Characterization of RNA transcripts from the alpha tubulin gene cluster of *Leptomonas seymouri*

Vivian Bellofatto and George A.M.Cross

The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399, USA

ABSTRACT
       A tandem cluster of alpha tubulin genes was identified in the trypanosomatid protozoan Leptomonas seymouri. One repeat unit and the first gene of the cluster, with its upstream flanking region, were cloned and analyzed for their transcriptional and coding capacities. The 12-15 copies per cell of the 4 kb repeat encode a stable 2 kb transcript, which contains a mini-exon at its 5' end and two closely spaced polyadenylation sites. Transcription of the alpha tubulin gene cluster in isolated nuclei was unidirectional. Intergenic and coding regions were transcribed at the same rate, and nascent intergenic and coding region transcripts were quantitatively linked. These results are consistent with the possibility that the primary transcripts are polycistronic, or that there is a single small intercistronic gap.

INTRODUCTION
       The discontinuous synthesis of mRNAs is a distinct biochemical feature

of the trypanosomatid flagellate protozoa. A mature mRNA results from the

joining of the products of two transcription units. The initial 39

nucleotides [1], the mini-exon [2] or spliced leader [3], of a small (80-140

nucleotide, depending on the cell species) capped RNA (medRNA; mini-exon

donor RNA) are spliced to form the 5' end of all mature mRNAs studied so

far. Several models have been proposed to explain how the chimeric mRNAs

are produced, but recent data are consistent with post-transcriptional

trans-splicing of coding transcripts and medRNA [4-6]. Tandemly clustered

genes, of which several examples exist in trypanosomatids, might generate

polycistronic primary transcripts and RNA trans-splicing could then play a

pivotal role in regulating the generation of individual mature mRNA

molecules. Transcription of the variant surface glycoprotein gene expression

site in Trypanosoma brucei is polycistronic, as shown by recent UV

inactivation studies [7], but this could be a special case since these

transcripts are the product of an alpha-amanitin-insensitive RNA polymerase

[8,9].

To understand the process by which trypanosomatid mRNA synthesis is regulated, we are developing Leptomonas seymouri as a model for genetic manipulation, since the more complex trypanosomatids are not yet amenable to the cultivation and selection techniques commonly used for microbial genetics. We show that the alpha tubulin genes of L. seymouri are clustered as a single continuous unit, which yields a 2 kb mRNA containing a mini-exon sequence at its 5' end. Our hybridization data suggest that the alpha tubulin genes and the corresponding intergenic regions are co-transcribed.

MATERIALS AND METHODS

Cells

Leptomonas seymouri (ATCC 30220), originally isolated from the midgut of the hemipteran Dysdercus suturellus, was grown at $27^{o}C$ in Bone and Steinert's medium [10]. Cell cultures were maintained in logarithmic growth between 10^5 and 3×10^7 cells/ml.

Nucleic acids

DNA was isolated as described [11]. RNA was prepared from cultures at 5×10^6 cells/ml by guanidinium thiocyanate SDS lysis and poly(A) RNA was isolated as described [11]. For the copy number estimation, kinetoplast DNA was removed from genomic DNA preparations according to Borst [12]. The alpha tubulin clone pST3 was obtained by screening L. seymouri BamHI restriction fragments, cloned in pUC12, with a nick-translated alpha tubulin probe from T. brucei. The 1.4 kb HindIII DNA fragments, containing L. seymouri alpha tubulin flanking sequences, were electroeluted from agarose gels, ligated to pUC12 and transformed into E. coli DH1 [13]. Colonies were screened according to Grunstein and Hogness [14]. Subcloning into vectors pGEM-3 and pGEM-4 (Promega Biotec) was conducted using standard procedures.

Nuclease S1 mapping and primer extension analysis

Nuclease S1 mapping [15] was done as described in [16] except that 5' or 3' ^{32}P-labelled DNA probes and 40 ug total RNA or 5 ug poly(A) RNA were used, and solution hybridization was done at $57^{o}C$. The method used for primer extension was as described [17]. The extended DNA fragment was electroeluted from a denaturing polyacrylamide gel and sequenced by the method of Maxam and Gilbert [18].

Nuclear run-on transcription

The method used was essentially as described for T. brucei [8]. Leptomonas cells (2×10^{10}) in mid log phase were collected by

centrifugation at $25^{\circ}C$ and disrupted by abrupt N_2 pressure change in a Stansted cell disrupter. Nuclei were collected by centrifugation (2600 x g, 10 min), resuspended in 1 ml transcription buffer [8] containing 1.6 mCi [^{32}P]UTP (Amersham) and incubated at $27^{\circ}C$ for 5 min. RNase-free DNase I (Promega Biotec) was added (10 ug/ml), the nuclei were incubated at $37^{\circ}C$ for 5 min and then lysed with 2% SDS lysis buffer, deproteinated with 10 ug/ml protease K and phenol-chloroform extracted [8]. The lysate was divided into four aliquots and each was put through a 1 ml G-50 spin column to remove unincorporated rNTPs. The eluates were 90% acid precipitable and contained about one cpm/nucleus.

^{32}P-labelled nuclear RNA (1 x 10^7 cpm) was hybridized with excess (10 ug), _in vitro_ transcribed RNA containing sequences complementary or equivalent to the 5' most 273 bases of the alpha tubulin mRNA coding regions immobilized on slots of nitrocellulose. The hybridization buffer contained 0.45 M NaCl and 0.045 M sodium citrate (3 x SSC), 50% formamide, 2 mM EDTA, 8 mM Tris-HCl, pH 7.5, 0.2% sodium dodecyl sulfate (SDS) 1 mg/ml poly(A), 250 ug/ml tRNA. Incubations were at $42^{\circ}C$ for 48 h. All components were RNase-free. Unselected ^{32}P-labelled RNAs were carefully removed from hybridization bags, diluted two-fold with hybridization buffer and hybridized with immobilized RNA on slots at $42^{\circ}C$ for 48 h. Filters were washed in 2 x SSC, briefly with and then without 0.1% SDS at room temperature, then in 2 x SSC with 200 ug/ml RNase A for 30 min at $37^{\circ}C$ and finally in 0.2 x SSC at $55^{\circ}C$ for 1 h before autoradiography.

In vitro synthesized RNA

RNA transcripts were synthesized from linearized pGEM-3 or pGEM-4 templates as described by Promega Biotec protocols. [^{3}H]UTP (Amersham, 0.6 uCi/uMole) was added to facilitate RNA quantitation and transcripts were run on 2.2 M formaldehyde 1% agarose gels to ascertain RNA quality.

RESULTS

Alpha tubulin genes are arranged as a tandem repeat

A genomic library containing BamHI-generated restriction fragments of L. seymouri DNA cloned into the unique BamHI site of pUC12 was screened with a nick-translated probe containing alpha and beta tubulin sequences from T. brucei. All hybridized clones contained either alpha or beta tubulin sequences as determined by subsequent hybridization with alpha- or beta-specific probes. The five alpha tubulin clones had identical 4.0 kb inserts, as determined by restriction analysis; clone pST3 was used for
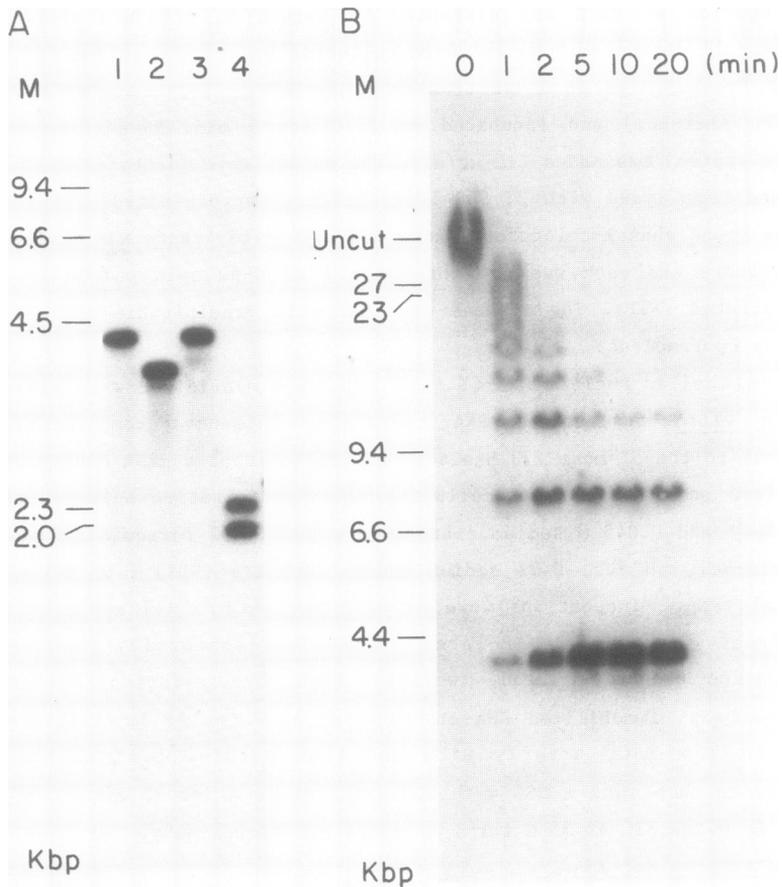
Figure 1. Organization of alpha tubulin genes. (A) Total L. seymouri DNA was digested with BamHI (1), PstI (2), SalI (3) or SphI (4), electrophoresed on a 1% agarose gel, blotted to nitrocellulose, and hybridized to the Leptomonas DNA insert of labelled pST3. The hybridized PstI fragment is 3.6 kb. The 0.4 kb PstI fragment, which is also part of the repeat unit, is a faint band near the bottom of the autoradiograph. The size markers are lambda DNA digested with HindIII. (B) Genomic DNA (30 ug) was digested with SalI for increasing time intervals after which an aliquot containing 5 ug of DNA was placed into 1% SDS. Digests were run on a 0.4% agarose gel, blotted onto nitrocellulose and probed with ^{32}P-labelled pST3. Markers are lambda HindIII digestion products.

further analysis. To confirm the identity of the alpha tubulin gene, a small

amount of sequence data were generated. Comparison of the 84 amino-terminal

amino acids revealed 93% or 90% homology to the alpha tubulin genes of T.

brucei [19] and Leishmania enrietti [20] respectively, and strong similarity

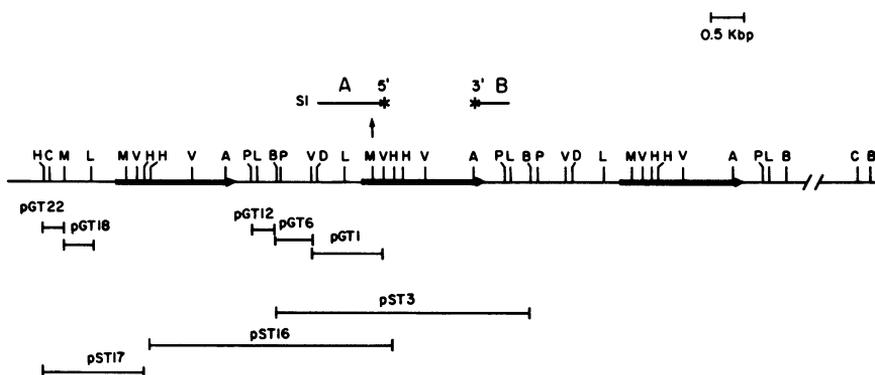to the alpha tubulin genes of other organisms [21,22].

Figure 2. Genomic map and clones of the alpha tubulin gene cluster. The narrow line represents genomic DNA and the superimposed arrow represents the alpha tubulin mRNA, as deduced from Northern blot and S1 nuclease mapping data. The lines above the map show the probes used for S1 nuclease mapping and primer extension; the asterisks represent the ^{32}P-labelled termini. Bracketed lines below the map indicate clones isolated from the designated areas. Clones pST3, pST16 and pST17 are pUC12 derivatives. All pGT clones are pGem-3 derivatives. Insert sizes are 280 bp (pGT22), 420bp (pGT18), 285bp (pGT12), 550bp (pGT6) and 1050bp (pGT1). Shown are all the BamHI (B), PvuII (V), HindIII (H), PstI (P) and Cla (C) sites and a subset of the MaeIII (M), BglI (L), DdeI (D), and AvaII (A) sites. The arrowhead designates the position of the 20 bp HaeIII-MaeIII fragment used for primer extension sequencing (Fig. 6).

The organization of alpha tubulin sequences within the L. seymouri genome was investigated by probing Southern blots of restriction enzyme-digested genomic DNA with ^{32}P-labelled pST3. Single hybridized fragments of 4.0 kb in lanes 1 and 3 (Fig. 1A) or summing to 4.0 kb (lane 4) indicated that either a single 4.0 kb copy of the alpha tubulin gene, flanked by equally spaced BamHI, SalI and a subset of SphI sites, is present in the genome, or that there are multiple clustered copies of this sequence. To distinguish between these two possibilities, genomic DNA was partially digested with SalI, fractionated by electrophoresis on a 0.4% agarose gel, transferred by the Southern technique and probed with pST3. The ladder of hybridized fragments indicated that about 12 tandem copies of the tubulin gene are present in L. seymouri (Fig. 1B). By comparing intensity of hybridization between completely BamHI digested genomic DNA and pST3 DNA using the pST3 probe on a Southern blot, a value of 12-15 alpha tubulin genes per haploid genome was obtained assuming a genome size of 5×10^4 kb [23] (data not shown). If all copies of alpha tubulin sequences are present in a tandem array then several 6-base sequence recognition restriction
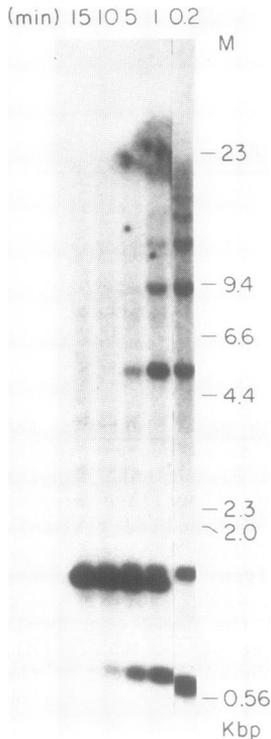
Figure 3. Identification of 5' flanking alpha tubulin sequences. L. seymouri DNA was completely digested with ClaI to generate the intact alpha tubulin gene repeat unit, and then was incubated with HindIII for increasing amounts of time. DNA was separated on a 0.4% agarose gel and transferred to nitrocellulose. The probe was the 280 bp insert from pGT22 (Fig. 2). The 0.6 kb hybridized fragment is an upstream ClaI fragment that contains sequences which flank the upstream HindIII site. Markers are HindIII-digested ⸱ lambda DNA.

enzymes, known not to recognize sequences within the alpha tubulin gene repeat, should generate only large restriction fragments that hybridize to pST3. Seven such restriction enzymes generated a null digestion pattern for alpha tubulin (data not shown). The absence of any smaller hybridizing fragments detected by the alpha tubulin probe indicated that there are no detectable orphon copies of alpha tubulin genes in the L.seymouri genome. A restriction map representing the essential features of alpha tubulin gene organization is depicted in Fig. 2.

To ensure that pST3 contained an entire alpha tubulin gene repeat and did not lack sequences that might be present on an uncloned BamHI fragment, we isolated an alpha tubulin clone (pST16; Fig. 2) from a partial L. seymouri HindIII genomic library. Restriction analysis demonstrated that pST16 and pST3 are permuted replicas of each other.

Characterization of the 5' flanking tubulin gene

To identify 5' flanking sequences of the alpha tubulin gene cluster, genomic DNA was digested with HindIII, an enzyme that cleaves at two closely spaced sites within the alpha tubulin coding region (Fig.2), electrophoresed
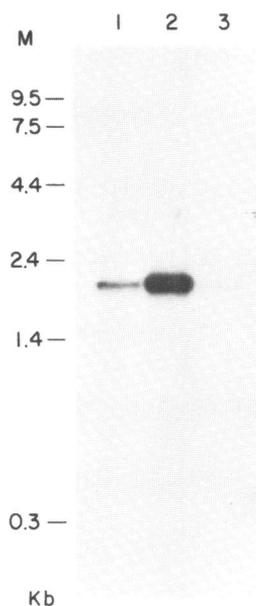
Figure 4. Identification of alpha tubulin mRNA.
8 ug total RNA (lane 1), 3.8 ug poly(A) RNA
(lane 2), or 8 ug non-poly(A) RNA (lane
3), were electrophoresed on a 1.5% agarose-2.2 M
formaldehyde gel, blotted onto Nytran
(Schleicher and Schuell) and hybridized with
labelled pST3. Markers were from the Bethesda
Research Laboratories' RNA marker kit.

on a 1% agarose gel, Southern blotted and probed with pST3. In addition to
a strongly hybridized 3.9 kb fragment, representing the tubulin gene repeat
(minus the internal 0.1 kb HindIII fragment), two less intensely hybridized
bands of 10.4 kb and 1.4 kb were detected (data not shown). The region of
the gel containing the 1.4 kb fragment, which hybridized to a 5' and not to
a 3' coding region probe, was electroeluted and cloned into pUC12. By
restriction analysis, this clone (pST17) was shown to diverge from pST3
approximately 500 bp upstream from the 5' end of the mature alpha tubulin
mRNA (Fig. 2). To determine if pST17 contained sequences 5' to the tubulin
repeat units or contained part of an undetected orphon or pseudo alpha
tubulin gene, a restriction fragment containing sequences unique to pST17
was used to probe partially restricted genomic DNA immobilized on a Southern
blot. A hybridized ladder of fragments, starting at 1.4 kb and increasing
by 3.9 kb increments, was obtained (Fig. 3). This result indicated that the
1.4 kb HindIII clone (pST17) is located adjacent to the alpha tubulin gene
repeat.

Characterization of the mature 5' and 3' ends of alpha tubulin mRNA

pST3 hybridized to a single 2.0 kb RNA species present in total RNA and
poly(A) RNA preparations (Fig. 4). To identify the 5' terminus of the alpha
tubulin mRNA, nuclease S1 mapping and primer extension analysis were done.
A map of fragments used for these and all subsequent experiments is shown in
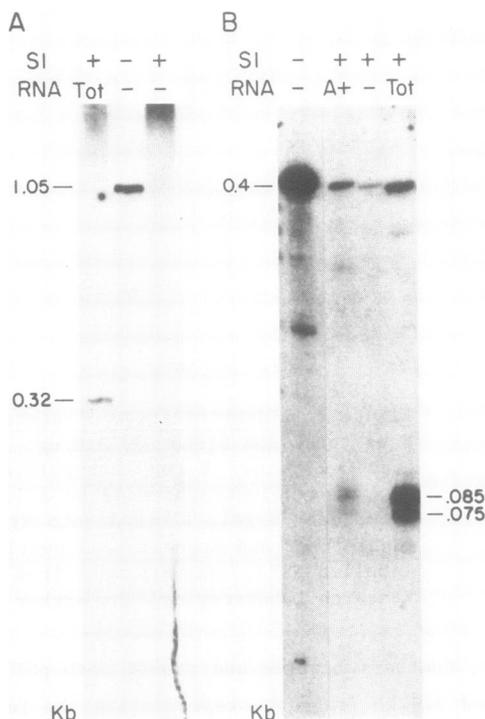
Figure 5.   Identification of the 5' and 3' ends of alpha tubulin mRNA.   (A) A 1.1 kb PvuII-PvuII fragment contained within the Leptomonas DNA inset of clone pST3 was 5' labelled with [^{32}P]ATP and then digested with DdeI to generate a uniquely labelled 1.05 kb fragment, (probe A), 0.1 ug of which was hybridized to 30 ug total L. seymouri RNA.   Nuclease S1 was added and the reaction products were electrophoresed on a 5% polyacrylamide 7 M urea gel and autoradiographed. (B) A 0.4 kb AvaII-BglI DNA fragment, ^{32}P-labelled at the 3' end of the AvaII site, was hybridized to poly(A) or total RNA, treated with nuclease S1, electrophoresed on a 8% polyacrylamide 7 M urea gel and autoradiographed.

Fig. 2.   DNA probe A (Fig. 2) was hybridized with total L. seymouri RNA, treated with nuclease S1 and analyzed on a denaturing acrylamide gel.   A single protected DNA fragment indicated that the 5' end of the coding exon is 320 bases from the labelled PvuII site (Fig. 5A, lane 1).   Identical results were obtained when 5 ug of poly(A) RNA was used in the reaction.   To determine the sequence of the 5' terminus of the alpha tubulin mRNA, a 20 bp MaeIII-HaeIII fragment, located approximately 100 bases from the 5' end of the tubulin mRNA as defined by the S1 experiment, was 5' end-labelled at the MaeIII site, hybridized to total mRNA, and extended with AMV reverse

```
                                                        -20
  px    5'                                    AACGCTATATAAGTATCAGTTT
  gene  5'    GCGCACGCACAACCACATAAAATAAAACAAAACGGTTGCGCATCCACCGCCCCCTCC

                    +I
  px          CTGTACTTTATTGCTCAACGCCTTCTTCAAACCTCTAACACACTCTCTTCAACTCTT
  gene        CACCCAAACGCAGCTCAACGCCTTCTTCAAACCTCTAACACACTCTCTTCAACTCTT
                      ↟

  gene        TAAACCCCTTCCAAGATGCGTGAGGCTATCTGCATTCACATCGGCCAGGCTGGCTAC
                         M  R  E  A  I  C  I  H  I  G  Q  A  G  Y

  gene        CAAATTGGTAACGCTTGCTCCCAGCTCTTCTGC
              Q  I  G  N  A  C  S  Q  L  F  C
```

Figure 6. <u>Alpha tubulin mRNA splice site and mini-exon sequence of L. seymouri</u>
The genomic sequence was determined by Maxam and Gilbert sequencing of clone
pST3. The primer-extended sequence (px) was deduced from similar sequencing
of the extended product using the complementary 20 bp primer (overlined) as
indicated in Fig. 2. The arrow represents the mini-exon addition site and the
underlined 'AG' is the canonical sequence present at the 3' end of introns.

transcriptase. The single product observed on a denaturing acrylamide gel

was eluted and sequenced. The sequence obtained was colinear with the

genomic alpha tubulin cloned DNA for 70 nt then diverged, revealing 35 nt of

a mini-exon sequence at the 5' end of the alpha tubulin mRNA (Fig. 6). The

ability to deduce a 35 nt sequence by primer extension suggests that the <u>L.</u>

<u>seymouri</u> mini-exon is longer than 35 nt, as in <u>T. brucei</u> [1]. The mini-exon

sequence was confirmed with genomic clones, and subsequent experiments have

shown that it is transcribed as an 85 nt RNA from a set of reiterated mini-

exon genes unrelated to the alpha tubulin gene cluster (data not shown).

     The position of the 3' terminus of the alpha tubulin mRNA was

determined by S1 analysis using a 0.4 kb <u>AvaII-BglI</u> DNA fragment labelled

with ^{32}P at the 3' end of an <u>AvaII</u> site (probe B, Fig.2). Probe B was

hybridized to poly(A) or total RNA, treated with nuclease S1 and analyzed

(Fig. 5B). The data indicate that either there are two closely spaced

poly(A) addition sites within a single repeated unit or that there are two

classes of nearly identical alpha tubulin genes within the gene cluster,

each possessing a slightly differently located poly(A) addition site.

     S1 mapping data, coordinated with the pST3 restriction map and sequence

data and Northern hybridization using the pST3 probe, demonstrated that

tubulin mRNA is encoded by 2.0 kb of a 4.0 kb repeat unit. Hybridization of

Northern blots containing total and poly(A) RNA with ^{32}P-labelled probes

generated from restriction fragments within the intergenic regions

demonstrated that no stable mRNA is encoded within this region (data not

shown). The absence of any "extra" protected fragments (or primer extended

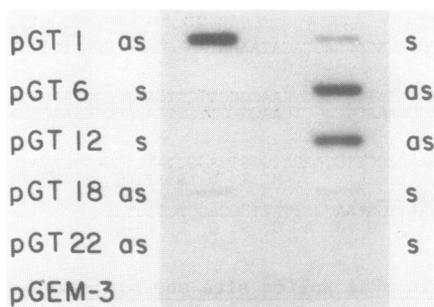sequences) from the nuclease S1 and primer extension experiments is

Figure 7. Analysis of ^{32}P-labelled nascent transcripts. Slots were loaded with 5 ug of in vitro synthesized RNA representing regions of the alpha tubulin gene cluster as shown in Fig. 2. S-labelled slots contain RNA equivalent to alpha tubulin mRNA and pre-mRNA sequences. AS-labelled slots contain RNA complementary to the alpha tubulin mRNA and pre-mRNA sequences.

consistent with these data. However, sequences immediately 5' to the mini-exon alpha tubulin mRNA junction are "intron-like" in certain respects (see Fig. 6). The AG dinucleotide in the genomic sequence at the junction site between mini-exon and mature alpha tubulin mRNA sequences is a canonical intron-exon junction in eukaryotic cells [24]. In addition, the pyrimidine-rich region from positions -9 to -27 upstream from the mini-exon alpha tubulin mRNA junction in L. seymouri (see Fig.6) is similar to the genomic sequence that preceeds many T. brucei steady-state mRNAs, and may serve as a recognition signal for RNA splicing enzymes [4].

Transcription of alpha tubulin genes in isolated nuclei

To detect transcription within the alpha tubulin repeat, restriction fragments that contained intergenic and intragenic sequences, as defined by S1 and Northern analyses, were cloned into pGEM-4, and antisense and sense RNAs were synthesized using T7 or SP6 RNA polymerase on templates linearized at a unique restriction site within the polylinker. These RNAs were immobilized on nitrocellulose filters and hybridized with ^{32}P-labelled transcripts generated by labelling nascent RNA chains in isolated nuclei. Fig. 7 shows that RNA synthesis occurs between the DNA sequences that encode mature alpha tubulin mRNA and that this transcription is directionally equivalent to that of the alpha tubulin mRNA. Counting of hybridized radioactivity per 100 bases of immobilized RNA, and densitometric tracings of hybridized slots, demonstrated that the rate of transcription was the same in intergenic and intragenic regions (Table I). A clone containing only unique DNA (pGT22) and a second subclone (pGT18) of pST17 (Fig. 2), which

Table 1. Rate of transcription through the alpha tubulin gene cluster.

Template	Complementary Immobilized RNA (nt)	Experiment Number	Density* Units per 100 nt	^{32}P cpm per 100 nt
pGT22	280	1	<0.1	0
		2	<0.1	0
pGT18	420	1	10	
		2	2.0	14
pGT12	285	1	62	
		2	23	52
		3	110	178
pGT6	550	1	46	
		2	12	33
		3	190	132
pGT1	1050	1	48	
		2	11	23
		3	134	84

All values were corrected for background, which was hybridization to pGEM-3 transcripts without inserts. *Densitometric values are in arbitrary units.

contained approximately 170 bp of alpha tubulin repeat sequences and 250 bp of unique upstream sequences, were used as DNA templates to detect antisense and sense RNA transcripts for nascent RNA mapping studies. Neither RNA strand from the unique sequence of pGT22 detected ^{32}P-labelled nuclear run-on RNA. Increasing the specific activity of nuclear RNA two-fold by adding twice the normal amount of [^{32}P]UTP to the transcription reactions did not change this result. Although RNA-RNA hybridization to detect nascent transcription is a sensitive assay, transcription of a single copy DNA sequence at the same rate as the tubulin gene cluster would be only slightly above background. Therefore, within the limits of sensitivity of this assay, there appears to be no transcription from the unique upstream region, indicating that the promoter for alpha tubulin gene transcription may be within this region.

To analyze the continuity of intergenic and coding-region transcription, ^{32}P-labelled nascent nuclear RNA was hybrid-depleted on nitrocellulose filters containing RNA complementary to or equivalent to the 5' end coding region of alpha tubulin mRNA. Equal amounts of acid-precipitable ^{32}P-labeled RNA that did not hybridize to the immobilized
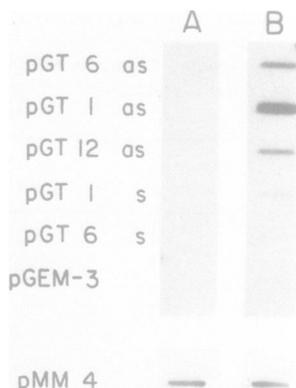
Figure 8. <u>Hybrid depletion of nascent transcripts.</u> ^{32}P-labelled RNA (3 x 107 cpm) hybridized to probes in lane A was first hybrid-selected by immobilized RNA (10 ug) complementary to 273 nt of the 5' portion of the alpha tubulin mRNA. The immobilized RNA was generated by <u>in vitro</u> transcription of AhaIII-restricted pGT1. Radioactive RNA hybridized to probes in lane B was first incubated with immobilized RNA sequences (10 ug) equivalent to the 5' end of the alpha tubulin mRNA. Clones listed indicate the templates used to generate antisense or sense RNA bound on filter. pMM4 is a clone of the genomic copy of the medRNA gene of <u>L. seymouri</u>. Hybridization to pMM4 by RNA from both hybrid depletion indicated that the RNAs in each experiment remained intact.

RNAs were then used to probe filters containing RNA complementary to or equivalent to the transcribed intergenic region (Fig. 8). Results showed that ^{32}P-labelled RNA transcribed from the intergenic regions of the alpha tubulin sequences was removed from the ^{32}P-labelled RNA population by RNA complementary to alpha tubulin mRNA. Thus, RNA transcription through the intergenic regions wascontigious with the alpha tubulin mRNA coding sequence.

DISCUSSION

We have determined that alpha tubulin gene expression includes transcription of the sequences between the RNA coding regions. Regions of the transcripts representing intergenic sequences were very unstable and detectable only in nascent transcripts. These observations may explain why intergenic transcripts could not be detected during analysis of the alpha tubulin gene cluster in <u>Leishmania enrietti</u> [20,25] or, until recently, in the <u>T. brucei</u> alpha and beta tubulin gene cluster [26,27]. However, the demonstration of small amounts of oligomeric RNAs [28] or intergenic RNAs

[26] does not prove their putative precursor status. In our analyses of the L. seymouri alpha tubulin gene cluster, all of the intergenic transcripts were covalently linked to mRNA coding sequences. The crucial evidence for this comes from the observation that, if a coding region probe was prehybridized to nascent RNA, no nascent RNA remained to hybridize to intergenic probes. In addition, intergenic and coding regions were transcribed at similar rates in isolated nuclei. Thus, all coding-sequence RNA must be linked to intergenic sequences. We envision two possible transcriptional modes that could generate these data. Coordinate transcription of 4 kb monocistronic units (the mature mRNA is 2 kb), initiating and terminating at specific sequences within the intergenic regions, could occur. A short, unique gap in transcription somewhere within each intergenic region, which would be diagnostic of this scenario, would not have been distinguished by our experiments, but multiple random breaks would have been. This type of coordinated transcription would be similar to that observed in the transcription of tandem rRNA gene repeats in at least two higher eukaroytes [29,30]. Alternatively, polycistronic transcription may occur. In this case, our inability to identify multimeric copies of alpha tubulin mRNAs on northern blots (even when probed with intergenic DNAs) would be explained by co-transcriptional polyadenylation and/or medRNA addition.

Analysis of transcription through a region of identical or nearly identical gene repeats, as is the case of the alpha tubulin gene cluster in L. seymouri, cannot be accurately done by UV inactivation analysis. Primary transcripts can, however, be characterized by first identifying the sequences in the gene cluster that are transcribed into nascent RNA, relating the rates of nascent RNA transcription from one region to another, and finally establishing and quantitating the relationship of nascent RNA to mature RNA. To delineate the transcription unit(s) within the alpha tubulin genes of L.seymouri, we initially defined the boundaries of steady state RNA, and determined that, within the limitations of the experimental system, similar quantities of nascent RNA were generated from the intergenic and intragenic regions. We therefore concluded that the primary transcripts of this gene cluster are at least 4 kb long. Thus, the read-through RNAs of other trypanosome genes [28,31] likely represent true precursors and not aberrant reactions. Our data are consistent with transcription of the alpha tubulin genes as a long polycistronic or monocistronic unit containing, in either case, extensive regions that must be exo or endonucleolytically

trimmed, during generation of the functional mRNA. The inability to detect any potential precursor molecules in steady-state alpha tubulin mRNA may reflect the rapid processing, by polyadenylation and mini-exon splicing, of alpha tubulin precursors into mRNA. However, proof of the nature and location of the regulatory signals for RNA transcription and splicing in trypanosomes, as well as proof that transcription elongation in isolated nuclei reflects in vivo events, must await the development of systems for DNA transformation of trypanosomatids, which is the reason for our current work on L. seymouri.

Details of the structure and organization of L. seymouri mini-exon genes and medRNA transcripts will be the subject of another publication. Crithidia fasciculata [32], Leishmania enrietti [33] and L. seymouri have an identical mini-exon sequence, which differs significantly from the published genomic mini-exon sequence of Leptomonas collosoma [34], which may or may not represent a functional sequence.

REFERENCES

1. Freistadt, M.S., Cross, G.A.M., Branch, A.D. and Robertson, H.D. (1987) Nucleic Acids Res. 15, in press.
2. Boothroyd, J.C. and Cross, G.A.M. (1982) Gene 20, 281-289.
3. Parsons,M., Nelson, R.G., Watkins, K.P. and Abagian, N. (1984) Cell 38, 309-316.
4. Sutton, R.E. and Boothroyd, J.C. (1986) Cell 47, 527-535.
5. Murphy, W.J., Watkins, K.P. and Agabian, N. (1986) Cell 47,517-525.
6. Laird, P.W., Zomerdijk, J.C.B.M., deKorte, D. and Borst, P. (1987) EMBO J. 6, 1055-1062.
7. Johnson, P.J., Kooter, J.M. and Borst, P. (1987) Cell 51, 273-281.
8. Kooter, J.M. and Borst, P. (1984) Nucleic Acids Res. 12, 9457-9471.
9. Kooter, J.M., van der Spek, H.J., Wagter, R., d'Oliveira, C.E., van der Hoeven, F., Johnson, P.J. and Borst, P. (1987) Cell 51, 261-272.
10. Bone, G.J. and Steinert, M. (1956) Nature 178, 308.
11. Cully, D.F., Ip, H.S. and Cross, G.A.M. (1985) Cell 42, 173-182.
12. Borst, P., van der Ploeg, M., van Hoek, J.F., Tas, J. and James, J. (1982) Mol. Biochem. Parasit. 6, 13-23.
13. Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.
14. Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965.
15. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
16. Feingold, J., Bellofatto, V., Shapiro, L. and Ameniya, K. (1985) J. Bact. 163, 155-166.
17. Singh, K., Caray, M., Saragusti, S. and Botchan, M. (1985) Nature 314, 553-556.

18. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
19. Kimmel, B.E.,Samson, S., Wu, J., Hirschberg, R. and Yarbrough, L.R. (1985) Gene 35, 237-248.
20. Landfear, S.M., Miller, S.I. and Wirth, D.F. (1986) Mol. Biochem. Parasit. 21, 235-245.
21. Lemischka, I. and Sharp, P.A. (1982) Nature 300, 330-335.
22. Cowan, N.J., Dobner, P.R., Fuchs, E.V. and Cleveland, D.W. (1983) Mol. Cell. Biol. 3, 1738-1745.
23. Leon, W., Fouts, D.L. and Manning, J. (1978) Nucleic Acids Res. 5, 491-504.
24. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Ann. Rev. Biochem. 55, 1119-1150.
25. Landfear, S.M. and Wirth, D.F. (1985) Mol. Biochem. Parasitol. 15, 61-82.
26. Imboden, M.A., Laird, P.W., Affolter, M. and Seebeck, T. (1987) Nucleic Acids Res. 15, 7357-7368.
27. Thomashow, L.S., Milhausen, M., Rutter, W.J. and Agabian, N. (1983) Cell 32, 35-43.
28. Gonzalez, A., Lerner, T.J., Huecas, M., Sosa-Pineda, B., Nogueira, N. and Lizardi, P.M. (1985) Nucleic Acids Res. 13, 5789-5803.
29. Henderson, S. and Sollner-Webb, B. (1986) Cell 47, 891-900.
30. McStay, B. and Reeder, R.H. (1986) Cell 47, 913-920.
31. Gibson, W.C., Swinkels, B.W. and Borst, P. J. Mol. Biol. in press.
32. Muhich, M.L., Hughes, D.E., Simpson, A.M. and Simpson, L. (1987) Nucleic Acids Res. 15, 3141-3153.
33. Miller, S.I., Landfear, S.M. and Wirth, D.F. (1986) Nucleic Acids Res. 14, 7341-7361.
34. Milhausen, M., Nelson, R.G., Sather, S., Selkirk, M. and Agabian, N. (1984) Cell 38, 721-729.