An Intrinsic Enhancer Regulates Splicing of the Twintron of *Drosophila melanogaster* prospero Pre-mRNA by Two Different Spliceosomes

Petra Scamborova, Anthony Wong, and Joan A. Steitz*

Department of Molecular Biophysics and Biochemistry, Yale University Howard Hughes Medical Institute, New Haven, Connecticut 06536-9812

Received 12 September 2003/Returned for modification 28 October 2003/Accepted 18 November 2003

We have examined the alternative splicing of the *Drosophila melanogaster* prospero twintron, which contains splice sites for both the U2- and U12-type spliceosome and generates two forms of mRNA, pros-L (U2-type product) and pros-S (U12-type product). We find that twintron splicing is developmentally regulated: pros-L is abundant in early embryogenesis while pros-S displays the opposite pattern. We have established a Kc cell in vitro splicing system that accurately splices a minimal pros substrate containing the twintron and have examined the sequence requirements for pros twintron splicing. Systemic deletion and mutation analysis of intron sequences established that twintron splicing requires a 46-nucleotide purine-rich element located 32 nucleotides downstream of the U2-type 5' splice site. While this element regulates both splicing pathways, its alteration showed the severest effects on the U2-type splicing pathway. Addition of an RNA competitor containing the wild-type purine-rich element to a Kc extract abolished U2-type splicing and slightly repressed U12-type splicing, suggesting that a trans-acting factor(s) binds the enhancer element to stimulate twintron splicing. Thus, we have identified an intron region critical for prospero twintron splicing as a first step towards elucidating the molecular mechanism of splicing regulation involving competition between the two kinds of spliceosomes.

Introns are removed from pre-mRNA by two transesterification steps catalyzed by a large multicomponent complex known as the spliceosome. Five small nuclear RNAs, U1, U2, U4/U6, and U5, and more than 60 polypeptides form the active U2-type spliceosome (5). Spliceosome assembly proceeds in an ordered fashion that is directed by the recognition of conserved sequence motifs within the pre-mRNA by small nuclear RNAs and protein factors. These sequences allow efficient and accurate removal of introns and are located at the 5' and 3' splice sites as well as the branch point (43).

A small subset of introns is removed via a unique and divergent spliceosome (U12-type), whose composition and splice site recognition signals differ (22, 56). U11, U12, and U4atac/U6atac are the functional analogues of U1, U2, and U4/U6 small nuclear RNPs in the U12-type spliceosome; U5 is the only small nuclear RNP shared by both types of spliceosomes (22, 32, 34, 61). As in its U2-type counterpart, interactions of conserved sequence elements at the 5' and 3' splice sites and branch point with the components of the U12-type spliceosome direct the correct recognition and removal of introns from the pre-mRNA.

In addition to the intrinsic quality of the splice sites themselves (19), splice site selection can depend on other properties of the pre-mRNA, such as exon sequences and relative splice site proximity (28, 49), RNA secondary structure (38, 58), exon size (27, 51), and intronic sequences (4). For many pre-mRNAs, the splicing reaction produces only a single product from the pre-mRNA transcript. However, other genes undergo alternative splicing, a process in which splice sites in a single primary transcript are differently paired to generate two or more mRNAs encoding multiple protein isoforms with slightly altered or opposing functional properties (18). It was recently estimated that nearly 60% of all human genes undergo at least one alternative splicing event (36).

A remarkable variety of alternative splicing patterns has so far been observed, involving differential 5' or 3' splice site selection, alternative exon selection, and intron retention (18). In many instances, alternative splicing is regulated in a developmental or tissue-specific manner. Such complex patterns have been suggested to derive from numerous distinct mechanisms (72). For many examples of alternatively spliced genes, exonic (21, 31, 67) or intronic (3, 9, 11, 20, 26, 29, 52, 68) cis-acting RNA sequences that positively or negatively regulate splice site choice have been identified. Binding of specific proteins to these regulatory sequences, in turn, dictates the specificity and efficiency of splicing resulting in promotion or repression of each splicing event (20).

The *Drosophila melanogaster* prospero (*pros*) pre-mRNA provides a rare and unusual example of alternative splicing. The *pros* gene locus encodes a homeodomain-containing neuronal transcription factor, Prospero, which is involved in control of axon outgrowth and in cellular specification in the developing *Drosophila* central nervous system (14, 40, 66). The second intron of *pros* pre-mRNA contains two complete sets of splice sites, an arrangement referred to as twintron. One set of splice sites is of the U2-type (GT-AG termini), whose usage leads to the production of the *pros-L* mRNA isoform. The U2-type splice sites are nested inside an intron defined by a second set of splice sites of the U12 type (with AT-AC termini), whose usage leads to the production of the *pros-S* mRNA isoform, 87 nucleotides shorter than *pros-L* (Fig. 1A). Excision of the
FIG. 1. Splicing of the pre-mRNA is temporally regulated during Drosophila embryogenesis and early larval development. (A) Schematic of the alternatively spliced second intron of D. melanogaster prospero. The U2-type intron is 730 nucleotides long. The U12-type intron contains 59 and 28 nucleotides (dotted boxes) from the flanking the U2-type intron. The sequences encoding flanking boxes. The positions of primers for reverse transcription-PCR analysis are indicated. (B) Input RNA levels correlate with reverse transcription-PCR product levels. In vitro-transcribed pros-L RNAs, derived from pros-S cDNA, were mixed at various ratios, reverse transcribed, and PCR amplified. The DNA amplicons were then electrophoretically separated, and the quantities of amplicons were measured by PhosphorImager analysis. (C) Stage-specific splicing of pros pre-mRNA during embryogenesis. The DNA amplicons were resolved on 12% denaturing polyacrylamide gels. (D) Quantification of pros-L mRNA products at various times throughout Drosophila development. The DNA amplicons were quantified by PhosphorImager analysis.

Fundamental to understanding the alternative splicing of the prospero twintron is an elucidation of the mechanism(s) involved in splice site selection. The existence of two competing splicing pathways within one intron raises the question of whether the two splicing pathways are coregulated or separately controlled through different mechanisms. We began by analyzing the temporal splicing profile of the two alternatively spliced pros mRNA isoforms and found that pros-L predominates during the first half of Drosophila embryogenesis while pros-S is more abundant at later stages. We also set out to identify possible regulatory elements within the twintron. The splicing of pros minigene constructs containing both sets of splice sites was analyzed both in vitro and in vivo. The results show that twintron splicing depends on a 46-nucleotide intronic sequence, called the purine-rich element, located 32 nucleotides downstream of the U2-type 5’ splice site. Its deletion or modification decreases or completely abolishes U2-type splicing. Lastly, we find that the same element is involved in the control of U12-type splicing, since mutation of the element induces a marked decrease in the splicing efficiency of that pathway. Several hypotheses to explain how this element exerts its regulatory functions are discussed.

MATERIALS AND METHODS

Drosophila embryo RNA preparation. Synchronized Drosophila embryos were collected from a wild-type Canton S population cultured at 25°C on apple juice collection plates coated with a paste of brewer’s yeast at 3-h intervals, the first 1-h collection being discarded. The embryos were washed, homogenized by tissue-teacer, and total RNA extracted by Trisol reagent according to the manufacturer’s directions.

Minigene constructs. All DNA constructs were created with standard cloning procedures. For cloning the wild-type pros minigene, two DNA fragments of pros were PCR amplified from wild-type total Drosophila genomic DNA as follows. Fragment I, nucleotides 4296 to 4509 (accession AF190403), and fragment II, nucleotides 4798 to 4984, had XhoI and XbaI sites introduced at the 5’ and 3’ termini, respectively. Fragment I included 118 nucleotides of exon 2 and 109 nucleotides of downstream intronic sequence, while fragment II included 63 nucleotides of exon 3 and 123 nucleotides of upstream intronic sequence. The resulting 413-nucleotide wild-type pros construct was cloned between the XhoI and XbaI sites of the pBluescript (SK+)/pEMD (Stratagene) downstream of a T7 promoter.

Several pros intron deletion and replacement mutants were constructed, denoted Δx-y for deletions and XxY for mutations, where x indicates the nucleotide at which the deletion or mutation starts and y indicates the last nucleotide deleted or mutated. The deletion and replacement mutant minigenes were generated with the QuickChange Mutagenesis system (Stratagene). Plasmids pGEM-2V6 and pGEM-2TB/R (50), containing the Drosophila ftz and tra genes, respectively, were a generous gift from Emmanuel Laborier and Donald Rio (UC Berkeley).

For generation of in vivo splicing constructs, each in vitro pros XhoI-XbaI fragment was PCR amplified from pBluescript (SK+)/pEMD and inserted into the SpeI and XhoI sites of the Drosophila expression vector pMTV5-HisC (Invitrogen) downstream of a metallothionein promoter and upstream of a simian virus 40 polyadenylation signal. All plasmid constructs were verified by restriction enzyme mapping and DNA sequencing (Keck DNA sequencing lab, Yale University).

In vitro transcription. pros-L splicing substrate constructs were linearized by cleaving at the 3’ restriction site with XhoI to create templates for in vitro transcription by T7 RNA polymerase, performed essentially as described by Tarn and Steitz (61) with [α-32P]UTP and 1 μg of linearized plasmid. After RQI DNase treatment (Promega), RNA transcripts were gel purified on 5% denaturing polyacrylamide gels and quantified by Cerenkov counting.
Competitor RNA fragments were transcribed under the same conditions except that reactions contained 400 μM each ribonucleoside triphosphate and 0.5 μCi of [α-32P]UTP (Amersham). The trace-labeled competitor RNA fragments were gel purified and quantified by Cerenkov counting.

RTG primer extensions. The Drosophila ΔSc transcription unit was transcribed and treated as described above.

Calculation for pros-S. The percent splicing efficiency; rather it increased background, probably because of RNA degradation. The concentrations of 2'-O-methyl oligonucleotides U2b (complementary to the sequences in pros-L exon 3 sequences nucleotides 280 to 290 (Fig. 2A) oligonucleotides 4974 to 4984, accession AF190403). The reaction was performed in 33 μl with the RTG First Strand Synthesis kit (Amersham) at 37°C for 2 h.

cDNA obtained by reverse transcription was amplified with 0.1 μg/μl primers complementary to the sequences in pros exon 2 (PR514; 5'-AGATGCGACGTCGAG-3') and pros exon 3 sequences nucleotides 280 to 290 (Fig. 2A) (nucleotides 4974 to 4984, accession AF190403) and exon 3 (PR323; 5'-GGTGCATCGGTGTCAATGTA-3'; nucleotides 262 to 281 (Fig. 2A) or nucleotides 4956 to 4975, accession AF190403), in a final volume of 100 μl with 2.5 U of AmpliTaq Gold polymerase (Perkin-Elmer), 50 mM Tris (pH 8.3), 7.5 mM MgCl2, 7.5 mM dithiothreitol, 10 mM MnCl2, 0.08 μg/ml bovine serum albumin, and 2.4 mM each deoxyribonucleotide triphosphate.

To quantify the amplification products, [γ-32P]ATP 5' end-labeled primer PR514 (200,000 cpm/μl) was added to the standard 100-μl PCR. Since pilot experiments showed that the PCR amplification rates of the endogenous pros transcripts and of the minigenic-derived pros transcripts were linear between the 22nd and 32nd cycles, we used the following cycle parameters for PCR amplification: an initial denaturation at 95°C for 10 min, followed by 30 cycles at 95°C for 1 min, 60°C for 1 min, and 72°C for 2 min. Amplicons (5% of the PCR mixture) were separated on a 12% polyacrylamide gel, dried, and exposed. Bands were quantitated with a Molecular Dynamics PhosphorImager. Each quantification was performed on at least 3 experiments. Representative reverse transcription-PCR splicing assays are presented in the figures.

Quantification of each band was corrected by a corresponding background value. In embryos, the percent pros-L inclusion for all time points was calculated as [cpm pros-L/cpm pros-L + cpm pros-S] × 100. For in vitro splicing assays, the percent pros-L spliced was calculated as [cpm pros-L/cpm pre-mRNA] × 100. Calculations for pros-S were done similarly to pros-L.

For the in vitro splicing competition assay, 1 μl of trace-labeled RNA competitor (1 to 100 fmol) was added together with 1 fmol of pre-mRNA to the splicing mix. Splicing reactions were carried out as described above.

Cell culture and DNA transfection. Schneider (S2) cells were grown at 25°C in Shelds and Sang M3 medium (Sigma) supplemented with 12.5% fetal bovine serum (Gibco-BRL) and 5% penicillin/streptomycin. Schneider cells were transiently transfected in 35-mm dishes with 1.5 μg of plasmid DNA and 50 μl of Lipofectin reagent (Invitrogen) per dish, according to the manufacturer's protocol. Expression was induced 24 h after transfection by adding 0.7 mM cAMP, 0.7 mM caffeine, and 0.6 mM cAMP, 0.6 mM caffeine. Cells were harvested 24 h postinduction and total cellular RNA was isolated with Trizol reagent according to the manufacturer's instructions. Contaminating genomic DNA was removed by treating with RNase-free DNase I (Promega) in 5 mM MgCl2 and 50 mM Tris, pH 8.0, at 37°C for 1 h.

Reverse protection assay. The template for the RNase protection probe (185 nucleotides) was subcloned into the wild-type pros pMT-5/V5-HisC construct into the pBluescript SK+ vector (Stratagene) in the reverse orientation with restric-
genesis (12 h; Fig. 1C and Fig. 1D, 9 to 12 h), the levels of the two spliced isoforms are similar. Although a maternal contribution of \textit{prospero} mRNA cannot be formally excluded, it is unlikely to have impact beyond the first 2.5 h of embryogenesis when zygotic transcription peaks (C. Hashimoto, personal communication). These data suggest that the choice of splicing pathway is developmentally regulated and strictly controlled during embryogenesis.

**Design of \textit{prospero} minigenes and in vitro splicing assay.** We defined the minimum \textit{pros} RNA sequence required to direct both types of splicing in a homologous \textit{Drosophila} Kc nuclear extract. A minigene substrate (wild-type \textit{pros}) that contains shortened exons 2 (from 234 nucleotides to 118 nucleotides) and 3 (from 132 nucleotides to 63 nucleotides) and a shortened intron 2 (from 729 nucleotides to 232 nucleotides) but conserves both sets of splice sites and their spacing with respect to one another was constructed (Fig. 2A). It includes the splicing enhancer element, the purine-rich element, defined in this study (Fig. 2A, green). This wild-type \textit{pros} pre-mRNA construct was processed so poorly in both \textit{Drosophila} Kc and human HeLa nuclear extracts that \(32^P\)-labeled spliced products could not be detected under a variety of conditions tested (data not shown), even though \textit{Drosophila} \textit{ftz} and \textit{tra} pre-mRNAs both spliced efficiently (data not shown), as previously reported (50). Therefore, the use of reverse transcription-PCR was necessary to assess \textit{pros} spliced products generated in \textit{Drosophila} Kc nuclear extract.

Reverse transcription-PCR analysis of the in vitro splicing of the parental wild-type \textit{pros} substrate (Fig. 3) demonstrated that the twintron was recognized correctly to generate both \textit{pros}-L and \textit{pros}-S spliced RNAs, resulting from the U2- and U12-type splicing pathways, respectively. Maximal splicing efficiency was obtained with 3 mM MgCl\(_2\) and 30 mM KCl (data not shown), where about 2.5% of pre-mRNA became spliced after 3 h at 24°C, compared with 6% of the human P120 U12-type splicing substrate that becomes spliced after 4 h in a
In vitro splicing of *prospero* substrates in *Drosophila* Kc nuclear extract. (A) Reverse transcription-PCR analysis of unspliced (unspl; lane 1) and wild-type *pros*-S in Kc nuclear extract in the presence of the U12a or U2b 2-O-/*H11032* pros-L and pros-S and amplified DNA fragments were fractioned with primers PR514 and PR323. Amplified cDNA products, pros-L resolved on 12% polyacrylamide gels. The pre-mRNA and the spliced products, pros-S, are indicated. (B) Splicing of the *pros*-splicing substrates listed in Fig. 2B were spliced in vitro, reverse transcription-PCR amplified, and analyzed as in A. (C and D) Graphical representation of the data in Fig. 3B showing the percentage of spliced product for each substrate tested, calculated as percent pros-L (pre-mRNA) (left-hand scale) and percent pros-S/pre-mRNA (right-hand scale). Measurements were made directly from the gels by PhosphorImager analysis. Each value is an average of at least three independent splicing reactions; error bars indicate the standard deviation associated with each average value. Black bars, products of U2-type splicing; gray bars, products of U12-type splicing.

**Fig. 3.** In vitro splicing of *prospero* substrates in *Drosophila* Kc nuclear extract. (A) Reverse transcription-PCR analysis of unspliced (unspl) and wild-type *pros*-S (lane 2), from 2'-O-methyl oligonucleotide blocking experiments (Fig. 3A, lanes 3 and 4) and from the results of splice site mutations (Fig. 3B, lanes 1 and 13; Fig. 3C, MUTATAC and MUTGU AG). When the nuclear extract was preincubated with a 2'-O-methyl oligonucleotide, U2b, complementary to the branch point binding region of U2 small nuclear RNA, U2-type splicing was blocked (15), as seen by the absence of *pros*-L product, while wild-type levels of *pros*-S spliced product were observed (Fig. 3A, lane 4). Conversely, blocking the competing U12-type splicing pathway with a 2'-O-methyl oligonucleotide, U12a, directed against the comparable region of U12 small nuclear RNA (15) produced wild-type levels of *pros*-L spliced product but no *pros*-S product (Fig. 3A, lane 3).

As previously observed with the U12-dependent P120 intron, the oligonucleotide complementary to U2 small nuclear RNA actually stimulates the level of in vitro U12-type splicing (61). Likewise, mutations in the U2-type splice sites produced wild-type levels of *pros*-S spliced product but no *pros*-L product (Fig. 3B, lane 13; Fig. 3C, MUTGUAG), whereas mutations in the U12-type splice sites produced the opposite effect (Fig. 3B, lane 1; Fig. 3C, MUTATAC). Thus, the two splicing pathways operate independently; blocking one splicing pathway by mutation does not affect the other. In addition to the *pros*-L and *pros*-S cDNA products, we frequently observed variable amounts of nonspecific bands running between the two; these were not present in embryo reverse transcription-PCR preparations (compare Fig. 1C, lanes 1 through 12, with Fig. 3B, lanes 1 through 13) and thus most likely are amplified from RNAs present in the nuclear extract.

**Defining sequence elements required for *prospero* twintron splicing in vitro.** Given the unusual structure of twintron, we anticipated that *pros* splicing would be dependent upon *cis*-acting sequences in addition to its two sets of splice sites. Most striking is a purine-rich sequence in the 5’ part of the intron, containing 11 GA repeats flanked by more heterogeneous purine-rich sequences (Fig. 2A). To examine the possible role of this intronic sequence in *pros* splicing, replacements (Rx-y) and deletions (Δx-y; numbering here and henceforth refers to the distance downstream from the 5’ U2-type splice site of the *pros* twintron) were introduced into the twintron substrate, as depicted in Fig. 2B, and their splicing patterns were examined in Kc nuclear extract. We calculated the percentage of each spliced product relative to the remaining pre-mRNA. These data, from at least four experiments, are plotted in Fig. 3C and D.

The R106-135 mutation replaces nucleotides that are pyrimidine rich in the middle of the intron with random sequence (Fig. 2A, yellow). The R106-135 substrate reproducibly showed no significant decrease in the splicing of *pros* twintron or in the ratio of *pros*-L to *pros*-S (Fig. 3C, R106-135) compared to
wild-type pros (Fig. 3B, lane 2; Fig. 3C, wild-type pros). Similarly, the R136-177 mutation, which replaces 3' intron sequences (Fig. 2A, purple), retains substantial levels of both splicing pathways (Fig. 3C, R136-177). A substrate containing a 10-nucleotide deletion immediately upstream of the purine-rich region, Δ22-32, also showed no effect on splicing levels (Fig. 3B, lane 8; Fig. 3C, Δ22-32). These results indicate that intronic regions outside of the purine-rich region contribute little or nothing to pros twintron splicing regulation.

Similar analyses of the purine-rich intronic region suggested that multiple copies of GA repeats might have an important enhancer function (Fig. 3B, lanes 4 and 5). To test this hypothesis, we first replaced the purine-rich region with a random sequence (39% purine) in construct R32-77 (Fig. 2B); this largely abolished the major-class splicing pathway and caused a marked decrease in minor-class splicing (Fig. 3B, lane 4; Fig. 3C, R32-77). Deletions of the intronic purine-rich region which remove the 11 GA repeats and an additional 19 or 24 nucleotides of flanking purine-rich sequence (constructs Δ32-77 and Δ37-77, Fig. 2B) showed a marked reduction in splicing levels compared to wild-type pros (Fig. 3B, lane 3; Fig. 3C, Δ32-77; Fig. 3D, Δ37-77) although the reduction was not as great as that of the R32-77 and R38-77 replacement constructs (Fig. 3B, lane 4; Fig. 4B, lane 2; Fig. 3C, R32-77; Fig. 4C, R38-77). These results suggest that the 46-nucleotide purine-rich region contains a sequence or structure that functions as an essential activating element in twintron splicing. We therefore designated the purine-rich region an intronic enhancer and refer to it as the purine-rich element.

A purine-rich element downstream of the major 5' splice site enhances splicing of both pathways. To further define the role of the purine-rich element, we made additional constructs deleting or mutating parts of the element (Fig. 2B). Constructs Δ40-77 and Δ42-77 had no significant effect on pros twintron splicing (Fig. 3D), while construct R32-36+Δ37-77, in which the first five nucleotides of the purine-rich element are replaced by a nonpurine sequence and the remainder of the purine-rich element is deleted, showed a marked decrease in pros-L spliced product (90% decrease) as well as in pros-S product (45% decrease) (Fig. 3D). Replacing the first 6 or 8 nucleotides with a heterologous non-purine-rich sequence, constructs R32-37 and R32-69, did not rescue splicing efficiency either (Fig. 3B, lane 7; Fig. 3C, R32-37 and R32-69). These experiments confirm that essential regulatory sequences are positioned within the purine-rich element (nucleotides 32 to 77 of intron 2) and that they are required for the splicing of both types of intron, although the effects are more pronounced on the U2-type splicing pathway (pros-L) than on the U1-type (pros-S).

Since the sequence downstream of the purine-rich element contains several GTA repeats flanked by purines and is itself quite purine rich (71.4% versus 93.5% for the purine-rich element), this region may substitute for the purine-rich element in deletion and substitution mutants. To test this hypothesis, we constructed two substrates, Δ32-105 and Δ40-105, that contained complete or partial deletions of the purine-rich element together with a deletion in the GTA-repeat region immediately downstream (Fig. 2B). Splicing analysis revealed the largest negative effects on both splicing pathways seen so far (Fig. 3B, lanes 11 and 12), with splicing levels of pros-L and pros-S diminished to roughly 10% for Δ32-105 (Fig. 3B, lane 11; Fig. 3D). Keeping the first 8 nucleotides of the purine-rich element, in Δ40-105, raised the U2-type splicing level to roughly 20% to 30% (Fig. 3B, lane 12; Fig. 3D). Thus, the first eight nucleotides of the purine-rich element appear to be a primary determinant of enhancer activity, with the downstream purine-rich region participating in splicing regulation of pros when the purine-rich element is deleted, but not when it is replaced. Taken together with the results of the Δ32-77 and Δ37-77 substrates, these data suggest that the purine-rich element functions as a bipartite element in which nucleotides 32 to 37 and 77 to 105 influence the overall efficiency of both splicing pathways, whereas nucleotides 37 to 77 primarily influence the ratio of pros-S to pros-L.

Purine-rich element is not replaceable by a heterologous splicing enhancer. The prospero purine-rich element is comprised of a 46-nucleotide-long stretch of purines interrupted by only three C residues. It therefore resembles the GAR (R = purine) repeats present in many previously characterized exonic splicing enhancers (26, 31, 60, 69). To ask whether the purine-rich element can be replaced by a known purine-rich enhancer element competent to facilitate splicing of other introns by binding to a known SR protein, we chose two such elements and cloned them into a substrate lacking the entire purine-rich element (see Fig. 2B, constructs R-SR and R-ASF/SF2). The first sequence is an enhancer found in the human fibronectin alternate ED1 exon, while the second was identified as an ASF/SF2 binding sequence by SELEX (59). When placed in the context of the pros purine-rich element, neither completely restored pros-L splicing to wild-type levels (Fig. 3B, lanes 9 and 10; Fig. 3D, R-SR and R-ASF/SF2), suggesting that specific elements within the purine-rich element or the spacing of repeats are important for full enhancement of pros twintron splicing.

First 20 nucleotides of the purine-rich element enhance splicing. To define the regulatory sequences within the pros purine-rich element more precisely, mutant constructs were made in which 5-nucleotide portions of the region were kept intact and the rest of the purine-rich element replaced with mutant sequences, thereby maintaining the spacing (Fig. 4A). When tested for their ability to activate prospero splicing in vitro (Fig. 4B), we observed that constructs R38-77 and R42-77, which contained only the first 6 or 10 nucleotides of the purine-rich element, respectively, decreased splicing to ~70% for the pros-S spliced product and to ~15% for pros-L product (Fig. 4B, lanes 2 and 3; Fig. 4C, R38-77 and R42-77). When the first 15 nucleotides of the purine-rich element were kept intact (construct R47-77), splicing proceeded with higher efficiency, but still significantly below wild-type levels (pros-S at ~80% and pros-L at ~45%) (Fig. 4B, lane 4; Fig. 4C, R47-77).

We also examined RNA substrates retaining the first 20 to 40 nucleotides of the purine-rich element and found that they were spliced at levels similar to the wild-type pros substrate (Fig. 4B, lanes 5 to 9). Here, pros-S was produced at ~95% and pros-L at 85 to 90% (Fig. 4C, R52-77, R57-77, R62-77, R67-77, and R72-77). Conversely, the R32-36+Δ37-77 RNA substrate, which has the first 5 nucleotides of the purine-rich element replaced, is spliced even less efficiently than the R38-77 substrate (Fig. 3D, R32-36+Δ37-77; Fig. 4C, R38-77). From these analyses, we conclude that the activating sequence essential for
FIG. 4. First 20 nucleotides of the purine-rich element are necessary for pros in vitro splicing. (A) Purine substitutions that substitute a heterologous sequence for the purine-rich element while preserving the spacing found in the parent substrate, wild-type pros. The wild-type purine-rich element sequence is underlined. (B) The substrates in A were spliced in Kc nuclear extract, followed by reverse transcription-PCR. The pre-mRNA and the spliced products (pros-L and pros-S) are indicated. (C) Graphic representation of pros-L and pros-S spliced products. Each percentage, calculated as described for Fig. 3C and D, is from a minimum of three independent experiments.
pros twintron splicing is located in the 5' half of the purine-rich element, 32 to 52 nucleotides downstream of the U2-type 5' splice site. This region includes the eight nucleotides identified as important above and contains six copies of the GA repeat at its 3' end.

**Trans-acting factors mediate enhancement of pros splicing via the purine-rich element.** To ask whether the purine-rich element interacts with trans-acting factors to enhance pros twintron splicing, we carried out competition experiments with increasing concentrations of a 46-nucleotide RNA containing the entire purine-rich element (purine-rich element RNA). Partial inhibition (to 25%) of U2-type splicing was observed with as little as twofold molar excess, while almost complete inhibition (to 10%) was observed with 100-fold excess of purine-rich element RNA (Fig. 5A, lanes 5 to 7; Fig. 5B). For U12-type splicing, addition of a twofold molar excess of purine-rich element RNA resulted in a 50% decrease in splicing (Fig. 5A, lane 5; Fig. 5B), while addition of a 100-fold molar excess further decreased splicing to about 40% (Fig. 5A, lane 7; Fig. 5B). In contrast, competition with a non-purine-rich RNA (mPRE RNA), containing sequences that did not substitute for the purine-rich element in vitro splicing experiments (Fig. 2B, R32-77; Fig. 3B, lane 4), did not significantly reduce the splicing efficiency of wild-type pros RNA even at the highest concentration tested (Fig. 5A, lane 4). These results are consistent with the idea that the RNA competitor sequesters, by mass action, one or more trans-acting factors required for splicing enhancement of the pros twintron.

**Purine-rich element enhances prospero splicing in vivo.** The results reported in the preceding sections identified an intronic purine-rich element that acts as an enhancer of pros splicing in vitro. To determine whether the purine-rich element affects pros splicing in vivo as it does in vitro, we transiently transfected Drosophila S2 cells with minigenes carrying the same intronic sequence variants as analyzed in vitro. The 413-bp prospero minigene substrate used in vitro splicing was PCR amplified from the pBS vector and subcloned into a Drosophila expression vector, pMT/V5-HisC, downstream of the metallothionein promoter and upstream of the simian virus 40 polyadenylation signal (Fig. 6A). The Schneider (S2) cell line was used in these assays because Drosophila Kc cells could not be efficiently transfected under a variety of conditions tested (data not shown). After transient transfection and induction with copper sulfate, total RNA was isolated and the splicing pattern of the minigene-derived mRNA molecules monitored by RNase protection (Fig. 6B). Because the S2 cell line is derived from blood precursor cells and pros is primarily expressed in developing neurons and in the midgut, we did not expect or detect any endogenous pros expression (data not shown).

Similar to in vitro splicing patterns, in vivo analyses identified reverse transcription-PCR amplicons of two spliced products (pros-L and pros-S), as well as of the pre-mRNA, based on electrophoretic mobility compared to cDNA plasmid standards cloned into the pMT/V5-HisC vector and transcribed into S2 cells (data not shown). In addition, U2- and U12-type splice site mutations cloned into the same vector revealed that interrupting one splicing pathway did not impede the other splicing pathway in vivo (Fig. 6C, mATAc and mGUAG). In strong contrast to the in vitro results, in vivo splicing of wild-type pros in S2 cells produced a small amount of pros-S (2% of pre-mRNA) and a large amount of pros-L product (300% of pre-mRNA). Nevertheless, splicing of the pros twintron was greatly reduced for minigenes lacking the purine-rich element (Δ32-77, Fig. 6B, lane 3; Fig. 6C) or where the purine-rich element was mutated (R32-77, Fig. 6B, lane 2; Fig. 6C). Therefore, the purine-rich element plays a similar role in pros splicing activation in vivo as it does in vitro.

The results of RNase protection analyses of RNAs obtained from transfected S2 cells are summarized in Fig. 6C, and representative splicing patterns are shown in Fig. 6B. Because the splicing efficiency of the U12-type splicing pathway was very low and thus difficult to quantify accurately, only the results for U2-type spliced products are discussed. As in vitro, mutations outside of the purine-rich element (Δ22-32, R106-135, and R136-177) did not significantly affect splicing efficiency in vivo (Fig. 6C). Likewise, deletion mutations within the purine-rich element (R32-36+Δ37-77 and Δ32-105) severely reduced U2-type splicing (Fig. 6C), while replacement mutants within the purine-rich element, constructs R32-37 and R32-69 reduced the pros-L splicing levels by 80% (Fig. 6D). In contrast with constructs Δ40-77 and Δ42-77, where the first eight or ten nucleotides of the purine-rich element were intact, the levels of pros-L were similar to that of wild-type pros (Fig. 6C), as in vitro. Likewise, double deletions encompassing the purine-rich element and downstream GTA repeat region showed a severe splicing phenotype, with the pros-L levels decreased by 95% when compared to wild-type pros (Δ32-105 and Δ40-105, Fig. 6B, lanes 5 and 6; Fig. 6C), again similar to the splicing phenotype observed in vitro. Lastly, the chimeric constructs R-SR and R-ASF, where the purine-rich element was substituted with heterologous purine-rich enhancer sequences, did not splice better in vivo (Fig. 6B, lane 4; Fig. 6D; data not shown) than in vitro (Fig. 3B), underscoring the sequence-specific functioning of the purine-rich element.

We also tested minigene replacement mutants in which increasing lengths of the purine-rich element were kept intact while replacing the rest with a nonpurine sequence to maintain the total length of the intron (Fig. 6B, lanes 9 to 16). Analysis of these additional mutants showed that the first 20 nucleotides of the purine-rich element are required for minimal pros-L splicing in vivo and the first 30 nucleotides are necessary for spliced levels equivalent to those of wild-type pros (Fig. 6B, lanes 12 to 14; Fig. 6D, R38-77 through R67-77), in good agreement with the in vitro data.

Overall, the in vivo analyses support the in vitro results. The S2 cell line, like the Kc nuclear extract, specifically requires the purine-rich element for splicing of the pros twintron, suggesting that a trans-acting factor(s) recognizing the purine-rich element is present in a variety of cell types.

**DISCUSSION**

Regulation of prospero twintron splicing is important for determining the balance between pros-S and pros-L, two isoforms encoded by the Drosophila melanogaster pros gene that arise from usage of the U12- and U2-type splicing pathways, respectively. In this study, we show that these two isoforms are expressed in a stage-specific manner; the U2-type spliced product predominates in the early stages of Drosophila embryogenesis, while the U12-type spliced product is abundant during the second half of embryo development. We have also identified a
FIG. 5. Inhibition of in vitro splicing by competition with purine-rich element RNA. (A) Kc nuclear extract was incubated with pre-mRNA and increasing amounts of an RNA containing the purine-rich element sequence (lanes 5 to 7) or a nonspecific sequence (Fig. 2B, R32-77; mPRE, lanes 2 to 4) that does not substitute for the purine-rich element in in vitro splicing assays. After splicing, the RNAs were analyzed by reverse transcription-PCR as above. The molar excess of competitor is indicated at the top of each lane. The wild-type prospero pre-mRNA and products, pros-L and pros-S, are indicated. (B) Graphical representation of the data in A showing the percentage of pros-L and pros-S spliced products in the presence of RNA competitors. Each value was calculated as percent pros-L (or pros-S)/pre-mRNA; the value of wild-type pros with no competitor was set at 100%.
A. Diagram showing the structure of a gene with exons labeled as Exon2, PRE, and Exon3. The diagram also includes a RNase protection probe and the RNA 3' end at 40 bases (pA).

B. Gel images showing the expression patterns labeled as pros-L (137 nts) and pros-S (108 nts) with markers R1, R2, R4, R5, R6, R7, R8, and wt pros.

C. Bar graph showing the expression levels of pros-L (300-400 units) with WT PROS, DEL22-52, DEL21-15, DEL12-17, DEL12-17, DEL12-17, DEL12-17, DEL12-17, DEL12-17, DEL12-17, DEL12-17, DEL12-17, DEL12-17, DEL12-17, and DEL12-17.

6-nucleotide cis-acting element, the purine-rich element, located 32 nucleotides downstream of the 5′ U2-type splice site that enhances pros twintron splicing both in vitro and in vivo. Its deletion or mutation has two effects, the most noticeable being the complete inhibition of the U2-type splicing pathway and the second a more subtle downregulation of the U12-type splicing pathway. Based on competition data, we conclude that the purine-rich element acts by binding a trans-acting factor(s) which remains to be identified.

Implications of differential temporal expression of prospero isoforms. Stage- and tissue-specific pre-mRNA splicing events are important aspects of gene regulation that allow an organism to produce related proteins from a single gene (18). Alternative splicing of the prospero twintron produces two isoforms, pros-L and pros-S mRNAs, via the usage of two distinct splicing pathways, the U2- and U12-type, respectively. Removal of the larger U12-type intron excises 29 amino acids encoded by pros-L (removed by the U2-type pathway), five of which are part of the N terminus of the pros homeodomain.

We began by investigating the stage-specific profile of prospero pre-mRNA splicing in the Drosophila embryo and early larva and found that one form (pros-L) predominates in early embryogenesis while the other (pros-S) predominates in middle to late embryogenesis and early larval development. Similar stage-specific splicing has been reported for another alternatively spliced Drosophila homeotic gene, ultrabithorax (ubx), where the RNA levels have been shown to fluctuate during Drosophila embryonic development with the long ubx isoforms (Ia and Ib) predominating during the first 7.5 h and a short isoform (IV) predominating from 18 to 22.5 h of Drosophila embryogenesis (35, 44).

What functional differences might derive from the structural differences between the proteins encoded by the two pros mRNAs? Pros protein has been reported to regulate multiple target genes in different lineages and different stages of neuronal development by repressing or activating their transcription (10, 66). It is therefore possible that the two pros isoforms interact with different coactivators in different tissues, and at different developmental stages. Recent structural studies on the Pros-L C-terminal domain (encompassing the homeodomain and Prospero domain) suggested three regions of potential Prospero-DNA contacts, among them, the N-terminal arm of the homeodomain appears to contact the minor groove of the DNA (53). The Pros homeodomain is essential for the sequence-specific DNA binding function of Prospero and its N terminus has been shown to associate with two homeobox proteins, Engrailed (12) and Deformed (25). Because the Pros-S homeodomain contains a different N terminus, it is tempting to speculate that the two Pros proteins exhibit different DNA-binding capacities that lead to target gene selectivity during neurogenesis in a tissue or stage-specific manner. As a result, production of various amounts of the two protein isoforms would be expected at different stages of development or in different tissues.

Two additional examples of twintrons in the second intron of the prospero gene are found in D. pseudoobscura and D. virilis. Sequence alignment further reveals a homologous purine-rich element in a similar location in D. pseudoobscura; it is located 40 nucleotides downstream of the 5′ U2-type splice site and contains the same number of GA repeats flanked on both sides by purines. The 5′ half of the purine-rich region is 100% conserved between the two species (apart from three missing residues, which could be due to sequencing error) while the 3′ end is less conserved. D. virilis (which is more diverged from D. melanogaster than D. pseudoobscura [48]) also contains a purine-rich region located 28 nucleotides downstream of its U2-type 5′ splice site, but this region contains only five GA repeats. These observations support the functional importance of the purine-rich element as a regulatory cis-element in twintron splicing.

Additional Prospero orthologs, called Prox1, have been identified in Caenorhabditis elegans (6), chicken (64), mouse (64), and human (73). Their protein expression patterns indicate that they are critical for the development and function of the developing central nervous system, eye, and midgut (64). While Prox1 is very well conserved in vertebrates (64), including 102 amino acids at the 3′ terminus of the Prox1 homoeodomain which are 62% identical to corresponding residues of Drosophila Prospero (45), the first seven amino acids of the Prox1 homoeodomain do not share homology with Drosophila Prospero. Strikingly, all of the vertebrate proxi genes contain a single U12-type intron located near the N terminus of the homeobox in exactly the same position as the U12-type intron in Drosophila prospero (64). However, no alternatively spliced variants of Prox1 have been identified (45), and we have found no hints of U2-type splicing signals in the second intron of mouse or human Prox1. Furthermore, sequence analysis of human prospero has not revealed any pronounced intronic purine-rich sequences. This suggests that the twintron and its intronic purine-rich element are unique to Drosophila species, where they may play a special role in development.

Splicing of prospero is enhanced by an intronic purine-rich element. To our knowledge, prospero provides the only example of a twintron whose splicing can proceed via either the U12- or U2-type pathway. Our documentation of temporal specificity in embryos lends further support to the idea that regulatory signals within the pros pre-mRNA act as sensors of developmental and/or tissue-specific cues. We generated a minimal splicing substrate and established an in vitro splicing system in Kc nuclear extract as a means to map and characterize cis-regulatory elements in a systematic way. Deletion and replacement mutant analysis of the intron led to identification of a 46-nucleotide purine-rich regulatory element, denoted the purine-rich element, that is located 32 nucleotides downstream of the U2-type 5′ splice site and is required for

FIG. 6. Splicing of prospero substrates in Drosophila S2 cells. (A) Organization of the pros minigene used for transfection. The line underneath the construct shows the probe used in the RNase protection assay. Vector sequences are not drawn to scale. (B) Representative gels of RNase-protected fragments analyzing the products of transient transfection. The pros-S product is always detected as a doublet, often seen in RNase protection analyses. (C and D) The bar graphs show the levels of pros-L splicing for each substrate tested. Error bars indicate the standard deviation of the measurements from four to six in vivo splicing experiments. The percent of pros-L was defined as [[(total counts in the pros-L product)/(total counts in the pre-mRNA)]].
twintron splicing both in vitro and in vivo. Complete replacement of the purine-rich element impairs U2-type splicing both in vitro and in vivo. The presence of multiple GA repeats is consistent with a particle, redundant element structure, a common feature of many intronic regulatory sequences (26). Indeed, in vitro studies indicated that only the first 20 nucleotides of the purine-rich element are required for pros twintron splicing. Furthermore, replacements of the purine-rich element by purine-rich SR protein binding sites did not completely restore splicing, while the deletion mutants in which only the first 8 or 10 nucleotides of the purine-rich element were kept did.

Although in vitro splicing in Kc cell extract yielded high levels of pros-S and low levels of pros-L products, the same transcripts spliced in vivo in the Schneider (S2) cell line showed the opposite effect. The linearity of the transfection data of Patel et al. (47), obtained in S2 cells under the same conditions, excludes the possibility that the observed differences result from artificial saturation of the U12-type splicing machinery by high levels of pre-mRNA expressed from the transfected constructs. Although the basis of reversed levels of pros-S and pros-L in vivo compared to the in vitro system is unknown, reduction in pros-S levels and increase in pros-L levels similar to those in S2 cells are also observed when the pros substrate is spliced in vitro in HeLa nuclear extracts (data not shown).

A likely explanation for this discrepancy stems from the origin of the two cell lines. Although both were derived from late-stage Drosophila embryos, Kc cells are of hematopoietic origin (1), whereas the Schneider cell line is of macrophage origin (54). Antagonistic splicing factors have been shown to influence alternative splicing in vivo in a concentration-dependent manner (7, 70). For instance, in vivo variations in hnRNP A1 protein levels have been shown to influence 5‘ splice site choice (70). Factors can also vary between cell types (16, 71), such as levels of ASF/SF2 differing nearly 20-fold between heart and testis (24) and levels of X16/SRp20 protein being high in pre-B cells and thymus but not abundant in other cells (2). Therefore, the enhancement of pros-L splicing observed in S2 cells could be due to various levels of hnRNP SR proteins, or other splicing factors.

The location of the purine-rich element and its enhancer function resemble the situation with several previously characterized intronic enhancer elements (3, 4, 8, 9, 11, 17, 29, 30, 33, 52, 57). Intronic enhancer sequences are often purine-rich, but are diverse in sequence and distinct from exonic enhancers. For instance, purine-rich regions act as important regulatory signals for neural splicing of mammalian c-src, calcinon/CGRP, and agrin (39, 41, 68). In the c-src gene, the intronic enhancer located just downstream of the N1 exon is required for the N1 exon to be positively selected (4); it assembles a complex of proteins, including hnRNP F, hnRNP H, and KSRP (41).

While intronic enhancers generally have a complex architecture, the prospero purine-rich element sequence appears distinct in that it is composed almost exclusively of purines. Furthermore, its deletion or modification had two effects: the almost complete abolition of major-class splicing and the down regulation of minor-class splicing. Specifically, the mutant in which the entire 46-nucleotide of the purine-rich element was replaced with a non-purine-rich sequence yielded no observable major class product, while deletions within purine-rich element decreased major class splicing to a lesser degree. The purine-rich region immediately downstream of the purine-rich element was able to partially replace portions of the purine-rich element in the deletion mutants (perhaps by being brought closer to the 5‘ splice sites) but not in the replacement mutants. Thus, activating sequences in addition to the purine-rich element may exist in the vicinity of the pros twintron.

There are at least two possible mechanisms by which the prospero purine-rich element may exert its stimulatory effect: one is that it participates in formation of a secondary structure that improves the accessibility of splicing factors to the splice sites, and another is that the element serves as a target of a trans-acting factor(s). Competition experiments (Fig. 6B) showed that the stimulatory effect of the purine-rich element can be titrated by competitor purine-rich element RNA, strongly supporting the latter mechanism. We thus propose that the purine-rich element serves as a target sequence for the formation of a protein enhancer complex. The data further suggest that the purine-rich element functions as a bipartite element in which nucleotides 32 to 37 and 77 to 105 influence the overall splicing efficiency of both splicing pathways, while nucleotides 37 to 77 primarily influence the ratio of pros-S to pros-L.

Future research will focus on the factors that bind the purine-rich element and flanking sequences, their binding locations and how they function to influence spliceosome assembly and promote twintron splicing in general.

ACKNOWLEDGMENTS

We thank C. Hashimoto, Y. Huang, R. Lytle, L. Otake, and K. Tycowski for comments on the manuscript; members of the Steitz laboratory for helpful discussions; and L. Raymond for invaluable help in making the Kc nuclear extracts.

This work was supported by National Institutes of Health grant GM26154 to J.A.S., who is an investigator of the Howard Hughes Medical Institute.

REFERENCES


