888–896 Nucleic Acids Research, 1997, Vol. 25, No. 4

Functional analysis of the polypyrimidine tract in pre-mRNA splicing

Candace J. Coolidge, Raymond J. Seely and James G. Patton*

Department of Molecular Biology, Box 1820 Station B, Vanderbilt University, Nashville, TN 37235, USA

Received September 17, 1996; Revised and Accepted December 18, 1996

ABSTRACT

The polypyrimidine tract is one of the important *cis*-acting sequence elements directing intron removal in pre-mRNA splicing. Progressive deletions of the polypyrimidine tract have been found to abolish correct lariat formation, spliceosome assembly and splicing. In addition, the polypyrimidine tract can alter 3′-splice site selection by promoting alternative branch site selection. However, there appears to be great flexibility in the specific sequence of a given tract. Not only the optimal composition of the polypyrimidine tract, but also the role of the tract in introns with no apparent polypyrimidine tracts or where changes in the tract are apparently harmless are uncertain. Accordingly, we have designed a series of *cis*-competition splicing constructs to test the functional competitive efficiency of a variety of systematically mutated polypyrimidine tracts. An RT/PCR assay was used to detect spliced product formation as a result of differential branch point selection dependent on direct competition between two opposing polypyrimidine tracts. We found that pyrimidine tracts containing 11 continuous uridines are the strongest pyrimidine tracts. In such cases, the position of the uridine stretch between the branch point and 3′-splice site AG is unimportant. In contrast, decreasing the continuous uridine stretch to five or six residues requires that the tract be located immediately adjacent to the AG for optimal competitive efficiency. The block to splicing with decreasing polypyrimidine tract strength is primarily prior to the first step of splicing. While lengthy continuous uridine tracts are the most competitive, tracts with decreased numbers of consecutive uridines and even tracts with alternating purine/pyrimidine residues can still function to promote branch point selection, but are far less effective competitors in 3′-splice site selection assays.

INTRODUCTION

Eukaryotic pre-mRNA splicing involves the excision of introns from nascent transcripts and the ligation of exons forming mature mRNA (reviewed in 1–4). Accurate removal of introns takes place in a two step reaction that requires several *cis*-acting sequences and *trans*-acting factors. The *trans*-acting factors that recognize the conserved *cis* sequences form a large protein–RNA complex termed the spliceosome. Five small nuclear RNAs and their associated proteins (U1, U2, U4, U5 and U6 snRNPs) constitute the major components of the spliceosome (5,6), in addition to a large number of non-snRNP proteins (7,8). Spliceosomes are dynamic structures that assemble in a step-wise fashion (for reviews see 9–11).

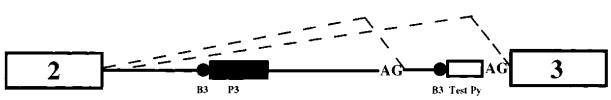
The *cis* sequences required for splicing of the major class of mammalian introns include a consensus 5′-splice site, a branch point with an adjacent polypyrimidine tract and the consensus 3′-splice site. The two dinucleotides that define the 5′- and 3′-boundaries of introns are invariant and when mutated splicing is greatly reduced or completely abolished. However, the requirement for some of the other conserved sequences appears to be quite versatile, at least in metazoans. Studies have shown that branch sites with perfect complementarity to a region of U2 snRNA are optimal for utilization (12–14). Indeed, in yeast (*Saccharomyces cerevisiae*) a strictly conserved consensus sequence surrounds the branch point (UACUA**A**C, branch point bold) that is optimal for pairing with U2 snRNA (15–17). However, in most metazoan introns, perfect complementarity does not exist and the sequence context surrounding the branch point becomes important in its utilization (12,13,18,19). In addition, metazoan recognition of branch point sequences can be affected by the adjacent polypyrimidine tract (18,20–25). Progressive deletions of the polypyrimidine tract have been found to abolish lariat formation, spliceosome assembly and splicing (18,20,26,27). Not only does the polypyrimidine tract increase the efficiency of branch point utilization (18,22,28,29), it can also function in the selection of alternative branch sites and thus 3′-splice site recognition (21,22, 30–38).

In contrast to mammalian introns, yeast have a highly conserved branch point sequence but generally lack a clear polypyrimidine tract. Nevertheless, it has been found that many yeast introns are enriched for uridines adjacent to the 3′-splice site AG, particularly at the –9 position preceding the AG (39). Increasing the number of uridines in this region greatly enhances 3′-splice site utilization in yeast (40).

Despite the important role of the polypyrimidine tract in splicing, there appears to be great flexibility in the specific sequence of a given tract. For certain substrates, the introduction of purines into the polypyrimidine tract is detrimental to splicing only if the length of the tract is shortened and if there is a reduction in the number of consecutive uridine residues (18,23,24). Also, the introduction of purines immediately downstream of the

*To whom correspondence should be addressed. Tel: +1 615 322 4738; Fax: +1 615 343 6707; Email: pattonjg@ctrvax.vanderbilt.edu

A



B

cis parent	UCUUCUCCUCCUCUCCCUCUCCC
G_{17}	GGGGGGGGGGGGGGGGG
U_{17}	UUUUUUUUUUUUUUUUU
U_{23}	UUUUUUUUUUUUUUUUUUUUUUU
C_{16}	CCCCCCCCCCCCCCCC
$(GU)_{11}$	GUGUGUGUGUGUGUGUGUGUGUU
$U_{11}G_{12}$	UUUUUUUUUUUGGGGGGGGGGGG
$G_{12}U_{11}$	GGGGGGGGGGGGUUUUUUUUUUU
U_6G_{17}	UUUUUUGGGGGGGGGGGGGGGGG
$G_{17}U_6$	GGGGGGGGGGGGGGGGGGUUUUUU
$G_{11}U_6G_6$	GGGGGGGGGGGUUUUUUGGGGGG
$G_6U_5G_{12}$	GGGGGGUUUUUGGGGGGGGGGGG
$G_6U_{11}G_{12}$	GGGGGGUUUUUUUUUUUGGGGGGGGGGGG

Figure 1. *cis*-Competition splicing substrates. (**A**) To create the *cis*-competition splicing substrates, a pre-mRNA substrate derived from α-tropomyosin (pGC+DX) was altered by placing an artificial 3′-splice site AG downstream of the original branch point/polypyrimidine tract (B3P3) of intron 2 from α-tropomyosin. The branch site was then duplicated and placed adjacent to the test pyrimidine tract. The original 3′-splice site AG immediately follows the test tract. The two possible spliced patterns are indicated by the dashed lines. (**B**) The sequences of the various test pyrimidine tracts are as shown.

branch point is apparently more detrimental than similar substitutions close to the 3′-splice site (18). In addition, uridine and cytidine do not appear to function equivalently within a polypyrimidine tract (21,23,25). Consequently, not only the optimal composition of the polypyrimidine tract, but also the role of the tract in introns where there are no recognizable tracts or where changes in the polypyrimidine tract are apparently harmless are in question. We designed a series of *cis*-competition splicing constructs to test the functional competitive efficiency of a variety of systematically mutated polypyrimidine tracts. An RT/PCR assay that detects spliced product formation as a result of direct competition between two opposing polypyrimidine tracts has been employed. Using constructs that contain variable numbers of continuous uridines, we have found that if the continuous stretch of uridines is sufficiently long, the position of the uridine stretch relative to the branch point or 3′-splice site AG is not important. However, if the continuous uridine stretch is decreased to five or six residues, such tracts are less competitive and the splicing of such substrates requires that the uridines be directly adjacent to the 3′-splice site AG. Pyrimidine tracts consisting of alternating uridines and guanosines can also support splicing, but tracts containing long continuous stretches of uridine are functionally the strongest competitors.

MATERIALS AND METHODS

Constructs

The parental pre-mRNA splicing substrate, pGC+DX, was made as described (41). The *cis*-competition constructs were prepared by altering pGC+DX as follows. An artificial 3′-splice site AG was inserted into the *Acc*I site (81 nt downstream of the B3P3 branch point) by insertion of *Afl*II linkers. Using reverse PCR techniques (42,43), a second identical branch point was then placed 12 nt downstream of the artificial AG adjacent to a second polypyrimidine tract that continues to the original 3′-splice site AG (Fig. 1). All subsequent substrates were prepared using specific oligonucleotides and reverse PCR followed by confirmatory DNA sequencing.

In vitro transcription and splicing

Transcripts for *in vitro* splicing were produced using SP6 polymerase on templates linearized at the *Bam*HI site (44,45). Transcription reactions were performed in 20 µl reactions containing 1× transcription buffer (Promega), 10 mM DTT, 0.5 mM each ATP, CTP and UTP, 0.125 mM GTP, 5 µCi [α-^{32}P]CTP (400 Ci/mmol; Amersham), 0.5 mM cap analog, 20 U SP6 RNA polymerase (Promega), 20 U rRNasin (Promega) and 1.0 µg linearized template DNA. After incubation for 1 h at 37°C, reactions were phenol/chloroform extracted and pre-mRNAs were ethanol precipitated.

In vitro splicing reactions were carried out using HeLa cell nuclear extract as described (44,45). Following splicing, reactions were digested with proteinase K, phenol/chloroform extracted and RNAs precipitated with ethanol (44,45). Splicing reactions were analyzed on 8% denaturing polyacrylamide gels or subjected to reverse transcription as described below.

RT/PCR assay

Reverse transcription reactions were carried out in 10 µl reactions containing 1× AMV reverse transcription buffer (Promega), 1 mM dNTPs, 1 mM DTT, 250 ng Pycomp 3 and 20 U rRNasin (Promega), using ~40% of the RNA from a given splicing reaction. After a 10 min pre-incubation at 65°C, 10 U AMV reverse transcriptase (Promega) were added, followed by incubation at 42°C for 2 h. Samples were then diluted 1:100 and frozen.

PCR amplification of the resultant cDNAs was performed with a labeled oligonucleotide to facilitate quantitation. Labeling of 50 ng TMX3 was carried out using 1× kinase buffer (Promega), 5 µCi [γ-^{32}P]ATP (3000 Ci/mmol, DuPont) and 5 U T4 polynucleotide kinase (Promega) for 30 min at 37°C. Labeling reactions were then heated for 2 min at 90°C and used directly in PCR reactions. PCR was performed in 50 µl reactions containing 10 mM Tris, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 0.2 mM dNTPs, 350 ng each Pycomp 2 and TMX3, 50 ng [^{32}P]TMX3, 2 µl diluted reverse transcription reaction and 5 U Taq DNA polymerase. After 20 cycles, products were separated on 3% agarose gels and dried down under vacuum at 50°C for 3 h. Products were then visualized using a PhosphorImager 445SI (Molecular Dynamics) and quantitative analysis was performed using IPLabGel (Signal Analytics).

Pycomp 3 and TMX3 are complementary to α-tropomyosin (α-TM) exon 3 and Pycomp 2 is complementary to α-TM exon 2 (Fig. 3) with the following sequences:

Pycomp 3: 5′-CCTGGGCATCTTTGAGAGCC-3′;
Pycomp 2: 5′-GGGCGTCGGAGGACGAGC-3′;
TMX3: 5′-CGAAGCTTGTATTTGTCCAGTTCATCTTC-3′.

RESULTS

Cis-competition splicing construct

To determine the functional definition of a polypyrimidine tract, we designed a splicing competition construct in which two polypyrimidine tracts were placed adjacent to identical branch

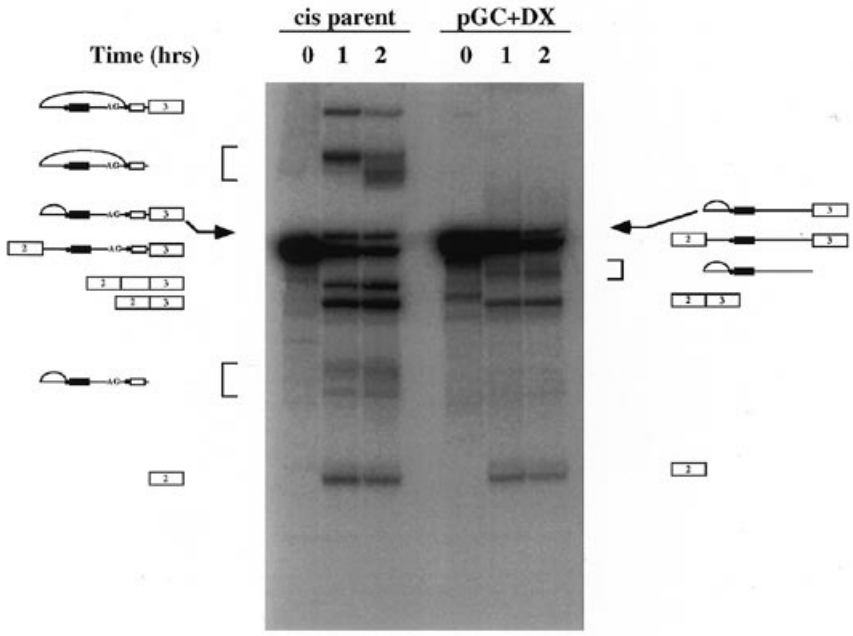


Figure 2. *In vitro* splicing of the parental substrates. ^{32}P-Labeled pre-mRNA substrates derived from pGC+DX and the *cis*-parent were spliced in HeLa nuclear extract followed by analysis on 8% polyacrylamide gels at the times indicated. The intron size of pGC+DX is 46 nt larger than that of the competition construct, so that the lariat intermediates and lariat products of splicing migrate at different positions between the two substrates (as indicated). The two possible mRNA products of splicing are as depicted.

point sequences and allowed to compete for branch point selection. The competition construct was derived from exons 2 and 3 and their intervening sequence from the rat α-TM gene. The intron between these two exons is unique in that the normal branch point is located 177 nt upstream of the 3′-splice site AG and only 41 nt downstream of the 5′-splice site of exon 2 (41). In wild-type α-TM, these exons are mutually exclusive due to the short distance between the 5′-splice site and the branch point (41). However, this inhibition can be relieved by simply increasing the distance between these elements, as has been incorporated here. Splicing experiments using such a construct (pGC+DX) have shown that once the branch point is identified, the 3′-splice site is chosen as the first AG downstream of the branch point (28,46). Insertion of upstream AG dinucleotides leads to their exclusive use, despite the presence of the normal 3′-splice site further downstream. Therefore, we inserted an upstream AG such that wild-type branch point selection would exclusively utilize this AG. We then inserted a second, identical downstream branch point with an adjacent polypyrimidine tract followed by the wild-type 3′-splice site AG (Fig. 1A). By incorporating identical branch point sequences, the choice of branch points and subsequent 3′-splice site selection is primarily due to direct competition between the two polypyrimidine tracts, determined by the relative strength of each tract, and can be monitored by the difference in size between the two spliced products. Therefore, systematically mutated polypyrimidine tracts (Fig. 1B) were placed in the downstream position (test position) and their competitive behavior was determined. Since all the test polypyrimidine tracts compete against the same upstream tract (P3), a hierarchy of usage can be determined allowing functional definition of the strength of a variety of polypyrimidine tracts.

cis-Competition of polypyrimidine tracts

As a first test of the competitive behavior of different polypyrimidine tracts, we prepared a construct containing 23 random pyrimidines in the test position (*cis*-parent; Fig. 1B). This construct was subjected to *in vitro* splicing to determine whether the test tract could compete against the P3 tract. For comparison and determination of spliced products and intermediates, the splicing of pGC+DX was performed in parallel. As shown in Figure 2, not only is the test tract of the *cis*-parent efficiently utilized, as shown by downstream branch point selection, but it actually out-competes the P3 tract. Quantitative analysis of the levels of each spliced product showed that the test tract was used ~8- to 9-fold more than the P3 tract. This was unexpected given the strength of P3 (22), but may be due to the relative positions of the two competing tracts. The distance between the upstream branch point and the first AG dinucleotide is 69 nt, while it is only 23 nt between the downstream branch point and the wild-type AG. Perhaps this distance difference allows the test tract to compete more effectively. Alternatively, since the test tract is adjacent to exon 3 of α-TM, exon sequences may increase utilization of the branch point associated with the test tract (reviewed in 10,11,47; see Discussion).

The competitive behavior of the parental test tract provided the opportunity to determine the competitive behavior of a variety of test tracts (Fig. 1). By comparing such tracts against a set competitor (P3) and in the same sequence context, variables such as the position effects just noted are eliminated and a functional hierarchy can be determined. Due to variable splicing efficiency between constructs and difficulty in quantitating spliced products from direct analysis of splicing, an RT/PCR strategy was adopted

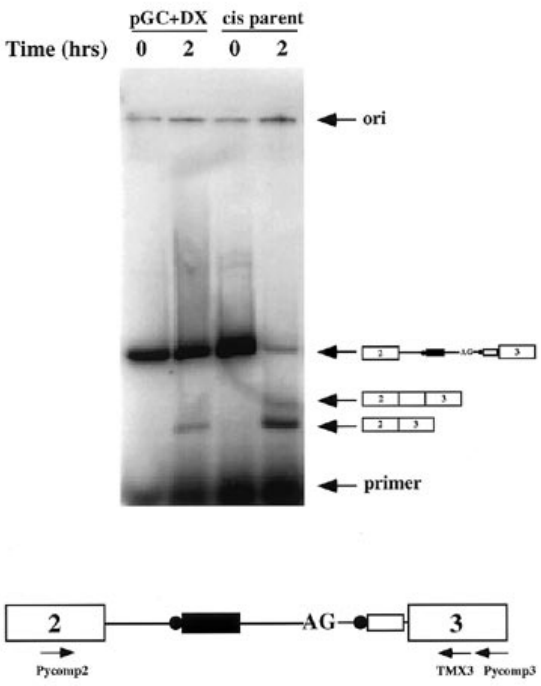


Figure 3. Spliced product analysis. *In vitro* splicing of the pGC+DX and the *cis*-parent substrates was carried out for either 0 or 2 h followed by reverse transcription with a primer complementary to exon 3 (Pycomp 3). The cDNAs were then amplified by PCR with primers complementary to exons 2 and 3 (Pycomp 2 and TMX3). The three possible PCR products are indicated on the right. The origin (ori) of electrophoresis is as marked.

to simplify quantitation. Following *in vitro* splicing, RNA was isolated and reverse transcribed using an oligonucleotide complementary to exon 3 and common to both products of splicing (Fig. 3). Competitive PCR was then used to amplify the resultant cDNAs with primers common to both spliced products so as to avoid any bias in the amplification procedure. The PCR amplification was performed with a ^{32}P-labeled primer and the products were analyzed on 3% agarose gels (Fig. 3). Three amplified bands are expected: unspliced pre-mRNA and the two spliced products corresponding to 3′-splice site selection of either the upstream AG or the downstream AG. Quantitative analysis was then performed to determine the ratio between the two products followed by comparison of all the test polypyrimidine tracts based on their competitive behavior against the common P3 tract. It is the comparison of the ratio of spliced products between constructs that allows hierarchical placement of the various test pyrimidine tracts. Reassuringly, quantitation of the ratio of spliced products using the RT/PCR assay matched the levels detected when splicing gels were directly analyzed (Fig. 2).

Functional analysis of polypyrimidine tract strength

Uridine and cytidine residues are not functionally equivalent. Test polypyrimidine tracts consisting entirely of uridine, cytidine or guanosine residues were made to determine the competitive efficiency of tracts composed of individual nucleotides (see Fig. 1). As expected, when a construct consisting of 17 guanosines (G_{17}) was placed in the test position, the upstream branch point was chosen 100% of the time, indicating that the wild-type polypyrimidine tract (P3) was preferred (Fig. 4). In contrast, constructs containing 17 and 23 uridines in the test position utilized the downstream AG ~95% of the time, whereas 16 consecutive cytidines in the test position led to use of the downstream AG only 8% of the time. Therefore, it seems clear that an effective polypyrimidine tract must contain a minimal number of pyrimidines but that uridine and cytidine do not function equivalently, as has been previously demonstrated (23,25). Due to the lack of competition with polypyrimidine tracts containing only cytidine residues (Fig. 4; 23), the remaining test polypyrimidine constructs were made using uridine as the variable pyrimidine residue in the test tracts interspersed with varying levels of guanosine. Since branch points are typically adenosines (48), we avoided their use to prevent cryptic branch point activation (49).

Tracts containing alternating pyrimidine and purine residues are functional. Since the second intron of the human apolipoprotein AII gene contains an alternating GU polypyrimidine tract (50), we wished to determine the competitive efficiency of such a tract in our system. When we used a polypyrimidine tract consisting of alternating guanosine and uridine residues $(GU)_{11}$, we found that it was chosen 61% of the time (Fig. 4). Thus, uridine in every other position within a polypyrimidine tract forms an entirely functional polypyrimidine tract, though the competitive efficiency of such a tract is only slightly higher than that of P3 and lower than seven of our 13 test tracts. Perhaps the total number of uridines (11) within the test tract of $(GU)_{11}$ promotes use of the downstream branch point despite the lack of a continuous stretch of pyrimidines.

The position of the polypyrimidine tract is more important than its length when the number of pyrimidines is limiting. A variety of pre-mRNAs, particularly those that undergo alternative splicing, have been found to utilize branch points abnormally distant from the 3′-splice site AG (33,41,51,52). In many of these introns the polypyrimidine tract adjacent to the branch point is quite long and there is often a significant enrichment for pyrimidines between the branch point and the 3′-splice site AG, especially next to the AG. It has been suggested that in introns with distant branch points, the second step of splicing proceeds more readily if a pyrimidine tract is adjacent to the 3′-splice site AG (18). However, in the case of α-TM intron 2, complete spliceosomes and the first step of splicing occur in the complete absence of the downstream pyrimidine tract and AG (28). Thus, the exact requirement for a pyrimidine tract adjacent to the 3′-splice site AG remains unclear. A goal of this study was to use a different, and hopefully more effective, approach to this question by systematically mutating the polypyrimidine tract in a *cis*-competition assay to determine not only what elements functionally compose the polypyrimidine tract and relate to its strength, but also to determine the effect of separating the polypyrimidine tract from either the branch point or the 3′-splice site AG. Therefore, we prepared several competition constructs designed to address these questions by systematically altering the uridine content of the test polypyrimidine tract. This experimental strategy allows for precise analysis of the effect of length and position of the polypyrimidine tract on its functional strength.

One of the first sets of constructs made to answer such questions contained eleven continuous uridines at various positions within the test polypyrimidine tract ($U_{11}G_{12}$, $G_{12}U_{11}$ and $G_6U_{11}G_{12}$). Each of these constructs maintain a competitive test pyrimidine tract regardless of the position of the uridine stretch relative to either the 3′-splice site or the branch point, as seen from the

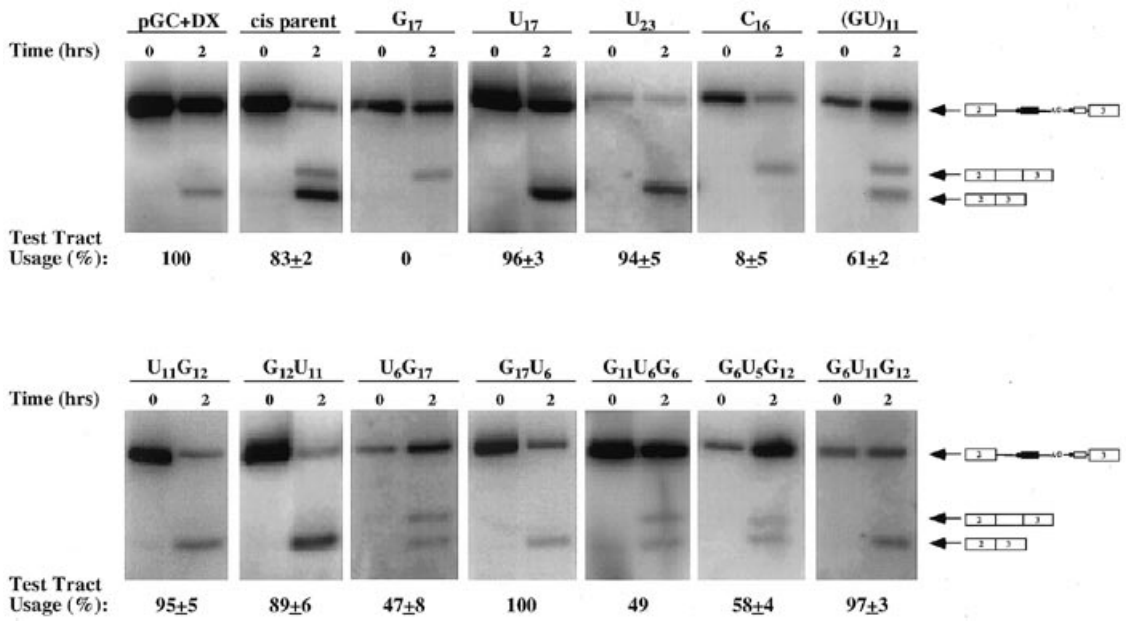


Figure 4. Quantitation of polypyrimidine tract competition constructs. Constructs containing the test polypyrimidine tracts detailed in Figure 1 were spliced*in vitro* followed by the RT/PCR assay described in Figure 3. The amounts of the two spliced products were analyzed using a PhosphorImager 445SI (Molecular Dynamics) and the ratio of the two possible spliced products was determined. The percent of test tract use is shown below each gel for the various constructs. The mean± SD of spliced product utilizing the downstream AG (i.e. downstream branch point and test polypyrimidine tract) was derived from multiple independent splicing reactions.

generally >89% usage of the test tract in all three cases (Fig. 4). However, when the number of consecutive uridines was reduced to six, as in the U_6G_{17} and $G_{17}U_6$ constructs, the frequency of downstream branch point selection became dependent upon the position of the uridines within the test tract. The $G_{17}U_6$ tract was chosen 100% of the time, whereas the U_6G_{17} tract was chosen only 47% of the time, suggesting that, as the number of consecutive uridines decreases, the location of the pyrimidine tract becomes increasingly important; short tracts immediately adjacent to the 3′-splice site are utilized more than twice as much as short tracts adjacent to the branch point. In addition, it appears that under limiting conditions the uridines must be directly adjacent to the 3′-splice site AG, since insertion of as few as six guanosines between the continuous uridine stretch and the 3′ AG resulted in decreased utilization (49%) of the branch point associated with the test tract in the $G_{11}U_6G_6$ construct. Apparently, U-rich pyrimidine tracts with consecutive uridines above a certain length function independent of position relative to the 3′-splice site AG or branch point sequence, whereas decreased uridine stretches are position dependent.

To further investigate these results, we tested two different length continuous uridine tracts sandwiched between purine blocks. When the number of consecutive uridines was 11, the downstream tract was used 97% of the time, whereas when the number of consecutive uridines was decreased to five ($G_6U_5G_{12}$), the test tract was chosen only 58% of the time. This agrees with the above results ($G_{17}U_6$ versus U_6G_{17}) in that the position of the polypyrimidine tract becomes increasingly important when the number of continuous pyrimidines is limiting but that tracts containing as few as five continuous uridines are sufficient in length to function as pyrimidine tracts, albeit at reduced levels in competitive environments. While position and length both play key roles in branch point selection, the proximity of the pyrimidine tract to the branch point does not appear to be overwhelmingly important, whereas proximity to the 3′-splice site appears to be imperative when the number of consecutive uridines is decreased.

A suboptimal polypyrimidine tract decreases the efficiency of the first step of splicing

Two classes of introns have been described based on the dependence of the 3′-splice site AG for the first step of splicing (18,39,53). Pre-mRNAs with a branch point followed by a short pyrimidine tract are typically dependent on the presence of the AG, whereas those containing a long pyrimidine tract do not require the AG to undergo the first step. In cases where the branch point is distant from the AG, sequences between the branch point and the AG, particularly a pyrimidine tract adjacent to the AG, can also affect the second step of splicing (18). To assess the effect of polypyrimidine tract composition on the first step of splicing and, subsequently, the rate of conversion of the first step to the second step, we subjected selected constructs to *in vitro* splicing and directly analyzed the lariat intermediates and spliced products. The specific constructs chosen contain two different lengths and positions of continuous uridine tracts. When six uridines were positioned adjacent to the branch point (U_6G_{17}), there was an inherently lower efficiency of the first step of splicing as compared with the amount of lariat intermediate detected when 11 uridines were adjacent to the branch point ($U_{11}G_{12}$; Fig. 5). Also, it appears that the U_6G_{17} tract is blocked after the first step of splicing, yet Figure 4 shows that roughly equivalent amounts of both spliced products are formed. When quantitated, the ratio of intermediate accumulation for the U_6G_{17} construct is equal to the ratio of product formation for the two competing events, suggesting that once the first step block is overcome, the second step of splicing proceeds rapidly and is not

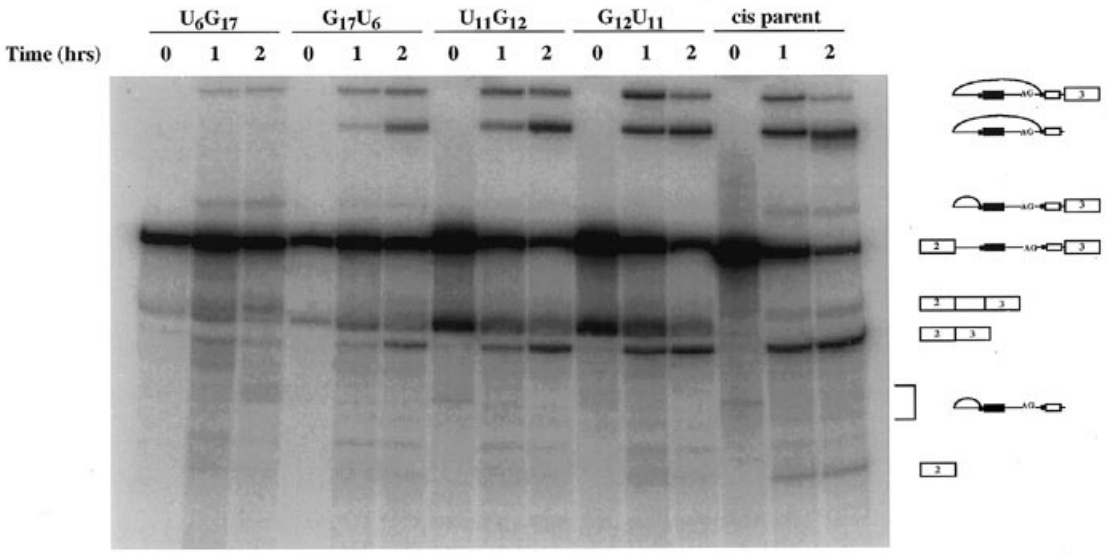


Figure 5. Analysis of the efficiency of the first and second steps of splicing. *In vitro* splicing was carried out as in Figure 2 with the indicated splicing substrates.

biased by branch point choice. When the number of consecutive uridines was increased from six to 11, the first step of splicing was much more efficient and there was no apparent block to the second step. This is consistent with results from the RT/PCR assay that showed that a continuous stretch of 11 uridines serves as a strong polypyrimidine tract, regardless of position. Decreasing numbers of continuous uridines decrease the efficiency of the first step of splicing.

DISCUSSION

To address what features of a mammalian polypyrimidine tract determine its functional strength, we have utilized a *cis*-competition assay designed to measure the ability of competing polypyrimidine tracts to promote branch point selection. All of the test tracts are placed in the exact same position relative to one another and compete against an identical polypyrimidine tract, so that the test tracts can be ordered with regard to their ability to promote branch point selection and subsequent 3′-splice site selection. The use of competition assays and systematic pyrimidine changes allowed us to confirm and extend previous experiments designed to address similar issues (18,23,25,29). Our data suggest that first, uridines are the preferred pyrimidine. Second, the proximity of the polypyrimidine tract to the 3′-splice site is most important when pyrimidines are limiting, whereas sufficiently strong polypyrimidine tracts are relatively position independent. Third, the polypyrimidine tract composition can affect the efficiency of the first step of splicing. Fourth, polypyrimidine tract strength is not determined solely by length. Lastly, a tract consisting of alternating pyrimidines and purines is functional.

There are at least two important caveats to our results. The first is the assumption that selection of an upstream branch point does not lead to selection of the downstream (second) 3′-splice site AG. If this were the case, there should be no correlation between the strength of a pyrimidine tract and branch point selection, as either AG could be selected, regardless of branch point selection. This does not appear to be the case, as there is a strong correlation between the strength of the pyrimidine tract and branch point selection. In addition, the amount of spliced product formed equals the amount of the corresponding lariat when splicing gels are directly analyzed. Finally, the first AG downstream of the branch point is almost always selected using α-TM-derived substrates (28,46).

The second possible complication with our *cis*-competition substrates concerns the possibility that secondary structures could affect splicing, especially since our test tracts could contain potential G:U base pairing arrangements. Such secondary structures placed between the branch point and the AG dinucleotide can block splicing after the first step and alter 3′-splice site selection (28,54,55). However, direct analysis of splicing gels (Figs 2 and 5 and data not shown) do not suggest such blocks to splicing with our substrates. Quantitative analysis focusing on lariat intermediate and lariat product formation completely agree with similar quantitative analysis using the RT/PCR spliced product assay, which would not be expected if there was a block to splicing at an intermediate stage due to unusual secondary structures. Thus, it appears that our *cis*-competition substrates derived from α-TM allow competitive splicing analysis free from at least these two potential difficulties.

Threshold levels of uridine determine polypyrimidine tract strength

The data shown in Figure 4 clearly show that polypyrimidine tracts with 11 continuous uridines are highly competitive pyrimidine tracts regardless of distance between the branch point and polypyrimidine tract. Limiting the number of continuous uridines to six demands that these uridines be placed immediately adjacent to the 3′-splice site AG to optimally function as a competitive pyrimidine tract. Tracts containing five or six continuous uridines can compete moderately well as long as they are positioned closer to the 3′ AG than the branch point, but are ineffective competitors if located adjacent to the branch point. By comparison of known polypyrimidine tracts, it seems that the threshold level of continuous uridines needed to allow optimal function is eight. A commonly used, efficiently spliced pre-mRNA

substrate derived from the adenovirus 2 major late promoter contains a stretch of eight continuous uridines located 4 nt from the 3′-splice site AG (56). Insertion of a single adenosine within the continuous uridine stretch leads to a near total loss of splicing and spliceosome assembly (23). Similarly, the *Drosophila sex-lethal* gene contains a regulated intron with eight continuous uridines and breaking the string leads to reduced splicing efficiency (23,25,31,57). Increasing the number of continuous uridines has also been used to increase the splicing efficiency of a variety of pre-mRNA substrates (18,23,25,40). Combining the current data with these previous results, it appears that a pyrimidine tract with eight or more continuous uridines constitutes a strong, competitive pyrimidine tract. However, introns containing less than eight continuous uridines can still maintain functional pyrimidine tracts (Fig. 4), as demonstrated by the substrate containing alternating uridines and guanosines, as well as many other substrates (48). Functional pyrimidine tracts do not absolutely require continuous uridines, but comparison with other substrates shows that increased continuous stretches of uridines increases splicing efficiency and competitiveness.

Rather than the number of consecutive uridines determining functional strength, an alternative hypothesis could be that the total number of uridines present in a given pyrimidine tract is also important in determining strength. For example, a pre-mRNA substrate derived from β-globin intron 1 is relatively efficiently spliced, yet the polypyrimidine tract does not contain any more than four continuous uridines (58). Similarly, model substrates derived from β-globin intron 1 are efficiently spliced with continuous uridine stretches of no more than three (18). However, in both cases, additional uridines (and cytidines) are found as part of the pyrimidine tract. As with the $(GU)_{11}$ substrate, it is possible that the total uridine content may relate to splicing efficiency. Thus, a continuous stretch of uridines may be optimal, but the total percentage of uridines could also determine functional competitiveness. Since strictly cytidine-containing pyrimidine tracts are apparently non-functional (23, this study), examination of the uridine content appears to be of greatest importance when designing or locating a functional pyrimidine tract. However, it should be stressed that as far as functional pyrimidine tracts are concerned, continuous uridine tracts versus total uridine content are not necessarily mutually exclusive arrangements. In a two exon, one intron substrate devoid of competition, a weak pyrimidine tract could promote branch point selection but appear non-functional under competition conditions.

Polypyrimidine tract binding proteins and 3′-splice site selection

Several RNA binding proteins have been found to preferentially bind the polypyrimidine tracts of metazoan introns, including heterogeneous ribonucleoprotein C (hnRNP C; 59), intron binding protein (60,61), polypyrimidine tract binding protein (PTB; 44,56), PTB-associated splicing factor (PSF; 45), U2 snRNP auxiliary factor (U2AF; 62,63) and the *Drosophila* splicing regulator sex-lethal (31,64–66). Genetic selection experiments designed to identify the optimal RNA binding sequence have been performed for four of these proteins (67–69). Each protein displays unique but partially overlapping pyrimidine-rich binding sites. The selected sequences for all of these proteins agree with the hypothesis that uridine content is a major determinant in pyrimidine tract strength. Since it appears that 3′-splice site selection is partly determined by competitive binding of these and perhaps other proteins, it may be that pyrimidine tracts with exceptionally strong or long U-rich tracts might bind such proteins too avidly, disallowing competition. For mammalian splicing, U2AF binds to the polpyrimidine tract with high affinity (63) and regulation of splicing is apparently allowed by having multiple proteins compete for binding to the pyrimidine tract with variable concentrations among the different U2AF competitors, particularly sex-lethal and PTB (68–71). Consistent with the competition model, it appears there is a dynamic rearrangement of polypyrimidine tract binding proteins during spliceosome assembly and during both steps of splicing (72). The dynamic rearrangement of proteins bound to the polypyrimidine tract are all consistent with an important role for the polypyrimidine tract in splicing and spliceosome assembly, consistent with our *cis*-competition results.

The ability of the $(GU)_{11}$ pyrimidine tract to compete for splicing is puzzling given the continuous uridine preference for the above-mentioned factors, particularly the essential splicing factor U2AF. All genetic selection and binding assays suggest that U2AF would not bind the $(GU)_{11}$ pyrimidine tract to enable U2 entry into the spliceosome (63,68,69,73). A possible explanation for the ability of the $(GU)_{11}$ tract to compete for splicing derives from the combinatorial nature of the *cis*-acting elements that direct 3′-splice site selection. While our experiments have attempted to isolate the pyrimidine tract and dissect its individual role in splicing, it is clear that the strength of the adjacent branch point also plays an important role in splice site selection (12,22,74). Indeed, the branch point/polypyrimidine tract is perhaps most correctly viewed as a single functional unit, with contributions from both elements determining strength (22). Thus, a weak pyrimidine tract can be offset by a strong branch point and vice versa. However, functional definition of strength does not end with combinatorial action between just these two elements. Exon enhancers must also be considered, as the presence of such enhancers can clearly rescue splicing from otherwise weak 3′-splice site signals (75–80). The presence of multiple exon enhancer sequences allows a normally weak female-specific 3′-splice site in the *Drosophila doublesex* pre-mRNA to compete for splicing against the male-specific 3′-splice site. Exon 3 of α-TM contains purine-rich sequences that may function as an enhancer (S.Mayer, personal communication), thereby promoting splicing using the downstream pyrimidine tract. This may well account for the seemingly surprising finding that many of our constructs were able to out-compete the upstream P3 tract. Proximity of the downstream tract to a possible exon enhancer could allow weaker tracts to compete more efficiently. Such combinatorial action is a possible explanation for the ability of introns with seemingly no pyrimidine tract to undergo splicing and may explain why it has been somewhat difficult to accurately define the specific sequence requirements for a strong pyrimidine tract. Consistent with such a hypothesis, equilibrium binding assays have recently shown that U2AF and PTB bind polypyrimidine tracts with very similar affinities (81). Since the nuclear concentration of PTB is much higher than U2AF, it seems likely that other factors, including U1 snRNP and SR proteins, contribute to U2AF binding (79,80,82). Thus, combinatorial action between *cis*-acting elements and the *trans*-acting factors that interact with these sequences likely enables great diversity in the functional strength of various 3′-splice sites and could account for the difficulty in assigning pyrimidine tract

strength by direct sequence analysis. Nevertheless, when such variables are held constant, it is possible to derive certain rules and preferences, as has been done here.

ACKNOWLEDGEMENTS

This work was supported by a grant from the National Institutes of Health, R01 GM50418. Phosphorimager analysis was made possible by funds provided by the National Science Foundation, BIR-9419667. R.J.S. was supported by training grant HL07751.

REFERENCES

1 Moore,M.J., Query,C.C. and Sharp,P.A. (1993) In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–358.
2 Madhani,H.D. and Guthrie,C. (1994) *Annu. Rev. Genet.*, **28**, 1–26.
3 Krämer,A. (1995) In Lamond,A.I. (ed.), *Pre-mRNA Processing*. R.G.Landes, New York, NY, pp. 35–64.
4 Adams,M.D., Rudner,D.Z. and Rio,D.R. (1996) *Curr. Opin. Cell Biol.*, **8**, 331–339.
5 Steitz,J.A., Black,D.L., Gerke,V., Parker,K.A., Krämer,A., Frendeway,D. and Keller,W. (1988) In Birnstiel,M.L. (ed.), *Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, Berlin, Germany, pp. 115–154.
6 Lührmann,R. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, Berlin, Germany, pp. 71–99.
7 Lamm,G.M. and Lamond,A.I. (1993) *Biochim. Biophys. Acta*, **1173**, 247–265.
8 Krämer,A. (1996) *Annu. Rev. Biochem.*, **65**, 367–409.
9 Rymond,B.C. and Rosbash,M. (1992) *The Molecular and Cellular Biology of the Yeast Saccharomyces: Gene Expression*, Vol. II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
10 Reed,R. (1996) *Curr. Opin. Genet. Dev.*, **6**, 215–220.
11 Black,D.L. (1995) *RNA*, **1**, 763–771.
12 Reed,R. and Maniatis,T. (1988) *Genes Dev.*, **2**, 1268–1276.
13 Zhuang,Y., Goldstein,A.M. and Weiner,A.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2752–2756.
14 Wu,J. and Manley,J.L. (1989) *Genes Dev.*, **3**, 1553–1561.
15 Pikielny,C.W., Teem,J.L. and Rosbash,M. (1983) *Cell*, **34**, 395–403.
16 Langford,C.J., Klinz,F.-J., Donath,C. and Gallwitz,D. (1984) *Cell*, **36**, 645–653.
17 Parker,R., Siliciano,P.G. and Guthrie,C. (1987) *Cell*, **49**, 229–239.
18 Reed,R. (1989) *Genes Dev.*, **3**, 2113–2123.
19 Nelson,K.K. and Green,M.R. (1989) *Genes Dev.*, **3**, 1562–1571.
20 Ruskin,B. and Green,M.R. (1985) *Nature*, **317**, 732–734.
21 Fu,X.-Y., Ge,H. and Manley,J.L. (1988) *EMBO J.*, **7**, 809–817.
22 Mullen,M.P., Smith,C.W.J., Patton,J.G. and Nadal-Ginard,B. (1991) *Genes Dev.*, **5**, 642–655.
23 Roscigno,R.F., Weiner,M. and Garcia-Blanco,M.A. (1993) *J. Biol. Chem.*, **268**, 11222–11229.
24 Norton,P.A. (1994) *Nucleic Acids Res.*, **22**, 3854–3860.
25 Bouck,J., Fu,X., Skalka,A.M. and Katz,R.A. (1995) *Mol. Cell. Biol.*, **15**, 2663–2671.
26 Frendeway,D. and Keller,W. (1985) *Cell*, **42**, 355–367.
27 Bindereif,A. and Green,M.R. (1986) *Mol. Cell. Biol.*, **6**, 2582–2592.
28 Smith,C.W.J., Porro,E.B., Patton,J.G. and Nadal-Ginard,B. (1989) *Nature*, **342**, 243–247.
29 Freyer,G.A., O'Brien,J.P. and Hurwitz,J. (1989) *J. Biol. Chem.*, **264**, 14631–14637.
30 Gallego,M.E., Balvay,L. and Brody,E. (1992) *Mol. Cell. Biol.*, **12**, 5415–5425.
31 Sosnowski,B.A., Belote,J.M. and McKeown,M. (1989) *Cell*, **58**, 449–459.
32 Helfman,D., Roscigno,R.F., Mulligan,G.J., Finn,L.A. and Weber,K.S. (1990) *Genes Dev.*, **4**, 98–110.
33 Goux-Pelletan,M., Libri,D., d'Aubenton-Carafa,Y., Fiszman,M., Brody,E. and Marie,J. (1990) *EMBO J.*, **9**, 241–249.
34 Libri,D., Goux-Pelletan,M., Brody,E. and Fiszman,M.Y. (1990) *Mol. Cell. Biol.*, **10**, 5036–5046.
35 Dominski,Z. and Kole,R. (1991) *Mol. Cell. Biol.*, **11**, 6075–6083.
36 Norton,P.A. (1994) *J. Cell Sci.*, **107**, 1–7.
37 Pret,A.M. and Fiszman,M.Y. (1996) *J. Biol. Chem.*, **271**, 11511–7.
38 Hwang,D.-Y. and Cohen,J.B. (1996) *Mol. Cell. Biol.*, **16**, 3012–3022.
39 Parker,R. and Patterson,B. (1987) In Inouye,M. and Dudock,B.S. (eds), *Molecular Biology of RNA: New Perspectives*. Academic Press, San Diego, CA, pp. 133–149.
40 Patterson,B. and Guthrie,C. (1991) *Cell*, **64**, 181–187.
41 Smith,C.W.J. and Nadal-Ginard,B. (1989) *Cell*, **56**, 749–758.
42 Imai,Y., Matshushima,Y., Sugimura,T. and Terada,M. (1991) *Nucleic Acids Res.*, **19**, 2785.
43 Coolidge,C.J. and Patton,J.G. (1995) *Biotechniques*, **18**, 763–764.
44 Patton,J.G., Mayer,S.A., Tempst,P. and Nadal-Ginard,B. (1991) *Genes Dev.*, **5**, 1237–1251.
45 Patton,J.G., Porro,E.B., Galceran,J., Tempst,P. and Nadal-Ginard,B. (1993) *Genes Dev.*, **7**, 393–406.
46 Smith,C.W.J. and Nadal-Ginard,B. (1993) *Mol. Cell. Biol.*, **13**, 4939–4952.
47 Berget,S.M. (1995) *J. Biol. Chem.*, **270**, 4411–2414.
48 Senepathy,P., Shapiro,M.B. and Harris,N.L. (1990) *Methods Enzymol.*, **183**, 252–278.
49 Ruskin,B., Greene,J.M. and Green,M.R. (1985) *Cell*, **41**, 833–844.
50 Shelley,C.S. and Baralle,F.E. (1987) *Nucleic Acids Res.*, **15**, 3787–3799.
51 Gattoni,R., Schmitt,P. and Stevenin,J. (1988) *Nucleic Acids Res.*, **16**, 2389–2408.
52 Helfman,D.M. and Ricci,W.M. (1989) *Nucleic Acids Res.*, **17**, 5633–5650.
53 Reich,C.I., VanHoy,R.W., Porter,G.L. and Wise,J. (1992) *Cell*, **69**, 1159–1169.
54 Chebli,K., Gattoni,R., Schmitt,P., Hildwein,G. and Stevenin,J. (1989) *Mol. Cell. Biol.*, **9**, 4852–4861.
55 Deshler,J.O. and Rossi,J.J. (1991) *Genes Dev.*, **5**, 1252–1263.
56 Garcia-Blanco,M.A., Jamison,S.F. and Sharp,P.A. (1989) *Genes Dev.*, **3**, 1874–1886.
57 Inoue,K., Hoshijima,K., Sakamoto,H. and Shimura,Y. (1990) *Nature*, **344**, 461–463.
58 Krainer,A.R., Maniatis,T., Ruskin,B. and Green,M.R. (1984) *Cell*, **36**, 993–1005.
59 Swanson,M.S. and Dreyfuss,G. (1988) *EMBO J.*, **7**, 3519–3529.
60 Tazi,J., Alibert,C., Temsamani,J., Reveillaud,I., Cathala,G., Brunel,C. and Jeanteur,P. (1986) *Cell*, **47**, 755–766.
61 Gerke,V. and Steitz,J.A. (1986) *Cell*, **47**, 973–984.
62 Zamore,P.D. and Green,M.R. (1989) *Proc. Natl. Acad. Sci.*, **86**, 9243–9247.
63 Zamore,P.D., Patton,J.G. and Green,M.R. (1992) *Nature*, **355**, 609–614.
64 Cline,T.W. (1978) *Genetics*, **90**, 683–698.
65 Boggs,R.T., Gregor,P., Idriss,S., Belote,J.M. and McKeown,M. (1987) *Cell*, **50**, 739–747.
66 Nagoshi,R.N., McKeown,M., Burtis,K.C., Belote,J.M. and Baker,B.S. (1988) *Cell*, **53**, 229–236.
67 Görlach,M., Burd,C.G. and Dreyfuss,G. (1994) *J. Biol. Chem.*, **269**, 23074–23078.
68 Singh,R., Valcarcel,J. and Green,M.R. (1995) *Science*, **268**, 1173–1176.
69 Perez,I., Lin,C.-H., McAfee,J.G. and Patton,J.G. (1996) submitted for publication.
70 Valcarcel,J., Singh,R., Zamore,P.D. and Green,M.R. (1993) *Nature*, **362**, 171–175.
71 Lin,C.-H. and Patton,J.G. (1995) *RNA*, **1**, 234–245.
72 Gozani,O., Patton,J.G. and Reed,R. (1994) *EMBO J.*, **13**, 3356–3367.
73 Zamore,P.D. and Green,M.R. (1991) *EMBO J.*, **10**, 207–214.
74 Noble,J.C.S., Pan,Z.-Q., Prives,C. and Manley,J.L. (1987) *Cell*, **50**, 227–236.
75 Tanaka,K., Watakabe,A. and Shimura,Y. (1994) *Mol. Cell. Biol.*, **14**, 1347–1354.
76 Xu,R., Teng,J. and Cooper,T.A. (1993) *Mol. Cell. Biol.*, **13**, 3660–3674.
77 Watakabe,A., Tanaka,K. and Shimura,Y. (1993) *Genes Dev.*, **7**, 407–418.
78 Lavigueur,A., La Branche,H., Kornblihtt,A.R. and Chabot,B. (1993) *Genes Dev.*, **7**, 2405–2417.
79 Lynch,K.W. and Maniatis,T. (1996) *Genes Dev.*, **10**, 2089–2101.
80 Wu,J.Y. and Maniatis,T. (1994) *Cell*, **75**, 1061–1070.
81 Patton,J.G., Dye,B.T., Barnard,D.C. and McAfee,J.G. (1997) In Richter,J.D. (ed.), *Analysis of mRNA Formation and Function*. Academic Press, San Diego, CA.
82 Hoffman,B.E. and Grabowski,P.J. (1992) *Genes Dev.*, **6**, 2554–2568.