Molecular and Biochemical Parasitology, 34 (1989) 155–166
Elsevier

MBP 01130

Structure, genomic organization and transcription of the bifunctional dihydrofolate reductase-thymidylate synthase gene from *Crithidia fasciculata*

Dallas E. Hughes, Opeolu A. Shonekan and Larry Simpson

Department of Biology and Molecular Biology Institute, University of California, Los Angeles, CA, U.S.A.

(Received 12 September, 1988; accepted 29 November 1988)

The bifunctional dihydrofolate reductase-thymidylate synthase (DHFR-TS) gene from the monogenetic kinetoplastid protozoan, *Crithidia fasciculata*, was isolated and characterized. The gene is located on a single chromosome of approximately one megabase, and shows significant sequence similarity to other eukaryotic and prokaryotic DHFR and TS genes. There is a single low-abundance polyadenylated DHFR-TS transcript of approximately 3100 nt. One major miniexon splice site was identified by primer extension analysis. The 5′ flanking region of the gene is divergently transcribed and shows strong similarities to a consensus DHFR promoter as well as to other eukaryotic 'housekeeping' gene promoter regions. A sequence downstream of the DHFR promoter consensus region is complementary to the 3′ end of the *C. fasciculata* miniexon-derived RNA. This suggests a means by which the two separately transcribed RNAs may be juxtaposed for *trans*-splicing. In the 3′ flanking region of the DHFR-TS gene, there is a sequence that is present in all of the chromosomes from this species and also from *Leishmania tarentolae*.

Key words: *Crithidia fasciculata;* dihydrofolate reductase-thymidylate synthase; *trans*-splicing; mini-exon

Introduction

Kinetoplastid protozoa exhibit several interesting phenomena in terms of gene expression including the transposition of nonexpressed variant surface protein genes to telomeric expression sites in the African trypanosomes [1,2], the *trans*-splicing of a separately-encoded 39-nt RNA (miniexon) onto the 5′ end of all kinetoplastid mRNAs [3–7], and the 'editing' of mitochondrial transcripts by insertion and deletion of uridines at specific sites [8–10]. Many nuclear genes are tandemly repeated and there is some evidence for the occurrence of large multicistronic transcriptional units [11,12]. However, no RNA Pol II initiation site, transcriptional control signal, or any primary RNA Pol II transcript has yet been identified. There is no in vitro system available for functional studies on gene expression.

In order to define the nature of a transcriptional control signal in a kinetoplastid, the dihydrofolate reductase-thymidylate synthase (DHFR-TS) gene from the monogenetic parasite, *Crithidia fasciculata,* was isolated and characterized. *C. fasciculata* is a model kinetoplastid protozoan which grows readily in defined media [13], forms colonies on agar, gives rise to drug-resistant mutants at a high frequency after mutagenesis [14], and has a standard *trans*-splicing apparatus for nuclear gene expression [15]. We have found that exogenous DNA can be introduced into *C. fasciculata* by the technique of electroporation and

Correspondence address: Dr. Larry Simpson, Department of Biology and Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90024, U.S.A.

Abbreviations: DHFR-TS, dihydrofolate reductase-thymidylate synthase; Pol II, RNA polymerase II; guanidium-HCl, guanidium hydrochloride; OFAGE, orthogonal-field-alternation gel electrophoresis; HSP85, heat shock protein 85; medRNA, miniexon-derived RNA.

Note: Nucleotide sequence data reported in this paper have been submitted to the GenBank™ Data Bank with the accession number J03973.

that this DNA is stable for 10–20 h (unpublished results). In addition to providing insight into the molecular mechanisms controlling gene expression in the kinetoplastids, the availability of a Pol II promoter segment would facilitate the construction of selectable markers for DNA transformation vectors.

In general, the DHFR gene belongs to the class of 'housekeeping' or 'growth control' genes, which characteristically show high evolutionary conservation, are of low copy number, often have more than one G+C-rich promoter region, are transcribed at a low level throughout most of the cell cycle, and have transcripts that display multiple 5′ termini [16–25]. The DHFR gene is often amplified in antifolate-resistant mammalian cells [26,27].

In protists, the DHFR gene is contiguous with the thymidylate synthase (TS) gene [16,28], expression of which results in a bifunctional protein that exhibits metabolic channeling [29]. In a methotrexate-resistant mutant of a related kinetoplastid, *Leishmania major*, the DHFR-TS gene is located on an amplified extrachromosomal circular element [30]. The 5′ sequence of the *L. major* gene has been examined, and no significant similarities to other DHFR consensus promoter elements were observed [31]. In this paper we report on the structure, genomic organization and transcription of the *C. fasciculata* DHFR-TS gene.

Materials and Methods

Cell culture. A clonal *Crithidia fasciculata* strain (Cf-C1) was grown in Brain-Heart Infusion medium (Difco) supplemented with 10 μg/ml hemin (Sigma) at 25°C with gentle rotation.

Nucleic acid isolation. C. fasciculata nuclear DNA was prepared by the Sarkosyl lysis method as described [32] and purified by CsCl centrifugation. Total cell RNA was isolated by the guanidium-HCl method [33]. Poly(A)$^+$ RNA was selected from total RNA by oligo-dT cellulose (Collaborative Research, Inc.) chromatography.

Nucleic acid manipulations. Agarose gel electrophoresis, staining, blotting, and hybridization were performed by standard procedures [34]. DNA sequencing was performed using the dideoxy chain termination method [35] on *Sst*I and *Sau*3A fragments of the 2396-bp DHFR-TS gene region subcloned into pUC18. Single-stranded probes were generated by primer extension on plasmid pGEM7Z containing Region 2 using the M13 forward and reverse primers and the large fragment of *Escherichia coli* DNA polymerase I.

Pulse field gel electrophoresis. C. fasciculata and *Leishmania tarentolae* cells were prepared for electrophoresis by washing in 0.15 M NaCl, 0.1 M EDTA, 0.01 M Tris-HCl (pH 7.9) and embedding in 0.7% low-melting agarose at 0.5×10^9 cells ml^{-1}. Small blocks were cut and treated with 1% Sarkosyl, 1 mg ml^{-1} proteinase K at 50°C for 40 h and chromosomes were separated on a 1.5% agarose gel by orthogonal-field-alternation gel electrophoresis (OFAGE) [36]. The gel was run at 250 V using an 80-s pulse time.

Primer extension analysis. The 5′ terminal end of the DHFR-TS transcript was determined by the primer extension technique [37], Briefly, 20 μg of poly(A)$^+$ RNA was hybridized with 5 ng of a 22-nt synthetic ^{32}P-5′-end-labeled primer corresponding to positions −37 to −58 of the DHFR-TS gene sequence (Fig. 2). The primer sequence was: 5′-TTATTGTTGCTGTTGATGGGGA-3′. Extensions were carried out with AMV reverse transcriptase at 42°C for 60 min.

Computer analysis. Analysis of the DNA sequences was performed using programs from the University of Wisconsin Genetics Computer Group software package. Alignments of the DNA and protein sequences were performed with the programs BESTFIT and GAP. Z-values were obtained for fifty random shuffles of two protein sequences using the global alignment program, SEQDP, from the Los Alamos software package [38]. The Z-value represents the number of standard deviations the alignment 'distance' of two test sequences deviates from the mean value for alignments of the randomly shuffled sequences. Z-values of greater than five standard deviation units are considered significant and suggest that the observed similarities are not due to chance. The relative size of the Z-value is a function of the relatedness of two sequences.

Results

Isolation and nucleotide sequence analysis of the C. fasciculata DHFR-TS gene. C. fasciculata nuclear DNA was digested with several different restriction enzymes and probed under low stringency hybridization conditions with a DHFR gene fragment from *L. major* [39]. A 2.4-kb *Hin*dIII fragment that hybridized to this probe was subcloned into pUC9. A partial restriction map of the gene and its flanking sequences is shown in Fig. 1.

The nucleotide sequence of the entire 2.4-kb *Hin*dIII gene fragment was determined on both strands using the dideoxy chain termination procedure. Fig. 2 shows the complete sequence along with the deduced amino acid sequence of the

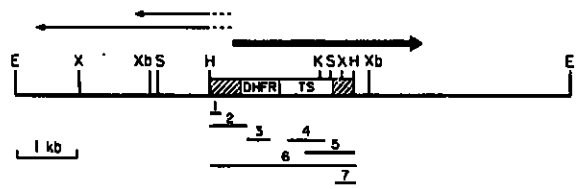

Fig. 1. Map of the *C. fasciculata* dihydrofolate reductase (DHFR)-thymidylate synthase (TS) gene region. The open boxes correspond to the open reading frame encoding the DHFR-TS. The hatched boxes indicate the 5′ and 3′ flanking regions that have been sequenced. The arrows indicate the transcript sizes and directions as determined by Northern analysis (arrow width indicates relative abundance). The lines under the map represent fragments used as probes. Enzyme abbreviations: E, *Eco*RI; X, *Xho*I; Xb, *Xba*I; S, *Sal*I; H, *Hin*dIII; K, *Kpn*I.

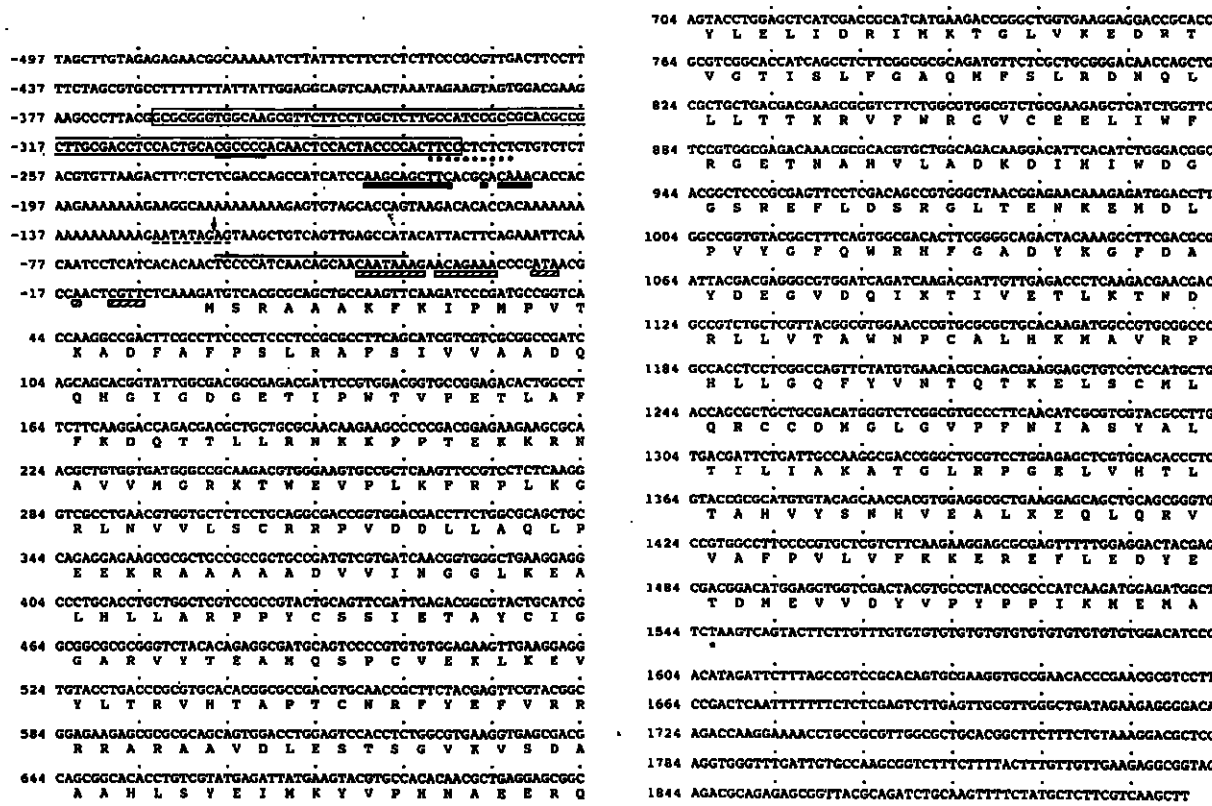

Fig. 2. Nucleotide sequence of the *C. fasciculata* DHFR-TS gene region. The boxed region contains sequences similar to other DHFR promoter regions (see Table II). The region complementary to the 3′ end of the *C. fasciculata* medRNA is underscored with a heavy line. A sequence element found upstream of the *C. fasciculata* miniexon gene is underscored with a dashed line. A sequence element found upstream of several kinetoplastid protein-coding genes is underscored with a dotted line. The 'degenerate' miniexon sequence is underscored with hatched boxes. The putative miniexon splice site is indicated by an arrow. A potential SP1 binding site is underscored with a thin line. The sequence used as a primer for the primer extension analysis is overlined.

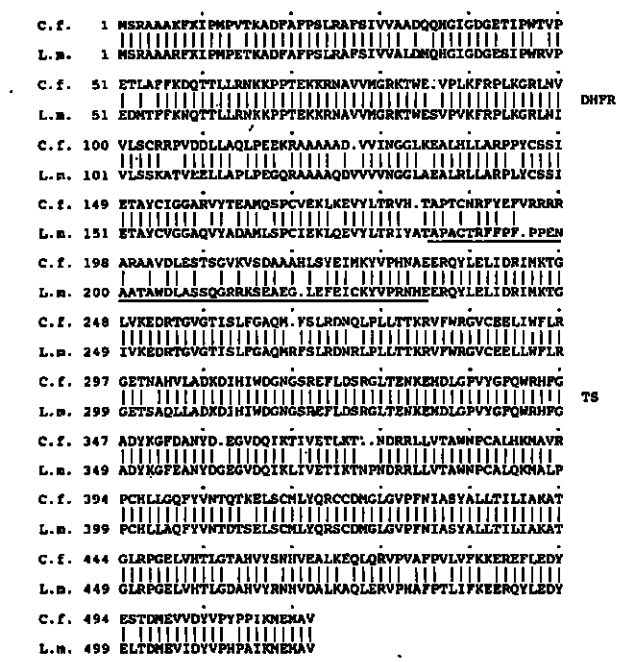

Fig. 3. Amino acid sequence alignment for the DHFR-TS of *Crithidia fasciculata* and *Leishmania major* using the GAP program of the University of Wisconsin Genetics Computer Group (UWGCG). The 'junction' region between the DHFR and TS domains is underlined.

1548-bp open reading frame corresponding to the DHFR-TS gene. The nucleotide sequence predicts an uninterrupted 516-amino-acid protein that, as in other protists, consists of a DHFR domain at the NH_2-end and a TS domain at the COOH-end.

Comparison of the *C. fasciculata* and the *L. major* amino acid sequences shows a 90% overall identity although the *C. fasciculata* protein is five residues shorter (Fig. 3). The region between the two domains (the junction region) is of similar length in both species and has several amino acid identities, although it is the least conserved portion of the DHFR-TS protein. As in the case of the *L. major* DHFR-TS gene, the junction region has no significant similarity to other DHFR or TS genes. Z-value comparisons of the TS domain of the *C. fasciculata* gene with the TS protein of other organisms reveal strong similarities (Table I). As expected, greater similarity was seen with the TS protein from other eukaryotes and the eukaryotic virus *Herpesvirus saimiri* than with the corresponding prokaryotic or bacteriophage proteins. Z-value comparisons of the *C. fasciculata* DHFR domain with other DHFR proteins also revealed significant similarities, although less conservation than in the case of the TS protein.

Genomic organization of the C. fasciculata DHFR-TS gene. Restriction digests of *C. fasciculata* nuclear DNA were probed with a subfragment from the TS domain (Probe 4, Fig. 1). In most cases there was weak hybridization to only one or two bands suggesting that the gene is of low copy (not shown). The probe hybridized with an equal intensity to two *Eco*RI + *Bam*HI fragments of approximately 9.3 and 11 kb. A probe located upstream of the coding region of the cloned gene (Probe 1, Fig. 1) hybridized only to the 9.3-kb band, suggesting that there are two separate genes. Further mapping indicated that another DHFR-TS gene was not located within 3 kb upstream or 4 kb downstream of the cloned gene (Fig. 1).

C. fasciculata chromosomes were separated by OFAGE (Fig. 4A) and probed with a subfragment from the DHFR domain (Probe 3, Fig. 1). This probe hybridized to a single band of approximately 1 megabase (Fig. 4C). This suggests that the DHFR-TS genes are located on the same

TABLE I

Z-values of the amino-acid alignments for the dihydrofolate reductase and thymidylate synthase of *Crithidia fasciculata* vs. other organisms

Organism	Z-value	
	DHFR	TS
Leishmania major	82.1	98.3
Bovine	18.6	–
Human	15.3	89.1
Chicken	14.5	–
Herpesvirus saimiri	–	80.4
Mouse	13.2	70.3
Pig	11.8	–
Escherichia coli	8.0	52.4
Neisseria gonorrhoeae	5.0	–
Streptococcus faecium	4.0	–
Lactobacillus casei	1.2	42.5
Bacteriophage T4	−0.9	36.9
Bacteriophage ϕ-3T	–	17.2

The Z-values are standard deviation units for fifty shuffles using the SEQDP program of Kanehisa [38]. Sequences which align with a Z-value of 5.0 or higher are considered to be significantly similar.

TABLE II

Comparison of the 5′ flanking region of the *C. fasciculata* DHFR-TS gene and the 5′ flanking sequence of other DHFR genes[a]

Species	Ref.	5′ Position[b]	Sequence[c,d]	3′ Position
C.f.	-	-366	GCGCGGGTGGCAAGCGTTCTTCGTCGCTTGCCATCCGGCGCACGCCGCTTGCGACGCTCCACTGCACGCCCGACAACTCCACTACC.CCACTTCC	-272
Mouse	(49)	-324	GCGCGAGGCGCGGCCGCACTTTCTCGCGCC.....CCGGCGCACGCCTC	-281
Mouse	(49)	-130	GCGCGGCGGGC	-120
Hamster	(48)	-633	CCGGGTCCGCCCGTTCTCAGTACTCCCCACCA	-603
Mouse	(49)	-199	CCTTAGCTGCACAA	-186
Human	(48)	-110	CCTCGCCTGCACAA	-97
Hamster	(48)	-593	CTACTGTACCTCCACTTCC	-575

[a]Sequences were aligned with the BESTFIT program of UWGCG.
[b]Relative to the translation initiation codon.
[c]Vertical lines indicate nucleotide identities with the *C. fasciculata* sequence.
[d]The boxed sequences are within the consensus DHFR promoter element.

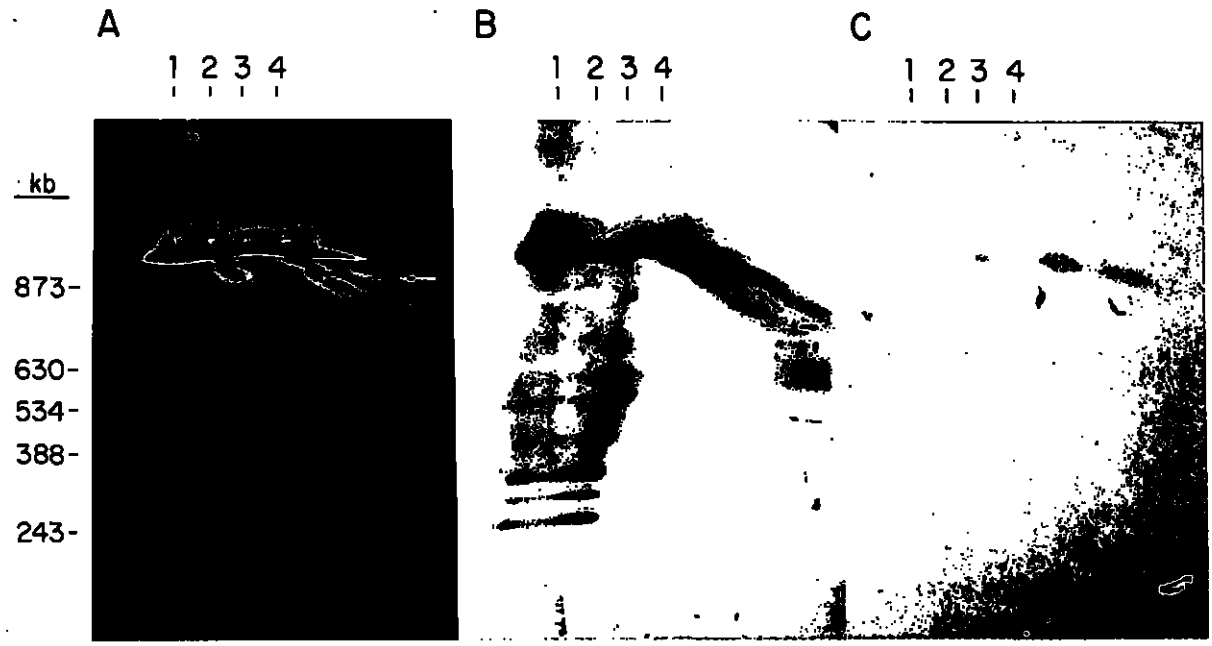

Fig. 4. Chromosomal localization of the DHFR-TS gene. (A): *Crithidia fasciculata* and *Leishmania tarentolae* chromosomes were separated by OFAGE and probed with: (B) Probe 6; or (C) Probe 3. Probes 2 and 4 hybridized to the same chromosome as Probe 3 (not shown), and Probe 7 hybridized to the same chromosomes as Probe 6 (not shown). Lane 1, *L. tarentolae*; lanes 2–4, *C. fasciculata*. The arrow in panel (A) indicates the chromosome that hybridized to Probe 3. Molecular weight standards were λDNA concatamers.

chromosome or are on two equivalently-sized chromosomes. Interestingly, when the same chromosome blot was reprobed with the entire 2.4-kb *Hin*dIII fragment (Probe 6, Fig. 1), strong hybridization was observed to apparently all the chromosomes of *C. fasciculata* and *Leishmania tarentolae*, a distantly related saurian kinetoplastid (Fig. 4B) This indicated that there is a sequence(s) on the fragment that is repetitive, dispersed and conserved with respect to both the *C. fasciculata* and *L. tarentolae* genomes. Subsequent analysis has localized this element to the 3′ untranslated region (Region 7, Fig. 1) where there is a 30-bp poly-GT stretch that is repetitive and highly conserved in many eukaryotic genomes [40].

Sequence analysis of the upstream region. Sequence analysis of the 5′ flanking region of the *C. fasciculata* DHFR-TS gene revealed several interesting features. There are no apparent TATA- or CCAAT-like elements that are known to function as transcriptional control signals for many eukaryotic genes. However, the region between −366 and −272 is strikingly similar to the DHFR promoter elements of mouse, human, and hamster (Table II). This includes a 50-bp region from −366 to −317 that is 72% identical to a 46-bp sequence from the repeated mouse DHFR promoter. Also, there are two short sequence elements with >75% identity with the 'GC' and 'CA' boxes of the consensus DHFR promoter [41] at positions −366 to −356 and −310 to −299, respectively. There is a potential SP1 binding site(CGCCCC) at positions −299 to −294 (Fig. 2).

The 5′ flanking region was compared to sequences upstream of other kinetoplastid protein-coding genes in order to search for sequence elements conserved within these organisms. A conserved sequence element, PyrPyrCCCTCTC (Fig. 2), was found between −45 and −436 bp with respect to the initiation codon of several kinetoplastid genes including the DHFR-TS gene of *L. major* [31], the HSP85 gene of *Trypanosoma cruzi* [42], the ATPase gene of *Leishmania donovani*

[43], and the alpha-tubulin genes of *Leishmania enrietti* [44] and *Trypanosoma brucei* [45]. This 'CT' motif has no similarity to other known regulatory signals and its significance is unknown. A similar motif CTCYTC was reported to be present 10–20 bp downstream of the miniexon addition site in several kinetoplastid genes [31].

The 5′ flanking sequence was also compared to the *C. fasciculata* miniexon gene sequence [15]. Beginning at position −222, there is a 10-bp sequence that is exactly complementary to the 3′ terminal region of the *C. fasciculata* medRNA (Fig. 2). This similarity extends 10 more bases to position −204 to give an overall 15/20 nucleotide sequence complementarity with the medRNA. A six bp element within this region is also complementary to a region just 3′ of the spliced leader in the *C. fasciculata* medRNA (not shown). In addition, a 9-mer AATATAGAG is found at positions −126 to −118 of the DHFR-TS gene that is also present 72 bp upstream of the miniexon + 1 site (Fig. 2). Finally, a 'degenerate', reverse complement miniexon sequence, with a 64% (23/36) nucleotide sequence identity to the *C. fasciculata* miniexon, is found at positions −43 to −8.

Transcription of the DHFR-TS gene region. *C. fasciculata* total cell, poly(A)$^+$, and poly(A)$^-$ RNA were probed with the entire 2.4-kb fragment (Probe 6, Fig. 1). One major poly(A)$^+$ transcript of approximately 3100 nt and at least three minor poly(A)$^+$ transcripts of approximately 3800, 2700, and 2200 nt were visualized (Fig. 5A, lane 1). Probe 3, specific for the DHFR coding region (Fig. 5A, lane 2), and Probe 4 (not shown), specific for the TS coding region, hybridized only to the 3100-nt transcript. Probe 5, which encompasses the C-terminal region of the TS domain and 356 bp of the 3′ untranslated region, hybridized to all four transcripts as well as several other minor poly(A)$^+$ species (Fig. 5A, lane 3).

The transcription of the 5′ flanking region, rep-

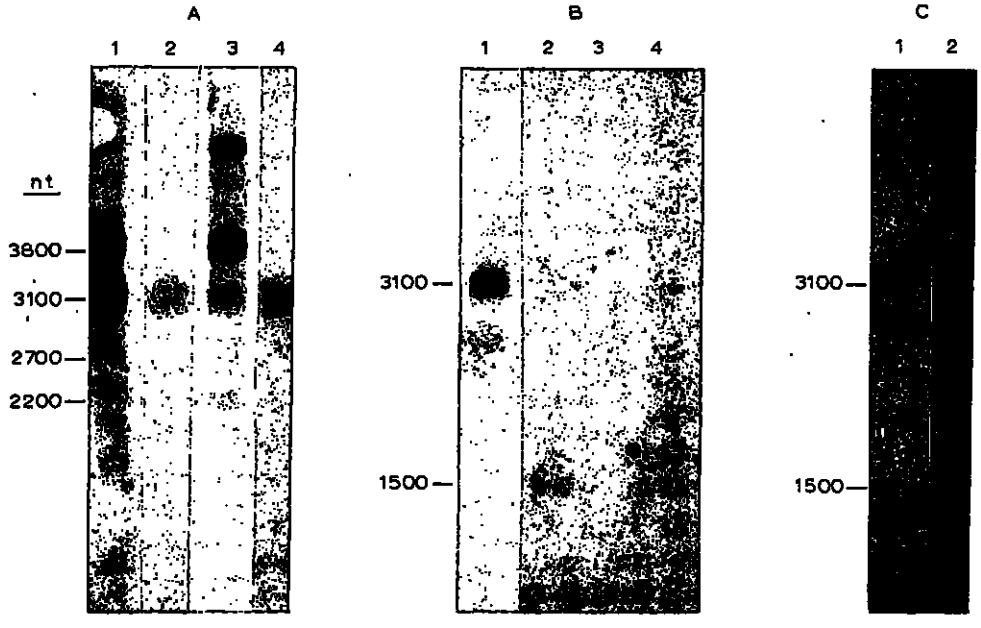

Fig. 5. Northern analysis of *C. fasciculata* RNA. Poly(A)$^+$ (panel (A), lanes 1–4; panel (B), lanes 1, 2, and 4; panel (C), lanes 1 and 2), or poly(A)$^-$ (panel (B), lane 3) RNA from *C. fasciculata* was electrophoresed on a 1.5% agarose formaldehyde gel, blotted onto nitrocellulose, and probed with: panel (A): lane 1, Probe 6; lane 2, Probe 3; lane 3, Probe 5; lane 4, Probe 2. Panel (B): lane 1, single-stranded probe from Region 2 that was generated in the opposite sense to the DHFR-TS transcript; lanes 2–4, single-stranded probe from Region 2 that was generated in the same sense as the DHFR-TS transcript. Panel (C): lane 1, same as panel (B) lane 1; panel (C) lane 2, same as panel (B) lanes 2–4.

resented by Probe 2, was complex and varied from one RNA preparation to another. In six out of eight independent RNA preparations, Probe 2 hybridized only to the 3100-nt transcript (Fig. 5A, lane 4). In two independent RNA preparations, however, Probe 2 also hybridized to a low-abundance 1500-nt poly(A)$^+$ transcript. A single-stranded probe generated from Region 2 in the same sense as the DHFR-TS transcript hybridized faintly to the 1500-nt poly(A)$^+$ transcript in addition to a 3100-nt transcript that was less abundant than the DHFR-TS transcript (Fig. 5B, lanes 2 and 4; Fig. 5C, lane 2), whereas a single-stranded probe from Region 2 that was complementary to the DHFR-TS transcript hybridized only to a 3100-nt transcript (Fig. 5B and C, lane 1). This observation suggested that the 5′ flanking region of the DHFR-TS gene is divergently transcribed, as is the DHFR gene from other eukaryotes. The abundance of the DHFR-TS transcript did not vary between RNA preparations as did the divergent transcript. None of the probes from the DHFR-TS gene region hybridized to poly(A)$^-$ RNA.

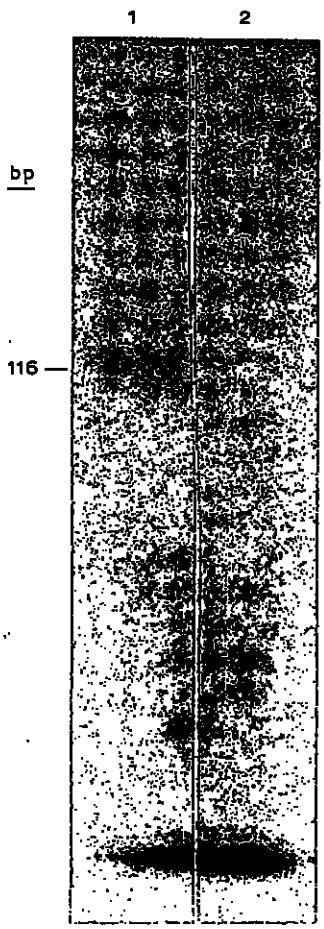

Fig. 6. Primer extension analysis of the 5′ end of the DHFR-TS transcript. Lane 1 is the major extension product. Lane 2 is the identical reaction without RNA. The molecular weight standard was a DNA sequence of Region 2 cloned using the same primer as for the RNA extension.

Primer extension analysis. All kinetoplastid mRNAs examined so far contain a 39-nt miniexon sequence at their 5′ end. Since *C. fasciculata* has been shown to have a highly abundant medRNA, we presumed that the DHFR-TS transcript contains a miniexon on its 5′ end. In order to map the miniexon splice site, we determined the 5′ end of the mRNA by primer extension analysis (Fig. 6) and inferred the miniexon splice site by subtracting 35 nt from the total length of the product (primer extension analysis will not identify the four modified nucleotides at the 5′ end). Using this approach, the splice site was mapped to an AG dinucleotide located within the 9-mer, AATATAGAG, that is also found upstream of the *C. fasciculata* miniexon gene.

Discussion

We have isolated and characterized the DHFR-TS gene from *C. fasciculata*. The 5′ flanking region of the gene is divergently transcribed and contains a region that is highly similar to the DHFR promoter elements from the mouse, hamster, and human DHFR genes. In addition, the *C. fasciculata* DHFR-TS mRNA has a 10-nt sequence element in the 5′ untranslated region that is complementary to the 3′ end of the *C. fasciculata* medRNA.

No kinetoplastid RNA Pol II promoter has yet been identified. Due to the absence of a functional assay for promoter activity, the identification of kinetoplastid promoters is currently dependent on the recognition of conserved sequence elements at sites suggested to be involved in transcriptional regulation. Such analysis is complicated by the fact that *trans*-splicing of medRNA

to the 5′ regions of primary transcripts appears to occur rapidly [46] and that primary transcripts may be multicistronic The *C. fasciculata* DHFR-TS 5′ flanking region has a G+C-rich region that displays striking similarity to other eukaryotic 'housekeeping' or 'growth control' gene promoter regions, including one that has 72% identity with a repetitive sequence element that functions as the mouse DHFR promoter [41]. In addition, short (15–30 bp) sequences are found within these regions which show similarity to promoter regions of the human adenosine deaminase [25] and c-Ha-*ras* genes (18) and to early-expressed SV40 genes [47] (not shown). It is thus possible that transcription initiation occurs between positions −366 and −272 (Fig. 2), possibly in a bidirectional manner.

Bidirectional promoters have been reported in other genes [48,49] including the mouse [50] and hamster [51] DHFR genes. This phenomenon probably reflects the orientation independence of the G+C-rich promoter elements [47,21]. In the case of the mouse DHFR gene, the divergent transcripts display heterogeneous start sites that are partially overlapping. It is unclear at this point whether the transcripts upstream of the *C. fasciculata* DHFR-TS gene overlap.

The 5′ flanking region of the amplified DHFR-TS gene of the related kinetoplastid *L. major* does not contain a DHFR consensus promoter element [31]. It is possible that the *L. major* gene utilizes a different promoter sequence than the *C. fasciculata* gene, or that a DHFR consensus promoter element lies further upstream of the *L. major* gene. Alternatively, the consensus element may not represent a true promoter for the *C. fasciculata* DHFR-TS gene. The latter possibility can be addressed only after the development of an in vitro transcription system.

The low abundance of the *C. fasciculata* DHFR-TS transcript is characteristic of 'housekeeping' genes, as they are often under negative transcriptional control or are rapidly degraded [17,21,23]. The size heterogeneity of the transcripts originating from the opposite strand of the DHFR-TS mRNA may reflect either processing intermediates or differences in their termini. The ability to detect divergent transcripts in only two out of eight RNA preparations implies that they are regulated differently from the DHFR-TS transcript. There appeared to be no correlation between the presence of the divergent transcripts and the growth stage of the cells, as all of the RNA preparations were performed at the late logarithmic stage. It is possible that minor variations in the preparation of the cells for lysis may be responsible for this effect.

The presence of a short sequence near the 5′ end of the DHFR-TS mRNA that is complementary to the *C. fasciculata* medRNA has interesting implications. The *trans*-splicing model proposed for the production of kinetoplastid nuclear mRNAs requires a mechanism by which the separately-encoded precursor RNAs are brought together for splicing. It is possible that the precursor RNAs are assembled into a splicing complex that is responsible for this function or brought together by co-complementarity with a snRNA as in conventional mRNA splicing. It has been shown, however, in the case of adenovirus, that introns containing short complementary sequences will undergo efficient *trans*-splicing in vitro via base pairing of these sequences [53,54]. Thus, the complementarity between the medRNA and 5′ region of the DHFR-TS mRNA suggests a model by which the precursor RNAs may be associated for splicing (Fig. 7). This model predicts a degradation product that is consistent with recent evidence for branched intron side-products in *T. brucei* [5,7,57]. Base complementarity between the medRNA and the 5′ flanking region of other kinetoplastid genes has not been observed as yet. Possibly, in other cases, assembly of the two separate transcripts occurs by a mechanism other than RNA:RNA duplex formation. Alternatively, the sequence complementarity may exist further upstream than has been examined.

There is one major miniexon splice site for the DHFR-TS transcript as determined by primer extension analysis. In contrast, S1 nuclease experiments identified two major and at least three minor splice sites for the *L. major* DHFR-TS transcript that is encoded from an extrachromosomal element. Interestingly, the *C. fasciculata* splice site is within a 9-nt sequence that is also found upstream of the *C. fasciculata* miniexon gene. In both cases, this sequence immediately follows a stretch of A residues. The significance

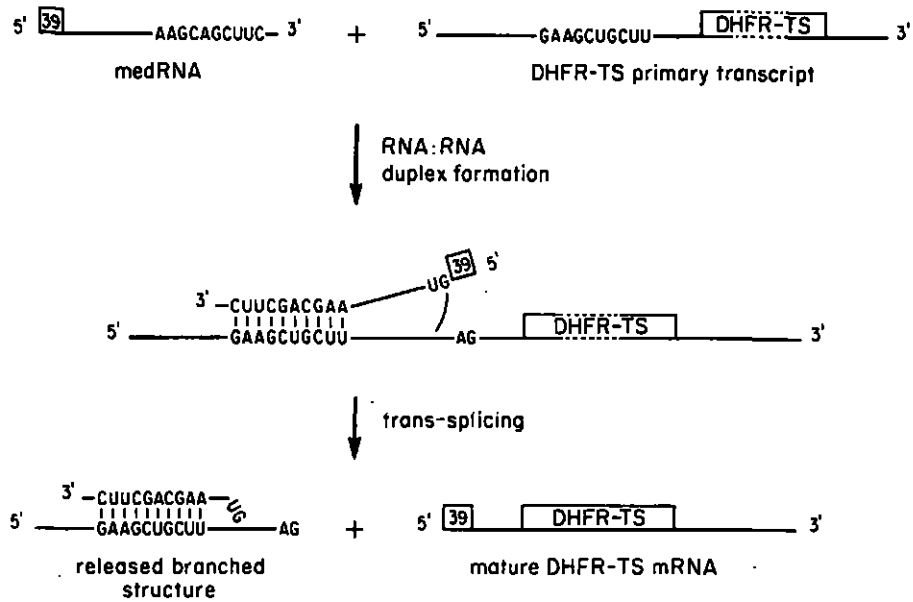

Fig. 7. Model for the initial assembly of the medRNA and the 3' DHFR-TS transcript via base complementarity. The 3' end of the separately encoded medRNA forms an RNA-RNA duplex with the 5' region of the primary transcript encoding the DHFR-TS. *Trans*-splicing then occurs to generate the mature transcript as well as a degradation product.

of this is unknown at this time.

The degenerate miniexon sequence found immediately upstream of the initiation codon shows complementarity to the 39-mer of the medRNA. This raises the intriguing possibility that this sequence element is involved in regulation of the DHFR-TS mRNA via anti-sense RNA [55,56]. In addition, it is probable that this sequence lies upstream of the gene encoding the divergent transcripts. This is consistent with the observation that several kinetoplastid genes contain a degenerate miniexon in the upstream region.

We have shown that the 5' flanking region of the *C. fasciculata* DHFR-TS gene is divergently transcribed and contains sequence elements that are highly similar to DHFR promoter elements from other cells. This suggests that the utilization of standard eukaryotic transcription initiation signals occurs in the kinetoplastids for the 3' exon of at least some mRNA-coding genes. In addition, the presence of a short sequence in the 5' transcribed region of the gene complementary to the *C. fasciculata* medRNA suggests a means by which two independent transcripts may become associated during *trans*-splicing in the kinetoplastids. This system should be useful for developing in vitro *trans*-splicing assays, as well as for studying the molecular mechanisms of transcriptional control in the kinetoplastids.

Acknowledgements

We thank Dr. Daniel V. Santi for the *L. major* DHFR-TS gene. Also, we thank Janet Shaw and Dr. David Campbell for helpful advice and discussions. This work was supported in part by NIH grant AI09102. O.A.S. is a recipient of a Research Training Grant from the Special Programme for Research and Training in Tropical Diseases (TDR) of the World Health Organization.

References

1 Borst, P. (1986) Discontinuous transcription and antigenic variation in trypanosomes. Annu. Rev. Biochem. 55, 701–732.

2 Culley, D., Ip, H. and Cross, G. (1985) Coordinate transcription of variant surface glycoprotein genes and an

expression site associated gene family in *Trypanosoma brucei*. Cell 42, 173–182.
3 Boothroyd, J. and Cross, G. (1982) Transcripts coding for variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5′ end. Gene 20, 281–289.
4 Campbell, D., Thornton, D. and Boothroyd, J. (1984) Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. Nature 311, 350–355.
5 Murphy, W., Watkins, K. and Agabian, N. (1986) Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans*-splicing. Cell 47, 517–525.
6 Perry, K., Watkins, K. and Agabian, N. (1987) Trypanosome mRNAs have unusual 'cap 4' structures acquired by addition of a spliced leader. Proc. Natl. Acad. Sci. USA 84, 8190–8194.
7 Sutton, R. and Boothroyd, J. (1986) Evidence for *trans*-splicing in trypanosomes. Cell 47, 527–535.
8 Benne, R., Van Den Burg, J., Brakenhoff, J., Sloof, P., Van Boom, J. and Tromp, M. (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell 46, 819–826.
9 Feagin, J.E., Abraham, J.M. and Stuart, K. (1988) Extensive editing of the cytochrome *c* oxidase III transcript in *Trypanosoma brucei*. Cell 53, 413–422.
10 Shaw, J.M., Feagin, J.E., Stuart, K. and Simpson, L. (1988) Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. Cell 53, 401–411.
11 Johnson, P., Kooter, J. and Borst, P. (1987) Inactivation of transcription by UV irradiation of *T. brucei* provides evidence for a multicistronic transcription unit including a VSG gene. Cell 51, 273–281.
12 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) The anatomy and transcription of a telomeric expression site for variant-specific surface antigens in *T. brucei*. Cell 51, 261–272.
13 Kidder, G.W. and Dutta, B.N. (1958) The growth and nutrition of *Crithidia fasciculata*. J. Gen. Microbiol. 18, 621–638.
14 Hughes, D., Schneider, C. and Simpson, L. (1982) Isolation and characterization of drug resistant mutants of *Crithidia fasciculata*. J. Parasitol. 68, 642–649.
15 Muhich, M., Hughes, D., Simpson, A. and Simpson, L. (1987) The monogenetic kinetoplastid protozoan, *Crithidia fasciculata*, contains a transcriptionally active, multicopy mini-exon sequence. Nucleic Acids Res. 15, 3141–3153.
16 Grumont, R., Washtien, W., Caput, D. and Santi, D. (1986) Bifunctional thymidylate synthase-dihydrofolate reductase from *Leishmania tropica*: sequence homology with the corresponding monofunctional proteins. Proc. Natl. Acad. Sci. USA 83, 5387–5391.
17 Hoffman, E.K., Trusko, S.P., Freeman, N.A. and George, D.L. (1987) Structural and functional characterization of the promoter region of the mouse c-Ki-*ras* gene. Mol. Cell. Biol. 7, 2592–2596.
18 Honkawa, H., Masahashi, W., Hashimoto, S. and Hashimoto-Gotoh, T. (1987) Identification of the principal promoter sequence of the c-Ha-*ras* transforming oncogene: deletion analysis of the 5′-flanking region by focus formation assay. Mol. Cell. Biol. 7, 2933–2940.
19 Ishii, S., Xu, Y., Stratton, R.H., Roe, A., Merlino, G.T. and Pastan, I. (1985) Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene. Proc. Natl. Acad. Sci. USA 82, 4920–4924.
20 Levanon, D., Lieman-Hurwitz, J., Dafni, N., Wigderson, M., Sherman, L., Bernstein, Y., Laver-Rudich, Z., Danciger, E., Stein, O. and Groner, Y. (1985) Architecture and anatomy of the chromosomal locus in human chromosome 21 encoding the Cu/Zn superoxide dismutase. EMBO J. 4, 77–84.
21 Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. 12, 7035–7056.
22 Mishoe, H., Brady, J.N., Radonovich, M. and Salzman, N.P. (1984) Simian virus 40 guanine-cytosine-rich sequences function as independent transcriptional control elements in vitro. Mol. Cell. Biol. 4, 2911–2920.
23 Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G., Brown, M.S., Goldstein, J.L. and Luskey, K.L. (1984) HMG CoA reductase: a negatively regulated gene with unusual promoter and 5′ untranslated regions. Cell 38, 275–285.
24 Shih, C., Linial, M., Goodenow, M.M. and Hayward, W.S. (1984) Nucleotide sequence 5′ of the chicken c-*myc* coding region: localization of a non-coding exon that is absent from myc transcripts in most avian leukosis virus-induced lymphomas. Proc. Natl. Acad. Sci. USA 81, 4697–4701.
25 Valerio, D., Duyvesteyn, M.G.C., Dekker, M.M., Weeda, G., Berkvens, T.M., van der Voorn, L., van Ormondt, H. and van der Eb, A.J. (1985) Adenosine deaminase: characterization and expression of a gene with a remarkable promoter. EMBO J. 4, 437–443.
26 Melera, P.W., Hession, C.A., Davide, J.P., Scotto, K.W., Biedler, J.L., Meyers, M.B. and Shanske, S. (1982) Antifolate-resistant Chinese hamster cell mRNA-directed overproduction of multiple dihydrofolate reductases from a series of independently derived sublines containing amplified dihydrofolate reductase genes. J. Biol. Chem. 257, 12939–12949.
27 Nunberg, J.H., Kaufman, R.J., Schimke, R.T., Urlaub, G. and Chasin, L.A. (1978) Amplified dihydrofolate reductase genes are localized to a homogeneously staining region of a single chromosome in a methotrexate resistant Chinese hamster ovary cell line. Proc. Natl. Acad. Sci. USA 75, 5553–5556.
28 Beverley, S., Ellenberger, T. and Cordingley, J. (1986) Primary structure of the gene encoding the bifunctional dihydrofolate reductase-thymidylate synthetase of *Leishmania major*. Proc. Natl. Acad. Sci. USA 83, 2584–2588.
29 Meek, T.D., Garvey, E.P. and Santi, D.V. (1985) Purification and characterization of the bifunctional thymidylate synthetase-dihydrofolate reductase from methotrex-

ate-resistant *Leishmania tropica*. Biochemistry 24, 678–686.
30 Beverley, S.M., Coderre, J.A., Santi, D.V. and Schimke, R.T. (1984) Unstable DNA amplifications in methotrexate-resistant *Leishmania* consist of extrachromosomal circles which relocalize during stabilization. Cell 38, 431–439.
31 Kapler, G.M., Zhang, K. and Beverley, S.M. (1987) Sequence and S1 nuclease mapping of the 5′ region of the dihydrofolate reductase-thymidylate synthase gene of *Leishmania major*. Nucleic Acids Res. 15, 3369–3383.
32 Simpson, L. and Berliner, J. (1974) Isolation of the kinetoplast DNA of *Leishmania tarentolae* in the form of a network. J. Protozool. 21, 382–393.
33 Berk, A. and Sharp, P. (1977) Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell 12, 721–732.
34 Masuda, H., Simpson, L., Rosenblatt, H. and Simpson, A. (1979) Restriction map, partial cloning and localization of 9S and 12S RNA genes on the maxicircle component of the kinetoplast DNA of *Leishmania tarentolae*. Gene 6, 51–73.
35 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.
36 Carle, G. and Olson, M. (1985) An electrophoretic karyotype for yeast. Proc. Natl. Acad. Sci. USA 82, 3756–3760.
37 Proudfoot, N., Shander, M., Manley, J., Gefter, M. and Maniatis, T. (1980) Structure and in vitro transcription of human globin genes. Science 209, 1329–1336.
38 Kanehisa, M.I. (1982) Los Alamos sequence analysis package for nucleic acids and proteins. Nucleic Acids Res. 10, 183–196.
39 Washtien, W.J., Grumont, R. and Santi, D.V. (1985) DNA amplification in antifolate-resistant *Leishmania*. J. Biol. Chem. 260, 7809–7812.
40 Hamada, H., Petrino, M.G. and Kakunaga, T. (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA 79, 6465–6469.
41 Yen, J.J. and Kellems, R.E. (1987) Independent 5′- and 3′-end determination of multiple dihydrofolate reductase transcripts. Mol. Cell. Biol. 7, 3732–3739.
42 Dragon, E., Sias, S., Kato, E. and Gabe, J. (1987) The genome of *Trypanosoma cruzi* contains a constitutively expressed, tandemly arranged multicopy gene homologous to a major heat shock protein. Mol. Cell. Biol. 7, 1271–1275.
43 Meade, J., Shaw, J., Lemaster, S., Gallagher, G. and Stringer, J. (1987) Structure and expression of a tandem gene pair in *Leishmania donovani* that encodes a protein structurally homologous to eukaryotic cation-transporting ATPases. Mol. Cell. Biol. 7, 3937–3946.
44 Landfear, S., Miller, S. and Wirth, D. (1986) Transcriptional mapping of *Leishmania enrietti* tubulin mRNAs. Mol. Biochem. Parasitol. 21, 235–245.
45 Sather, S. and Agabian, N. (1985) A 5′ spliced leader is added in *trans* to both alpha and beta-tubulin transcripts in *Trypanosoma brucei*. Proc. Natl. Acad. Sci. USA 82, 5695–5699.
46 Imboden, M.A., Laird, P.W., Affolter, M. and Seebeck, T. (1987) Transcription of the intergenic regions of the tubulin gene cluster of *Trypanosoma brucei*: evidence for a polycistronic transcription unit in a eukaryote. Nucleic Acids Res. 15, 7357–7368.
47 Everett, R.D., Baty, D. and Chambon, P. (1983) The repeated GC-rich motifs upstream from the TATA box are important elements of the SV40 early promoter. Nucleic Acids Res. 11, 2447–2464.
48 Saffer, J.D. and Singer, M.F. (1984) Transcription from SV40-like monkey DNA sequences. Nucleic Acids Res. 12, 4769–4788.
49 Spandidos, D.A. and Riggio, M. (1986) Promoter and enhancer like activity at the 5′-end of normal and T24 Ha-*ras*1 genes. FEBS Lett. 203, 169–174.
50 Farnham, P.J., Abrams, J.M. and Schimke, R.T. (1985) Opposite-strand RNAs from the 5′ flanking region of the mouse dihydrofolate reductase gene. Proc. Natl. Acad. Sci. USA 82, 3978–3982.
51 Mitchell, P.J., Carothers, A.M., Han, J.H., Harding, J.D., Kas, E., Venolia, L. and Chasin, L.A. (1986) Multiple transcription start sites, DNase I-hypersensitive sites, and an opposite-strand exon in the 5′ region of the CHO dhfr gene. Mol. Cell. Biol. 6, 425–440.
52 McGrogan, M., Simonsen, C.C., Smouse, D.T., Farnham, P.J. and Schimke, R.T. (1985) Heterogeneity at the 5′ termini of mouse dihydrofolate reductase mRNAs. J. Biol. Chem. 260, 2307–2314.
53 Konarska, M.M., Padgett, R.A. and Sharp, P.A. (1985) *Trans* splicing of mRNA precursors in vitro. Cell 42, 165–171.
54 Solnick, D. (1985) *Trans* splicing of mRNA precursors. Cell 42, 157–164.
55 Izant, J.G. and Weintraub, H. (1984) Inhibition of thymidine kinase gene expression by anti-sense RNA: a molecular approach to genetic analysis. Cell 36, 1007–1015.
56 Nepveu, A. and Marcu, K.B. (1986) Intragenic pausing and anti-sense transcription within the murine c-*myc* locus. EMBO J. 5, 2859–2865.
57 Ralph, D., Huang, J. and van der Ploeg, L.H.T. (1988) Physical identification of branched intron sideproducts of splicing in *Trypanosoma brucei*. EMBO J. 7, 2539–2545.