A new twist in trypanosome RNA metabolism: *cis*-splicing of pre-mRNA

GUNNAR MAIR,^{1,11} HUAFANG SHI,^{1,11} HONGJIE LI,¹ APPOLINAIRE DJIKENG,¹
HERNAN O. AVILES,² JOSEPH R. BISHOP,³ FRANCO H. FALCONE,⁴
CRISTINA GAVRILESCU,⁵ JACQUI L. MONTGOMERY,⁶ M. ISABEL SANTORI,⁷
LEAH S. STERN,⁸ ZEFENG WANG,⁹ ELISABETTA ULLU,^{1,10} and CHRISTIAN TSCHUDI¹

¹Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut 06520-8022, USA

²Indiana State University, Life Sciences Department, Terre-Haute, Indiana 47809, USA

³Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

⁴Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom

⁵Department of Microbiology and Immunology, Cornell University, Ithaca, New York 14850, USA

⁶The Walter and Eliza Hall Institute, Post Office, Royal Melbourne Hospital, North Melbourne 3050, Australia

⁷Instituto de Investigaciones en Ingeniería Genética y Biología Molecular, Vuelta de Obligado 2490, 1428 Buenos Aires, Argentina

⁸113B Veterans Administration Medical Center, San Francisco, California 94121, USA

⁹Department of Biological Chemistry, Johns Hopkins University, Baltimore, Maryland 21205, USA

¹⁰Department of Cell Biology, Yale University School of Medicine, New Haven, Connecticut 06520-8022, USA

ABSTRACT

It has been known for almost a decade and a half that in trypanosomes all mRNAs are *trans*-spliced by addition to the 5' end of the spliced leader (SL) sequence. During the same time period the conviction developed that classical *cis*-splicing introns are not present in the trypanosome genome and that the trypanosome gene arrangement is highly compact with small intergenic regions separating one gene from the next. We have now discovered that these tenets are no longer true. Poly(A) polymerase (PAP) genes in *Trypanosoma brucei* and *Trypanosoma cruzi* are split by intervening sequences of 653 and 302 nt, respectively. The intervening sequences occur at identical positions in both organisms and obey the GT/AG rule of *cis*-splicing introns. PAP mRNAs are *trans*-spliced at the very 5' end as well as internally at the 3' splice site of the intervening sequence. Interestingly, 11 nucleotide positions past the actual 5' splice site are conserved between the *T. brucei* and *T. cruzi* introns. Point mutations in these conserved positions, as well as in the AG dinucleotide of the 3' splice site, abolish intron removal *in vivo*. Our results, together with the recent discovery of *cis*-splicing introns in *Euglena gracilis*, suggest that both *trans*- and *cis*-splicing are ancient acquisitions of the eukaryotic cell.

Keywords: introns; kinetoplastidae; poly(A) polymerase; *trans*-splicing; *Trypanosoma brucei*; *Trypanosoma cruzi*

INTRODUCTION

Mature mRNA production in trypanosomes is a step-wise process that differs in several aspects from the biogenesis of mRNA in most eukaryotes (Ullu et al., 1996). First, trypanosome protein-coding genes are organized as polycistronic rather than monocistronic transcription units. Second, the 5' ends of all mature mRNAs are formed by *trans*-splicing, an RNA processing reaction, rather than by transcription initiation. The partners in *trans*-splicing are the polycistronic pre-mRNA, which

contains the mRNA coding regions preceded by a 3' splice site, and the spliced leader (SL) RNA, which provides the capped SL sequence and the 5' splice site. The first evidence for *trans*-splicing was obtained in trypanosomatid protozoa in 1982, when it was noted that an identical 39-nt SL sequence is present at the very 5' end of transcripts coding for different variant surface glycoproteins in *Trypanosoma brucei* (Boothroyd & Cross, 1982). Although first discovered in trypanosomes (Murphy et al., 1986; Sutton & Boothroyd, 1986; Laird et al., 1987), *trans*-splicing was subsequently found to be present in nematodes (Krause & Hirsh, 1987), euglenoids (Tessier et al., 1991), and certain trematodes (Rajkovic et al., 1990). Extensive analysis, both *in vivo* in trypanosomatids, and *in vitro* in nematodes, has clearly shown that *trans*-splicing proceeds

Reprint requests to: Dr. Christian Tschudi, Department of Internal Medicine, LCI 805, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520-8022, USA; e-mail: christian.tschudi@yale.edu.

¹¹The first two authors contributed equally to this work.

through a two-step reaction pathway that is directly analogous to *cis*-splicing, the removal of intervening sequences (Nilsen, 1993; Ullu et al., 1996). Furthermore, *cis*- and *trans*-splicing share a common set of small nuclear ribonucleoprotein (snRNP) cofactors, namely U2, U4, U5, and U6 snRNPs (Tschudi & Ullu, 1990; Hannon et al., 1991; Maroney et al., 1996). However, in contrast to *cis*-splicing, *trans*-splicing does not require U1 snRNP (Hannon et al., 1991).

Although nematodes, trematodes, and euglenoids carry out both *cis*- and *trans*-splicing, the notion developed over the years that trypanosomes lack intervening sequences and therefore the machinery to carry out *cis*-splicing. This was based on observations such as that the genes characterized so far did not have introns and that trypanosomes appeared to lack a homolog of the U1 snRNA (Mottram et al., 1989), which plays a crucial role in the initial recognition of the 5' splice site (Ares & Weiser, 1995). Although it was acknowledged that these were negative results, it nevertheless became an established fact that the genome of trypanosomes was devoid of intervening sequences and that *trans*-splicing was the only type of snRNP-mediated splicing present in these organisms.

We have now found that the poly(A) polymerase (PAP) genes of *T. brucei* and *Trypanosoma cruzi* are interrupted by intervening sequences, thus clearly establishing that *cis*-splicing is present in trypanosomatid protozoa.

RESULTS

The *T. brucei* poly(A) polymerase gene is split by an intervening sequence

Searching the available databases for trypanosome homologs of the polyadenylation machinery, we identified a *T. brucei* sequence tag with significant amino acid similarity to PAP from several different organisms. Using this information, we went on to isolate several *T. brucei* genomic clones and determined the sequence of the locus encoding the putative PAP (accession no. AF201909). This analysis revealed an open reading frame of 400 amino acids with 29% identity to residues 116–505 of bovine PAP (Raabe et al., 1991). However, the *T. brucei* PAP appeared to be lacking the amino terminus, including two conserved aspartate residues within the proposed catalytic domain (Asp-113 and Asp-115 in bovine PAP; see Martin & Keller, 1996). In particular, there was a TGA termination codon at the corresponding position of the Asp-115 codon. It was unlikely that this was due to a cloning or sequencing artifact, because the DNA sequence originally identified in the database also encoded a termination codon at that position. Using the available PAP genomic sequence as a probe in Southern blot analysis, we further established that the *T. brucei* PAP is a single-copy gene

in the genome (data not shown), thus eliminating the possibility that we were dealing with a pseudogene. We were intrigued by the fact that the termination codon TGA was immediately followed by an AG dinucleotide (TGAAG), the invariant sequence element of a 3' splice site. In addition, the AG dinucleotide was preceded by a polypyrimidine tract, suggesting the presence of a 3' splice acceptor region. Thus, even if there was no precedence, we considered the possibility that the *T. brucei* PAP gene was interrupted by an intervening sequence and that the amino terminus was encoded further upstream. To test this hypothesis, we performed 5'-end RACE with a PAP-specific primer and an oligonucleotide corresponding to part of the SL sequence, present at the 5' end of all mRNAs in trypanosomes. As shown in Figure 1, this resulted in the amplification of two cDNA products, and sequence analysis of the shorter fragment (193 bp) revealed that the SL sequence was attached to sequences immediately 3' to the AG dinucleotide of the putative 3' splice site. Thus, it appeared that *trans*-splicing of the SL sequence occurs at this position. On the other hand, the sequence of the 3' end of the longer cDNA fragment (471 bp) was identical to the PAP genomic sequence up to, but not including, the AG dinucleotide. The remaining sequence did not align with the available genomic sequence, but upon translation extended the open reading frame of PAP to include the conserved aspartate residues of the catalytic core, as well as the amino terminus. Further PCR analysis of genomic DNA with oligonucleotides specific for the PAP amino terminus and sequences located downstream of the AG dinucleotide revealed that the 5' portion of the *T. brucei* PAP was separated from the remainder of the PAP gene by approximately 600 bp (Figs. 1 and 2). Sequence analysis of genomic DNA confirmed this prediction and established that the *T. brucei* PAP gene is indeed interrupted by an intervening sequence of 653 bp (Fig. 2) and that the borders of the intron conform to the GT/AG rule of *cis*-splicing introns.

FIGURE 1. The *T. brucei* and *T. cruzi* PAP genes contain introns. PCR amplifications of cDNAs (cDNA) and genomic DNAs (gDNA) were carried out as described in Materials and Methods and the prominent bands in each lane were sequenced, leading to the structure shown in Figure 2.

The position of the PAP intron is conserved in *T. cruzi*

To test whether the presence of an intron in the PAP gene is a common feature in trypanosomatid protozoa, we took advantage of an EST encoding part of the *T. cruzi* PAP (accession number A1562587; 58% identity to amino acids 192–316 of the *T. brucei* PAP). Similar to the approach described above, 5'-end RACE was

used to obtain the 5' end of the *T. cruzi* PAP mRNA and then the genomic counterpart was PCR amplified (Fig. 1). This resulted in a genomic PCR fragment that was approximately 300 nt larger than what we predicted from the cDNA, suggesting that the *T. cruzi* PAP gene was also interrupted by an intervening sequence. Indeed, direct sequence analysis of the PCR fragments confirmed an intron of 302 bp in the *T. cruzi* PAP gene, with the position of the intron being identical to that in the *T. brucei* PAP gene (Fig. 2).

Comparison of the *T. brucei* and *T. cruzi* intron sequences revealed that, including the GT motif, 11 nt are conserved at the 5' splice site, whereas only the AG motif is conserved at the 3' splice site (Fig. 2). Interestingly, the first 5 nt of the 5' splice site (GTATG) are identical to the corresponding positions of the SL RNA 5' splice site. Although a polypyrimidine tract is found upstream of each 3' splice site, its position and extent is different in the two introns.

The *T. brucei* PAP intron is spliced in vivo

To begin to identify signals in the pre-mRNA required for *cis*-splicing, the *T. brucei* intron, flanked on the 5' and 3' sides by 93 and 101 nt of exon sequences, respectively, was fused to the CAT reporter gene and inserted into the pLew79 expression vector (Fig. 3A; construct WT). Expression of these sequences was driven by the PARP promoter and signals for 5' end processing by *trans*-splicing and 3' end formation and polyadenylation were provided by PARP sequences and the 3'UTR of aldolase, respectively (Wirtz et al., 1999). As a control for correct removal of the intron, a construct was assembled lacking the intervening sequence (Fig. 3A, construct cDNA). Both DNAs were transfected into cultured procyclic *T. brucei* cells by electroporation and RNA was prepared 3 h after transfection. In the following experiments RNA samples were equalized by monitoring expression of a cotransfected U6 snRNA gene (data not shown). To analyze the RNAs

FIGURE 3. In vivo splicing patterns of mutants in the 3' splice region. **A:** The structure of the expression constructs is shown. The *T. brucei* intron, including 93 (E1) and 101 (E2) nt of exon sequences, was placed between PARP sequences and the CAT reporter gene followed by aldolase 3' processing sequences (ALD). SL and A flags mark the positions of SL and poly(A) addition, respectively. **B:** RNA was prepared from procyclic *T. brucei* cells transfected with the indicated constructs and analyzed by reverse transcription followed by PCR amplification. The exact structure of the PCR products was determined by DNA sequencing and is shown schematically. The following mutations were tested: lane 4: the polypyrimidine tract, underlined in Figure 2B, was substituted in M(pPy) to 5'-AACGGG TCGCGACGGGAACGG-3'; lane 5: the AG dinucleotide of the 3' splice site was changed to CA in M(AG→CA); lane 6: the first 20 nt of exon 2 were changed in M(E2) to 5'-ATATACTCGCGACAACG ACT-3'. Drawings are not to scale.

generated, we employed 5'-end RACE using SL- and CAT-specific oligonucleotides.

When the cDNA construct was transfected into trypanosome cells, *trans*-splicing occurred at the PARP 3' splice site, as evident by a PCR fragment of the predicted size of 300 nt (Fig. 3B, lane 2) and direct

sequence analysis of this PCR product (data not shown). 5'-end RACE analysis of RNA isolated from cells transfected with the WT construct, containing the intron, produced four distinct PCR fragments of 178, 307, 543, and 960 nt (Fig. 3B, lane 3), which were subjected to DNA sequence analysis. As expected from its size, the 307-nt product originated from RNA, where the intron had been removed and the SL sequence was added at the PARP 3' splice site, thus demonstrating that the sequence information provided by the PAP intron and the flanking exon sequences was sufficient for accurate removal of the intervening sequence. The shortest PCR product was diagnostic of molecules where *trans*-splicing occurred at the 3' splice site of the intron, similar to what we observed in steady-state RNA (Fig. 1, lane 1). The 543-bp fragment represented RNA molecules that were *trans*-spliced at a cryptic site in the intron, whereas the 960-bp fragment originated from RNA molecules that were *trans*-spliced at the very 5' end at the PARP 3' splice site, but the intron was not removed by *cis*-splicing.

pre-mRNA signals required for removal of the intervening sequence

Having established an in vivo assay for *cis*-splicing, we next asked what signals in the pre-mRNA are required for accurate removal of the intron by introducing a series of mutations in the WT construct (Figs. 3 and 4). At the 3' splice site, we replaced the AG dinucleotide, the upstream polypyrimidine tract, and the first 20 nt of exon sequences. The most dramatic effect was obtained with the AG mutation, which resulted in no detectable *cis*-splicing and *trans*-splicing at the intron 3' splice site (Fig. 3B, lane 5). The polypyrimidine tract mutation led to the additional use of a cryptic 3' splice site in the intron and drastically reduced *trans*-splicing at the intron 3' splice site (Fig. 3B, lane 4). This mutation did not significantly affect the amount of *cis*-spliced product, although we reproducibly see a double band at that position, suggesting the use of an alternate GT

FIGURE 4. Effects of 5' splice site mutations on in vivo splicing. The nucleotide sequence of the mutated region is shown on the left and for each mutation only substituted nucleotides are indicated. DNA constructs were transfected into *T. brucei* cells and isolated RNA was analyzed by 5' end RACE.

or AG dinucleotide. In contrast, the exon mutation had no detectable effect (Fig. 3B, lane 6).

Our major target in this mutational analysis was the 5' splice site region, because comparison of the *T. brucei* and *T. cruzi* intron revealed that 11 nt of intron sequences are conserved. Based on only two examples, this level of sequence conservation could be fortuitous. Nevertheless, it was intriguing to test whether part or all of the 11 nt play a role in intron removal. As can be seen in Figure 4, substitutions of nucleotide positions 1–10 interfered with splicing and did not result in a detectable PCR product corresponding to accurate removal of the intron (lanes 4–7), whereas substitution of position 11 drastically reduced the amount of *cis*-spliced PCR product (lane 8). However, mutations in sequences not conserved between the *T. brucei* and *T. cruzi* 5' splice site region (positions 12–14) did not affect *cis*-splicing (Fig. 4, lanes 9–11).

DISCUSSION

Here, we have used data generated by the ongoing trypanosome genome sequencing projects and DNA transfection assays to demonstrate that *cis*-splicing introns are an integral part of the trypanosome genome and that trypanosomes have the machinery to carry out both *cis*- and *trans*-splicing. In addition, we showed that mutations of the conserved sequences at the 5' and 3' splice site of the intron severely reduce correct splicing in vivo. These results definitively resolve a long-standing issue regarding the apparent lack of *cis*-splicing introns in these organisms.

Although we have at present too few sequences available to establish consensus sequence features for trypanosome introns, it was nevertheless tantalizing to find that 11 nt were conserved at the 5' splice site between such divergent genera as *T. brucei* and *T. cruzi*. This high level of conservation is reminiscent of a class of rare eukaryotic introns, the AT-AC introns, where 8 nt at the 5' splice site are essentially invariant (Hall & Padgett, 1994). In contrast, the recently described spliceosomal introns in *Euglena gracilis* revealed only limited sequence conservation at the 5' splice site (Breckenridge et al., 1999). The importance of the 11-nt conserved trypanosome sequence was underscored by our mutational analysis, as changes in this sequence inhibited splicing of the intron in vivo. Analogous to the *cis*-splicing reaction in other eukaryotes, our prediction would be that these sequences interact through base pairing with the 5' terminal sequence of the U1 snRNA (Ares & Weiser, 1995). Indeed, the identification of a candidate U1 snRNA in *Crithidia fasciculata* and *Leishmania tarentolae* (Schnare & Gray, 1999), two related trypanosomatid protozoa, demonstrated a potentially extensive base-pair interaction of the 5' splice site of the PAP intron with the 5'

end of U1 snRNA (Fig. 5). We are currently in the process of isolating the *T. brucei* U1 snRNA, which will allow us to directly test whether compensatory mutations in the U1 snRNA 5' end can suppress the observed effect of the 5' splice site mutations.

We were somewhat surprised that the 3' splice site of the PAP intron was used to a significant level by the *trans*-splicing machinery. At present we do not know whether this shortened mRNA is translated and if it is, whether a functional protein is generated. However, we find it very unlikely that the translation product will be a functional PAP, because it is missing the catalytic domain. Regardless of this uncertainty, our finding raises important issues regarding the decision to *cis*- or *trans*-splice. Previous experiments in *Caenorhabditis elegans* have shown that a 3' *trans*-splice site can be suppressed by introducing a 5' splice site 50 nt upstream, thereby creating a *cis*-splicing intron (Conrad et al., 1993). Thus, in this case an upstream 5' splice signal dictated the *cis/trans* decision. However, it was also evident from these experiments that the further upstream the 5' splice site was placed, the less effective was the switching from a *trans*- to a *cis*-splice site. It was also noted in these studies that the context into which the splice site was positioned plays a crucial role in its utilization. What we observed in the PAP intron of *T. brucei* and *T. cruzi* is that there appears to be competition between the two types of splicing (Fig. 1). An analysis of the RNA constituents of nematode *cis*- and *trans*-spliceosomes provided some insight into this phenomenon. One of the surprising findings in these studies was the presence of the SL RNA in both the *trans*- and *cis*-spliceosomes, raising the intriguing possibility that the *cis/trans* decision takes place after spliceosome assembly (Maroney et al., 1996). However, this issue will need to be addressed in more detail experimentally.

From a phylogenetic point of view, the discovery of *cis*-splicing in trypanosomatid protozoa unifies the theme in that to date every organism with *trans*-splicing also carries out *cis*-splicing. Furthermore, our results, together with the recent evidence for *cis*-splicing introns with GT/AG consensus borders in *E. gracilis* (Breckenridge et al., 1999), where *trans*-splicing was described almost ten years ago (Tessier et al., 1991), suggest that both *trans*-splicing and *cis*-splicing are ancient acquisitions of the eukaryotic cell.

FIGURE 5. Diagram of the potential interaction that could form between the 5' splice site of the *T. brucei* PAP intron and the 5' end of the putative *Crithidia fasciculata* U1 snRNA (Schnare & Gray, 1999). The exon–intron boundary is indicated by a slash.

MATERIALS AND METHODS

Isolation of the *T. brucei* and *T. cruzi* poly(A) polymerase genes

Database searches were performed with the N-terminal half of bovine PAP (amino acids 1–400; see Raabe et al., 1991), which resulted in the identification of a *T. brucei* sequence of 433 bp on chromosome 1 with significant homology to PAP (clone trypKe1.p1p). (Sequence data for *T. brucei* chromosome 1 was obtained from The Sanger Centre website at http://www.sanger.ac.uk/Projects/T_brucei/. Sequencing of *T. brucei* chromosome 1 was accomplished as part of the Trypanosoma Genome Network with support by The Wellcome Trust.) The corresponding sequence was PCR amplified from *T. brucei* Y Tat 1.1 genomic DNA, used as a probe to screen a genomic phage library as described previously (Silva et al., 1998), and the PAP genomic locus was sequenced. Additional partial PAP sequences were identified in the *T. brucei* database at The Institute for Genome Research. PAP cDNAs were PCR amplified from oligo d(T)-primed cDNA with the SL-specific oligonucleotide Eco-SL (5'-GGGAATTC CGCTATTATTAGAACAGTTTCT-3') and TBPAP4 (5'-ATCG ACAGCTGTGCCTCTG-3'), located 136 bp downstream of the 3' splice site. The major products shown in Figure 1 were sequenced after agarose gel purification using the QIAquick PCR purification Kit (QIAGEN). Sequences encoding the 5' end of the PAP gene were PCR amplified from genomic DNA with a 5' primer (5'-GGAAGAAACACTTTATTG-3'), immediately adjacent to the SL addition site, and TBPAP4.

A search of the GenBank EST database using the basic local alignment search tool (BLAST; see Altschul et al., 1997) with the primary amino acid sequence of *T. brucei* PAP led to the identification of a *T. cruzi* PAP EST clone (accession number AI562587) with 58% identity to amino acids 192–316 of the *T. brucei* PAP. PAP-specific oligos were designed based on this EST sequence and for 5'-end RACE analysis, mRNAs were primed with a PAP-specific oligo TCPAP2 (5'-GCCCC GAGACTCTAAAGATG-3'), and cDNA was amplified by nested PCR using an SL-specific oligo for the 5' end (5'-CGCTATT ATTGATACAGTTTCTGTAC-3') and a PAP-specific oligo TCPAP3 (5'-CATAGCCCATAACTTCACGGAC-3') for the 3' end. The corresponding genomic DNA was then PCR amplified with TCPAP5 (5'-ATGATGGAGTTGCTGTACGGG-3') located 40 nt downstream of the SL addition site and TCPAP2.

DNA constructs

The wild-type intron construct was PCR generated by first amplifying a 847-bp region from procyclic *T. brucei* genomic DNA, containing 93 bp 5' exon, 653 bp intron, and 101 bp 3' exon from the PAP gene, using primers PAP9 (5'-AACAAGCTTGCACTCCCAAAACACTCAG-3') and PAP10 (5'-CAACGGTGGTATATCCAGTGCTGGCAGCGACTACAG ACAC-3'; the underlined sequence is complementary to the CAT reporter gene), and the CAT reporter gene with primers CAT1 (5'-CACTGGATATACCACCGTTG-3') and CAT-IN01 (5'-AACGGATCCCATATCACAGCTCACCG-3'). Then, the final construct was assembled by combining an aliquot of the two PCR reactions followed by amplification with primers PAP9 and CAT-IN01. The gel-purified PCR product was digested

with *Hind*III and *Bam*HI and cloned into pLew79 (Wirtz et al., 1999) between the PARP promoter and the 3' UTR of the aldolase gene. The corresponding intron-lacking region was amplified from cDNA and combined with the CAT reporter to yield the cDNA construct. Mutagenesis was performed by two sequential PCRs and mutations were verified by DNA sequencing.

Transient DNA transfection and RNA analysis

Transient transfection of procyclic forms of *T. brucei rho-desiense*, RNA isolation, and primer extension analysis were done essentially as described (Fantoni et al., 1994; Matthews et al., 1994). 5'-end RACE was carried out following the manufacturer's protocol (Gibco BRL). First-strand synthesis was initiated with the CAT-IN oligonucleotide (5'-CCCATATCAC CAGCTCACCG-3') 233 nt downstream from the CAT initiation codon. This was followed by PCR amplification with the SL-specific oligonucleotide Eco-SL (see above) and CAT-5 (5'-GCCATTGGGATATATCAACGGTGG-3'), located 22 nt downstream from the CAT initiation codon. 5'-end PCR products were analyzed by agarose gel electrophoresis and sequenced after purification using the QIAquick PCR purification Kit (QIAGEN). All transfections were performed independently at least three times and each PCR analysis was repeated at least three times with similar results. The relative band intensities shown do not change if the number of PCR cycles is altered.

ACKNOWLEDGMENTS

H.O.A., J.R.B., F.H.F., C.G., J.L.M., M.I.S., L.S.S., and Z.W. were participants of the 1999 Biology of Parasitism Course at the Marine Biology Laboratory in Woods Hole, Massachusetts, where part of this work was done. The Biology of Parasitism Course is in part supported by Burroughs Wellcome. At Yale University this study received support from National Institutes of Health (NIH) Grant AI28798 to E.U., who is the recipient of a Burroughs Wellcome Fund Scholar Award in Molecular Parasitology. Sequence data for *T. brucei* chromosome 1 was obtained from The Sanger Centre website at http://www.sanger.ac.uk/Projects/T_brucei/. Sequencing of *T. brucei* chromosome 1 was accomplished as part of the Trypanosoma Genome Network with support by The Wellcome Trust. Sequence data for *T. brucei* was also obtained from The Institute for Genomic Research website at <http://www.tigr.org>. Sequencing of *T. brucei* was accomplished with support from NIH.

Received October 25, 1999; returned for revision
November 4, 1999; revised manuscript received
November 8, 1999

REFERENCES

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
- Ares M Jr, Weiser B. 1995. Rearrangement of snRNA structure during assembly and function of the spliceosome. *Prog Nucleic Acid Res Mol Biol* 50:131–159.

- Boothroyd JC, Cross GA. 1982. Transcripts coding for variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5' end. *Gene* 20:281–289.
- Breckenridge DG, Watanabe Y, Greenwood SJ, Gray MW, Schnare MN. 1999. U1 small nuclear RNA and spliceosomal introns in *Euglena gracilis*. *Proc Natl Acad Sci USA* 96:852–856.
- Conrad R, Liou RF, Blumenthal T. 1993. Conversion of a *trans*-spliced *C. elegans* gene into a conventional gene by introduction of a splice donor site. *EMBO J* 12:1249–1255.
- Fantoni A, Dare AO, Tschudi C. 1994. RNA polymerase III-mediated transcription of the trypanosome U2 small nuclear RNA gene is controlled by both intragenic and extragenic regulatory elements. *Mol Cell Biol* 14:2021–2028.
- Hall SL, Padgett RA. 1994. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J Mol Biol* 239:357–365.
- Hannon GJ, Maroney PA, Nilsen TW. 1991. U small nuclear ribonucleoprotein requirements for nematode *cis*- and *trans*-splicing in vitro. *J Biol Chem* 266:22792–22795.
- Krause M, Hirsh D. 1987. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* 49:753–761.
- Laird PW, Zomerdijs JC, de Korte D, Borst P. 1987. In vivo labelling of intermediates in the discontinuous synthesis of mRNAs in *Trypanosoma brucei*. *EMBO J* 6:1055–1062.
- Maroney PA, Yu YT, Jankowska M, Nilsen TW. 1996. Direct analysis of nematode *cis*- and *trans*-spliceosomes: A functional role for U5 snRNA in spliced leader addition *trans*-splicing and the identification of novel Sm snRNPs. *RNA* 2:735–745.
- Martin G, Keller W. 1996. Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases. *EMBO J* 15:2593–2603.
- Matthews KR, Tschudi C, Ullu E. 1994. A common pyrimidine-rich motif governs *trans*-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes & Dev* 8:491–501.
- Mottram J, Perry KL, Lizardi PM, Luhrmann R, Agabian N, Nelson RG. 1989. Isolation and sequence of four small nuclear U RNA genes of *Trypanosoma brucei* subsp. *brucei*: Identification of the U2, U4, and U6 RNA analogs. *Mol Cell Biol* 9:1212–1223.
- Murphy WJ, Watkins KP, 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.
- Nilsen TW. 1993. *trans*-splicing of nematode premessenger RNA. *Annu Rev Microbiol* 47:413–440.
- Raabe T, Bollum FJ, Manley JL. 1991. Primary structure and expression of bovine poly(A) polymerase. *Nature* 353:229–234.
- Rajkovic A, Davis RE, Simonsen JN, Rottman FM. 1990. A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proc Natl Acad Sci USA* 87:8879–8883.
- Schnare MN, Gray MW. 1999. A candidate U1 small nuclear RNA for trypanosomatid protozoa. *J Biol Chem* 274:23691–23694.
- Silva E, Ullu E, Kobayashi R, Tschudi C. 1998. Trypanosome capping enzymes display a novel two-domain structure. *Mol Cell Biol* 18:4612–4619.
- Sutton RE, Boothroyd JC. 1986. Evidence for *trans* splicing in trypanosomes. *Cell* 47:527–535.
- Tessier LH, Keller M, Chan RL, Fournier R, Weil JH, Imbault P. 1991. Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by *trans*-splicing in *Euglena*. *EMBO J* 10:2621–2625.
- Tschudi C, Ullu E. 1990. Destruction of U2, U4, or U6 small nuclear RNA blocks *trans* splicing in trypanosome cells. *Cell* 61:459–466.
- Ullu E, Tschudi C, Gunzl A. 1996. *trans*-splicing in trypanosomatid protozoa. In: Smith DF, Parsons M, eds. *Molecular biology of parasitic protozoa*. Oxford, United Kingdom: Oxford University Press. pp 115–133.
- Wirtz E, Leal S, Ochatt C, Cross GA. 1999. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol Biochem Parasitol* 99:89–101.