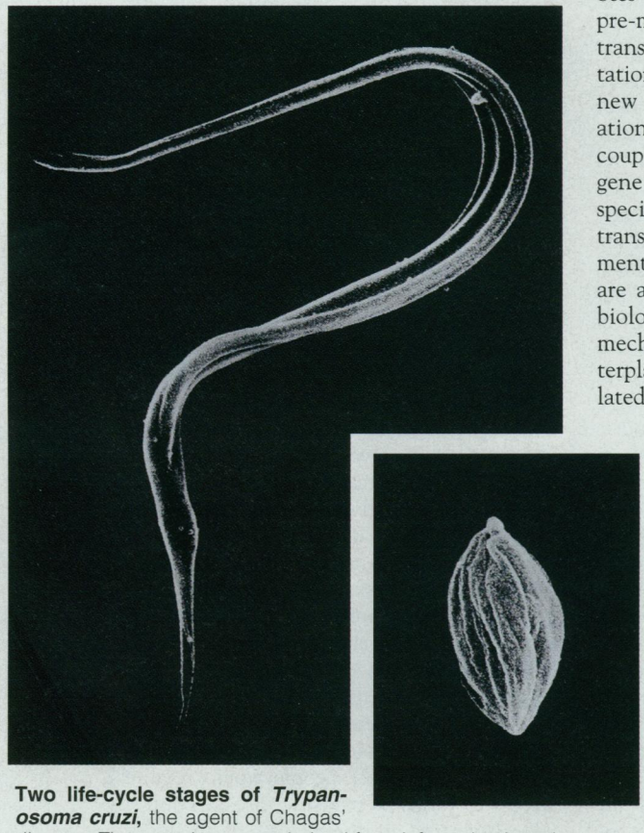
Unusual Strategies of Gene Expression and Control in Parasites

Timothy W. Nilsen

Some parasites display highly unusual mechanisms of mRNA maturation and modes of gene regulation in comparison to those of higher eukaryotes. The details of these mechanisms have come primarily from in vivo studies in the trypanosomids and in vitro manipulation of cell-free systems derived from the parasitic nematode *Ascaris lumbricoides*. The first eukaryotic multicistronic transcription units were found in trypanosomatids, and they also exhibit trans-splicing and RNA editing (see accompanying Perspective by Simpson and Maslov). Unlike transcriptional gene regulation in other eukaryotes, gene regulation in these organisms is largely dictated by post-transcriptional mechanisms.

At first blush, trypanosomatid mRNA appears indistinguishable from the standard eukaryotic mRNAs. It has a single coding sequence flanked by 5' and 3' untranslated regions; the 5' ends are capped and the 3' ends are polyadenylated. Yet, the biogenesis of this RNA differs in almost every respect from that in most eukaryotes. Trypanosomatid genes exhibit an eccentric mode of transcription: These organisms are unusual because they contain a number of tandem arrays of closely spaced genes (*1*), which do not seem to contain promoter elements either within or at the extremities of their sequences. These findings, coupled with short-wavelength ultraviolet light inactivation studies and subsequent transfection experiments, have clearly established that most protein-coding genes are synthesized as part of multicistronic primary transcripts (*1*). The generation of mature mRNA requires that individual coding sequences and their associated untranslated regions be excised from the primary transcript. This is accomplished by two RNA-processing reactions: intermolecular (trans)-splicing and polyadenylation. Trans-splicing is an unusual RNA processing reaction that precisely joins exons derived from separate transcripts (*2, 3*). As discussed below, trans-splicing in trypanosomatids (and other organisms) is closely related to the process whereby introns are excised from nuclear pre-mRNAs in other eukaryotes. The substrates of trans-splicing are a small non-polyadenylated RNA, the spliced leader RNA (SL RNA), which contains the 5' exon (SL) and splice donor site, and the acceptor pre-mRNA, which contains the splice acceptor site and 3' exon. Through mechanisms not yet fully understood, the SL RNA and pre-mRNA efficiently associate in the nucleus, and the SL is spliced to the body of the mRNA, thereby creating a discrete 5' end. Because the SL RNA is capped, trans-splicing also serves as an unconventional but effective mechanism of ensuring that each mature mRNA has a 5' terminal cap. Trans-splicing seems to be an obligatory step in the maturation of all trypanosomatid mRNAs—all mRNAs appear to contain the SL sequence (*4*) and no protein-coding genes in these organisms contain the usual eukaryotic cis intervening sequences.



Two life-cycle stages of *Trypanosoma cruzi*, the agent of Chagas' disease. The parasites were derived from infected cultured mammalian cells. The upper, elongated trypomastigote form is normally the blood-borne, infective stage, and the lower, rounded amastigote is the replicative, intracellular stage. [Courtesy of Edith Robbins, New York University]

Maturation of the 3' end of mRNAs is thought to occur (as it does in other eukaryotes) through endonucleolytic cleavage and polyadenylation. Until quite recently, it was not known how trypanosomatid polyadenylation sites were specified; no consensus sequence analogous to the AAUAAA hexanucleotide characteristic of polyadenylation sites in higher eukaryotes has been identified in trypanosome pre-mRNAs. The availability of efficient transfection systems (*5*) coupled with mutational analyses has provided intriguing new insight into this question. Polyadenylation of an upstream gene is functionally coupled to trans-splicing of a downstream gene (*6, 7*); indeed, poly A site selection is specified by the position of the downstream trans-splice acceptor site, and sequence elements recognized by the splicing machinery are also required for polyadenylation. The biological significance and the biochemical mechanisms underlying the functional interplay between these two apparently unrelated RNA-processing reactions will be of considerable interest.

Do the unconventional mechanisms of mRNA synthesis and maturation reflect similarly unusual gene regulation in trypanosomatids? The similarity of multicistronic transcription units to operons in prokaryotic organisms has led to the suggestion that both reflect mechanisms to coordinately regulate cotranscribed genes. Although this idea has intrinsic appeal, it remains unclear whether it is generally applicable in trypanosomatids. In some cases, multicistronic transcription units contain genes encoding functionally related proteins, whereas in other multicistronic units, the functional relationship, if any, of cotranscribed genes is unknown (*1*). Furthermore, all well-characterized bacterial operons are tightly regulated by control of transcription initiation or elongation. In trypanosomes, control of transcription initiation has not been observed, and there is only one instance of transcriptional regulation of elongation, the variant surface glycoprotein (VSG) transcription unit (see accompanying Perspective by Borst and Rudenko). Despite the apparent dearth of transcriptional control, the abundance of specific mRNAs

The author is in the Center for Molecular Parasitology, Department of Molecular Biology and Microbiology, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA.

Three organisms that trans-splice. Two parasitic metazoan worms, *Schistosoma mansoni* (**left**) and *Ascaris lumbricoides* (**middle**), and a nonparasitic protist, *Euglena* (**right**). [Photos by R. Kessel and G. Shih (right and left) and A. M. Siegelman (middle)/Visuals Unlimited]

(including those within the same transcription unit) can differ greatly, and there are several examples of developmentally regulated mRNAs (*1*). In principle, mRNA abundance could be determined by trans-splicing, polyadenylation, transport to the cytoplasm, or stability. In fact, the mRNAs for VSG are destabilized during the transformation from the bloodstream form to the procyclic form of the organism (*8*), and the 3' untranslated regions of several genes participate in their developmental regulation (*9*), although we do not yet know how.

Because trypanosomatids are one of the earliest branches in the eukaryotic phylogenetic tree (*10*), it is tempting to speculate that their promiscuous transcription and exclusively posttranscriptional regulatory mechanisms predated the evolution of the complex transcriptional regulatory networks typical of higher eukaryotic cells. If this is true, the apparent oddities of trypanosomatid gene expression would not be confined to these organisms. Indeed, the numerous examples of posttranscriptional gene control in higher eukaryotes may represent vestiges of what were once primary regulatory circuits.

Trans-splicing is also more prevalent than previously suspected. Although its overall phylogenetic distribution is not known, trans-splicing occurs in nematodes, both free-living and parasitic, some parasitic trematodes (*Schistosoma*), and in *Euglena*, a nonparasitic photosynthetic protist (*11*). Even mammalian cells (and extracts) can catalyze trans-splicing when presented with heterologous substrates, but there is (as of yet) no evidence for endogenous vertebrate SL RNAs (*12*). In contrast to trypanosomatids, the other organisms that trans-splice also process pre-mRNAs by cis-splicing, and in most cases the same pre-mRNAs are substrates for both reactions (*3*). The fact that *Euglena*, which branched at least as early as the trypanosomatids in eukaryotic evolution (*10*), possesses both cis- and trans-splicing machinery suggests that both reactions evolved concurrently (and, if true, that trypanosomatids must have lost their cis introns). Coevolution of cis- and trans-splicing would explain the fundamental mechanistic similarities between the two processes (see below).

Trypanosomatids use trans-splicing to process multicistronic pre-mRNAs. Does trans-splicing perform a similar function in other organisms? Unfortunately, essentially nothing is known about transcription units encoding trans-spliced mRNAs in *Euglena* and *Schistosoma*. Furthermore, despite the extensive (and growing) literature on gene structure and organization in *Caenorhabditis elegans* (a free-living nematode), surprisingly little is known about transcription in this organism. A notable exception is a group of multicistronic transcription units in *C. elegans* that require trans-splicing for maturation of their mRNAs (*13*). These studies provide the first functional parallel between trans-splicing in the trypanosomatids and trans-splicing in a metazoan and suggest that multicistronic transcription will be a general feature of gene expression in organisms that carry out trans-splicing.

Although our understanding of the prevalence and function of trans-splicing remains somewhat fragmentary, a clear picture of its biochemical mechanism is beginning to emerge. Trans-splicing seems to be a close mechanistic cousin of cis-splicing (*14*). Both processes use a two-step phosphoryl transfer reaction pathway that generates analogous products and intermediates. The splice donor and acceptor sites of trans-splicing conform to cis-splicing consensus sequences, and both reactions occur in large ribonucleoprotein complexes with at least some common constituents. Although cis- and trans-splicing are unambiguously related, they are not identical. Trans-splicing may require novel small nuclear RNAs not found in the cis-spliceosome and may depend on unique protein-protein or RNA-RNA interactions not required for cis-splicing (but needed to facilitate the efficient association of SL RNAs with pre-mRNAs) (*15*, *16*).

What is the evolutionary origin and function of SL RNAs? These RNAs (unique to trans-splicing) bear the hallmarks characteristic of the U small nuclear RNAs that are required for cis-splicing (*17*). But unlike these U snRNAs, SL RNAs are consumed during the trans-splicing reaction. They may be merely exon delivery vehicles or may participate in the trans-splicing process itself. It is widely speculated that the U snRNAs required for cis-splicing are descendants of functional domains of self-splicing introns. If true, where do the SL RNAs fit in?

Molecular analyses in parasitic organisms and their free-living relatives have been exceptionally rewarding in revealing unexpected strategies of gene expression and control. Recent technical breakthroughs including transfection in other parasites, such as malaria and *Toxoplasma* (*18*), promise to yield more surprises.

References

1. C. Clayton, *Prog. Nucleic Acids Res. Mol. Biol.* **43**, 37 (1992).
2. N. Agabian, *Cell* **61**, 1157 (1990).
3. T. W. Nilsen, *Annu. Rev. Microbiol.* **47**, 413 (1993).
4. J. A. Walder *et al.*, *Science* **233**, 569 (1986).
5. V. Bellofatto and G. A. M. Cross, *ibid.* **244**, 1167 (1989); A. Laban and D. F. Wirth, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9119 (1989); G. M. Kapler, C. M. Coburn, S. M. Beverley, *Mol. Cell. Biol.* **10**, 1084 (1990).
6. K. R. Matthews, C. Tschudi, E. Ullu, *Genes Dev.* **8**, 491 (1994).
7. J. H. LeBowitz, H. Q. Smith, L. Rusche, S. M. Beverley, *ibid.* **7**, 996 (1993).
8. E. Pays *et al.*, *EMBO J.* **9**, 3145 (1990).
9. D. Jeffries, P. Tebabi, E. Pays, *Mol. Cell. Biol.* **11**, 338 (1991); M. Hug *et al.*, *Mol. Biochem. Parasitol.* **61**, 87 (1993).
10. M. L. Sogin, J. H. Gunderson, H. J. Elwood, R. A. Alonso, D. A. Peattie, *Science* **243**, 75 (1989).
11. M. Krause and D. Hirsh, *Cell* **49**, 753 (1987); A. Rajkovic, R. E. Davis, J. N. Simonsen, F. M. Rottman, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8879 (1990); L.-H. Tessier *et al.*, *EMBO J.* **10**, 2621 (1991).
12. J. P. Bruzik and T. Maniatis, *Nature* **360**, 692 (1992).
13. J. Spieth, G. Brooke, S. Kuersten, K. Lea, T. Blumenthal, *Cell* **73**, 521 (1993).
14. G. J. Hannon, P. A. Maroney, J. A. Denker, T. W. Nilsen, *ibid.* **61**, 1247 (1990); K. P. Watkins and N. Agabian, *Genes Dev.* **5**, 1859 (1991); C. Tschudi and E. Ullu, *Cell* **61**, 459 (1990); Y.-T. Yu, P. A. Maroney, T. W. Nilsen, *ibid.* **75**, 1049 (1993).
15. K. P. Watkins, J. M. Dungan, N. Agabian, *Cell* **76**, 171 (1994).
16. G. J. Hannon, P. A. Maroney, Y.-T. Yu, G. E. Hannon, T. W. Nilsen, *Science* **258**, 1775 (1992).
17. J. P. Bruzik, K. van Doren, D. Hirsh, J. A. Steitz, *Nature* **335**, 559 (1988); Z. Palfi, A. Günzl, M. Cross, A. Bindereif, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9097 (1991); K. van Doren and D. Hirsh, *Nature* **335**, 556 (1988); J. D. Thomas, R. C. Conrad, T. Blumenthal, *Cell* **54**, 533 (1988).
18. D. Saldati and J. C. Boothroyd, *Science* **260**, 349 (1993); K. Kim, D. Soldati, J. C. Boothroyd, *ibid.* **262**, 911 (1993); R. Goonewardene *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5234 (1993).