Nucleotide sequence of the *Leishmania donovani* medRNA gene

Keith Wilson, Sheri Hanson, Scott Landfear¹ and Buddy Ullman*

Department of Biochemistry and Molecular Biology and ¹Department of Microbiology and Immunology Oregon Health Sciences University, Portland, OR 97201-3098, USA

Submitted September 17, 1991

EMBL accession no. X62143

The trans-splicing reaction of kinetoplastid protozoa attaches a separately encoded 39 nucleotide mini-exon onto the 5' ends of all mRNAs (1). The presence of this defined sequence affords the possibility of facile mapping of the 5' termini of mature mRNAs in these primitive pathogens by amplification of RNA sequences using the polymerase chain reaction. The subcloning and sequencing of the resulting amplified DNAs circumvents the necessity of less accurate and more time consuming traditional methods for mapping 5' ends of transcripts, such as S1 nuclease mapping or primer extension. Although the sequences of the mini-exons in kinetoplastids are highly conserved (2), particularly among *Leishmania* species (3), only the first 15 nucleotides of the *L. donovani* mini-exon sequence have been reported.

To develop a polymerase chain reaction based amplification system for examining the 5' termini of *L. donovani* mRNAs and to compare the sequences and components of the regions surrounding the mini-exon domain with that in other kinetoplastids, the *L. donovani* medRNA gene was cloned and a 429 bp *ApaI* fragment containing the complete medRNA transcript was sequenced in both directions (Figure 1). The 39 nucleotide mini-exon sequence was identical to those previously described from other *Leishmania* species, whereas the spliced leader intron differed from that of *L. enriettii* in 4 nucleotide positions and contained a one nucleotide deletion. Outside the transcript regions, there was essentially no homology between the intergenic regions of the *L. donovani* and *L. enriettii* mini-exon gene repeats with the exception of an oligo dT tract following the 3' end of the medRNA (nucleotides 157–165), a feature which is also present in the mini-exon gene from *Trypanosoma brucei* (4) and *Trypanosoma cruzi* (5).

To determine whether the medRNA gene of *L. donovani* is arranged in a tandem repeat, as in the *L. enriettii* gene, genomic DNA was partially digested with *RsaI*, an enzyme that cuts once in the *ApaI* fragment.

Southern blots of *RsaI* digested DNA revealed the characteristic ladder of hybridization signals, each a multimer of the 429 bp unit, indicating that the mini-exon genes are arranged in a tandemly repeated array (Figure 2). This arrangement of the mini-exon gene sequences in *L. donovani* was confirmed by complete digestion of genomic DNA with *NcoI* (Figure 2), *ApaI*, *AvaI*, and *BglII*, each of which excised a DNA fragment of 429 bp.

ACKNOWLEDGEMENTS

This work received financial support from grant No. AI23682 (B.U.) and AI25920 (S.L.) from the National Institute of Allergy and Infectious Disease.

REFERENCES

1. Agabian, N. (1990) *Cell* **61**, 1157–1160.
2. Miller, S.I., Landfear, S.M. and Wirth, D.F. (1986) *Nucl. Acids Res.* **14**, 7341–7360.
3. Scholler, J.K., McArdle, S., Reed, S. and Kaneniotto, R. (1989) *Nucl. Acids Res.* **17**, 7999.
4. Campbell, D., Thornton, D. and Boothroyd, J. (1984) *Nature* **311**, 350–355.
5. Milhausen, M., Nelson, R.G., Sather, S., Selkirk, M. and Agabian, N. (1984) *Cell* **38**, 721–729.

```

L.d.  GCG CCC GTG CCG GGA GAA CGT CCG CCG GAA TCG GCG CTG TGG GTG TGA 51
L.e.  TC GGT TGC T T A T GTG A T G AC C G CGA GCT GAC G G T

L.d.  CCG CTT GCA ACT AAC GCT ATA TAA GTA TGA GTT TGT GTA CTT TAT TGG TAT 102
L.e.  A A GG TT

L.d.  GCG AAA C-TT CCG GAA CCG GTC TTC CCG CAA CAT TTT GGA AGC GCG CAG GCG 153
L.e.  C T G G A CG

L.d.  CTC TTT TTT TTT GTG TGC TGT TGT GTG TGG CCG CCG GTC CCG CCG GCG CCG 204
L.e.  --- G C C GT GCC T C GCC C T T T CCT GTG

L.d.  CCG CCG CTT CCG TAG CCG TGG CCG CCG CCG CCG GAG GCG GCG CCA CCG CCG 255
L.e.  GCT GTC TC GG GGC CC TC GT C AG G G CC C TGC ATC T G

L.d.  GCG GCA GGC GTC CCG CCA GGC GGC GCG TGA CTG GCG CCG CCG GCA CCG GCG 306
L.e.  T CCG G CCG G GGC T GG CA C C AAC G TAG GC G GCT TG

L.d.  GTG CGT GCG CCG GCG CGT TTC CGT CTC CCG GCG GCG CCG CCG GCG GCG GTG 357
L.e.  GTG C T CT CG A G C G C C GCA ACG CT T G T GCG CGT GC GT

L.d.  TGT GCG ACG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG 408
L.e.  C G CG T C GGT GTT TTT TG T G GG C G GT T C T G CCG GT G T

L.d.  GCG GCG ATG GTG GTG ACG CCG 429
L.e.  TG T GG CG CG T T G T CAA GCG GCG TG 438

```

Figure 1. Nucleotide sequence of the *L. donovani* medRNA gene and comparison with *L. enriettii* medRNA gene (2). The mini-exon sequence is double underlined, while the intron is underlined once and the 3' end is indicated on the basis of the *L. enriettii* medRNA sequence.

Figure 2. Genomic Southern blot of *L. donovani* digested with two restriction endonucleases that have a unique site within the *ApaI* fragment. Lanes A-D = *RsaI*, 2, 5, 15 and 120 min digestions respectively; Lane E = *NcoI*, 120 min digestion.

* To whom correspondence should be addressed