ma brucei. Cold Spring Harbour Symp. quant. Biol. *45* (1981) 935–943.
2 Certa, U., Malaria vaccine. Experientia *47* (1991) 157–163.
3 Clegg, J. A., Smithers, S. R., and Terry, R. J., Acquisition of human antigens by *Schistosoma mansoni* during cultivation in vitro. Nature *232* (1971) 653–654.
4 Cross, G. A. M., Identification, purification and properties of clone specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei.* Parasitology *71* (1975) 393–417.
5 Goldring, O. L., Clegg, J. A., Smithers, S. R., and Terry, R. J., Acquisition of human blood group antigens by *Schistosoma mansoni.* Clin. exp. Immun. *26* (1976) 181–187.
6 Gray, A. R., Antigenic variation in clones of *Trypanosoma brucei.* 1. Immunological relationship of the clones. Ann. trop. Med. Parasit. *59* (1965) 27–36.
7 Hockley, D. J., and McLaren, D. J., *Schistosoma mansoni*: changes in the outer membrane of the tegument during development from cercaria to adult. Int. J. Parasit. *3* (1973) 13–25.
8 Kemp, W. M., Merritt, S. C., and Rosier, J. G., *Schistosoma mansoni*: identification of immunoglobulins associated with the tegument of adult parasites from mice. Exp. Parasit. *45* (1979) 81–87.
9 Khalife, J., Capron, M., Capron, A., Grzych, J., Butterworth, A. E., Dunne, D. W., and Ouma, J. H., Immunity in human schistosomiasis mansoni. Regulation of protective mechanisms by IgM blocking antibodies. J. exp. Med. *164* (1986) 1626–1640.
10 Khorana, H. G., Nucleic acid synthesis in the study of the Genetic Code, in: Nobel Lectures: Physiology or Medicine, vol. 4. American Elsevier, 1973.
11 Le Page, R., Studies on the variable antigens of *Trypanosoma brucei.* Trans. Roy. Soc. trop. Med. Hyg. *61* (1967) 139–140.
12 McLaren, D. J., Vaccines and Vaccination Strategies. Symposia of the British Society for Parasitology, vol. 26. Parasitology *98* (1989) S1–S99.
13 McLaren, D. J., and Hockley, D. J., Blood flukes have an outer double membrane. Nature *269* (1977) 147–149.
14 McLaren, D. J., and Terry, R. J., The protective role of acquired host antigens during schistosome maturation. Parasite Immun. *4* (1982) 129–148.
15 Meselson, M., and Stahl, F. W., The replication of DNA in *E. coli.* Proc. natl Acad. Sci. USA *44* (1957) 671–687.
16 Morrison, W. I., Immunological control of ticks and tick-borne parasitic diseases of livestock, in: Vaccines and Vaccination Strategies, Symposia of the British Society for Parasitology, vol. 26. Parasitology *98* (1989) S69–85.
17 Nirenberg, M., The Genetic Code. Scient. American March 1963, pp. 80–94.
18 Parkhouse, R. M. E., ed., Parasite Evasion of the Immune Response. Symposia of the British Society for Parasitology, vol. 21. Parasitology *88* (1984) 571–682.
19 Perez, H., and Terry, R. J., The killing of adult *Schistosoma mansoni* in the presence of antisera to host antigenic determinants and peritoneal cells. Int. J. Parasit. *3* (1973) 499–503.
20 Poynter, D., Jones, B. V., Nelson, A. M. R., Peacock, R., Robinson, J., Silverman, P. H., and Terry, R. J., Symposium of husk. Recent experiences with vaccination. Vet. Rec. *72* (1960) 1078–1096.
21 Simpson, A. J. G., Parasites and Molecular Biology. Symposia of the British Society for Parasitology, vol. 23. Parasitology *92* (1986) S1–S139.
22 Simpson, A. J. G., Walker, T., and Terry, R., An introduction to recombinant DNA technology, in: Parasites and Molecular Biology. Symposia of the British Society for Parasitology, vol. 23. Parasitology *92* (1986) S7–S14.
23 Smithers, S. R., and Terry, R. J., Naturally acquired resistance to experimental infections of *Schistosoma mansoni* in the rhesus monkey (*Macaca mulatta*). Parasitology *55* (1965) 701–710.
24 Smithers, S. R., and Terry, R. J., Resistance to experimental infection with *Schistosoma mansoni* in rhesus monkeys induced by the transfer of adult worms. Trans. R. Soc. trop. Med. Hyd. *61* (1967) 517–533.
25 Smithers, S. R., and Terry, R. J., Immunity to schistosomiasis. Ann. N.Y. Acad. Sci. *160* (1969) 826–840.
26 Smithers, S. R., Terry, R. J., and Hockley, D. J., Host antigens in schistosomiasis. Proc. R. Soc. B. *171* (1969) 483–494.
27 Steinert, M., and Pays, E., Genetic control of antigenic variation. Br. med. Bull. *41* (1985) 149–155.
28 Taylor, D. W., Schistosome vaccines. Experientia *47* (1991) 152–157.
29 Terry, R. J., Transmission of antimalarial immunity (*Plasmodium berghei*) from mother rats to their young during lactation. Trans. R. Soc. trop. Med. Hyg. *50* (1956) 41–46.
30 Torpier, G., Capron, A., and Ouasi, M., Receptor for IgG (Fc) and human β-2 microglobulin on *S. mansoni* schistosomulae. Nature *287* (1979) 447–449.
31 Trager, W., and Jensen, J. B., Human malaria parasites in continuous culture. Science *193* (1976) 673–675.
32 Vickerman, K., and Luckins, A., Localization of variable antigens in the surface coat of *Trypanosoma brucei* using ferritin-labelled antibody. Nature *224* (1969) 1125–1126.
33 Watson, J. D., and Crick, F. H. C., Molecular structure of nucleic acid. A structure for deoxyribose nucleic acid. Nature *171* (1953) 737–738.
34 Yi, Y., Omer-Ali, P., Kelly, C., Simpson, A. J. G., and Smithers, S. R., IgM antibodies recognizing carbohydrate epitopes shared between miracidia and schistosomula of *Schistosoma mansoni* block in vitro killing. J. Immunol. *137* (1986) 3946–3954.

0014-4754/91/020115-04$1.50 + 0.20/0

mRNA processing in the Trypanosomatidae

K. Perry and N. Agabian

Intercampus Program, Molecular Parasitology, University of California, San Francisco, and University of California, Berkeley, Laurel Heights Campus, San Francisco (California 94143-1204, USA)

Summary. Members of the Trypanosomatidae, which include the African trypanosomes, the American trypanosomes and the leishmanias, cause disease in vast proportions in man and his livestock and are a major detrimental factor to the social and economic well-being of the third world. Current research using the techniques of molecular biology has revealed two unusual types of mRNA processing in these protozoans; these are the addition of a shared leader sequence to the 5′ ends of nuclear mRNAs by a mechanism of *trans* splicing, and the insertion and deletion of specific uridine residues in mitochondrial transcripts by RNA editing. The presence of these two mRNA processing pathways in the Trypanosomatidae has profound consequences for the organization and expression of their genetic information.

Key words. Trans splicing; spliced leader addition; RNA editing.

The Trypanosomatidae are a family of protozoan parasites which have long fascinated biologists because of their unique cellular properties and their ability to survive and proliferate in antagonistic host environments. Members of this family, which branched very early from the main eukaryotic lineage, have adapted to hosts as divergent as amphibians, birds, fish, insects, mammals, and even plants. Many are digenetic parasites of both invertebrates and vertebrates and have complex life cycles exhibiting both extracellular and intracellular forms. In recent years the Trypanosomatidae have come under the scrutiny of molecular biologists, who, in their attempts to better understand these parasites, have revealed startling facts about the organization and expression of their genetic information. This review will focus on an aspect of trypanosomatid gene expression that has presented new challenges to the basic dogma of how eukaryotic genes are transcribed and expressed – that of mRNA processing.

The majority of eukaryotic mRNAs undergo multiple rounds of modification prior to their transport to the cytoplasm and subsequent translation. The processing events generally include 5′ capping, methylation of both 5′ and internal nucleotides, removal of intervening sequences and addition of poly(A) tails. In the trypanosomatids, two other types of mRNA modification have now been identified; these are the addition of a 5′ leader sequence to nuclear mRNAs by *trans* splicing[2] and the insertion of internal U residues in mitochondrial mRNAs by a process that has been termed 'RNA editing'[71]. Although much has been done to characterize and understand these processes, both remain puzzling as to their origin and function in the cell.

Trans splicing of nuclear mRNAs

The experimental studies which eventually led to the discovery of *trans* splicing addressed the mechanism of antigenic variation in the African trypanosome, *T. brucei*. Several investigators observed that the mRNAs encoding variant surface glycoproteins (VSGs) began with a common 5′ leader sequence that is not present in genomic clones of the genes[9,67]. The leader sequence, which has been variously termed the 'spliced leader' or 'mini-exon' because it is not encoded contiguously with structural gene sequences, has 39 nucleotides and contains a highly modified cap structure[24,51]. Although early speculations held that this 5′ leader was specific to VSG gene expression, it was subsequently shown that every nuclear mRNA in *T. brucei* begins with this sequence[18,48].

Southern analyses indicate that the mRNA spliced leader (SL) is encoded within a 1.4 kb unit which is present in approximately 200 copies per trypanosome genome[17,36,44]. These SL genes are not scattered throughout the genome in a manner which would suggest that each is associated with one or a few of the trypanosome protein coding genes; rather, they are organized in a few large tandem arrays. The organization of the 1.4 kb units rules out the possibility that the spliced leader and individual structural genes are transcribed into continuous pre-mRNAs. This hypothesis is most rigorously supported by molecular karyotype determinations in which the 1.4 kb SL genes and VSG genes are mapped to chromosomes that have been resolved by pulse field gradient gel electrophoresis. These determinations reveal that VSG genes which reside on chromosomes with no detectable copies of the 1.4 kb SL gene are transcribed and give rise to mRNAs which contain a 5′ leader[2,25,68]. Therefore, transcription of these genes occurs independently and in a manner which permits the joining of RNAs that are encoded by genes on separate chromosomes. The SL gene transcript has been identified in Northern analyses of total *T. brucei* RNA[13,30,37]; the transcript is a small RNA with a discrete size of approximately 135-147 nucleotides. This RNA (the SL RNA or medRNA) is encoded continuously within the 1.4 kb repeat unit and contains the 39 base sequence at its 5′ end[13,30,37].

Mechanism of spliced leader addition

The leader sequence is donated from the SL RNA to pre-mRNAs by a process of *trans* splicing[43,60]; this reaction is shown schematically in figure 1. The mechanism of *trans* splicing in trypanosomes appears to be similar biochemically to that of *cis* splicing in other eukaryotes insofar as the reaction precursors and intermediates have been described. The 5′ splice site of the SL RNA and 3′ splice sites of pre-mRNAs fit the consensus sequences (5′/GU3′ and 5′AG/3′ respectively) for *cis*-spliced sites and the branch residue is primarily an adenine[49,61] as is the case for *cis* splicing. On the basis of these descriptions, it is reasonable to speculate that, by analogy with the well-characterized *cis*-splicing pathway, in the first step of the reaction a branchsite residue in the pre-mR-

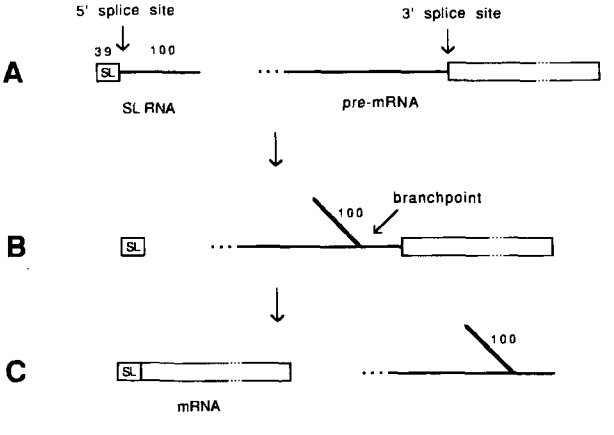

Figure 1. Model of *trans* splicing in *Trypanosoma brucei*. Substrates for the reaction are the 139 nt SL RNA and a pre-mRNA (A). The first step involves cleavage of the 5′ splice site and formation of a branched Y structure (B). In the second step of the reaction, the two exons are joined to produce a mature mRNA and the introns are released as a discrete species (C).

NA participates in a nucleophilic attack on the 5′ splice site, leading to cleavage of the splice site and the generation of a 2′–5′ branched intermediate (fig. 1 B). Because the two RNA substrates for *trans* splicing are not linked as they are in *cis* splicing, this branched intermediate has the shape of a Y structure rather than a lariat. The existence of this structure was demonstrated by Murphy et al.[43] and by Sutton and Boothroyd[60] who showed that the intron portion of the SL RNA (the 3′ 100 nucleotides) is associated by a 2′–5′ phosphodiester bond with high molecular weight RNA and is released by treatment with HeLa cell debranching enzyme. Because the intron portion of the SL RNA is released from this RNA as a discrete sized species, it is clearly not covalently linked at its 3′ end to the remainder of the intron as would be expected for a *cis*-spliced lariat.

Resolution of the branched intermediate occurs when the 3′ splice site is cleaved and the two exons ligated (fig. 1 C). The products of *trans* splicing are therefore a mature mRNA and a branched intron-structure that contains the 3′ 100 nucleotides of the SL RNA bound to the intron portion of the pre-mRNA. The intron regions of different pre-mRNA species appear to differ in size as well as in sequence and thus branched intron-structures are found in the cell as a heterogeneous-sized population of RNAs[52]. These structures are presumably substrates for the debranching enzyme activity that has been identified in trypanosomes[61], and are likely to be rapidly degraded.

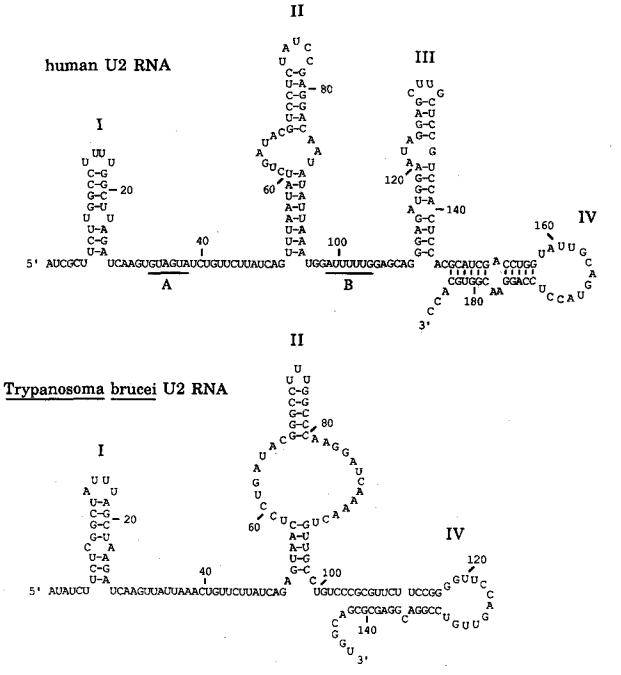

Figure 2. Potential stem-loop structures I, II and IV are conserved between the *Trypanosoma brucei* and human U2 RNAs. Structures are drawn as suggested by Shuster and Guthrie[56]. The underlined regions A and B of the human U2 RNA indicate the branchpoint recognition site and Sm binding site respectively. These regions are not conserved in the trypanosome U2.

The machinery of trans splicing

Although the exact details of *trans* splicing remain unknown, there are indications that the cellular machinery that carries out the reaction is related to the *cis*-splicing machinery of yeast and higher eukaryotes. *Cis* splicing takes place in the context of a large RNA-protein complex termed the 'spliceosome'[8, 34, 54]. The spliceosome assembles on the nascent transcript and is composed of several distinct ribonucleoprotein particles (RNPs); each of these particles appears to be required during a specific step or stage of the splicing pathway. In addition to their protein components, individual particles contain one or two of the small nuclear class of U RNAs. These U snRNAs have several characteristic features, including a trimethylguanosine cap structure and a high content of uridine residues (hence the name U RNA).

Five U RNAs are required for splicing in yeast and metazoan cells; three of their analogs have been identified in trypanosomes[40, 64, 65]. These include the 148 nt trypanosome U2 RNA analog, which shows extensive homology to other U2 RNAs over the first 100 nt (fig. 2). This '5′ domain' is highly conserved and is thought to contain sequence or structural determinants that are required for the function of the U2 RNP in pre-mRNA processing[8, 34, 47, 54]. The 3′ region of the trypanosome U2 RNA is far less conserved, and potential secondary structure comparisons suggest that it is lacking an internal single-stranded region and a stem-loop structure (stem-loop III) that is present in other U2 RNAs that have been studied (fig. 2)[56]. Interestingly, it is this single-stranded region of the U2 RNAs of *S. cerevisiae* and higher eukaryotes that contains the Sm binding site which is required for assembly of the RNA into an RNP particle. This site is present in many of the sn URNAs involved in *cis* splicing, is required for binding of a set of proteins (the Sm antigens) that are common to the various particles containing these URNAs, and is necessary for these RNAs to acquire their characteristic TMG cap structure. Interestingly, none of the trypanosome U RNAs that have been identified contain an Sm binding sequence[40, 64, 65]; also, *T. brucei* does not appear to contain proteins that react with anti-Sm antibodies (S. Michaeli, unpublished). The trypanosome URNAs do appear to be bound by protein(s), however, as they migrate in Cs_2SO_4 density gradients as RNP particles. As trypanosomes are phylogenetically far removed from other organisms whose sn URNPs have been examined, it will be interesting from an evolutionary point of view to compare the proteins of the trypanosome URNP particles to those of mammalian URNPs.

Trypanosome U4 and U6 RNAs have been identified and were found to be 57% and 69%, respectively, identical to their *Drosophila* analogs[40, 64]. The *T. brucei* U4 and U6 RNA sequences show extensive intermolecular complementarity, consistent with the hypothesis put forth for higher eukaryotes that U4 and U6 RNAs are complexed

by base-pairing within a single RNP particle[10, 26]. Although the precise function of these RNAs in mRNA splicing is not known, they may interact with the 5′ exon, perhaps holding it in place in the spliceosome once the 5′ splice site has been cleaved.

The hypothesis that *trans* splicing takes place within a large ribonucleoprotein complex such as that described for *cis* splicing is supported by results obtained by Miller and Wirth[39]. They showed by density centrifugation of nuclear extracts of a related trypanosomatid, *Leishmania enriettii*, that SL RNA is present in a 60S peak and that this peak contains protein as well as RNA. The *L. enriettii* U2 RNA cosediments with the SL RNA in this 60S fraction, and may be part of a structure analogous to the spliceosome described in *cis*-splicing systems[8, 34, 54]. Definitive proof that *trans* splicing is carried out by URNP complexes will require the development of an in vitro system; however, in the current absence of such a system these in vivo results have provided working models that may be tested in the future.

While the similarities that have been noted between *trans* and *cis* splicing are likely to reflect a common evolutionary origin, the differences between the two pathways are, perhaps, far more intriguing. Differences are to be expected, because the *cis*-splicing machinery of mammalian cells is very inefficient at *trans* splicing artificial substrates in vitro[29, 58]. The efficiency of the reaction is increased by the introduction of complementarity between the artificial substrates[29], suggesting that the mammalian *cis*-spliceosome does not require that its substrates be continuous, but that it has no inherent mechanism for efficiently bringing together two discontinuous substrates. Such a mechanism is likely to exist in trypanosomes because no extensive complementarity between the SL RNA and mRNA precursors has been demonstrated, and yet *trans* splicing appears to occur very rapidly in these organisms.

Perhaps the most significant dissimilarity between the two types of splicing is that, for *trans* splicing, every 5′ exon is identical. This property may be necessary and may provide for a mechanism which allows the efficient recognition of discontinuous substrates. Recognition could be promoted either by the interaction of the SL RNA with specific proteins, or by base-pairing of the SL RNA with other RNAs that are involved in splicing. Michaeli et al. have obtained evidence for the former type of interaction; they find that the SL RNA is present as part of a unique ribonucleoprotein particle with a sedimentation coefficient of approximately 10S[3, 35a]. This particle, termed the SL RNP, has several properties of a U snRNP, including its density and its ability to withstand fractionation by Cs_2SO_4. These observations suggest that during *trans* splicing the 5′ exon-containing RNA in some respects resembles a U RNA.

The hypothesis that the SL RNA has properties of a U RNA has been expanded by Bruzik et al., who showed that the SL RNAs from several different species can be folded into very similar secondary structures even though their primary sequences are not conserved[11]. Each potential secondary structure contains three stem-loops and in each the 5′ splice site is base-paired with upstream exon sequences. For most types of RNA splicing, base-pairing across the 5′ site is thought to activate the phosphodiester bond for the first step of the reaction. During *cis* splicing, it is the U1 RNA that base-pairs with the pre-mRNA 5′ splice site; in *trans* splicing the potential for intramolecular base-pairing in the SL RNA may obviate a need for an additional U1 RNA activity. In this regard, it is interesting that no U1 RNA has been identified in trypanosomes[40]. Thus, the SL RNA may play a bifunctional role in the cell in that it may act both as a substrate for *trans* splicing and also to promote its own catalysis.

Branchpoint recognition in trypanosomes

The potential for the SL RNA to interact directly by base-pairing with other RNAs involved in *trans* splicing has been suggested by RNA sequence comparisons. In general, the models that have been proposed have implicated the SL RNA as having a mechanistic role early in the formation of the *trans*-splicing complex and, in particular, in the recognition of the branchsite residue. One of the more striking observations to come from these RNA sequence comparisons is the complementarity between the SL RNA and the trypanosome U2 RNA[40, 65]. In *T. brucei brucei* and *T. brucei gambiense* this complementarity includes a large block of potential uninterrupted base-pairs between the central portion of the SL and a 10 nucleotide region of the U2 RNA. The notion that the SL RNA and U2 RNA interact directly is attractive because it suggests a model whereby during *trans* splicing, the U2 RNA, acting as a 'molecular adaptor', brings the 5′ splice site into contact with the branchpoint region. However, the model is not strongly supported by the analyses of SL RNA and U2 RNA sequences in other trypanosomatids. In *Leishmania enriettii* and *Leptomonas collosoma* the complementarity between the SL RNA and U2 RNA is not as extensive as for the *T. brucei brucei* and *T. brucei gambiense* RNAs and the largest block of potential base-pairs is interrupted[25a, 39].

A second model for SL RNA base-pairing interactions has been proposed by Layden and Eisen[33]. They suggest that limited regions of complementarity allow the SL RNA to interact directly with pre-mRNAs at the position of the branchsite. This model is not supported by recent branchpoint mapping experiments, however, which indicate that for the α- and β-tubulin pre-mRNAs, the branched residues and the predicted SL RNA complementarity sites are not coaligned[49].

The question as to the mechanism of branchpoint recognition is particularly puzzling in *T. brucei* because regions adjacent to SL addition sites have none of the conserved features that are present upstream of the 3′ splice sites of

cis-spliced pre-mRNAs. These features include, for *S. cerevisiae*, a highly conserved sequence (UACUAAC) that is complementary to nucleotides 34–39 of the U2 RNA. Base pairing between this region of the U2 RNA, which is absolutely conserved among the higher eukaryotes that have been studied, and pre-mRNAs appears to be partly responsible for branchpoint recognition in yeast[47]. Interestingly, the analogous region of the trypanosome U2 RNA differs from the consensus sequence in five out of six positions. There is no potential for interaction between this region of the U2 RNA and the tubulin branchsites that have been identified[49], and the region is not conserved among the U2 RNAs of other trypanosomes that have been studied, suggesting that there are no functional constraints placed upon it[25a, 39]. An additional hypothesis for branchpoint recognition has been suggested for mammalian cells, which do not contain highly conserved UACUAAC regions; this hypothesis suggests that polypyrimidine tracts located near the 3′ splice site may be required for the initial interaction of the pre-mRNA with the U2 RNA[14, 31, 53]. An examination of trypanosome sequences reveals that many genes do have pyrimidine-rich regions upstream of their 3′ splice sites; however, these polypyrimidine tracts are not conserved among all the genes that have been studied.

Our inability to identify conserved regions of pre-mRNAs that may act as recognition sequences for the *trans*-splicing machinery suggests two hypotheses, each of which would distinguish *trans* splicing in trypanosomes from *cis* splicing. A first hypothesis is that the *trans*-splicing machinery may recognize sequences that are removed from the 3′ splice site; this could be either a direct recognition or a recognition through other factors. For example, *trans* splicing may be coupled to another process such as transcription and may identify the appropriate 3′ splice sites only through its interaction with the transcription complex. A second hypothesis is that recognition sequences near the 3′ splice sites of trypanosome pre-mRNAs do exist but are not well conserved because the fidelity of *trans* splicing need not be as high as the fidelity of *cis* splicing of protein coding regions. SL addition occurs in the nontranslated regions of trypanosome pre-mRNAs and there is not a need for extreme precision during the reaction in order to maintain open reading frames. This hypothesis is supported by the fact that many trypanosome genes have multiple potential SL addition sites.

Functional significance of trans splicing

While much understanding has been gained of the mechanism of *trans* splicing, its function in the cell remains unclear. Analogs of the SL RNA have been found in every trypanosomatid that has been studied, indicating that *trans* splicing evolved before the divergence of this large family[5, 16, 37, 38, 42, 45, 65]. In figure 3 is shown a

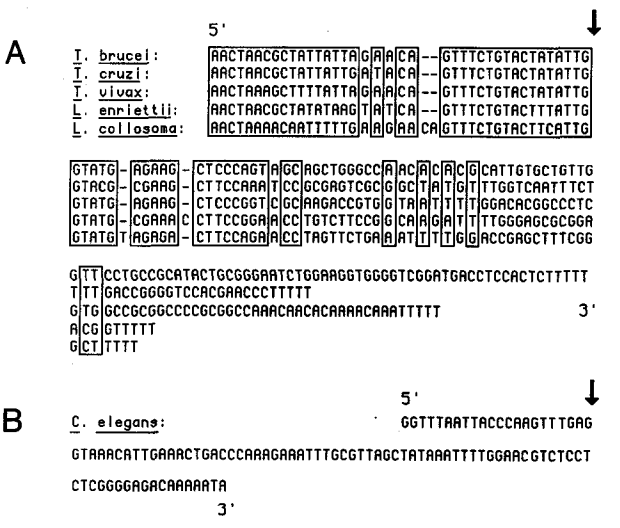

Figure 3. *A* Sequence comparison of the SL RNAs of several trypanosomatid species. Regions that are conserved between the species are boxed; dashes have been inserted to preserve the alignment and the 5′ splice site is indicated by an arrow. *B* Nucleotide sequence of the *C. elegans* SL RNA showing the position of the 5′ splice site (arrow).

comparison of the SL RNAs that have been sequenced from different species including *T. brucei*[16, 37], *T. cruzi*[16, 37], *T. gambiense*[65], *T. vivax*[16], *Crithidia fasciculata*[42], *Leishmania enriettii*[38], *Leptomonas collosoma*[16] and *Leptomonas seymouri*[5]. A comparison of these sequences reveals that the spliced leader and a few of the nucleotides immediately downstream of the splice junction are highly conserved. The conserved nucleotides downstream of the splice site may be involved in recognition of the SL RNA as a substrate for *trans* splicing; for example, they may provide binding sites for the proteins of the SL RNP. Because the conserved nucleotides of the spliced leader are present on both the SL RNA and on mature mRNAs, they may be required for *trans* splicing and/or for other aspects of mRNA metabolism such as transport from the nucleus, translation and stability.

Trans splicing may have arisen as a consequence of the unusual organization and expression of the trypanosome genome. What little data has been obtained has indicated that patterns of transcription in trypanosomes may be very different from those of other eukaryotes. Perhaps the most unusual feature that has been suggested is that trypanosome precursor transcripts are polycistronic. Polycistronic transcripts have been detected for several housekeeping genes; these genes are reiterated in tandem arrays in the trypanosome genome[41, 66]. There is also evidence to suggest that transcription of genes which are not tandemly repeated may be polycistronic. This has been suggested for active VSG genes, all of which are located near the ends of telomeres and which may be transcribed as part of huge units which give rise to many poly(A)$^+$ RNAs[28].

A function of *trans* splicing may be to resolve these polycistronic transcripts into mature, monocistronic

mRNAs. Spliced leader addition provides a means for cleaving these precursor RNAs within the 5′ nontranslated regions of the genes. Resolution of the 3′ ends of the genes is then likely to occur by cleavage and addition of a poly(A) tail. An additional function of *trans* splicing may be to provide 5′ cap structures to the mature mRNAs. The SL RNA has a highly modified 5′ end including a m^7G cap which it donates to the mRNAs during *trans* splicing[24, 51]. In other eukaryotes, m^7G caps increase the stability of mRNAs and are necessary for their recognition by the translation machinery.

Trans splicing in nematodes

The mechanism and function of spliced leader addition has garnered much interest with the more recent observation that the process is not restricted to the order Kinetoplastidae. Several species of nematodes, both parasitic and free-living, are now known to contain *trans*-spliced mRNAs[4, 32, 46, 62]. For *Caenorhabditis elegans*, the approximate fraction of the total cellular mRNAs that contain an identical, 22 nucleotide spliced leader is 10–15%[4]. Several mRNAs which contain the SL have been identified. These encode two ribosomal proteins, ubiquitin, GAPDH, a heat shock protein and three actins[4, 32]. There is no obvious argument as to why specific mRNAs in *C. elegans* should receive a leader sequence and others not.

The *C. elegans* leader sequence is donated by a 100 nt SL RNA which is encoded within the 5S RNA repeat unit[32]. The SL RNA is transcribed independently of the 5S RNA and, in *Ascaris suum*, the transcribing polymerase appears to be RNA polymerase II[35]. The sequence of the nematode SL RNA, also shown in figure 3, shares little homology with the trypanosome SL RNA. However, the 22 nt SL is conserved among nematodes of several genera, including *C. briggsae*[4] and *Panagrellus redivivus* (other free-living nematodes)[4], *Ancylostoma caninum* (a parasitic nematode of dogs; R. Swoboda, unpublished), Anisakis spp. (parasitic nematodes of fish)[4], *Ascaris suum* (a parasitic nematode of pigs)[4, 46], *Brugia malayi* (the causative agent of elephantiasis in humans)[62], *Haemonchus contortus* (a parasitic nematode of sheep)[4] and *Trichinella spiralis* (the causative agent of trichinosis; R. Swoboda, unpublished). An additional, conserved feature of the various nematode SL RNAs, one that distinguishes them from their trypanosome counterparts, is the presence of an Sm antigen binding site. As mentioned above, this site is present in many of the small nuclear U RNAs involved in *cis* splicing and is required for the binding of a set of proteins (the Sm antigens) that are common to the various particles containing these U RNAs[8, 34]. As a consequence of their having an Sm binding site, nematode SL RNAs are precipitable with anti-Sm antibodies and have a trimethylguanosine cap structure[63]. Thus, in nematodes, the SL RNA resembles a U RNA even more then does the trypanosome SL RNA.

Whether the differences between *trans* splicing in trypanosomes and *trans* splicing in nematodes are merely a reflection of evolutionary distance or whether they reflect inherent differences between the two systems is unknown. The analysis of *trans* splicing in nematodes is complicated by the fact that *cis* splicing of mRNAs also occurs in these organisms and the same transcript can be a substrate for both pathways. This raises the question as to how the machinery distinguishes between sites that should be *trans* spliced and sites that should be *cis* spliced. It will be interesting to determine whether *trans*- and *cis* splicing in nematodes are completely distinct pathways or whether they share common components.

Editing of mitochondrial mRNAs

When *trans* splicing in vivo was first discovered in the trypanosomes, the observation was unprecedented; nevertheless, the process has been somewhat tractable from the experimental point of view because of its similarity to *cis* splicing. A second phenomenon of mRNA processing that has been described in these parasites is that of RNA editing. The origin and mechanism of this process are far more perplexing than those of *trans* splicing because no related pathways have been described in other organisms.

Editing of mRNAs has thus far been observed with a number of mitochondrial transcripts from three distantly related trypanosomatids: *T. brucei, Crithidia fasciculata* (an insect trypanosome) and *Leishmania tarentolae* (a parasite of lizards)[1, 6, 7, 20, 22, 23, 55, 69]. The outcome of the editing process is evident upon comparison of the sequences of the mature transcripts with their corresponding coding sequences in the mitochondrial genomes; specific U residues that are not encoded in the genome are inserted into transcripts and other U residues that are encoded in the genome are deleted from the transcripts. Addition/ deletion of U residues occurs either singly or in multiples and, depending on the transcript being studied, editing may affect both coding sequences and noncoding sequences, including poly(A) tails.

The extent to which editing occurs varies widely among different mitochondrial transcripts; for example, in *T. brucei* the cytochrome oxidase I (COI), NADH dehydrogenase 4 (ND4) and ND5 transcripts appear to remain unedited[55] while transcripts such as those encoding apocytochrome b (CYb) receive a limited number of uridine insertions[21, 22] and, at the opposite end of the spectrum, the 3′ end of the cytochrome c oxidase subunit III (COIII) transcript receives an astonishing 58% of its total nucleotides by the RNA editing process[20]. The extent to which editing occurs also differs among analogous transcripts from the three species that have been studied. This is most clearly demonstrated by the dramatic contrast between the COIII mRNA in *T. brucei*, and the *C. fasciculata* and *L. tarentolae* analogs, which

are edited only to a limited extent in their 5′ regions[20]. The discovery of RNA editing has provided explanations for many of the unusual features of trypanosomatid mitochondrial genomes. In addition to the observation that several normally conserved genes such as those encoding subunits 6, 8 and 9 of ATP synthetase are absent from these genomes, is the puzzling observation that, of the genes that are present, several lack ATG initiation codons and the COII, MURF2, and MURF3 genes encode internal reading frameshifts[15,57]. As will be discussed below, RNA editing corrects many of these genomic errors with the result that fully functional messages are produced.

Editing was first observed in the COII transcripts of *C. fasciculata* and *T. brucei*[7]. The COII gene is peculiar in the trypanosomes in that it harbors a −1 frameshift at amino acid 170 in each of the species studied, yet other characteristics of the gene suggest that it is functional[15,27,50]. These include the fact that the predicted amino acid sequence of the COII gene product is highly conserved, both in the region in which the frameshifts occur and in the remainder of the gene, and the nucleotide substitutions that have arisen occur predominantly at the third base position and are primarily silent or conservative. Benne et al.[7] showed that the frameshift in the genomic COII sequence is suppressed by an addition to the transcript of four uridines. The uridines are added at the same positions in *C. fasciculata* and *T. brucei* and the amino acid sequences predicted for the corrected region of the mRNAs are identical. Similar frameshifts are present in the MURF3 (mitochondrial unidentified reading frame 3) gene of *L. tarentolae* and, in this case, the open reading frame is restored by the insertion of 5 U residues into the transcript[55].

Creation of in-frame initiation codons by RNA editing has been demonstrated for the CYb, MURF2 and COIII transcripts[22,55]. In general, these transcripts are heavily edited in the 5′ coding sequences. For example, for the MURF2 transcripts, 28 uridines are added to and 4 deleted from a region that spans only 24 nucleotides of the *L. tarentolae* coding sequence, 28 uridines are added to a region that spans 18 nucleotides of the *C. fasciculata* coding sequence, and 26 nucleotides are added to and 4 deleted from a region that spans 30 nucleotides of the *T. brucei* coding sequence (fig. 4)[55]. Many of the editing changes are conserved between the three species and, as a result, the amino acid sequences predicted for these portions of the edited transcripts are identical. Editing of the *T. brucei* CYb transcript was unexpected because an in-frame AUG is coded for in the genome[21]. However, RNA editing creates a second AUG at the same position as the AUGs created in the *L. tarentolae* and *C. fasciculata* transcripts[22]. It is possible that, by regulating RNA editing in *T. brucei*, both AUG codons may be utilized, giving rise to two CYb proteins which differ by 20 amino acids at the N-terminus.

Evidence for regulation of RNA editing

Regulation of RNA editing has been demonstrated during the developmental stages of *T. brucei*[21,23]. This digenetic species requires both a mammalian host and an insect vector to complete its differentiative life cycle[19,70]. In the mammalian host, the trypanosome lives primarily in interstitial fluids including blood, and produces energy by glycolysis. Two stages of the parasite cycle occur in the vertebrate host; in the slender form the single mitochondrion is reduced, with few cristae present, and the cell completely lacks Krebs cycle enzymes and cytochromes. The stumpy form has an incomplete respiratory system with some enzyme activities associated with the Krebs cycle, but it also produces energy solely by glycolysis. It

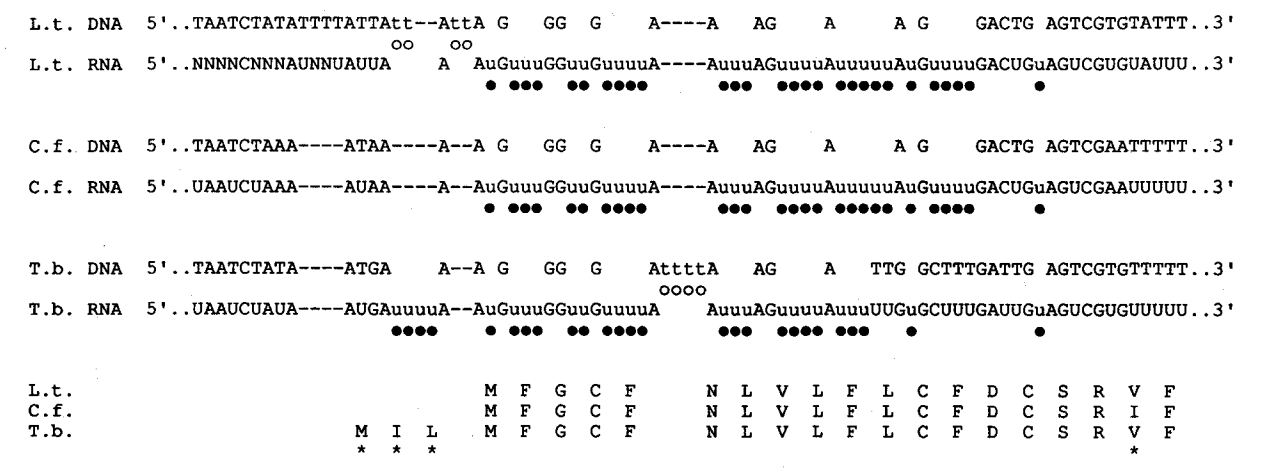

Figure 4. Comparison of the MURF2 DNA, RNA, and amino acid sequences from *L. tarentolae*, *C. fasciculata*, and *T. brucei*. A portion of the 5′ end sequences of the MURF2 transcripts are shown with the corresponding genomic DNA sequences aligned above each RNA sequence. Sites of uridine addition are indicated by gaps in the DNA sequences and by closed circles below the RNA sequences. Sites of uridine deletion are indicated by gaps in the RNA sequences and open circles below the DNA sequence. Dotted lines indicate gaps introduced to facilitate the alignment of sequences between species. The N-terminal amino acid sequences predicted from the edited transcripts are shown below. Asterisks indicate positions of non-identity between these predicted sequences.

is the stumpy form which is thought to be infective to the tsetse fly vector. Upon ingestion into the midgut of this insect host, the trypanosome develops into the procyclic form, which has a fully functional respiratory system.

An analysis of mitochondrial transcripts from these various life cycle stages suggests that, for specific transcripts, there is a stage specificity of uridine addition that reflects the stage specificity of synthesis of the encoded proteins. For example, for the CYb and COII genes, edited transcripts are fully present in procyclic forms, they are present to a limited extent in stumpies, and they are absent from slender bloodstream forms[21, 23]. Unedited transcripts are found in all three life cycle stages although their abundance may differ between the stages. These observations have suggested that a potential role of RNA editing in trypanosomes is to regulate the translation of mitochondrial proteins. In this respect, it is of note that of the six genes whose transcripts are known to be edited, five lack appropriate initiation ATG codons. Moreover, editing in the 5′ initiation region of these transcripts is generally well conserved among the species studied whereas editing of other regions of the transcripts, for example in the coding region of the COIII mRNA, is not conserved. However, a model in which the role of editing is simply to control mitochondrial mRNA translation does not begin to address questions as to the origins of such extensive editing as is found in the *T. brucei* COIII transcript[20].

Mechanism of RNA editing

The mechanism of RNA editing has not yet been determined but it appears likely that the process occurs post-transcriptionally. This hypothesis is primarily supported by the detection of unedited and partially edited transcripts of mitochondrial genes, although it has not been proven that these transcripts are actual precursors of mature, edited mRNAs. Unedited transcripts have been detected in RNA sequencing experiments[21, 55] and in Northern analyses using specific oligonucleotides as probes[7, 21, 22, 69]. Partially edited COIII transcripts have been detected in Polymerase Chain Reaction experiments in which one of the two primers used was specific for an edited region of the transcript and the other primer was specific for an unedited region of the transcript[1]. Depending on the set of primers used, these experiments can theoretically detect transcripts that are edited only in a 5′ and not in a 3′ region or transcripts that are edited only in a 3′ and not in a 5′ region. Only the latter class of partially edited mRNAs have been detected in these experiments, suggesting that the process of RNA editing occurs in a 3′ to 5′ direction. This hypothesis is also supported by sequence analyses of cDNA clones of several partially edited transcripts; these analyses reveal that the transcripts are invariably edited at their 5′ ends and not at their 3′ ends[1, 69].

The unedited/edited junctions of the partially edited COIII transcripts are interesting in that they are not sharp transitions from the sequences encoded by the genome to the sequences of the mature mRNAs. Rather, these junctions are regions which may differ both from the gene and the fully edited RNA sequences at several sites. At some of the sites in the unedited/edited junctions, fewer uridines may be present than the number added to the mature transcript. At other sites, a greater number of uridines may be present than in the mature transcript, and the partially edited transcripts may contain uridine additions at sites that are not edited at all in the mature transcript. These junctions may represent regions where editing was in process at the time of RNA isolation.

Taken together, these observations suggest that the process of RNA editing involves several enzymatic activities including those of an endonuclease, a uridyl transferase and an RNA ligase. A general model for RNA editing which describes these activities is shown in figure 5 and is based on a model proposed by Stuart[59]. For the sake of simplicity, the RNA editing activities in the model have been attributed to a single complex termed the 'editosome', although it is not known whether such a complex actually exists in the cell. Regardless of the nature of the editing machinery, the sequence of events that is proposed for the RNA editing process is: assembly of the editosome on the unedited transcript, recognition of a region of the transcript to be edited, cleavage of a site or sites within that region, addition or deletion of uridines at the site(s), and ligation.

If the hypothesis is correct that editing occurs in a 3′ to 5′ direction, then assembly of the editosome may take place either at the 3′ end of the transcript or at the 3′-most region to be edited. For many transcripts, these

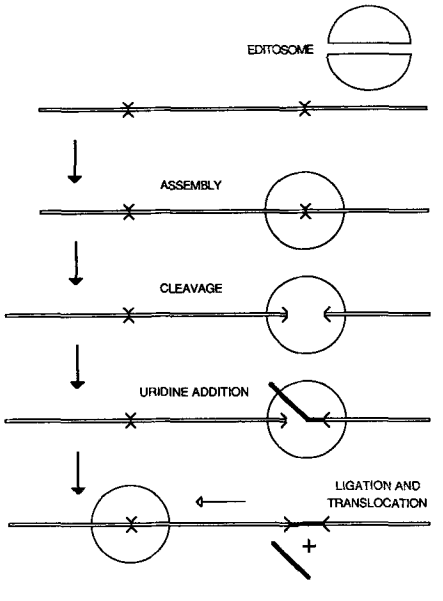

Figure 5. Model for RNA editing proposed by Stuart[59]. RNA is shown as horizontal bars, added uridines are indicated by black bars and sites to be edited are marked with crosses (X).

two possibilities cannot be distinguished because RNA editing appears to affect the poly(A) tails[6,7,12,20,69]. Although it has been suggested that insertion of uridines into poly(A) tails functions to stabilize the mRNAs, it is equally possible that the editing of these regions is merely a reflection of the sequence recognition requirements of the editosome. That is to say, the primary sequence requirements that direct the editing machinery to a particular region may be contained within the tracts of A residues in the poly(A) tails. Once the editing machinery is assembled, it may begin to insert and/or delete U residues regardless of whether editing is required in the region. The hypothesis that poly(A) tails contain recognition sequences for RNA editing is supported by the observation that mitochondrial genome sequences that are targets for editing are rich in purines. One example of this is apparent upon comparison of the edited and unedited sequences of the COIII gene of *T. brucei*[20]. The unedited, genomic sequence contains runs of purines punctuated by T residues and occasional C residues. In some cases, as few as four or five purines separate T residues in the sequence, suggesting that these limited purine tracts may be all that is required for initial recognition by the editing machinery.

Once the editosome is assembled, the endonuclease and uridyl transferase activities appear to be somewhat random. This is suggested by analyses of several partially edited COIII transcripts[1]. Abraham et al. found several cases in which too many or too few U residues had been added to sites that require RNA editing, and they found one case in which two U's were added to a site that is not edited at all in the mature mRNA. In addition to these full-length, partially edited mRNAs, several transcripts have been identified in *T. brucei* that terminate abruptly at an editing site with a long, 5′ stretch of U's[1,21]. In one example, the 5′ end of a partial CYb transcript has a run of 36 U's although the edited transcript has only 8 uridines at that position[21]. If these transcripts are true intermediates in the editing process, then it appears that in *T. brucei* the uridyl transferase activity initially adds long tracts of uridines which are subsequently chewed back, and that several rounds of editing may be required to obtain the correct number of uridine additions in the mature mRNA.

Because of the great diversity of edited sites, it is unlikely that the information as to the correct sequences of these transcripts is either inherent to the editosome or is present in the primary structure of the unedited RNAs. It is conceivable that secondary structure characteristics of the regions direct the editing machinery although an analysis of edited regions did not reveal any structures that might function in this manner[55]. An alternative hypothesis is that the editosome is working with a template. Such a template is unlikely to be DNA because no nuclear or mitochondrial sequences corresponding to the edited COII, COIII and CYb transcripts of *T. brucei* have been identified[7,20,21,55,69]. It is also unlikely that the editing process is directed by an anti-sense RNA template[22]. A more intriguing possibility is that a plus-strand RNA template directs editing by a means other than traditional Watson-Crick base-pairing. Plus-strand templates could simply be identical to mature, edited mRNAs; these have been detected in the cell at low levels under all conditions that have been studied.

Regardless of its mechanism, RNA editing is a remarkable phenomenon with consequences for molecular evolution that we can only begin to understand. The identification of the *T. brucei* COIII gene, and its comparison to the *C. fasiculata* and *L. tarentolae* analogs, has provided us with a spectacular example of the plasticity that the process has conferred upon the trypanosomatid mitochondrial genome. Our knowledge of RNA editing has explained several puzzling aspects of the trypanosomatid mitochondrial genome and may equip us to tackle other difficult questions of trypanosome biology in the future.

1 Abraham, J. M., Feagin, J. E., and Stuart, K., Characterization of cytochrome c oxidase III transcripts that are edited only in the 3' region. Cell *55* (1988) 267–272.

1a Agabian, N., *Trans* splicing of nuclear pre-mRNAs. Cell *61* (1990) 1157–1160.

2 Agabian, N., Nelson, R. G., Parsons, M., Milhausen, M., Sather, S., Selkirk, M., Watkins, K. P., Myler, P., Rothwell, V., and Stuart, K., Antigenic variation and the role of the spliced leader sequence in trypanosomatid gene expression, in: Genome Rearrangement, pp. 153–172. Alan R. Liss, Inc., New York 1985.

3 Agabian, N., Perry, K. L., and Murphy, W. J., RNA joining and trypanosome gene expression, in: Molecular Biology of RNA: New Perspectives, pp. 151–171. Academic Press, Inc., San Diego 1987.

4 Bektesh, S., Van Doren, K., and Hirsh, D., Presence of the *Caenorhabditis elegans* spliced leader on different mRNAs and in different genera of nematodes. Genes Dev. *2* (1988) 1277–1283.

5 Bellofatto, V., Cooper, R., and Cross, G. A. M., Discontinuous transcription in *Leptomonas seymouri*: presence of intact and interrupted mini-exon gene families. Nuc. Acids Res. *16* (1988) 7437–7456.

6 Benne, R., RNA editing in trypanosome mitochondria. Biochim. biophys. Acta *1007* (1989) 131–139.

7 Benne, R., van den Burg, J., Brankenhoff, J. P. J., Sloof, P., van Boom, J. H., and Tromp, M. C., Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell *46* (1986) 819–826.

8 Birnstiel, M. L., (ed.), Small nuclear ribonucleoprotein particles. Springer-Verlag, Heidelberg 1988.

9 Boothroyd, J. C., and Cross, G. A. M., Transcripts coding for the variant surface glycoproteins of *Trypanosoma brucei* have a short identical exon at their 5′ end. Gene *20* (1982) 281–289.

10 Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., and Lührmann, R., Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. EMBO J. *3* (1984) 1357–1363.

11 Bruzik, J. P., Van Doren, K., Hirsh, D., and Steitz, J. A., *Trans* splicing involves a novel form of small nuclear ribonucleoprotein particles. Nature *335* (1988) 559–562.

12 Campbell, D., Spithill, T., Samaras, T., Simpson, A., and Simpson, L., Apparent uridine addition in the poly A tail of *Leishmania major* ND1 transcript. Molec. biochem. Parasit. *36* (1989) 197–199.

13 Campbell, D. A., Thornton, D. A., and Boothroyd, J. C., Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. Nature *311* (1984) 350–355.

14 Chabot, B., and Steitz, J. A., Multiple interactions between the splicing substrate and small nuclear ribonucleoproteins in spliceosomes. Molec. cell. Biol. *7* (1987) 281–293.

15 De la Cruz, V. F., Neckelmann, N., and Simpson, L., Sequences of six genes and several open reading frames in the kinetoplast maxicircle DNA of *Leishmania tarentolae*. J. biol. Chem. *254* (1984) 15136–15147.

16 De Lange, T., Berkvens, T. M., Veerman, H. J. G., Frasch, A. C. C., Barry, J. D., and Borst, P., Comparison of the genes coding for the

common 5′ terminal sequence of messenger RNAs in three trypanosome species. Nuc. Acids. Res. *12* (1984) 4431–4443.
17 De Lange, T., Liu, A. Y. C., Van der Ploeg, L. H. T., Borst, P., Tromp, M. C., and Van Boom, J. H., Tandem repetition of the 5′ mini-exon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription? Cell *34* (1983) 891–900.
18 De Lange, T., Michels, P. A. M., Veerman, H. J. G., Cornelissen, A. W. C. A., and Borst, P., Many trypanosome mRNAs share a common 5′ terminal sequence. Nuc. Acids Res. *12* (1984) 3777–3790.
19 Englund, P. T., Hajduk, S. L., and Marini, J. C., The molecular biology of trypanosomes. A. Rev. Biochem. *51* (1982) 695–726.
20 Feagin, J. E., Abraham, J. M., and Stuart, K., Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. Cell *53* (1988) 413–422.
21 Feagin, J. E., Jasmer, D. P., and Stuart, K., Developmentally regulated addition of nucleotides within apocytochrome b transcripts in *Trypanosoma brucei*. Cell *49* (1987) 337–345.
22 Feagin, J. E., Shaw, J. M., Simpson, L., and Stuart, K., Creation of AUG initiation codons by addition of uridines within cytochrome b transcripts of kinetoplastids. Proc. natl Acad. Sci. USA *85* (1988) 539–543.
23 Feagin, J. E., and Stuart, K., Developmental aspects of uridine addition within mitochondrial transcripts of *Trypanosoma brucei*. Molec. cell. Biol. *8* (1988) 1259–1265.
24 Freistadt, M. S., Cross, G. A. M., Branch, A. D., and Robertson, H. D., Direct analysis of the mini-exon donor RNA of *Trypanosoma brucei*: detection of a novel cap structure also present in messenger RNA. Nuc. Acids Res. *15* (1987) 9861–9879.
25 Guyaux, M., Cornelissen, A. W. C. A., Pays, E., Steinert, M., and Borst, P., *Trypanosoma brucei*: a surface antigen mRNA is discontinuously transcribed from two distinct chromosomes. EMBO J. *4* (1985) 995–998.
25a Hartshorne, T., and Agabian, N., A new U2 RNA secondary structure provided by phylogenetic analysis of trypanosomatid U2 RNAs. Genes Dev. (1991) in press.
26 Hashimoto, C., and Steitz, J. A., U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. Nuc. Acids Res. *12* (1984) 3283–3293.
27 Hensgens, L. A. M., Brakenhoff, J., De Vries, B. F., Sloof, P., Tromp, M. C., Van Boom, J. H., and Benne, R., The sequence of the gene for cytochrome c oxidase subunit I, a frameshift containing gene for cytochrome c oxidase subunit II and seven unassigned reading frames of *Trypanosoma brucei* mitochondrial maxicircle DNA. Nuc. Acids Res. *12* (1984) 7327–7344.
28 Johnson, P. J., Kooter, J. M., and Borst, P., Inactivation of transcription by UV irradiation of *T. brucei* provides evidence for a multicistronic transcription unit including a VSG gene. Cell *51* (1987) 273–281.
29 Konarska, M. M., Padgett, R. A., and Sharp, P. A., *Trans* splicing of mRNA precursors in vitro. Cell *42* (1985) 165–171.
30 Kooter, J. M., De Lange, T., and Borst, P., Discontinuous synthesis of mRNA in trypanosomes. EMBO J. *3* (1984) 2387–2392.
31 Kramer, A., Analysis of RNase A resistant regions of adenovirus 2 major late precursor-mRNA in splicing extracts reveals an ordered interaction of nuclear components with the substrate RNA. J. molec. Biol. *196* (1987) 559–573.
32 Krause, M., and Hirsh, D., A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. Cell *49* (1987) 753–761.
33 Layden, R. E., and Eisen, H., Alternate *trans* splicing in *Trypanosoma equiperdum*: implications for splice site selection. Molec. cell. Biol. *8* (1988) 1352–1360.
34 Maniatis, T., and Reed, R., The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. Nature (London) *325* (1987) 673–678.
35 Maroney, P. A., Hannon, G. J., and Nilsen, T., Transcription and cap trimethylation of nematode spliced leader RNAs in a cell-free system. Proc. natl Acad. Sci. USA *87* (1990) 709–713.
35a Michaeli, S., Roberts, T. G., Watkins, K. P., and Agabian, N., Isolation of distinct small ribonucleoprotein particles containing the spliced leader and U2 RNAs of *Trypanosoma brucei*. J. biol. Chem. *265* (1990) 10582–10588.
36 Michiels, F., Matthyssens, G., Kronenberger, P., Pays, E., Dero, B., Van Assel, S., Darville, M., Cravador, A., Steinert, M., and Hamers, R., Gene activation and reexpression of a *Trypanosoma brucei* variant surface glycoprotein. EMBO J. *2* (1983) 1185–1192.
37 Milhausen, M., Nelson, R. G., Sather, S., Selkirk, M., and Agabian, N., Identification of a small RNA containing the trypanosome spliced leader: a donor of shared 5′ sequences of trypanosomatid mRNAs? Cell *38* (1984) 721–729.
38 Miller, S. I., Landfear, S. M., and Wirth, D. F., Cloning and characterization of a *Leishmania enriettii* gene encoding a RNA spliced leader sequence. Nuc. Acids Res. *14* (1986) 7341–7360.
39 Miller, S. I., and Wirth, D. F., *Trans* splicing in *Leishmania enriettii* and identification of ribonucleoprotein complexes containing the spliced leader and U2 equivalent RNAs. Molec. cell. Biol. *8* (1988) 2597–2603.
40 Mottram, J., Perry, K. L., Lizardi, P. M., Lührmann, R., Agabian, N., and Nelson, R. G., Isolation and sequence of four small nuclear U RNA genes of *Trypanosoma brucei* subsp. *brucei*: Identification of the U2, U4 and U6 RNA analogs. Molec. cell. Biol. *9* (1989) 1212–1223.
41 Muhich, M. L., and Boothroyd, J. C., Polycistronic transcripts in trypanosomes and their accumulation during heat shock: evidence for a precursor role in mRNA synthesis. Molec. cell. Biol. *8* (1988) 3837–3846.
42 Muhich, M. L., Hughes, D. E., Simpson, A. M., and Simpson, L., The monogenetic kinetoplastid protozoan, *Crithidia fasciculata*, contains a transcriptionally active, multicopy mini-exon sequence. Nuc. Acids Res. *15* (1987) 3141–3153.
43 Murphy, W. J., Watkins, K. P., and Agabian, N., Identification of a novel Y branch structure as an intermediate in Trypanosome mRNA processing: evidence for *trans* splicing. Cell *47* (1986) 517–525.
44 Nelson, R. G., Parsons, M., Barr, P. J., Stuart, K., Selkirk, M., and Agabian, N., Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphons in *Trypanosoma brucei*. Cell *34* (1983) 901–909.
45 Nelson, R. G., Parsons, M., Selkirk, M., Newport, G., Barr, P. J., and Agabian, N., Sequences homologous to variant antigen mRNA spliced leader in Trypanosomatidae which do not undergo antigenic variation. Nature (London) *308* (1984) 665–667.
46 Nilsen, T. W., Shambaugh, J., Denker, J., Chubb, G., Faser, C., Putnam, L., and Bennett, K., Characterization and expression of a spliced leader RNA in the parasitic nematode *Ascaris lumbridcoides* var. *suum*. Molec. cell. Biol. *9* (1989) 3543–3547.
47 Parker, R., Siliciano, P. G., and Guthrie, C., Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. Cell *49* (1987) 229–239.
48 Parsons, M., Nelson, R. G., Watkins, K. P., and Agabian, N., Trypanosome mRNAs share a common 5′ spliced leader sequence. Cell *38* (1984) 309–316.
49 Patzelt, E., Perry, K. L., and Agabian, N., Mapping of branch sites in trans-spliced pre-mRNAs of *Trypanosoma brucei*. Molec. cell. Biol., *9* (1989) 4291–4297.
50 Payne, M., Rothwell, V., Jasmer, D. P., Feagin, J. E., and Stuart, K., Identification of mitochondrial genes in *Trypanosoma brucei* and homology to cytochrome c oxidase II in two different reading frames. Molec. Biochem. Parasit. *15* (1985) 159–170.
51 Perry, K. L., Watkins, K. P., and Agabian, N., Trypanosome mRNAs have unusual 'cap 4' structures acquired by addition of a spliced leader. Proc. natl Acad. Sci. USA *84* (1987) 8190–8194.
52 Ralph, D., Huang, J., and Van der Ploeg, L. H. T., Physical identification of branched intron side-products of splicing in *Trypanosoma brucei*. EMBO J. *7* (1988) 2539–2545.
53 Ruskin, B., and Green, M. R., Specific and stable intron-factor interactions are established early during in vitro pre-mRNA splicing. Cell *43* (1985) 131–142.
54 Sharp, P. A., Splicing of messenger RNA precursors. Science *235* (1987) 766–771.
55 Shaw, J. M., Feagin, J. E., Stuart, K., and Simpson, L., Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. Cell *53* (1988) 401–411.
56 Shuster, E. O., and Guthrie, C., Two conserved domains of yeast U2 snRNA are separated by 945 nonessential nucleotides. Cell *55* (1988) 41–48.
57 Simpson, L., Neckelmann, N., de la Cruz, V. F., Simpson, A. M., Feagin, J. E., Jasmer, D. P., and Stuart, K., Comparison of the maxicircle (mitochondrial) genomes of *Leishmania tarentolae* and *Trypanosoma brucei* at the level of nucleotide sequence. J. biol. Chem. *262* (1987) 6182–6196.
58 Solnick, D., *Trans* splicing of mRNA precursors. Cell *42* (1985) 157–164.
59 Stuart, K., RNA editing: new insights into the storage and expression of genetic information. Parasit. Today *5* (1989) 5–8.

60 Sutton, R. E., and Boothroyd, J. C., Evidence for *trans* splicing in trypanosomes. Cell *47* (1986) 527–535.
61 Sutton, R. E., and Boothroyd, J. C., Trypanosome *trans*-splicing utilizes 2′–5′ branches and a corresponding debranching activity. EMBO J. *7* (1988) 1431–1437.
62 Takacs, A. M., Denker, J. A., Perrine, K. G., Maroney, P. A., and Nilsen, T. W., A 22-nucleotide spliced leader sequence in the human parasitic nematode *Brugia malayi* is identical to the *trans*-spliced leader exon in *Caenorhabditis elegans*. Proc. natl Acad. Sci. USA *85* (1988) 7932–7936.
63 Thomas, J. D., Conrad, R. C., and Blumenthal, T., The *C. elegans trans*-spliced leader RNA is bound to Sm and has a trimethylguanosine cap. Cell *54* (1988) 533–539.
64 Tschudi, C., Krainer, A. R., and Ullu, E., The U6 small nuclear RNA from *Trypanosoma brucei*. Nuc. Acids Res. *16* (1988) 11 375.
65 Tschudi, D., Richards, F. F., and Ullu, E., The U2 RNA analogue of *T. brucei gambiense*: implications for a splicing mechanism in trypanosomes. Nuc. Acids Res. *14* (1986) 8893–8903.
66 Tschudi, C., and Ullu, E., Polygene transcripts are precursors to calmodulin mRNAs in trypanosomes. EMBO J. *7* (1988) 455–463.
67 Van der Ploeg, L. H. T., Liu, A. Y. C., Michels, P. A. M., De Lange, T., Borst, P., Majumder, H. K., Weber, H., Veeneman, G. H., and Van Boom, J., RNA splicing is required to make the messenger RNA for variant surface antigen in trypanosomes. Nuc. Acids Res. *10* (1982) 3591–3604.
68 Van der Ploeg, L. H. T., Cornelissen, A. W. C. A., Michels, P. A. M., and Borst, P., Chromosome rearrangements in *Trypanosoma brucei*. Cell *39* (1984) 213–221.
69 Van der Spek, H., Van den Burg, J., Croiset, Al, Van den Broek, M., Sloof, P., and Benne, R., Transcripts from the frameshifted MURF3 gene from *Crithidia fasciculata* are edited by U insertion at multiple sites. EMBO J. *7* (1988) 2509–2514.
70 Vickerman, K., Developmental cycles and biology of pathogenic trypanosomes. Br. med. Bull. *41* (1985) 105–114.
71 Weiner, A. M., and Maizels, N., RNA editing: guided but not templated. Cell *61* (1990) 917–920.

0014-4754/91/020118-11$1.50 + 0.20/0

The molecular epidemiology of parasites

G. Hide and A. Tait

Wellcome Unit of Molecular Parasitology, University of Glasgow, Bearsden Road, Glasgow (Scotland)

Summary. The explosion of new techniques, made available by the rapid advance in molecular biology, has provided a battery of novel approaches and technology which can be applied to more practical issues such as the epidemiology of parasites. In this review, we discuss the ways in which this new field of molecular epidemiology has contributed to and corroborated our existing knowledge of parasite epidemiology. Similar epidemiological questions can be asked about many different types of parasites and, using detailed examples such as the African trypanosomes and the *Leishmania* parasites, we discuss the techniques and the methodologies that have been or could be employed to solve many of these epidemiological problems.
Key words. Molecular epidemiology; epidemiology; parasites; DNA probes; trypanosomes; *Leishmania*; parasitic nematodes.

1 Introduction

One of the central areas of investigation considered in the study of parasites and the diseases they cause concerns the dynamics of interaction of the parasite, its host(s) and the environment. The study of such interactions – epidemiology – seeks to describe, both qualitatively and quantitatively, the rules which govern the ways in which a given parasite is transmitted from one host to other hosts, and how this influences the spread of disease within and between populations of hosts. Such studies on transmission must encompass a wide range of factors in order to document fully the epidemiology of a given parasitic disease.

One of the primary problems in parasite epidemiology is the correlation of a given disease with its causative organism. In some cases such an association can be clearly made. For example, in disease due to the bovine lungworm, *Dictyocaulus viviparus*, the clinical signs of bronchitis and pneumonia can be readily associated with the presence of adult worms in the trachea and bronchi. In contrast, the association between parasite and disease may be much less obvious, as is the case with the morphologically identical trypanosomes of the *Trypanosoma brucei* complex, the causative agents of African sleeping sickness. Here, apparently identical parasites have different specificities for the host and therefore different ranges.

Having established what is the causative agent of a disease, the epidemiologist must then consider the ways in which transmission occurs. Parasites can utilise very straightforward transmission routes; for example, *Trypanosoma equiperdum* is transferred from one host to the next by direct venereal transmission. Other parasites may require more complex life cycles, like the malaria parasite, *Plasmodium*, which uses the mosquito as a vector for transmission, or the blood fluke, *Schistosoma*, which requires an intermediate host to complete its life cycle. Mechanisms of transmission are almost as diverse as the parasites that use them.