4658–4665 *Nucleic Acids Research, 1997, Vol. 25, No. 22*

Both the polypyrimidine tract and the 3′ splice site function prior to the first step of splicing in fission yeast

Charles M. Romfo and Jo Ann Wise*

Department of Molecular Biology and Microbiology, Case Western Reserve University, School of Medicine, 10900 Euclid Avenue, Cleveland, OH 44106-4960, USA

Received July 3, 1997; Revised and Accepted September 25, 1997

ABSTRACT

While it is known that several *trans*-acting splicing factors are highly conserved between *Schizosaccharomyces pombe* and mammals, the roles of *cis*-acting signals have received comparatively little attention. In *Saccharomyces cerevisiae*, sequences downstream from the branch point are not required prior to the first transesterification reaction, whereas in mammals the polypyrimidine tract and, in some introns, the 3′ AG dinucleotide are critical for initial recognition of an intron. We have investigated the contribution of these two sequence elements to splicing in *S.pombe*. To determine the stage at which the polypyrimidine tract functions, we analyzed the second intron of the *cdc2* gene (*cdc*2-Int2), in which pyrimidines span the entire interval between the branch point and 3′ splice site. Our data indicate that substitution of a polypurine tract results in accumulation of linear pre-mRNA, while expanding the polypyrimidine tract enhances splicing efficiency, as in mammals. To examine the role of the AG dinucleotide in *cdc*2-Int2 splicing, we mutated the 3′ splice junction in both the wild-type and pyrimidine tract variant RNAs. These changes block the first transesterification reaction, as in a subset of mammalian introns. However, in contrast to the situation in mammals, we were unable to rescue the first step of splicing in a 3′ splice site mutant by expanding the polypyrimidine tract. Mutating the terminal G in the third intron of the *nda*3 gene (*nda*3-Int3) also blocks the first transesterification reaction, suggesting that early recognition of the 3′ splice site is a general property of fission yeast introns. Counter to earlier work with an artificial intron, it is not possible to restore the first step of splicing in *cdc*2-Int2 and *nda*3-Int3 3′ splice site mutants by introducing compensatory changes in U1 snRNA. These results highlight the diversity and probable redundancy of mechanisms for identifying the 3′ ends of introns.

INTRODUCTION

The role of the 5′ splice site is fundamentally similar in all organisms investigated, whereas the functions of the *cis*-acting elements at the 3′ ends of introns are idiosyncratic (reviewed in 1). The integrity of the branch point recognition sequence (BPS) is critical for events that precede the first transesterification reaction in both yeast and mammals, but mutations affect the assembly of different pre-splicing complexes (2–5). These two extensively investigated systems display even more radical differences in the timing and stringency of their requirements for the polypyrimidine tract and 3′ splice site. *Saccharomyces cerevisiae* introns generally lack a prominent polypyrimidine tract, consistent with the observation that small deletions and point mutations in this region have no discernible effect on splicing efficiency or accuracy (see 6 for example). On the other hand, using a sensitive *cis*-competition assay, it was possible to demonstrate that an adjacent run of uracils promotes preferential use of a particular 3′ splice site (7). Because the U-rich element is more common in introns with a long BPS to 3′ splice site distance (8) and expansion of this interval with U-poor sequences results in lariat accumulation (9), it was proposed to contribute to 3′ splice site recognition for the second step of splicing (7,8). Similarly, the 3′ AG is dispensable for the first step of splicing in *S.cerevisiae*, since mutating this dinucleotide in several different introns resulted in lariat accumulation (10 and references therein).

In vertebrate introns, the interval between the branch point and 3′ junction is most often pyrimidine-rich throughout, and mutational analysis indicates that purine substitutions in this region become detrimental when the number of consecutive uracil residues drops below a required threshold of five or six (11,12). Extensive evidence indicates that the principal function of the 3′ polypyrimidine tract in mammalian introns is to promote formation of early pre-splicing complexes (reviewed in 1), although with an appropriate construct it was possible to demonstrate a role in facilitating the second transesterification reaction (13). In contrast to the situation in *S.cerevisiae*, a subset of mammalian introns cannot proceed through the first step of splicing if the 3′ AG dinucleotide is deleted or mutated (reviewed in 1). The work of Reed (13) suggests that AG-dependent and AG-independent mammalian introns may be distinguished by the presence of a more extensive polypyrimidine tract in the latter class.

While it is clear that the basic mechanism of pre-mRNA splicing is conserved between budding yeast and mammals, the paucity of multi-intronic precursors in *S.cerevisiae* led to the proposal that the fission yeast *Schizosaccharomyces pombe* might provide a valuable model system for investigating this key cellular process (see 14,15 for example). Indeed, molecular

*To whom correspondence should be addressed. Tel: +1 216 3681876; Fax: +1 216 368 2010; Email: jaw17@po.cwru.edu

analysis has demonstrated that both the RNA and protein components of the splicing machinery share more extensive sequence and structural similarity between fission yeast and metazoa (see 16–19 for example). It is of obvious interest to determine whether the parallels between *S.pombe* and mammals extend to the organization and function of *cis*-acting signals within introns. Based on the fact that fission yeast introns generally lack an extensive run of pyrimidines between the branch point and 3′ splice site, it has been repeatedly asserted that a polypyrimidine tract is not required for splicing in this organism (15,18,20,21). However, this conclusion did not take into account the fact that the branch point to 3′ junction distance in *S.pombe* is much shorter than in the two extensively studied systems (15). Moreover, when considered as a whole, this region does show a pyrimidine bias, in some cases 100% (see Results). While the role of the polypyrimidine tract has not been experimentally investigated in fission yeast, a previous study indicated that the AG dinucleotide is required prior to the first transesterification reaction, at least for splicing of an artificial intron (22).

In the work presented here, we have assessed the contribution of sequences downstream of the branch point to splicing of natural *S.pombe* introns. Our results indicate that both the polypyrimidine tract and the 3′ AG contribute to fission yeast splicing at a stage prior to the first transesterification reaction. These observations reinforce the view that studies in this organism may shed light on the mechanisms employed by mammalian cells to deliver the 3′ splice site to the catalytic center of the spliceosome. However, while our results in general more closely parallel observations in mammalian cells than in *S.cerevisiae*, subtle differences may point to features of 3′ splice site recruitment unique to fission yeast.

MATERIALS AND METHODS

Analysis of the 3′ end architecture of *S.pombe* introns

Sequences for 203 fission yeast introns deposited in GenBank through the first quarter of 1995 were extracted and their 3′ ends analyzed for distance and sequence constraints as described in the legend to Figure 2.

Plasmid construction and mutagenesis

To express a portion of the *cdc*2 gene encompassing the second intron and surrounding exons in *S.pombe* under control of the *nmt*1 promoter, we first introduced recognition sequences for the restriction enzymes *Nde*I and *Bam*HI into the second and third exons respectively of the coding sequence (23), carried on the vector pTZ19R. Mutagenesis was carried out with reagents supplied commercially (Amersham Corp., Arlington Heights, IL) using the oligonucleotides cdc2-Nde and cdc2-Bam (see Table 1). A 271 nt *Nde*I–*Bam*HI fragment was then subcloned into the vector pREP2 (24) to generate the plasmid pREP2-cdc2. The primary transcript from this construct (Fig. 1A) contains a 206 nt 5′ exon (68 nt of *nmt*1 sequence and 138 nt of *cdc*2 sequence), a 71 nt intron and a 215 nt 3′ exon (61 nt of *cdc*2 sequence and 154 nt of *nmt*1 sequence). To analyze the role of the polypyrimidine tract in splicing of this intron, the DNA sequence corresponding to the 3′ end of the intron was modified using wild-type pREP2-cdc2 as template, unless otherwise indicated, and the following oligonucleotides, whose sequences are shown in Table 1: R-short, R-long, Y-long, M-short, M-long (template, Y-long), Y-short, Y-short+1 and Y-short+2. In addition to modifying the *cdc*2-Int2 polypyrimidine tract, most of these oligonucleotides have the potential to create one or more mutations in the 3′ splice site. The identities of all mutations were confirmed by sequencing double stranded plasmid DNA with the primer nmt1-poly(A) (Table 1) using the Sequenase enzyme and protocol (US Biochemical Co., Cleveland, OH).

Construction of pREP2-nda3, in which a transcript consisting of the third intron of the *S.pombe nda*3 gene together with flanking exon sequences is expressed under control of the *nmt*1 promoter, was described previously (25). Mutagenesis of this plasmid was performed using the oligonucleotides Y-short/nda3 and Y-long/nda3 (Table 1).

Table 1. Oligonucleotides used for directed mutagenesis and PCR

Oligonucleotide	Sequence
cdc2-Nde	5′-AGA*CATA**TG***TTGTCAGGG-3′
cdc2-Bam	5′-CAT*GGAT**CC***GAAAAAATACATG-3′
R-short/cdc2	5′-CCAAAAGT**YY**T**TTTTTTTT**GTTAGAAACAAAG-3′
R-shAG/cdc2	5′-CCAAAAGT**YY**TTTTTTTTTGTTAG-3′
R-long/cdc2	5′-GTAAAATATCCAAAAGT**YKTTTTTTTTT**TTTTTTTTGTTAGAAAC-3′
Y-long/cdc2	5′-GTAAAATATCCAAAAGT**YYA**AAAAAAAAAAAAAAAGGGTTAGAAAC-3′
M-short/cdc2	5′-ATATCCAAAAGT**YYAGC**A**GT**AT**C**GTTAGAAACAAAG-3′
M-long/cdc2	5′-ATATCCAAAAGT**YYAC**A**GT**AT**CA**G**C**AGTATATCGTTAGAAACAAAG-3′
Y-short/cdc2	5′-GTAAAATATCCAAAAGT**YYR**AAAAAAGGGTTAGAAAC-3′
Y-sh+1/cdc2	5′-GTAAAATATCCAAAAGT**YYR**AAAAAAAGGGTTAGAAAC-3′
Y-sh+2/cdc2	5′-GTAAAATATCCAAAAGT**YYA**AAAAAAAAGGGTTAGAAAC-3′
nmt1-poly(A)	5′-AAACCCTAGCAGTACTGGCAAG-3′
Y-short/nda3	5′-GAAGTTCCATGATAT**Y**TA**XZP**AAAAGTCAGTAAAAAAACAAGCAA-3′
Y-long/nda3	5′-GCCTCAGAAGTTCCATGATAT**Y**TAAAAAAAAAAAAAAGTCAGTAAAAAAACAAGCAA-3′

Restriction sites are italicized, mutant nucleotides are shown in bold and inserted nucleotides are underlined. K, 50% T, 50% G; Y, 50% T, 50% C; R, 50% A, 50% G; Z, 75% A + 25% G; P, 75% A + 25% T; X, 75% A + 25% C.

Schizosaccharomyces pombe transformation, RNA preparation and splicing assays

The plasmids described in the preceding section were introduced into *S.pombe* strain DS2 (h^+, *ade6-210*, *leu1-32*, *ura4-d18*; 21) by the lithium acetate method using ~1 µg purified plasmid DNA, which yielded ~10^5 transformants. Strains harboring *cdc*2-Int2 variants were grown to late log phase in 100 ml appropriate selective medium at 30°C and RNA prepared as described (26). To assay the effects of polypyrimidine tract substitutions on *cdc*2-Int2 splicing, primer extensions were performed on 20 µg total RNA using the primer nmt1-poly(A), labeled at its 5′ end using [γ-^{32}P]ATP and polynucleotide kinase. Reverse transcriptase reactions were carried out in 10 µl RT buffer (50 mM Tris–HCl, pH 8.3, 60 mM sodium chloride, 10 mM dithiothreitol, 20 mM magnesium acetate, 400 nM each deoxynucleoside triphosphate). After annealing 250 ng labeled primer to the RNA, 5 U AMV reverse transcriptase were added and the samples were incubated at 37°C for 5 min and then at 54°C for 20 min. Following ethanol precipitation the products were resuspended in 5 µl RNase A (2 mg/ml) and incubated at 37°C for 30 min. An equal volume of formamide loading buffer was then added to each sample and the products were resolved in 6% polyacrylamide–7 M urea gels. The relative amounts of precursor, lariat intermediate and mature mRNA were quantitated on a Molecular Dynamics PhosphorImager (data not shown). For each mutant described, primer extension splicing assays were performed multiple times on RNA prepared from at least two independent transformations; the results were highly reproducible.

RESULTS

The polypyrimidine tract of *cdc*2-Int2 is critical for efficient splicing

Given the importance of the polypyrimidine tract for splicing in mammals, we wanted to assess whether sequences between the branch point and 3′ splice site might play a similar role in early intron recognition in *S.pombe*. To this end, we performed an extensive mutational analysis of the polypyrimidine tract within the second intron of the cell division cycle gene *cdc*2 (hereafter designated *cdc*2-Int2), in which the entire interval between the branch point and 3′ splice site is spanned by pyrimidine nucleotides. Because there are four introns in the native *cdc*2 primary transcript (23), we created a single intron construct to facilitate primer extension analysis of splicing *in vivo* (Fig. 1A). As an initial test of the importance of the polypyrimidine tract for splicing of *cdc*2-Int2, we used site-directed mutagenesis (see Materials and Methods for details) to replace the eight pyrimidines with eight adenosines (R-short allele); G residues were excluded from the purine stretch to avoid creation of an AG dinucleotide that could be used as an alternative 3′ splice site. Following transformation of this construct into fission yeast, we used a primer extension assay to analyze the profile of splicing products. As demonstrated in Figure 1B, the complete substitution of purines for pyrimidines between the branch point and 3′ splice site produces a dramatic accumulation of precursor relative to wild-type *cdc*2-Int2 (cf. lanes 1 and 3), providing strong evidence that the polypyrimidine tract makes an important contribution to efficient splicing of this intron. Importantly, the unspliced RNA in the R-short mutant accumulates as linear precursor, as in mammalian cells and extracts; absolutely no

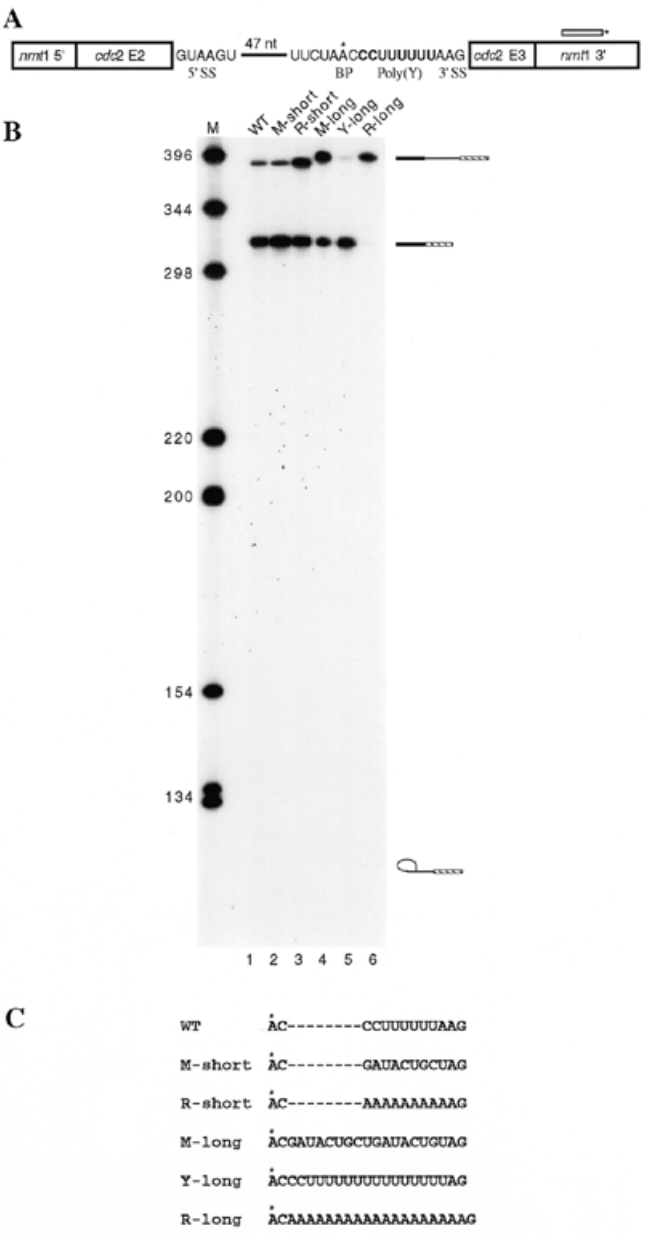

Figure 1. (**A**) Architecture of the *cdc*2-Int2 pre-mRNA. The transcripts analyzed contain the second intron of the *cdc*2 gene and flanking exons surrounded by sequences derived from the *nmt*1 (*no m*essage on *t*hiamine; 24) gene (see Materials and Methods for details of the construction). The sequences of the splicing signals are as shown, with the polypyrimidine tract indicated in bold. The box above the 3′ exon indicates the oligonucleotide used for primer extension assays of splicing. (**B**) Primer extension splicing assays on *cdc*2-Int2 mutants with altered base composition and distance between the branch point and 3′ splice site. Total RNA was isolated from *S.pombe* cells transformed with the indicated *cdc*2-Int2 allele and subjected to primer extension analysis. The predicted sizes of the cDNA products, whose positions are indicated alongside the gel, are: precursor, 388 nt; mature, 317 nt; lariat intermediate, 121 nt (see Materials and Methods). M, molecular weight markers (1 kb ladder); lane 1, wild-type (WT); lane 2, mutant in which the polypyrimidine tract is replaced by an unbiased mixture of nucleotides (M-short); lane 3, mutant in which the polypyrimidine tract is replaced by adenosines (R-short); lane 4, mutant in which the branch point to 3′ junction distance is doubled and the interval filled with an unbiased mixture of nucleotides (M-long); lane 5, mutant with a long branch point to 3′ junction distance containing exclusively pyrimidines (Y-long); lane 6, mutant with a long branch point to 3′ junction distance containing exclusively purines (R-long). (**C**) Sequences of the polypyrimidine tract variants analyzed in (B). The branch point A is indicated by an asterisk.

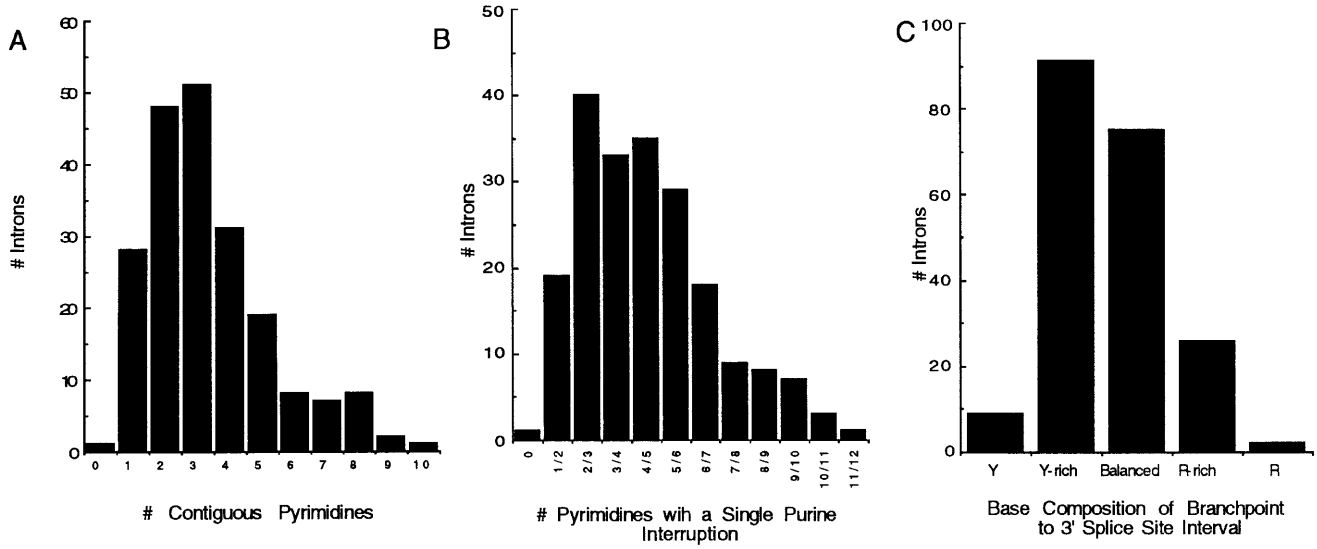

Figure 2. (**A**) Histogram displaying the size distribution of the longest continuous pyrimidine tracts at the 3′ ends of *S.pombe* introns. The terminal nucleotide of the branch point consensus (UACUAAC) and the 5′ nucleotide of the 3′ splice site consensus (UAG) were not included in the compilation. This and subsequent panels depict the results of analyzing 203 fission yeast introns (see Materials and Methods) and was generated using CricketGraph. (**B**) Histogram displaying the size distribution of polypyrimidine tracts interrupted by a single purine at the 3′ ends of *S.pombe* introns. (**C**) Histogram displaying the nucleotide composition of the branch point to 3′ splice site interval in *S.pombe* introns. Y, exclusively pyrimidine; Y-rich, >60% pyrimidine; Balanced, 40–60% pyrimidine; R-rich, >60% purine; R, exclusively purine.

signal is visible in the region of the gel where lariat intermediate should run. The formal possibility that splicing is also blocked at the second step but the lariat intermediate is selectively degraded within the nucleus cannot be positively excluded. However, there is no discernible band at the appropriate position even upon a 5-fold longer exposure of the gel shown in Figure 1B or more dramatic overexposure of other similar gels. Thus the difference between the half-lives of linear precursor and lariat intermediate would have to be several hundred-fold to account for our results. We therefore conclude that in *S.pombe*, sequences between the branch point and 3′ splice site play an important role prior to the first step of splicing.

The branch point to 3′ splice site distance in *cdc*2-Int2 (12 nt) exceeds the average for *S.pombe* by a single residue (15). To determine whether the close apposition of the 3′-terminal splicing signals is critically important, we doubled the length of the purine tract in the *cdc*2-Int2 R-short mutant to create the R-long allele. This manipulation reduced the level of mature mRNA to a nearly undetectable level (Fig. 1B, lane 6), implying that the short distance between the branch point and 3′ splice site is an important determinant of splicing efficiency in the context of a purine-rich 3′ end. A long branch point to 3′ splice site distance is not always fatal, however, since an allele with the same spacing but exclusively pyrimidine nucleotides in this interval (Y-long) is spliced very efficiently (Fig. 1B, lane 5). We elected to incorporate exclusively uridines, because the natural 3′ end of *cdc*2-Int2 is U-rich (23) and C residues were not as effective as U residues in promoting splicing of a viral intron (11). Notably, the Y-long allele accumulates virtually no pre-mRNA, in contrast to wild-type *cdc*2-Int2 (Fig. 1B, cf. lanes 1 and 5), which is all the more remarkable in that only four of the 203 introns identified to date in *S.pombe* have a longer distance between the branch point and 3′ splice site consensus sequences (J.A.Wise, unpublished observations). This observation reinforces the conclusion that a polypyrimidine tract between the branch point and 3′ splice site is critical for splicing of *cdc*2-Int2 and further implies that strengthening this element can compensate for the deleterious effect of expanding the distance (see below). Note that these experiments do not address the role of the polypyrimidine tract in 3′ splice site selection, since activation of a cryptic site is unlikely due to the fact that the next AG in the 3′ exon is 30 nt downstream of the natural 3′ splice site (23).

Further support for the notion that efficient splicing in fission yeast requires either a short branch point to 3′ junction distance or pyrimidine-rich sequences between these two signals is provided by analysis of two alleles containing random sequences in this interval (M-short and M-long). Of these, the short variant is spliced much more efficiently (Fig. 1B, cf. lanes 2 and 4). In fact, the ratio of precursor to mature mRNA for this allele, which contains alternating purines and pyrimidines but no AG dinucleotide between the branch point and original 3′ splice site (see Fig. 1C), is comparable to that of wild-type *cdc*2-Int2. Thus, as long as the 3′ signals are closely spaced, a long run of contiguous pyrimidines is not absolutely required for efficient splicing of this intron.

A potential caveat to the conclusions drawn above is that either the local or global secondary structure of the intron may be perturbed in one or more of the polypyrimidine tract variants analyzed. However, we made every effort to avoid introducing stable hairpins between the branch point and 3′ splice site, and the strong correlation between 3′ pyrimidine content and the efficiency of intron excision suggests that structural changes, if they occur, do not dramatically affect splicing.

How can our finding that the polypyrimidine tract plays a critical role in splicing of *cdc*2-Int2 be reconciled with the results of statistical analyses indicating that this feature is absent from fission yeast introns (15,20,21)? To address this question, we first examined an expanded database of fission yeast introns and found that, consisent with previous reports, most contain four or fewer contiguous pyrimidines in the short interval between the branch point and 3′ splice site (Fig. 2A). Modifying the analysis to allow a single purine interruption, as is often seen in mammalian polypyrimidine tracts, increased the pyrimidine tract length by

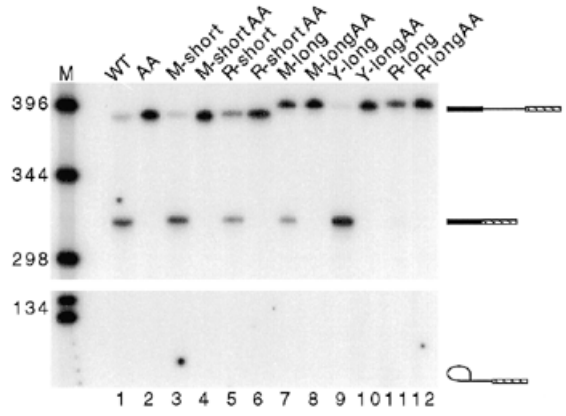

Figure 3. Primer extension splicing assays on wild-type *cdc*2-Int2 and polypyrimidine tract variants in combination with 3′ splice site mutations. RNA analysis was performed on each indicated *cdc*2-Int2 allele as described in the legend to Figure 1B. Lane 1, wild-type (WT); lane 2, mutant in which an A is substituted for the terminal G; lane 3, M-short variant; lane 4, M-short variant with a G→A change at the 3′ splice site; lane 5, R-short variant; lane 6, R-short variant with a G→A change at the 3′ splice site; lane 7, M-long variant; lane 8, M-long variant with a G→A change at the 3′ splice site; lane 9, Y-long variant; lane 10, Y-long variant with a G→A change at the 3′ splice site; lane 11, R-long variant; lane 12, R-long variant with a G→A change at the 3′ splice site. To conserve space, the lower 2/3 of this and several subsequent gels is not shown except for a small slice surrounding the position where the lariat should run.

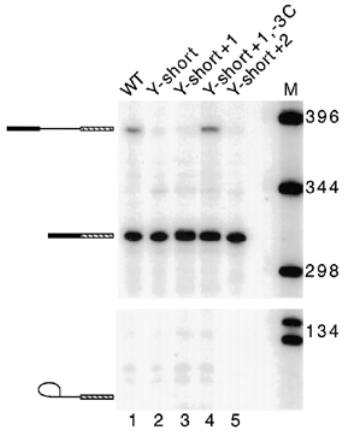

Figure 4. Primer extension splicing assays on *cdc*2-Int2 transcripts with mutations at position –3. RNA preparation and analysis was performed as described in the legend to Figure 1B. Lane 1, wild-type (WT); lane 2, mutant in which the A at –3 was changed to U (Y-short); lane 3, –3U mutant with a single uridine insertion in the poly(Y) tract (Y-short+1); lane 4, –3C mutant with a single uridine insertion in the poly(Y) tract (Y-short+1/–3C); lane 5, –3U mutant with a two uridine insertion in the poly(Y) tract (Y-short+2/–3U); M, molecular weight markers (1 kb ladder).

only a single nucleotide (Fig. 2B). On the other hand, a recent study indicates that, even in mammals, a very short (5–6 nt) polypyrimidine tract can support efficient splicing if it is located immediately adjacent to the 3′ splice site (12). Finally, examination of the base composition of the branch point to 3′ splice site interval in fission yeast in its entirety indicates that: (i) the vast majority of fission yeast introns contain either a balanced purine:pyrimidine ratio or are biased toward pyrimidines; (ii) eight introns in addition to *cdc*2-Int2 contain exclusively pyrimidine nucleotides in this region; and (iii) purine-rich sequences are very rare (Fig. 2C). Thus the effects on splicing of the mutations we have introduced into *cdc*2-Int2 parallel the natural sequence distribution at the 3′ ends of fission yeast introns.

*cdc*2-Int2 requires an intact 3′ splice site to proceed through the first transesterification reaction

We next set out to determine whether the first step of splicing in *cdc*2-Int2 depends on the integrity of the 3′ splice site. To address this question, we replaced the invariant guanosine of the 3′ splice site in each of our alleles with adenosine. As demonstrated in Figure 3, this substitution in an otherwise wild-type intron produces accumulation of exclusively linear precursor RNA (cf. lanes 1 and 2); absolutely no lariat intermediate is visible even upon prolonged exposure of this or other similar gels (data not shown). We then tested the effect of mutating the 3′ splice site of the Y-long variant, in which the polypyrimidine tract is double the length of that in wild-type *cdc*2-Int2. Again, the only species that accumulates is linear precursor (Fig. 3, cf. lanes 9 and 10). The first step of splicing is also completely blocked for three other 3′ splice site mutants (CG, GG and GA) in combination with this extended polypyrimidine tract (data not shown). Finally, we examined the effect of mutating the 3′ splice site in the remaining polypyrimidine tract variants (Fig. 3, cf. lanes 3, 5, 7 and 11 with 4, 6, 8 and 12; data not shown). As for the wild-type and Y-long alleles, in no case is lariat intermediate detectable. In aggregate, these data conclusively demonstrate that splicing of *cdc*2-Int2 is blocked prior to the first transesterification reaction by 3′ splice site mutations regardless of the length of its polypyrimidine tract.

The identity of the nucleotide at position –3 dramatically affects *cdc*2-Int2 splicing efficiency

An unusual feature of *cdc*2-Int2 is the presence of an A at position –3 (Fig. 1A), which is a U in the vast majority of fission yeast introns. The identity of this nucleotide is known to be of critical importance not only in nematodes, where it is virtually invariant (27), but also in vertebrates, where it is more variable (28). To determine the extent to which the A at position –3 influences *cdc*2-Int2 splicing efficiency, we mutated it to a U in an otherwise wild-type allele (designated Y-short). Remarkably, as demonstrated in Figure 4, this change virtually eliminates the precursor accumulation observed with wild-type *cdc*2-Int2 (cf. lanes 1 and 2). Two –3U variants that contain polypyrimidine tracts extended by one or two uridines (Y-short+1 and Y-short+2) are also spliced with very high efficiency (Fig. 4, lanes 3 and 5). The dramatically improved splicing of an allele in which the only change is at position –3 raises the possibility that this substitution alone might be sufficient to account for the effect of expanding the polypyrimidine tract in the Y-long allele (Fig. 1B, lane 5). However, the relatively poor splicing of the M-long variant (Fig. 1B, lane 4) indicates that the distance between the branch point and 3′ splice site also plays a major role in determining splicing efficiency in *S.pombe*.

The strong bias toward U at position –3 in fission yeast introns contrasts both with mammals, where C is the predominant nucleotide (reviewed in 29), and *S.cerevisiae*, where C and U appear with nearly the same frequency (reviewed in 30). To determine the functional relevance of this difference, we mutated position –3 in the Y-short+1 allele to a C. This change produces significant precursor accumulation relative to an allele with the same spacing, but the consensus nucleotide at the antepenultimate

position (Fig. 4, cf. lanes 3 and 4). Thus the effects of –3 substitutions on *cdc*2-Int2 splicing parallel the natural distribution at this nucleotide in fission yeast introns. Notably, as with the polypyrimidine tract mutants, the block in splicing of the –3C allele occurs prior to the first transesterification reaction.

Splicing of the third intron in the β-tubulin gene is also AG-dependent

The work presented here as well as other studies from our laboratory indicate that inefficient splicing of *cdc*2-Int2 is due to features other than its suboptimal polypyrimidine tract, including its large size for this organism (71 nt) and the presence of a strong secondary structure at the 5′ splice site (25; C.J.Alvarez, C.M.Romfo and J.A.Wise, in preparation). These observations raised the possibility that the requirement for an intact 3′ splice site prior to the first transesterification reaction in this intron might reflect poor recognition of the other splicing signals during pre-spliceosome assembly (see Discussion). We therefore wanted to determine whether splicing of a more typical *S.pombe* intron would also require an intact 3′ AG for the first transesterification reaction. To this end, we examined the effect of mutating the 3′ splice site in the third intron of the gene encoding β-tubulin, *nda*3 (hereafter designated *nda*3-Int3), whose length (44 nt) corresponds to the peak of a histogram displaying intron length in this organism (15; R.VanHoy and J.A.Wise, unpublished data). In contrast to *cdc*2-Int2, this intron is spliced very efficiently when expressed from the *nmt*1 promoter (25; Fig. 5, lane 1). Upon mutating the 3′ splice site of an otherwise wild-type *nda*3-Int3, we observe accumulation of linear pre-mRNA, not lariat intermediate (Fig. 5, lane 4), i.e. splicing of this intron can also be classified as AG-dependent. Because *nda*3-Int3 is also typical of *S.pombe* introns in that the interval between its branch point and 3′ splice site is not composed exclusively of pyrimidines (31), we next tested the effect of converting the two purines in this region to uridines (Y-short allele). This variant is spliced with wild-type efficiency (Fig. 5, lane 2) and also shows a first-step block upon mutating its 3′ splice site (lane 5). The same behavior is observed for a variant in which the length of the *nda*3-Int3 polypyrimidine tract is doubled (Y-long, lane 6). The efficient splicing of this allele is notable in that over 1/3 of the intron is located between the branch point and terminal G. Because expansion of the polypyrimidine tract in *nda*3-Int3 does not relieve the first step block imposed by mutations at the 3′ splice site (Fig. 5, lane 6), we conclude that the AG-dependence of *cdc*2-Int2 (see above) is not due to its anomalous features, but rather is likely to reflect a general feature of *S.pombe* introns.

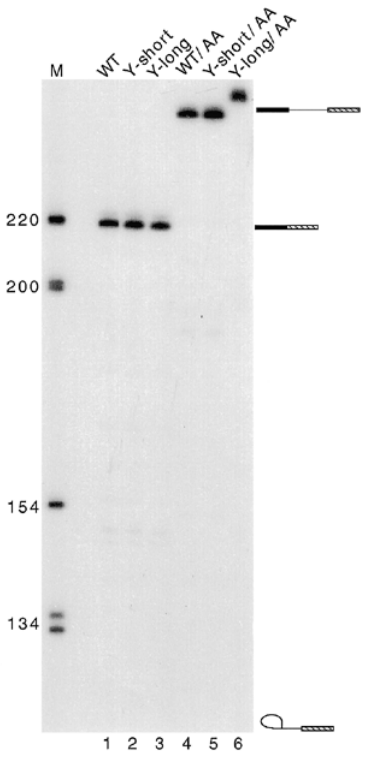

Figure 5. Primer extension splicing assay on wild-type *nda*3-Int3 and polypyrimidine tract variants with and without 3′ splice site mutations. RNA preparation and analysis was performed as described in the legend to Figure 1B. The predicted sizes of the cDNA products derived from *nda*3 RNA species are: precursor, 261 nt; mature, 220 nt; lariat intermediate, 116 nt (25). Lane 1, wild-type *nda*3-Int3 (WT); lane 2, mutant in which the two purine nucleotides between the branch point and 3′ junction were changed to uridines; lane 3, mutant in which the length of the polypyrimidine tract was doubled; lanes 4–6, same alleles assayed in lanes 1–3 but with an additional G→A mutation in each case at the terminal intron nucleotide.

The first step of splicing cannot be rescued in *cdc*2-Int2 and *nda*3-Int3 3′ splice site mutants by compensatory alleles of U1 snRNA

We previously used genetic suppression experiments with an AG-dependent artificial intron to provide evidence that the 3′ splice site is recognized at an early stage of spliceosome assembly by base pairing to the invariant U1 dinucleotide (C7U8), which lies just downstream of the 5′ splice site recognition sequence (22). Because *cdc*2-Int2 meets the operational definition of AG dependence despite the presence of a very strong (for *S.pombe*) polypyrimidine tract, an obvious experiment was to test whether the first step block imposed by mutating its 3′ terminal G to an A could be relieved by introducing a compensatory base change in U1 snRNA (C7U). As shown in Figure 6A, lane 3, restoring complementarity between U1 snRNA and the mutant 3′ splice site does not allow production of a detectable level of lariat intermediate. However, because we had previously observed that an extra copy of even the wild-type U1 gene inhibits splicing of this intron (25), it was conceivable that the lack of first step rescue was due to the exquisite sensitivity of *cdc*2-Int2 to the levels of this snRNA.

To ascertain whether a compensatory change in U1 could rescue the first step of splicing for a 3′ splice site mutant of an intron comparable in size to the artificial intron previously examined, we turned to *nda*3-Int3. In addition to having a length typical of natural *S.pombe* introns, our earlier work demonstrated that splicing of *nda*3-Int3 is not affected by the presence of a plasmid-borne snRNA gene (25). As shown in Figure 6B, lane 3, restoring the potential for pairing between U1 and the 3′ splice site also does not produce detectable levels of lariat intermediate for this intron. Control experiments show that non-compensatory changes in U1 snRNA have no effect on the product profile for a 3′ splice site mutant of either intron (cf. Fig. 6A and B, lanes 2 and 4). Finally, we also tested the effect of a compensatory change in U1 snRNA on splicing of two polypyrimidine tract variants for each intron. As for the wild-type introns, we do not observe production of lariat intermediate (Fig. 6A and B, lanes 7 and 11). We conclude that base pairing between U1

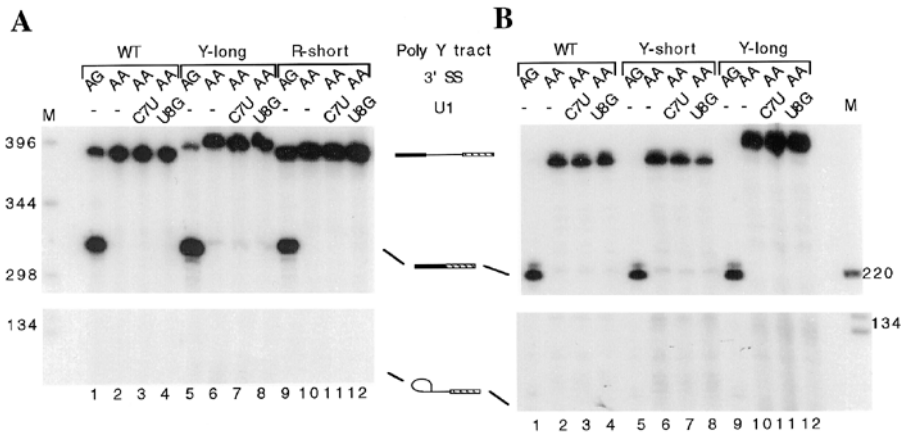

Figure 6. (**A**) Primer extension splicing assays of *cdc*2-Int2 3′ splice site mutants in the presence of mutant alleles of U1 snRNA. RNA preparation and analysis was performed as described in the legend to Figure 1B. Lane 1, wild-type (WT) *cdc*2-Int2; lane 2, *cdc*2-Int2 3′ splice site mutant; lane 3, 3′ splice site mutant in combination with a compensatory U1 allele; lane 4, 3′ splice site mutant in combination with a non-compensatory U1 allele; lanes 5–8, as in lanes 1–4 but using the Y-long allele of *cdc*2-Int2, in which the length of the polypyrimidine tract is doubled; lanes 9–12, as in lanes 1–4 but using the R-short allele of *cdc*2-Int2, in which purines replace the pyrimidines between the branch point and 3′ splice site. (**B**) Primer extension splicing assays of *nda*3-Int3 3′ splice site mutants in the presence of mutant alleles of U1 snRNA. Lane 1, wild-type (WT) *nda*3-Int3; lane 2, *nda*3-Int3 3′ splice site mutant; lane 3, 3′ splice site mutant in combination with a compensatory U1 allele; lane 4, 3′ splice site mutant in combination with a non-compensatory U1 allele; lanes 5–8, as in lanes 1–4 but using the Y-short allele of *nda*3-Int3, which contains exclusively pyrimidines between the branch point and 3′ splice site; lanes 9–12, as in lanes 1–4 but using the Y-long allele of *nda*3-Int3, in which the length of the polypyrimidine tract is doubled.

and the 3′ splice site is not sufficient to allow either *cdc*2-Int2 or *nda*3-Int3 to proceed through the first step of splicing.

DISCUSSION

The most notable conclusion to emerge from the work presented here is that a polypyrimidine tract located between the branch point and 3′ splice site of an *S.pombe* intron functions in a manner that is fundamentally similar to its counterpart in mammalian introns. In particular, our demonstration that this element promotes the first transesterification reaction echoes results from a number of studies that employed mammalian cell extracts (see 10–12 for example) and contrasts with data from budding yeast, where recognition of the 3′ run of U residues contributes principally to the second step of splicing (8,9). More recently, polypyrimidine tract function has been investigated in *Drosophila* and *Caenorhabditis elegans*, two organisms in which a significant fraction of the introns are small and lack extensive runs of pyrimidines between the branch point and 3′ splice site (32–34), similar to the vast majority of the introns in *S.pombe*. Based on statistical analysis of information content (34), it was suggested that a polypyrimidine tract would be required for splicing of large but not small fly introns. However, counter to this notion, Kennedy and Berget (35) demonstrated that mutations in pyrimidine-rich elements reduce or eliminate splicing of the small *mle* intron. Interestingly, in wild-type *mle*, polypyrimidine tract function appears to be supplied by C-rich sequences located between the 5′ splice site and the branch point, in contrast to the more traditional placement of the corresponding element in fission yeast *cdc*2-Int2. Notably, the upstream pyrimidine tracts in *mle* can be functionally replaced by a U-rich sequence between the branch point and 3′ splice site. Mutating the pyrimidine-rich elements blocks the first transesterification reaction in *Drosophila*, further underscoring the parallels with fission yeast.

In worms, a highly conserved sequence (UUU) located immediately upstream of the 3′ splice site (CAG) is critical for splicing and was proposed to function in a capacity similar to the longer but more variable polypyrimidine tracts found in other organisms (27), including *S.pombe*. A particularly intriguing aspect of polypyrimidine tract function in *C.elegans* is that this element can direct accurate and efficient 3′ splice site selection even when the terminal G of the intron is mutated to an A (27 and references therein). This is in striking contrast to our observation that the corresponding change in either *cdc*2-Int2 or *nda*3-Int3 of *S.pombe* completely blocks splicing even in the presence of an improved 3′ pyrimidine run. The inability of an expanded polypyrimidine tract to bypass the requirement for a 3′ AG prior to the first step of splicing provides an exception to the parallel between early intron recognition in mammals and fission yeast. We believe that this difference is secondary to the similarities but likely to be meaningful, since mammalian introns with fairly modest pyrimidine runs are partially AG-independent (36,37).

The stringent and apparently general requirement for recognition of the 3′ AG prior to any of the chemical events in splicing may be related to the unexpected observation that all 5′ splice site mutations block the first step in *S.pombe* (C.J.Alvarez and J.A.Wise, in preparation). In other organisms, ranging from *S.cerevisiae* to mammals and nematodes, G1A and U2A/G substitutions produce exclusively lariat intermediate accumulation. We believe that the early splicing block imposed by any change in a *cis*-acting signal may reflect the existence of a mechanism in fission yeast for initial recognition of the intron as a unit. This proposal is consistent with the extremely strong preference for selecting the pair of splice junctions that are in closest proximity (C.J.Alvarez, C.M.Romfo and J.A.Wise, in preparation). The contrast between our results and data from other organisms highlight the evolutionary diversity of mechanisms for defining intron termini. In all likelihood, the functions of 3′ splicing signals are at least partially overlapping, with different elements playing prominent roles at distinct stages of the pathway in different organisms.

Of key importance to deciphering the idiosyncratic roles of the *cis*-acting elements at the 3′ ends of introns is identification and characterization of the *trans*-acting factors responsible for their recognition. In contrast to the nucleotides surrounding the sites of the two transesterification reactions, which are recognized at least in part through base pairing interactions with snRNAs, the polypyrimidine tract appears to function solely as a protein binding site (reviewed in 1). In mammals, this element is recognized at a very early stage of the splicing pathway by the large subunit of the heterodimeric splicing factor U2AF (U2AF^{65}), which plays a prominent role in promoting correct identification of the poorly conserved branch site sequence (38). Although budding yeast contains an ortholog, Mud2p, this protein shows rather weak similarity to its human counterpart and is not required for viability (39). In *S.cerevisiae*, the U-rich element preceding the 3′ splice site is recognized by a different protein, Prp8p, only after the first transesterification reaction (40). We have recently analyzed the contribution of fission yeast U2AF^{65}, which shares extensive sequence similarity with its human counterpart and is required for viability (18), to splicing of a variety of fission yeast introns (C.M.Romfo and J.A.Wise, in preparation). Our data provide evidence that this factor is required for splicing of all pre-mRNAs in *S.pombe*, including those with purine-rich 3′ ends, and suggest that introns such as *cdc*2-Int2, whose splicing is inherently inefficient, are under selective pressure to maintain a strong polypyrimidine tract (i.e. a good U2AF^{65} binding site).

The factors responsible for early recognition of 3′ splice sites are less well defined. Several years ago, based on experiments conducted on an artificial intron, we proposed that U1 snRNA contributes to this process via base pairing to the terminal AG (22). It was clear from the outset that this interaction was not necessary for all splicing events, since all introns in *S.cerevisiae* and a subset of those in mammals could proceed through the first transesterification reaction when the 3′ splice site was mutated or even deleted (10,13). The results from the present study indicate that, although the requirement for an intact 3′ AG to proceed through the first step of splicing appears to be general in *S.pombe*, restoring potential base pairing between U1 and the 3′ splice site does not promote the first transesterification reaction in *cdc*2-Int2 or *nda*3-Int3. The most likely interpretation is that the 3′ splice site of these natural introns must be recognized at an early stage of spliceosome assembly by a factor other than or in addition to U1 snRNA. What could be the identity of this component? One candidate might be a fission yeast homolog of the 72 kDa protein recently shown to crosslink to the 3′ splice site at an early step in spliceosome assembly in a mammalian extract (41). Alternatively, the AG dinucleotide may be part of the recognition sequence for U2AF^{65}. This possibility was suggested on the basis of experiments with a crude preparation of the protein (42), but has not been reinvestigated with recombinant U2AF^{65}. In this regard it may be relevant that all three RNA recognition motifs are shared between fission yeast and mammalian U2AF^{65} (18), whereas only the third RNA recognition motif shows similarity in *S.cerevisiae* (39).

In conclusion, the work presented here extends the parallels between splicing in metazoa and *S.pombe*. Further investigation will be necessary to determine whether the complete repertoire of functions attributed to 3′ splicing signals and the factors that recognize them in mammals is conserved in fission yeast.

ACKNOWLEDGEMENTS

The authors thank Tim Nilsen and Jim Bruzik for helpful discussions and critical comments on the manuscript. This research was supported by grant #GM 38070 to J.A.W. from the National Institutes of Health.

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 Ruby,S.W. and Abelson,J. (1988) *Science*, **242**, 1028–1035.
3 Séraphin,B. and Rosbash,M. (1989) *Cell*, **59**, 349–358.
4 Nelson,K.K. and Green,M.R. (1989) *Genes Dev.*, **3**, 1562–1571.
5 Champion-Arnaud,P., Gozani,O., Palandjian,L. and Reed,R. (1995) *Mol. Cell. Biol.*, **15**, 5750–5756.
6 Fouser,L.A. and Friesen,J.D. (1987) *Mol. Cell. Biol.*, **7**, 3511–3519.
7 Patterson,B. and Guthrie,C. (1991) *Cell*, **64**, 181–187.
8 Parker,R. and Patterson,B. (1987) In Inouye,M. and Dudock,B. (eds), *Molecular Biology of RNA, New Perspectives*. Academic Press, New York, NY. pp. 133–149.
9 Cellini,A., Felder,E. and Rossi,J.J. (1986) *EMBO J.*, **5**, 1023–1030.
10 Séraphin,B. and Kandels-Lewis,S. (1993) *Cell*, **73**, 803–812.
11 Roscigno,R.F., Weiner,M. and Garcia-Blanco,M.A. (1993) *J. Biol. Chem.*, **268**, 11222–11229.
12 Coolidge,C.J., Seely,R.J. and Patton,J.G. (1997) *Nucleic Acids Res.*, **25**, 888–896.
13 Reed,R. (1989) *Genes Dev.*, **3**, 2113–2123.
14 Russell,P. and Nurse,P. (1986) *Cell*, **45**, 781–782.
15 Zhang,M.Q. and Marr,T.G. (1994) *Nucleic Acids Res.*, **22**, 1750–1759.
16 Brennwald,P., Liao,X., Holm,K., Porter,G. and Wise,J.A. (1988) *Mol. Cell. Biol.*, **8**, 1580–1590.
17 Porter,G., Brennwald,P. and Wise,J.A. (1990) *Mol. Cell. Biol.*, **10**, 2874–2881.
18 Potashkin,J., Naik,K. and Wentz-Hunter,K. (1993) *Science*, **262**, 573–575.
19 Wentz-Hunter,K. and Potashkin,J. (1996) *Nucleic Acids Res.*, **24**, 1849–1854.
20 Csank,C., Taylor,F.M. and Martindale,D.W. (1990) *Nucleic Acids Res.*, **18**, 5133–5141.
21 Prabhala,G., Rosenberg,G.H. and Kaufer,N.F. (1992) *Yeast*, **8**, 171–182.
22 Reich,C.I., VanHoy,R.W., Porter,G.L. and Wise,J.A. (1992) *Cell*, **69**, 1159–1169.
23 Hindley,J. and Phear,G.A. (1984) *Gene*, **31**, 129–134.
24 Maundrell,K. (1990) *J. Biol. Chem.*, **265**, 10857–10864.
25 Alvarez,C.J., Romfo,C.M., VanHoy,R.W., Porter,G.L. and Wise,J.A. (1996) *RNA*, **2**, 404–418.
26 Wise,J.A. (1991) *Methods Enzymol.*, **194**, 405–415.
27 Zhang,H. and Blumenthal,T. (1996) *RNA*, **2**, 380–388.
28 Aebi,M., Hornig,H., Padgett,R.A., Reiser,J. and Weissmann,C. (1986) *Cell*, **47**, 555–565.
29 Green,M.R. (1991) *Annu. Rev. Cell Biol.*, **7**, 559–599.
30 Rymond,B.C. and Rosbash,M. (1992) In Broach,J.R., Pringle,J.R. and Jones,E.W. (eds), *The Molecular and Cellular Biology of the Yeast Saccharomyces, Gene Expression*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. pp. 143–192.
31 Hiraoka,Y., Toda,T. and Yanagida,M. (1984) *Cell*, **39**, 349–358.
32 Blumenthal,T. and Thomas,J. (1988) *Trends Genet.*, **4**, 305–308.
33 Fields,C. (1990) *Nucleic Acids Res.*, **18**, 1509–1512.
34 Mount,S.M., Burks,C., Hertz,G., Stormo,G.D., White,O. and Fields,C. (1992) *Nucleic Acids Res.*, **20**, 4255–4262.
35 Kennedy,C.F. and Berget,S. (1997) *Mol. Cell. Biol.*, **17**, 2774–2780.
36 Reed,R. and Maniatis,T. (1985) *Cell*, **41**, 95–105.
37 Ruskin,B. and Green,M.R. (1985) *Nature*, **317**, 732–734.
38 Zamore,P.D., Patton,J.G. and Green,M.R. (1992) *Nature*, **355**, 609–614.
39 Abovich,N., Liao,X.C. and Rosbash,M. (1994) *Genes Dev.*, **8**, 843–854.
40 Umen,J.G. and Guthrie,C. (1995) *Genes Dev.*, **9**, 855–868.
41 Chiara,M.D., Gozani,O., Bennett,M., Champion-Arnaud,P., Palandjian,L. and Reed,R. (1996) *Mol. Cell. Biol.*, **16**, 3317–3326.
42 Ruskin,B., Zamore,P.D. and Green,M.R. (1988) *Cell*, **52**, 207–219.