The monogenetic kinetoplastid protozoan, *Crithidia fasciculata*, contains a transcriptionally active, multicopy mini-exon sequence

Michael L.Muhich*, Dallas E.Hughes, Agda M.Simpson and Larry Simpson⁺

Department of Biology and Molecular Biology Institute, University of California, Los Angeles, CA 90024, USA

Received October 27, 1986; Revised and Accepted January 27, 1987

ABSTRACT

A repeated sequence from the *Crithidia fasciculata* nuclear genome has been isolated which is homologous to the mini-exon genes of other kinetoplastid protozoa. Sequence analysis of the 417 bp monomeric unit confirmed the presence of a 35 nt sequence within the repeat that is 77% homologous with the *Trypanosoma brucei* 35-mer mini-exon or spliced leader sequence. The repeat is present at approximately 250 copies per cell and is organized into one, or a few, large head to tail tandem clusters predominantly on a single chromosome. The mini-exon repeat unit hybridizes to a major 84 nt and a minor 87 nt poly (A)- steady state transcript, the first 35 nts of which comprise the mini-exon sequence found at the 5' end of mRNAs in several other kinetoplastid species. The 3'-termini of the transcripts map to positions on the DNA sense strand directly preceding a stretch of 8 thymidine residues. *Crithidia* represents the most primitive kinetoplastid species which apparently possesses a discontinuous type of mRNA processing, implying that this represents a conserved feature in possibly all genera of kinetoplastid protozoa.

INTRODUCTION

One of the most interesting and unique features of the molecular biology of kinetoplastid protozoa is the presence of a 35 nt mini-exon or spliced leader sequence at the 5' ends of apparently all nuclear-derived messenger RNAs (reviewed in 1). This mini-exon sequence is encoded by a repeated nuclear DNA sequence which varies in size from 1.35 kb to 0.4 kb in different species, and is present in tandem arrays of around 200 copies. The sequence of the mRNA mini-exon is known in several kinetoplastid species: *Trypanosoma brucei* (2-4), *T. vivax* (3), *T. cruzi* (3, 5), *Leptomonas collosoma* (5, 6), and *Leishmania enrietti* (7). The sequence of the complete repeat unit is known for several stocks of *T. brucei* (8, 9), *T. vivax* (3), *T. cruzi* (3) and *L. enrietti* (7), and a partial sequence is known for *L. collosoma* (5). Hybridization evidence for the existence of a mini-exon repeat has been reported for *T. equinum*, *T. congolense*, *Leishmania tropica*, *Herpetomonas muscarum*, *Phytomonas davidi* and *Crithidia fasciculata* (3, 5, 6).

Transcription of the mini-exon repeat leads to a steady state transcript, termed mini-exon derived RNA (medRNA), which varies in size from 140 nt in *T. brucei* to 105 nt in *T. cruzi*, 95 nt in *L. collosoma* (1), and 85 nt in *L. enrietti* (7). The mechanism for the transfer of the 35 nt mini-exon sequence from the medRNA to the 5' ends of mRNAs is not known, although two models have been proposed involving priming and trans (bimolecular) splicing. Evidence for the latter mechanism has recently been obtained (10, 11). The biological function of the mini-exon sequence in the kinetoplastid protozoa is not known. We show in this paper the sequence of the mini-exon repeat in *C. fasciculata*.

a primitive, monogenetic kinetoplastid protozoan and perform a comparative analysis of this sequence with all other reported mini-exon repeat sequences as an approach to understanding the biological role of this phenomenon and the evolution of the different genera within the family, Trypanosomatidae.

MATERIALS AND METHODS

Cell culture

Cultures of *C. fasciculata* were grown in either Kidder-Dutta defined medium (12) or Brain-Heart Infusion medium (Difco) supplemented with 10 ug/ml hemin, at 25 C with gentle shaking.

RNA and DNA isolation

Total cell RNA was prepared from exponential phase cultures by the guanidine HCl method of Berk and Sharp (13) and then incubated with 20 ug/ml DNase I (iodoacetate-treated) for 30 min at 37 C. Poly(A)+ enriched RNA was prepared by poly(U) Sephadex (Bethesda Research Labs) chromatography of whole cell RNA.

C. fasciculata nuclear DNA was prepared from stationary phase cells by lysis with 1% sarkosyl and digestion with 1 mg/ml pronase CB (autodigested, 30 min., 37 C) at 60 C for 2 hours. Nuclear DNA was separated from the bulk of the kinetoplast network DNA by CsCl/ethidium bromide isopycnic centrifugation.

Nuclease digestions and nucleic acid hybridizations

S1 nuclease (BRL) protection assays were performed as described previously (14). Mung bean nuclease (P-L Biochemicals) digestions of *C. fasciculata* genomic DNA were performed under conditions previously described (15). Filter blotting and nucleic acid hybridization methods were as described (16). DNA probes were ³²P-labeled *in vitro* by either nick translation (17) or T4 polynucleotide kinase treatment (14). ³²P-labeled RNA complementary to the mini-exon transcript was synthesized *in vitro* by SP6 polymerase as described by the supplier (Promega Biotec).

Orthogonal field agarose gel electrophoresis

Orthogonal field gel electrophoresis (OFAGE) was performed in an apparatus designed by C. Brunk based on the apparatus of Carle and Olson (18). Cells were embedded in 0.7% low melting agarose at 4 X 10⁸ cells/ml, and 5 X 4 X 1 mm blocks were treated with 1% Sarkosyl and 1 mg/ml proteinase K in 0.5 M EDTA (pH 8.0) at 50 C for 48 hours prior to electrophoresis. Gels were stained in 0.5 ug/ml ethidium bromide and destained for 24 hr at 4 C before photography. Lambda ladders were obtained by incubation of linearized lambda monomeric DNA (50 ug/ml) embedded in 0.45% low melting agarose at 37 C in 0.1 M EDTA, 10 mM Tris HCl (pH 7.9), for 17 hr. The agarose blocks were then used for OFAGE electrophoresis. A value of 50 kb per lambda monomer was used for the calculation of band size.

DNA sequence analysis

DNA sequence determinations were performed using the dideoxy-chain termination method (19) as described previously (20). DNA sequences were analyzed with the SEQH and SEQA programs of Goad and Kanehisa (21), and the BESTFIT and GAP programs of the University of Wisconsin Genetics computer group using a VAX 11/750 computer.

```

      GC
      |
1  AAGCTTCCGG AAACAACCGG CACAAATTTT GAGGCGGAAG CTGCTTTT
      |
51  TTGTGTCCGG GGGGGCCTCC TTGGGGTCCC CTGTCCAGC CCCAGCCGTT
      |
101 CGCCACCACA TAGGAATTTG CGAAGGACCC CAAAAATCC CGGTCCCGG
      |
151 GCGAGTTGTC CCAACTTTTT CAAACCTCAT GAAGAGCTAG TTGCGTCATT
      |
201 GAAAAGTTCG TGTGCAGAAA CCCCCTCCCC CACGTTTGTA CAATGGAAGA
      |
251 GTTTACGATA CAGGTTTTCT CACGGTTTTG AGGTGTTTTT TCGAAAAACA
      |
301 AAAAAATATAG AGGTGTATAG CGCTTATTTT TGACACCCCC CTCAAAACAT
      |
351 GCTGGGGGTA TAGGTCCTTC CAACTAACGC TATATAAGTA TCAGTTTCTG
      |
401 TACTTTATTG GTATAAG

```

Figure 1. Nucleotide sequence of the 417 bp Cf-2 HindIII mini-exon repeat from *C. fasciculata*. The box indicates the position of the 35 nt mini-exon. The arrows indicate the positions of the 3'-termini of the major, 84 nt (large arrow) and the minor, 87 nt (small arrow) medRNAs as mapped by S1 protection analysis using the XmaI mini-exon repeat, clone Cf-3 (the single XmaI site is indicated in the Cf2 sequence). The Cf-3 sequence contained an additional GC at positions 82 and 83 (the first A of the mini-exon is considered +1) on the Cf-2 mini-exon transcription unit. The oligo-T 8 mer adjacent to the 3' end of the medRNA is indicated, as are two 7 mer repeats of unknown significance, and three TATA sequences located 5' of the medRNA initiation site.

RESULTS

The *C. fasciculata* genome encodes a 35 nt mini-exon sequence

Hybridization of a synthetic 22 nt probe complementary to the *T. brucei* mini-exon (Fig. 7) (kindly provided by K. Stuart) occurred under reduced stringency conditions to *C. fasciculata* genomic digests, as expected from previous results by De Lange et al (3). The 22 nt *T. brucei* mini-exon probe hybridized (in 0.75 M NaCl, 20% formamide, 37 °C) with a 400 bp fragment of *C. fasciculata* nuclear DNA which had been digested with either HindIII or XmaI (data not shown).

To isolate the potential mini-exon encoding fragment from the *C. fasciculata* genome, a size-selected library of HindIII digested *C. fasciculata* genomic DNA fragments averaging 400 bp in length was constructed in M13 mp10. Eighteen plaques hybridizing with the *T. brucei* mini-exon probe were recovered. Sequence analysis of the 417 bp insert of clone Cf-2 (Fig. 1) revealed the presence of a 35 nt sequence which is 77 percent (27 out of 35 nt) homologous with the *T. brucei* mini-exon sequence.

Sequence analysis of a second cloned XmaI-released fragment, Cf-3, confirmed the presence of an identical 35 nt mini-exon sequence, and revealed an additional GC dinucleotide between positions 39 and 40 of the Cf-2 mini-exon repeat sequence (Fig. 1), implying the existence of minor sequence heterogeneity among the mini-exon repeats.

The *C. fasciculata* mini-exon repeat codes for a small, poly(A)- steady state transcript (medRNA)

The *C. fasciculata* mini-exon repeat hybridized with an approximately 85 nt transcript which was present in total *Crithidia* RNA but absent from the poly(A)+ RNA fraction (Fig. 2). The absence of a smear of high molecular weight poly A+ mRNA species hybridizing with the mini-exon repeat, as would be expected if cellular mRNAs were 5'-terminated with the 35mer mini-exon, probably is due to the fact that the entire 417 nt

Figure 2. Northern analysis of *C. fasciculata* RNA. Twenty micrograms of total cell RNA and six micrograms of poly(A)+ enriched RNA were electrophoresed through a 1.5% agarose-2.2 M formaldehyde gel, blotted onto nitrocellulose, and probed at a high stringency with ³²P-labeled Cf-2 insert. Following autoradiography, the filters were briefly boiled to remove the probe and then rehybridized with a ³²P-labeled *Leishmania tropica* beta-tubulin probe (kindly provided by T. Spithill) as a control for the absence of poly(A)+ RNA degradation.

Figure 3. Determination of the steady state size of the mini-exon repeat transcript by S1 protection analysis. 32 P-labeled RNA complementary to the mini-exon transcript was synthesized *in vitro* by SP6 polymerase transcription of a SP6 promoter/Cf-3 mini-exon template. The labeled, SP6 run-off transcripts were annealed with 5 ug of total *Crithidia* RNA and the mixture digested with the indicated units of S1 nuclease. The digestion products were sized on a 6% acrylamide-7 M urea sequencing gel adjacent to a dideoxy-sequencing ladder of a *Leishmania tarentolae* maxicircle fragment of known sequence.

repeat was used as a probe instead of the 35mer alone and to the stringent hybridization conditions used.

To more accurately size the medRNA and to map the transcript termini, total cell RNA was hybridized with uniformly labeled (SP6 RNA polymerase-transcribed) RNA complementary to the mini-exon transcript, and then digested with S1 nuclease. The run-off probe RNA was synthesized from a Cf-3 template that encompassed the entire mini-

exon repeat (Fig. 3). Two S1 resistant products were observed (Fig. 3) at increasing S1 nuclease concentrations: a predominant 84 nt species and a minor 87 nt species. An analogous S1 experiment performed with probe RNA synthesized from an SP-6 Cf-2 template (linearized with XmaI) in which the 5'-proximal *C. fasciculata* mini-exon repeat sequence begins at the unique HindIII site (see Figs 1 and 3), yielded a single 45 nt

A Units HindIII / μ g *C. fasciculata* DNA

B Units Nuclease/ μ g *C. fasciculata* DNA

Figure 4. Organization of the mini-exon repeats in the *Crithidia* genome. (A) One microgram of *C. fasciculata* nuclear DNA was digested with the indicated units of HindIII (which cleaves once within each repeated unit) for one hour at 37°C, electrophoresed through 1.5% agarose, blotted onto nitrocellulose and probed with ³²P-labeled Cf-2 insert. (B) One microgram of *C. fasciculata* DNA was incubated with 10 units of mung bean nuclease at the indicated formamide concentrations for 30 min. at 50°C; gel electrophoresis and filter hybridization were as described in (A). A series of HindIII partial digests of *C. fasciculata* DNA were co-electrophoresed as marker DNA.

protected fragment (data not shown). This localizes the steady-state 5'-terminus of the *C. fasciculata* medRNA to nt position 376 on the Cf-2 sequence (Fig. 1), which corresponds to nt +1 of the *C. fasciculata* mini-exon sequence as defined by the alignment with the *T. brucei* 35-mer sequence (9). Moreover, the unique protection product obtained with SP6-Cf-2 probe indicates that the heterogeneity in size of the S1 resistant species observed with the SP6-Cf-3 probe (Fig. 3) is a result of heterogeneity at the 3'-end of the steady state medRNA. Heterogeneity at the 3'-terminus of the medRNA could be the result of true length variation resulting from either termination/processing variations or from differences in the 3'-terminal sequences of transcribed mini-exon repeats. The observed heterogeneity may also be an end nibbling artifact of the S1 nuclease digestion.

As shown above, a two nt difference in the putative transcribed regions of two isolated *C. fasciculata* mini-exon repeats was observed (Fig. 1); however, the expression of these specific repeats *in vivo* remains an open question. The close agreement between the size estimates for the medRNA, as determined by Northern (Fig. 2) and S1 (Fig. 3) analyses, argues that if significant sequence heterogeneity does occur among the transcribed mini-exon repeats which give rise to the bulk of the steady state medRNA population, then it probably occurs within the portion of the mini-exon repeat transcription unit corresponding to the 3'-steady state terminus.

Genomic organization and copy number of the *C. fasciculata* mini-exon repeats

To investigate the organization of the mini-exon repeats within the *Crithidia* genome, nuclear DNA was partially digested with HindIII, which cleaves once within each repeat unit (Fig. 1), electrophoresed on agarose and then transferred to nitrocellulose. Fig. 4A shows that incubation of this blot with ³²P-labeled Cf-2 insert yielded a "laddered" hybridization profile, in which the size of the Cf-2 homologous bands increased in multiples of the mini-exon repeat length, reflecting a tandem, head-to-tail organization of the mini-exon repeats within the *Crithidia* genome. Limit digestion with HindIII (1-20 units/ug DNA, Fig. 4A) indicates that greater than 95% of the mini-exon homologous sequences are contained within repeats having a monomeric length of approximately 420 bp.

The number of mini-exon repeats present in the *C. fasciculata* genome was determined by quantitative Southern hybridization as shown in Fig. 5. A densitometric scan of the autoradiograph indicated that the 0.05 ug Cf-2 RF and the 2 ug *C. fasciculata* nuclear DNA bands contained approximately equivalent numbers of mini-exon hybridizing repeat units. The 417 bp mini-exon repeat represents approximately 5.5% of the mass of the Cf-2 molecule and hence, 0.14% of the approximately 72.5×10^3 kb of DNA in the diploid *Crithidia* nucleus (22), which corresponds to approximately 250 copies of the 417 bp repeat per diploid cell.

Release of mini-exon repeats with mung bean nuclease

The tandem, head-to-tail arrangement of mini-exon repeats was also apparent from digestions of *Crithidia* nuclear DNA with mung bean nuclease. As shown in Fig. 4B, mung bean nuclease, which has been found to cleave duplex DNA from *Plasmodium* and *Trypanosoma*, under specific conditions, at sites before and after genes (23, 24), cleaves at specific, identically located sites within the tandemly clustered mini-exon repeats of *Crithidia*. Cleavage within the mini-exon repeat by mung bean nuclease is precise, as shown by comparing the mung bean enzyme and HindIII generated hybridization profiles in Fig. 4B. The nature of the mung bean nuclease cleavage site was not investigated.

Figure 5. Copy number of the mini-exon repeat in the *Crithidia* genome. The indicated masses of *C. fasciculata* nuclear and Cf-2 RF DNAs were digested to completion with HindIII, electrophoresed through 1.5% agarose, blotted onto nitrocellulose, and probed with an excess of 32 P-labeled Cf-2 insert.

Chromosomal distribution of the mini-exon repeats

C. fasciculata chromosomes were separated by the OFAGE method (18, 25), blotted to nitrocellulose, and the blot probed with 32 P-labeled Cf-2 insert. Approximately 12 ethidium bromide-staining chromosomal bands can be visualized by this method, ranging from 470 kb to more than 850 kb in size. The approximate sizes of all but the upper two or three bands were obtained by calibration with a lambda ladder. The mini-exon repeat probe hybridizes predominantly to a single 710 kb chromosome band, and to a minor extent to a 800 kb chromosomal region (Fig. 6). As is the case with chromosome gel patterns of other kinetoplastid protozoa, the intensity of the bands is not always in molar ratio, implying either a comigration of different chromosomes of the same size or the presence of multiple copies of a single chromosome.

Comparative sequence analysis of the *C. fasciculata* mini-exon repeat

The *C. fasciculata* Cf2 mini-exon repeat sequence was aligned with the published mini-exon repeats from *T. brucei* (3), *T. vivax* (3), *T. cruzi* (3), *Leishmania enrietti* (7), and *Leptomonas collosoma* (5) by the UWGCG BESTFIT program. The alignments are shown in Fig. 7, with the mini-exon consensus sequence presented beneath the alignments. A 13 nt sequence within the 35 nt mini-exon sequence is conserved in all six species. The variation in the *L. collosoma*, *T. brucei*, *T. vivax* and *T. cruzi* sequences occurs in the 5' portion of the 35mer. The *L. enrietti* sequence is identical to the *C. fasciculata* sequence for a span of 43 nt in the miniexon region. It is interesting that sequence conservation extends 3' into the repeat unit sequence for all species. A 5' AACT motif is present in all species but *L. collosoma*. However, if a two nt gap is inserted into the aligned miniexon sequences of these five species, then there is an 5' AACT in the *L. collosoma* sequence which aligns with all others. No gaps were inserted into the alignments of the 35mer sequences, pending identification of the 5' end of the *L. collosoma* medRNA transcript. The universally conserved GT dinucleotide immediately 3' of the 35mer sequence is consistent with the eukaryotic splice junction sequence at the 5' end of introns (26).

Probe: *C.f.* MINIEXON REPEAT

Figure 6. OFAGE separation of *C. fasciculata* chromosomes. The gel was run at 120 mA - 250 V (buffer = 0.5 X TBE) for 26 hours at 14 C, stained with ethidium bromide (right panel), blotted to nitrocellulose and the filter probed with ³²P-labeled Cf-2 mini-exon repeat (left panel). The arrow indicates the position of the major hybridizing chromosome in the ethidium stained gel. The approximate sizes of the chromosomal bands were determined by calibration with lambda ladders (not shown) to range from approximately 480 kb to more than 850 kb. The upper two or three chromosome bands are above the calibration range of the lambda ladder.

Figure 7. Sequence alignments of the *C. fasciculata* mini-exon repeat with repeats from five other species. The alignments were done using the UWGCG BESTFIT compute program, and the 35-mer mini-exon consensus sequence was obtained using the UWGCG CONSENSUS program at a certainty level of 75% at each position. Note that the mini-exon consensus sequence is written using the IUB ambiguity codes (36) (W = A or T, R = A or G, K = G or T). Matches with the *C. fasciculata* sequence are indicated by vertical lines. The nt numbers refer to the published sequences in each case. The abbreviations of the species and the literature sources are as follows: LEP=L. collosum (5), TC= T. cruzi (3), TV=T. vivax (3), TB= T. brucei (3, 8), LE= L. enrietti (7), CF= C. fasciculata.

An appropriate conserved eukaryotic promotor sequence could not be identified at the expected locations 5' of the 35mer mini-exon sequence. There are, however, several TATA sequences localized at -17, -60 and -70 nt, and a CAAT sequence located 65 nt upstream from the -70 TATA sequence. However, the sequence surrounding the CAAT sequence does not exhibit significant homology with the eukaryotic consensus sequence (27). In addition, there is no sequence homology at corresponding locations in the reported mini-exon repeat sequences from *T. cruzi* and *T. vivax* although the *T. brucei* sequence does show the TATA and CAAT motifs at approximately similar locations.

DISCUSSION

C. fasciculata is an insect hemoflagellate that has been extensively studied as a model kinetoplastid protozoan due to its hardiness in culture and ability to grow in a simple defined medium, short cell division time, high plating efficiency on agar and lack of pathogenicity. It is, however, unique in terms of morphology (the choanomastigote form is not found in any other genus) and appears to be widely diverged from other species in terms of several molecular and biochemical characters. A recent analysis of the mitochondrial ribosomal RNA sequences of *Crithidia*, *Leptomonas*, *Leishmania* (*Leishmania tarentolae*, a lizard parasite) and *Trypanosoma* has indicated that *Crithidia* represents the earliest branching point within the family (de la Cruz, Ferreira, Morel, Lake and Simpson, unpublished results). We have shown in this paper that *C. fasciculata* contains a mini-exon repeat sequence that is homologous to the sequences reported from four other kinetoplastid species and that this sequence is transcribed into a small medRNA. It has not yet been demonstrated, but is highly likely, that the medRNA donates the 35mer mini-exon sequence to the 5' end of all nuclear-derived cellular mRNAs in *Crithidia*, as in other trypanosomatids (28-31).

Alignments of the mini-exon repeat sequences from five species with the *Crithidia* sequence show the presence of a strongly conserved region and a variable region within the 35mer sequence in the case of *T. brucei*, *T. vivax*, *T. cruzi* and *L. collosoma*. The alignments extended 3' and 5' for all species, possibly implying a functional role for flanking sequences in either transcription or RNA processing. The extent of mini-exon repeat sequence divergence of *Crithidia* from each of the other species is approximately the same, except in the case of *L. enrietti*, in which there is a stretch of 43 identical nucleotides encompassing the 35mer mini-exon region.

The *C. fasciculata* 417 nt repeat does not contain the alternating adenosine-cytosine repeats noted in the *T. brucei* sequence (8). There is a run of 8 T's immediately downstream of the 3' end of the medRNA transcription unit, which may function as a termination signal, since the *T. brucei*, *T. cruzi*, *L. collosoma* and *L. enrietti* medRNAs also terminate adjacent to T-rich sequences and there is a similar T-rich sequence located at the expected terminus of the *T. vivax* transcriptional unit (3). The *C. fasciculata* and the *L. enrietti* oligo-T sequences show an identity of 12 nt including a 3' GTGT tetramer, and these represent the shortest T-rich putative termination signals yet found in the kinetoplastid family.

The striking similarities between *C. fasciculata* and *L. enrietti* in terms of the size of the repeat unit (417 nt, 438 nt), the size of the medRNA transcripts (84 nt, 85 nt), the sequence of the 35mer minixon region, and the sequence of the 3' region of the

medRNA are unexpected in view of our mitochondrial ribosomal RNA alignment analysis which places Crithidia (the published sequences of the 9s and 12s RNAs from the Hoeijmakers C. fasciculata cell line were used for this analysis) (32) at a deeper phylogenetic branch point than Leishmania tarentolae (de la Cruz et al, unpublished results). The evidence that the C. fasciculata line (Cf-C1) we work with is actually Crithidia is as follows: 1. The cell morphology is the expected choanomastigote form and not the promastigote form characteristic of Leishmania. 2. The 2.5 kb size of the kinetoplast minicircle DNA is the expected size of the Crithidia minicircle. However, there are differences in the minicircle sequence heterogeneity in that the Cf-C1 line contains a fairly homogeneous minicircle population (33), whereas the C. fasciculata line used in previous work by the Borst laboratory (34, 35) contains a heterogeneous minicircle population. 3. The cloned C. fasciculata miniexon repeat (Cf3) hybridizes well with multiple genomic DNA preparations obtained at different times from our Crithidia cell line. 4. The chromosome pattern on an OFAGE gel is characteristic of this species and differs from that of Leishmania. These facts imply that the sequence similarities observed are not due to cross contamination of cultures. The miniexon sequence similarities between these two species remain to be explained.

Eukaryotic promotor consensus sequences which are conserved in all six species could not be identified; identification must await the development of an in vitro transcription system which would allow a functional deletion analysis.

The specific cleavage of the 400 nt mini-exon repeats from genomic DNA by mung bean nuclease under conditions known to cleave before and after genes in Plasmodium and T. brucei is suggestive that the nuclease is recognizing a DNA sequence or DNA structure closely related to that found adjacent to structural genes. However, the nature of the enzyme recognition site is not known for any case.

We conclude that the presence of a transcribed mini-exon repeat in the monogenetic kinetoplastid, C. fasciculata, makes it likely that this is a common feature of all members of the family, Trypanosomatidae, and that the appearance of this unique mode of mRNA synthesis was at an early point in the evolution of these protozoa.

ACKNOWLEDGEMENTS

This work was supported in part by NIH research grant AI09102 and by a California Biotechnology Research and Education Program award to L.S. (4-443832-20509). M.L.M. was the recipient of a U.S. Public Health Service National Research Award (GM-07104) and a Dr. Ursula Mandel Scholarship from UCLA.

* Current address: Department of Medical Microbiology, Stanford University School of Medicine, Stanford CA 94305

+ Corresponding author.

REFERENCES

1. Borst, P. (1986) Ann. Rev. Biochem. 55, 701-732.
2. Boothroyd, J. and Cross, G. (1982) Gene 20, 281-289.
3. De Lange, T., Berkvens, T., Veerman, H., Frasch, A., Barry, D. and Borst, P. (1984) Nucl. Acids Res. 12, 4431-4443.

4. Kooter, J.M., De Lange, T. and Borst, P. (1984) *EMBO J.* 3, 2387-2392.
5. Milhausen, M., Nelson, R., Sather, S., Selkirk, M. and Agabian, N. (1984) *Cell* 38, 721-729.
6. Nelson, R., Parsons, M., Selkirk, M., Newport, G., Bart, P. and Agabian, N. (1984) *Nature* 308, 665-667.
7. Miller, S., Landfear, S. and Wirth, D. (1986) *Nucl. Acids Res.* 14, 7341-7360.
8. Dorfman, D. and Donelson, J. (1984) *Nucl. Acids Res.* 12, 4907-4920.
9. Campbell, D., Thornton, D. and Boothroyd, J. (1984) *Nature* 311, 350-355.
10. Murphy, W., Watkins, K. and Agabian, N. (1986) *Cell* 47, 517-525.
11. Sutton, R. and Boothroyd, J. (1986) *Cell* 47, 527-535.
12. Kidder, G. W., and Dutta, B. N. (1958) *J. Gen. Microbiol.* 18, 621-638.
13. Berk, A. and Sharp, P. (1977) *Cell* 12, 721-732.
14. de la Cruz, V.F., Lake, J.A., Simpson, A.M. and Simpson, L. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1401-1405.
15. Muhich, M. and Simpson, L. (1986) *Nucl. Acids Res.* 14, 5531-5556.
16. Muhich, M., Neckelmann, N. and Simpson, L. (1985) *Nucl. Acids Res.* 13, 3241-3260.
17. Muhich, M., Simpson, L. and Simpson, A. (1983) *Proc. Natl. Acad. Sci. USA* 80, 4060-4064.
18. Carle, G. and Olson, M. (1984) *Nucl. Acids Res.* 12, 5647-5664.
19. Sanger, F. Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
20. de la Cruz, V., Neckelmann, N. and Simpson, L. (1984) *J. Biol. Chem.* 259, 15136-15147.
21. Goad, W. and Kanehisa, M. I. (1982) *Nucl. Acids Res.* 10, 247-264.
22. Borst, P., van der Ploeg, M., van Hoek, J.F., Tas, J. and James, J. (1982) *Mol. Biochem. Parasitol.* 6, 13-23.
23. McCutchan, T., Hansen, J., Dame, J. and Mullins, J. (1984) *Science* 225, 625-628.
24. Brown, K., Bretano, S. and Donelson, J. (1986) *J. Biol. Chem.* 261, 10352-10358.
25. Schwartz, D. and Cantor, C. (1984) *Cell* 37, 67-75.
26. Ruskin, B., Krainer, A., Maniatis, T. and Green, M. (1984) *Cell* 38, 317-331.
27. Efstratiadis, A., Posakony, J., Maniatis, T., Lawn, R., O'Connell, C., Spritz, R., DeRiel, J., Forget, B., Weissman, S., Slightom, J., Blechl, A., Smithies, O., Baralle, F., Shoulders, C. and Proudfoot, N. (1980) *Cell* 21, 653-668.
28. De Lange, T., Michels, P., Veerman, H., Cornelissen, A. and Borst, P. (1986) *Nucl. Acids Res.* 12, 3777-3790.
29. Parsons, M., Smit, J., Nelson, R., Stuart, K. and Agabian, N. (1984) *Mol. Biochem. Parasitol.* 10, 207-216.
30. Walder, J., Eder, P., Engman, D., Bretano, S., Walder, R., Knutzon, D., Dorfman, D. and Donelson, J. (1986) *Science* 233, 569-571.
31. Cornelissen, A., Verspiieren, M., Toulme, J., Swinkels, B. and Borst, P. (1986) *Nucl. Acids Res.* 14, 5605-5614.
32. Sloof, P., Van den Burg, J., Voogd, A., Benne, R., Agostinelli, M., Borst, P., Gutell, R. and Noller, H. (1985) *Nucl. Acids Res.* 13, 4171-4190.
33. Birkenmeyer, L., Sugisaki, H. and Ray, D.S. (1985) *Nucl. Acids Res.* 13, 7107-7118.
34. Kleisen, C., Borst, P. and Weijers, P. (1976) *Eur. J. Biochem.* 64, 141-151.
35. Hoeijmakers, J. and Borst, P. (1982) *Plasmid* 7, 210-220.
36. Cornish-Bowden, A. (1985) *Nucl. Acids Res.* 13, 3021-3030.