

7. J. Carvajal *et al.*, *Nature Genet.* **14**, 157 (1996).
8. Mouse and rat homologs were also found among sequences in the expressed sequence tag (EST) database.
9. Cells (DY150), mutagenized with ethylmethane sulfonate were subjected to a streptonigrin enrichment screen followed by replica plating on low-iron [BPS(5)] and high-iron [yeast extract, peptone, and dextrose (YPD)] plates as described by C. C. Askwith *et al.* [*Cell* **76**, 403 (1994)]. Low-iron medium was generated by using the iron chelator bathophenanthrolinedisulfonate (BPS). Numbers in parentheses indicate exogenous iron added back in micromolar; for example, BPS (5) contains 5 μ M iron. The *bm-8* mutant grew well on YPD but exhibited a marked growth defect on iron-restricted plates.
10. The mutant *bm-8* was transformed with a multiplicity yeast genomic library, and the plasmid pTF63/p24 was recovered by its ability to partially restore growth of *bm-8* on low-iron medium. The complementing region was narrowed by subcloning to a 1.1-kb Hind III fragment. M13 forward and reverse primers were used to sequence into the insert. The region was located on chromosome IV and shown to contain the entire open reading frame for YDL120w.
11. The complementing Hind III fragment was cloned into the centromeric vector M1188 YCP-URA pRS305, and the resulting construct was shown not to complement *bm-8* on low-iron medium. The Hind III fragment was also cloned into the yeast integrative plasmid YIP-LEU2 M1091. This construct was transformed into *bm-8*, and recombinants were selected on plates of complete minimal medium without leucine. The integration of *LEU2* next to *YFH1* was confirmed by Southern blot analysis. Recombinants were then mated to wild-type strain DY1457 and the diploids sporulated. Tetrad analysis revealed that the *LEU2* marker segregated independently of the *bm-8* phenotype of poor growth on low-iron medium.
12. *YFH1* was disrupted by insertion of a Bam HI fragment of the *HIS3* gene into the PflMI site of *YFH1*, creating plasmid MB1. A Sal I-Eco RI insert containing the disrupted *YFH1* reading frame was isolated and transformed into wild-type strain DY150. *YFH1* disruption candidates were selected on CM-his⁻ medium. Recombinants were verified by Southern blot analysis.
13. M. Babcock *et al.*, data not shown.
14. K. Nakai, *Bull. Inst. Chem. Res. Kyoto Univ.* **69**, 269 (1991).
15. A *YFH1* PCR product was generated with the added restriction sites Xba I and Cla I. The forward primer was 5'-GCTCTAGAGTGTAGCAATGATTAAGC-3', and the reverse primer was 5'-CCATCGATTTGGCTTTTGAAGATGGC-3'. The PCR product, digested with Xba I and Cla I, was cloned into the vector pGFP-C-FUS [R. K. Niedenthal *et al.*, *Yeast* **12**, 773 (1996)]. The resulting Yfh1p-GFP construct was transformed into *yfh1::HIS3*. The construct was able to complement the growth defect on YPD.
16. R. Stearman, D. S. Yuan, Y. Yamaguchi-Iwai, R. D. Klausner, A. Dancis, *Science* **271**, 1552 (1996).
17. Y. Yamaguchi-Iwai *et al.*, *EMBO J.* **14**, 1231 (1995).
18. L. Schols *et al.*, *Eur. J. Neurol.* **3**, 55 (1996).
19. E. Bonfoco *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7162 (1995).
20. J. P. Blass *et al.*, *N. Engl. J. Med.* **295**, 62 (1976); U. J. Dijkstra *et al.*, *Ann. Neurol.* **13**, 325 (1983); D. A. Stumpf *et al.*, *Neurology* **32**, 221 (1982). Several abnormalities were reported and thought to be the primary defect in FRDA. These include lipoamide dehydrogenase deficiency, pyruvate carboxylase deficiency, and mitochondrial malic enzyme deficiency. Subsequent studies did not confirm that any of these abnormalities could be the primary defect in FRDA.
21. R. O. Morgan *et al.*, *Can. J. Neurol. Sci.* **6**, 227 (1979).
22. S. Chamberlain and P. D. Lewis, *J. Neurol. Neurosurg. Psychiatry* **45**, 1136 (1982).
23. J. B. Lamarche *et al.*, in *Handbook of Cerebellar Diseases*, R. Lechtenberg, Ed. (Dekker, New York, 1993), pp. 453-457. Stainable iron was identified in cardiomyocytes from all 15 FRDA patients that were examined and none from 10 controls. No iron deposits were present in the liver, in endocrine tissues, or in the reticuloendothelial system of FRDA patients.
24. G. Barbeau, *Can. J. Neurol. Sci.* **7**, 455 (1980).
25. A. E. Harding, *Adv. Neurol.* **61**, 1 (1993).
26. M. Ben Hamida *et al.*, *Neurology* **34**, 2179 (1993).
27. S. M. Thomas *et al.*, *Biochim. Biophys. Acta* **1002**, 189 (1989).
28. A. Filla *et al.*, *Am. J. Hum. Genet.* **59**, 554 (1996); A. Dürr *et al.*, *N. Eng. J. Med.* **335**, 1169 (1996); L. Montermini *et al.*, *Ann. Neurol.* **41**, 675 (1997).
29. C. C. Askwith and J. Kaplan, *J. Biol. Chem.* **272**, 401 (1997).
30. B. S. Glick and L. A. Pons, *Methods Enzymol.* **260**, 213 (1995).
31. Supported by grants from NIH (DK49219 and DK30534 to J.K., NS34192 to M.P.), by the Muscular Dystrophy Association (M.P.), and by the Fondation Notre Dame (M.P.). D.S. was supported by a NIH postdoctoral Hematology training grant (T32DK07115), and L.M. was supported by a Medical Research Council of Canada postdoctoral fellowship.

13 March 1997; accepted 6 May 1997

Bimolecular Exon Ligation by the Human Spliceosome

Karin Anderson and Melissa J. Moore*

Intron excision is an essential step in eukaryotic gene expression, but the molecular mechanisms by which the spliceosome accurately identifies splice sites in nuclear precursors to messenger RNAs (pre-mRNAs) are not well understood. A bimolecular assay for the second step of splicing has now revealed that exon ligation by the human spliceosome does not require covalent attachment of a 3' splice site to the branch site. Furthermore, accurate definition of the 3' splice site in this system is independent of either a covalently attached polypyrimidine tract or specific 3' exon sequences. Rather, in this system 3' splice site selection apparently occurs with a 5' \rightarrow 3' directionality.

Introns in nuclear pre-mRNAs are removed by the spliceosome, a 60S complex composed of the pre-mRNA, four small nuclear RNAs (snRNAs) (U1, U2, U4-U6, and U5), and numerous associated proteins (1). Within this complex, intron excision occurs in two chemical steps, each comprising a single transesterification (2, 3). First, the 2'-OH of the branch site adenosine attacks the 5' splice site to release the 5' exon and form a lariat intron intermediate. The newly freed 3'-OH of the 5' exon then attacks the 3' splice site to join the exons

and displace the intron. However, the mechanism by which the spliceosome precisely targets these chemical reactions to bona fide splice sites in a milieu of potential cryptic sites is not clear.

Splice site recognition is especially problematic in mammals, in which genes are usually interrupted by multiple introns, often much longer than the exons, and splice site sequences are not highly conserved. The mammalian 5' splice, branch, and 3' splice (4) site consensus sequences are /GURAGU, YNYURAC, and $Y_{\geq 10}$ NYAG/, respectively (where / denotes a splice site; N, any nucleotide; R, purine; Y, pyrimidine; and underlining, the most highly conserved positions) (1, 5). The branch and 3' splice sites are usually separated by 11 to 40 nucleotides (nt) be-

cause, in mammals, the polypyrimidine tract (PPT) portion of the 3' splice site consensus is required for branch site definition (6).

Although the mechanisms that underlie recognition of the 5' splice and branch sites have been studied extensively (7), much less is known about how the exact site of exon ligation is determined (8). In budding yeast, the 3' splice site consensus is simply YAG/; there is no conserved PPT. However, even in these organisms, which have far shorter and far fewer introns than do mammals, trinucleotide recognition alone is clearly insufficient to ensure 3' splice site fidelity, given that 1 of every 32 random trinucleotides is YAG. In most introns in yeast and all introns in mammals, the site of exon ligation is the first YAG downstream of the branch site (9), which suggests that branch site proximity is a key factor in 3' splice site selection. Yet, in *Saccharomyces cerevisiae*, a second YAG downstream of the first can be the preferred exon ligation site if it is preceded by a region rich in pyrimidines (10). Thus, sequence context also contributes substantially to 3' splice site use in yeast.

The relative contributions of branch site proximity and sequence context have been more difficult to evaluate in mammalian introns because the YAG most proximal to the branch site is generally preceded by a pyrimidine-rich region: the PPT. Because the PPT normally abuts the splice site YAG, it could be inferred that an adjacent

W. M. Keck Institute for Cellular Visualization, Department of Biochemistry, Brandeis University, Waltham, MA 02254, USA.

*To whom correspondence should be addressed. E-mail: mmoores@binah.cc.brandeis.edu

PPT is a key element for 3' splice site definition in mammals. However, because the PPT is required for branch site recognition before lariat formation (6, 7), its additional contribution, if any, to defining the site of exon ligation has been difficult to determine. One study has suggested that efficient use of exon ligation sites distant from the branch site (and its attendant PPT) requires both the YAG consensus and a second, adjacent PPT (11), whereas another model posits that mammalian 3' splice sites are defined by an as yet uncharacterized "scanning" mechanism that initiates at the branch site and does not involve the PPT per se (12).

Our understanding of the recognition events that lead to the first step of splicing has been greatly advanced by mechanistic analysis of bimolecular reactions in which the 5' splice site is contributed by one molecule and the branch and 3' splice sites by a second molecule (13–16). These assays, analogous to the naturally occurring trans-splicing performed by kinetoplastids and nematodes (17), not only facilitated determination of the minimal sequences required for 5' splice site use (14, 15) but also revealed the importance of specific cis- and trans-acting elements for spliceosome assembly (14–16). In addition, they demonstrated that the 5' splice and branch sites need not be physically attached before the first step of splicing (13–16).

We now show that a parallel approach can be taken for the second step of splicing. In human splicing extracts, the 3' splice site YAG can be contributed by an RNA molecule (the 3' substrate) different from that containing the 5' splice and branch sites (the 5' substrate). For this study, the 5' substrate was a 256-nt adenovirus major late (AdML) pre-mRNA derivative containing a 129-nt 5' exon, consensus 5' splice and branch sites, and an uninterrupted 28-nt PPT, but no sequences 3' to the PPT (Fig. 1A). When incubated under splicing conditions, this RNA is assembled into splicing complexes (18) and undergoes 5' splice site cleavage and lariat formation (Fig. 1B, lanes 6 to 14). Because it lacks both a YAG consensus and an attendant 3' exon, splicing intermediates accumulate (19). However, when the reactions were supplemented with either of two RNAs containing consensus 3' splice sites (3' substrates) (Fig. 1A), substantial amounts of products accumulated with electrophoretic mobilities expected for ligated exons (Fig. 1B, lanes 15 to 22 and 24 to 27).

Both 3' substrates contained the same 14-nt PPT and CAG/3' splice site as full-length AdML pre-mRNA (Fig. 1A), but they included different 3' exons. The TNT

exon contains purine-rich splicing enhancer elements that promote spliceosome assembly through interactions with SR proteins (20, 21), whereas the AdML 3' exon contains no known enhancers (14). With both 3' substrates, bimolecular exon ligation was markedly efficient in that >50% of the 5' exons liberated by lariat formation

were converted into ligated exons. In the reactions shown, the 3' substrates were at a five-fold molar excess relative to the 5' substrate (input concentrations of 175 and 35 nM), but even at a 1:1 substrate ratio ligated exons were readily detectable with both 3' substrates (18).

Because we knew of no reports of such a

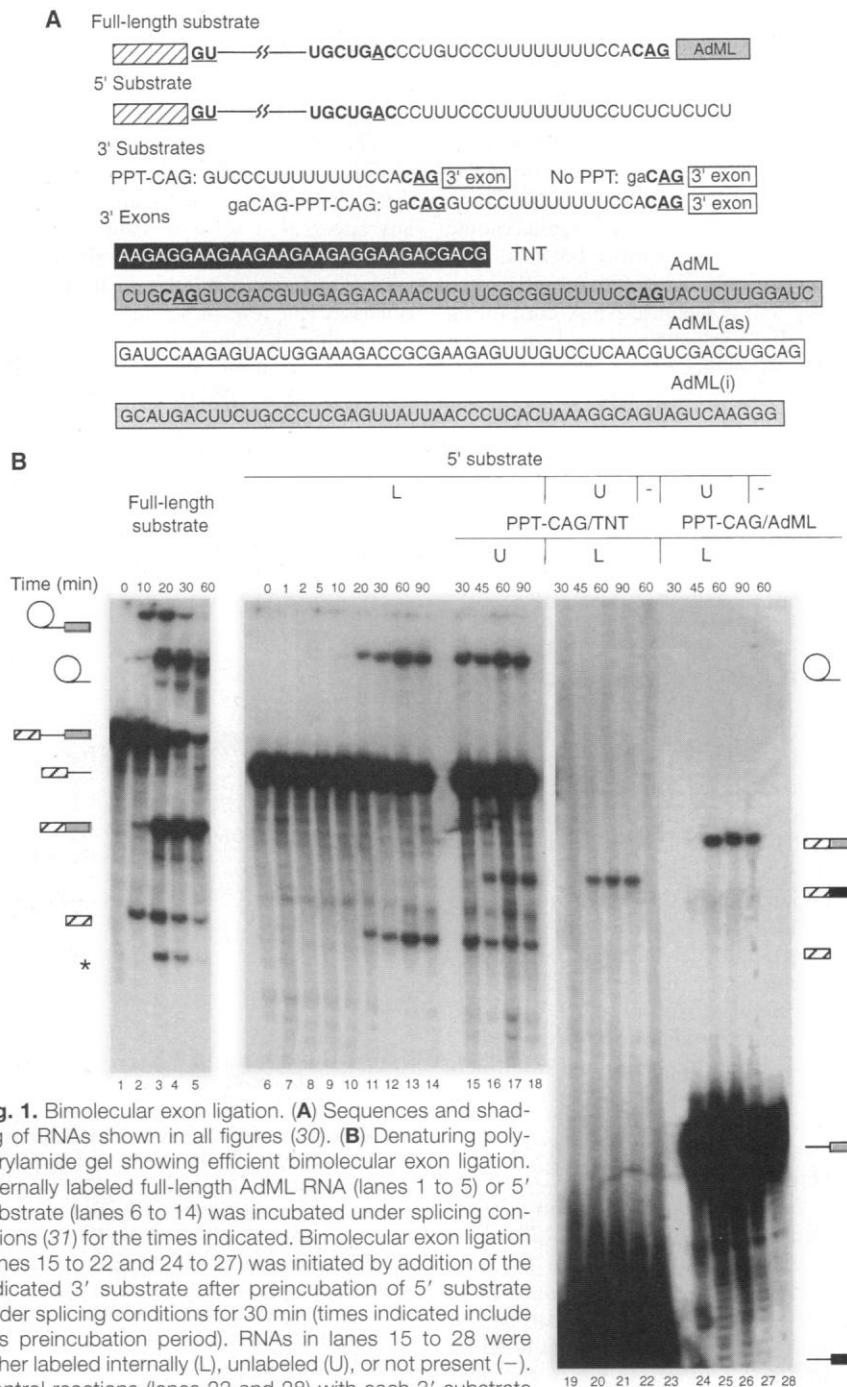


Fig. 1. Bimolecular exon ligation. (A) Sequences and shading of RNAs shown in all figures (30). (B) Denaturing polyacrylamide gel showing efficient bimolecular exon ligation. Internally labeled full-length AdML RNA (lanes 1 to 5) or 5' substrate (lanes 6 to 14) was incubated under splicing conditions (31) for the times indicated. Bimolecular exon ligation (lanes 15 to 22 and 24 to 27) was initiated by addition of the indicated 3' substrate after preincubation of 5' substrate under splicing conditions for 30 min (times indicated include this preincubation period). RNAs in lanes 15 to 28 were either labeled internally (L), unlabeled (U), or not present (-). Control reactions (lanes 23 and 28) with each 3' substrate but without 5' substrate were performed under splicing conditions for 60 min. All lanes are from the same gel; lanes 19 to 28 required longer exposure. RNA species are as depicted to either side with exon shading as in (A). The species denoted with an asterisk, detected only with full-length AdML, may be an exonucleolytic degradation product resulting from protection of this substrate at the branch point (3, 11, 32).

bimolecular exon ligation reaction for the spliceosome, we undertook several experiments to confirm the identities of the new bands in Fig. 1B. Reciprocal labeling demonstrated that both 5' and 3' substrates contributed to the products (Fig. 1B) (18). No products were obtained when either the 5' substrate (Fig. 1B, lanes 23 and 28) or 3' substrate (Fig. 1B, lanes 6 to 14) was omitted, or when the former terminated before the branch site (18). Product electrophoretic mobility varied as expected with the lengths of both the 5' (18) and 3' (Fig. 1B) (18) exons. Finally, reverse transcription polymerase chain reaction (RT-PCR)-based sequencing confirmed that the products from both 3' substrates contained the expected splice junctions (18, 22). These results indicate that the human spliceosome can mediate exon ligation between one RNA molecule that undergoes the first step internally and a separate RNA containing only a PPT, 3' splice site, and downstream exon. Therefore, accurate 3' splice site definition can occur independently of attachment or linear proximity to the branch site.

If these bimolecular reactions are catalyzed by the same active site as that which mediates exon ligation on full-length substrates, then there should be a similar requirement for the YAG consensus at the 3' splice site. To examine whether such a requirement exists, we changed the adenosine in the 3' splice site CAG of both 3' substrates to guanosine, thereby generating PPT-CGG/TNT and PPT-CGG/AdML RNAs. When PPT-CGG/TNT RNA was incubated under bimolecular splicing conditions, a product with a slightly greater electrophoretic mobility than that obtained with PPT-CAG/TNT RNA did accumulate, but at a much reduced rate (Fig. 2A). In PPT-CGG/TNT RNA, all available AG dinucleotides are preceded by a purine, which is generally detrimental to exon ligation efficiency in full-length substrates, AAG less so than GAG (11, 12). Thus, the small amount of product formed with PPT-CGG/TNT RNA likely resulted from inefficient usage of a downstream AAG as the exon ligation site.

The original PPT-CAG/AdML 3' substrate (Fig. 1A) contains three CAG triplets, but bimolecular exon ligation was apparent only at the 5'-most triplet (23). This site is analogous to that used in the full-length AdML RNA (Fig. 1B, lanes 1 to 5) in that it is the first YAG downstream of the PPT. When this site was mutated to CGG in PPT-CGG/AdML RNA, the next CAG 6 nt downstream was efficiently activated as the exon ligation site (Fig. 2, B and C; Fig. 3) (24). Mutation of the 5'-most CAG to GAG also switched the site of exon ligation to the downstream CAG, but

at a reduced efficiency (Fig. 2C). When both CAG sequences were mutated, no splicing to either site was detected (18, 25). All of these effects parallel those seen with full-length substrates, in which mutations that affect the AG dinucleotide usually shift the exon ligation site to the next YAG downstream, and a GAG triplet is not only skipped but can also reduce the usage of the downstream sites (11, 12). Taken together, these data demonstrate that bimolecular exon ligation and intramolecular splicing have the same requirements for the YAG consensus and are thus likely mediated by the same active site.

The choice of exon ligation site in this system is necessarily independent of linear distance from the branch site, as well as of any effects that sequences surrounding the 3' splice site might have on the first step (26). Thus, it was possible to investigate in isolation the role of sequence context on mammalian 3' splice site selection. Successive truncation of the PPT in the PPT-CAG/AdML 3' substrate affected neither the efficiency of bimolecular exon ligation nor the choice of exon ligation site (18). Both TNT and AdML 3' substrates containing only 5 nt (5'-gaCAG/) (27) upstream of the exon ligation site were spliced with similar efficiency (Fig. 3A, lanes 2 and 3; Fig. 3B, lanes 1 and 3) and with the same

selectivity for the upstream CAG (Fig. 2C, lanes 1 and 3; Fig. 3, A and B, lanes 1 to 4) as the original PPT-containing 3' substrates. These observations indicate that there is no inherent requirement for a covalently attached PPT for accurate 3' splice site definition and use by the human spliceosome, even though consensus sequences suggest otherwise.

Although our results showed that a covalently attached PPT is not essential, they provided no information as to whether an adjacent PPT can enhance the use of a downstream YAG relative to that of an upstream one lacking a PPT, analogous to the situation in yeast (10). We therefore constructed a 3' substrate in which the sequence 5'-gaCAG was appended to the 5' end of PPT-CAG/AdML RNA (thereby generating gaCAG-PPT-CAG/AdML RNA) (Fig. 1A). Of all possible exon ligation sites in this substrate, the 5'-most CAG was used at least 20 times as efficiently as any other site (Fig. 3B, lane 5). Mutation of this site to CGG activated the next CAG downstream (the one immediately adjacent to the PPT), but not the third CAG (Fig. 3B, lane 6). Thus, in all of the 3' substrates tested, the predominant exon ligation site was the 5'-most CAG, regardless of its position relative to any PPT.

To determine whether the sequence 5'-gaCAG was generally sufficient to specify a 3' splice site, we appended these nucleotides to the 5' termini of two other RNAs (Fig. 1A): AdML(as), the exact antisense of the AdML exon, and AdML(i), an RNA of similar length but derived from the body of

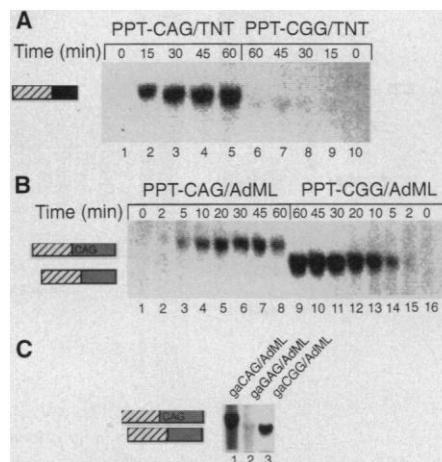


Fig. 2. Requirement of a YAG sequence for bimolecular exon ligation. Denaturing polyacrylamide gels show ligated exon products (33) from bimolecular splicing reactions with various labeled 3' substrates: (A) PPT-CAG/TNT or PPT-CGG/TNT; (B) PPT-CAG/AdML or PPT-CGG/AdML; and (C) gaCAG/AdML, gaGAG/AdML, or gaCGG/AdML. Exon ligation was initiated by addition of the indicated substrate to splicing reactions in which unlabeled 5' substrate had been preincubated for 30 min, and incubation was then continued for the times indicated in (A) and (B) and for 30 min in (C). Simultaneous controls containing each 3' substrate but lacking the 5' substrate (as in Fig. 1B, lanes 23 and 28) yielded no detectable products (18). RNAs are shaded as in Fig. 1A.

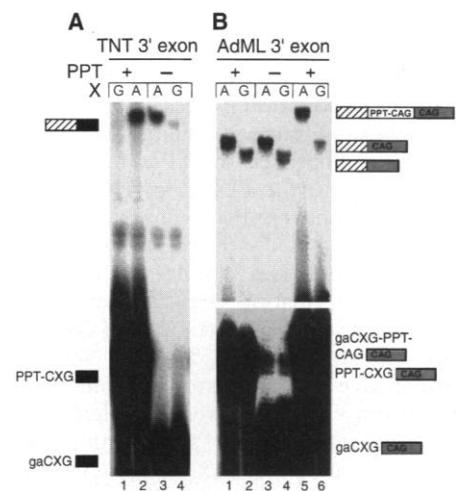


Fig. 3. Lack of requirement for a PPT in the 3' substrate. Unlabeled 5' substrate RNA was preincubated under splicing conditions for 30 min before addition of either (A) TNT or (B) AdML 3' substrates (Fig. 1A) and incubated for an additional 60 min (33). The identity of the nucleotide (A or G) at position X is as shown. The lower half of (B) is a lighter exposure than the upper half.

the AdML intron. Both the resultant RNAs were substrates for bimolecular exon ligation (Fig. 4, lanes 1 to 4 and 9 to 12), albeit at different efficiencies (28). As with the TNT and AdML 3' exons, addition of an upstream PPT to the AdML(as) RNA had no marked effect, positive or negative, on its use as an exon ligation substrate (Fig. 4, lanes 5 to 8).

Our data suggest that when exon ligation is physically uncoupled from branch formation, the only sequence necessary for recruitment of an exogenous RNA as a 3' splice donor is a suitably positioned YAG. No attached PPT or special sequences in the 3' exon are needed. Yet, in previous bimolecular assays in which the 5' splice and branch sites were on separate molecules, efficient intermolecular branch formation in HeLa splicing extracts required either an enhancer sequence or downstream 5' splice site in the 3' exon (14, 16). It is generally believed that these elements help to recruit trans-acting factors to the PPT and branch site, thereby promoting spliceosome assembly before the first chemical step (21). In the bimolecular reactions described here, however, it is the second step of splicing, not the first, that occurs intermolecularly. In this system, the 5' substrate serves as a template for spliceosome assembly and branch formation independent of any 3' splice site or 3' exon. Therefore, the minimal 3' substrates used here are unlikely to contribute to spliceosome assembly (26) and so only need the sequence elements

required for exon ligation.

Given that as few as 5 nt will suffice to specify a 3' splice site accurately, the spliceosomal factors responsible for defining the actual site of exon ligation do not require extensive interactions with upstream intronic sequences. This situation differs from mammalian branch site recognition, in which multiple sequence elements contribute (6, 7), but it parallels the minimum requirements for 5' splice site recruitment in first-step bimolecular reactions, in which only seven intronic nucleotides and no special sequences in the 5' exon (14–16) are required. However, if a YAG trinucleotide is the only sequence element needed for 3' splice site definition at the time of exon ligation, the issue of accuracy still requires clarification. Perhaps the apparent 5' → 3' directionality of 3' splice site selection in this system is important in this regard (29).

Finally, if bimolecular exon ligation can occur so readily, why was this activity not noted previously, and what prevents it from happening inappropriately *in vivo*? The answer probably lies in structural differences between the truncated 5' substrate used here and the full-length substrates encountered by the spliceosome *in vivo*. Inclusion of additional sequences downstream of the 5' substrate PPT markedly inhibits bimolecular exon ligation (18), presumably by limiting access of the 3' substrates to the YAG recognition site. Thus, on a full-length pre-mRNA, invasion of exogenous 3' exons is effectively excluded. However, the fact that bimolecular exon ligation can occur at all suggests the possibility that splicing of some transcripts *in vivo* might involve use of a YAG on a molecule separate from that containing the 5' splice and branch sites.

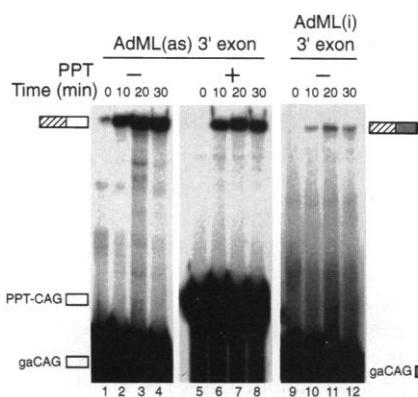


Fig. 4. The sequence 5'-gaCAG is sufficient for bimolecular 3' splice site definition. Denaturing polyacrylamide gel shows bimolecular exon ligation with AdML(as) and AdML(i) 3' substrates (shaded as in Fig. 1A). Internally labeled 3' substrate was added to splicing reactions after preincubation of unlabeled 5' substrate for 30 min, and incubation was then continued for the times indicated. The small amount of product observed in lane 1 is spillover from lane 2. Controls without 5' substrate yielded no products (18). RNA species are as indicated to either side of the figure. All lanes are from the same gel; lanes 9 to 12 required longer exposure.

REFERENCES AND NOTES

1. M. J. Moore, C. C. Query, P. A. Sharp, in *The RNA World*, R. F. Gestland and J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1993), pp. 303–357; T. W. Nilsen, *Cell* **78**, 1 (1994); A. Krämer, in *Pre-mRNA Processing*, A. I. Lamond, Ed. (Landes, Austin, TX, 1995), pp. 35–64.
2. R. A. Padgett, M. M. Konarska, P. J. Grabowski, S. F. Hardy, P. A. Sharp, *Science* **225**, 898 (1984); M. J. Moore and P. A. Sharp, *Nature* **365**, 364 (1993); K. L. Maschhoff and R. A. Padgett, *Nucleic Acids Res.* **21**, 5456 (1993).
3. B. Ruskin, A. R. Krainer, T. Maniatis, M. R. Green, *Cell* **38**, 317 (1984).
4. In this report, the term “3' splice site” refers to the site of exon ligation only, in contrast to many other papers in which this same term is used to refer collectively to all of the sequences near the 3' end of an intron (that is, the branch site, PPT, and site of exon ligation).
5. P. Senapathy, M. B. Shapiro, N. L. Harris, *Methods Enzymol.* **183**, 252 (1990).
6. R. F. Roscigno, M. Weiner, M. A. Garcia-Blanco, *J. Biol. Chem.* **268**, 11222 (1993) and references therein.
7. R. Reed, *Curr. Opin. Genet. Dev.* **6**, 215 (1996).
8. J. Umen and C. Guthrie, *RNA* **1**, 869 (1995).
9. All examples of alternative 3' splice site use in meta-

10. B. Patterson and C. Guthrie, *Cell* **64**, 181 (1991); J. G. Umen and C. Guthrie, *Genes Dev.* **9**, 855 (1995).
11. R. Reed, *Genes Dev.* **3**, 2113 (1989).
12. C. W. J. Smith, E. B. Porro, J. G. Patton, B. Nadal-Ginard, *Nature* **342**, 243 (1989); C. W. J. Smith, T. T. Chu, B. Nadal-Ginard, *Mol. Cell. Biol.* **13**, 4939 (1993).
13. D. Solnick, *Cell* **42**, 157 (1985); M. M. Konarska, R. A. Padgett, P. A. Sharp, *ibid.*, p. 165; J. P. Bruzik and T. Maniatis, *Nature* **360**, 692 (1992); A. Ghetti and J. N. Abelson, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11461 (1995); Z. Pasman and M. A. Garcia-Blanco, *Nucleic Acids Res.* **24**, 1638 (1996).
14. M. D. Chiara and R. Reed, *Nature* **375**, 510 (1995).
15. B. B. Konforti and M. A. Konarska, *RNA* **1**, 815 (1995).
16. J. P. Bruzik and T. Maniatis, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7056 (1995).
17. T. W. Nilsen, *Infect. Agents Dis.* **1**, 212 (1992); *Mol. Biochem. Parasitol.* **73**, 1 (1995).
18. K. Anderson and M. J. Moore, unpublished data.
19. Similar results have been observed for other such truncated splicing substrates [D. Frenthewey and W. Keller, *Cell* **42**, 355 (1985); B. Ruskin and M. R. Green, *Nature* **317**, 732 (1985)].
20. J. Ramchatesingh, A. M. Zahler, K. M. Neugebauer, M. B. Roth, T. A. Cooper, *Mol. Cell. Biol.* **15**, 4898 (1995).
21. X. D. Fu, *RNA* **1**, 663 (1995); A. Krämer, *Annu. Rev. Biochem.* **65**, 367 (1996); J. L. Manley and R. Tacke, *Genes Dev.* **10**, 1569 (1996); J. Valcarcel and M. R. Green, *Trends Biochem. Sci.* **21**, 296 (1996).
22. The ligated exon products from both 3' substrates were purified by electrophoresis and subjected to RT-PCR. The RT-PCR products were then cloned and sequenced. Correct splice junction usage was also observed on direct sequencing of the AdML/TNT RT-PCR product.
23. All splicing gels were quantitated with a Molecular Dynamics PhosphorImager. Even on the darkest exposures, no product was detected from exon ligation at either downstream CAG of PPT-CAG/AdML, with the limit of detection being 1/17 that of the 5'-most CAG.
24. The site of exon ligation was confirmed by RT-PCR sequencing as in (22).
25. Splicing the 3'-most CAG has not been observed. Although the reason for this is unknown, the use of this site would result in a 3' exon of only 12 nt. Several studies on 3' exon length with full-length pre-mRNA constructs have shown that terminal 3' exons containing <20 nt are inefficient second-step substrates [A. Parent, S. Zeitlin, A. Esfstratiadis, *J. Biol. Chem.* **262**, 11284 (1987); A. D. Turnbull-Ross, A. J. Else, I. C. Eperon, *Nucleic Acids Res.* **16**, 395 (1988); P. J. Furdon and R. Kole, *Mol. Cell. Biol.* **8**, 860 (1988)].
26. To date, there is no indication that any of the 3' substrates described here has any effect on the efficiency by which the 5' substrate undergoes the first step (18).
27. This sequence corresponds to the last 4 nt of the AdML intron plus a 5' terminal guanosine for transcription initiation.
28. Most of the AdML(as) sequence is part of the coding region for AdML DNA polymerase, and so is normally part of an exon, whereas the AdML(i) sequence is not exonic. Therefore, splicing efficiency in this system may correlate with the presence of exonic sequences.
29. We are currently investigating this phenomenon (K. Anderson and M. J. Moore, in preparation).
30. Plasmids pAdMLpar and pAdMLDAG were as described [O. Gozani, J. G. Patton, R. Reed, *EMBO J.* **13**, 3356 (1994)]. The sequences of these plasmids have been extensively modified compared to the original adenovirus L1-IVS1-L2 sequences from which they were derived. The beginning of the second exon of pAdMLpar, and therefore the beginning of the downstream exons of all of the AdML 3' substrates described here, differs from wild-type AdML

exon L2 in that /CUCGCG was replaced by /CUG-CAGGUCGAC, thereby creating Pst I and Sal I restriction sites in the exon. For both the full-length AdML and AdML/5'SS-BS substrates, the 5' exon, 5' splice site (†), and intron sequences up to the branch point (A) are:

```
5' -GGGAGACCGCGCAGACGACGUCGUCGUCUACUG
UCAUCUAGUGAUAUCAUCGAUGAAUUCGAGCUCGGUACC
CCGUCUCUCCUCAUCUCUCCGCAUCGUCUCGUCGAG
GGCCAGCUGUUGGG/GUGAGUACUCCUCUCAAAAGCGG
GCAUGACUUCGCCCUCGAGUUAUUAACCCUCACUAAAG
GCAGUAGUCAAGGGUUCUUGAAGCUCUUCGUCUGAC-3'
```

In Figs. 2B, 2C, 3, and 4, a truncated version of the 5' substrate was used in which nucleotides 5 to 46 of the 5' exon had been deleted. The TNT exon sequence is identical to the optimized exon 5 mutation N of chicken cardiac troponin T (20). Full-length AdML RNA was transcribed with T7 polymerase from pAdMLpar that had been linearized with Bam HI. All other RNAs were transcribed from T7 templates generated by PCR with either self-complementary primers (TNT 3' substrates) or primers complementary to either pAdMLpar (AdML 3' substrates) or pAdMLDAG (AdML/5'SS-BS RNA). RNAs were labeled internally with either [α - 32 P]ATP (adenosine triphosphate) or [α - 32 P]UTP (uridine triphosphate). All AdML/5'SS-BS and 3' substrate RNAs in Figs. 1B, 2A, 2C, and 4 contained G(5')ppp(5')G caps. In Figs. 2B, 3A, and 3B, the 3' substrate RNAs were capped with GMP (guanosine monophosphate).

31. Splicing reactions were performed at 30°C in volumes

of 5 or 10 μ l containing 40% (w/v) HeLa nuclear extract (HeLa cells were obtained from Cellex Biosciences, Minneapolis, MN), 70 to 85 mM KCl, 2 mM MgCl₂, 1 mM ATP, and 5 mM creatine phosphate. When present, the 5' substrate was 35 nM and the 3' substrates were 175 nM. RNAs were extracted and separated by denaturing polyacrylamide gel electrophoresis. All gels were subjected to autoradiography and quantitated with a Molecular Dynamics PhosphorImager.

32. A. R. Krainer, T. Maniatis, B. Ruskin, M. R. Green, *Cell* **36**, 993 (1984); J. C. S. Noble, H. Ge, M. Chaudhuri, J. L. Manley, *Mol. Cell. Biol.* **9**, 2007 (1989).
33. The slight fuzziness of the bands in Figs. 2B and 3B was likely due to heterogeneity at the 3' end of the 3' substrate. Blurred bands were observed only in experiments with the truncated version of the 5' substrate and AdML 3' substrates separated on low-percentage gels (8 to 10% polyacrylamide). The AdML 3' substrate terminates at a Bam HI restriction site (Fig. 1A), which is known to cause extensive 3' end heterogeneity in run-off transcripts [for example, see figure 2B, in M. J. Moore and P. A. Sharp, *Science* **256**, 992 (1992) and references therein].
34. We thank R. Reed for plasmids pAdMLpar and pAdMLDAG; T. Cooper and A. Zahler for cTNT sequences; and L. Davis, J. Gelles, C. Miller, C. Query, M. Rosbash, and P. Sharp for critical reading of the manuscript. This work was supported by NIH grant GM53007, a Packard Fellowship, and a Searle Scholarship.

13 September 1996; accepted 28 March 1997

Trans-kingdom Transposition of the *Drosophila* Element *mariner* Within the Protozoan *Leishmania*

Frederico J. Gueiros-Filho and Stephen M. Beverley*†

Transposable elements of the *mariner*/Tc1 family are postulated to have spread by horizontal transfer and be relatively independent of host-specific factors. This was tested by introducing the *Drosophila mauritiana* element *mariner* into the human parasite *Leishmania major*, a trypanosomatid protozoan belonging to one of the most ancient eukaryotic lineages. Transposition in *Leishmania* was efficient, occurring in more than 20 percent of random transfectants, and proceeded by the same mechanism as in *Drosophila*. Insertional inactivation of a specific gene was obtained, and a modified *mariner* element was used to select for gene fusions, establishing *mariner* as a powerful genetic tool for *Leishmania* and other organisms. These experiments demonstrate the evolutionary range of *mariner* transposition in vivo and underscore the ability of this ubiquitous DNA to parasitize the eukaryotic genome.

Transposons of the *mariner*/Tc1 family are ubiquitous elements of eukaryotic genomes, occurring in virtually every taxon examined (1–3). Phylogenetic studies of *mariner* elements have provided compelling evidence for the occurrence of horizontal transfer across species during evolution, traversing distances as far as that separating insects and flatworms (1, 2, 4). This suggested that *mariner* could

transpose independently of host-specific factors, a belief bolstered by studies of transposition activity in vitro (5). Hence, *mariner* was advanced as a potentially general tool for stable transformation and insertional mutagenesis in eukaryotic genomes after heterologous expression (2, 3, 6, 7). However, thus far this prediction has only been fulfilled in transfers among relatively closely related species within the order Diptera, as seen with the *Drosophila* elements *mariner*, *hermes*, *hobo*, and *minos* (7–9).

We decided to probe the evolutionary limits of *mariner*'s ability to transpose in vivo by introducing it into *Leishmania major*, a human pathogen belonging to the flagellate order Kinetoplastida, one of the earliest branching

eukaryotic lineages (10). Success here could also provide genetic methods to study processes of virulence and pathogenesis in leishmaniasis, a widespread tropical disease that can frequently be fatal and for which satisfactory vaccines or chemotherapy are lacking. Although methods for stable DNA transfection and expression of foreign genes are well established in *Leishmania*, nonhomologous insertion of DNA has not been observed in stable transfections of this diploid organism (11). Mobilization of *mariner* would thus provide a powerful tool for insertional mutagenesis in this pathogen.

The 1.3-kb *Mos1 mariner* element from *Drosophila mauritiana* contains a single open reading frame (ORF) encoding the transposase, flanked by 28–base pair (bp) inverted repeats (12). An intact *mariner* element was inserted in one *Leishmania* vector (pX63PAC-Mos1, Fig. 1A) (13), and a helper plasmid was used to provide transposase. *Leishmania* and other trypanosomatid protozoa synthesize mRNAs by a trans-splicing mechanism, where a 39-nucleotide mini-exon is added to the 5' end of every mRNA (14). Accordingly, the *mariner* transposase coding region was inserted in a *Leishmania* expression vector (pX63TKNEO-TPASE, Fig. 1A) (13) downstream of a trans-splice acceptor site because *Drosophila* genes lack these RNA signals. The two plasmids were then introduced into *Leishmania major* line +/Δ1 (15). *Leishmania* plasmids are maintained as stable episomes while under drug pressure (G418 and puromycin for the NEO and PAC markers, respectively) but are slowly lost during growth in the absence of selection (16).

Transfectant colonies were analyzed for *mariner* transposition first by Southern (DNA) blot hybridization (17). Despite the lack of any selection for transposition, 5 of 22 colonies (23%) showed new *mariner*-hybridizing bands (Fig. 1B). No evidence of transposition was obtained in Southern blot analysis of 52 colonies containing only pX63PAC-Mos1 (18). The *mariner* insertion site from several of the new fragments arising in the presence of transposase was obtained by inverse polymerase chain reaction (PCR) (19). Sequence analysis showed that they contained *mariner*, followed by a TA dinucleotide and sequences not present in the donor plasmid DNA (Fig. 1C) (19). Southern blot hybridizations with the new *mariner*-flanking sequences showed that they were of *Leishmania* origin (20). Moreover, their fragment size had increased by 1.3 kb in the colony that gave rise to the PCR product (20), as expected for bona fide transposition.

The frequency of *mariner* insertion into a specific locus was measured for dihydrofolate reductase–thymidylate synthase (*DHFR-TS*). The parental +/Δ1 line used in the studies above is heterozygous, having one copy of

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.

*To whom correspondence should be addressed. E-mail: beverley@borcim.wustl.edu
†Address after 1 July 1997: Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110, USA.