Physical and transcriptional analysis of the *Trypanosoma brucei* genome reveals a typical eukaryotic arrangement with close interspersion of RNA polymerase II- and III-transcribed genes

Maria A. Marchetti[1], Christian Tschudi[1], Erika Silva[1] and Elisabetta Ullu[1,2,*]

[1]Department of Internal Medicine and [2]Department of Cell Biology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520-2088, USA

ABSTRACT

To further our understanding of the structural and functional organization of the *Trypanosoma brucei* genome, we have searched for and analyzed sites in the genome where Pol II transcription units meet Pol III genes. Physical and transcriptional maps of cosmid clones spanning the Pol III-transcribed U2 small nuclear RNA (snRNA) and U3 snRNA/7SL RNA gene loci demonstrated that single-copy Pol II genes are closely associated with Pol III-transcribed genes, being separated from each other by 0.6–3 kb. At the U3/7SL transcriptional domain, two Pol II transcription units converged from either side of the chromosome towards the Pol III genes, suggesting that at least for the chromosome containing the U3 snRNA and 7SL RNA genes, there exist two distinct initiation sites for Pol II. Furthermore, in all cases the Pol III genes hallmark the end of Pol II transcription units, suggesting perhaps a functional role for this genetic arrangement. Lastly, we asked whether the environment within a Pol III transcriptional domain allowed expression of pre-mRNA. To test this we inserted a CAT gene cassette, seemingly promoterless but endowed with pre-mRNA processing signals, in the chromosome between the U3 snRNA and 7SL RNA genes. Interestingly, abundant CAT mRNA was produced suggesting that the Pol III genes in the immediate vicinity did not prevent access of presumably Pol II to the CAT gene cassette. We propose that either CAT mRNA is synthesized by Pol II run-through transcription or by Pol II initiation upstream from the CAT gene.

INTRODUCTION

The genome of *Trypanosoma brucei* is transcribed by three RNA polymerase activities which appear to be analogous to the three classical eukaryotic RNA polymerases. The sensitivity of trypanosomal RNA polymerases to the transcription inhibitor α-amanitin is similar to that of vertebrate RNA polymerases, namely RNA polymerase I (Pol I) is insensitive to the inhibitor, whereas Pol II and Pol III are highly and moderately sensitive, respectively. The division of labor among the trypanosome RNA polymerases and their mode of transcription follow what has been described for classical eukaryotic RNA polymerases with the following exceptions. First, Pol I transcribes the genes coding for the large ribosomal RNAs and, in addition, the genes coding for stage-specific surface antigens (1). Second, the small nuclear RNA (snRNA) genes coding for U2, U3 and U4 snRNA are transcribed by Pol III (2,3) and not by Pol II, as is the case in most other eukaryotes (4). Third, the mode of transcription of protein coding genes is polycistronic, rather than monocistronic. A corollary of this latter feature is that 5′ end formation of mRNA molecules is achieved post-transcriptionally by *trans*-splicing (reviewed in 5), an RNA processing reaction, and not by transcription initiation as is the case in most other eukaryotes.

At present our knowledge of the long range structural organization of transcription units in *T.brucei* is rather limited. Most of what we know about this topic is derived from the study of developmentally regulated genes coding for the variant surface glycoproteins (VSG) of bloodstream forms and the procyclic acidic repetitive proteins (PARP) of insect form trypanosomes. Transcriptionally competent VSG genes are located adjacent to telomeres and the active VSG gene resides in a polycistronic expression site with the promoter 40–50 kb upstream (reviewed in 6). In contrast, PARP genes belong to a small transcription unit of a few kb in size (7,8). Promoters for VSG genes, as well as PARP genes, resemble Pol I promoters in several respects (1). On the other hand, we know much less about the functional organization of Pol II-transcribed genes, namely all the house-keeping protein coding genes. In particular, the identification of specific transcription initiation sites has yet to be achieved. One exception is the recently described putative heat shock promoter, but the specific sequences involved in transcription initiation have not yet been defined (9). The failure to identify Pol II promoters in the trypanosome genome has led to speculations that

*To whom correspondence should be addressed at: Department of Internal Medicine, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520-2088, USA. Tel: +1 203 785 3563; Fax: +1 203 785 3864; Email: elisabetta.ullu@yale.edu

perhaps Pol II promoters might be sparse, and there has even been the extreme suggestion that each chromosome harbors only one Pol II promoter.

In the case of Pol III genes, the study of Pol III transcription units revealed some degree of clustering of small nuclear and cytoplasmic RNA genes and tRNA genes. This was brought to light by an *in vivo* and *in vitro* analysis of the architecture of the U2 (2), U3, and U6 snRNA, and 7SL RNA gene promoters (10,11). In all cases examined, a divergently oriented tRNA gene (or a tRNA-like gene in the case of the U2 gene) was required for expression of the companion small RNA gene. In particular, the integrity of the A and B boxes of each tRNA intragenic promoter was essential for expression of the companion small RNA gene. At present we do not understand the selective advantage behind the genetic and functional linkage between Pol III-transcribed genes in *T.brucei*, although a necessity for coordinate regulation of these genes cannot be ruled out.

Thus, so far Pol I-, Pol II- and Pol III-transcribed genes have been studied as isolated units, and to our knowledge no attempt has been made to look at the overall organization of transcription units in trypanosomes. In particular, the polycistronic nature of Pol II genes and the apparent non-existence of promoters raises the question of whether Pol II transcription units are physically and functionally separated from other genes or whether they are intermingled with them. To further our knowledge on this subject, we decided to take as starting points two well-characterized Pol III loci, namely the one coding for the U2 snRNA (12) and the other coding for the closely linked U3 snRNA and 7SL RNA (13). Physical and transcriptional mapping of cosmid clones encompassing the U2 and U3/7SL RNA gene loci revealed that Pol II transcription units are intimately interspersed with Pol III genes. In all of the three cases examined, Pol III genes constitute the 3′ boundary of Pol II transcription units with the 3′ ends of the terminal mRNAs being between 0.6 and 3 kb from the closest Pol III gene. Interestingly, using a seemingly promoterless CAT expression cassette we showed that the environment within a Pol III transcriptional domain is permissive to expression of pre-mRNA.

MATERIALS AND METHODS

DNA analysis

Cosmid library construction. We constructed a cosmid library of Sau3A digested genomic DNA from our trypanosome procyclic strain YTAT1.1 in plasmid SuperCos1 (Stratagene). The packaged material was sent to Dr Sara Melville at the Institute of Pathology in Cambridge (UK) where 5000 independent clones were arranged in microtiter dishes from which high density filters were generated. pU2cos-1 and pU3cos-1 were isolated by screening the library with gene-specific probes.

Physical mapping. Physical mapping of the cosmid clones was carried out following the procedure suggested by Stratagene. Briefly, partial and complete restriction digests of cosmid DNAs were generated and Southern blots were hybridized to end-labeled T3 and T7 oligonucleotides which flank the *Bam*HI cloning site.

Probes

Total ^{32}P-labelled RNA for transcriptional analysis was synthesized in permeable procyclic trypanosome cells and purified as previously described (14,15). For each hybridization experiment

we used the RNA synthesized from ~2.5 to 5.0 × 10^8 cells during a 20 min incubation. Double- and single-stranded PCR probes were synthesized using standard conditions. Radiolabeled cDNA was synthesized using as a template poly(A)$^+$ RNA isolated from procyclic trypanosomes, and oligo(dT) as a primer.

Identification of cDNA clones and determination of the direction of transcription

To identify pc8 and pc17, *Sal*I fragment 2 of pU2cos-1 DNA (Fig. 1A) was used as a probe to screen a cDNA library from procyclic trypanosomes. Sequence analysis of pU2cos-1 DNA confirmed that both cDNAs were derived from fragment 2, and that the corresponding genes were transcribed towards the RNA polymerase III transcribed domain (Fig. 1A). To characterize the RNA polymerase II transcribed regions of pU3cos-1 DNA, we first constructed a library of random 2–3 kb DNA fragments in plasmid vector pCRII™ (InVitrogen). DNAs from 100 independent clones were spotted onto nitrocellulose filters which were then hybridized to the repetitive sequence INGI (16), to a U3-snRNA coding region probe and to cDNA. Out of 25 cDNA positive subclones, we identified four inserts containing an internal *Eco*RV site, namely p76, p85, p33 and p91, and we placed and oriented them on the cosmid map by hybridization to *Eco*RV digests with single- and double-stranded probes. cDNA clones for all probes were then selected by screening a cDNA library. To determine the direction of transcription of 33, 91, 76 and 85 mRNAs, we synthesized single-stranded probes using as primers the M13 forward or reverse oligonucleotides which flanked the inserts on either side. Each strand synthesized from each subclone was hybridized to northern blots of total RNA. In all four cases, only one of the strands hybridized to the corresponding RNA. By definition, the hybridizing strand represents the template strand for transcription by RNA polymerase II. The orientation of transcription was then determined by combining the results of the northern blots with the orientation of the subclones on the physical map.

Construction of the pU3CAT vector

A 5 kb *Sal*I fragment containing a portion of the 85 gene and the U3/7SL RNA gene loci was subcloned from pU3cos-1 into the pCRII^{TM} vector to generate pU3. At the same time by PCR mutagenesis we inserted a unique *Pac*I restriction site between the two tRNA genes located upstream of the U3 and 7SL RNA genes using as a template plasmid pG7SL, which was previously described (10) and contains a tagged version of the 7SL RNA gene. Using the unique restriction sites *Bam*HI and *Not*I we exchanged the U3/7SL RNA genes of the *Pac*I-containing version of pG7SL for the homologous sequences of pU3, to generate pU3P. pU3P DNA was linearized at the *Pac*I site where the CAT reporter gene cassette was inserted to generate the final vector named pU3CAT. The cassette was derived from plasmid pβαCAT3′P and contains βα-tubulin sequences that provide signals for *trans*-splicing, the CAT coding region and the PARP signals for 3′-end formation

Co-transfection of procyclic trypanosomes and selection of stable cell lines

Trypanosoma brucei rhodesiense YTAT1.1 procyclic trypanosomes were cultured as previously described (17). 100 µg of pU3CAT
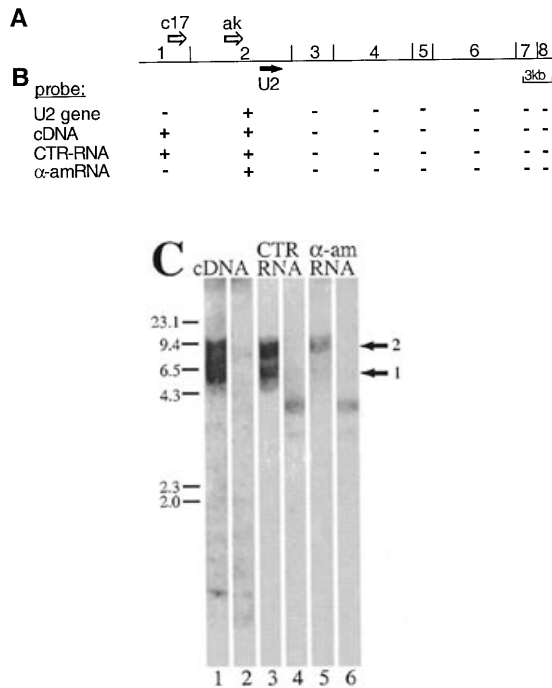
Figure 1 — A/B/C (pU2cos-1)

probe:	1	2 (U2)	3	4	5	6	7	8
U2 gene	−	+	−	−	−	−	−	−
cDNA	+	+	−	−	−	−	−	−
CTR-RNA	+	+	−	−	−	−	−	−
α-amRNA	−	+	−	−	−	−	−	−

(Arrows: c17 above lane 1, ak above lane 2; 3kb marker at lanes 7/8. Panel C: Southern blot lanes 1–6 with λDNA *Hin*dIII markers 23.1, 9.4, 6.5, 4.3, 2.3, 2.0; solid arrows 2 and 1.)

Figure 1. Physical and transcriptional map of pU2cos-1. (**A**) Physical map of pU2cos-1 DNA. The size of the insert is ~37 kb and vertical bars indicate the position of *Sal*I sites. Open arrows indicate the approximate position and direction of transcription of two cDNAs derived from this locus. A solid arrow indicates the direction of transcription of the U2 snRNA gene. (**B**) Transcriptional map of pU2cos-1 DNA. The symbols + or − stand for positive or negative hybridization of the restriction fragments to the four different probes namely the U2 gene, cDNA, total RNA (CTR-RNA) and α-amanitin RNA (α-am RNA). (**C**) Southern blot hybridization of pU2cos-1 DNA. Cosmid DNA was digested with *Sal*I and *Not*I (lanes 1, 3 and 5). pSPR3A1, a plasmid clone containing the U2 gene, was digested with *Eco*RI and *Hin*dIII (lanes 2, 4 and 6). DNA samples in triplicate were electrophoresed through a 0.8% agarose gel, transferred to a nitrocellulose membrane and hybridized to the three different probes, as indicated above the lanes. Bars on the side indicate the position of λDNA *Hin*dIII restriction fragments, as molecular weight markers in kb. Solid arrows indicate the positions of *Sal*I fragments 1 and 2 of pU2cos-1 DNA.

Figure 2 — A/B/C (pU3cos-1)

probe:	1	2	3	4	5	6	7 (U3 7SL)	8 (tRNAs)
U3 gene	−	−	−	−	−	−	+	−
cDNA	+	+	+	−	−	+	+	+
CTR-RNA	+	+	+	−	−	+	+	+
α-amRNA	−	−	−	−	−	−	+	+

(Arrows: c33 c91 above lane 3 region, c85 above lane 6, c76 TBX92 above lane 8; 2kb marker at lane 1. Panel C: Southern blot lanes 1–6 with markers 23.1, 9.4, 6.5, 4.3, 2.3, 2.0; solid arrows 7 and 8.)
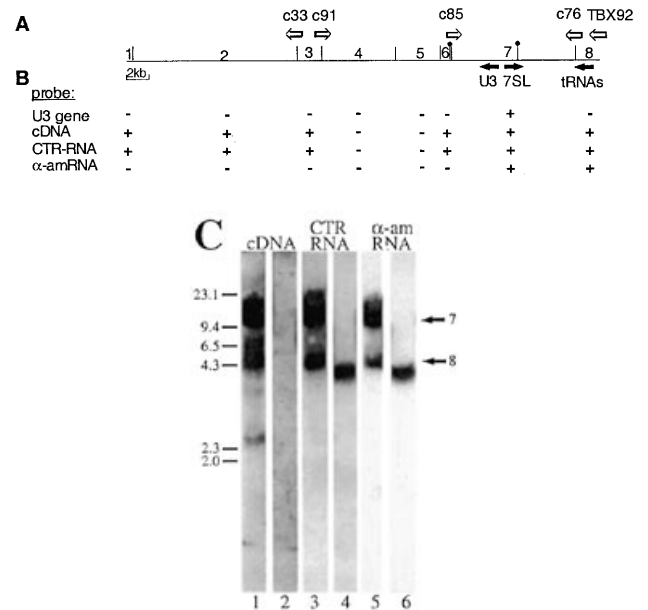
Figure 2. Physical and transcriptional map of the pU3cos-1. (**A**) Physical map of pU3cos-1 DNA. The size of the insert is ~32 kb and the vertical bars indicate the position of *Eco*RV sites. Open arrows indicate the approximate position and direction of transcription of four cDNAs derived from this locus. Solid arrows indicate the direction of transcription of U3 snRNA, tRNA$_{gln}$ and 7SL RNA genes. A vertical bar with a solid circle indicates *Sal*I sites used to subclone the region of interest. (**B**) Transcriptional map of pU3cos-1 DNA. Symbols are as defined in the legend of Figure 1B. (**C**) Southern blot hybridization of pU3cos-1 DNA. Lanes 1, 3 and 5, pU3cos-1 DNA digested with *Eco*RV and *Not*I; lanes 2, 4 and 6, pG7SL (a subclone containing the U3 snRNA and 7SL RNA genes), digested with *Bam*HI. Arrows on the side indicate the position of *Eco*RV fragments 7 and 8. Electrophoresis, hybridization conditions and molecular weight markers were as described in the legend of Figure 1C.

DNA linearized at the *Sma*I site were mixed with 5 µg of pXS2GFP DNA (18) linearized at the *Mlu*I site, and the mixture was electroporated into procyclic trypanosome cells as described. After transfection the cells were selected with 15 µg/ml of G418 and cloned by limited dilution in microtiter dishes (19). Two cell lines were selected and named C1 and C2.

Hybridizations

Southern and northern blots were performed according to standard procedures (20). For hybridization of radiolabelled total RNA we used aqueous solutions at 65°C for 24–48 h.

CAT assay

Freeze–thaw lysates from ~10^7 trypanosomes were assayed for CAT activity as described using [^{14}C]chloramphenicol (20). Reactions were incubated for 30 min at 37°C.

RESULTS

Pol II transcription units are in proximity to Pol III-transcribed snRNA genes

To further our understanding of the long range organization of Pol III genes, we chose the U2 snRNA (12) and U3 snRNA/7SL RNA gene loci, which are single-copy (13,21) and map to different chromosomes (data not shown). Cosmid clones encompassing the U2 and U3/7SL loci were isolated and one clone each, namely pU2cos-1 and pU3cos-1, was selected for further detailed analysis. Initial hybridization experiments suggested that the inserts contained Pol II-transcribed regions in addition to Pol III genes. Thus, we decided to establish a detailed transcriptional map of the two cosmid clones. Figures 1A and 2A show long range physical and transcriptional maps, for pU2cos-1 and pU3cos-1, respectively. The transcriptional maps were obtained by hybridization of Southern blots of restricted cosmid DNA to the following radiolabelled probes: (i) control RNA (CTR-RNA), represents total ^{32}P-labelled RNA synthesized in permeable procyclic trypanosome cells during a 20 min incubation time period (14,15). This probe is enriched for newly synthesized RNAs and detects chromosomal regions transcribed by Pol I, II or III. (ii) α-amanitin (α-am) RNA, which is radiolabelled RNA synthesized in permeable trypanosome cells in the presence of 5 µg/ml of the transcription inhibitor α-amanitin. Under these

conditions Pol II is >90% inhibited, whereas Pol I and Pol III are not affected (14). Thus, this probe can detect regions transcribed by Pol III or Pol I. (iii) cDNA which was synthesized from procyclic poly(A)$^+$ RNA using oligo(dT) as a primer, and thus it is enriched for mRNA complementary sequences. In addition, we synthesized an RNA probe in permeable trypanosomes in the presence of 1 mg/ml α-amanitin, which will only allow transcription by Pol I, but the experiment was negative.

For pU2cos-1 DNA (Fig. 1C) we found that only *Sal*I fragments 1 and 2 gave positive hybridization signals. In particular, fragment 1 hybridized to cDNA and control RNA but not to α-am RNA, suggesting the presence of a Pol II transcription unit. Fragment 2 hybridized to all three probes. This indicated that this region was likely to be transcribed by both Pol II and Pol III. Indeed, this restriction fragment contains the U2 snRNA gene which is transcribed by Pol III. The specificity of the hybridization was demonstrated by the hybridization of plasmid pSPR3A1, a clone containing the U2 snRNA coding region (14). As expected the U2 snRNA gene hybridized to CTR-RNA (lane 4), α-am RNA (lane 6) but not to cDNA (lane 2). Downstream from *Sal*I fragment 2, we were unable to detect significant hybridization with any of the three probes.

In the case of the U3 snRNA gene locus (Fig. 2C), two prominent hybridizing bands were detected with all three probes. These corresponded to *Eco*RV fragments 7 (which contains the U3 snRNA, 7SL and associated tRNA genes) and 8, whose positions are indicated in Figure 2A. As in the case of the U2 cosmid, these hybridization results indicated potential interspersion of Pol III and Pol II transcription units. In addition, in longer exposures of the autoradiogram we detected Pol II-transcribed regions in the areas spanning *Eco*RV fragments 1, 2, 3 and 6.

Stable mRNAs are transcribed from single-copy genes surrounding U2 and U3/7SL RNA loci

We next searched for stable mRNAs originating from these regions and used a variety of approaches that are described in detail in Materials and Methods. Briefly, cDNA clones were isolated and then characterized by DNA sequencing, northern and Southern blot analysis. For pU2cos-1, two cDNA clones specifying two different loci were characterized, namely pc17 and pc8 (Fig. 1A). Sequence analysis revealed that pc8 codes for the enzyme adenylate kinase (22), whereas no match was found for pc17 in the available databases.

In the case of pU3cos-1 four cDNA clones were selected, termed pc33, pc91, pc85 and pc76 (Fig. 2A). Clones pc33 and pc91 were not characterized further, because we discovered that the genomic region spanning fragments 2, 3 and 4 of pU3cos-1 was unstable in cultured procyclic *T.brucei* cells. As a result the sequence arrangement present in the isolated cosmid clone did not correspond to that found in genomic DNA (data not shown). In light of this observation, we decided to concentrate our further analysis on the portions of pU3cos-1 DNA downstream of fragment 4 till the end of the insert. The relevant portions of this region were defined by the pc85 and pc76 cDNA clones, which were separated by ~10 kb of DNA containing the U3 snRNA, 7SL RNA and associated tRNA genes. So far, the sequence of pc85 and pc76 did not reveal a convincing match in database searches. However, by sequencing the ends of the pU3cos-1 cosmid insert (Fig. 2A), we recognized a previously identified 'silent' open reading frame (ORF) called TBX92 (23), which instead is
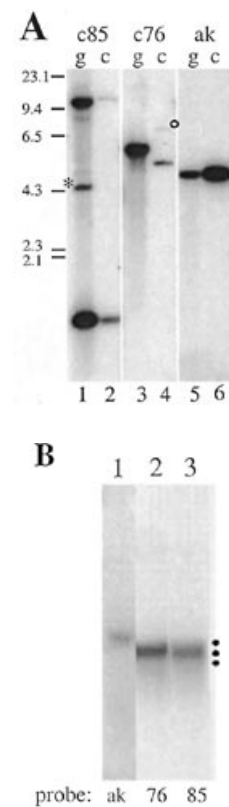

Figure 3. The ak, 76 and 85 genes produce stable mRNAs in procyclic trypanosomes. (**A**) Southern hybridization of genomic and cosmid DNAs. g, procyclic genomic DNA digested with *Eco*RV (lanes 1, 3 and 5); c, cosmid DNAs digested with *Eco*RV plus *Not*I (lanes 2 and 4, pU3cos-1 DNA); lane 6, pU2cos-1 DNA. The blot was hybridized to radiolabelled gene-specific probes as indicated above the lanes. AK stands for adenylate kinase. An asterisk indicates a cross-hybridizing band of unknown origin. A dot indicates the cosmid vector DNA which hybridized to the probe. The molecular weight of λDNA fragments digested with *Hin*dIII are indicated in kb. (**B**) Northern hybridization. Total RNA (10 μg) from procyclic trypanosomes was electrophoresed through a 1.2% agarose–2.2 M formaldehyde gel, transferred to a nitrocellulose membrane and hybridized to radiolabelled gene-specific probes as indicated below each lane. The positions of the large ribosomal RNAs (2250, 1850 and 1350 nt) are indicated by dots on the side.

abundantly expressed in our procyclic trypanosome strain (data not shown).

In order to establish whether the genes coding for adenylate kinase (ak), pc76 and pc85 were single-copy and to confirm the structure of the cosmid DNAs, Southern blots of restricted cosmid and trypanosome genomic DNA were hybridized with gene specific probes (Fig. 3A). The ak probe recognized a single band of identical molecular weight in genomic (lane 5) and in cosmid DNA (lane 6). In the case of pc76 there was also a single hybridizing band, but the cosmid hybridizing band (lane 4) was smaller than that seen in the genome (lane 3). This can be explained because with this particular restriction digest the pc76 gene was contained within an end fragment of the cosmid clone. However, other Southern hybridizations confirmed that the cosmid-derived and the genomic pc76 gene were indistinguishable (data not shown). For pc85 the digestion with *Eco*RV split the 85 coding region, thus producing two distinct hybridizing bands both in genomic (lane 1) and in cosmid (lane 2) DNA. In summary, the
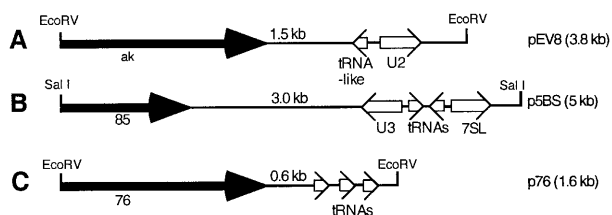
Figure 4. Schematic representation of the linkage between RNA polymerase II and RNA polymerase III transcription units at the U2, U3/7SL and tRNA$_{gln}$ gene loci. The diagrams depict the following plasmid clones: (**A**) pEV8 which is an *Eco*RV subclone of pU2cos-1; (**B**) p5BS, which is a *Sal*I subclone from pU3cos-1; (**C**) p76, which is a subclone derived from a random fragment library of pU3cos-1 DNA (see Materials and Methods). The inserts were sequenced and the arrangement of the various genes is shown. Solid black arrows indicate the direction of transcription of the RNA polymerase II transcribed genes. The open arrows indicate the direction of transcription of genes transcribed by RNA polymerase III. The distance between the RNA polymerase II and the closest Pol III gene is shown for each locus. The drawing is not to scale.

gene loci identified by the ak, pc76 and pc85 clones were single-copy in the trypanosome genome and the arrangement in the cosmid clones was identical to that in the genome.

Hybridization of ak, pc76 and pc85 probes to northern blots of total trypanosome RNA indicated that the corresponding genes were transcribed to produce stable mRNAs (Fig. 3, lanes 1–3). Using single-stranded DNA probes these experiments also established the direction of transcription as shown in Figures 1A and 2A (see Materials and Methods for technical details). For the U2 cosmid, we found that ak and pc17 were transcribed in the same direction and both of them pointed towards the U2 gene. At the U3/7SL RNA locus pc85 and pc76 were instead transcribed from opposite directions of the chromosome, and both transcription units pointed towards the U3/7SL RNA genes. The direction of transcription of these genes was also confirmed by direct sequence analysis (see below).

The 3′ termini of the loci defined by ak, pc76 and pc85 are in close proximity to Pol III genes

The physical mapping showed that ak and the 85 gene are in close proximity to Pol III genes. To determine the exact distance between the 3′ ends of ak and of the 85 gene and their corresponding downstream Pol III genes, we sequenced several kb of DNA from each of the two loci (DDBJ/EMBL/GenBank accession nos AF047722 and AF047723). The results of these sequence analyses are schematically shown in Figure 4. For the ak gene there are ~1.5 kb of DNA between the gene 3′ end and the tRNA-like gene located upstream of the U2 coding region. In the case of the 85 gene, the distance between its 3′ end and the U3 snRNA gene is ~3.0 kb. These 'spacer' sequences were analyzed for the presence of ORFs, but only relatively small ORFs were detected. In addition, in the case of the 76 locus we sequenced a subclone derived from random cloning of the cosmid DNA (see Materials and Methods; DDBJ/EMBL/GenBank accession no. AF0447724), and serendipitously discovered that there was a tRNA$_{gln}$ ~600 nt downstream from the 76 gene 3′ end, as defined by sequencing the corresponding cDNA. This tRNA$_{gln}$ gene is part of a cluster of tRNA genes which was previously sequenced (24). The 0.6 kb of DNA separating the 3′ end of the 76 gene from the tRNA$_{gln}$ gene did not show any potential ORF.

To determine whether the DNA downstream of ak and the 85 gene was transcribed to produce stable RNAs, which had escaped our previous analysis, we performed northern hybridizations with specific probes, and also RT–PCR with antisense oligonucleotides complementary to the putative ORFs. The results of these analyses were negative (data not shown), suggesting that both the ak and the 85 genes, like the 76 gene, represented the 3′ terminal genes of their respective Pol II transcription units. Close inspection of the DNA sequence separating the Pol II and Pol III genes did not reveal any unusual base compositions or repeated sequences. Thus, we concluded that Pol II transcription units can be intimately interspersed with Pol III genes and that in all three cases examined Pol III genes constituted the 3′ boundary of Pol II transcription units.

A promoterless CAT gene cassette is expressed when inserted at the U3/7SL RNA locus

The finding that Pol III genes were located at the 3′ end of three different Pol II transcription units suggested to us that this genetic organization might have a functional meaning. One possibility was that Pol III genes were functioning like road blocks for Pol II transcription and somehow contribute to Pol II transcription termination. We decided to probe whether the environment within a Pol III transcriptional domain was permissive to expression of pre-mRNA by inserting a promoterless reporter gene in between the U3 snRNA and the 7SL RNA genes. The detailed structure of this locus is shown in Figure 5A. The U3 snRNA and 7SL RNA genes are divergently transcribed and separated by two tRNA genes, each of which provides promoter functions to its companion small RNA gene (10). The 3′ ends of the tRNA genes are separated by ~100 nt (13). Within this 100 nt we inserted a promoterless chloramphenicol acetyl transferase (CAT) gene cassette to generate the targeting vector pU3CAT (Fig. 5A). The CAT gene cassette consisted in a 5′ to 3′ direction of β–α tubulin sequences (from the nucleotide after the β-tubulin termination codon to the nucleotide preceding the α-tubulin ATG) which provided all the necessary signals for *trans*-splicing (25), the CAT coding region and the 3′ untranslated region and polyadenylation site of the PARP A gene. Additionally, the 7SL RNA gene of pU3CAT was modified by insertion of a 'tag' to monitor its expression (10).

The pU3CAT construct was targeted to the U3/7SL RNA locus as shown in Figure 5A by cotransfection with plasmid pXS2GFP (Fig. 5C). PXS2GFP carries the neomycin phosphotransferase gene as a selectable marker and was targeted to the tubulin locus. Two clonal cell lines, named C1 and C2, were chosen for further analysis. Correct insertion of the pU3CAT targeting vector by homologous recombination should result in the genetic arrangement shown in Figure 5B. Next, we digested genomic DNA with *Not*I and *Eco*RI and hybridized the corresponding Southern blot with a 7SL coding region probe (Fig. 5). The C1 and C2 DNAs (lanes 5 and 6) gave rise to two hybridizing bands. The largest one had the same molecular weight as the wt locus (lane 4, control DNA), whereas the smaller one of ~5.5 kb in size had the molecular weight predicted from the structure of the recombined chromosome (Fig. 5B). PhosphorImager quantitation of the hybridizing bands showed that for the C1 and C2 cell lines there was twice as much radioactivity associated with the band diagnostic of the wt locus as compared with the band diagnostic of the recombined locus. This was consistent with the duplication of the 7SL RNA gene
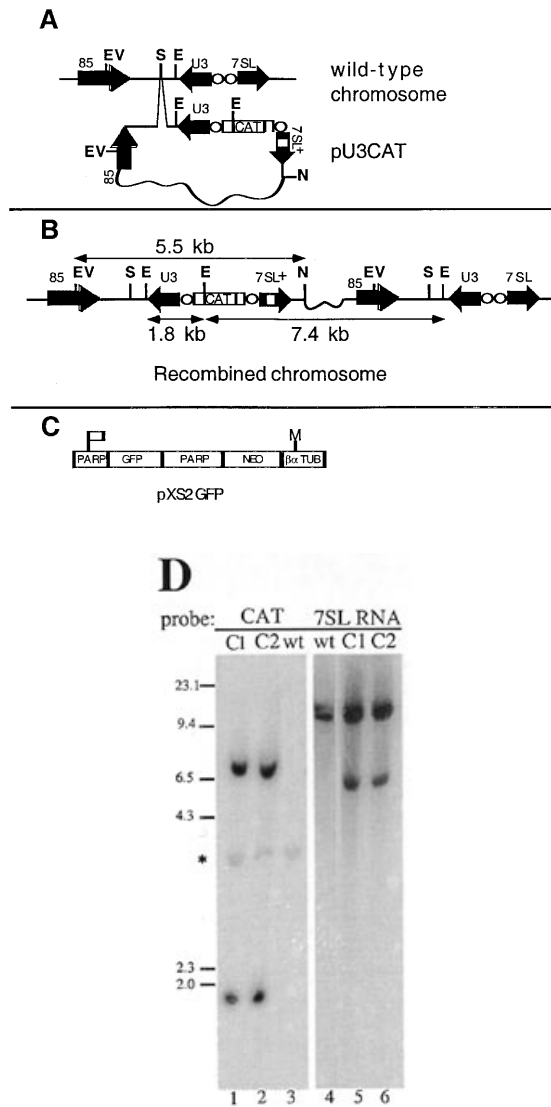
Figure 5. Targeting of the pU3CAT vector to the U3/7SL RNA gene locus by homologous recombination. (**A**) Structure of the U3/7SL RNA gene locus and of the pU3CAT vector. Restriction sites for *Eco*RV (EV), *Sma*I (S), *Eco*RI (E) and *Not*I (N) are indicated by vertical bars. The *Not*I site is derived from the polylinker of the plasmid vector. Open circles indicate tRNA genes, a wavy line indicates plasmid sequences and the open box indicates the tag within the 7SL RNA gene. Plasmid pU3CAT contains the reporter gene CAT flanked by sequences required for pre-mRNA processing (see text for details). pU3CAT was linearized at the *Sma*I site and transfected into procyclic trypanosomes. (**B**) Predicted structure at the U3/7SL RNA gene locus after integration of pU3CAT DNA. Symbols are as described above. The distances in kb between adjacent diagnostic restriction sites are indicated above and below the diagram. (**C**) The pXS2GFP vector. pXS2GFP has the plasmid backbone of pXS2 and contains the green fluorescent protein (GFP) coding region and the neomycin phosphotransferase gene (neo) arranged as a dicistronic transcription unit under the control of the PARP promoter (indicated by a flag). pXS2GFP was digested with *Mlu*I (M) and targeted to the tubulin locus. (**D**) Southern blot analysis of genomic DNA from wild-type, C1 and C2 stables cells lines. Genomic DNA was digested with *Not*I and *Eco*RV (lanes 4–6) or *Eco*RI (lanes 1–3). Lanes 3 and 4, genomic DNA from wild-type cells (wt); lanes 1 and 5, genomic DNA from C1 stable cell line (C1); lanes 2 and 6, genomic DNA from C2 stable cell line (C2). The DNA was electrophoresed through a 0.7% agarose gel and transferred to a nitrocellulose membrane. Subsequently the nitrocellulose membranes were hybridized to a CAT coding region probe (lanes 1–3) or to a 7SL RNA probe (lanes 4–6). The 3.5 kb band indicated by an asterisk (*) and present in lanes 1–3 is a background band whose origin is unknown. The molecular weights of λDNA fragments digested with *Hin*dIII are indicated in kb.

locus which was produced upon insertion of the pU3CAT construct into the chromosome. Next we tested whether the C1 and C2 cell lines contained single or multiple insertions of pU3CAT DNA. In this case the DNAs were digested with *Eco*RI and hybridized to a CAT coding region probe. Whereas the control DNA showed no hybridization to the probe (lane 3), the C1 and C2 DNAs displayed hybridization to two bands, ~7.4 and 1.8 kb in size, consistent with a single insertion. The recombined chromosome structure was also confirmed by PCR with specific oligonucleotide primers (data not shown).

In order to determine whether the insertion of the CAT gene cassette had effects on the expression of the neighbouring Pol III genes, we had introduced a 'tag' into the 7SL RNA gene of the pU3CAT vector (10). The hybridization of total RNA from the C1 cell line with a 7SL RNA probe is shown in Figure 6A. RNA derived from control cells (lane 1) showed one hybridizing band of ~270 nt, which is the size of *T.brucei* 7SL RNA. In the case of the C1 cell line (lane 2), two 7SL RNA bands were visible. One corresponded to wt 7SL RNA, and the second larger one represented the tagged 7SL RNA transcript, as we have previously shown (10). However, the amount of tagged 7SL RNA in the C1 cell line was substantially lower than wt 7SL RNA. A similar result was obtained with the C2 cell line (data not shown). This could be due to several reasons and we have not investigated this phenomenon further. Nevertheless, from a qualitative point of view it appeared that insertion of the CAT gene cassette did not prevent expression of the downstream 7SL RNA gene.

Next we tested whether we could detect CAT activity in the C1 and C2 cell lines. Extracts were prepared and tested for CAT activity using [^{14}C]chloramphenicol and thin layer chromatography (Fig. 6B). Relative to control cells (lane 1) extracts from both the C1 and C2 cell lines (lanes 2 and 3) had considerable CAT activity, as it is evident by the conversion of [^{14}C]chloramphenicol to the two monoacetylated forms. Lastly, we performed northern blot analysis of C1 and C2 total RNAs with a CAT coding region probe (Fig. 6C). CAT mRNA of the expected molecular weight of ~1.0 kb was observed in both cell lines.

DISCUSSION

In this paper we report on the analysis of the physical and transcriptional organization of the *T.brucei* genome with respect to domains transcribed by RNA polymerase II and RNA polymerase III. We found that at the U2 and U3/7SL RNA gene loci Pol II transcription units are interspersed with and closely juxtaposed to the Pol III genes. We think this arrangement is likely to be widespread in the trypanosome genome because we have found that ~5% of the cosmid clones of our cosmid library hybridize to radiolabelled gel-purified tRNAs (data not shown). This high number of hybridizing clones is consistent with a scattered distribution of tRNA genes throughout the trypanosome genome. Since there are many other Pol III genes encoded in the genome, it is conceivable that they are also similarly distributed and that Pol II transcription units are often interrupted by Pol III genes. This is exactly what we discovered at the U2 and U3/7SL gene loci which were only chosen on the basis of the fact that they are single copy. Moreover, we found that the U2, U3 and U6 snRNA genes map to different chromosomes (data not shown), further strengthening the notion that Pol III genes are not clustered at a few chromosomal locations. This is not surprising since Pol II/Pol III interspersion is a characteristic of the
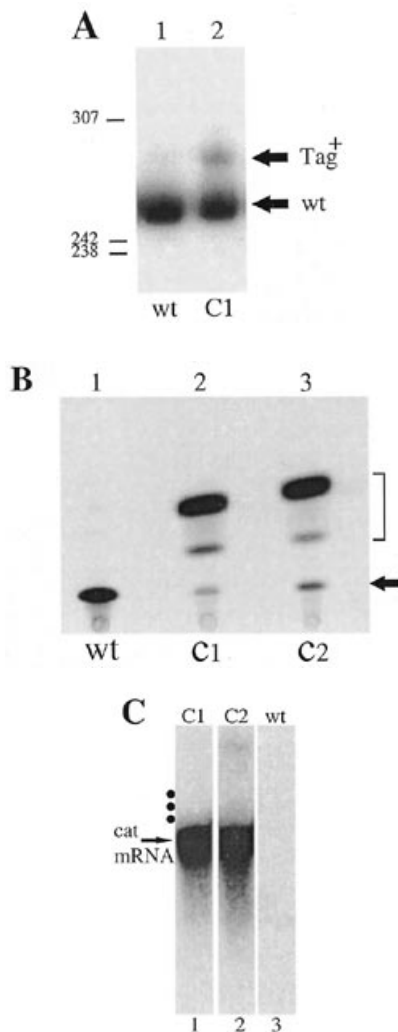
Figure 6. Characterization of C1 and C2 cell lines. (**A**) Expression of the 'tagged' 7SLRNA. Total RNA was separated on a 5% denaturing polyacrylamide gel, transferred to a nylon membrane and hybridized to a 5′-end-labelled oligonucleotide complementary to positions 117 to 136 of the 7SL RNA coding region. Lane 1, control RNA; lane 2, C1 RNA. The positions of the wt and tagged (tag[+]) 7SL RNA are indicated by arrows. Relevant ^{32}P-labelled *Msp*I fragments of pBR322 are indicated as molecular weight markers in bp. (**B**) CAT activity assay. Trypanosome cell lysates were assayed for CAT activity as described in Materials and Methods using [^{14}C]chloramphenicol. Acetylated chloramphenicol (indicated by a bar) was separated from unmodified chloramphenicol (indicated by an arrow) using thin layer chromatography. Lane 1, extract from wild-type cells (wt); lane 2, extract from C1 cells; lane 3, extract from C2 cells. (**C**) Expression of CAT mRNA. Total RNA was fractionated as described in the legend of Figure 3B and hybridized to a radiolabelled CAT gene probe. Lane 1, C1 RNA; lane 2, C2 RNA; lane 3, wt RNA. Dots indicate the large ribosomal RNAs as described in the legend of Figure 3B.

eukaryotic genome. In the sequenced yeast genome tRNA genes are present on all chromosomes and the distance between neighbouring tRNA genes varies greatly from a few to >150 kb. Once the sequence of the *T.brucei* genome is on its way, it will become clearer whether the degree of interspersion of Pol II and Pol III genes is indeed similar in trypanosomes and yeast. Furthermore, the distance between the 3′ end of the 85, 76 and ak genes and the downstream Pol III genes (Fig. 4) is very similar to what is found in the yeast genome for Pol II/Pol III genes,

suggesting that also in this respect the trypanosome genome is not unique. Instead, what was unexpected is that in all three cases examined the direction of Pol II transcription was towards the Pol III genes, or in other words that the Pol III genes constituted the 3′ boundary of Pol II transcription units. Of course one trivial explanation for this observation is that this is just a serendipitous finding due to the small number of loci which were analyzed. On the other hand, it is possible that this is a valid observation and that Pol III genes serve a functional purpose by being at the end of Pol II transcription units. Because Pol III genes are known to form stable complexes in association with transcription factors (26) and to inhibit nucleosome positioning (27), it is possible that the chromatin structure surrounding Pol III genes has some special features. It is tempting to speculate that the presence of actively transcribed Pol III genes might act as a road block for transcription by RNA polymerase II, and that transcription by Pol II might terminate or attenuate before reaching the Pol III genes. The fact that we were unable to detect any Pol II transcription downstream of the U2 gene in pU2cos-1 DNA might support the 'attenuation/termination' hypothesis. It would make sense that also in the case of the 85 and 76 transcription units, transcription by Pol II would attenuate or terminate before reaching the Pol III genes. This is because the consequences of Pol II transcribing through Pol III genes would be to produce antisense transcripts for U3, 7SL and the associated tRNA genes, and also antisense transcripts of the adjacent Pol II genes, if transcription were to proceed even downstream from the Pol III genes. We have searched for these putative antisense transcripts by hybridization of steady-state RNA with single-stranded sense probes, but the results have been negative (data not shown). Although we cannot exclude that antisense transcripts are produced and rapidly degraded, it might be harmful to the cell to synthesize antisense transcripts of small nuclear and cytoplasmic RNAs which provide essential housekeeping functions for pre-ribosomal RNA processing (U3), co-translational insertion of membrane proteins in the ER membrane (7SL) and translation (tRNA).

An interesting problem specific to protein coding genes located at the end of Pol II transcription units, is how the process of 3′ end formation of mature mRNA takes place. This is because we (25) and others (28–29) have shown that for genes that are internal to a polycistronic transcription unit, the process of 3′ end cleavage/ polyadenylation of pre-mRNA is tightly and functionally coupled to active *trans*-splicing at a 3′ splice site (3′ SS), which is usually located 100 to a few hundred nucleotides downstream from the poly(A) site. We have analyzed the sequences downstream of 85, 76 and ak genes for the presence of 3′ splice acceptor regions, namely an AG dinucleotide preceded by a polypyrimidine tract. We have found several potential 3′ SS candidates in all three cases, and we are now investigating whether 3′ end formation of the ak gene is dependent upon *trans*-splicing at a downstream 3′ SS.

To test whether the environment within a Pol III transcriptional domain was permissive for transcription of pre-mRNA, we inserted a CAT reporter gene cassette in between the tRNAs upstream of the U3 and 7SL RNA genes. The sequences upstream of the CAT coding region were derived from the β–α tubulin genes and contained the signals required for β-tubulin mRNA polyadenylation and α-tubulin mRNA *trans*-splicing (25). We chose these sequences because they have been characterized in detail and because to our knowledge nobody had reported promoter activity associated with the tubulin genes, so our assumption was that the pU3CAT construct was promoterless.

For the interpretation of the experiment it is important to consider the following. First, the CAT cassette was not linked to a selectable marker, and therefore no direct pressure was applied to select cells that had integrated the CAT cassette. Rather, selection of trypanosomes cells that had taken up exogenous DNA was achieved by co-transfecting plasmid pXS2GFP, which carries the neomycin phosphotransferase gene, and targeting the dominant selectable marker to the tubulin locus. Second, after integration by homologous recombination the structure of the chromosome upstream and immediately downstream of the CAT cassette was unchanged relative to the wild-type locus. As shown by northern blot analysis and by enzymatic activity assays, abundant CAT mRNA and protein were produced in C1 and C2 cell lines. Since in these cells expression of G418 resistance was independent from expression of the CAT reporter cassette, we think it is unlikely that our results can be explained by selection of random mutations that activated expression of the CAT gene. Preliminary experiments using α-amanitin sensitivity of newly-synthesized RNA indicate that the CAT cassette is transcribed by Pol II. In a way this was not surprising since it was unlikely that Pol III could transcribe through the β–α-tubulin sequences and the CAT coding region, given the presence of numerous T-tracks which are known to function as terminators for Pol III transcription (26). Two hypotheses can be put forward to explain expression of the CAT cassette placed in between Pol III genes, namely the run-through and the specific-initiation hypothesis. As discussed above it is possible, although we think unlikely, that Pol II transcription coming from the 85 gene does not halt before the Pol III genes but runs through them. This would generate pre-mRNA transcripts containing antisense U3 sequences linked to sense transcripts of the CAT gene cassette. *Trans*-splicing and polyade-nylation would then generate mature CAT mRNA. We have searched by RT–PCR for pre-mRNA transcripts containing U3 antisense sequences linked to the CAT coding region, but our attempts did not meet with any success (data not shown). The alternate possibility to explain expression of the CAT gene is that Pol II transcription initiated upstream from the CAT coding region. This could have happened because either the β–α sequences themselves that were linked to the 5′ end of the CAT gene or sequences further upstream from the site of insertion provided initiation site(s) for Pol II transcription. This is an attractive possibility that we are now in the process of testing.

ACKNOWLEDGEMENTS

REFERENCES

1 Chung,H., Lee,M. and Van der Ploeg,L. (1992) *Parasitol. Today*, **8**, 414.
2 Fantoni,A., Dare,A.O. and Tschudi,C. (1994) *Mol. Cell. Biol.*, **14**, 2021–2028.
3 Günzl,A., Tschudi,C., Nakaar,V. and Ullu,E. (1995) *J. Biol. Chem.*, **270**, 17287–17291.
4 Hernandez,N. (1992) In McKnight,S. and Yamamoto,K. (eds), *Transcription of Vertebrate snRNA, Genes and Related Genes. Transcriptional Regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
5 Ullu,E., Tschudi,C. and Günzl,A. (1996) In Smith,D.F. and Parsons,M. (eds), *Molecular Biology of Parasitic Protozoa*. Oxford University Press, Oxford, pp. 115–133.
6 Pays,E., Vanhamme,L. and Berberof,M. (1994) *Annu. Rev. Microbiol.*, **48**, 25–52.
7 Rudenko,G., Le Blancq,S., Smith,J., Lee,M.G., Rattray,A. and Van der Ploeg,L.H. (1990) *Mol. Cell. Biol.*, **10**, 3492–3504.
8 Clayton,C.E., Fueri,J.P., Itzhaki,J.E., Bellofatto,V., Sherman,D.R., Wisdom,G.S., Vijayasarathy,S. and Mowatt,M.R. (1990) *Mol. Cell. Biol.*, **10**, 3036–3047.
9 Lee,M.G. (1996) *Mol. Cell. Biol.*, **16**, 1220–1230.
10 Nakaar,V., Dare,A.O., Hong,D., Ullu,E. and Tschudi,C. (1994) *Mol. Cell. Biol.*, **14**, 6736–6742.
11 Nakaar,V., Gunzl,A., Ullu,E. and Tschudi,C. (1997) *Mol. Biochem. Parasitol.*, **88**, 13–23.
12 Tschudi,C., Richards,F.F. and Ullu,E. (1986) *Nucleic Acids Res.*, **14**, 8893–8903.
13 Mottram,J.C., Bell,S.D., Nelson,R.G. and Barry,J.D. (1991) *J. Biol. Chem.*, **266**, 18313–18317.
14 Ullu,E. and Tschudi,C. (1990) *Nucleic Acids Res.*, **18**, 3319–3326.
15 Ullu,E. and Tschudi,C. (1995) *J. Biol. Chem.*, **270**, 20365–20369.
16 Kimmel,B.E., ole-MoiYoi,O.K. and Young,J.R. (1987) *Mol. Cell. Biol.*, **7**, 1465–1475.
17 Tschudi,C. and Ullu,E. (1988) *EMBO J.*, **7**, 455–463.
18 Bangs,J.D., Brouch,E.M., Ransom,D.M. and Roggy,J.L. (1996) *J. Biol. Chem.*, **271**, 18387–18393.
19 Wirtz,E., Hartmann,C. and Clayton,C. (1994) *Nucleic Acids Res.*, **22**, 3887–3894.
20 Ausubel,F., Brent,R., Kingston,R., Moore,D., Seidman,J., Smith,J. and Struhl,K. (eds) (1987) *Current Protocols in Molecular Biology*. John Wiley and Sons, Inc., USA.
21 Mottram,J., Perry,K.L., Lizardi,P.M., Luhrmann,R., Agabian,N. and Nelson,R.G. (1989) *Mol. Cell. Biol.*, **9**, 1212–1223.
22 Shahjahan,M., Yamada,M., Tanaka,H. and Nakazawa,A. (1991) *Gene*, **107**, 313–317.
23 Schlaeppi,K. and Seebeck,T. (1989) *Nucleic Acids Res.*, **17**, 5399.
24 Campbell,D.A. (1989) *Nucleic Acids Res.*, **17**, 9479.
25 Matthews,K.R., Tschudi,C. and Ullu,E. (1994) *Genes Dev.*, **8**, 491–501.
26 Geiduschek,E.P. and Tocchini-Valentini,G.P. (1988) *Annu. Rev. Biochem.*, **57**, 873–914.
27 Morse,R.H., Roth,S.Y. and Simpson,R.T. (1992) *Mol. Cell. Biol.*, **12**, 4015–4025.
28 Vassella,E., Braun,R. and Roditi,I. (1994) *Nucleic Acids Res.*, **22**, 1359–1364.
29 Hug,M., Hotz,H.R., Hartmann,C. and Clayton,C. (1994) *Mol. Cell. Biol.*, **14**, 7428–7435.