Trans splicing in trypanosomes – archaism or adaptation?

PETER W. LAIRD

In trypanosomes, a single transcription unit usually covers several protein-coding genes. The primary transcript is cut up by trans-splicing and polyadenylation machineries to generate individual mature mRNAs. All nuclear mRNAs acquire the same capped 39 nucleotide sequence at their 5' end as a consequence of the trans-splicing event. Trans splicing is used in the synthesis of some mRNAs in nematodes and chloroplasts. These unusual systems are clearly related to cis-splicing systems, but it remains an intriguing question whether they are merely exotic offshoots of cis splicing or archaic remnants of cis-splicing progenitors.

Discontinuous RNA synthesis involves the covalent linkage of RNA segments from two different transcription units. It differs from various other forms of transcript modification, such as polyadenylation, in that the alteration in the RNA is specified by a template. As Table 1 shows, linkage of the two RNA segments can be either co- or post-transcriptional. Examples of co-transcriptional linkage can be found in several viral systems[1]: influenza virus obtains primers for mRNA synthesis by cleaving 10–13 nucleotides from the 5' ends of newly synthesized host mRNAs ('cap snatching'), while corona viruses use a primer encoded by the viral genome itself.

Post-transcriptional linkage of RNA segments could in theory occur by simple end-to-end ligation or by *trans* splicing; so far, only examples of *trans* splicing have been found. Post-transcriptional linkage of RNA segments was discovered in 1984 in African trypanosomes[2–4], but it has since been shown to occur in other kinetoplastids, in nematodes[5] and in chloroplasts[6–8] as well.

Discontinuous mRNA synthesis in trypanosomes

Several early observations led to the conclusion that mRNA synthesis in the mammalian parasite *Trypanosoma brucei* is discontinuous[2–4]. Attention initially focused on the genes encoding the surface coat proteins involved in the antigenic variation of *T. brucei*. The variant-specific surface glycoprotein (VSG) gene transcripts were found to carry a short 5' sequence that was not encoded immediately adjacent to the otherwise intronless main portion of the gene. This sequence was the same for each of the VSG mRNAs studied.

It was assumed that this short sequence was encoded by a separate exon and spliced onto the remainder of the mRNA. Boothroyd and Cross[9] coined the name 'mini-exon' for this sequence in 1982; since then the term 'spliced leader' has also come into use.

The mini-exon was initially thought to be involved in the regulation of VSG gene expression, but it soon became clear that other mRNAs also carried the mini-exon sequence. The virtually total inhibition of *in vitro* translation of *T. brucei* RNA by hybridization with oligodeoxynucleotides complementary to the mini-exon sequence indicates that probably all trypanosome mRNAs carry this short sequence at their 5' ends[10,11].

There are about 200 copies of the mini-exon encoded in clusters in the genome, each mini-exon being part of a 1.35 kb sequence that is tandemly repeated. Hence, many trypanosome chromosomes contain no mini-exons at all, even though they embody genes that give rise to mRNAs that do contain a mini-exon. The 1.35 kb mini-exon repeats show very little sequence conservation between different kinetoplastids, except within the mini-exon sequence itself. The only detectable transcript from the mini-exon repeats is a 140 nucleotide RNA that carries the mini-exon at its 5' end. This RNA is known either as the spliced leader RNA (SLRNA) or the medRNA, which stands for mini-exon-derived RNA (as seen from DNA) or mini-exon-donor RNA (as seen from RNA).

TABLE 1. Discontinuous mRNA synthesis

(A) Co-transcriptional (primed RNA synthesis)
- Influenza viruses and Bunya viruses: 'cap snatching'
- Corona viruses: priming by virus-derived primers

(B) Post-transcriptional (*trans* splicing)
- Trypanosomes and other Kinetoplastida: all mRNAs
- *Caenorhabditis elegans*: a minority of mRNAs
- Chloroplasts: a small minority of mRNAs

The mini-exon sequences on both the mRNA and the medRNA carry a normal 7-methyl-guanosine cap in addition to modified nucleotides at the first four and at the sixth position[12,13]. The medRNA has the high turnover expected for an mRNA precursor; its half-life *in vivo* is four minutes[14].

Together these results established that the mini-exon sequence is transcribed separately from the rest of the mRNAs and therefore that mRNA synthesis in *T. brucei* is discontinuous.

The mechanism of mRNA synthesis in *T. brucei*

The mini-exon sequence in medRNA and the remainder of the mRNA in pre-mRNAs are both flanked by sequences similar to nuclear pre-mRNA splice sites (Fig. 1) and this suggested at an early stage that splicing was involved in the final linkage of mini- and main exon sequences. Several linkage mechanisms incorporating a splicing step could be envisaged. A normal *cis*-splicing reaction would suffice if the medRNA were first linked to the pre-mRNA, either by the use of medRNA as a transcription primer or by simple end-to-end ligation of the medRNA and (cleaved) pre-mRNA.

Alternatively, the splicing reaction could take place without prior covalent joining of the two RNAs – in other words, a splicing reaction in *trans*. If such a reaction mimicked nuclear pre-mRNA *cis* splicing, it would result in splicing intermediates with a simple

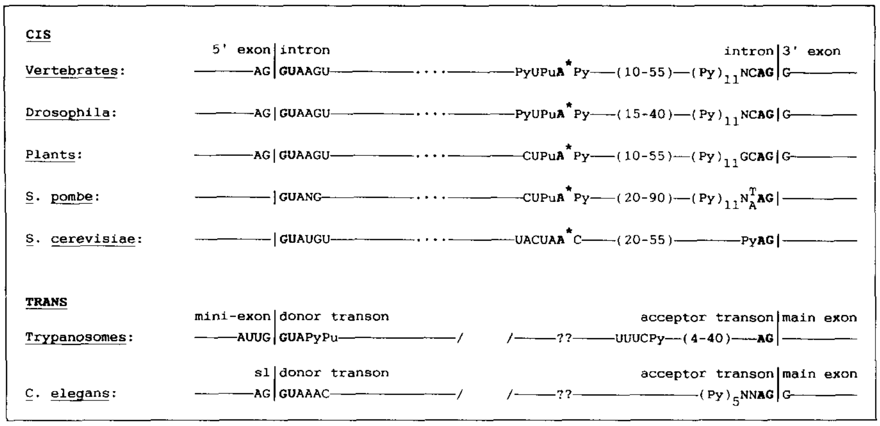

FIG **1**

A comparison of eukaryotic nuclear mRNA splice signals. *Cis*- and *trans*-splicing signals of several eukaryotic groups and species are listed as schematic RNA molecules. The nucleotides shown indicate conserved sequences, except for the *C. elegans* sequence, which is based on one splice donor sequence and two actin splice acceptor sequences. Horizontal lines indicate nonconserved sequences. The intron borders are designated by vertical bars. Dots indicate the variable intron lengths and the asterisks denote the adenosine residues involved in branch formation. The 100 nucleotide 3' part of the medRNA and the upstream part of the pre-mRNA are not intron sequences in the true sense of the word. Zaita *et al.* have therefore coined the term 'transon' to describe such segments[6].

forked structure (Fig. 2) instead of the lariat structures found in conventional nuclear pre-mRNA *cis*-splicing intermediates.

As attempts to establish *in vitro* linkage met with failure, such diagnostic intermediates proved to be the key to the elucidation of the mechanism. Incubation of trypanosome RNA with a debranching enzyme that hydrolyses the 2'–5'-phosphodiester bonds found in lariat branches yields small amounts of a 100 nucleotide RNA corresponding to the medRNA minus the mini-exon sequence[15–17] (Fig. 2). This is the result predicted if forked *trans*-splicing intermediates are present at low amounts in steady-state RNA. Branched molecules containing the 100 nucleotide segment have also been detected by their anomalous migration in two-dimensional gels[18].

These results together with other circumstantial evidence such as the short half-life of medRNA[14,17] and the lack of detectable full-length medRNA covalently linked to high-molecular-weight RNA[14,15,17] have led to a consensus that *trans* splicing is the mechanism linking mini-exon and main exon sequences. Our present understanding of trypanosome mRNA synthesis is summarized in Fig. 2. The 50 or so trypanosome genes analysed to date are generally tightly packed and lack introns[2]. In most cases, gaps in nascent transcription between genes cannot be found and the transcription units appear to span multiple genes. Even so, long multicistronic primary transcripts are usually not detected, possibly because of co-transcriptional processing, as shown in Fig. 2.

Little is known of the factors responsible for carrying out the reactions. Homologs of several components known to be involved in nuclear *cis* splicing in other eukaryotes have been found in trypanosomes. These include U2 (Ref. 19), U4 and U6 snRNAs. *Trans* splicing therefore probably involves a spliceosome-like particle. A trypanosome homolog for U1 – the snRNP associated with the 5' splice site in mammalian mRNA splicing – has not been identified, which raises the possibility that the medRNA itself, being a small capped RNA transcribed by RNA polymerase II (Ref. 20), takes on the role of this snRNA. Such a double role of substrate and snRNA, first proposed by Sharp[21], is supported by the recent discovery that the medRNAs of various kinetoplastids can bind mammalian Sm protein, a component of snRNPs[22]. It remains to be seen how the typical assembly of snRNPs in the cytoplasm and re-entry into the nucleus can be reconciled with the very short half-life of medRNA[14,17]. The intrinsic low efficiency of a bimolecular *trans*-splicing reaction compared with a monomolecular *cis*-splicing reaction would be circumvented if the medRNA were a preassembled

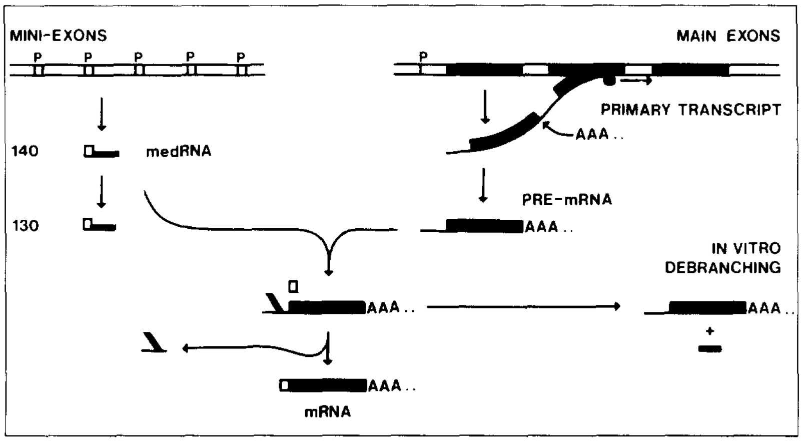

FIG **2**

Schematic representation of discontinuous mRNA synthesis in *T. brucei*. The 140 nucleotide mini-exon-derived RNA (medRNA) is transcribed from mini-exons in tandem repeats. The short capped mini-exon sequence in the medRNA, indicated by an open block, is *trans* spliced onto polyadenylated pre-mRNAs, which have been processed probably co-transcriptionally from nascent primary transcripts of multicistronic units. The 130 nucleotide species is a 3'-truncated medRNA, which accumulates *in vivo* when mRNA synthesis is inhibited, but probably does not participate in the *trans*-splicing reaction[14]. The *trans*-splicing reaction proceeds through a forked intermediate, which contains the 3' part of medRNA as a 2'–5'-phosphodiester-linked branch, which can be released *in vitro* by a treatment with debranching enzyme. The final product is the mature polyadenylated mRNA with the capped mini-exon sequence at its 5' end. (From Ref. 14.)

universal component of the spliceosome in trypanosomes.

Mini-exons and discontinuous mRNA synthesis appear to be common to all major genera of the order Kinetoplastida, and it seems likely that discontinuous mRNA synthesis is a universal characteristic of Kinetoplastida. Other peculiar features of this order include the presence within the mitochondrion of the kinetoplast, an unusual network of kinked catenated DNA circles that lends the order its name, and the presence of microbodies called glycosomes, which contain most of the enzymes involved in glycolysis[2].

Relevance to other systems

The first evidence that *trans* splicing could occur came from *in vitro* studies in the mammalian nuclear pre-mRNA splicing system. Noncovalently linked exon sequences can be joined at low efficiency in the HeLa cell extract system[23,24].

In contrast to cofactor-dependent *trans* splicing in HeLa cell extracts and in trypanosomes, in some situations *trans* splicing can occur *in vitro* in the absence of cofactors such as snRNPs (so-called *trans* self-splicing reactions). Both group I and group II self-splicing intron systems have been found to be capable of autocatalytic splicing in *trans* in artificial systems[25,26].

Group II *trans* self-splicing reactions can also occur *in vivo*. Exon 1 of the tobacco chloroplast ribosomal protein S12 gene is transcribed separately from exons 2 and 3. The exon 1 sequence is then spliced in *trans* to the exon 2 sequence[6,7]. A similar situation is found for the *psaA* gene in the plastid DNA of the green alga *Chlamydomonas reinhardtii*[8]. In both cases the separate transcripts each contain non-exon RNA segments which, after annealing with each other, result in a structure resembling a group II self-splicing intron. As in trypanosomes, the non-exon RNA segments cannot be viewed as true intron sequences. Zaita *et al.* have coined the term transon to describe such discontinuous gene-associated sequences[6].

The system most closely related to the *trans*-splicing mechanism in trypanosomes is the *trans*-splicing event in the mRNA synthesis of three out of four actin genes and of a number of other genes in the nematode *Caenorhabditis elegans*[5]. In these cases, a common 22 nucleotide leader sequence encoded near the 5S rRNA genes is spliced *in trans* from a 100 nucleotide spliced leader RNA onto the 5' ends of the nascent mRNAs. The similarity to trypanosome mRNA synthesis is striking. Three differences are that the *C. elegans* spliced leader contains a trimethylguanosine cap, that *C. elegans* uses this mechanism for only a subset of genes and that the nematode also possesses genes with normal introns, whereas introns have not been found in trypanosome genes. How the *C. elegans* spliced leader is prevented from splicing into the middle of genes at normal splice acceptor sites is not known.

Origin of *trans* splicing

Current consensus now holds that our DNA–RNA–protein world evolved from a self-replicating RNA system. Similarities between self-splicing reactions and nuclear pre-mRNA splicing reactions[27] suggest that these systems may have a common ancestor, which could have evolved before the development of translation.

In trying to understand why trypanosomes rely on such an outlandish method of mRNA synthesis, it is worth considering the possible evolutionary origin of *trans* splicing. The variety in the examples of *trans* splicing described above appears to reflect the variation in corresponding *cis*-splicing systems. Of the four categories of *cis* splicing, two have naturally occurring counterparts *in trans* (Table 2). The particular *trans*-splicing reactions have more in common with their related *cis*-splicing systems than with each other. This similarity between the *cis* and *trans* reactions suggests a close evolutionary relationship between the two. Their common evolutionary ancestor could in principle have been either a primitive *cis*- or *trans*-splicing system.

A *cis*-splicing ancestor

So far, speculation in the literature has focused mainly on the possibility that *trans*-splicing systems are evolutionary side-tracks derived from ancestral *cis*-splicing systems (see, for example, Refs 21, 22, 28). The trypanosome and nematode *trans*-splicing

TABLE 2. Four types of *cis*-splicing reactions and their *trans*-splicing counterparts

Four types of *cis* splicing	Splicing mechanism	*Trans* splicing counterparts
1 - tRNA splicing	Simple, enzymatic 3' Cyclic phosphate intermediate RNA structure important	No examples
2 - Group I self-splicing	Simple, autocatalytic GTP-dependent RNA structure important	Artificial examples
3 - Group II self splicing	Simple, autocatalytic Lariat intermediate RNA structure important	Artificial examples Chloroplast rps 12 Alga plastid *psaA*
4 - Nuclear pre-mRNA splicing	Complex, enzymatic Lariat intermediate RNA structure less important	Artificial examples Trypanosomes: all mRNAs Nematodes: some mRNAs

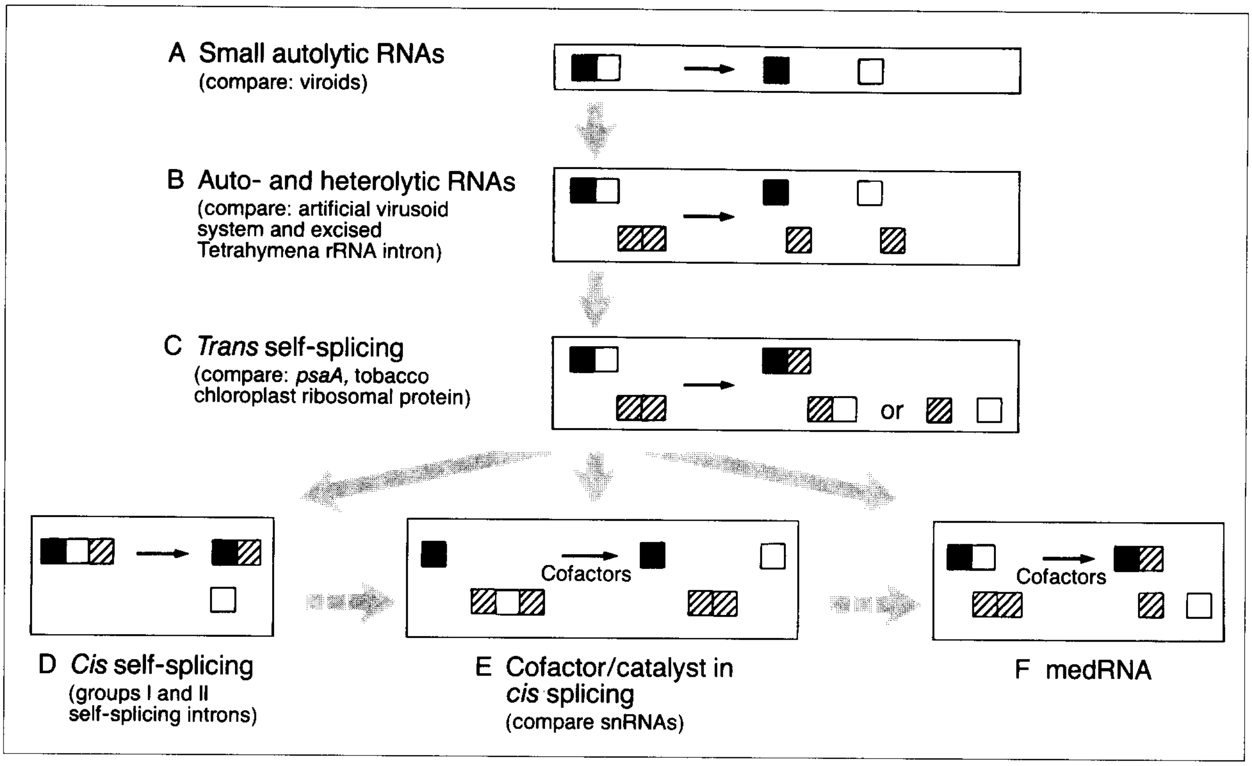

FIG **3**

Proposed evolutionary pathway for *cis*- and *trans*-splicing systems. The four basic steps described in the text are depicted with present-day possibly vestigial examples of the various intermediates listed. Various RNA segments are depicted by differently shaded blocks.

medRNAs and SLRNAs would then represent 5' exons that have become separated from their splice acceptors. In this view, the similarities between the medRNAs and small nuclear RNAs could be due to a fusion of these derailed 5' exons to snRNA sequences[21]. An even more provocative notion is that small nuclear RNAs arose directly from such liberated 5' exons, ultimately losing their splice donor site in addition to their splice acceptor site, but retaining their catalytic function[22].

A *trans*-splicing ancestor

According to an alternative scheme for splicing evolution, *trans*-splicing systems were the first to arise during evolution and then in turn gave rise to the more sophisticated *cis*-splicing systems. I shall attempt to trace a possible evolutionary pathway through vestigial examples of RNA catalysis (see also Fig. 3).

As a first step in the evolution of self-splicing, the most simple autocatalytic reaction conceivable is self-cleavage. Plant viroid RNAs[29] and satellite transcripts from the newt[30] are capable of autocatalytic cleavage and would represent present-day counterparts of such RNAs (Fig. 3A).

A next step would be the evolution of an RNA molecule capable of cleaving not only itself, but also another RNA molecule (Fig. 3B). The catalytic center for the cleavage capacity of the *Tetrahymena* ribozyme and of the viroids can be physically separated from the substrate. The prokaryotic RNase P RNAs are natural examples of RNAs capable of catalytic cleavage of other RNA molecules.

The reverse reaction of cleavage is ligation. If the energy derived from the cleavage reaction is conserved in a normal 5'–3' bond, in a 5'–2' branch or as a 2'–3' cyclic phosphodiester bond, the cleavage should in principle be reversible. This indeed appears to be the case for the *Tetrahymena* ribozyme and RNAs complementary to a plant virus satellite RNA[31]. The prevalence of 5'–2' branches and 2'–3' cyclic phosphodiester bonds in modern splicing intermediates may reflect the requirement of autocatalytic cleavage reversibility in the early development of splicing systems.

Catalytic cleavage by an RNA of itself and of another RNA, followed by a reversal of the reactions, but with an exchange of RNA segments constitutes *trans* self-splicing; the examples of *trans* self-splicing in the tobacco chloroplast ribosomal protein and *C. reinhardtii psaA* genes mentioned earlier might be relics of such systems (Fig. 3C). Reactions of this type are the RNA equivalents of reciprocal and nonreciprocal DNA recombination and could have played a major role in generating sequence diversity and in recombining advantageous RNA segments in an RNA world.

Evolution could now have followed several courses. Linkage of the catalytic RNA to the target RNA would yield RNAs capable of *cis* self-splicing. Group I and group II self-splicing introns would be modern representatives of such RNAs (Fig. 3D).

Alternatively, the catalytic RNA could have been further refined, so that it could function as a catalyst without being a substrate – in other words an RNA capable of catalysing *cis* splicing in other RNAs (Fig. 3E). Such an RNA could play a major role in bringing together RNA segments encoding primitive protein domains, as proteins started to evolve. The evolution of proteins would yield catalysts with a higher degree of versatility and specificity, which could assist the catalytic RNAs in the *cis*-splicing reactions, allowing the RNAs to degenerate to the small nuclear RNAs involved in non-self-catalysed nuclear pre-mRNA splicing as we know it today.

How does cofactor-dependent *trans* splicing in trypanosomes fit into this scheme? The medRNA resembles the autocatalytic molecules capable of splicing themselves onto other RNAs, but as in the case of small nuclear RNAs, the medRNA has apparently lost its autocatalytic activity (Fig. 3F).

The fact that trypanosomes also contain molecules homologous to snRNAs could be interpreted to mean that at a certain stage in evolution, organisms contained both types of RNA. Most eukaryotes retained the snRNAs but lost the *trans*-splicing donors, while kinetoplastids and nematodes retained both types of molecules. The kinetoplastids apparently deleted introns in protein-coding regions, perhaps for reasons of genome streamlining as in the case of prokaryotes, or maybe to avoid interference with the *trans*-splicing machinery.

Role of the mini-exon

Irrespective of whether trypanosome *trans* splicing is derived from a primordial *trans*-splicing system or simply more recently from a *cis*-splicing system, the question remains why kinetoplastids have selected and retained this system of mRNA synthesis. A compelling reason is not immediately obvious. The answer is probably linked to the multicistronic transcription units of protein-coding genes: *trans* splicing is an elegant way to provide individual mRNAs from a multicistronic transcript with a 5' cap.

It does not seem likely that the function of the multicistronic transcription units is to alleviate the rate-limiting step of transcription initiation, since each mRNA still requires initiation of transcription of a mini-exon. At best the reservoir of medRNA molecules could provide a buffer of pre-initiated transcripts upon a sudden change of environment.

An alternative hypothesis states that trypanosome mRNAs must have a single common 5' sequence because of a restrictive mRNA transport or translation system. The 5' sequence could efficiently be kept homogeneous by gene conversion events in the mini-exon repeats. However, even if such systems were shown to be dependent on the mini-exon sequence, this would not prove that this was the original function of the mini-exon sequence; transport or translation machineries may be expected to have adapted to this universal sequence anyhow.

Neither of the explanations for the method of mRNA synthesis in trypanosomes suggested above is very convincing, since they are both cumbersome solutions to problems easily avoided or solved in other eukaryotes. At present, the most plausible scenario to me is that the primitive ancestral eukaryotes used both *cis* and *trans* splicing. The kinetoplastid lineage slowly became dependent upon *trans* splicing in combination with the development of multicistronic transcription units while most other eukaryotes weaned themselves off dependency on this activity. *Trans* splicing was then retained by necessity in trypanosomes and lost due to redundancy in most other eukaryotes.

Acknowledgements

I thank Dr Piet Borst and his collaborators for their advice. I was supported by a grant from the Netherlands Foundation for Chemical Research (SON-11-26-26) with financial aid from the Netherlands Organization for Scientific Research (NWO).

References

1 Krug, R.M. (1985) *Cell* 41, 651–652
2 Borst, P. (1986) *Annu. Rev. Biochem.* 55, 701–732
3 Boothroyd, J.C. (1985) *Annu. Rev. Microbiol.* 39, 475–502
4 Van der Ploeg, L.H.T. (1986) *Cell* 47, 479–480
5 Krause, M. and Hirsch, D. (1987) *Cell* 49, 753–761
6 Zaita, N., Torozawa, K., Shinozaki, K. and Sugiura, M. (1987) *FEBS Lett.* 210, 153–156
7 Koller, B., Fromm, H., Galun, E. and Edelman, M. (1987) *Cell* 48, 111–119
8 Choquet, Y. *et al.* (1988) *Cell* 52, 903–913
9 Boothroyd, J.C. and Cross, G.A.M. (1982) *Gene* 20, 281–289
10 Cornelissen, A.W.C.A. *et al.* (1986) *Nucleic Acids Res.* 14, 5605–5614
11 Walder, J.A. *et al.* (1986) *Science* 233, 569–571
12 Freistadt, M.S., Cross, G.A.M., Branch, A.D. and Robertson, H.D. (1987) *Nucleic Acids Res.* 15, 9861–9879
13 Perry, K.L., Watkins, K.P. and Agabian, N. (1987) *Proc. Natl Acad. Sci. USA* 84, 8190–8194
14 Laird, P.W., ten Asbroek, A.L.M.A. and Borst, P. (1987) *Nucleic Acids Res.* 15, 10087–10103
15 Murphy, W.J., Watkins, K.P. and Agabian, N. (1986) *Cell* 47, 517–525
16 Sutton, R.E. and Boothroyd, J.C. (1986) *Cell* 47, 527–535
17 Laird, P.W., Zomerdijk, J.C.B.M., De Korte, D. and Borst, P. (1987) *EMBO J.* 6, 1055–1062
18 Ralph, D., Huang, J. and Van der Ploeg, L.H.T. (1988) *EMBO J.* 7, 2539–2545
19 Tschudi, C., Richards, F.F. and Ullu, E. (1986) *Nucleic Acids Res.* 14, 8893–8903
20 Laird, P.W., Kooter, J.M., Loosbroek, N. and Borst, P. (1985) *Nucleic Acids Res.* 13, 4253–4266
21 Sharp, P.A. (1987) *Cell* 50, 147–148
22 Bruzik, J.P., Van Doren, K., Hirsch, D. and Steitz, J. (1988) *Nature* 335, 559–562
23 Solnick, D. (1985) *Cell* 42, 157–164
24 Konarska, M.M., Padgett, R.A. and Sharp, P.A. (1985) *Cell* 42, 165–171
25 Inoue, T., Sullivan, F.X. and Cech, T.R. (1985) *Cell* 43, 431–437
26 Jacquier, A. and Rosbash, M. (1986) *Science* 234, 1099–1104
27 Cech, T.R. (1986) *Cell* 44, 207–210
28 Blumenthal, T. and Thomas, J. (1988) *Trends Genet.* 4, 305–308
29 Forster, A.C. and Symons, R.H. (1987) *Cell* 50, 9–16
30 Epstein, L.M. and Gall, J.G. (1987) *Cell* 48, 535–543
31 Buzayan, J.M., Gerlach, W.L. and Bruening, G. (1986) *Nature* 323, 349–353

P.W. LAIRD IS IN THE DIVISION OF MOLECULAR BIOLOGY H-8, NETHERLANDS CANCER INSTITUTE, PLESMANLAAN 121, 1066 CX AMSTERDAM, THE NETHERLANDS; PRESENT ADDRESS: DIVISION OF MOLECULAR GENETICS H-4, AT THE ABOVE INSTITUTE.