J. Mol. Biol. (1986) **188**, 393–402

Tubulin mRNAs of *Trypanosoma brucei*

M. Imboden[1], B. Blum[1], T. DeLange[2]†, R. Braun[1] and Th. Seebeck[1]

[1] *Institut für Allgemeine Mikrobiologie Baltzerstrasse 4, CH-3012 Bern, Switzerland*

[2] *The Netherlands Cancer Institute Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands*

(Received 15 August 1985, and in revised form 9 December 1985)

The tubulin genes of *Trypanosoma brucei* are located in a single, tightly packed cluster of ten tandemly arranged alternating α and β-genes. No tubulin genes are detected outside the clustered array. Therefore, the cluster can be assumed to be the locus of tubulin gene expression. Single bands of α and β-tubulin mRNAs are observed in cultured procyclic as well as in bloodstream trypanosomes. Both α and β-tubulin mRNAs have distinct 5′ termini, which carry a 35-nucleotide mini-exon sequence. The 3′ termini of both mRNA populations are heterogeneous.

1. Introduction

The cytoskeleton of the parasitic hemoflagellate *Trypanosoma brucei* consists predominantly of microtubule-based structures (Vickerman & Preston, 1976); namely, a tight layer of pellicular microtubules that encloses the entire cell body, and the flagellar axoneme. The trypanosomal tubulins have been isolated and characterized (Stieger *et al.*, 1984; K. Gull, personal communication). The architecture and biochemistry of the trypanosomal microtubules persist without significant alterations throughout the complex life-cycle of the organism. Thus, genes that code for microtubular components can be regarded as *bona fide* household genes and might be suitable substrates for analyzing the mechanisms of gene expression in trypanosomes.

The tubulin genes of *T. brucei* are arranged in tightly packed clusters of tandemly repeated alternating α and β-genes (Seebeck *et al.*, 1983; Thomashow *et al.*, 1983). The basic repeat unit of each cluster has a length of 3·7 kb† and contains one α and one β-gene, whose coding regions are separated from each other by a few hundred base-pairs only. The nucleotide sequence of the entire repeat unit has been determined (Kimmel *et al.*, 1985). Analyses of the genome of trypanosomes by pulsed field gradient electrophoresis have presented evidence that the tubulin gene cluster is probably diploid and that it is located in large chromosomes (Gibson *et al.*, 1985).

Recent studies of *T. brucei* mRNAs have led to the observation that many of them carry an identical 35-nucleotide sequence at their 5′ terminus (Boothroyd & Cross, 1982; Van der Ploeg *et al.*, 1982; DeLange *et al.*, 1983; Parsons *et al.*, 1984*a*). This mini-exon has so far been studied mainly in conjunction with the transcription of the variable surface glycoprotein genes. The functional role of this structure is not understood, though it is already clear that its presence is not correlated with the transposition-mediated mode of gene expression exhibited by many of these genes.

This study describes the transcriptional organization of the tubulin gene repeat unit and analyzes the tubulin mRNA in *T. brucei*. A single species of each α and β-tubulin mRNA is detected, and both carry the mini-exon sequence at their 5′ termini.

2. Materials and Methods

(a) *Trypanosomes*

Trypanosomes (*Trypanosoma brucei brucei* stock 427, variants 117b and 118a, described by Cross, 1975) were isolated from infected rat blood. Procyclics of stock 427 and of stock STIB 366 were grown in SDM-79 (Brun & Schönenberger, 1979).

(b) *Pulsed field gradient electrophoresis*

The PFG apparatus was constructed according to the instructions of Schwartz & Cantor (1984). Trypanosomes

† Present address: Department of Microbiology and Immunology, University of California, San Francisco, CA 94143, U.S.A.

† Abbreviations used: kb, 10^3 bases or base-pairs; bp, base-pair(s); PFG, pulsed field gradient electrophoresis; SDS, sodium dodecyl sulfate; DTT, dithiothreitol; VSG, variable surface glycoproteins.

0022-2836/86/070393-10 $03.00/0

© 1986 Academic Press Inc. (London) Ltd.

were lysed *in situ* in agarose blocks (Van der Ploeg *et al.*, 1984*a*) and the DNA was digested with restriction enzymes following the procedure of Gibson *et al.* (1985). Running conditions were similar to those described by Van der Ploeg *et al.* (1984*a*).

(c) *RNA blotting*

Trypanosomes grown *in vitro*, or total infected rat blood, were mixed with 10 vol. ice-cold 3 M-LiCl, 6 M-urea and were homogenized in an Omnimix homogenizer for 1 min at full speed. RNA was precipitated from this mix overnight on ice and collected by centrifugation (12,000 revs/min in a Sorvall HB-4 rotor at 0°C). The precipitate was dissolved in 10 mM-Tris·HCl (pH 7·5), 10 mM-EDTA, 0·2% (w/v) SDS, and treated with proteinase K, followed by extraction with phenol and precipitation with ethanol. Poly(A)$^+$ RNA was then isolated as described (Van der Ploeg *et al.*, 1982). Portions (2 μg) of poly(A)$^+$ RNA were treated with glyoxal and size-separated according to the procedure of Thomas (1980).

Alternatively, RNA was extracted from trypanosomes by the guanidinium/isothiocyanate procedure (Harding *et al.*, 1978), poly(A)$^+$ RNA was isolated according to Ryffel & McCarthy (1975), and the RNA was size-fractionated on formaldehyde-containing agarose gels (Hall *et al.*, 1983).

After transfer to nitrocellulose, RNA was hybridized with ^{32}P-labeled DNA probes at 42°C in 5 × SSC (SSC is 0·15 M-NaCl, 0·015 M-trisodium citrate, pH 7), 50% formamide, 1% SDS, 5 × Denhardt's solution, 100 μg/ml depurinized calf thymus DNA/ml and washed according to Thomas (1980).

(d) *S_1 nuclease mapping*

S_1 mapping was done essentially according to Weaver & Weissmann (1979). The S_1 nuclease digestion was performed at room temperature (Miller & Sollner-Webb, 1981). The reaction was stopped by the addition of ammonium acetate and EDTA to 0·4 M and 0·01 M, respectively, followed by several extractions with phenol/$CHCl_3$. The remaining RNA was then hydrolyzed in 0·2 M-NaOH for 15 min at 60°C. After neutralization, 6 μg of yeast tRNA was added as carrier, and the samples were precipitated with ethanol. The dry nucleic acids were then dissolved in 3 μl of water and 3 μl of loading buffer (96% formamide, 20 mM-EDTA, pH 8·3), heated for 10 min in a boiling waterbath and analyzed on 6% or 8% polyacrylamide gels (Sanger *et al.*, 1977).

(e) *Primer extension sequencing*

(i) *Plasmid-derived primer*

Appropriately 5′ end-labeled fragments were isolated by gel electrophoresis (Maxam & Gilbert, 1980) and further purified by DEAE-cellulose chromatography, extraction with phenol/$CHCl_3$ and precipitation with ethanol. A sample (10 ng) of purified fragment was then hybridized with 10 μg of poly(A)$^+$ RNA (Weaver & Weissmann, 1979) in a final volume of 10 μl under a layer of paraffin oil. After annealing for 3 to 16 h at 57°C, the nucleic acids were precipitated with ethanol twice and the final pellet was washed several times with 80% ethanol. Primer extension was performed in 100-μl reactions containing 50 mM-Tris·HCl (pH 8·3), 6 mM-$MgCl_2$, 40 mM-KCl, 0·03 M-β-mercaptoethanol and 83 units of reverse transcriptase for 1 h at 37°C. The following nucleotide triphosphate concentrations were used: the full reaction contained each of the 4 deoxy-triphosphates at 200 μM. The 4 sequencing reactions were each adjusted to 7 μM in the respective dideoxy-derivative and 25 μM in its corresponding deoxy-triphosphate, while the other 3 deoxy-triphosphates were maintained at 200 μM. The reactions were stopped and the products extracted and analyzed exactly as described above for S_1-mapping.

(ii) *Synthetic primer*

Portions (100 ng) of 5′ end-labeled oligonucleotide were annealed with 5 μg of poly(A)$^+$ RNA in 83 mM-Tris·HCl (pH 8·3), 0·25 M-KCl, 16 mM-$MgCl_2$, 33 mM-DTT in a final volume of 15 μl at 50 to 55°C for 2 h. Then 3 μl of the annealed nucleic acids were mixed with 3 μl of nucleotide triphosphates and 1 μl of reverse transcriptase (4 units), and the reaction was left to proceed for 30 min at 42°C. Nucleotide triphosphates were present at 250 μM for the full reaction, and at 25 μM (dideoxy-derivative) and 63 μM (corresponding deoxy-triphosphate) for each sequencing reaction. Reactions were stopped by adding 4 μl of loading buffer (see S_1 mapping) and incubation for 10 min in a boiling waterbath.

(f) *DNA labeling and sequencing procedures*

Nick-translation of DNA, 5′ end-labeling with polynucleotide kinase and 3′ end-labeling with Klenow polymerase were all done according to Maniatis *et al.* (1982). 3′ end-labeling of recessed ends with [^{32}P]-cordycepin was performed by boiling the fragment for 1·5 min in water immediately before adding all other components of the reaction. The labeling reaction itself was done according to the instructions of the supplier of the enzyme (Bethesda Research Laboratories). DNA sequencing was done according to Sanger *et al.* (1977), using ^{35}S-labeled adenosine triphosphate.

(g) *Plasmid pTBtu9A*

The structure of this plasmid has been published (Seebeck *et al.*, 1983). It contains 1 entire tubulin gene repeat unit (*Bam*HI–*Bam*HI) cloned into the *Bam*HI site of pBR322.

(h) *Materials*

Reverse transcriptase was obtained from Cambridge Biotechnology Laboratories and terminal transferase from Bethesda Research Laboratories, while all other enzymes were from Boehringer-Mannheim. Radioactive nucleotides were purchased from New England Nuclear and unlabeled deoxy and dideoxy-triphosphates from Pharmacia. The synthetic primers were synthesized by K. Kaluza at the Department of Microbiology of the Eidgenoessische Technische Hochschule in Zürich. Their sequences were as follows: α-primer: 5′-A-C-C-T-A-C-T-G-T-T-G-C-T-A-A-T-A-T-A-A-3′; β-primer: 5′-C-T-G-G-T-T-A-C-C-G-C-A-T-T-G-G-C-C-A-G-3′.

3. Results

(a) *The tubulin gene cluster*

The tubulin gene clusters of *T. brucei* have been shown to be located on large chromosomes (Gibson *et al.*, 1985; Van der Ploeg *et al.*, 1984*b*). Since considerable variations in the size of the DNA fragments containing the tubulin gene cluster were

observed in this study between different trypanosome stocks, it was necessary to investigate the chromosomal organization of tubulin genes in the particular stock (STIB 366) used in our experiments by PFG. Figure 1(a) presents the pattern of undigested and of *Xho*I or *Hpa*I-digested DNA from stock STIB 366 as visualized by staining with ethidium bromide and by hybridization with the tubulin-specific probe pTBtu9A. In undigested DNA, tubulin sequences are found both in the DNA remaining close to the slot and in the DNA migrating in the "megabase band" (Gibson *et al.*, 1985). This double localization of the tubulin genes is attributed to non-specific trapping of DNA in the slot. Similar variable distributions of DNA between slot and megabase band have been observed for a number of trypanosomal genes (Gibson *et al.*, 1985; Van der Ploeg *et al.*, 1984*b*). The restriction endonucleases *Xho*I and *Hpa*I, which both do not cleave within the tubulin gene cluster of the trypanosome stocks examined by Seebeck *et al.* (1983) and Thomashow *et al.* (1983), generate single hybridizing bands of about 70 kb and 40 kb, respectively.

Since the hybridization conditions used allow the easy detection of single copy genes (compare Fig. 1(b), bands of 7 and 12 kb), Figure 1(a) clearly demonstrates that in stock STIB 366 no single copies of tubulin genes are found outside the cluster. This observation establishes that this tubulin gene cluster is indeed the locus of tubulin gene transcription, though it is not clear if the entire cluster, or only a subpopulation of its tubulin genes, is transcriptionally active.

Digestion of trypanosomal DNA within two restriction enzymes, *Eco*RI and *Bam*HI, which both cut within the tubulin gene repeat, revealed upon hybridization with the tubulin-specific probe pTBtu9A the presence in each digest of two single copy bands, in addition to the reiterated 3·7 kb repeat unit. The two minor bands generated by *Eco*RI digestion (12 and 7 kb) represent the 5′ end (12 kb fragment) and the 3′ end (7 kb fragment), while in the *Bam*HI digest, the 9 kb band represents a 5′ end and a 6 kb band represents the 3′ end of the tubulin gene cluster (unpublished results).

(b) *The α and β-tubulin mRNAs differ in size*

As an initial characterization of tubulin mRNA, poly(A)$^+$ RNAs from cultured procyclics of stock STIB 366 (Fig. 2(a)), from two variants of bloodstream trypanosomes (117b and 118a) and from cultured bloodstream forms of stock 427 (Fig. 2(b)) were fractionated on denaturing gels and hybridized with DNA probes specific for α and β-tubulin. The results given in Figure 2 demonstrate that α and β-tubulin mRNA migrate as single bands of about 1950 and 2300 nucleotides in length, respectively. No size difference of the tubulin mRNAs is detected between bloodstream and procyclic trypanosomes, nor between different stocks. The different intensities in the α and β mRNA bands may indicate that α-tubulin mRNA is more abundant than β-tubulin mRNA. This quantitative difference, which is reproducibly observed, is most clearly visible in Figure 2(a), where mRNA has been hybridized with the pTBtu9A probe, which contains one complete copy each of the α and the β-tubulin genes.

(c) *Tubulin mRNAs carry a 35-nucleotide mini-exon sequence*

In order to map the 5′ termini of the tubulin mRNA populations, they were analyzed by S_1-mapping and primer extension with reverse transcriptase (for a map of the fragments used in these and all subsequent experiments, see Fig. 3).

For mapping the 5′ terminus of β-tubulin mRNA, a 1·7 kb *Pvu*II–*Pvu*II fragment containing the first 129 nucleotides of the β coding sequence, the upstream spacer and the 3′ moiety of the preceding α-gene was used (see Fig. 3). S_1-mapping with this fragment resulted in a single band of 190 bp length (Fig. 4), placing the 5′ terminus of the β mRNA at around position −60 from the first nucleotide of the initiation codon. The precise localization of the β mRNA terminus was determined by co-electrophoresis of the S_1-generated fragment with a Maxam–Gilbert sequencing ladder, which was produced from the same fragment after an additional asymmetric cleavage with *Bgl*II. The initial nucleotide was found to be either the G residue numbered +1 in Figure 5 or, if this residue is contributed by the mini-exon sequence (see bottom of Fig. 5), the C residue numbered +2. The position of these residues with respect to the initial AUG codon of the β-tubulin gene is −60 and −59, respectively. Consistent results were obtained in independent experiments (data not shown) using a number of different restriction fragments from the β gene sequence (see summary in Fig. 3). These additional experiments have served to establish that the faint band visible in Figure 4, slot H_5, with an apparent size of about 360 nucleotides is an artifact, and does not reflect an additional start-point of β-tubulin mRNA.

For primer extension analysis of the 5′ terminus of the β-tubulin mRNA, the 5′-labeled *Pvu*II–*Pvu*II fragment was further cleaved with *Bst*EII (see Fig. 3), and the resulting fragment corresponding to nucleotides 98 to 189 of the mRNA (39 to 130 of the coding sequence) was used as a primer for reverse transcriptase. Using the identical labeled 5′ terminus for both S_1-mapping and primer extension allowed a direct comparison of the length of the fragments created in the two types of experiments. Lane R_t of Figure 4 demonstrates that the cDNA synthesized off this primer migrates as a distinct band, which is 35 nucleotides larger than the fragment protected from S_1. This observation strongly suggested the presence of a 35-nucleotide sequence at the 5′ terminus of β-tubulin mRNA not contiguously encoded with the β-tubulin gene.

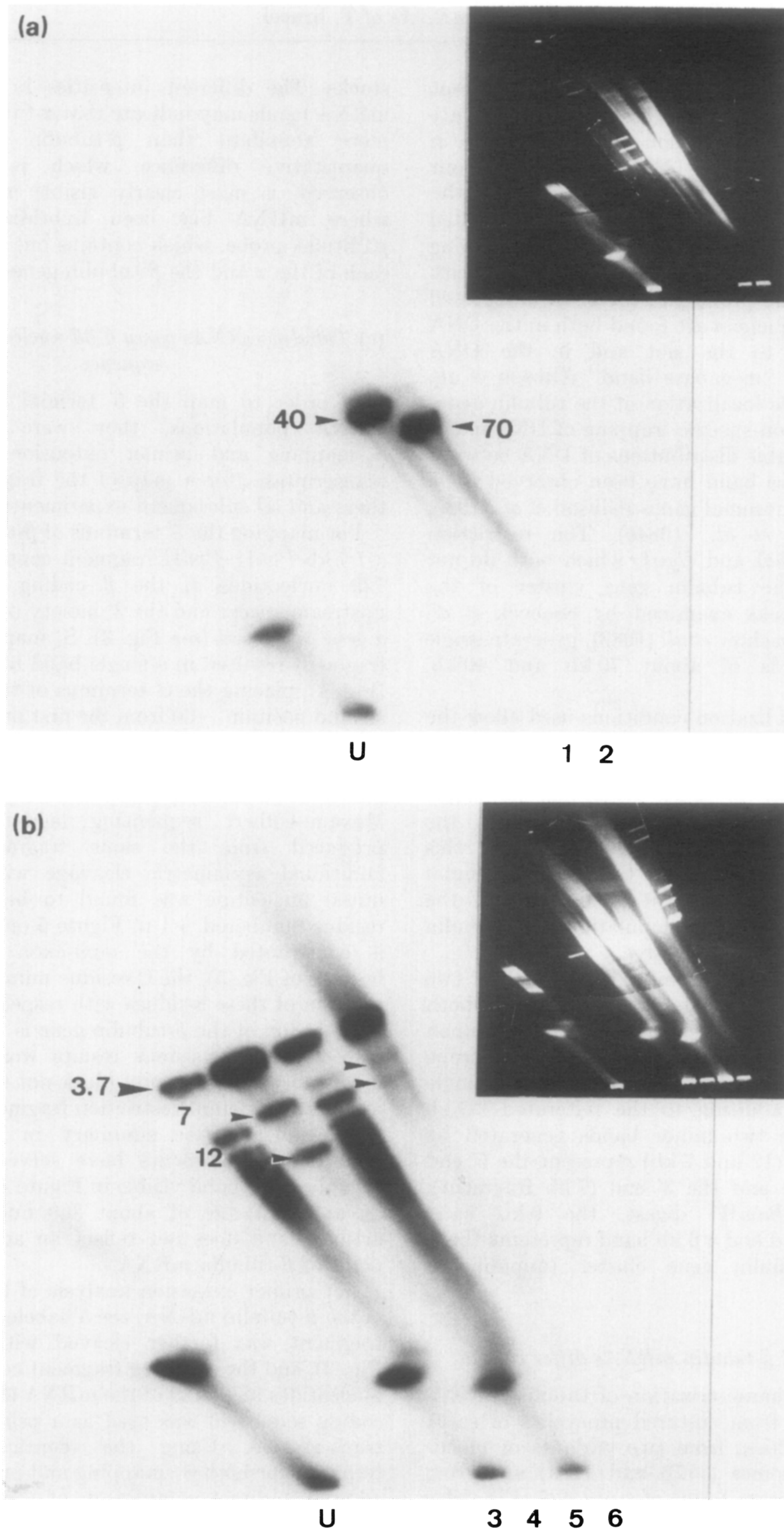

Figure 1. PFG analysis of the tubulin gene cluster. DNA of procyclic trypanosomes (stock STIB 366) was digested *in situ* with restriction enzymes (a) that do not or (b) do cleave within the tubulin gene cluster. DNA was hybridized with pTBtu9A. Insets in each autoradiogram show the ethidium-bromide-stained gel. λ DNA and λ DNA digested with *Hin*dIII were used as size markers. Running conditions were: 450 V (vertical, inhomogeneous field); 190 V (horizontal, homogeneous field); pulse time 45 s; running time 20 h. (U) undigested DNA; (1) 200 units of *Hpa*I; (2) 200 units of *Xho*I; (3) 20 units of *Eco*RI; (4) 180 units of *Eco*RI; (5) 20 units of *Bam*HI; (6) 180 units of *Bam*HI. The tubulin cluster repeat unit (large arrow) and the end-fragments of the cluster (small arrows) are indicated (fragment sizes are indicated in kb). Running direction is from bottom to top.

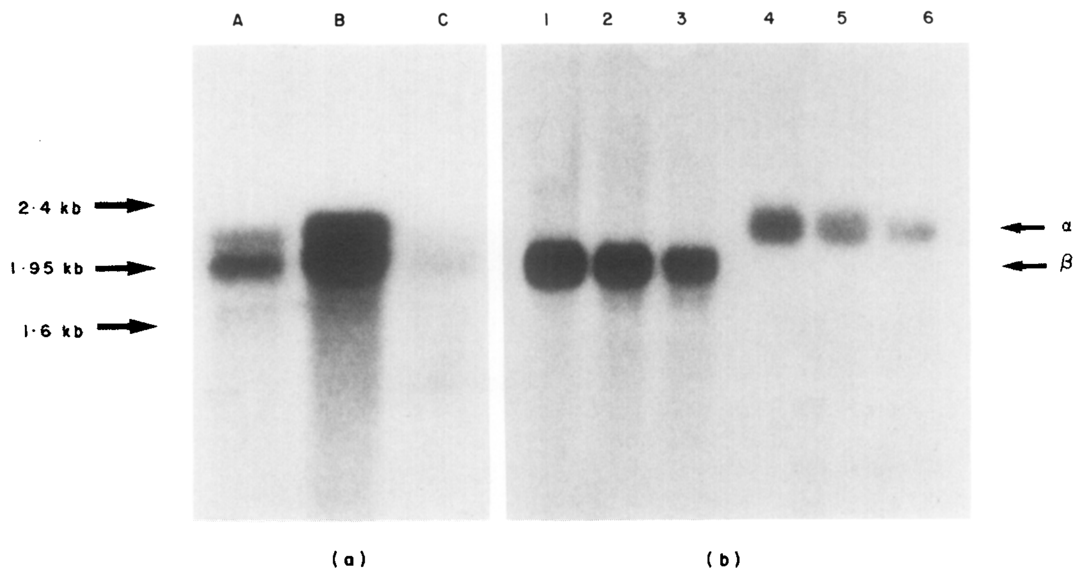

Figure 2. Northern blots of tubulin mRNA. Portions (3 to 5 μg) of RNA were fractionated on denaturing gels as detailed in Materials and Methods. (a) RNA extracted from cultured procyclics of stock STIB 366 and analyzed on a formaldehyde-containing gel. The filter was hybridized with pTBtu9A, containing one complete α and β-gene. Lane A, total RNA; lane B, poly(A)$^+$ RNA; lane C, poly(A)$^-$ RNA. (b) RNA extracted from bloodstream forms of stocks 117b (slots 1 and 4), 118a (slots 2 and 5) and 427 (slots 3 and 6). RNAs were fractionated after denaturation in glyoxal. Slots 1 to 3 and 4 to 6 were hybridized with probes specific for α and β-tubulin, respectively. The sizes indicated at the left margin are those of the trypanosomal ribosomal RNAs, which were used as internal markers (Imboden, 1983).

The precise structure of this 5′ terminus was further analyzed by dideoxy-sequencing of the β mRNA, using the same primer (Fig. 4). The sequence obtained (Fig. 5) shows the presence of a 35-nucleotide mini-exon sequence at the 5′ terminus of the β-tubulin mRNA. Fully compatible results were obtained when a synthetic 20mer (nucleotides +24 to +45 of the coding sequence) was used as a primer (data not shown).

At the very end of the cDNA, a small number of non-specific bands can be observed. Their occurrence is strongly dependent on the reaction

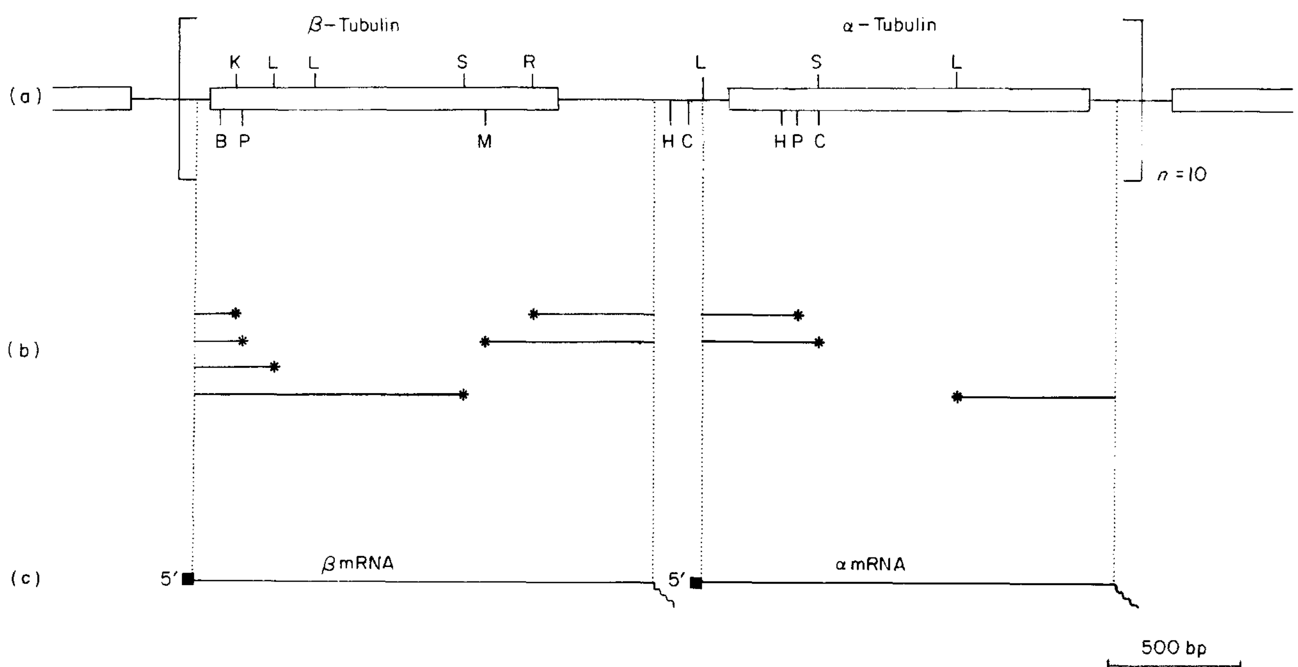

Figure 3. Structure of tubulin gene repeat and outline of S_1-mapping. (a) One complete 3·7 kb tubulin gene repeat unit is indicated. The open bars delineate the coding regions. Restriction enzyme cleavage sites are indicated by the following symbols: B, *Bst*EII; C, *Hinc*II; H, *Hha*I; K, *Kpn*I; L, *Bgl*II; M, *Bam*HI, P, *Pvu*II; R, *Eco*RI; S, *Sal*I. Only the relevant restriction sites are indicated. (b) 5′-labeled (——*) and 3′-labeled (*——) fragments protected in S_1-mapping experiments are indicated. The vertical broken lines indicate the endpoints of protection from S_1. (c) Map of tubulin mRNA. (5′ ■) mini-exon; (~~~) poly(A) tails.

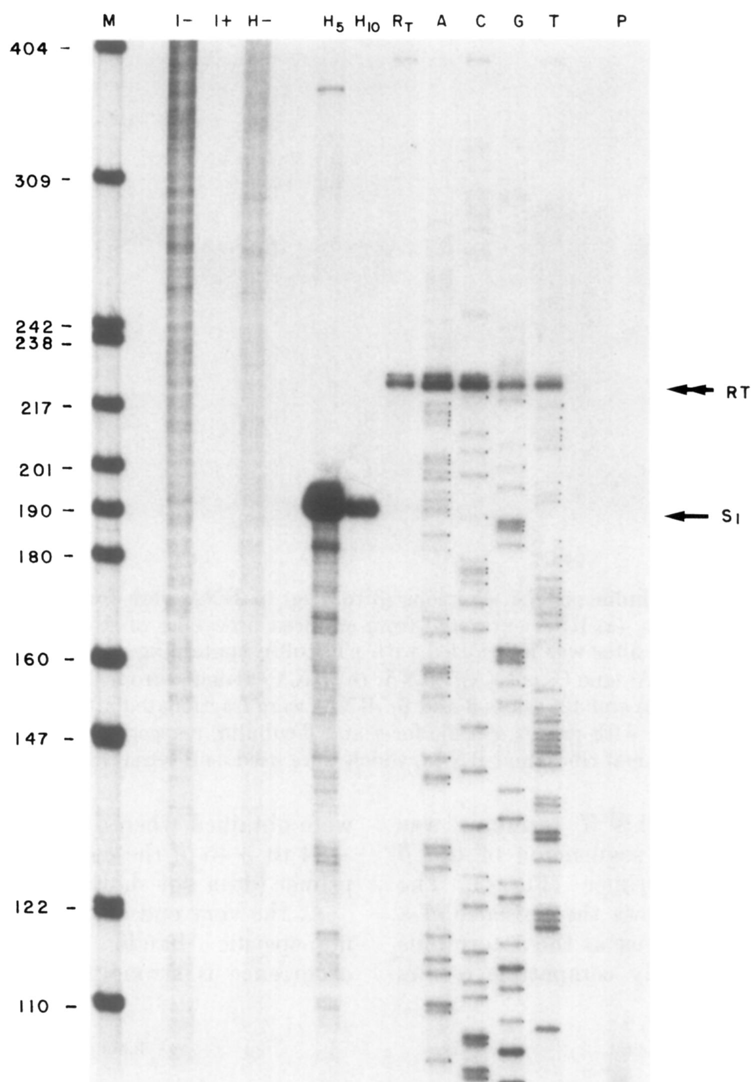

Figure 4. Characterization of the 5′ terminus of the β-tubulin mRNA. S_1-mapping was done with the 1·7 kb *Pvu*II–*Pvu*II fragment, and the 92 bp fragment derived therefrom by an additional *Bst*EII cleavage was used for primer extension. S_1-mapping experiment: (I−) input DNA fragment; (I+) input DNA fragment digested with 10,000 units of S_1 nuclease; (H−) DNA/RNA hybrid without S_1 digestion; (H_5) DNA/RNA hybrid digested with 5000 units of S_1; (H_{10}) DNA/RNA hybrid digested with 10,000 units of S_1. Primer extension: (R_T) reverse transcriptase complete reaction; (A) A-reaction; (C) C-reaction; (G) G-reaction; (T) T-reaction; (P) full reaction, but without RNA. The fragment sizes indicated on the left are those of *Hpa*II-digested pBR322. RT and S_1 indicated on the right denote the endpoint bands generated by reverse transcriptase and S_1 nuclease in their respective reactions.

conditions, and they are probably due to the idling reaction of the reverse transcriptase at the end of the mRNA. A closer inspection of the autoradiogram presented in Figure 4 reveals the presence of an additional sequence extending far beyond the heavy band that delineates the 5′ end of the mini-exon. This additional sequence most probably represents an artifact of the type previously reported in reverse transcriptase sequencing experiments (Michels *et al.*, 1983). It is not observed when the synthetic, rather than the plasmid-derived, primer is used. Furthermore, this additional sequence could not be correlated with any sequences surrounding the tubulin genes.

α-Tubulin mRNA was analyzed by a similar strategy. After defining the 5′ terminus by S_1-mapping experiments (see Fig. 3), the 5′ portion of the α-tubulin gene was sequenced using the Sanger procedure (Sanger *et al.*, 1977). A synthetic primer complementary to nucleotides 42 to 61 of the upstream spacer region was then used for reverse transcriptase sequencing of the mRNA. The 5′ terminus of α mRNA could thus be located at position −106 from the initial codon of the α-tubulin gene. As observed before with β-tubulin mRNA, the mini-exon sequence was also found at the 5′ terminus of α-tubulin mRNA (see Fig. 5).

Figure 5 shows a comparison between the genomic DNA sequences (derived from the genomic tubulin repeat clone pTBtu9A) and the sequences obtained from reverse transcriptase sequencing of α and β-tubulin mRNA. It demonstrates the absence of mini-exon sequences in the immediate vicinity of the tubulin genes. No mini-exon sequences have

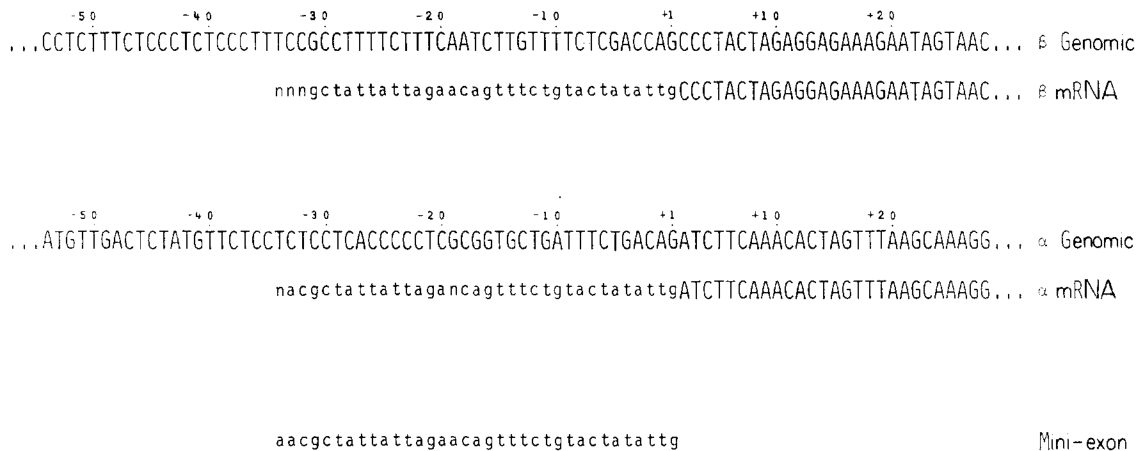

Figure 5. Compilation of tubulin gene and mRNA sequences. Genomic α and β sequences were derived from a cloned tubulin gene repeat (pTBtu9A). mRNA sequences are those obtained by primer extension sequencing and the mini-exon sequence is from Campbell *et al.* (1984). Sequence hyphens are omitted for clarity.

been found elsewhere in the tubulin gene repeat, nor within at least 12 kb upstream of the 5′ end of the tubulin gene cluster (unpublished results; U. Pauli, personal communication).

(d) *3′ Mapping of tubulin mRNAs*

Having established the location and structure of the 5′ termini of the α and β-tubulin mRNAs, their respective 3′ ends were localized by a series of S_1-mapping experiments using various 3′-labeled restriction fragments (for a summary, see Fig. 3).

The 3′ end of α-tubulin mRNA was analyzed by using the 1·1 kb long *Bgl*II–*Bgl*II fragment that contains the 3′ end of the α-tubulin gene (see Fig. 3). The fragment was 3′ end-labeled with Klenow polymerase and purified by gel electrophoresis prior to use in the S_1-mapping. The results given in Figure 6(a) place the 3′ end of α-tubulin mRNA at approximately 580 nucleotides downstream from the *Bgl*II site, i.e. about 90 nucleotides downstream from the termination codon. In analogy to what had been observed with the β mRNA, the 3′ end of the α-tubulin mRNA is also somewhat heterogeneous. In underexposed autoradiograms, two major and one minor S_1-generated bands can be discerned at 595, 580 and 560 nucleotides, respectively.

For mapping the 3′ end of the β-tubulin mRNA, the *Eco*RI fragment containing the end of the β-tubulin gene (from pTBtu9A) was 3′-labeled with [^{32}P]cordycepin and S_1-mapping was performed as detailed in Materials and Methods. Figure 6(b) demonstrates that the 3′ terminus of β-tubulin mRNA is located about 460 nucleotides downstream from the *Eco*RI site, i.e. about 350 nucleotides downstream from the termination codon. Closer inspection of the autoradiograms revealed that there are most likely two major 3′ termini of the RNA, about ten nucleotides apart, suggesting a slight heterogeneity at the 3′ end of the mRNA. These results were confirmed by similar mapping experiments in which 3′-labeled *Bam*HI fragments were used, and which located the 3′ end of the mRNA at about 650 nucleotides downstream from the *Bam*HI restriction site (data not shown), which is in excellent agreement with the result of the *Eco*RI mapping. The two sets of mapping are fully consistent, since sequencing data have demonstrated that the *Bam*HI site in the β-tubulin gene is 188 bp upstream from the *Eco*RI site (Imboden, unpublished results; Kimmel *et al.*, 1985).

In an attempt to specifically detect possible minor transcripts that might extend across the β/α spacer, further S_1-mappings were done with two fragments containing parts of this spacer region. The 410 bp *Hha*I–*Hha*I fragment and the 500 bp *Hinc*II–*Hinc*II fragment cover the last 106 bp and 50 bp, respectively, of β/α spacer region, in addition to the 5′ moiety of the following α gene (for location of the respective restriction sites, see Fig. 3). However, when these fragments were analyzed by S_1-mapping, no protection of either of them could be detected (data not shown). These observations support the conclusion that the β-tubulin mRNA indeed ends about 460 bp downstream from the *Eco*RI site.

In sum, our observations suggest a picture where both α and β-tubulin mRNAs carry at their 5′ termini a 35-nucleotide mini-exon sequence, followed by a untranslated leader sequence (105 and 60 nucleotides, respectively), the coding sequence and finally an untranslated 3′ tail (see Fig. 3(c)).

4. Discussion

The tubulin genes of *T. brucei* are arranged in tight clusters of tandemly arranged alternating α and β-genes (Seebeck *et al.*, 1983; Thomashow *et al.*, 1983). Analysis of the genome of trypanosomes with pulsed field gradient electrophoresis (Schwartz & Cantor, 1984) has confirmed this chromosomal organization (Gibson *et al.*, 1985; this study). The PFG analysis of restriction-enzyme-digested DNA (see Fig. 1) now clearly demonstrates that no

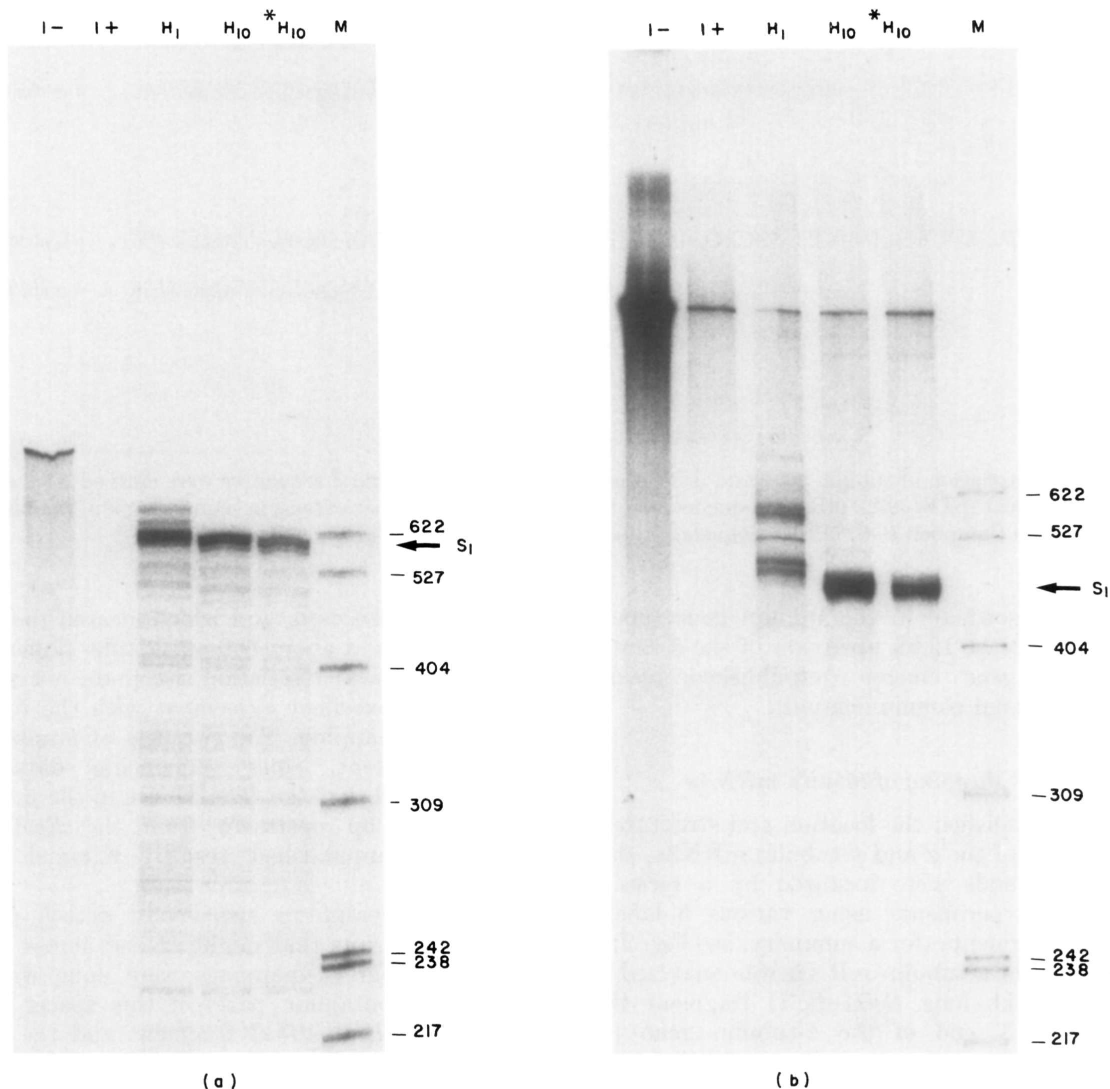

Figure 6. S_1-mapping of the 3′ termini of β and α-tubulin mRNAs. (I−) input DNA fragments; (I+) input DNA fragments, digested with 10,000 units of S_1 nuclease; (H_1) DNA/RNA hybrid digested with 1000 units of S_1; (H_{10}) hybrid digested with 10,000 units of S_1; ($^*H_{10}$) same, but with a different batch of RNA; (M) marker fragments (pBR322 *Hpa*II fragments). Horizontal arrows indicate the position of protected fragments. (a) α-mRNA; (b) β-mRNA.

tubulin genes are found outside the cluster. These observations establish that the tubulin gene cluster is indeed the locus of transcription. Digestion with *Hpa*I generates a single band, which contains the entire tubulin gene cluster, and which migrates slightly ahead of intact lambda phage DNA, i.e. exhibits a length of about 40 kb. Since *Hpa*I does not cleave within the tubulin cluster (Seebeck *et al.*, 1983), but cuts in the immediate vicinity of the 5′ and 3′ ends of the cluster (unpublished results), this 40 kb *Hpa*I fragment represents the net size of the tubulin cluster. Based on a unit repeat length of 3·7 kb, our data indicate the presence of no more than ten tubulin gene repeats per cluster. This value is in good agreement with previous estimates obtained by entirely different means (Seebeck *et al.*, 1983; Thomashow *et al.*, 1983). The genome of *T. brucei* is diploid for the tubulin gene cluster, and the allelic clusters can be distinguished, in some trypanosome stocks, by a heterologous *Xho*I restriction site (Gibson *et al.*, 1985; our own unpublished results). In undigested DNA, the tubulin gene clusters are located on large (i.e. $>700\times10^3$ M_r) chromosomal DNA.

Transcription of the tubulin genes yields a single class of mature mRNAs for each, α and β-tubulin, with respective lengths of 1950 and 2300 nucleotides. S_1-mapping experiments revealed homogeneous 5′ termini and slightly heterogeneous 3′ ends of both mRNAs. Primer extension sequencing has established the presence of the mini-exon sequence at the 5′ termini of both α and β-tubulin mRNA (see Fig. 3(c)). These observations, which are in good agreement with an independent study (Sather & Agabian, 1985), support earlier contentions that the mini-exon sequence is not exclusive to the VSG mRNAs, but that it is present on most, if not all, trypanosomal mRNAs (DeLange *et al.*, 1984; Parsons *et al.*, 1984*a*,*b*). The detection of the mini-exon sequence

on calmodulin mRNAs (Tschudi *et al.*, 1985) further supports this view. In addition, evidence for the presence of a mini-exon sequence at the 5′ terminus of tubulin mRNA from *Leishmania* has been presented (Landfear & Wirth, 1985).

Mapping of the 3′ termini of the two tubulin mRNAs demonstrated a single, slightly heterogeneous terminus of each mRNA. This is in contrast to the report by Sather & Agabian (1985), where two different termination sites are proposed for the β-tubulin mRNA. Their minor site closely agrees with the termination site defined in this study. However, their major termination site may either reflect a difference in trypanosomal stocks, or it may represent an S_1-mapping artifact. The major termination site proposed by Sather & Agabian (1985) is located within a stretch of DNA that contains 87% A+T. Such A+T-rich regions have been demonstrated to easily cause difficulties in S_1-mapping experiments (Miller & Sollner-Webb, 1981).

Our data on the 5′ terminus of the trypanosomal mRNAs give no indication on the initiation point of transcription. It is not clear if the mini-exon sequence is added directly to the original 5′ terminus of the primary transcript, or if its addition occurs *via* a joining reaction, which may eliminate the 5′ moiety of the primary transcript. A common feature of genes transcribed into mini-exon-containing mRNAs, including the tubulin mRNAs, is the invariable presence of a 5′-A-G-3′ in the DNA sequence at the junction site between mini-exon and mRNA sequence. The DNA sequences upstream from the 5′ termini of both tubulin mRNAs are remarkably pyrimidine-rich (87% between positions −1 to −50 of the β gene; 70% in the same stretch of the α-gene). Similar conspicuous pyrimidine clusters are observed in analogous regions of other trypanosomal genes (calmodulin genes *1* to *3* (Tschudi *et al.*, 1985), variable surface glycoprotein genes *221* (Bernards *et al.*, 1984) and *118* (Van der Ploeg *et al.*, 1982)). The locations of these regions suggest that they might be involved in the initiation of transcription, despite the absence of canonical TATA boxes upstream from the mRNA 5′ terminus. Very similar arrangements of pyrimidine-rich stretches have been identified upstream from the start of transcription of some mammalian genes (3-hydroxy-3-methylglutaryl coenzyme A reductase, Reynolds *et al.*, 1984; hypoxanthine phosphoribosyl transferase, Melton *et al.*, 1984), which are noted for their apparent lack of canonical transcription signals. Non-TATA transcription initiation signals have been proposed for the tubulin genes of *Chlamydomonas* (Brunke *et al.*, 1984). A pyrimidine-based set of signals may hence represent an alternative code of communication between eukaryotic genes and RNA polymerases.

However, no evidence for a defined initiation of transcription within at least a few hundred base-pairs upstream from the 5′ terminus of trypanosomal genes has been found. Hence the pyrimidine cluster may rather play a crucial role in the mini-exon/mRNA joining reaction. In the latter case, the 5′ termini of the mature mRNAs would be created by this RNA joining, rather than by the RNA polymerase initiation reaction.

A distinction between these two formal possibilities of trypanosomal tubulin mRNA generation can clearly not be made at present. More work is obviously required before we can gain a reasonably detailed insight into the mode and mechanics of the trypanosomal gene expression machinery.

We are very grateful to A. Räber of the University Electronics Workshop for his enthusiastic collaboration in the construction of the PFG equipment; to H. Hennecke and K. Kaluza for supplying us with synthetic primers; to Gary McMaster for his most valuable technical instructions; to Wolf Heyer for critical reading of the manuscript; and to U. Kurath, K. Behrens and R. Gardi for their skilful technical assistance. The continuous interest and critical comments of P. Borst are gratefully acknowledged. This work was funded by the Swiss National Science Foundation, the UNDP/World Bank/WHO Special program for Research and Training in Tropical Diseases and by a fellowship to M.A.I. from the Studentship Trust of the Swiss Chemical Industry.

References

Bernards, A., DeLange, T., Michels, P. A. M., Liu, A. Y. C. & Borst, P. (1984). *Cell,* **36**, 163–170.

Boothroyd, J. C. & Cross, G. A. M. (1982). *Gene*, **20**, 281–289.

Brun, R. & Schönenberger, M. (1979). *Acta Trop.* **36**, 289–292.

Brunke, K. J., Anthony, J. G., Sternberg, E. J. & Weeks, D. P. (1984). *Mol. Cell. Biol.* **4**, 1115–1124.

Campbell, D. A., Thornton, D. A. & Boothroyd, J. C. (1984). *Nature (London)*, **311**, 350–355.

Cross, G. A. M. (1975). *Parasitology,* **71**, 393–402.

DeLange, T., Liu, A. Y. C., Van der Ploeg, L. H. T., Borst, P., Tromp, M. C. & Van Boom, J. H. (1983). *Cell,* **34**, 891–900.

DeLange, T., Michels, P. A. M., Veeneman, H. J. G., Cornelissen, A. W. C. A. & Borst, P. (1984). *Nucl. Acids Res.* **12**, 3777–3790.

Gibson, W. C., Osinga, K. A., Michels, P. A. M. & Borst, P. (1985). *Mol. Biochem. Parasitol.* **16**, 231–242.

Hall, A., Marshall, C. J., Spurr, N. K. & Weiss, R. A. (1983). *Nature (London)*, **303**, 396–400.

Harding, J. D., Przybyla, A. E., McDonald, R. J., Pictet, R. L. & Rutter, W. J. (1978). *J. Biol. Chem.* **253**, 7531–7537.

Imboden, M. (1983). Masters thesis, University of Bern.

Kimmel, B. E., Samson, S., Wu, J., Hirschberg, R. & Yarbrough, L. R. (1985). *Gene,* **35**, 237–248.

Landfear, S. M. & Wirth, D. F. (1985). *Mol. Biochem. Parasitol.* **15**, 61–82.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning, A Laboratory Manual,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Maxam, A. & Gilbert, W. (1980). *Methods Enzymol.* **65**, 499–560.

Melton, D. W., Konecki, D. S., Brennard, J. & Caskey, C. T. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 2147–2151.

Michels, P. A. M., Liu, A. Y. C., Bernards, A., Sloof, P., Van der Bijl, M. M. W., Schinkel, A. H., Menke, H. H., Borst, P., Venneman, G. H., Tromp, M. C. & Van Boom, J. H. (1983). *J. Mol. Biol.* **166**, 537–556.
Miller, K. G. & Sollner-Webb, B. (1981). *Cell,* **27**, 165–174.
Parsons, M., Nelson, R. G., Watkins, K. P. & Agabian, N. (1984*a*). *Cell,* **38**, 309–316.
Parsons, M., Nelson, R. G., Stuart, K. & Agabian, N. (1984*b*). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 684–688.
Reynolds, G. A., Basu, S. K., Osborne, T. F., Chin, D. J., Gil, G., Brown, M. S., Golfstein, J. L. & Luskey, K. L. (1984). *Cell,* **38**, 275–285.
Ryffel, G. U. & McCarthy, B. J. (1975). *Biochemistry,* **14**, 1379–1385.
Sanger, F., Nicklen, S. & Coulson, A. R. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463–5467.
Sather, S. & Agabian, N. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 5695–5699.
Schwartz, D. C. & Cantor, C. R. (1984). *Cell,* **37**, 67–75.
Seebeck, T., Whittaker, P. A., Imboden, M. A., Hardman, N. & Braun, R. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 4634–4638.
Stieger, J., Wyler, T. & Seebeck, T. (1984). *J. Biol. Chem.* **259**, 4596–4602.
Thomas, P. S. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 5201–5205.
Thomashow, L. S., Milhausen, M., Rutter, W. J. & Agabian, N. (1983). *Cell,* **32**, 35–43.
Tschudi, C., Young, A. S., Ruben, L., Patton, C. L. & Richards, F. F. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 3998–4002.
Van der Ploeg, L. H. T., Liu, A. Y. C., Michels, P. A. M., DeLange, T., Borst, P., Majumber, H. K., Weber, H., Veeneman, G. H. & Van Boom, H. (1982). *Nucl. Acids Res.* **10**, 3591–3604.
Van der Ploeg, L. H. T., Schwartz, D. C., Cantor, C. R. & Borst, P. (1984*a*). *Cell,* **37**, 77–84.
Van der Ploeg, L. H. T., Cornelissen, A. W. C. A., Barry, J. D. & Borst, P. (1984*b*). *EMBO J.* **3**, 3109–3115.
Vickerman, K. & Preston, T. M. (1976). In *Biology of the Kinetoplastida* (Lumsden, W. H. R. & Evans, D. A., eds), pp. 35–130. Academic Press, New York.
Weaver, R. F. & Weissmann, C. (1979). *Nucl. Acids Res.* **7**, 1175–1193.

Edited by B. Mach