deletion relative to 427, starting 20nt downstream of the termination codon. No stable RNAs have been detected from the rest of the intergenic region, which is AT-rich and repetitive in nature. Homology with the segment between the Ua gene and aldolase starts abruptly at nt 5881, beginning exactly at the position of Ua polydenylation. The sequence repeats almost exactly down to the aldolase coding region.

The repeated area ends 585nt 3′ to the second I gene, labelled I′ in figure 1. 67 nt downstream of this point is another long open reading frame (D). It can encode a protein of 186 amino acids (20,633Da), yet again rather hydrophilic and basic (18 Arg, 11 Lys, 7 Asp, 12 Glu). We do not know if this region is transcribed.

Mapping of mini-exon addition points

All trypanosomatid mRNAs examined so far carry at their 5′ends a 39nt sequence that is added post-transcriptionally by trans-splicing (18, 19). Preliminary experiments using sandwich hybridization and RNAaseH mapping suggested that this was true for the aldolase, Ua, and I RNAs (data not shown). The 5′-ends of the aldolase and U RNAs were mapped by primer extension. A 20nt primer corresponding to the 5′-untranslated region of aldolase up to and including the initiation codon (primer 11 in Figure 3, nt 2348−2329) was hybridised to bloodstream RNA and extended by reverse transcriptase, yielding a product that comigrated with nt 2237, or position −108 relative to the AUG (Figure 5). (The minor band at position −100 is an artifact that was present in all reverse transcription lanes in this experiment; it is not usually seen.) The splice acceptor site is always an AG dinucleotide. Our longest aldolase cDNA starts at nt 2281, just downstream from AGs at nt 2277 and 2271. During primer extension, only the 3′ 35nt of the mini-exon is reverse-transcribed, as the 5′-most four nucleotides are methylated (20). The AG at 2271 is therefore probably the splice acceptor site. Similar results were obtained using two other primers (not shown). That means that the polyadenylation site of Ua is only

176−180nt upstream of the start of the mature aldolase mRNA. Primer extension using the Ua-specific primer 6, nt 1473−1454, yielded a product comigrating with nt 1418−97, suggesting that the mini-exon addition site for Ua is at nt 1453−4. There is an AG dinucleotide at 1452. As expected, primer 4, which spans the acceptor site, yielded no extension products.

Mapping of the I transcript 5′-end was more problematic. As the cDNA was rather short relative to the RNA, we were unsure

a

```
   1  CAATCAAAACTCCCTCACCTCCCTCTCCAAACAGGTGAGTTGGCGACACTTCTTTCGCCA

                 MetLeuLeuCysAspSerValProSerProLeuSerLeuLeuLeuArg
  61  AACCCCGGAGAGATGCTGCTGTGTGATTCAGTTCCATCCCCACTCTCACTGTTGCTACGG

      ArgCysAlaThrValLysMetSerGluAsnGlyThrSerArgArgGlyGlyAlaThrArg
 121  CGTTGTGCCACGGTTAAGATGTCTGAAAATGGAACTTCGAGACGGGGTGGAGCAACTCGC

      AsnGluSerValValGlySerSerLeuCysSerPheAlaAsnArgProAlaProValSer
 181  AATGAATCCGTGGTGGGCTCTTCATTATGCAGCTTCGCAAACCGTCCAGCGCCAGTATCC

      AspThrGlnPheIleArgSerAlaValGlyLysAsnSerCysPheLeuIleGluIleVal
 241  GATACTCAGTTTATTCGTTCGGCTGTGGGGAAGAATTCTTGCTTCCTGATCGAGATAGTG

      ArgTrpLeuArgAspValProSerAspGluAlaLysAspValAlaLysGluLysGlySer
 301  CGGTGGTTGCGGGATGTACCTTCTGATGAGGCAAAGGATGTAGCAAAGGAGAAAGGCAGC

      AlaTyrSerAsnSerGluGlyThrLysGlnGlnAsnGlyArgThrAsnGlyGluSerAsn
 361  GCTTATAGCAATTCCGAGGGTACGAAGCAACAAAATGGAAGAACAAATGGGGAAAGTAAT

      ArgGluGlnArgLeuLeuLeuGlyLeuSerThrSerAsnGluIleIleAspHisValVal
 421  AGGGAGCAGCGGCTTTTGTTGGGGTTGTCTACTTCAAATGAAATTATCGACCACGTTGTT

      HisAspLeuProPheGlyThrAlaValAlaGlnLeuIleArgGluLeuValGlySerAsn
 481  CACGATCTCCCCTTTGGCACCGCCGTTGCACAATTAATACGTGAACTCGTTGGCAGTAAC

      LeuPheThrGluGluGlnLeuHisArgTrpIleGlyArgSerLeuLysAspValGluSer
 541  TTGTTCACAGAGGAACAGTTGCACAGGTGGATCGGGCGCTCACTTAAAGATGTAGAGAGT

      LysProProAlaThrIleGlyCysPheAlaAlaLeuLeuValPheValLeuLeuHisSer
 601  AAACCACCCGCAACAATCGGATGTTTTGCGGCACTTCTTGTTTTCGTTCTTCTTCACAGT

      ArgTrpGluLeuProAlaAspIleArgGluMetThrValGlnLeuCysSerGlnAsnLys
 661  AGGTGGGAACTCCCAGCAGATATACGAGAGATGACAGTCCAACTCTGCTCGCAGAACAAA

      GlnGlnGlnAspCysValGlnLeuIleAsnIleLeuGlnArgGlyAlaAM
 721  CAGCAGCAGGACTGTGTTCAACTGATTAACATTTTGCAGCGGGGCGCTTAGCGGTGACGC

 781  ACTCGCGCAAACGTTGAGTGGTGCCGCATTTGGAAGACATTTGACATTTCCTTGATGGGT
 841  GCTGAGTGGAGTTGTCTTCCCGAGGGACTTTGTGCCAAACAGTGTAACAAATTGCTAAAA
 901  TCGAATAACGCGCCTCTGGGCAGAACAATCGCGCATGAGGTTTCCCTCGTGACTTCACTG
 961  GAACAATAGGTGTTATTATGGAAGAGTGTGCCCCAGCAGGAGGAAAAAAGGGGGTGAGGG
1021  AAAACCACGTCACAACACAAGCCTTATATTCGTGTAAAAAATAATAACAATAATTTTTCA
1081  TTTACCAGTTGTATCAGTCTTTACTCTTTAAAAAAAAAATGCTTCGACTCCTGACACGTA
                                  DraI
1141  CCCTTAGCATCAATAAAGCGGTTGCTTACTACTCGTGTTATTATTCGCTCGCCTGTCCAA
1201  AGCGTGAGGAAGCGAGCTTTGTTGCTGTGCATGTTTGTCCGCTTTGAGTTTATTTAGTCC
1261  CCGTCCTTTGTGTCGTATTGGAAGTAACGTTTTCCGTGTCCAAGCAAATTTATTTGGTGC
1321  TAATGTAACATGAATAATGAAAGTCGAATCGCCCTCTCGATGTGTTTGCTCCCCTGTACA
1381  AGCGCTTCCTCTTCTTTTCGTGCCTACATTAACTTTTATTTATGTATTGCATTATTTACG
                                        *                  [4--

                        MetArgAsnTyrAsnAsnPheAsnArgValTrpLysAlaPro
1441  TGCTTACGAAAGGGGTTAAGATGCGCAACTACAACAACTTCAACCGTGTTTGGAAGGCCC
      -----------[6=======4]---------6]

        ArgArgProPheGluLysGluArgLeuAspArgGluMetLysLeuCysGlyGlnTyrGly
1501  CACGACGCCCCTTCGAGAAGGAGCGTCTCGACCGCGAGATGAAGCTCTGCGGGCAGTACG

        LeuArgCysLysArgGluIleTrpArgValAsnMetThrLeuSerLysMetArgArgThr
1561  GTCTCCGCTGCAAGCGCGAGATTTGGCGCGTCAACATGACACTCTCAAAGATGCGTCGCA

        AlaArgLeuLeuLeuThrLeuProGluAsnHisProArgArgLeuLeuGluGlySerAla
1621  CCGCGCGTCTTTTGCTGACGCTTCCGGAAAACCACCCGCGCCGCCTTCTTGAGGGCTCTG

        IleMetArgArgCysHisGlyTyrGlyPheLeuAspGluAspLysAspLysLeuAspTyr
1681  CAATTATGCGCCGCTGCCACGGATATGGTTTCCTCGACGAAGATAAGGACAAACTCGATT

        ValLeuSerLeuThrValProAspIleLeuGluArgArgLeuGlnThrValValPheLys
1741  ATGTGCTTTCGCTCACCGTTCCCGACATTCTCGAGCGCCGCCTTCAAACGGTCGTCTTCA

        HisGlyLeuAlaLysSerValHisHisSerArgValLeuIleGlnGlnArgHisIleAla
1801  AGCATGGCCTCGCAAAGTCCGTCCACCACTCCCGTGTTCTTATCCAACAGCGGCACATTG

        ValAlaLysGlnIleValThrIleProSerPheIleValArgValSerSerGluHisHis
1861  CCGTTGCTAAGCAGATTGTTACGATTCCTTCCTTCATTGTCCGCGTGAGTTCTGAGCACC

        IleAlaPheAlaAspAlaSerProPheGlyAsnGlyArgProGlyArgValLysArgVal
1921  ATATCGCCTTTGCGGATGCGTCGCCATTCGGCAACGGCCGACCGGGTCGCGTGAAGCGTG

        LysArgAsnAlaAlaLysLysGlySerGlyGlyGlyAspAspAspGluOC
1981  TGAAGAGGAACGCTGCGAAGAAGGGTAGCGGTGGCGGTGATGATGACGAGTAAAGATGTG
```

```
2041  AATGTGTCAAAAACCCTCATTACTGACGTTCTTTCTACTTTCTTTATTCTTTCTGAGTAA
2101  AATCTTTTCTCTCCATCAGCATTGCTGGGTCTGTTGCGCGACCAGGACGCGGGTGCTCAA
2161  GCTGTGTAGCGCACGCGTTTCCTTACATATTTCTCTAACAGGCACGGAAGCCTAACAAAT
2221  ACACTTGGCTTATTTTTTTGCCCCCTCATGTCTTGTACAAATATTTGCGATAGCTTAGCT
                    *

2281  ATCAGCCACATTAATCAAACAAGTATACCAACAAGCCCGAAAACATAAACTCAACTGCAA
                                                    [11----------

           MetSerLysArgValGluValLeuLeuThrGln
2341  CGAAGATGTCCAAGCGTGTTGAAGTTCTGCTTACCCAA....aldolase gene.....
      ------11]

4021  TCCGAAAATATTTTGATGCACCAAATAAGTGAACATACTGACGAAATCAAAGAGTTGGAG

4081  GATGGATAGGGAGGCCTCATTGGCAGTGGTAAATTGATTTGACTTGAATATGATACCCTC
                StuI
4141  TTACTGTTTCATTGCCACTACAAGTTGGTTTCCTTCCCCGGATGAATGGTTGTGAGGACA
4201  TCCTTTTATGTGCCTTTGCGCCCCTTTATTTTCACTTCTTGTACCCTTATTTTACTCGCC
4261  ATGCCATCGATACTTTGCTTGTGGTGGCGGTCGTGTGGGTTACCTGTTCGCTTGTTAAGG
                                  *                  (*)

                                                MetAspAspGlyValGluAlaLysProLeuCysLeuThr
4321  TGGTTGATTCCTTTGGTAGAAGATGGATGATGGTGTAGAAGCCAAGCCCCTCTGCTTGAC
                                                      [2---------(*)-2]

        ArgGluGlnIleAspLysGlnValGluArgLeuSerArgArgProGluGlnArgThrLeu
4381  ACGGGAACAAATTGATAAGCAAGTAGAGCGGCTGTCGCGCCGGCCGGAGCAGCGAACGCT

        ProAspProProheProValCysProThrValArgMetSerLysGluGlnLeuGluGlnVal
4441  ACCGGATCCGTTTCCTGTTTGTCCGACTGTGCGAATGTCGAAGGAGCAACTTGAGCAAGT
          BamHI[8-------------------8]

        ThrLysArgValPheTyrHisTyrSerGluLysHisAlaGluAlaLeuArgLeuAlaGlu
4501  TACGAAACGTGTGTTTTATCATTATTCGGAGAAACACGCCGAAGCGTTGCGACTTGCGGA

        GluArgArgGluLysGluCysGlyValAlaSerThrValLeuSerAlaSerAspValAsp
4561  AGAAAGGCGTGAGAAGGAATGCGGGGTTGCATCCACTGTATTGAGCGCGAGTGATGTGGA

        AspIleValLysArgLeuTyrTyrGluGlyMetGluArgValLysValGlyArgLysGlu
4621  TGACATCGTGAAGCGTTTGTACTATGAGGGAATGGAGCGTGTGAAGGTGGGTCGCAAGGA

        AlaSerAspArgLeuLeuPheLysSerThrLysValLeuProValIleSerLeuLysArg
4681  GGCTAGCGACCGTCTGTTGTTTAAGTCAACCAAGGTTCTTCCGGTTATCTCGCTAAAGAG

        PheValAsnAspMetTyrLeuArgGlyLeuGluArgGluLysLysLysGluGluLysLeu
4741  GTTTGTCAACGACATGTATCTACGCGGTTTGGAGCGGGAAAAGAAGAAGGAGGAAAAGCT

        TyrGluLysTyrIleLeuProThrGluIleProAsnLeuArgIleSerLysSerGlnAla
4801  GTACGAGAAGTACATTCTCCCTACAGAGATTCCCAATCTGAGGATATCGAAGTCTCAAGC

        AlaGluSerAlaMetArgLeuSerArgArgHisGluOC
4861  CGCGGAGTCAGCAATGCGCTTGTCACGTCGACATGAATAATCACTGTGTATCATTTATCC
                                  SalI
4921  GGGTGTAGTGACGGTGAGTTTCTCGTGGCTGAAGGGGGTTCCTTGCACCATTTGTACTGT
4981  GCACTGCGTAACTAGCATTTTGCCTCTGTGTTGCTCCTTAAGGGCATAGGTTCATGTAGA
5041  TTAGCGTCCCGGAGGAATTAGTGTGGAATTGTGTTGGAGAAGCTTCTTGCCTCCGTTTTA
                                                HindIII
5101  GACCACGATTGTGATAGCCATGACTGTGTGGACGTTTTCCGTGTGTCCCGTGATTGGTGC
5161  CCTGTGTTTCCACCCGAAGGTGATTTCCTCCTTATCATTTGATTCTTCACTATGATATTC
5221  ATTTTCACCTTTTAATTTTTTCGTGTGTATGCATTTCACGTAAGGAGTTTAACTAGGGGT
5281  AGTCAAGGCGAGCTGATGGAAGACGATGCTGATGCCTACCTTCGGCGGGTGTCTCAACTG
5341  AAGCGGAAACACTTCAGAAAGGATTCTGTCACTGCATATCATCCCACGCCGGTGGAGGTG
5401  GAAACAGTCGACACCAAACCTATACATACCGTGGCTCGAAGGCTACCAATGAAGAAACGA
            SalI
5461  AAGCGGATGTGTGATGATGCAAGATTGGGGTGTAATAGTTAGGTAATGTTTATCATTTCT
5521  AATAATGCCGTTTACTTTTATTTTTTTGGTAGTCGGACTGTGATTTTTATTGATATAACT
5581  GTAAATGATTTTGGTGTGTTCTTATATAGCTTTGGAGCATAGTTTCGCGGCTTTTGCAAT
5641  TTTTTTTTGTTTGCCACTTTTTTTTTTTGACTTTGAGCTGGTGGCTAATATAATGATTAA
5701  TTATTTACCCGTTCCCAATATATATATATATATTTACCATATTTAGATGTTCGTGGTT
5761  TGGGCTTCTGGTTGCCATTGCACGACTTCGATGACAATTACACTTTAAAGCGAACGTCAT
5821  TGTGAATTGTACCGAGAAAGGGGGGTTATAAAAGAGTAGCTGACCAGGCTACTAGATTCA
                                                                |1

5881  TTTCTCTCCATCAGCATTGCTGGGTCTGTTGCGCGACCAGGACGCGGGTGCTCAAGCTGT
5941  GTAGCGCACGCGTTTCCTTACATATTTCTCTAACAGGCACGGAAGCCTAACAAATACACT
6001  TGGCTTATTTTTTTGCCCCCTCATGTCTTGTACAAATATTTGCGATAGCTTAGCTATCAG
6061  CCACATTAATCAAACAAGTATACCAACAAGCCCGAAAACATAAACTCAACTGCAACGAAG

        MetSerLysArgValGluValLeuLeu
6121  ATGTCCAAGCGTGTTGAAGTTCTGCTT....aldolase gene.....
```

Figure 3a Sequence of the region around the aldolase genes. The sequence starts 273nt upstream of the EcoRI site at position 2.4kb on the map. A translation of the long open reading frames is included but the main body of the aldolase mRNA sequence has been deleted for brevity as it is already published. The positions of oligonucleotides used in 5′-end analysis are indicated by the dashed lines enclosed in parentheses; and the AG dinucleotide splice acceptor sites are double underlined. The migration positions of primer extension products denoted by asterisks; these positions are in each case about 35nt upstream of the putative splice acceptor. The 5′ point at which the downstream aldolase locus diverges from the upstream one is marked by a vertical line, |1. Bases 1−273 were sequenced in the sense direction only.

which initiation codon was used. Extension of primer 2 (nt 4375−4361) (not shown) and primer 8 (nt 4470−4451) yielded only faint bands (Figure 5) whose lengths predicted mini-exon

b

```
     ProAspProPheProValCysProThrValArgMetSerLysGluGlnLeuGluGlnVal
   1 CCGGATCCGTTTCCTGTTTGTCCGACTGTGCGAATGTCGAAGGAGCAACTTGAGCAAGTT

     ThrLysArgValPheTyrHisTyrSerGluLysHisAlaGluAlaLeuArgLeuAlaGlu
  61 ACGAAACGTGTGTTTTATCATTATTCGGAGAAACACGCCGAAGCGTTGCGACTTGCGGAA

     GluArgArgGluLysGluCysGlyValAlaSerThrValLeuSerAlaSerAspValAsp
 121 GAAAGGCGTGAGAAGGAATGCGGGGTTGCATCCACTGTATTGAGCGCGAGTGATGTGGAT

     AspIleValLysArgLeuTyrTyrGluGlyMetGluArgValLysValGlyArgLysGlu
 181 GACATCGTGAAGCGTTTGTACTATGAGGGAATGGAGCGTGTGAAGGTGGGTCGCAAGGAG

     AlaSerAspArgLeuLeuPheLysSerThrLysValLeuProValIleSerLeuLysArg
 241 GCTAGCGACCGTCTGTTGTTTAAGTCAACCAAGGTTCTTCCGGTTATCTCGCTAAAGAGG

     PheValAsnAspMetTyrLeuArgGlyLeuGluArgGluLysLysLysGluGluLysLeu
 301 TTTGTCAACGACATGTATCTACGCGGTTTGGAGCGGGAAAAGAAGAAGGAGGAAAAGCTG

     TyrGluLysTyrIleLeuProThrGluIleProAsnLeuArgIleSerLysSerGlnAla
 361 TACGAGAAGTACATTCTCCCTACAGAGATTCCCAATCTGAGGATATCGAAGTCTCAAGCC

     AlaGluSerAlaMetArgLeuSerArgArgHisGluOC
 421 GCGGAGTCAGCAATGCGCTTGTCACGTCGACATGAATAATCACTGTGTATCATTTATCCG

 481 GGTGTAGTGACGGTGAGTTTCTCGTGGCTGAAGGGGGTTCCTTGCACCATTTGTACTGTG
 541 CACTGCGTAACTAGCATTTGCCTCTGTGTTGCTCCTTAAGGGCATAGGTTCATGTAGATT
 601 AGCGTCCCGGAGGAATTAGTGTCGAATTGTGTTGGAGAAGCTTCTTGCCTCCGTTTTAGA
 661 CCACGATTGTATAGCCATGACTGTGTGGACGTTTTCCGTGTGTCCCGTGATTGGTGCCCT
 721 GTGTTTCCACCCGAACCTGGATTTCCTCCTATCATTGATTCTTCACTATGATATTCATTT
 781 TCACCTTTTAATTTTTCGTGTGTATGCATTTCACGTAAGGAGTTTAACTAGGGTAGTCAA
 841 GGCGAGCTGATGGAAGACGATGCTGATGCCTACTTCGGCGGGTGTCTCAACTGAAGCGGA
 901 AACACTTCAGAAAGGATTCTGTCACTGCATATCATCCCACGCCGGTGGAGGTGGAAACAG
 961 TCGACACCAAACCTATACATACCGTGGCTCGAAGGCTACCAATGAAGAAACGAAAGCGGA
1021 TGTGTGATGATGATGAATTTGACGTTGTTTCTCTTGCGAGGGACGCTGTCAATGTTTCTT
                |2
                                MetGlyProSerThrSerProLeuValGlyCysAsnGly
1081 CTTGTGGAGTTTTTGTCGGCGGATGGGTCCTTCTACCAGTCCATTGGTTGGTTGTAATGG

      AspArgLysValLeuGlnGlnProValLysIleThrLeuMetGluLysTrpPheProGly
1141 TGATAGAAAAGTATTGCAACAACCGGTGAAAATTACGCTGATGGAAAAGTGGTTTCCGGG

      AspValAlaGluSerThrLeuAspProAlaGlnAlaProLeuTyrGlnTyrThrGluGly
1201 AGACGTTGCAGAATCGACTCTGGACCCTGCCCAAGCACCGCTGTACCAGTACACGGAAGGT

      ValAspGluAlaAsnGluGlyCysProArgAlaPheValThrLeuLysAlaGlyValGlu
1261 TGTCGACGAAGCGAATGAAGGATGCCCGAGAGCGTTTGTTACCTTAAAAGCCGGGGTGGA

      GluAspAlaLeuLysTyrSerPheValGlyCysCysSerAsnPheAspGluValValSer
1321 GGAGGATGCGTTGAAGTACTCTTTTGTGGGCTGCTGCAGTAATTTTGACGAAGTTGTCTC

      ArgMetThrLysGluSerTrpIleGluLeuArgAsnThrArgGlyGlnIleSerSerArg
1381 CAGAATGACCAAAGAGAGTTGGATTGAGCTTCGCAACACCCGAGGGCAAATATCGTCGCGT

      AlaSerLeuArgArgLysLysGlyAlaValArgAlaMetLysTyrIleLeuSerThrAsn
1442 GCTTCCTTGCGAAGGAAAAAGGGTGCAGTGAGGGCCATGAAGTACATTTTGTCAACAAAT

      ValGlyGluHisValProLysSerSerIleLeuArgHisTrpAsnGluTyrLeuLeuIle
1502 GTTGGGGAACATGTGCCGAAATCCTCAATTTTGCGGCACTGGAACGAATATCTTTTGATT

      LeuAlaArgAsnArgAspSerSerLeuProPheArgPheSerPheGlySerLysValArg
1562 CTGGCCCGGAACAGGGATTCCAGCCTACCCTTTCGCTTTTCGTTTGGGTCGAAAGTGCGG

      AlaLeuProLeuSerSerLeuProSerLeuAlaAlaHisOC
1622 GCATTACCGCTCTCTTCACTGCCGTCGTTAGCCGCGCACTAAAATTGTTTAAGCAAGGCA

1682 AAAAACCAATGGCGTACTCGTTTTATGATGCTCAGGAGGAGGCTGAGCCAGTTGTTGTGT
1742 TGAGGCTGCTACAAGTGAAGAAGACGAGTGCGTTGACAGCGGCGAGGTGCTTTCTTGTGA
1802 AGGGAGCGTTGAAAAATCGGCAAAGGAGAGTGAAAAGATCGCAGTAGGGACTGGGGAAGG
1862 AAACGATGATAATTCAGTGAGGCCGCCTTCCCTAGTGGAGCGCCTTCCTCGACTCGGCGG
1922 CTGTTCAGAGGATGTAATATCGCTTCTTTCGTTCTTGCGCAGTATTAATGCCAGTGGACT
1982 ATCTGGACGAGGATGGAGGGATGATTGGCGTTCCGCTGGTTACGGCGTTTGGTTGTATGC
2042 CTGCTTGATTGCTGTTGACACACCCTTTGACCCGGACTTGGACCGACTTGTTCATGAGTT
2102 TTTCCGTACTTGTTGCTTACAATTACGGGTACTCGGGGAGGCACACAACGTCCGCGGAGA
2162 TGATCGTGGGGCGATATTGAAAGTTTTCCCATCCCGCAAAGATGACCCCCCTCAGTCGTC
2222 ACAATACAACTCCCTTGGTGATGTGAGGAAGGAGGATGTGTTGGCACTCCACACCATTAT
2282 CGTTGTGCTAGCCAAGTTATTCCGCCTAAATGGTAATCGTTTGGTTCCGCTGTAAAAGGA
2342 TTTGAGGTACTGCAGTTACAACGTTCGAGTGGGCACTATACCAGATGGCAGTCGGGGCAT
2402 AGGTGAGAGCGACTGCTAATGAAACGGAAAGGAGAGTAGTTACTTCAACTATAAGGTTTC
2462 ATGGACAACGTTTGCACCTGGTTTCAGTGTCACTTACGTGCGGTGGGTTGCAGTTTGCTT
2522 TCCTCCTTTTCCGAGGATGTCAGCAGACTTTCATGTATGGTGTGCTACCTTCGGCTAAGA
2582 ATCTTGACGCTCATTGCCATCCAATGACTCCCAATGTCTGTGCGGACGACACATCCTAAC
2642 AAGGGAAACACGTTGTTATTTTCCCCTTTTAGGTTTGTTTAATTTGGGACCTTGGATCTC
2702 GAAACTCTAGGAGGACTTTTGAGCTCAAGTTTTATGCGAATGTGGTGATCTAAACACGCT
2762 TTGTTATCCTGTGGATGGGGGTATGGTATCGGTCAACTTGTGAAGAAATGGACCCCGGGT
2822 ACCGAGACTCGAATTCCCTATAGTGAGTCGTATTAAATTCGTAATCATGTCATAGCTGTT
2882 TCCTGTGTGAAATTGTTATCCCGTATACAATTCCACA
```

Figure 3b. Downstream sequence starting at the downstream BamHI site at position 10.3 on the map. The point of divergence with the intergenic region is marked |2. This sequence was obtained from overlapping clones but was not fully confirmed for both strands.

```
GCAGAGTAACAGATCGGCACGCCCATGCGCAACTACAACA
ACTTCAACCGCGTGTGGAAGGCCCCACGGCGACCCTTCGA
GAAGGAGCGTCTCGACCGTGAGATGAAGCTCTGCGGGCAG
TACGGTCTGCGCTGCAAGCGCGAGATTTGGCGCGTCAACA
TGACACTCTCAAAGATGCGTCGCACCGCGCGTCTTTTGCT
GACGCTTCCGGAAAACCACCCGCGCCGCCTTCTTGAGGGC
TCTGCAATTATGCGCCGCTGCCACGGATATGGTTTCCTCG
ACGAAGATAAGGACAAACTCGATTATGTGCTTTCGCTCAC
CGTTCCCGACATTCTCGAGCGCCGCCTTCAAACGGTCGTC
TTCAAGCATGGCCTCGCAAAGTCCGTTCACCACTCCCGTG
TTCTTATCCAACAGCGGCACATTGCTGTTGCTAAGCAAAT
CGTTACGATTCCTTCCTTTATCGTGCGCGTGAGCTCTGAG
CACCATATCGCCTTTGCGGATGCGTCGCCATTCGGCAATG
GTCGCCCAGGTCGCGTGAAGCGTGTGAAAAGGAACGCTGC
GAAGAAGGGTAGCGGTGGTGGTGATGATGACGAGTAAGCA
CCAAGGGGTCATATAGCGCTTGACAGAGAGAAGACGTGAC
GCGTTGGGCGACAAGTTGGACTCGCATGCCATCGGAATCC
CATTATGTGAACTTGCGCCTTTCCCCCATTTGTTTGTCTG
CTTTTTTCTTTTACAAAGAGGCAGCCACAGGGGTGAAAAG
GTGGGCAACAAAGTGGAGGTGATGGGGCGAGTGAGTGCAC
TTGGACAAACTGAAGAGCGAAAGAATAACCGAAAGCACGT
CGGGCAAGAGTAAGTCAGGTATGATGCTATTGGCGTGTGA
GGCAGTACTAGCGGTGATGCAGCCACTTTATAATTTTTTG
GGTGCCTTCCCCATTTTTTGTGCACGTTTCTTGTGCTTCT
TTTTTTTTTTTTACTGTTGAAAAAAAAAA
```

Figure 4. Sequence of the Ub cDNA. The initiation and termination codons are double underlined. Nucleotides that differ from Ua are underlined.

acceptor sites at positions where there was no acceptor AG dinucleotide. We concluded that the low abundance of the RNA, possibly combined with some secondary structure that inhibited reverse transcriptase, might be the problem. To facilitate detection of the full length primer extension product, we amplified it by the polymerase chain reaction, using a 35 nt mini-exon primer for the 5′-end. Results are shown in figure 6. One clear 184 nt product was seen from primer 8, no equivalent of which was ever seen using reverse transcriptase alone. This would comigrate with nt 4286 on a sequencing ladder, and predicts mini-exon addition at 4321. There is an AG at 4318−9. As spurious bands are often seen in PCR reactions, we confirmed the identity of the 184 nt product by blot hybridisation: it, but none of the other bands on the gel, hybridises specifically with a ^{32}P-end-labelled oligonucleotide 2 probe (data not shown). This result indicates that the intergenic RNA splice acceptor is at nt 4319, and limits the distance between the polyadenylation site of aldolase and the start of the intergenic mRNA to 193nt. This start site predicts an mRNA of 836nt, not including polyadenylation, so is in good agreement with the size estimate by gel electrophoresis.

The presence of the mini-exon on the Ua and aldolase transcripts was also checked by PCR (Figure 6). The aldolase product (using oligo 11 and mini-exon) was 105 nt as opposed to the predicted 108; that for the upstream transcript was 55 nt (predicted 56).

Laird (21) has described a general sequence motif that occurs 5′ to splice acceptor sites in *T. brucei*. A computer search found the concensus at appropriate positions near all the open reading frames (Table 1). As in the other cases studied, the distances between the various conserved parts of the motif are highly variable.

ORF	predicted protein			transcript		
	length (aa)	M_r (kDa)	net charge	mature trancripts	splice acceptor site TTTCPy-$(T)_n$-aAG	
V	232	25.6	+3	?	-20	-4
U	190	22.0	+19	Ua, Ub	-66	-10
ALD	372	41.0	+10	ALD	-168/-156	-74
I	185	21.7	+6	I	-172	-25
D	186	20.6	+6	?	-55/-28	-15

Table 1. A summary of the properties of the open reading frames around the aldolase genes of *T. brucei*. For the calculation of net charge of the predicted protein, Asp and Glu are taken as −1, Lys and Arg as +1 and all other amino acids as zero. Assignment of splice acceptor sites is based on data in the text (for U, ALD and I) and on a computer search for the concensus sequence postulated by Laird [21]: TTTCPy-$(T)_n$-aAG*$(A)_n$in which $(T)_n$ indicates a variable length of T-rich sequences, $(A)_n$ a variable length of A-rich sequences and the asterisk the splice point. The positions of the nucleotides underlined in this putative splice acceptor sequence are given relative to the start codon.

Measurement of transcription rate

The rate of transcription across the aldolase locus was measured. Small clones (Figure 1 and Table 2), were hybridized to [^{32}P]-labelled RNA synthesised in isolated nuclei. With the exception of the clones including region U, of which there was one other homologous copy elsewhere in the genome (the Ub gene), all the clones used hybridised only to the aldolase locus. Various other transcribed regions were included as controls; no hybridization was seen to vector DNA. To compare results from different preparations, autoradiograms were exposed so as to give similar signals from ribosomal and tubulin DNAs. Very little difference (i.e. not more than about two-fold) in aldolase transcription was ever seen when bloodstream and procyclic nuclei were compared. This was true whether the bloodstream nuclei were allowed to elongate at 27°C or 37°C. Varying the potassium concentration between 20 and 100mM had little effect on aldolase transcriptionalthough 100mM potassium appeared to inhibit ribosomal RNA synthesis. As there is at least six times more steady-state aldolase mRNA in bloodstream forms as in procyclic forms (5)(e.g Figure 2) most regulation must be post-transcriptional.

The close packing of exons in the aldolase locus meant that transcriptional analysis of large restriction fragments could not reveal discontinuities in transcription. To test for continuity of transcription, we therefore hybridized nascent RNA to very short cloned DNAs (Figure 7b). This approach enabled us to map transcriptional activity of both exons and intergenic regions in great detail, but the use of very small clones inevitably led to very faint hybridization signals. However, by using these very small genomic fragments we were able to confirm that all the sequences concerned appear to be transcribed, whether or not they are represented in mature mRNA. Quantitation of the hybridization by scintillation counting proved impossible. From scanning the autoradiograms from several independent experiments, results for the larger clones (U25, U23, U10, U11, I12, I24, CALD) indicated that the apparent rate of transcription varied no more than 2.5-fold over the region from 0kb to 8kb on the map in Figure 1. Given the technical limitations of the experiments, these results are consistent with transcription of the

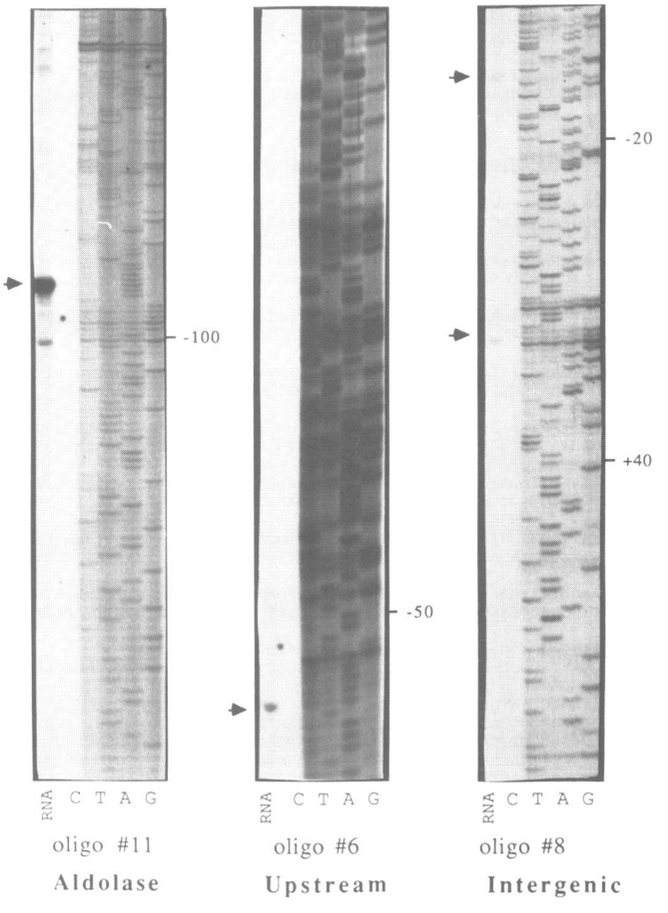

Figure 5. Primer extension analysis. Primers (50ng) were annealed to polyadenylated bloodstream trypanosome RNA (5μg) and elongated using reverse transcriptase. Sequencing reactions using the same primers on supercoiled cloned template DNAs served as markers. All reactions were run on the same 8% sequencing gel. The oligonucleotides used are indicated below each panel together with the nucleotide labels for sequencing lanes. The positions on the sequence relative to the ATG (A = +1) for each open reading frame are indicated next to the sequence.

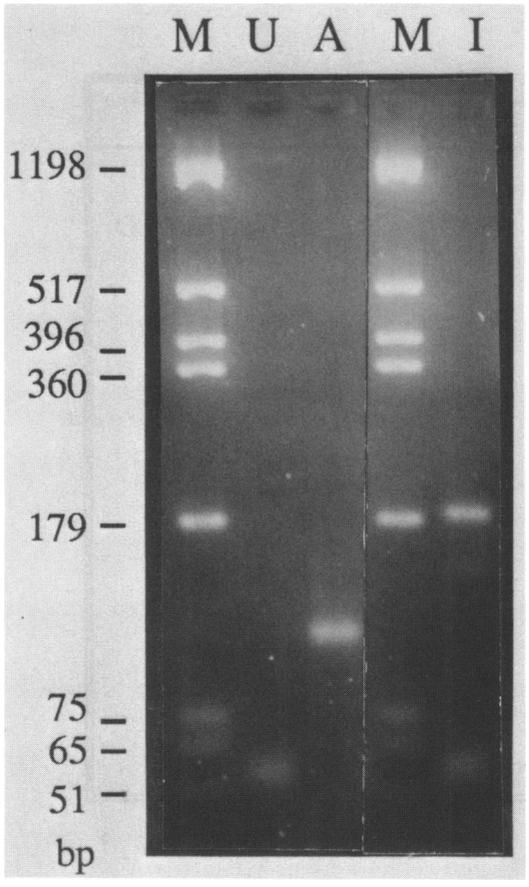

Figure 6. Definition of 5'-ends by the polymerase chain reaction. Primer extension products (as in Figure 5) were amplified using a sense-strand mini-exon primer and the products run on a 3.5% Nu-Sieve agarose gel with restriction-digested plasmid markers and visualized with ethidium bromide. Products for upstream (U), aldolase (A) and intergenic (I) transcripts are shown.

whole aldolase locus as a poly-cistronic precursor, with subsequent post-transcriptional regulation.

DISCUSSION

To understand how trypanosomes regulate their energy metabolism throughout the life cycle it is clearly necessary to know how they control the levels of cytoplasmic mRNA. In trypanosomes, as in other organisms, steady-state mRNA levels could be influenced not only by the rates of mRNA transcription and degradation, but also by the efficiency of splicing, polyadenylation and export from the nucleus.

The trypanosome genes whose transcription has been investigated most thoroughly encode the variant surface glycoproteins. Various groups have studied the transcription of VSG genes and the regions upstream in nuclei isolated from trypanosomes after nitrogen cavitation. In general, transcription appears to progress at a constant rate over 20−60kb upstream of the telomeric VSG genes and, like higher eucaryotic polymerase I transcription, is resistant to 1mg/ml alpha-amanitin. Examples include variant 221 (60kb) (22); variant 1.8 (27kb) (23), and variant 117 (at least 40kb) (24). Pays *et al* (25) and Johnson *et al.* (26) tried inactivating transcription with uv irradiation before isolating the nuclei; results confirmed the estimate for the size of the 221 transcription unit and suggested that that of the AnTat 1.3A gene is 47kb long. In each case, the long transcripts appear to be precursors for several mRNAs of low abundance in addition to the very abundant VSG transcript, indicating that regulation of relative mRNA levels is post-transcriptional. Nevertheless, the transcription of VSG loci shows strong developmental regulation, being undetectable in procyclic parasites.

The genes encoding glycosomal phosphoglycerate kinase (whose mRNA is abundant in bloodstream forms) and the cytoplasmic isozyme (more abundant in procyclics) are only separated by 300 base-pairs, and are preceded by a closely-related gene that is expressed as mRNA at a low level throughout the life cycle (9). However, analyses of transcription in isolated nuclei indicate that the rate of transcription is constant across all three genes (12). Putative precursor RNAs could also be detected by blot hybridization of total RNA; the presence of transcripts spanning the intergenic regions was shown by S1 nuclease protection experiments. The authors concluded that regulation of the steady-state levels of mRNA is effected post-transcriptionally.

The *T. gambiense* calmodulin genes are found in tandem arrays of three or four genes.(11). Tschui and Ullu (11) were able to detect trancripts that appeared to contain at least two gene copies by blot hybridization and by RNAase and SI nuclease protection analysis using probes from the intergenic region and from regions represented in mature RNA. They also observed a constant rate of transcription across the whole region in permeabilised cells. Possible high molecular weight precursor RNAs have also been seen for a number of other trypanosomatid RNAs (e.g. (27, 28))

Muhich and Boothroyd (29) took advantage of the fact that, as in *Drosophila* cells (30), RNA splicing in trypanosomes is inhibited by heat shock. They found that heat shock causes accumulation of multimeric precursors, detected by both blot hybridization and RNAase protection, originating from the alternating alpha−beta tubulin tandem repeat. Even though these precursors were of much higher abundance than those found from other genes or in normal cells, they could not be chased into mature RNA. In heat-shocked *Drosophila* cells, at least some precursor RNAs are exported from the nucleus, precluding subsequent splicing and analysis of a precursor-product relationship (31). Possibly the same thing happens in trypanosomes.

The evidence for polycistronic transcription in trypanosomes is all indirect. A constant rate of transcription across several genes and their intergenic regions has been seen in a number of cases, and low-abundance RNAs traversing the region between one cistron and the next have been detected. Termination of eucaryotic RNA polymerase II usually occurs downstream of the polyadenylation site (32). Trypanosome genes are so closely packed that it is quite conceivable that all the putative precursor RNAs could be aberrant species resulting from inefficient termination. The results of u.v. inactivation experiments are more convincing but the technique is unfortunately only appropriate for abundant transcripts.

Our attempts to measure the rate of transcription across the aldolase locus were only partially successful. In order to distinguish between transcription of intergenic regions and of exons we had to use very small segments of cloned genomic DNA as probes. Transcription of the whole region is not very active so quantitation was extremely difficult. For the smallest clones the hybridization was barely above background, even for

Table 2. List of clones used in transcription analysis.

Clone	Site 1		Site 2		length (nt)	copy no	transcripts
	enzyme	position	enzyme	position			
U1	EcoRI	274	XhoI	1771	1497	1	Ua,Ub
U2	XhoI	1771	XhoI	2710	939	2-3	Ua,Ub,ALD
U4	XhoI	1771	MluI	1901	403	2-3	Ua,Ub
U10	EcoRI	274	DraI	1108	834	1	-
U11	DraI	1108	XhoI	1771	663	2	Ua,Ub
U12	TaqI	468	TaqI	902	434	1	-
U13	TaqI	902	TaqI	1125	223	1	-
U16	TaqI	1125	TaqI	1345	220	1	-
U17	Taq1	1358	TaqI	1514	156	1-2	Ua,Ub
U23	EcoRI	~-1230	EcoRI	274	~1500	?	?
U25	-2230	EcoRI	~-1230	-	~1000	?	?
cALD		2009	StuI	3820	1811	2	ALD
I1	SalI	4888	SalI	5398	520	2	I
I2	BamHI	4445	SalI	4888	443	2	I
I12	BamHI	4445	HindIII	5081	636	2	I
I13	StuI	4093	BamHI	4445	352	2	ALD,I
I24	SalI	5398	RsaI	5830	422	1	

sequences represented in mature mRNA (e.g. U17 in figure 7B). We could nevertheless conclude that the level of transcription across 9kb of the aldolase transcription unit was approximately constant, despite the fact that the steady-state levels of the product mRNAs differed markedly. Also it is clear that intergenic regions were transcribed (e.g. I24 in Figure 7B). In preliminary experiments, polyadenylated RNAs that appeared to contain aldolase and intergenic sequences and were longer than the mature mRNAs were found in heat-shocked bloodstream trypanosomes (C. Clayton, unpublished results), but the latter result was poorly reproducible and no putative precursors could be detected by blot hybridization of normal RNA. Attempts to demonstrate such precursors by PCR have so far been foiled by our inability completely to remove genomic DNA from our RNA preparations. As DNA transformation has now been established in *T. brucei* (C. Clayton and J. Fueri, manuscript submitted) we have suspended our search for precursors in favour of greater efforts to analyse transcriptional control regions *in vivo*.

Supposing that aldolase transcription is in fact polycistronic, control of mRNA levels must be by regulation of the rates of polyadenylation, trans-splicing, export from the nucleus and degradation. Ample precedent exists for post-transcriptional regulation of RNA levels in eucaryotes: regulation of *cis*-RNA splicing in eucaryotes is well documented (33), as are controls of mRNA stability (e.g. (34)) which can be influenced in higher eucaryotes by sequences in the 3'-untranslated region (e.g. (35, 36)) and of transport from the nucleus (e.g. (37)). The close packing of the genes under investigation implies that all regulatory sequences for aldolase must be in a very small, defined region. Any signals for regulation of trans-splicing must be either in the 180nt between the Ua polyadenylation site and the aldolase splice acceptor site, or in the mature mRNA sequence. Similarly, any regulation of polyadenylation must come from either the

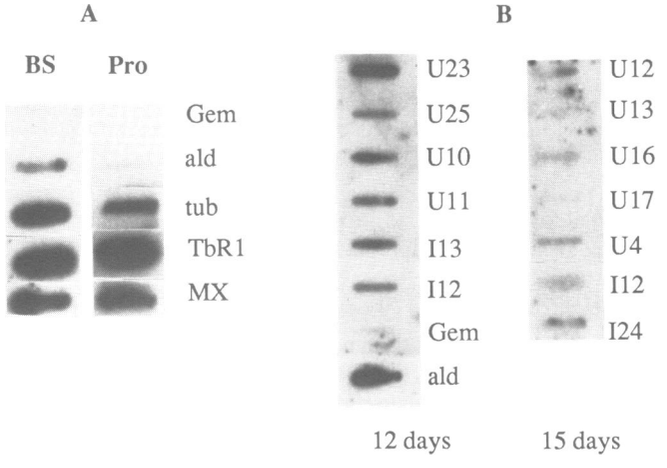

Figure 7. Aldolase transcription in isolated nuclei. A) Radioactive RNA made by bloodstream (BS) or procyclic (Pro) nuclei was hybridized at 10^7 cpm/ml (0.5ml) to a variety of cloned trypanosome genes for about 40h, except for the MX slots which were hybridized at 10^6 cpm/ml in 0.1ml. Blots were treated with RNAase (100μg/ml) and washed at 65°C with 0.5×SSC before autoradiography using pre-flashed film at −70°C for 64 hours. Clone names beside the slots are: MX: mini-exon; rRNA: TbR1; tub: alpha and beta tubulin; ald: aldolase cDNA; Gem: vector control. The exposure time was 16h. B) As in A except that bloodstream nuclei were used throughout; exposure times are shown beneath in parentheses.

3'-untranslated region (which should also include any sequences responsible for controlling mRNA stability) or the 193 nt gap between the aldolase and intergenic transcripts. Another approach which can give a hint of the location of promoters and control regions is to look for conserved consensus sequences in the

regions upstream of a number of different genes. This has not yet yielded much in trypanosomatids; although Glass *et al.* (38) suggested that the regions between the trypanosome HSP70 genes might contain heat shock promoters based on homology with such promoters in other species, such a function has yet to be demonstrated convincingly.

The only detailed analysis of mRNA turnover in *T. brucei* is that published by Ehlers *et al.* (39) who reported that the turnover of VSG mRNA was accelerated in bloodstream trypanosomes that were transforming into procyclic forms whereas tubulin mRNA was stabilized. The aldolase transcripts are an order of magnitude less abundant than these, so similar experiments, together with analyses of regulation of trans-splicing or polyadenylation will be much easier now we can increase the abundance of the RNA by substituting a more active promoter and reintroducing a chimaeric gene by DNA transfection. Another possible level of regulation, at the level of translation or organelle assembly (40) also remains to be investigated: aldolase protein is regulated 30-fold whereas the mRNA regulation is about 6-fold.

The inhibition of aldolase transcription by low levels of alpha amanitin suggests that it is transcribed by RNA polymerase II. It is therefore the first polymerase II gene to be shown to be interspersed with completely unrelated transcripts. We cannot be sure, though, that they really are unrelated until the encoded proteins have been located. The U and I proteins would have no signal peptide for membrane insertion, and the U transcripts are found on free polysomes (C. Clayton, unpublished results); neither U nor I protein has any detectable homology with anything in the existing databases. Most glycosomal enzymes have rather basic isoelectric points (41). It is intriguing that the predicted V, U, I and D proteins share this property; perhaps they are previously unidentified glycosomal proteins.

ACKNOWLEDGEMENTS

We wish to thank Dr. P. Hevezi (then at RU) for the mini-exon and alpha tubulin clones and Dr. D. Cambell (UCLA) for ribosomal clone TbRI. Dr. T. Chen (Rockefeller University) allowed us to use her laser densitometer. We thank Greg Wisdom and Eliot Davidowitz for technical assistance, members of the Cross lab (RU) and Tropical Medicine research unit (Brussels) for discussions and Dr. F.R. Opperdoes and Professor G.A.M. Cross for their advice and support. This work received financial support from NIH grant AI22229 (CC), a Hirschl Weil-Caulier award (CC) and the UNDP/World Bank/WHO special programme for research and training in tropical diseases (PAMM).

REFERENCES

1. F. R. Opperdoes, *Ann. Rev. Microbiol.* **41**, 127–151 (1987).
2. A. H. Fairlamb, F. R. Opperdoes, in *'Carbohydrate metabolism in cultured cells'* M. J. Morgan, Eds. (Plenum publishing corporation, 1986),
3. R. A. Aman, C. C. Wang, *Mol. Biochem. Parasit.* **21**, 211–220 (1986).
4. D. T. Hart, O. Misset, S. W. Edwards, F. R. Opperdoes, *Mol. Biochem. Parasit.* **12**, 25–36 (1984).
5. C. E. Clayton, *EMBO J.* **4**, 2997–3003 (1985).
6. M. Marchand, *et al.*, *Mol. Biochem. Parasit.* **29**, 65–76 (1988).
7. C. E. Clayton, in *Genetic Engineering* P. W. J. Rigby, Eds. (Academic Press, Ltd., 1988), pp. 2–56.
8. P. A. M. Michels, *et al.*, *EMBO J.* **5**, 1049–1056 (1986).
9. K. A. Osinga, *et al.*, *EMBO J.* **4**, 3811–3817 (1985).
10. M. R. Mowatt, C. E. Clayton, *Mol. Cell. Biol.* **7**, 2838–2844 (1987).
11. C. Tschudi, E. Ullu, *EMBO J.* **7**, 455–463 (1988).
12. W. C. Gibson, B. W. Swinkels, P. Borst, *J. Mol. Biol.* **201**, 315–325 (1988).
13. P. Borst, *Ann. Rev. Biochem.* **55**, 701–732 (1986).
14. M. R. Mowatt, C. E. Clayton, *Mol. Cell. Biol.* **8**, 4055–4062 (1988).
15. R. K. Saiki, *et al.*, *Science* **239**, 487–491 (1988).
16. V. Bellofatto, R. Cooper, G. A. M. Cross, *Nucleic Acids Res.* **16**, 7437–7456 (1988).
17. J. M. Kooter, P. Borst, *Nucleic Acids Res.* **12**, 9457–9472 (1984).
18. R. E. Sutton, J. C. Boothroyd, *Cell* **47**, 527–535. (1986).
19. L. H. T. Van der Ploeg, *Cell* **47**, 479–480. (1986).
20. M. S. Freistadt, G. A. M. Cross, H. D. Robertson, *J. Biol. Chem.* **263**, 15071–15075 (1988).
21. P. Laird, *PhD thesis, University of Amsterdam* 1988).
22. J. M. Kooter, *et al.*, *Cell* **51**, 261–272 (1986).
23. C. Shea, L. H. T. Van der Ploeg, *Mol. Cell. Biol.* **8**, 854–859 (1988).
24. C. P. Gibbs, G. A. M. Cross, *Mol. Biochem. Parasit.* **28**, 197–206 (1988).
25. E. Pays, *et al.*, *Cell* **57**, 835–845 (1989).
26. P. J. Johnson, J. M. Kooter, P. Borst, *Cell* **51**, 273–281. (1987).
27. V. Bellofatto, G. A. M. Cross, *Nucleic Acids Res.* **16**, 3455–3469 (1988).
28. M. F. Ben Amar, *et al.*, *Mol. Cell. Biol.* **8**, 2166–2176 (1988).
29. M. Muhich, J. C. Boothroyd, *Mol. Cell. Biol.* **8**, 3837–3846 (1988).
30. H. J. Yost, S. Lindquist, *Cell* **45**, 185–193 (1986).
31. H. J. Yost, S. Linquist, *Science* **242**, 1544–1548 (1989).
32. J. R. Nevins, *Ann. Rev. Biochem.* **52**, 441–466. (1983.).
33. R. E. Breitbart, A. Andreadis, B. Nadal-Ginard, *Ann. Rev. Biochem.* **56**, 467–495 1987).
34. H. D. Love, A. Allen-Nash, Q. Zhao, G. A. Bannon, *Mol. Cell. Biol.* **8**, 427–432. 1988).
35. T. R. Jones, M. D. Cole, *Mol. Cell. Biol.* **12**, 4513–4521 (1987).
36. G. Shaw, R. Kamen, *Cell* **6**, 659–667 (1986).
37. M. Moore, J. Schaack, S. B. Baum, R. I. Morimoto, T. Schenk, *Mol. Cell. Biol.* **12**, 4505–4512 (1987).
38. D. J. Glass, R. I. Polvere, L. H. T. van der Ploeg, *Mol. Cell. Biol.* **6**, (1986).
39. B. Ehlers, J. Czichos, P. Overath, *Mol. Cell. Biol.* **7**, 1242–1249 (1987).
40. A. F. Torri, S. L. Hajduk, *Mol. Cell. Biol.* **8**, 4625–4633 (1988).
41. O. Misset, O. J. M. Bos, F. R. Opperdoes, *Eur J. Biochem.* **157**, 441–453 (1986).