© 1990 Oxford University Press

Developmentally regulated transporter in *Leishmania* is encoded by a family of clustered genes

David A.Stein, Bradley R.Cairns+ and Scott M.Landfear*
Department of Microbiology and Immunology and Vollum Institute for Advanced Biomedical Research, The Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, OR 97201, USA

Received November 14, 1989; Revised and Accepted February 8, 1990 GenBank accession no. M26229

ABSTRACT

We have previously cloned a gene for a developmentally regulated transport protein from the trypanosomatid protozoan *Leishmania enriettii*. We demonstrate here that this transporter is encoded by a single family of tandemly clustered genes containing approximately 8 copies of the 3.6 kilobase repeat unit. Transcriptional mapping defines a contiguous 3.3 kilobase region of the repeat unit that encodes the mRNA. The 5′ end of the mature mRNA contains the spliced leader or mini-exon previously identified in kinetoplastid protozoa, while the 3′ ends of the mRNA are heterogeneous in sequence and in location of the polyadenylation site. We have identified genomic restriction fragments that flank the tandem repeat on the 5′ and 3′ sides and which may be linked to sequences required for expression of the gene family. Other species of *Leishmania* also contain sequences that hybridize to the cloned *L. enriettii* gene at high stringency.

INTRODUCTION

The trypanosomatid protozoa are unicellular eukaryotes that exhibit many unusual features of gene expression, including processing of mRNA *via trans*-splicing (1,2), transcriptional editing of mitochondrial mRNAs (3,4), and probably transcription of large polycistronic precursors of mature mRNAs (5–8). Consequently, studies on these organisms have generated new insights into mechanisms for expressing genetic material. However, very little is known about the regulation of gene expression in these protozoa. Many trypanosomatids possess multiple morphologically distinct stages in their life cycles and must alter the level of expression of certain genes during the transitions between these different life cycle stages. To understand how this regulation is acheived, it will be necessary to identify and clone stage-regulated genes from these protozoa.

Leishmania are parasitic trypanosomatids that have two life cycle stages. Promastigotes are flagellated forms of the protozoan that colonize the gut of the sandfly vector, and amastigotes are non-flagellated forms that are specialized for growth and survival inside the lysosomal vesicles of the vertebrate host macrophages (9). We have cloned a gene from *Leishmania enriettii* whose mRNA is much more abundant in promastigotes compared to amastigotes (10). This gene, designated Pro-1, encodes a protein whose sequence is similar to several known integral membrane transporters, including mammalian glucose and bacterial sugar transporters. This sequence similarity suggests that the Pro-1 gene product is also a transporter and may be required for the intracellular accumulation of a ligand that is utilized in the insect stage of the life cycle. To study how this gene is regulated during the *Leishmania* life cycle, we have investigated the structure of the Pro-1 gene and its arrangement within the parasite genome. We show here that multiple Pro-1 genes are clustered in a single tandem repeat, and we have determined the regions of this repeat which encode the mature mRNA. This mRNA contains at its 5′ end the spliced leader or mini-exon (11), which is apparently *trans*-spliced onto trypanosomatid mRNAs (1,2), and displays at least two polyadenylation sites. We have also identified genomic restriction fragments that flank the 5′ and 3′ ends of the tandem repeat. These flanking fragments could contain or be linked to sequences required for the transcription or regulation of the Pro-1 genes.

MATERIALS AND METHODS

Cells and nucleic acids

Promastigotes of *Leishmania enriettii* were grown as previously described (12). Isolation of genomic DNA and of RNA and methods for Southern and Northern blotting were performed as described (12). Inserts of plasmid clones were radiolabelled by the oligonucleotide priming method (13). Hybridizations of blots were performed at 42°C for at least 24 hr in 50% formamide, 5× SSC (1× SSC is 0.15 M NaCl, 0.015 M Na citrate, pH 7.0), 10× Denhardt's solution, 20 mM sodium phosphate pH 6.5, 0.1% SDS, 250 μg/ml denatured herring sperm DNA. Filters were washed twice at room temperature and then three times at 55°C in 0.1× SSC, 0.1% SDS.

* To whom correspondence should be addressed

+ Present address: Department of Biochemistry, Stanford University, Stanford, CA 94305, USA

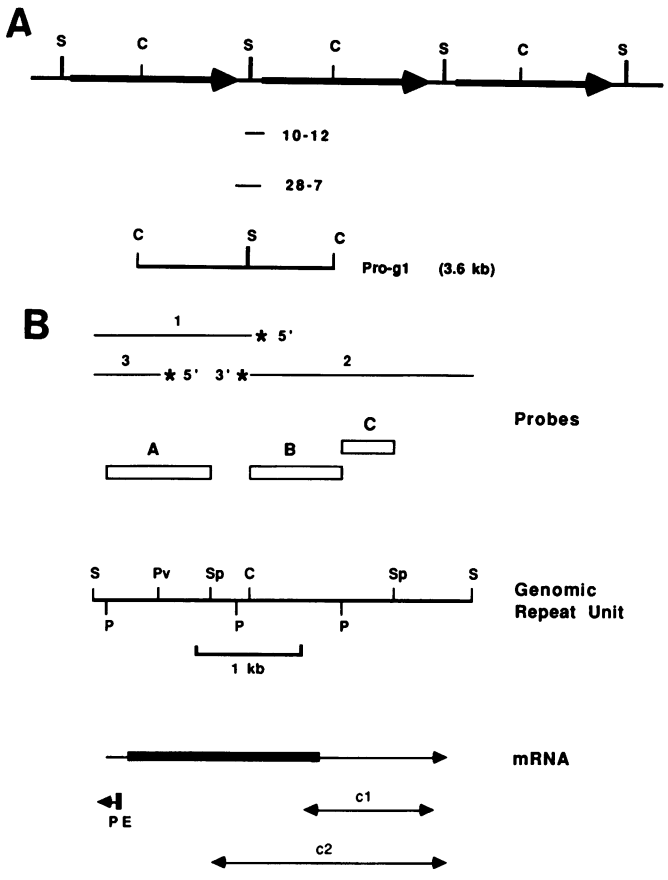

Figure 1. Restriction and transcription maps of the Pro-1 genomic repeat. (A) Tandem repeats of Pro-1 genes. Three repeat units of the Pro-1 tandem cluster are illustrated. The single *ClaI* (C) and *ScaI* (S) sites within each unit are separated by 3.6 kilobases. The heavy arrow indicates the region of the genomic DNA that encodes the Pro-1 mRNA, as determined by transcriptional mapping (see text). The Pro-g1 clone that contains a 3.6 kilobase *ClaI* fragment of the genomic repeat is illustrated below. Subclones 10-12 (230 base pairs) and 28-7 (380 base pairs) were generated by exonuclease III deletion of a subclone of Pro-g1 and were used to determine the mRNA coding and non-coding regions of the Pro-1 repeat (see text). (B) Map of the repeat unit. A restriction map (middle) of the Pro-1 repeat unit (3.6 kilobases) indicates the locations of *ScaI* (S), *ClaI* (C), *PstI* (P), *PvuII* (Pv) and *SphI* (Sp) sites within the genomic DNA. The map position of the Pro-1 mRNA is displayed below the restriction map (line with a single rightward pointing arrowhead). The thick bar internal to the mRNA is the protein coding region of the message, as determined by DNA sequencing (10). The double arrows (bottom) indicate the map positions of the cDNA clones c1 and c2. The heavy vertical bar marked PE is the oligonucleotide probe used for primer extension sequencing of the Pro-1 mRNA, and the leftward pointing arrowhead indicates the direction of primer extension. The locations of probes used for S1 mapping (5′ end-labelled probes 1 and 3 and 3′ end-labelled probe 2), and for Southern blotting (probes A, B and C) are indicated at the top.

cDNA library construction

A λgt10 cDNA library (14) was prepared from 10 μg of promastigote polyadenylated RNA by the method of Gubler and Hoffman (15), as detailed in the Invitrogen (San Diego, CA) Librarian X instruction manual.

S1 mapping and primer extension sequencing

S1 mapping using 5′ and 3′ end labelled probes was performed as described (12) using a hybridization temperature of 57°C. Protected fragments were resolved on either alkaline 1% agarose gels (16) or on 5% acrylamide/8M urea gels (17).

For primer extension sequencing, 10 pmoles of oligonucleotide primer was end labelled with γ^{32}P-ATP and T4 polynucleotide kinase (16) and hybridized to 15 μg of promastigote polyadenylated RNA in 200 mM Tris pH 8.3, 0.25 M KCl by incubating at 80°C for 3 minutes and then at 45°C for 30 minutes. The hybrid was then mixed with 11 units of avian myoblastosis virus reverse transcriptase (Life Sciences, St. Petersburg, FL) in a final buffer composition of 50 mM Tris pH 8.3, 100 mM KCl, 10 mM $MgCl_2$, 10 mM DTT and divided into four tubes containing 555 μM of each deoxynucleoside triphosphate and either 22 μM dideoxy GTP, 33 μM dideoxy ATP, 55 μM dideoxy TTP, or 33 μM dideoxy CTP. Primer extension was performed at 45°C for 30 minutes, and the extension products were ethanol precipitated and separated on a 6% acrylamide/8M urea sequencing gel.

Densitometry of autoradiograms

To estimate the copy number of the Pro-1 genes, Southern blots were exposed to pre-fogged Kodak XAR-5 film in the presence of a Dupont Cronex Lightning Plus intensifying screen. Autoradiograms were scanned with a Bio Rad model 620 video densitometer, and the optical densities underneath peaks were automatically integrated using the Bio Rad 1-D Analyst software and an IBM AT personal computer. Peak heights varied from 0.16 to 1.65 O.D. units, depending upon the scan and the peak integrated.

Orthogonal field alternating gel electrophoresis (OFAGE)

Cell blocks containing *L. enriettii* chromosomes were prepared by washing 4×10^7 cells per block in phosphate buffered saline (PBS) followed by resuspension in 50 μl PBS. Cells were mixed with an equal volume of 1% In Cert agarose (FMC Bio Products, Rockland, ME) and poured into a mold. Blocks were incubated at 55°C for 48 hours in 0.5 M EDTA pH 8.0, 1% sarkosyl, 2 mg/ml proteinase K, rinsed with 10 mM Tris pH 8.0, 1 mM EDTA and stored at 4°C. One half block was used per lane of the gel.

Electrophoresis was perfromed using a rotating gel apparatus (18) designed by Mr. Steven Smith (Los Lunas, NM). Gels were run at 15°C for 36 hours at 5 volts/cm with a 90 second dwell time. Following electrophoresis, gels were stained with ethidium bromide (1 μg/ml) and then incubated in 0.25 N HCl for 30 minutes followed by two 30 minute incubations in 0.5 N NaOH, 1.5 M NaCl (denaturing solution). Gels were blotted onto nylon filters using denaturing solution for the transfer.

Yeast chromosomal size markers (19) were prepared from strain CYY1 of *Saccharomyces cerevisiae* , a leu2 ura3 derivative of strain X2180A.

RESULTS

Pro-1 genes are clustered in a single tandem repeat

To probe the molecular mechanisms involved in regulating the expression of the Pro-1 gene during the parasite life cycle, it will first be necessary to determine the detailed structure of the gene. The sequence of the previously described (10) Pro-g1 genomic clone suggests that this gene is clustered in tandem repeats in the *Leishmania enriettii* genome. This Pro-g1 clone contains a 3.6 kilobase *ClaI* fragment of genomic DNA (Fig. 1A). A single translational reading frame, encoding a protein homologous to other known transport proteins, can be assembled from the sequence if the 5′ terminal *ClaI* site is joined with the 3′ terminal *ClaI* site of the cloned insert (Fig. 1A,B). Such a circularly permuted coding sequence would occur if the genes encoding the protein were arranged within the genome as tandem repeats

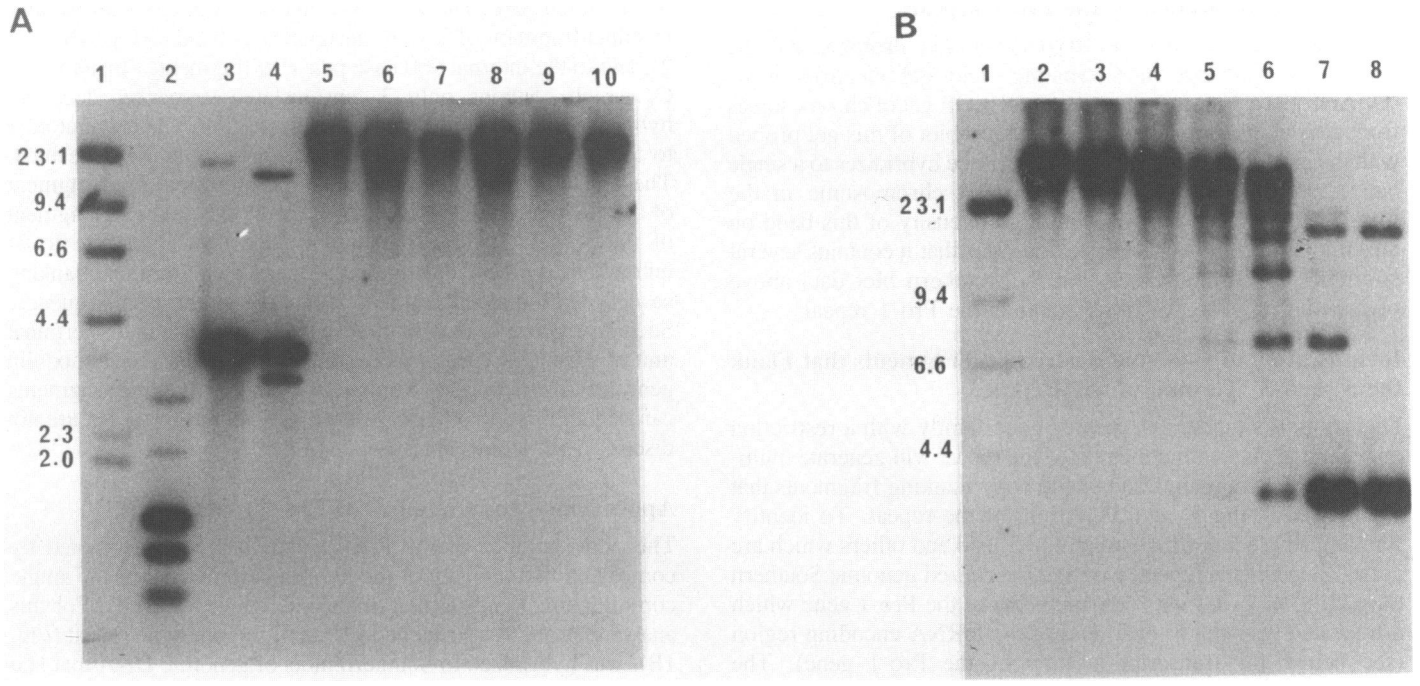

Figure 2. Genomic arrangement of Pro-1 genes. (A) 2 μg of genomic DNA from *L. enriettii* was digested with *PstI* (lane 2). *ClaI* (lane 3), *ScaI* (lane 4), *BamHI* (lane 5), *EcoRI* (lane 6), *HindIII* (lane 7), *SalI* (lane 8), *KpnI* (lane 9), and *SmaI* (lane 10), separated on a 1% agarose gel, and blotted onto a nylon membrane. The filter was hybridized to the radiolabelled insert of the Pro-g1 genomic clone (Fig. 1). Lane 1 contains 200 ng of *HindIII* digested λ DNA molecular weight markers, which were visualized by including labelled λ DNA in the hybridization mixture. The size of the molecular weight markers in kilobases is indicated at the left of the figure. The size of the major hybridizing band in the *ClaI* and *ScaI* digestions (lanes 3 and 4) is 3.6 kilobases. (B) 2 μg of genomic DNA was digested with 0 units (lane 2), 0.0036 units (lane 3), 0.018 units (lane 4), 0.045 units (lane 5) 0.090 units (lane 6), 0.18 units (lane 7), and 0.45 units of *ClaI* at 37° for 1 hour followed by electrophoresis on a 0.7% agarose gel. The gel was blotted onto a nylon filter and hybridized with the radiolabelled insert of the Pro-c1 cDNA clone (10). The individual bands in the partial digest in lane 6 are separated by 3.6 kilobase units. The molecular weight markers (lane 1) are the same as in part A. The 20 kilobase band in the complete digest (lane 8) is a 5′ flanking fragment of the repeat (see text).

containing a single *ClaI* site within the protein coding region. Tandemly repeated genes appear to be common in the trypanosomatids (7).

To confirm this arrangement of Pro-1 transporter genes within the *Leishmania enriettii* genome, we digested promastigote DNA with various restriction enzymes and probed Southern blots of these digests (Fig. 2A) with the insert of the Pro-g1 genomic clone. Digestions with *ClaI* and *ScaI* (lanes 3 and 4, Fig. 2A), enzymes which cut once within the Pro-g1 insert (Fig. 1A,B), generated an intensely hybridizing band of 3.6 kilobases and one or two less intensely hybridizing bands. This is the result expected for a tandem repeat of Pro-1 genes, with the 3.6 kilobase band representing multiple copies of the repeat unit and the lighter bands representing single copy terminal fragments of the repeat. In contrast, digestion with *BamHI*, *EcoRI*, *HindIII*, *SalI*, *KpnI* and *SmaI* (lanes 5 through 10 respectively, Fig. 2A), enzymes which do not cut within the Pro-g1 insert, generated a single hybridizing band larger than the 23.1 kilobase molecular weight marker, containing the entire uncleaved tandem repeat. The absence of additional hybridizing bands in these null digestions (lanes 5 through 10, Fig. 2A) indicates the absence of non-linked orphon copies (20) of the Pro-1 gene. Furthermore, the presence of only two putative terminal fragments (*vide infra*) for the *PstI* and *ScaI* digests (lanes 2 and 4 respectively, Fig. 2A) indicates that all the Pro-1 genes are clustered in a single repeat, because multiple separate repeats would generate additional terminal fragments.

Further confirmation for tandem clustering of Pro-1 genes was achieved by partially digesting genomic DNA with *ClaI* (Fig. 2B), an enzyme which cuts once within the repeat unit. Appropriate digestion conditions (lane 6, Fig. 2B) generate a ladder of Pro-1 hybridizing fragments which are separated by 3.6 kilobases in size, the length of the repeat unit. This result is predicted for partial digestion of tandemly reiterated elements, because oligomers of the repeat unit will differ from each other in size by one unit of the repeat.

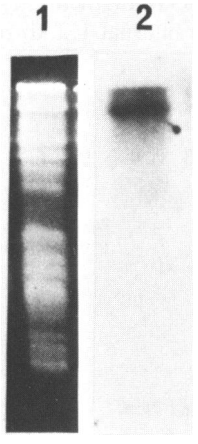

Figure 3. Chromosomal location of the Pro-1 repeat. Chromosomes from *L. enriettii* (lane 1) were separated by orthogonal field alternating gel electrophoresis, and the gel was stained with ethidium bromide. The gel was blotted and hybridized with the cloned Pro-1 probe (lane 2).

Chromosomal location of the Pro-1 repeat.

Individual chromosomes of trypanosomatid protozoa can be resolved by orthogonal alternating field gel electrophoresis (OFAGE)(21). Figure 3 displays an OFAGE gel of chromosomes from *Leishmania enriettii* and a Southern blot of this gel probed with the cloned Pro-1 gene. The Pro-1 probe hybridizes to a single band, which is above the largest yeast chromosome in the megabase region of the gel. The high intensity of this band on the ethidium bromide stained gel indicates that it contains several co-migrating chromosomes, but the Southern blot data above reveal that only one of these contains the Pro-1 repeat.

Identification of Genomic Restriction Fragments that Flank the 5′ and 3′ Termini of the Repeat.

Digestion of a tandemly repeated gene family with a restriction enzyme that cuts within the unit of the repeat will generate multi-copy internal fragments and single copy flanking fragments that are located at the 5′ and 3′ termini of the repeat. To identify flanking fragments which are at the 5′ end and others which are at the 3′ end of the repeat, we have hybridized genomic Southern blots (Fig. 4 A−C) with subfragments of the Pro-1 gene which are located near the 5′ or 3′ end of the mRNA encoding region (see below for transcription map of the Pro-1 gene). The restriction fragments used as probes are indicated in Figure 1B. Probe A is a 960 base pair *PstI-SphI* fragment that encodes 5′ untranslated and amino terminal protein coding regions of the Pro-1 gene. Probe B is a 820 base pair *ClaI-PstI* fragment that encodes carboxy terminal and 3′ untranslated regions of the gene. Probe C is a 460 base pair *PstI-SphI* fragment that encodes 3′ untranslated region of the Pro-1 gene.

A *PstI* digestion of genomic DNA (Fig. 2A, lane 2) generates three fragments that are internal to the repeat (1.5, 1.2 and 0.9 kb) and two terminal fragments of 2.75 and 2.1 kb. The 5′ probe A hybridizes to the 2.75 kb fragment, designated as band *a* (Fig. 4A, lane 2), and to the internal 1.2 kb fragment containing the 5′ end of the gene. The 3′ probe B hybridizes to the other putative terminal fragment of 2.1 kb, designated as band *b* (Fig. 4B, lane 2), and to the internal 900 base pair *PstI* fragment. Finally probe C, which contains only 3′ untranslated sequence, does not hybridize to either terminal fragment (Fig. 4C, lane 2) but only to the internal 1.5 kb *PstI* fragment from which it is derived. These results imply that band *a* is the 5′ terminal *PstI* fragment of the tandem repeat, and band *b* is the 3′ terminal *PstI* fragment of the repeat. Band *b* presumably does not hybridize to 3′ untranslated probe C due to divergence of the 3′ terminal flanking sequence of the repeat from the internal 3′ untranslated sequence. Such divergence of the 3′ untranslated sequence in the 3′ terminal unit of a tandem repeat has been demonstrated for the calmodulin genes of *T. brucei* (22). Similarly, 5′ and 3′ flanking fragments can be identified in other digests (Fig. 4, bands *c*−*e*) but are not discussed in detail here.

Approximate copy number of Pro-1 genes

The copy number of the Pro-1 genes has been estimated by comparing the intensity of the hybridization signal of the single copy flanking bands to that of the internal repeat units. For this analysis, probe A which encodes the 5′ region of the repeat (Fig. 1B), was hybridized to Southern blots of genomic DNA that had been digested with *ScaI*. Since *ScaI* cuts between protein coding units of the repeat (Fig. 1A,B) to release intact coding units, probe A hybridizes to both the 5′ and 3′ flanking *ScaI* fragments (Fig. 4A). Consequently, the intensity of the hybridization to the 3 kilobase flanking fragment (band *e*, Fig. 4A−C) could be compared to the intensity of the 3.6 kilobase internal repeats to estimate the copy number. This choice allowed us to compare bands of similar size and hence similar efficiency of transfer during the blotting process. Autoradiograms of a Southern blot containing 0.4 μg, 1.0 μg, 2.0 μg and 5.0 μg of *ScaI* digested genomic DNA were quantitated by densitometry, and the integrated intensities of the flanking and internal peaks were

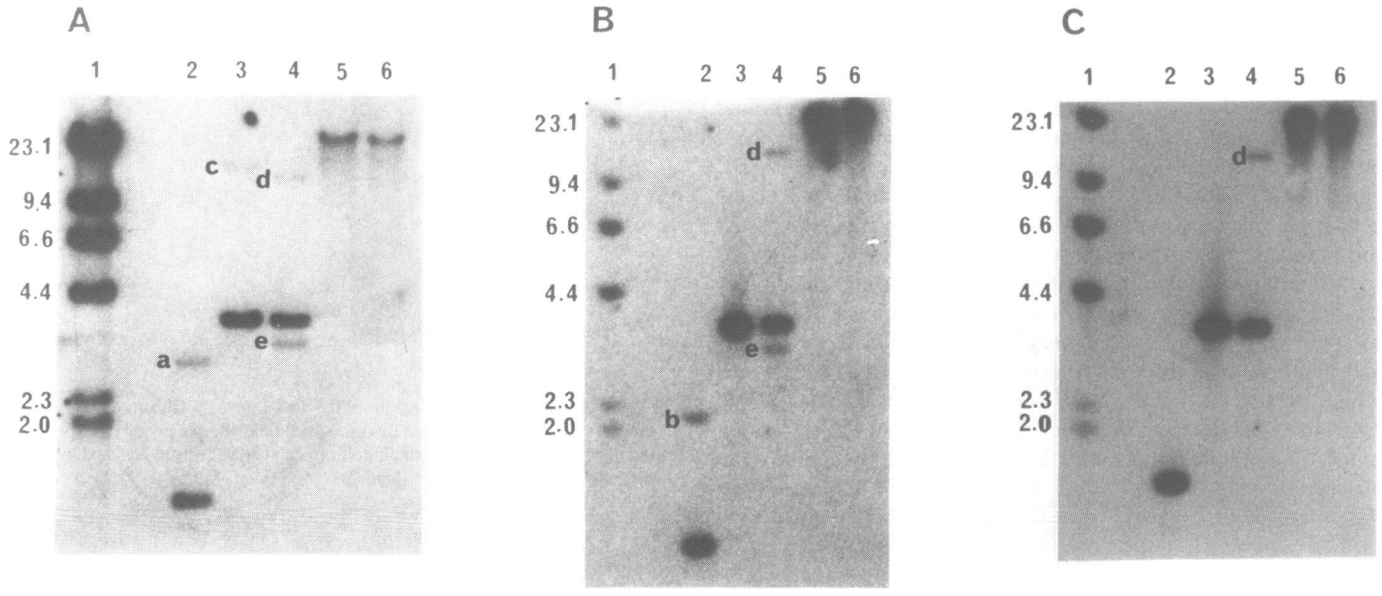

Figure 4. 5′ and 3′ flanking fragments of the Pro-1 genomic repeat. Genomic Southern blots were probed with restriction fragments of the Pro-g1 clone representing the 5′ and 3′ coding regions of the repeat unit. The probes used, shown in Figure 1, are: (A) Probe A, 5′ protein coding region; (B) Probe B, 3′ protein coding and untranslated region; (C) Probe C, 3′ untranslated region. The blots contained 2 μg of genomic DNA digested with *PstI* (lane 2), *ClaI* (lane 3), *ScaI* (lane 4), *EcoRI* (lane 5), *HindIII* (lane 6). Lane 1 contained 200 ng of *HindIII* digested λ DNA (sizes in kilobases indicated at left). Letters *a*−*e* refer to genomic restriction fragments that flank the tandem repeat, as discussed in the text.

compared. The ratios of these peak intensities range from 5.2 to 6.3, predicting a copy number of 7 to 8 when the two flanking fragments are added to the internal bands. Alternatively, comparison of the internal peak intensity of one lane (e.g the 0.4 μg lane) to the flanking peak intensity of another lane (e.g. the 2.0 μg lane) predicts a peak ratio of 7, or a copy number of 9, when intensities are corrected for the different amounts of DNA on each lane. This latter method compares absolute intensities which are close to each other in value, rather than those which differ several fold in optical density. Hence, we estimate a copy number of approximately 8 Pro-1 genes per haploid genome.

Transcriptional mapping of Pro-1 mRNA

One important step in the characterization of the Pro-1 gene family is to define the structure of the Pro-1 mRNA and to map the region of the genomic repeat unit that codes for this mRNA.

Initially, the Pro-1 mRNA was sized on denaturing agarose gels (16). Northern blots of *L. enriettii* promastigote RNA probed with the insert of the Pro-g1 genomic clone reveal a major hybridizing band of 3.1 kilobases (Fig. 5, size of mRNA measured from lower exposure of autoradiogram). Upon extended exposure of these blots, we also observe a lighter band of approximately 6 kilobases (arrowhead in Fig. 5). This higher molecular weight band could represent a precursor of the mRNA containing two units of the tandem repeat.

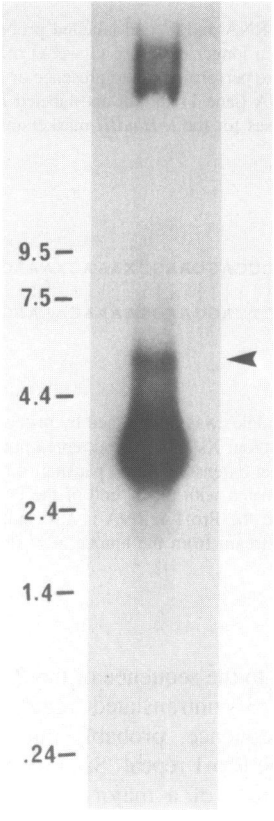

Figure 5. Northern blots of Pro-1 RNA. Size of Pro-1 mRNA. 5 μg of promastigote RNA was electrophoresed on a 1% agarose/2.2 M formaldehyde gel, blotted onto a nylon membrane and hybridized with the radiolabelled insert of the Pro-g1 genomic clone. The mobilities of RNA molecular weight markers (Bethesda Research Laboratories, Gaithersburg, MD) were determined by staining a marker lane with ethidium bromide; numbers at the left indicate the sizes of these markers in kilobases. The Pro-1 RNA band migrates at 3.1 kilobases, as determined from a lower exposure of this autoradiogram. The arrowhead at the right marks the 6 kilobase band that is visible upon long exposure of the autoradiogram.

To define the region of the Pro-1 repeat that encodes the mature mRNA, we performed S1 mapping experiments (23) using 5′ and 3′ end-labelled (16) restriction fragments of the Pro-g1 genomic clone. Initially, we used a 5′ end-labelled *ClaI-ScaI* restriction fragment (Fig. 1B, probe 1) to map the 5′ end of the transcript and a 3′ end-labelled *ClaI-ScaI* restriction fragment (Fig. 1B, probe 2) to map the 3′ end of the mRNA. Both probes were end-labelled at the *ClaI* site inside the mRNA coding region, and together, they encompass the entire repeat unit. Figure 6 demonstrates that hybridization to promastigote RNA protects a 1.4 kilobase fragment of probe 1 from S1 digestion (lanes 5−7), tentatively mapping the 5′ end of the mRNA to a position 1.4 kilobases upstream of the internal *ClaI* site (Fig. 1B). Similarly, hybridization to promastigote RNA protects a major 1.2 kilobase fragment (Fig. 6, lane 4) and a minor 1.8 kilobase fragment (Fig. 6, lanes 4 and 9, arrowheads) of probe 2 from S1 nuclease digestion. As discussed below, we believe that most of the Pro-1 mRNA terminates about 1.8 kilobases downstream from the *ClaI* site, and that the more intense 1.2 kilobase band may arise from sequence heterogeneity in the 3′ untranslated region of the multicopy gene family, which would result in nuclease sensitive mismatches between the 3′ end-labelled probe and the mRNA in the hybrid. Based on the larger 3′ protected fragment, these S1 mapping results predict a message of about 3.2 kilobases, consistent with the observed size of the Pro-1 mRNA as determined on Northern blots.

The 5′ end of the mRNA was further confirmed using a 570 base pair *ScaI-PvuII* fragment (Fig. 1B, probe 3) 5′ end labelled at the *PvuII* site. Following hybridization and S1 digestion, this probe generated a 425 base pair protected fragment (Fig. 6, lane 12) in agreement with the initial localization of the 5′ mRNA terminus. An additional protected fragment of 325 base pairs is present when hybridizations are performed either with (Fig. 6, lane 12) or without (Fig. 6, lane 11)*Leishmania* RNA and may be due to partial renaturation of the labelled probe (24).

Pro-1 mRNA contains the 5′ spliced leader sequence

To determine whether the Pro-1 mRNA contains the spliced leader observed on the 5′ termini of other *Leishmania* mRNAs (11) and to map precisely the 5′ end of the message, we have have sequenced the 5′ end of the mRNA by primer extension (25,26). A synthetic pentadecanucleotide (Fig. 1B, nucleotides 228−242 downstream from the *ScaI* site) was end-labelled, hybridized to promastigote polyadenylated RNA and extended with reverse transcriptase in the presence of dideoxynucleoside triphosphates. The sequence of these primer extension products (Fig. 7) indicates that the Pro-1 mRNA does contain the spliced leader and that this leader has been added onto the RNA at a 3′ consensus splice site. This RNA sequence also defines the map position for the 5′ end of the Pro-1 exon at 175 base pairs downstream from the *ScaI* site (Fig. 1B and Fig. 7) of the Pro-1 sequence (10), and confirms the approximate map position determined by S1 mapping.

Pro-1 mRNA contains multiple polyadenylation sites

To map precisely the 3′ ends of the Pro-1 mRNA, we have sequenced the 3′ termini of several Pro-1 cDNA clones. A λgt10 cDNA library (14) made from promastigote mRNA was screened with the Pro-g1 clone, and inserts from positive plaques were

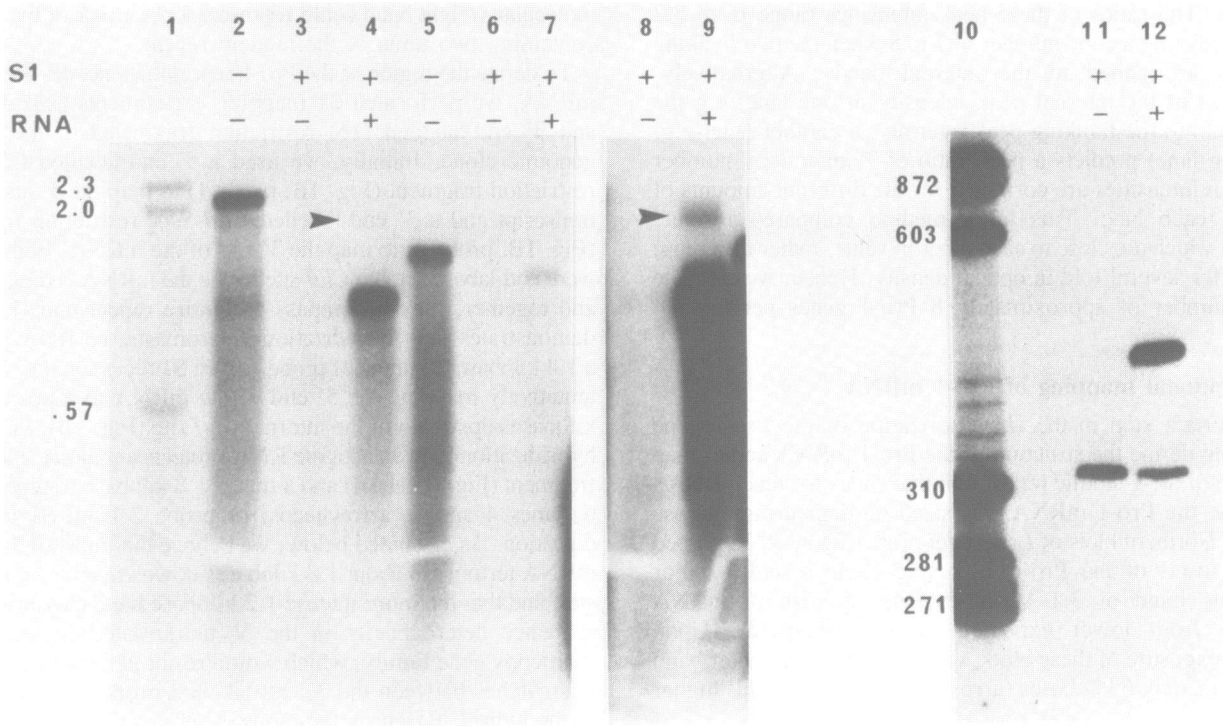

Figure 6. S1 mapping of Pro-1 mRNA. S1 mapping experiments were performed using promastigote total RNA and 5′ end labelled probes 1 (lanes 5–7) and 3 (lanes 11 and 12) and 3′ end labelled probe 2 (lanes 2–4 and 8–9), shown in Figure 1. Lanes 8 and 9 are a longer exposure (1 week) of lanes 3 and 4 (18 hour exposure). Hybridizations were performed in the presence and absence of RNA, and digestions of hybrids were performed in the presence or absence of S1 nuclease, as indicated at the top of the figure. Molecular weight markers were an end-labeled *HindIII* digest of λ DNA (lane 1) and an end-labeled *HaeIII* digest of φX174 DNA (lane 10). The sizes of the molecular weight markers are indicated at the left of each panel, in kilobases for the λ *HindIII* markers and in base pairs for the φX174 markers.

```
SL      AACUAACGCUAUAUAAGUAUCAGUUUCUGUACUUUAUUG
                ||||||||||||||||||||||||||||||||||
mRNA    XXXXXXCGCUAUAUAAGUAUCAGUUUCUGUACUUUAUUG^CUCUCUCAGCAAGUXAAACXAAAC
                                                ||||||||||||||| |||| ||||
Pro-1   CCCACCCACACGCGCGGTAACACTTTCACTGCGCTGCAG^CTCTCTCAGCAAGTAAAACAAAAC
                                                |
                                               175
```

Figure 7. Primer extension sequencing of the 5′ end of Pro-1 mRNA. The sequence of the Pro-1 mRNA (middle) was determined by primer extension sequencing. The internal positions marked X are those that could not be read unambiguously from the sequencing gel. The first four Xs at the 5′ end represent the four hypermethylated bases known to be located at the 5′ end of the spliced leader (36), but which cannot be distinguished by primer extension due to pausing of the reverse transcriptase at these positions. The sequence (11) of the *L. enriettii* spliced leader (SL) is indicated at the top, and the match with the 5′ end of the Pro-1 mRNA is indicated by vertical lines. The sequence of the Pro-1 genomic DNA (Pro-1) is displayed below, and the match with the Pro-1 mRNA is also indicated by vertical lines. The underlined AG is the 3′ consensus splice sequence. The C residue that is located 175 base pairs downstream from the unique *ScaI* site in the Pro-1 genomic repeat is marked. The caret indicates the position where the spliced leader is added onto the Pro-1 RNA.

subcloned (16) into the Bluescript plasmid vector (Strategene, San Diego, CA). The 3′ terminal sequences of two such cDNA clones (Fig. 1B, clones c1 and c2) showed poly A stretches of about 20 nucleotides preceded by Pro-1 sequence (data not shown). These polyadenylation sites occured at nucleotides 3276 and 3462 downstream from the *ScaI* site of the Pro-1 sequence. The location of these sites agrees with the assignment of the major 3′ terminus as predicted from the larger (1.8 kilobase) protected fragment in the 3′ S1 mapping experiment (Fig. 6, lanes 4 and 9). These results confirm the transcription map assigned from the S1 mapping data (Fig. 1B) and indicate that the Pro-1 mRNA contains at least two different polyadenylation sites.

Furthermore, we detected single base differences in the 3′ untranslated region of the mRNA when we compared the partial sequence of the c2 cDNA clone to the partial sequence of the c1 cDNA clone or to the sequence of the Pro-g1 genomic clone (10). Hence, the 3′ untranslated region of the mRNA is heterogenous in sequence, probably due to variations in the different units of the Pro-1 repeat. Such sequence variation could explain the occurence of a major S1 protected band of 1.2 kilobases in the 3′ mapping experiments (Fig. 6, lane 4), a location which is about 600 base pairs upstream from the polyadenylation sites defined from sequence of the cDNA clones.

A 380 base pair subclone of Pro-g1, designated 28-7 (Fig. 1A), hybridizes to Pro-1 mRNA, as determined by Northern blot analysis, while the 230 base pair subclone 10-12 does not (data not shown). These results are consistent with the transcription map (Fig. 1B) and demonstrate that a short region of the Pro-1 repeat, represented minimally by the 230 base pair insert of the 10-12 clone, is not present in mature mRNA.

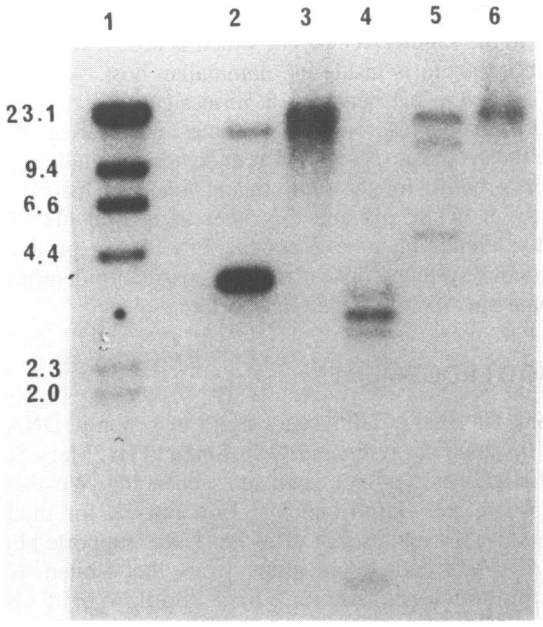

Figure 8. Sequences homologous to the Pro-1 gene are present in other species of *Leishmania*. A Southern blot containing 2 μg of *ClaI* digested *L. enriettii* DNA (lane 2), 5 μg of *ClaI* (lane 3) and *PstI* (lane 4) digested DNA from *L. donovani* clone DI700, 5 μg of *ClaI* digested DNA from *L. donovani* strain LRC-L353 (lane 5), and 5 μg of *ClaI* digested DNA from *L. major* strain WR303 (lane 6) was hybridized and washed under standard stringency conditions (Materials and Methods) with the insert of the Pro-g1 clone. Lane 1 contains a *HindIII* digest of λ DNA, with molecular weights in kilobases indicated at the left.

Detection of Pro-1 homologues in other species of *Leishmania*

To determine whether other species of *Leishmania* contain a gene that is structurally and presumably functionally related to the Pro-1 gene of *L. enriettii*, we have used the Pro-g1 insert to probe Southern blots containing genomic DNA from three *Leishmania* species that are human pathogens: *L. donovani, L. major and L. mexicana*. Figure 8 shows one such blot containing DNA from *L. enriettii* (lane 2), *L. donovani*, clone DI700 (27) (lanes 3,4), *L. donovani*, strain LRC-L353 (lane 5), and *L. major*, strain WR303 (lane 6). Even under the high stringency conditions used for this hybridization (Materials and Methods) all three strains contained sequences that cross-hybridized to the *L. enriettii* Pro-1 clone, and in other experiments (data not shown), *L. mexicana* DNA generated hybridization signals similar in intensity to those shown here. Structural differences among these related sequences are apparent from the polymorphisms in restriction sites and from the reduced intensity of hybridization of the heterologous sequences to the *L. enriettii* probe.

DISCUSSION

The life cycles of parasitic protozoa typically expose the organism to pronounced changes in their environments. Since membranes mediate the interaction of *Leishmania* parasites with the different physiological environments of the insect gut and the macrophage lysosome, there are likely to be membrane proteins which serve specialized roles in one life cycle stage and which are expressed in a stage-specific manner. We have isolated a gene, called Pro-1, that is expressed primarily in the insect stage of the parasite life cycle and that encodes a putative membrane transporter (10). We are interested in two broad questions with regard to the Pro-1 gene: i) what is the function of this gene in the parasite life cycle, specifically, what ligand does it transport; ii) how is the expression of this gene regulated during the parasite life cycle?

In the present study, we have shown that the Pro-1 genes are arranged in a tandem cluster containing approximately 8 copies of the repeat per haploid genome. Many of the genes that have been cloned from trypanosomatid protozoa are arranged in tandem repeats. A number of authors have suggested that such repeats may be transcribed from upstream promoters to yield polycistronic transcripts; these large transcripts would be processed to mature unit size mRNAs by polyadenylation at the 3′ ends and *trans*-splicing of the mini-exon sequence onto the 5′ ends. In the related protozoan *Trypanosoma cruzi*, transcripts of greater than unit size have been detected directly on Northern blots of total RNA (5). In *T. brucei*, they have been enriched by heat shock (α- and β-tubulin transcripts [7]), which apparently disrupts processing (28) of these putative precursors, or by selection of newly synthesized pulse labelled transcripts (calmodulin transcripts [8]). UV transcription mapping has defined a single 60 kilobase transcription unit in *T. brucei* containing a variant surface glycoprotein (VSG) gene and at least one other protein coding gene located upstream from the VSG gene (6).

Northern blots of *L. enriettii* RNA probed with the Pro-g1 clone reveal the presence of a 6 kilobase transcript which is much less abundant that the mature 3.1 kilobase mRNA and which could contain two units of the repeat. Muhich *et al.* (7) have demonstrated that application of an abrupt heat shock to bloodstream forms of *T. brucei* induces the accumulation of larger than unit size transcripts derived from the α-β-tubulin gene cluster. Although it has not been possible to directly demonstrate a precursor-product relationship between these larger transcripts and the mature tubulin mRNAs, the data are consistent with the presence of such polycistronic precursors. We have attempted similar heat shock experiments using our culture form promastigotes (grown at 27°), but we have not been able to detect an increase in abundance of the 6 kilobase transcript at any heat shock temperature from 32° to 42°. In trypanosomes the accumulation of putative precursors in response to heat shock is greatly diminished by treatments that stress the cells prior to the application of heat shock (7). It is possible that promastigotes that are grown in tissue culture medium are in some way pre-stressed and hence will not respond to heat shock in the same way that unstressed bloodstream trypanosomes respond. However, at present the origin of the 6 kilobase band detected on Northerns is uncertain. Ultimately, the identification of promoters and transcription units will require the application of a functional assay for transcription initiation, utilizing the transfection systems that are currently being developed (29,30).

Sequencing of the Pro-1 mRNA by primer extension demonstrates that this message contains the spliced leader or mini-exon sequence at its 5′ end, as observed for all other trypanosomatid mRNAs that have been investigated to date. In contrast to the 5′ end, the 3′ ends of Pro-1 mRNA are heterogeneous both in sequence and in the site of polyadenylation. Partial sequencing of the 3′ termini of two Pro-1 cDNA clones reveals that polyadenylation occurs at at least two different sites separated by 200 nucleotides and that single nucleotide differences occur within the 3′ untranslated regions of various Pro-1 mRNAs. We do not know at present whether a single unit of the repeat can be polyadenylated at different sites or whether each different 3′ terminus is derived from a different unit of the repeat. The heterogeneity in sequence probably arises from sequence

differences within the 3′ untranslated coding region of different units of the Pro-1 genomic repeat, as demonstrated previously for the repeated calmodulin (22) and procyclic acid repetitive protein (PARP) genes (31) of *T. brucei*.

The structures of two Pro-1 cDNA clones studied here indicate that the mRNAs from which they are derived are colinear with the genomic DNA within the 3′ untranslated region. The measured sizes of the cDNA inserts (1.2 and 2.3 kilobases respectively for c1 and c2) are the sizes that are predicted from the positions of the 5′ and 3′ terminal cDNA sequences within the genomic sequence, if colinearity of the cDNA and genomic DNA is assumed. Furthermore, the location of the polyadenylation sites of the two cDNA sequences is 3.1 and 3.3 kilobases downstream from the 5′ mRNA map site in the genomic DNA, in agreement with the measured size of 3.1 kilobases for the Pro-1 mRNA.

The sequences upstream from the two polyadenylation sites do not contain the consensus polyadenylation signal (32) from higher eukaryotes AATAAA; this sequence appears to be absent in the 3′ untranslated region of other genes in trypanosomatids. Similarly, we do not find a putative trypanosome polyadenylation signal (33), AAAATTYT (Y indicates pyrimidine), nor do we observe TTA repeats or T homopolymer tracts which may be involved in transcription termination of some trypanosome genes (34).

We have identified genomic restriction fragments that flank the 5′ and 3′ termini of the Pro-1 tandem repeat by hybridization to probes that are specific for the 5′ and 3′ regions of the Pro-1 gene. The identification of 5′ flanking sequences is especially important, as the promoter and *cis*-acting DNA elements which may be important for the regulation of Pro-1 gene expression could be located upstream from the Pro-1 repeat. We have recently cloned a 5′ flanking fragment (band *a*, Fig. 4A) and demonstrated by sequencing that it contains Pro-1 protein coding sequence at one end and sequence unrelated to the Pro-1 tandem repeat unit at the other end (data not shown). Ultimately, segments of this or other flanking sequences will have to be tested for promoter function by transfecting (29,30) *Leishmania* with constructs containing the flanking DNA fused to a reporter gene. The location and structure of RNA polymerase II promoters and the nature of the sequence elements involved in regulating transcription remain to be defined in the trypanosomatid protozoa. Since these organisms are very primitive eukaryotes which diverged from the main eukaryotic lineage very early in evolution (35), it will be important to determine how similar or different their transcription apparatus is in relation to other eukaryotes. The Pro-1 gene family is strongly regulated during the life cycle switch in *Leishmania* and will be a valuable probe for dissecting the molecular events involved in regulating gene expression during development. It remains to be determined whether this regulation is exerted at the transcriptional or post-transcriptional levels.

Using the Pro-1 gene as probe, we have detected genomic sequences in other species of *Leishmania* that are closely related to the Pro-1 gene of *L. enriettii*. This observation suggests that a structurally related gene exists in all species of *Leishmania* and presumably serves a similar biological function in the promastigote stage of the life cycle. The existence of developmentally regulated transporters may reflect the requirement of these parasites for a particular nutrient that is present in the sandfly vector but which is absent or not utilized when the parasite is inside its mammalian host. Alternatively, the difference in pH between the insect gut, which is close to neutrality, and the macrophage lysosomes, which are acidic, may require the synthesis of two different developmentally regulated transport proteins for the same ligand, one of which functions optimally at acidic pH and the other at neutral pH. We are currently attempting to express the Pro-1 gene product in a heterologous system where it may be possible to determine the functional specificity of the transporter.

ACKNOWLEDGEMENTS

We thank Dr. Buddy Ullman for a gift of genomic DNA from the DI700 clone of *Leishmania donovani* and Dr. Mary Stenzel-Poore for advice on primer extension sequencing. We also thank Dr. Margaret H. Baron and Dr. Eric Barklis for thoughtful comments on the manuscript. This work was supported by NIH grant AI25920 and by a grant from the United Nations Development Programme/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases.

REFERENCES

1. Murphy, W.J., Watkins, K.P. and Agabian, N. (1986) Cell 47, 517–525.
2. Sutton, R.E. and Boothroyd, J.C. (1986) Cell 47, 527–535.
3. Benne, R., Van Den Burg, J., Brakenhoff, J.P.J., Sloof, P., Van Boom, J.H. and Tromp, M.C. (1986) Cell 46, 819–826.
4. Feagin, J.E., Jasmer, D.P. and Stuart, K. (1987) Cell 49, 337–345.
5. Gonzalez, A., Lerner, T.J., Huecas, M., Sosa-Pineda, B., Nogueira, N. and Lizardi, P.M. (1985) Nucleic Acids Res., 12, 3777–3790.
6. Johnson, P.J., Kooter, J.M. and Borst, P. (1987) Cell 51, 273–281.
7. Muhich, H.L. and Boothroyd, J.C. (1988) Mol. Cell Biol. 8, 3837–3846.
8. Tschudi, C. and Ullu, E. (1988) EMBO J. 7, 455–463.
9. Zuckerman, A. and Lainson, R. (1979) In: Parasitic Protozoa (Kreier, J.P., ed.), Vol. I, pp. 57–133, Academic Press, New York.
10. Cairns, B.R., Collard, M.W. and Landfear, S.M. Proc. Natl. Acad. Sci. U.S.A. 86, 7682–7686.
11. Miller, S.I., Landfear, S.M. and Wirth, D.F. (1986) Nucleic Acids Res. 14, 7341–7360.
12. Landfear, S.M. and Wirth, D.F. (1985) Mol. Biochem. Parasitol. 15, 61–82.
13. Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem 132, 6–13.
14. Huynh, T.V., Young, R.A. and Davis, R.W. (1985) In: DNA Cloning Techniques, a Practical Approach (D. Glover, ed.), Vol. I, pp.49–78, IRL Press, Oxford.
15. Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263–269.
16. Maniatis, T., Fritsch , E.F. and Sambrook, J. (1982) Molecular Cloning, a Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
17. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
18. Anand, R. (1986) Trends in Genet. 2, 278–283.
19. Carle, G.F. and Olson, M.V. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 3756–3760.
20. Childs, G., Maxson, R., Cohn, R.H. and Kedes, L. (1981) Cell 23, 651–663.
21. Carle, G.F. and Olson, M. (1984) Nucleic Acids Res. 12, 5647–5664.
22. Tschudi, C., Young, A.S., Rubin, L., Patton, C.L. and Richards, F.F. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 3998–4002.
23. Weaver, R.F. and Weissman, C. (1979) Nucleic Acids Res. 7, 1175–1193.
24. Sharp, P. A., Berk, A.J. and Berget, S.M. (1980) Methods in Enzymol. 65, 750–768.
25. Levy, S., Sures, I. and Kedes, L. H. (1979) Nature 279, 737–739.
26. McGeoch, D.J. and Turnbull, N.T. (1979) Nucleic Acids Res. 5, 4007–4024.
27. Iovannisci, D.M., Kaur, K., Young, L. and Ullman, B. (1984) Mol. Cell. Biol. 4, 1013–1019.

28. Yost, H.J. and Lindquist, S. (1986) Cell 45, 185–193.
29. Bellofatto, V. and Cross, G.A.M. (1989) Science 244, 1167–1169.
30. Laban, A. and Wirth, D.F. (1989) Proc. Natl. Acad. Sci. U.S.A. 86,9119–9123.
31. Mowatt, M.R. and Clayton, C.E. (1988) Mol. Cell. Biol. 8, 4055–4062.
32. Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41, 349–359.
33. Borst, P. (1986) Ann. Rev. Biochem. 55, 701–732.
34. Glass, D.J., Polvere, R.I. and Van der Pleog, L.H.T. (1986) Mol. Cell. Biol. 6, 4657–4666.
35. Sogin, M.L., Elwood, H.J. and Gunderson, H.H. (1986) Proc. Natl. Acad. Sci. U.S. A. 83, 1383–1387.
36. Perry, K., Watkins, K.P. and Agabian, N. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 8190–8194.