Spring 5-2002

The splicing of U12-type introns is a rate-limiting step in gene expression

Abhijit A. Patel
Yale University.

The Splicing of U12-type Introns is a Rate-Limiting Step in Gene Expression

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

in Candidacy for the Degree of

Doctor of Philosophy

by

Abhijit A. Patel

Dissertation Director: Joan A. Steitz

May 2002

Abstract

The Splicing of U12-type Introns is a Rate-Limiting Step in Gene Expression

Abhijit A. Patel

Yale University

2002

Metazoan genes have recently been found to contain a novel class of introns that display non-canonical consensus sequences and are excised by a distinct splicing machinery. These introns occur very rarely, and have been called U12-type introns because recognition of their branch point sequences requires the U12 snRNP, which performs an analogous function to the U2 snRNP in splicing major-class introns. The persistence of two spliceosomes throughout virtually all of metazoan evolution suggests that the two spliceosomes play distinct and probably indispensable cellular roles. One enticing possibility is that the U12-type spliceosome functions in post-transcriptional regulation, serving as the rate-determining step in the splicing pathway of genes containing U12-type introns. To test this idea I have investigated the timing of U12-type intron removal relative to the removal of major-class introns from pre-mRNAs that contain both intron types. I have addressed this question using two different experimental approaches.

One way to evaluate the order of intron removal from a transcript is to document the relative amounts of unspliced intron sequences within a steady-state population of partially-processed transcripts: if the splicing of U12-type introns is rate limiting and occurs last, then their sequences should be more abundant than those of their major-class intron neighbors. I have developed an accurate assay based on the technique of quantitative RT-PCR to measure the relative abundance of unspliced introns within several genes. Here, I present results from the analysis of three human genes showing that in all three cases splicing of the U12-type intron proceeds more slowly than splicing of the U2-type introns from the same transcript.

The second experimental approach aimed to address the question of whether replacement of a naturally occurring U12-type intron with U2-type intron consensus sequences can affect the rate of production of mature mRNA and protein. Constructs were created which expressed either cyan or yellow fluorescent proteins only when completely spliced. The constructs contained either a U12-type or a U2-type intron in an arrangement which permitted correlation of color with type of splicing. By observing the relative intensities of the two fluorescent colors, it was possible to infer the relative efficiencies of the two splicing pathways within transfected *Drosophila melanogaster* tissue culture cells. Results of these experiments showed that replacement of a U12-type intron with canonical consensus sequences did indeed dramatically increase expression of the corresponding mRNA and protein.

These results provide direct evidence that *in vivo* gene expression can be altered by the presence of a U12-type intron and implicate the U12-type spliceosome as a potential target in the post-transcriptional regulation of gene expression.

Table of Contents

List of Figures, Tables, and Schemes

Chapter I

Introduction

Principles of Pre-mRNA Processing

Most eukaryotic genes are interrupted by non-coding sequences known as introns that are spliced from precursor messenger RNA molecules (pre-mRNAs) to produce mature mRNAs. Splicing of these pre-mRNAs occurs via two successive transesterification reactions mediated by a multicomponent complex called the spliceosome (Reviewed in Steitz et al., 1988; Moore et al., 1993; Staley and Guthrie, 1998). In the first step, cleavage of the 5' exon-intron junction is caused by nucleophilic attack on the phosphodiester bond at the junction by the 2' hydroxyl group of a branch point nucleotide. This generates a 3' hydroxyl group on the 5' exon, and a lariat intermediate with a 2'-5' phosphodiester bond between the 5' end of the intron and the branch point nucleotide. In the second step, the 3' hydroxyl group attacks the phosphodiester bond at the 3' intron-exon junction, displacing a lariat intron and producing ligated exons.

The two transesterification reactions occur within a catalytic complex, called the spliceosome, that is composed of five small nuclear RNAs (snRNAs) and approximately 100 proteins. These components exist in eukaryotic cells as RNA-protein complexes known as small ribonucleoprotein particles (snRNPs). (A number of non-snRNP protein factors are also involved in the splicing reaction.) For the majority of introns, the five snRNPs that mediate splicing are U1, U2, U4/U6, and U5. Assembly of these snRNPs

onto an intron requires recognition of conserved sequences at the 5' splice site (5' AG|GTRAGT 3') and in the region of the branch site and 3' splice site (5' YNCTNACT[Y]$_n$NYAG|G) (40>n>17, splice junctions are marked with vertical bars, invariant nucleotides are boldfaced, the branch point adenosine is underlined, Y = pyrimidine, R=purine) (Mount, 1982; Reed, 1996). Since introns are frequently several kilobases in length and the splicing consensus sequences have relatively little information content, accurate removal of introns requires remarkable precision in the recognition of the splice sites. This is achieved by a number of RNA-RNA, protein-RNA, and protein-protein interactions that are precisely orchestrated, both spatially and temporally. In addition to the snRNPs, a variety of gene-specific protein factors (many belonging to the serine/arginine repeat-containing SR protein family) are involved in splice-site recognition and spliceosome assembly (Kenan et al., 1991; Mayeda et al., 1992; Zahler et al., 1992). Often, more than one splicing pattern is found for a given gene *in vivo*, and in such cases, splice site choice is regulated by specific protein factors (Smith and Valcarcel, 2000; Black, 2000; Grabowski and Black, 2001).

For the majority of introns, the 5' splice site initially interacts with the U1 snRNP via base pairs that are formed between the intron's consensus sequences and the 5' terminal sequences of U1 snRNA (Steitz et al., 1988; Staley and Guthrie., 1998). This interaction, as well as the association of SR proteins such as SF1/ASF and U2AF with the branch point sequence and polypyrimidine tract, results in the formation of complex E (Michaud and Reed, 1991). Subsequently, the U2 snRNP recognizes and binds to the branch point sequence via base pairing interactions between the snRNA and intron (Wu and Manley, 1989; Nelson and Green, 1989) forming complex A. The U4/U6, U5 tri-

snRNP complex then associates to form complex B. Several rearrangements subsequently take place within the spliceosome, leading to formation of a catalytic core within which the transesterification reactions take place. These rearrangements include disruption of U1 base pairing with the 5' splice site and displacement of U1 from the spliceosome (Burge et al., 1999), unwinding of the base pairing interaction between U4 and U6 (Burge et al., 1999), formation of base pairs between a segment of U6 and the 5' splice site (Wassarman and Steitz, 1992; Moore et al., 1993), and formation of base pairs between U2 and U6 giving rise to two helices that are essential for catalysis (Madhani and Guthrie, 1992; Hausner et al., 1990; Datta and Weiner, 1991; Wu and Manley, 1991). As a result of these rearrangements, elements of the spliceosome are positioned to catalyze the first step of splicing, producing complex C which contains a bimolecular intermediate. The second transesterification reaction requires reconfiguration of the spliceosome to allow nucleophilic attack of the 3' hydroxyl of the 5' exon on the phosphodiester bond at the 3' intron-exon junction (Steitz and Steitz, 1993; Moore and Sharp, 1993). The U5 snRNP has been shown to base-pair with sequences in the 5' and 3' exons, and is believed to facilitate juxtaposition of the two exons for the second step of splicing (Sontheimer and Steitz, 1993; Newman, 1997). Following the second step, the ligated exons and a lariat intron are released, and the spliceosomal components dissociate and are recycled for use in the splicing of additional introns (Burge et al., 1999).

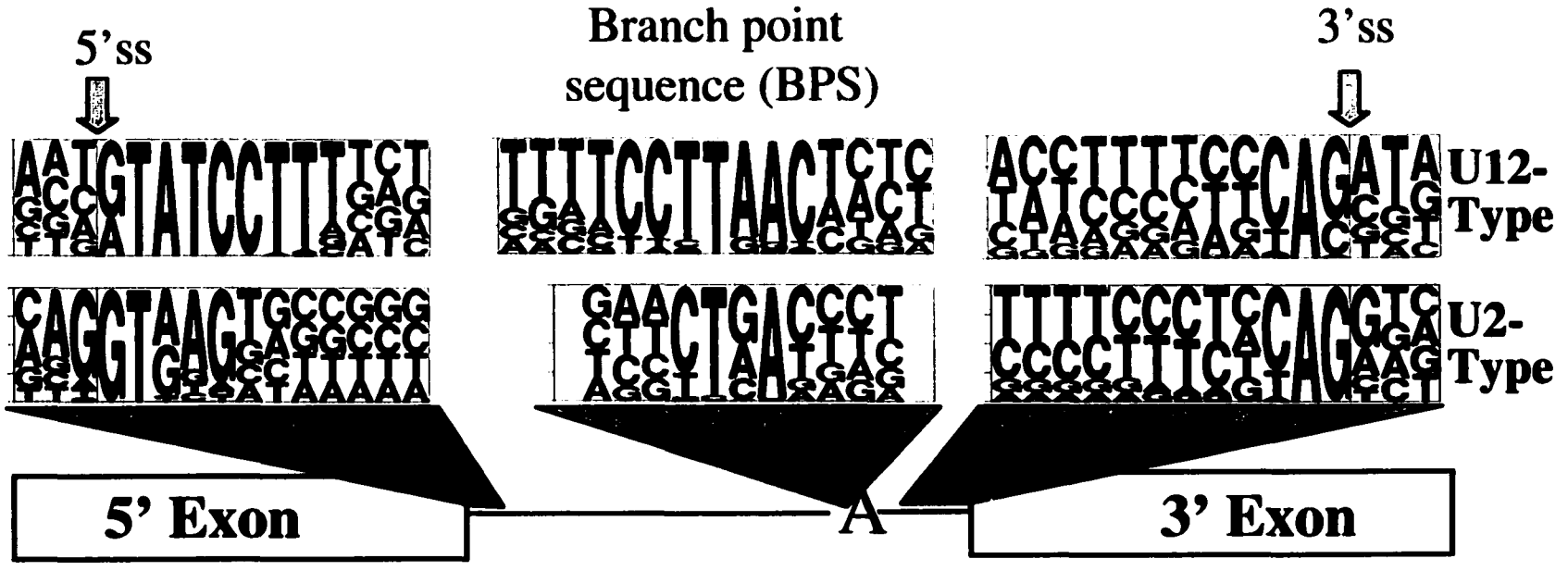
A Novel Class of Noncanonical Introns and a Divergent Spliceosome

A rare, divergent class of introns was identified by Jackson (1991) and Hall and Padgett (1994); they noticed a few introns with unusual AU and AC dinucleotides at their 5' and 3' termini, differing from the nearly invariant GU and AG termini of canonical introns. These minor-class introns were found to be further distinguishable from major-class introns based on highly conserved sequences at their 5' splice site and branch site, as well as by the lack of a polypyrimidine tract (Figure 1) (Hall and Padgett, 1994). Sequence complementarity suggested that the 5' splice sites and branch sites of these introns might interact with U11 and U12 snRNPs, respectively (Hall and Padgett, 1994). It was subsequently shown that the excision of minor-class introns is indeed mediated by a distinct low-abundance spliceosome. The U12-type spliceosome contains U11, U12, U4atac, and U6atac snRNPs, which are the functional analogues of the major-class U1, U2, U4, and U6 snRNPs, while U5 snRNP participates in the splicing of both U12-type and the classical (U2-type) intron (Figure 2) (Hall and Padgett, 1996; Tarn and Steitz, 1996a, 1996b). While AU-AC termini were initially considered to be a defining feature of minor-class introns, mutation to GU-AG termini does not interfere with splicing via the U12-dependent pathway (Dietrich et al, 1997). In fact, surveys of genomic databases have shown that the majority of naturally occurring U12-type introns have GU-AG termini (Sharp and Burge, 1997; Burge et al. 1998).

The splicing of U12-type introns is carried out in a similar manner to that of U2-type introns, with subtle, but probably important differences. The minor-class snRNAs possess relatively little sequence identity with their major class snRNA analogues, but

Figure 1. Comparison of Splicing Consensus Sequences in U12-type and U2-type Introns (adapted from Burge et al., 1999)

Consensus sequences are displayed using a Pictogram program representation, where the frequency of occurrence of each base at a particular sequence position is represented by the relative height of the corresponding letter, in descending order of frequency. The U12-type consensus sequences show greater information content at the 5' splice site and branch site, and typically lack a polypyrimidine tract. The typical distance between the branch site and 3' splice site is in the range of 10 to 20 nucleotides for U12-type introns, compared to 10 to 50+ for U2-type introns (Sharp and Burge, 1997).



From Burge, Padgett & Sharp (1998)
Molecular Cell, 2: 773-785.

Figure 2. Splicing Pathways of the U12-type and U2-type Spliceosomes (figure made by M. Frilander)

Comparison of the analogous pathways of spliceosome assembly and catalysis for both classes of introns. The minor-class snRNPs, U11, U12, U4atac, and U6atac are the functional analogues of the major class snRNPs, U1, U2, U4, and U6, respectively. The U5 snRNP participates in both splicing pathways. Note that U11 and U12 are pictured as entering the spliceosome as a di-snRNP complex. The branch point adenosine is indicated, and YY represents the polypyrimidine tract.

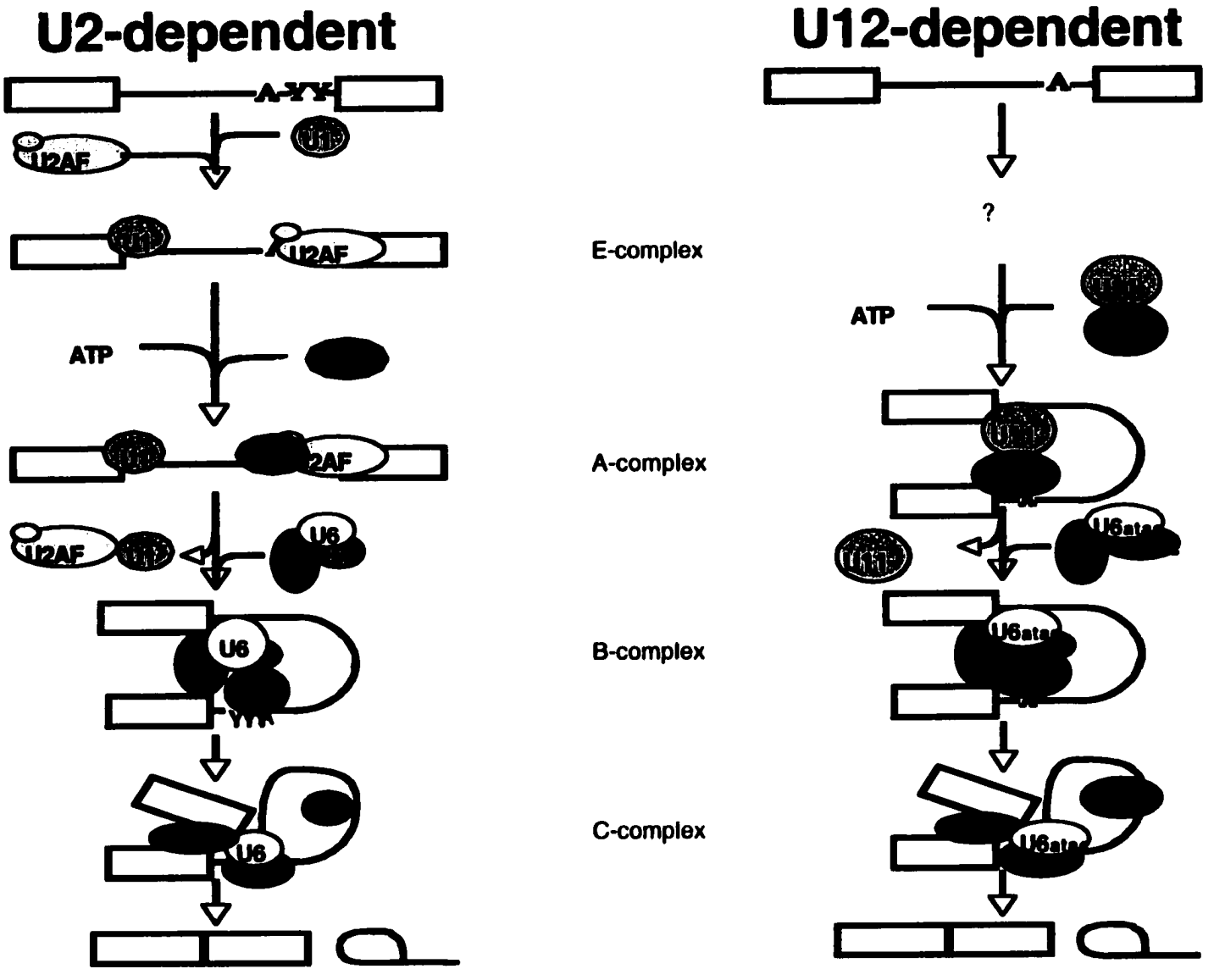
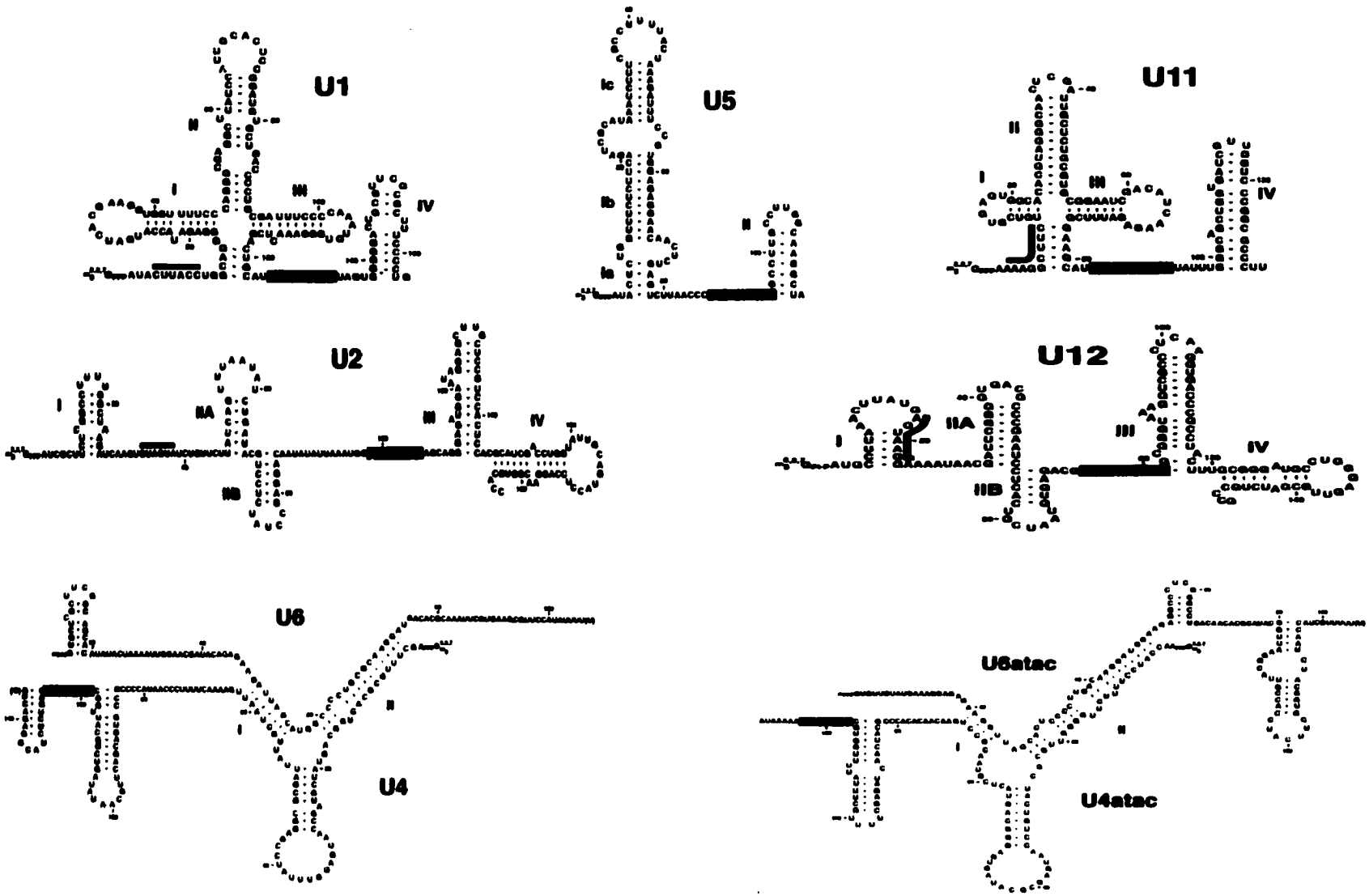


Figure 3. Primary and Secondary Structures of the snRNA Components of Both Spliceosomes (from Burge et al., 1999)

Nucleotide positions for each snRNA are indicated in Arabic numerals, whereas Roman numerals indicate intrastrand or interstrand RNA helices. Sm protein binding sites are shaded, and sites that base-pair with the intron consensus sequences are indicated by dark lines. All snRNAs are shown in 5' to 3' orientation with the exception of U4 and U4atac which are shown 3' to 5'.



From Burge, C. B., Tuschl, T., and Sharp, P. A. (1999) in The RNA World, CSHL Press, pp. 525-560.

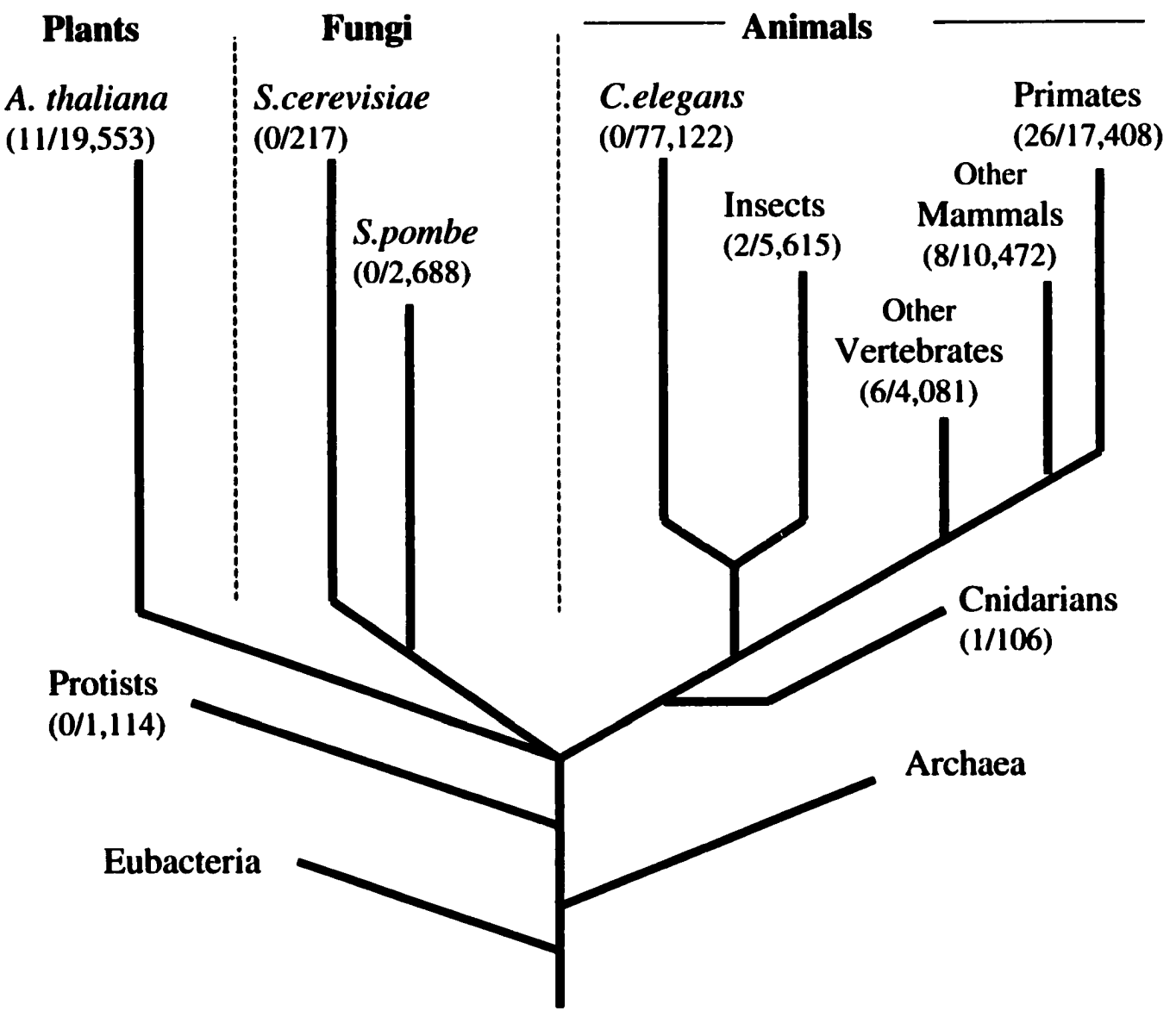
display remarkable similarity in their predicted secondary structures (Figure 3) (Tarn and Steitz, 1997). However, close examination of the sequences reveals that the most highly conserved core sequences of U6 are also present in U6atac. The minor class snRNPs are known to be approximately 100-fold less abundant than the major-class snRNPs in mammalian cell nuclei (Montzka and Steitz, 1988; Tarn and Steitz, 1996a). The U11 and U12 snRNPs, unlike U1 and U2, are known to interact with each other in HeLa cell extracts, forming a di-snRNP complex (Montzka-Wassarman and Steitz, 1992; Frilander and Steitz, 1999). In fact, while U1 and U2 interact separately with intron consensus sequences, initial recognition of minor-class introns *in vitro* has been shown to require simultaneous U11/5' splice-site and U12/branchpoint interactions, suggesting that U11 and U12 are delivered to the intron as a di-snRNP (Frilander and Steitz, 1999). Since U12-type introns do not possess a polypyrimidine tract, it is reasonable to expect U2AF not to be involved in U12-type splicing. However, other SR proteins have been shown to activate the splicing of U12-type introns (Hastings and Krainer, 2001).

A Phylogenetic Examination of U12-type Introns

Examples of U12-type introns are found in virtually all metazoan taxa, including vertebrates, insects, plants, and cnidarians, but are absent from simpler eukaryotes such as *C.elegans, S. cerevisiae, S. pombe*, and protists (Figure 4) (Burge et al, 1998). A total of 60 non-redundant U12-type introns were identified in a search of all available genomic sequences in 1998 (Burge et al, 1998), and a more recent search of the human genome yielded 404 examples (Levine and Durbin, 2001). Thus, the frequency of occurrence of

Figure 4. The Phylogenetic Distribution of U12-type Introns (from Burge et al., 1998)

The frequency of occurrence of U12-type introns in different taxa are shown. The number of identified U12-type introns out of the total number of complete introns analyzed is indicated below the name of each taxon. The taxa are organized according to evolutionary relationship trees based on 18S rRNA conservation. Eubacteria and Archea do not have spliceosomal introns.



U12-type introns relative to U2-type introns is in the range of 0.15% to 0.34% in vertebrates, and lower in other taxa (Burge et al, 1998; Levine and Durbin, 2000). These introns can be found at any position, and almost always co-exist with U2-type introns within the same gene. In humans, the mean length of U12-type introns (3600 bp) is similar to that of U2-type introns (4130 bp) (Levine and Durbin, 2001), but does not yet include introns exceeding 20kb.

By comparing introns at homologous positions (in terms of codon location and phase) in homologous genes from different species, Burge et al. (1998) were able to conclude that U12-type introns occurred much more frequently in early evolutionary history and were either lost or converted to U2-type introns over time. (Conversion from U2-type to U12-type is considered to be extremely improbable since U12-type consensus sequences are more highly constrained.) Further, they propose a fission/fusion model in which the U12-type and U2-type splicing systems evolved in separate lineages, each harboring a single splicing system, which later fused to become a progenitor of higher eukaryotes. Merging of the genetic material resulted in a genome containing both intron types and both spliceosomes. Evidence for this model is provided by the observation of four genes that each harbor two U12-type introns, out of 56 total U12-type intron-containing genes that they identified. They argue that given the extreme rarity of these introns, such clustering within certain genes is highly improbable if the introns were randomly inserted. Rather, they propose that the early progenitor organism had certain genes containing only U12-type introns that were contributed from the U12-type lineage. Over time, the majority of the introns within these genes were either lost or converted to

U2-type introns, but the U12-type introns that have persisted are found to be clustered within this subset of genes that originated from the U12-type organism.

Despite this tendency for conversion and loss, Burge et al. (1998) have identified several remarkable instances of U12-type introns that are conserved between homologous genes from highly diverged species. One such example is the Huntington's disease gene in which the last intron of sixty-six introns is a U12-type intron in both pufferfish (*Fugu*) and humans. Another particularly striking example is the second intron of the sodium channel α subunit gene that is conserved between humans and jellyfish, organisms that diverged at least 600-800 million years ago (Spafford et al., 1998). Perhaps even more surprising is the finding that in spite of the rarity of these introns, several examples were noted of U12-type introns at nonhomologous positions in paralogous genes (Burge et al., 1998).

Hypothesis: Could the Splicing of U12-type Introns be a Rate-Limiting Step in Gene Expression?

The conclusions of Burge et al. (1998) led me to wonder whether perhaps the few U12-type introns that have resisted conversion or loss, may have been retained over evolutionary time because their presence was important in some functional way for the genes that harbor them. Have the remaining introns remained because they play an indispensable role in the cell? One compelling possibility, I thought, was that the splicing of U12-type introns might be important for the regulation of expression of the genes that contain them. That is, if U12-type introns are spliced more slowly than

neighboring U2-type introns in the same pre-mRNA, their removal may represent the rate-determining step in the splicing pathway of those genes.

Thesis Plan:

This thesis aims to directly test the hypothesis that U12-type introns are excised more slowly from pre-mRNAs than neighboring U2 type introns, and may therefore be a point of regulation of expression of the encoded proteins. A number of existing observations hinted at the validity of this hypothesis. First, *in vitro* splicing of U12-type introns was consistently reported to be extremely inefficient by different groups using different splicing substrates (Wu and Krainer, 1996; Tarn and Steitz, 1996b). Also, the snRNPs unique to the minor-class spliceosome are approximated to be 100-fold less abundant than their major-class counterparts in mammalian cell nuclei (Montzka and Steitz, 1988; Tarn and Steitz, 1996a), suggesting that splicing rates might be reduced because of the lower concentration of spliceosomal components. Perhaps most informative were a series of papers from Irene Bozzoni and colleagues outlining their efforts to understand the low efficiency splicing of the "regulated third intron" of the *Xenopus laevis* ribosomal protein L1a gene, an intron that we now know to be of the U12 type (Bozzoni et al., 1984; Pierandrei-Amaldi et al., 1987; Caffarelli et al., 1987, 1992; Fragapane et al., 1992). They found that upon microinjection of cloned RPL1a genomic sequences into *Xenopus* oocytes, partially spliced pre-mRNAs containing this intron (and to a lesser extent, the second intron) accumulated in the nuclei and produced little extra

protein (Bozzoni et al., 1984). They even reported detecting the presence of endogenous RPL1a pre-mRNAs at low levels in uninjected oocytes (Bozzoni et al., 1984).

In order to establish whether similar observations would be found under physiologic conditions, I believed that it was important to examine whether U12-type introns were removed most slowly from endogenous cellular transcripts. Under conditions of overexpression, it is possible that the U12-type splicing machinery could be saturated, and thereby hinder the splicing of U12-type introns. Therefore, a method was developed for assessing the splicing rates of introns within endogenous pre-mRNAs *in vivo*. To investigate whether the hypothesis could be generalized, I evaluated the pre-mRNAs of several U12-type intron-containing genes. A quantitative RT-PCR-based method was developed to measure the relative amounts of each unspliced intron within a steady-state population of pre-mRNAs, since it can be demonstrated that the amount of each unspliced intron in a given gene is inversely proportional to its rate of splicing (assuming minimal degradation). Quantitative RT-PCR was utilized because of its extreme sensitivity, since unspliced introns were expected to be present in very small quantities.

I chose three human genes for analysis based on the fact that they had a small number of total introns, which simplified the experiments. By measuring and comparing the levels of each unspliced intron from these three endogenous human transcripts prepared from growing tissue culture cells, I was able to show that the U12-type introns are indeed spliced more slowly in each case. This work is described in Chapter II of this thesis.

While carrying out the above experiments, I realized that the methodology had certain important limitations. If it turned out to be true that U12-type splicing was a rate-limiting step in gene expression, one might expect the efficiency of U12-type splicing to be modulated within cells. Perhaps the activity of the minor-class spliceosome might be modulated during different stages of development or different stages of the cell cycle or in a tissue-specific manner. The RT-PCR assay that I developed was not well-suited to investigating such changes. Repeating the quantitation for several different RNA samples seemed extremely cumbersome, but this would have been necessary to detect changes in splicing efficiency within different tissues or at different developmental stages. Also, since RNA was prepared from a population of cells, any differences in splicing efficiency within that population (such as differences during the cell cycle) might be missed because of averaging of the measurements over the heterogeneous population. Therefore, I thought it necessary to develop a complementary assay for splicing efficiency that could be followed in real-time within individual cells.

I thus employed fluorescent reporter proteins, whose expression was dependent upon splicing of a U12-type versus a U2-type intron, to monitor splicing efficiency. These fluorescent protein-coding constructs were designed to correlate the production of two different colored fluorescent proteins with splicing of the two different intron types. Therefore, the ratio of the two different colored fluorescence intensities from each cell would reflect the relative efficiency of each type of splicing. Since the proteins were destabilized with "PEST" sequences (yielding a 2-hour half-life), dynamic changes in splicing efficiency would be observable within cells. The genomic sequences used in the constructs were from a segment of a U12-type intron-containing *Drosophila*

melanogaster gene. A *Drosophila* gene was chosen because I hoped to exploit the well-established P-element-mediated transgene insertion technology to create transgenic flies containing these constructs in each of their cells. In this way, I hoped to follow any changes in U12-type splicing activity within different tissues or at different stages of *Drosophila* development by observing fluorescence intensities. However, the transgenic flies failed to produce sufficient fluorescent protein to be detected with a fluorescence microscope, and after several attempts at troubleshooting, the creation of transgenic flies was abandoned. Nevertheless, the constructs were able to produce detectable fluorescence when transfected into *Drosophila* tissue culture cells (probably because of an increased copy number of the plasmids within each cell). Analysis of the fluorescence from these constructs within the transfected cells demonstrated that conversion of a U12-type intron to a U2-type intron does dramatically increase production of the mature mRNA and encoded protein. Chapter III of this thesis describes the experiments with the transfected *Drosophila* cells, and the unsuccessful experiments with the transgenic flies are described in Appendix A.

Chapter II

U12-type Introns Are Spliced more slowly than U2-type Introns from Endogenous Pre-mRNAs *in vivo*

Summary

In the second chapter, I describe experiments designed to investigate whether U12-type introns are excised more slowly from endogenous cellular pre-mRNAs than their U2-type intron neighbors. One way to evaluate the order of intron removal from a transcript is to document the relative amounts of unspliced intron sequences within a steady-state population of partially-processed transcripts: if the splicing of U12-type introns is rate-limiting and occurs last, then their sequences should be more abundant than those of their major-class intron neighbors. I have developed an accurate assay that employs quantitative RT-PCR to measure the relative abundance of unspliced introns within the transcripts of three human genes. This assay requires the use of *in vitro* transcribed RNAs that serve as internal standards to control for variations in RT-PCR amplification efficiency. I was able to demonstrate that each internal control was amplified with an efficiency that was virtually identical to that of the corresponding cellular RNA sequence. Then, co-amplification of known amounts of these control RNAs with fixed amounts of total RNA from growing human tissue culture cells revealed that unspliced U12-type introns did accumulate to a greater extent than unspliced U2-type introns within the transcripts of three human genes. Because U12-type introns were independently found to be the slowest introns to be spliced from these human genes, I contend that the same should hold true for other U12-type intron-containing genes, and

that the splicing of U12-type introns therefore appears to be important for the regulation of gene expression.

Results

A quantitative RT-PCR method to assess relative rates of intron removal

I devised a method for evaluating the relative rates of intron removal from a particular transcript by determining the abundance of intron sequences within a steady-state population of partially processed cellular pre-mRNAs. If the excision of a U12-type intron occurs more slowly than that of major-class introns in the same pre-mRNA, then the amount of unspliced U12-type sequences should be higher than each of the other intron sequences. Indeed, it can be mathematically demonstrated that the amount of each unspliced intron within a steady-state pre-mRNA population is inversely proportional to the relative rate of removal of that intron, assuming minimal pre-mRNA degradation (see Mathematical Scheme 1). Because the cellular levels of unspliced pre-mRNAs were expected to be very low, the extreme sensitivity of quantitative RT-PCR was exploited to analyze total RNA isolated from growing human tissue culture cells (HeLa or SK Hep).

As schematized in Figure 5, for each intron, reverse transcription was carried out using short, sequence-specific primers. PCR was then performed using primer pairs that amplify only unspliced sequences: each forward primer was complementary to 5' exon sequences and each reverse primer was complementary to downstream intron sequences. For each intron, control RNA standards were needed that possessed RT and PCR amplification efficiencies closely matched to those of the cellular RNAs being analyzed. Therefore, I synthesized RNAs that were identical to the target sequences to be amplified

Mathematical Scheme 1. Proof that the concentration of an unspliced intron is inversely proportional to its splicing rate:

A simplified reaction scheme is shown for the production of mRNA from DNA:

$$\text{DNA} \xrightarrow[\text{Transcription}]{k_{tx}} \underset{[US]}{\text{Unspliced pre-mRNA}} \xrightarrow[\text{Splicing}]{k_{sp}} \text{Spliced mRNA}$$

Where $[US]$ represents the concentration of a given unspliced intron,
k_{tx} represents the rate of transcription of that gene,
and k_{sp} represents the rate of splicing of that intron.

The following rate law can be written for the above scheme:

$$\frac{d[US]}{dt} = k_{tx} - k_{sp}[US]$$

$$\frac{d[US]}{dt} = -k_{sp}\left([US] - \frac{k_{tx}}{k_{sp}}\right)$$

$$\frac{d[US]/dt}{\left([US] - \frac{k_{tx}}{k_{sp}}\right)} = -k_{sp}$$

$$\frac{d\ln\left|[US] - \frac{k_{tx}}{k_{sp}}\right|}{dt} = -k_{sp}$$

$$\int d\ln\left|[US] - \frac{k_{tx}}{k_{sp}}\right| = -\int k_{sp}\,dt$$

Integrating both sides, where C represents an integration constant:

$$\ln\left|[US] - \frac{k_{tx}}{k_{sp}}\right| = -k_{sp}t + C$$

$$\left|[US] - \frac{k_{tx}}{k_{sp}}\right| = e^{C} \cdot e^{-k_{sp}t}$$

Since C is an arbitrary constant, we can set $C'=e^{C}$

$$[US] = \frac{k_{tx}}{k_{sp}} + C' e^{-k_{sp} t}$$

Regardless of the value of C', at steady state when $t \rightarrow \infty$:

$$\boxed{[US] = \frac{k_{tx}}{k_{sp}}}$$

Since there are many introns in a transcript, and each intron has the same rate of synthesis (k_{tx}), only the rates of splicing (k_{sp}) should determine the relative steady-state concentrations of the unspliced introns ([US]), assuming that there is minimal intron degradation by other pathways.

Therefore, at steady state:

$$[US_{Intron1}] = \frac{k_{tx1}}{k_{sp1}}, \qquad [US_{Intron2}] = \frac{k_{tx2}}{k_{sp2}}, \qquad [US_{Intron3}] = \frac{k_{tx3}}{k_{sp3}}, \quad etc...$$

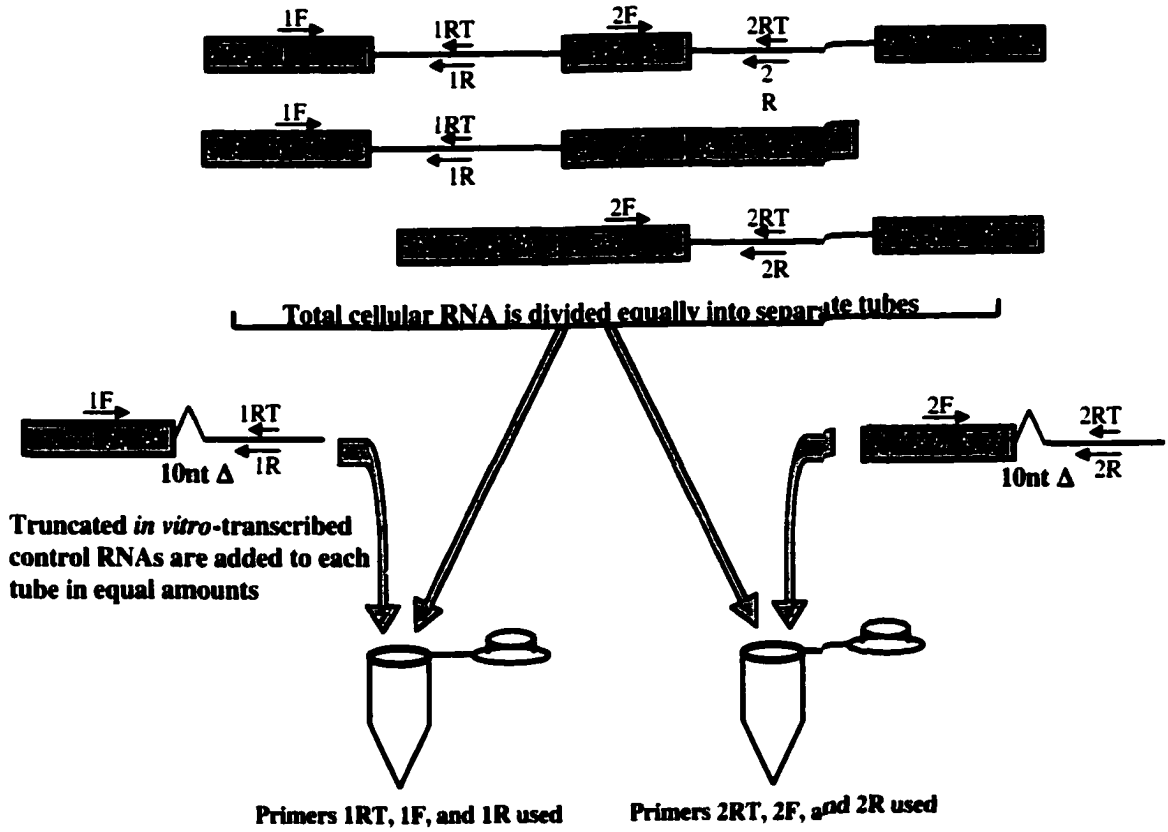
But since $k_{tx1}=k_{tx2}=k_{tx3}$=etc..

$$\boxed{Transcription\ Rate = [US_{Intron1}]k_{sp1} = [US_{Intron2}]k_{sp2} = [US_{Intron3}]k_{sp3} = etc...}$$

Thus, the relative concentration of a given unspliced intron is inversely proportional to its relative splicing rate.

Figure 5. Assessing *in vivo* Splicing Rates by Quantitative RT-PCR

Quantitation of unspliced introns from total cellular RNA containing pre-mRNAs of a hypothetical two intron-containing gene is schematized. *In vitro*-transcribed control RNA standards, truncated relative to the cellular RNA sequence by 10 nucleotides, are added in equal amounts to separate reaction tubes containing a fixed amount of total cellular RNA. Reverse transcription is performed using short primers complementary to intronic sequences (1RT or 2RT). cDNAs are then PCR-amplified using primer pairs that only amplify unspliced sequences (1F/1R or 2F/2R). The cellular and control RNA amplicons are then electrophoretically separated, and the quantity of each unspliced intron is inferred from the ratio of the amplicons produced.

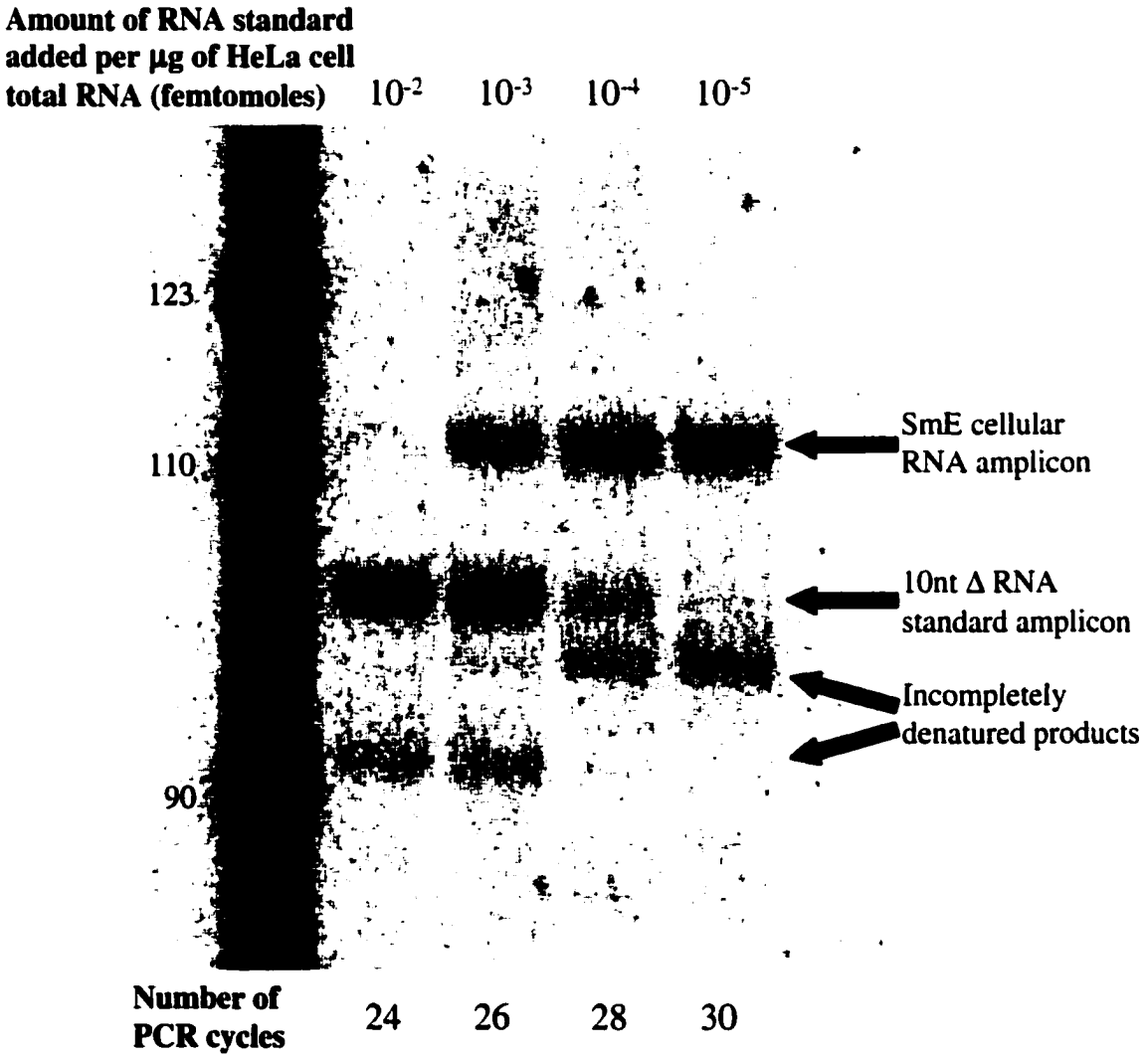


with the exception of an internal 10-nucleotide deletion. Because amplification variability can arise at the reverse transcription step due to RNA secondary structure impeding the polymerase (Freeman et al., 1999), I designed the RNA standards to mimic the local secondary structure of each target by including sequence from most of the upstream exon and >195nt of the intron. Co-amplification of the cellular RNA population with a known quantity of each RNA standard therefore yielded amplicons (105 to 164nt long) that differed in size by 10 nucleotides. By comparing their production after separation on a polyacrylamide gel, the amount of the cellular RNA sequence present was quantitated.

For each gene studied, the concentration of unspliced U12-type intron within a fixed amount of total cellular RNA was determined by titration with varying amounts of RNA standard (Figure 6). Then, that amount of RNA standard for each of the gene's introns was added to separate tubes containing identical amounts of total RNA (Figure 5). If the U2-type introns are spliced more rapidly than the U12-type intron, then those unspliced sequences will be less abundant than the sequences of the corresponding RNA standards, as reflected in the ratio of the co-amplified RT-PCR products generated in each tube.

Figure 6. Quantitation of the Unspliced U12-type Intron of SmE

The concentration of unspliced U12-type intron within total cellular RNA for each of the three human genes was determined by titration with varying amounts of RNA standard. As an example, determination of the concentration of the unspliced first intron (U12-type) of SmE is shown. Varying amounts of RNA standard were added to a fixed amount of total cellular RNA, and the samples were co-amplified by RT-PCR. PCR reactions were carried out to different numbers of cycles depending on the concentration of the added RNA standard. to ensure that PCR amplification was stopped in the exponential phase.



The most slowly spliced introns from endogenous human pre-mRNAs in vivo are U12-type

The three human U12-type intron-containing genes chosen for study had a relatively small number of total introns, making the analysis less cumbersome. These genes were: 1) SmE, a member of the core complex of proteins in spliceosomal snRNPs (Stanford et al., 1988; Accession: AL356980); 2) E2F2, a transcription factor involved in cell cycle regulation (Ivey-Hoyle et al., 1993; Accession: AL021154); and 3) INSIG1, a gene of unknown function that may be important in liver regeneration (Peng et al., 1997; Accession: U96876). These genes contain four, six, and four total introns, with the U12 type intron as the first, fourth, and second intron, respectively (Figure 7A).

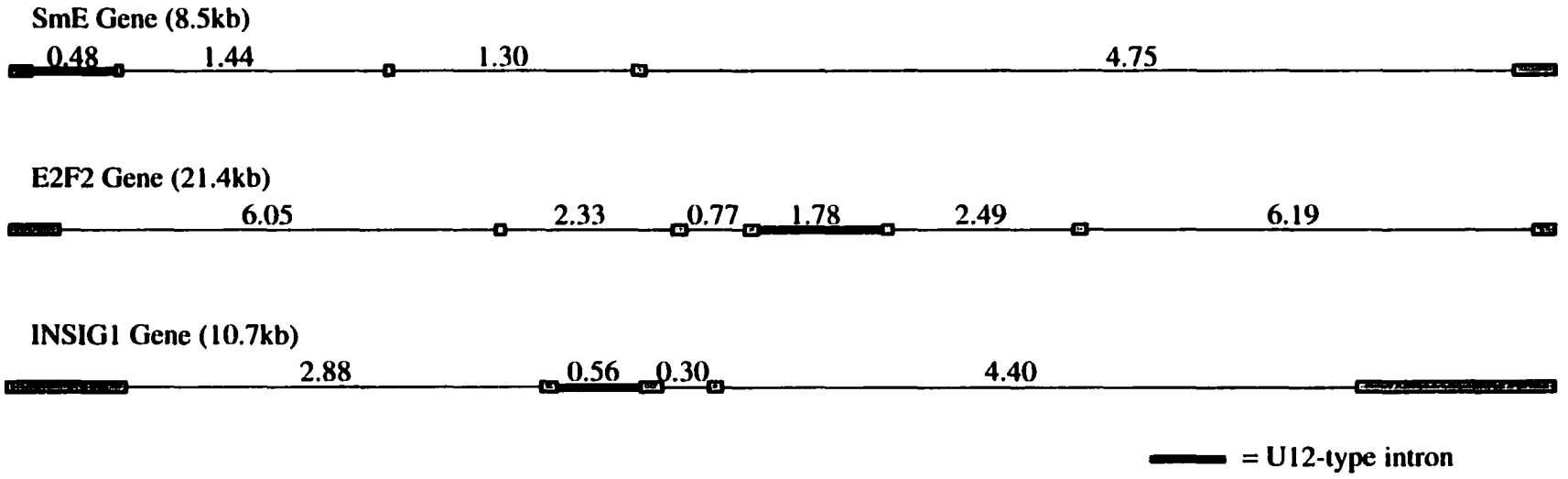
RT-PCR quantitation of the unspliced introns from these three genes in the cellular RNA population is presented in Figure 7B. Each lane shows two bands separated by 10 nucleotides. The upper band is the amplicon from the cellular RNA, while the lower band is that produced from the added RNA standard. Although the absolute intensities of the bands can vary from lane to lane depending on the amplification efficiencies of specific primer pairs, the ratio of intensities within any given lane provides an accurate measure of initial RNA quantities. Since equal amounts of the RNA standards are used in each reaction for a given gene, comparing the ratios across lanes provides a direct comparison of the relative amounts of each unspliced intron.

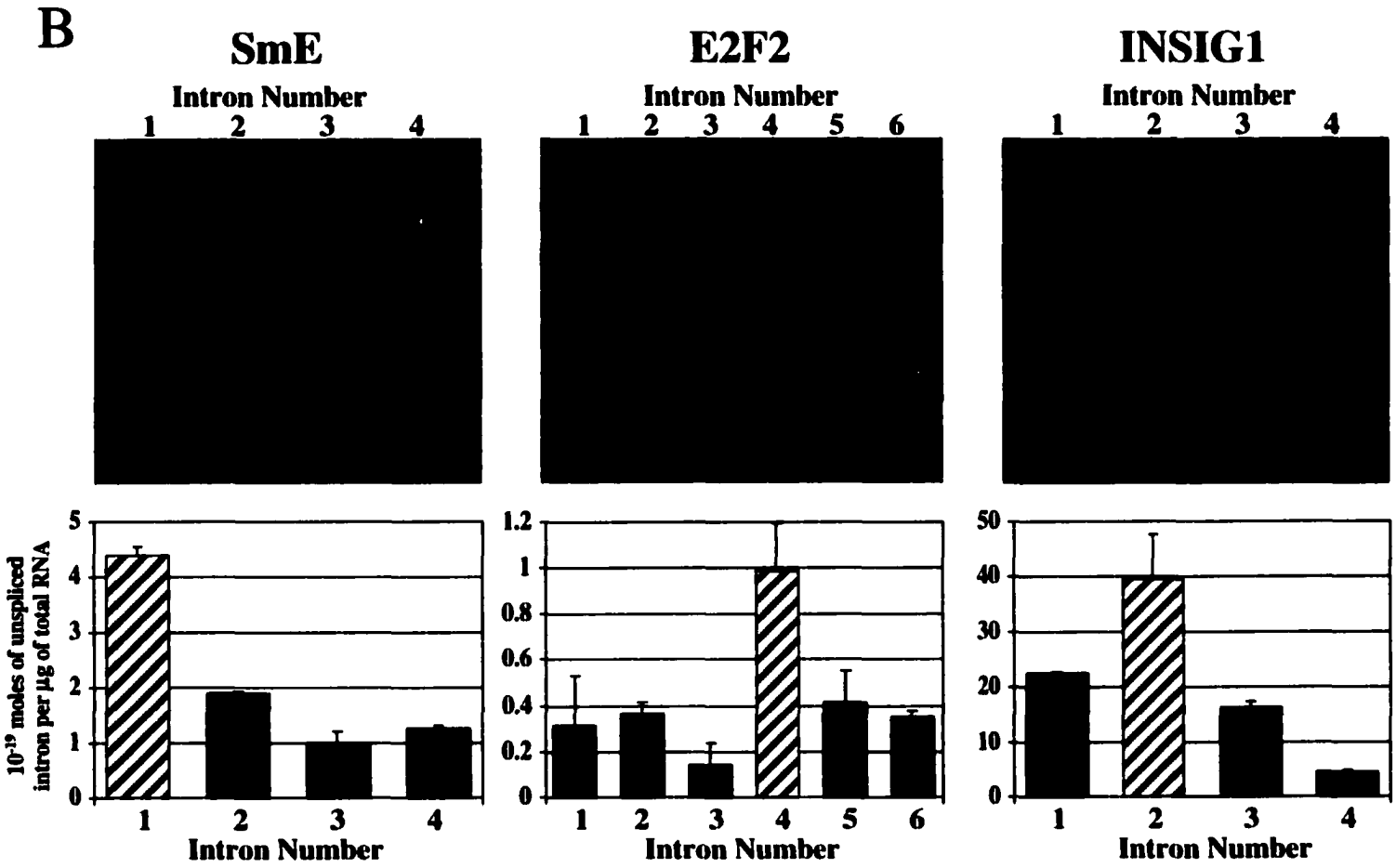
Graphical presentation of these intensity ratios (Figure 7B) shows that the unspliced U12-type intron sequences are more abundant than the neighboring U2-type intron sequences for all three pre-mRNAs analyzed. To confirm that the amplification

Figure 7. The U12-type Intron Is Excised most slowly from Three Endogenous Human Pre-mRNAs

(A) The intron-exon structure is depicted for the three human genes, SmE, E2F2, and INSIG1. Shaded boxes represent exons, thin lines represent U2-type introns, and thick lines represent U12-type introns, with intron lengths (in kb) indicated above. (B) Above are shown the gels from which the ratios of cellular to control (10 nucleotides shorter) amplicons were measured. The amounts of each unspliced intron within the pre-mRNA population from growing HeLa cells (for SmE and E2F2) or SK Hep cells (for INSIG1) are graphed below. Solid bars represent U2-type introns and hatched bars represent U12-type introns, with error bars indicating the standard deviation of two experiments. $4.56x10^{-19}$ moles, $1.01x10^{-19}$ moles, and $4.47x10^{-18}$ moles of control RNA standards were added per microgram of total cellular RNA for the analysis of introns from SmE, E2F2, and INSIG1, respectively.

A

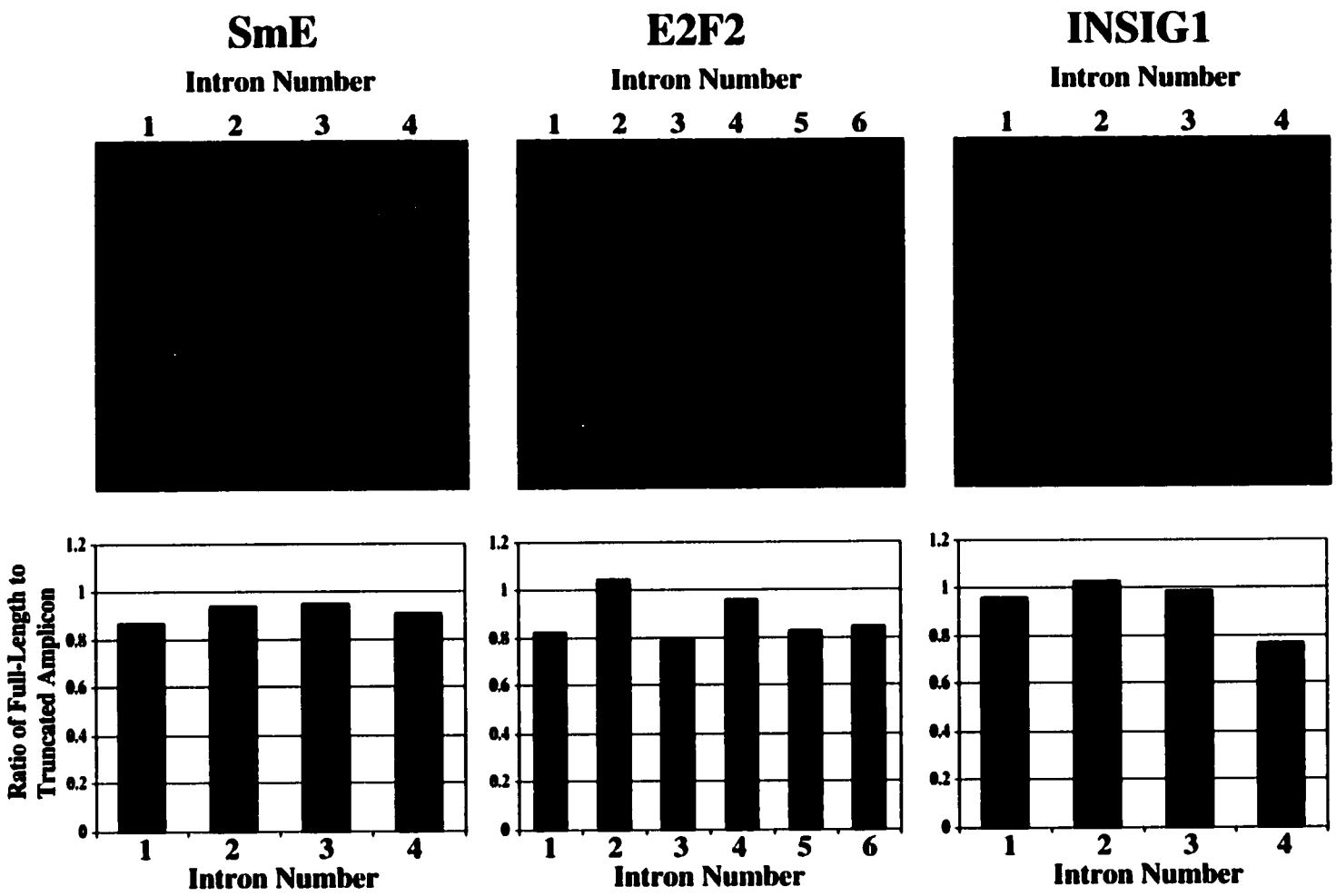




efficiencies for each intron and its corresponding 10-nucleotide deleted RNA standard were equivalent, equal amounts of *in vitro* transcribed full-length and truncated RNAs corresponding to each intron were co-amplified. The appearance of pairs of amplicons of approximately equal intensity in each lane (Figure 8) allows the conclusion that the U12-type introns within the *in vivo* transcripts of three human genes are indeed spliced more slowly than their U2-type neighbors.

Figure 8. Confirmation that Amplification Efficiencies of Truncated and Full-Length RNAs Are Equal

Control co-amplification of equal amounts of *in vitro*-transcribed full-length and 10-nucleotide truncated RNAs shows approximately equal production of amplicons (ratio of 0.91±0.08). Gels showing band intensities are displayed above graphs depicting the intensity ratio of each amplicon pair. 10^{-19} moles of each RNA species were RT-PCR amplified.



Discussion

The data presented substantiate the hypothesis that U12-type introns are spliced from pre-mRNA transcripts with slower kinetics than canonical introns, and therefore represent a probable point of post-transcriptional regulation of gene expression. This suggests that the persistence of U12-type introns within specific genes over large spans of evolutionary time could be a consequence of their function as post-transcriptional regulators of the genes that harbor them.

Although I analyzed endogenous splicing rates for the transcripts of only three human genes, I argue that our results are generalizable. The observed pattern of splicing can be extrapolated to other U12-type intron containing genes if it can be demonstrated that coincidental observation of the same pattern within the studied sample is highly improbable. The likelihood of finding by random chance that the U12-type intron is spliced most slowly from genes containing 4, 4, and 6 total introns is 1/4, 1/4, and 1/6, respectively. However, the probability of independently finding this pattern for all three genes is the product of the individual probabilities, which is 1/96. Therefore, the consistency I observe is unlikely to occur randomly and is thus probably representative of other genes containing U12-type introns.

Indeed, another U12-type intron containing gene, the ribosomal protein L1a gene from *Xenopus laevis*, was the focus of several papers from Irene Bozzoni and colleagues (Bozzoni et al., 1984; Pierandrei-Amaldi et al., 1987; Caffarelli et al., 1987, 1992; Fragapane et al., 1992). These papers investigated the low efficiency splicing of the "regulated third intron" of the RPL1a gene, an intron which, unbeknownst to the

investigators at the time, was of the U12-type. Not only did partially spliced pre-mRNAs containing this intron (and to a lesser extent, the second intron) accumulate and produce little extra protein upon microinjection of cloned RPL1a genomic sequences into *Xenopus* oocytes, but endogenous partially-spliced RPL1a pre-mRNAs were detected at low levels in uninjected oocytes (Bozzoni et al., 1984).

These reported findings provided the initial clues to support the validity of my hypothesis. Therefore, to confirm these findings using my more sensitive RT-PCR assay, I tried to measure the endogenous levels of unspliced introns from the *Xenopus laevis* RPL1a gene in total RNA prepared from oocytes. Preliminary experiments did indicate that the unspliced U12-type intron was indeed present at higher levels than the other eight U2-type introns in the same gene (data not shown). However, these experiments proved to be difficult to interpret because of the presence of snoRNAs, which can be removed via a pathway that competes with splicing (Cafarelli et al., 1994), within several of the RPL1a introns (including the U12-type intron). Additionally, the possibility of amplifying sequences from both gene copies in the pseudo-tetraploid genome of *Xenopus laevis* further complicated the interpretation of the quantitative RT-PCR results. Thus, human genes were chosen for subsequent analysis.

Among the genes analyzed, no obvious correlation could be discerned between intron length and efficiency of splicing (Figure 7). While the U12-type introns within these three chosen genes were, on average, shorter than the U2-type introns, in two out of three genes they are not the shortest intron. It is therefore unlikely that the slower splicing of these U12-type introns is related to their lengths. Also, no length correlation was observed among the splicing rates of the U2-type introns within these genes,

consistent with previously reported findings (Gardner et al., 1988; Gudas et al., 1990; Kessler et al., 1993; Wetterberg et al., 1996).

Also, no relationship was discernible between intron position and splicing efficiency. Since pre-mRNAs are transcribed in a 5' to 3' direction, and spliceosomes have been reported to assemble cotranscriptionally (Beyer and Osheim, 1988; Bauren and Wieslander, 1994; Kiseleva et al., 1994), one might expect the splicing of upstream introns to precede that of downstream introns. Wetterberg et al. (1996) have reported excision of introns from the Balbiani ring 3 (BR3) gene of the dipteran *Chironomus tentans* in an overall 5' to 3' polarity that was established during transcription, whereas individual neighboring introns were removed at very different rates, in a preferred order that was not necessarily 5' to 3'. My data show no such pattern, but since my assay observes unspliced introns from a steady-state pre-mRNA population, detection of such a pattern should not be expected. Even if the upstream introns are spliced more quickly within a given transcript, the total RNA population used in my assay also contains partially transcribed RNAs that increase the measured quantity of unspliced 5' introns, balancing any 5' to 3' splicing bias.

The limitations of this RT-PCR-based assay became apparent when I began thinking about whether the activity of the U12-type splicing machinery might be modulated in a physiologically important manner. I postulated that if the splicing of U12-type introns was a rate-limiting step in gene expression, then modulation of the efficiency of that rate-limiting step might be a way to up- or down-regulate the expression of the encoded proteins. U12-type splicing activity might be regulated in a tissue-specific manner or at different stages of the cell cycle or at different stages of

embryogenesis. If any of these possibilities were true, my RT-PCR method was not well-suited for assessing such changes. Averaging of the quantities of unspliced introns over a heterogeneous population of cells might overlook potential changes in splicing efficiency during the cell cycle. Also, analyzing RNA from different tissues or from embryos developed to different stages would have been difficult using the labor-intensive RT-PCR strategy. Moreover, it would be difficult to predict, *a priori*, which tissue samples or which embryonic stages would exhibit differences in U12-type splicing activity. Therefore, I attempted to develop a method to follow U12-type versus U2-type splicing activity in real-time within individual cells or within a whole organism.

Chapter III

Conversion of a Native U12-type Intron to a U2-type Intron Increases mRNA and Protein Production *in vivo*

Summary

Described in this chapter are experiments that aim to address the question of whether replacement of a naturally occurring U12-type intron with U2-type intron consensus sequences can affect the rate of production of mature mRNA and protein. I created constructs that express either cyan or yellow fluorescent proteins only when completely spliced. The constructs contain either a U12-type or a U2-type intron in an arrangement that permits correlation of fluorescent color with type of splicing. By observing the relative intensities of the two fluorescent colors, it is possible to infer the relative efficiencies of the two splicing pathways within transfected *Drosophila melanogaster* tissue culture cells. Observation of cells by fluorescence microscopy showed that replacement of U12-type consensus sequences with U2-type sequences does indeed result in dramatically increased expression of the encoded fluorescent protein. FACS analysis of these cells indicated that the slower splicing of the U12-type introns did not result from saturation of the U12-type splicing machinery, since it was observed that the ratio of the two different-colored fluorescent signals remained roughly constant within a population of cells producing dramatically different amounts of total fluorescence signal. Finally, direct analysis of the RNA transcripts produced from the transfected constructs showed that while mature mRNA rapidly accumulates from the U2-type construct, mRNA production is much slower from the U12-type construct, and

instead, significant degradation as well as accumulation of partially spliced pre-mRNAs are apparent.

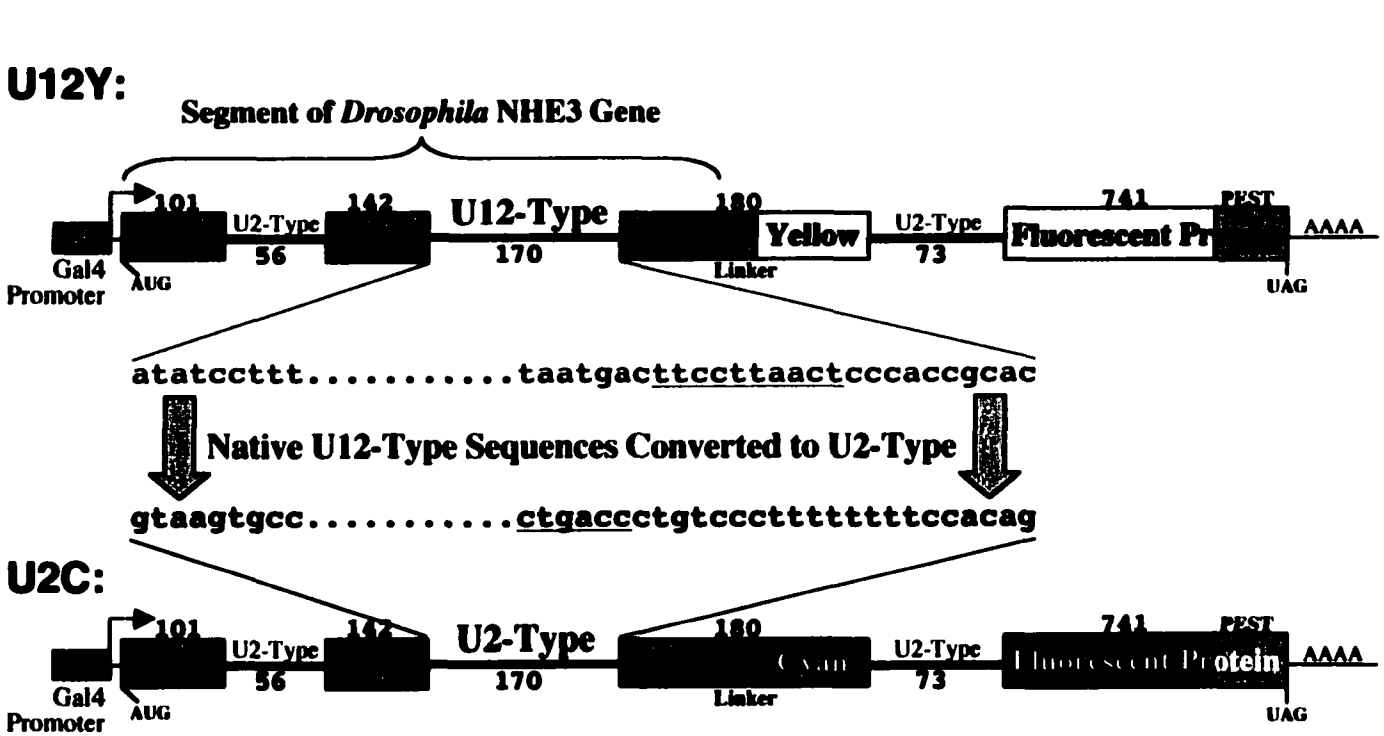
Results

Minigene constructs that report the relative efficiencies of U12-type versus U2-type splicing in vivo

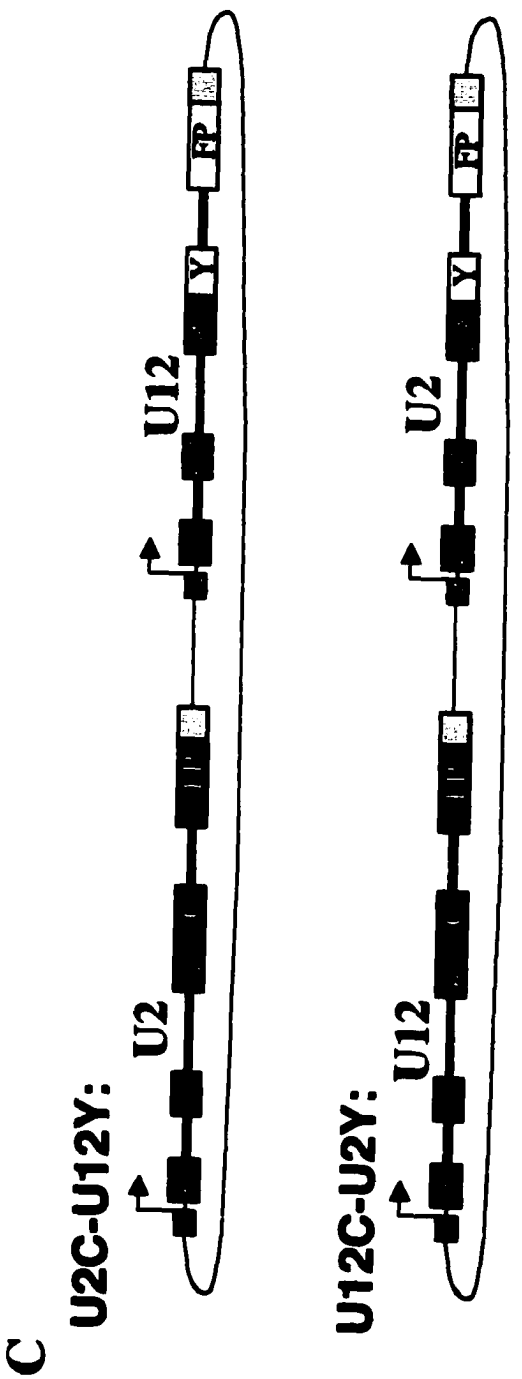
I next asked whether a U12- versus a U2-type intron in the same genetic context would lead to differential expression of the encoded protein. I designed plasmid constructs that would permit unambiguous correlation of reporter protein production with splicing efficiency. The constructs (Fig 9A) express either cyan or yellow fluorescent protein only when completely spliced, dependent on removal of a U12-type versus a U2-type intron. Since the spliced mRNAs are nearly identical in sequence (see Figure 9B), the relative intensity of the two fluorescent signals provides a readout of the relative efficiencies of the two splicing pathways within transfected tissue culture cells.
The fifth intron of the *Drosophila* sodium-hydrogen exchange channel gene, NHE3, was chosen for analysis because its adjacent exons and introns are small (see Fig 9A), facilitating the creation of small fusion constructs encoding fluorescent reporter proteins. Its sequences perfectly match the optimal U12-type consensus (C. Burge, personal communication). Since cooperativity between adjacent introns has been shown to influence the rate and fidelity of splicing (Robberson et al, 1990; Wu and Krainer, 1996), 496 nucleotides of the NHE3 gene, spanning the 3' portion of the fourth exon to the 5' portion of the sixth exon, were included in the construct. A start codon with Kozak consensus (Kozak, 1987) was added to the 5' end of the segment, and the segment was

Figure 9. Plasmids that Report the Efficiency of U12-type versus U2-type Splicing

(A) A segment of the *Drosophila melanogaster* NHE3 gene including introns 4 (U2-type) and 5 (U12-type) and their flanking exons was fused, via a linker, to a yellow (Y) or cyan (C) fluorescent protein (FP) coding sequence containing destabilizing "PEST" sequences to form U12Y or U12C. Intron 6 (U2-type) of NHE3 was inserted into the CFP or YFP coding sequences. A Gal4-responsive promoter and a Kozak consensus translation start site were directly upstream of the fusion protein. Exon and intron lengths are indicated above and below (not drawn to scale). To form U2Y (or U2C), consensus splicing sequences of the U12-type intron were mutated to U2-type sequences derived from an adenovirus intron, keeping the intron length constant. (B) The sequences of the protein-coding regions of the fusion constructs are shown, with exon sequences in uppercase and intron sequences in lowercase, and alternative sequences indicated. Red = U12-type consensus sequences; green = converted U2-type sequences; grey = NHE3 exon sequences; brown = linker sequences; orange = PEST sequences. The mutations converting cyan to yellow fluorescence are highlighted in cyan and yellow, respectively. The mutations introduced to change hydrophobic amino acids in the predicted transmembrane domain of exon 5 are boldfaced. (C) A plasmid containing the U2C and U12Y constructs in tandem (named U2C-U12Y) was created to ensure equal transfection of both constructs. The reciprocal plasmid (U12C-U2Y) with reversed color scheme controlled for inherent differences in CFP and YFP fluorescence yield.



B



fused, via a short linker, upstream of either yellow or cyan fluorescent protein cDNA (Miller et al., 1999). The fluorescent protein coding sequences were interrupted by insertion of the sixth intron from the NHE3 gene. Therefore, the U12-type intron in the construct was flanked by native exons and U2-type introns on both sides.

The consensus sequences within the U12-type intron were then mutated to U2-type consensus sequences derived from the first intron of the adenovirus major late transcription unit, an intron whose splicing has been well-characterized (Solnick, 1985). Changes were limited to mutations at the 5' splice site, branch site, 3' splice site, and the insertion of a 3' polypyrimidine tract (Figure 9A). The overall length of the intron was preserved, as were the internal, non-consensus sequences. Consequently, the mRNAs produced by splicing of the converted U2-type intron and the original construct containing the U12-type intron were identical, except for the nine codons that differ between YFP and CFP (Figure 9B).

Synthesis of pre-mRNA was driven by a Gal4-responsive promoter so that expression could be induced by producing Gal4 protein from either a copper-responsive metallothionein or a heat shock promoter on a separate, co-transfected plasmid. Versions of YFP and CFP containing protein destabilizing "PEST" sequences, which reduced the half-life of the fluorescent proteins from about 24 hours to 2 hours (Clontech product literature), were used so that the fluorescence signal would reflect differences in the rate of splicing rather than in protein accumulation. Also, the constructs were modified by mutating eight hydrophobic amino acids within the predicted transmembrane region (exon 5) of NHE3 (Figure 9B) to avoid possible membrane-targeting of the fusion protein. Finally, to guarantee that each cell was transfected with equal amounts of each

fluorescent protein gene, the two constructs were cloned in tandem into a single plasmid (Figure 9C). To normalize for inherent differences in the fluorescence intensities of YFP and CFP, a reciprocal construct was made with reversed color scheme (Figure 9C).

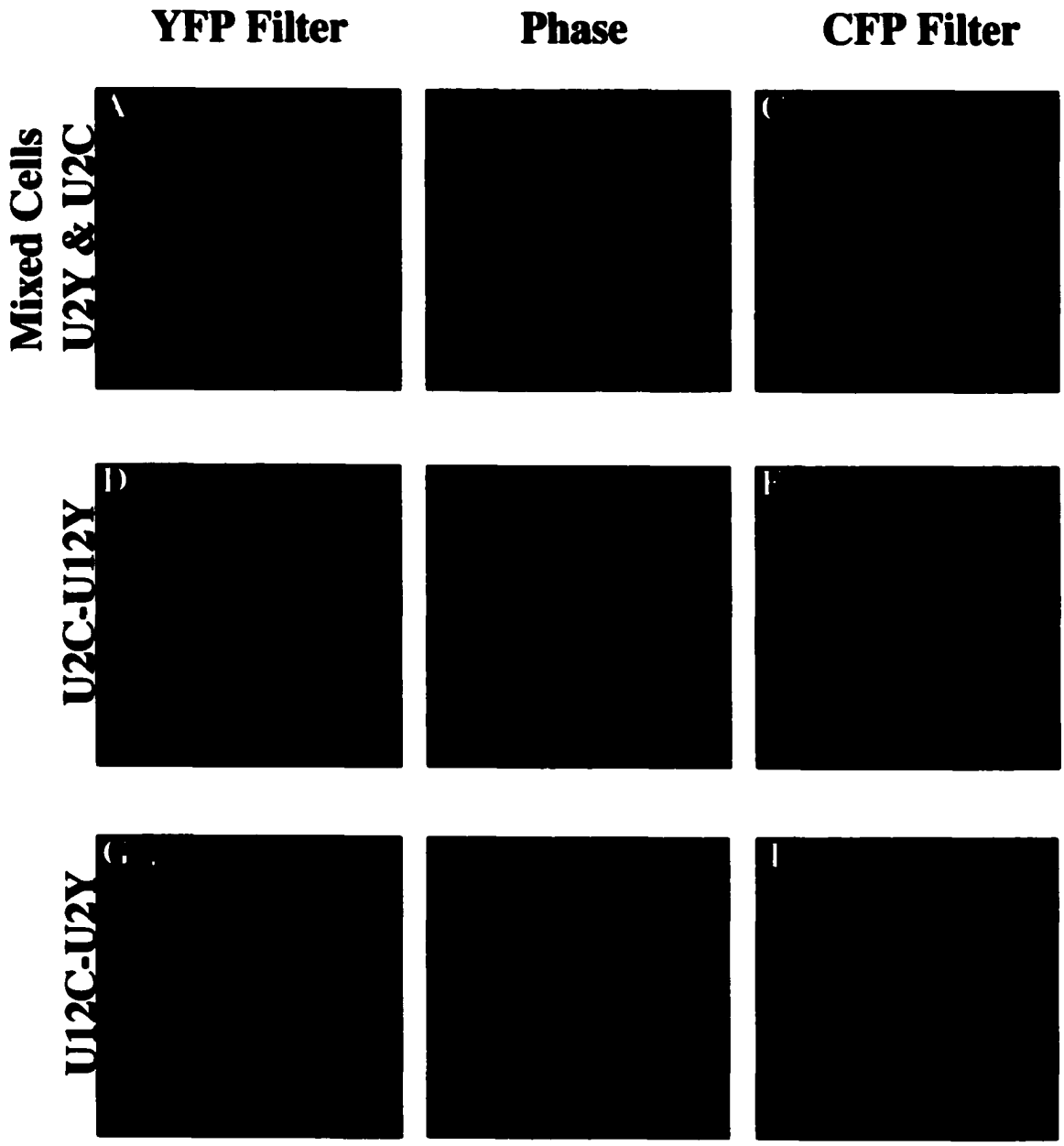
Replacement of natural U12-type intron sequences with U2-type consensus sequences increases protein expression in vivo

To confirm that the fluorescent signals from YFP and CFP were completely distinguishable (Miller et al., 1999), U2Y and U2C constructs (named on the basis of the middle intron and the fluorescent color) were co-transfected separately into *Drosophila* S2 cells with a metallothionein-Gal4 plasmid. Expression was induced 24 hours after transfection by adding cupric sulfate, and cells were viewed after 12 hours. As expected, fluorescence from any single cell was detected with either the cyan or the yellow filter set, but never with both (Figures 10A,B,C). Fluorescence from cells transfected with either the U2Y or U2C construct was, on average, quite robust, whereas fluorescence from cells transfected with the U12Y or U12C construct appeared much weaker (data not shown).

Next, the tandem U2C-U12Y plasmid (Figure 9C) was co-transfected with a metallothionein-Gal4 plasmid, and gene expression induced and assessed as above. While different cells within the transfected population exhibited different fluorescence intensities, any particular cell consistently showed at least a 5-fold higher fluorescent signal with the cyan filter than with the yellow filter (based on comparison of integrated pixel intensities from CCD-captured digital images) (Figures 10D,E,F). Transfection of

Figure 10. U2-dependent Splicing Produces Brighter Fluorescence than U12-dependent Splicing in *Drosophila* S2 Cells

(A-C) CFP and YFP fusion proteins are distinguishable. *Drosophila* S2 cells were separately co-transfected with either U2C or U2Y and metallothionein-Gal4, mixed and viewed 12 hours after cupric sulfate induction using a YFP filter (panel A), phase contrast (panel B), or a CFP filter (panel C). (D-I) S2 cells co-transfected with U2C-U12Y (panels D-F) or U12C-U2Y (panels G-I) and metallothionein-Gal4 were similarly induced and viewed using a YFP filter (panels D and G), phase contrast (panels E and H), or a CFP filter (panels F and I).



the reciprocal tandem U12C-U2Y plasmid (Figure 9C) yielded the opposite pattern: each transfected cell exhibited at least 5-fold more yellow fluorescence than cyan fluorescence (Figures 10G,H,I). Because all other differences between the constructs are controlled by the use of reciprocal plasmids, the altered expression levels can be ascribed to the identity of the middle intron.

To confirm that the fifth intron of the NHE3 gene is in fact spliced by the U12-type splicing machinery, excision of this intron from endogenous pre-mRNA in wild-type *Drosophila* larvae was compared to that in larvae homozygous for a P-element disruption of U6atac (characterized by Otake et al., 2002). The U6atac-disrupted larvae die at the third instar because of a deficiency in U12-type splicing. Analysis of total RNA by RT-PCR revealed that the U6atac-deficient larvae (at ~60 hours post egg-laying) accumulated a substantial amount of unspliced intron 5, whereas the control wild-type larvae accumulated very little (Figure 11), demonstrating excision of this intron by the U12-type spliceosome.

The 6- to 8-fold difference in fluorescent protein expression is not induced by saturation of the U12-type splicing machinery

The results from fluorescence microscopy, showing a consistent and significant difference in the levels of YFP and CFP from the reciprocal constructs containing U12-type and U2-type introns (Figures 10D through I), were extended by fluorescence-activated cell sorting (FACS). I first verified that cyan and yellow fluorescence signals can be detected independently by FACS. Compared to mock-transfected cells, which

Figure 11. U6atac-Disrupted *Drosophila* Larvae Confirm that the Fifth Intron of NHE3 is U12-type

Total RNA prepared from *Drosophila* larvae (~60 hours post egg-laying) that were either heterozygous or homozygous for a P-element disruption of U6atac was analyzed by RT-PCR for the presence of endogenous unspliced NHE3 intron 5. Homozygosity was assessed by the lack of a green fluorescent protein-producing balancer chromosome in the larvae. RT and PCR primers were complementary to exon sequences surrounding the 5th intron, so that excision of the intron produced a shorter amplicon. Two sets of sibling larvae were analyzed to establish reproducibility, and each set was amplified to either 26 or 28 PCR cycles to increase the range of detection. (This experiment was performed in collaboration with Matthew McCarthy, Yale College '02.)

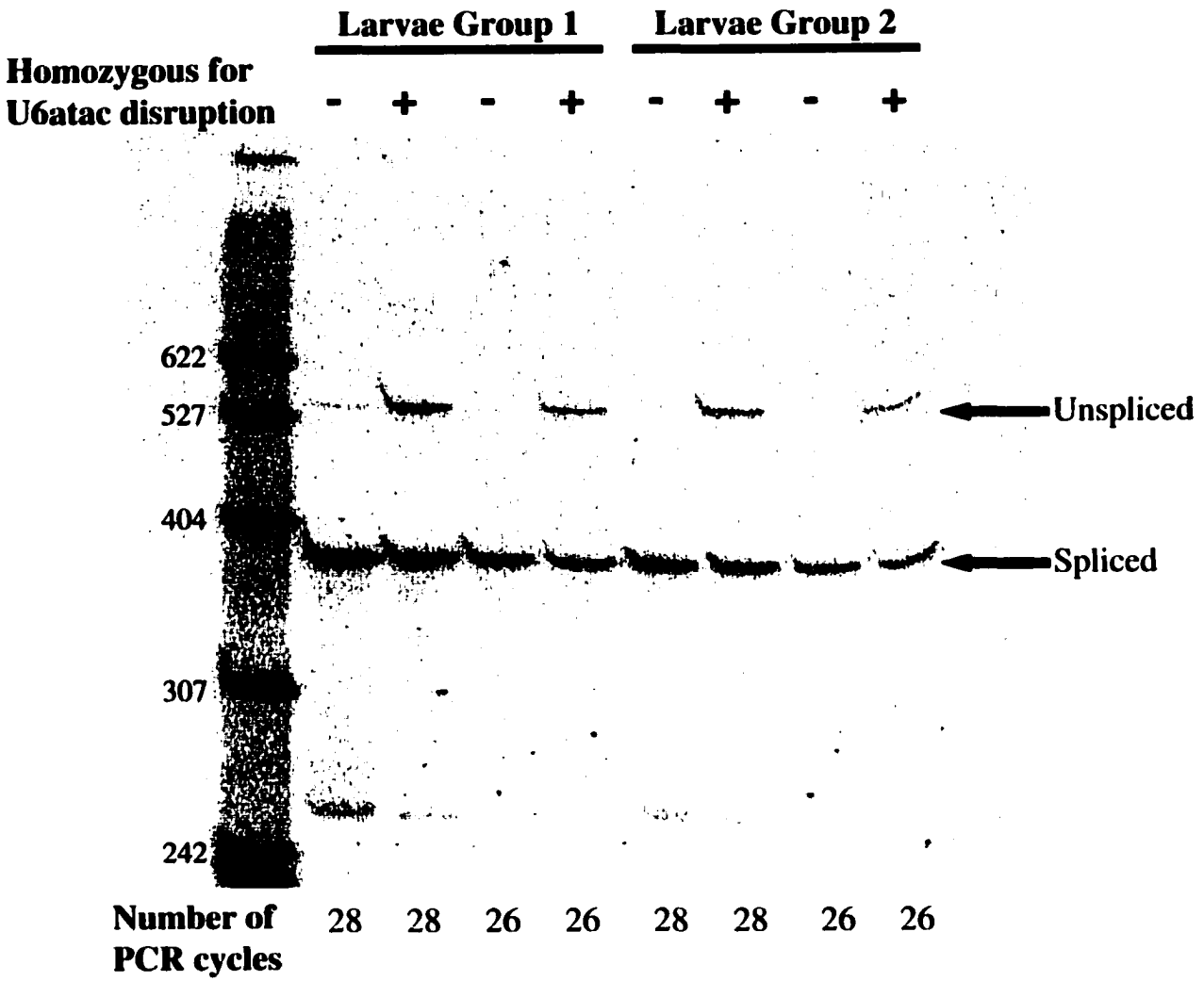


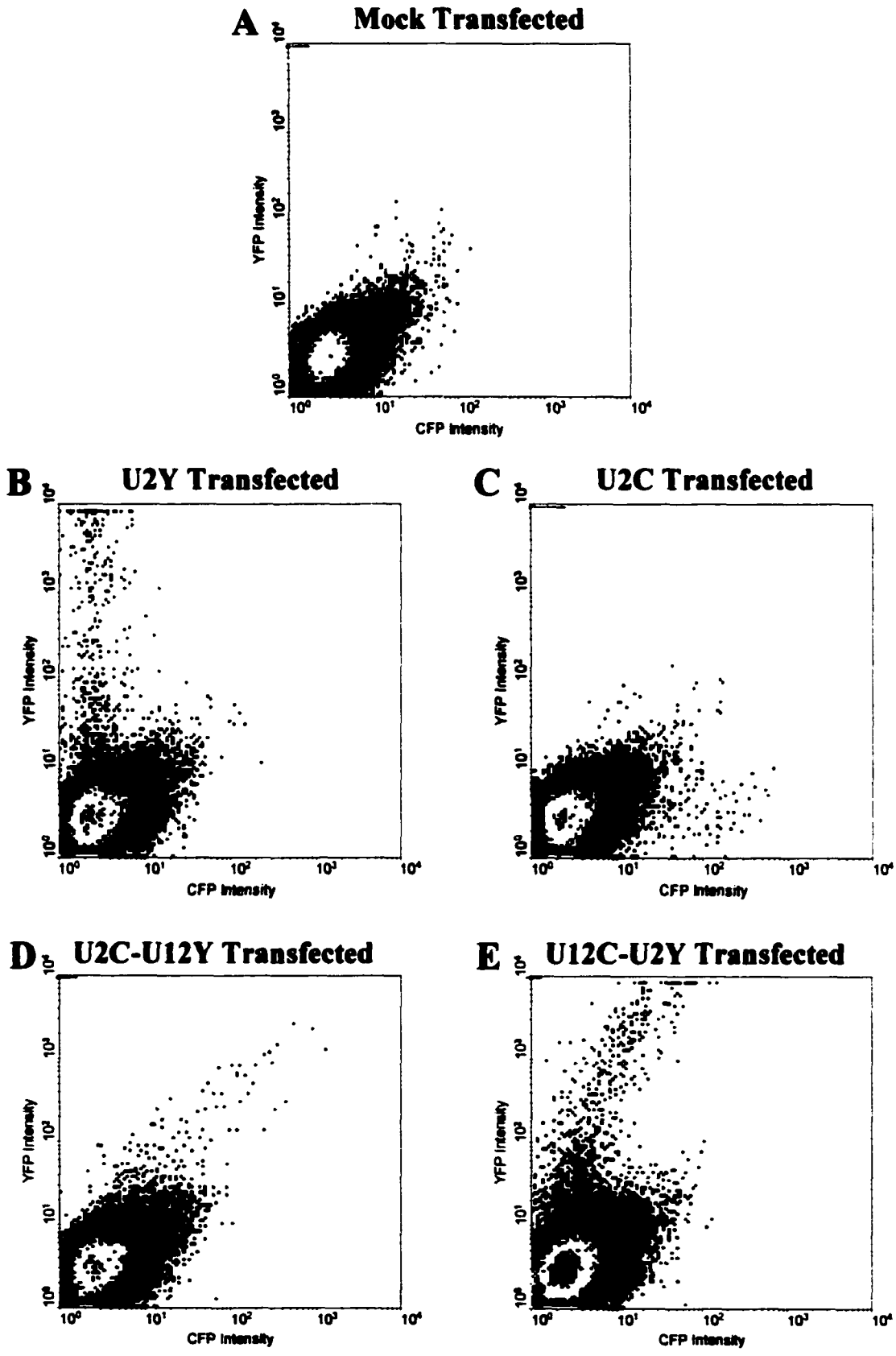
exhibit minimal intensity on each channel (Figure 12A), S2 cells transfected with U2C or U2Y (and metallothionein-Gal4, induced as above) showed significant fluorescence on the cyan or yellow channel, respectively, and negligible fluorescence on the other channel (Figure 12B, C). The large range of fluorescence intensities seen (note the logarithmically scaled axes) probably resulted from cells being transfected with different numbers of plasmids or in different physiological states.

Cells transfected with the tandem U2C-U12Y or U12C-U2Y plasmid displayed the same patterns of fluorescence as previously observed by microscopy: the intensity of the protein expressed from the U2-type intron containing gene was always several-fold greater than from that containing the U12-type intron (Figure 12D, E). This pattern manifests itself on the FACS plot as a roughly linear set of points through which a best-fit line can be drawn. As expected, the slope of the line for the U12C-U2Y transfected cells is greater than that for the U2C-U12Y transfected cells, because of increased relative expression of the protein from the mRNA generated by U2-dependent splicing only. The relative difference in expression can be calculated from the slopes of the two lines, taking into account the decreased CFP detection sensitivity and assuming that the nine amino acids that change the color of the fluorescent protein do not affect protein expression or stability. I conclude that the expression of fluorescent protein increases 6- to 8-fold when the U12-type intron is converted to a U2-type intron.

The linearity of the data points in Figures 12D and 12E also excludes the possibility that the observed differences result from saturation of the U12-type splicing machinery by high levels of pre-mRNA expressed from the transfected constructs. Because the scale of the axes are logarithmic, the fluorescence emitted by individual cells

Figure 12. FACS Analysis of Transfected S2 Cells

Density plots of yellow versus cyan fluorescence intensity for S2 cells co-transfected with metallothionein-Gal4 and various plasmids, obtained 12 hours after cupric sulfate induction. The flow cytometer detection optics are inherently less sensitive for CFP than for YFP. 50,000 cells were observed on each plot. (A) Mock transfected S2 cells show minimal cyan or yellow fluorescence. (B)U2Y-transfected cells show only yellow fluorescence. (C) U2C-transfected cells show only cyan fluorescence. (D) U2C-U12Y-transfected cells show greater cyan than yellow fluorescence (correcting for low CFP detection sensitivity). (E) U12C-U2Y-transfected cells show greater yellow than cyan fluorescence.



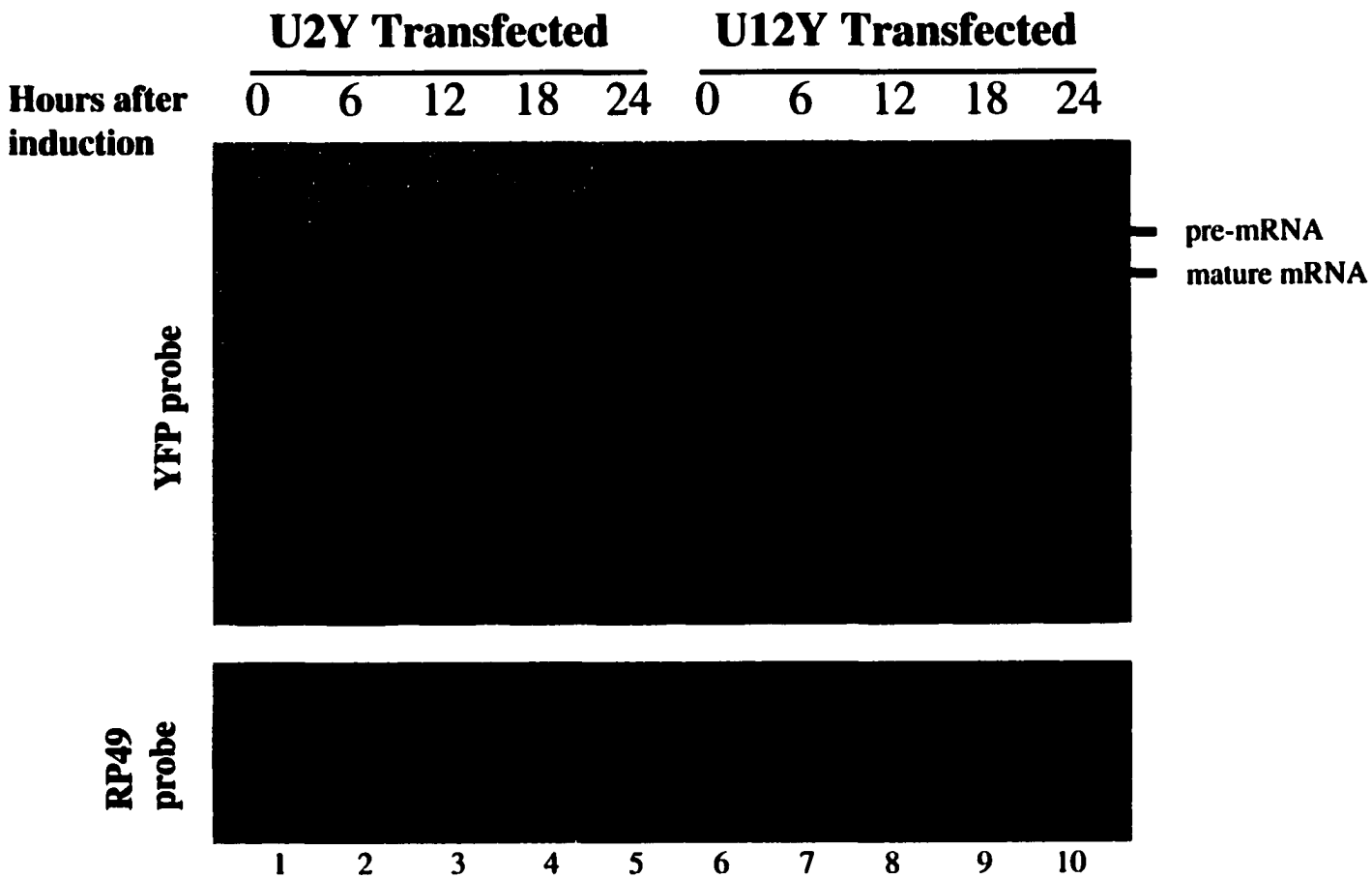
in the population differed by several orders of magnitude. Over this entire range, the proportion of yellow to cyan fluorescence was roughly constant. For example, a strongly fluorescent cell making 50-fold more U2-type reporter protein also made about 50-fold more U12-type reporter protein compared to a weakly fluorescent cell. This would not be the case if the splicing capacity of U12-type spliceosomes were saturated by expression of high levels of pre-mRNA and fluorescent proteins. Likewise, in a collaborative experiment with Matthew McCarthy (Yale College, '02) I found that when different amounts of U2C-U12Y or U12C-U2Y were individually transfected into S2 cells (over a 20-fold range), the ratio of cyan to yellow fluorescence in each cell remained approximately constant (within detection limits) despite great differences in the absolute fluorescence intensity (data not shown).

Presence of a U12-type intron hinders production of mature mRNA

To confirm that the observed differences in fluorescent protein expression were caused by differences in the efficiency of U2-type versus U12-type splicing, the effects of intron conversion were examined directly at the RNA level. S2 cells were co-transfected with metallothionein-Gal4 and either U2Y or U12Y, expression induced after 24 hours, and total RNA at various times after induction analyzed by Northern blot. Cells transfected with the U2Y construct showed rapid accumulation of mature mRNA, whereas cells transfected with the U12Y construct accumulated mature mRNA more slowly (Figure 13). Also, a more slowly migrating species was observed in the U12Y-transfected cells. The identity of this species as partially spliced pre-mRNA was

Figure 13. mRNA Production Is Slowed by Inefficient Splicing of a U12-type Intron

Northern blot of total RNA isolated from S2 cells co-transfected with U2Y (lanes 1-5) or U12Y (lanes 6-10) and metallothionein-Gal4 at the indicated times after induction of expression. The blot was successively hybridized with random-primed DNA probes complementary to the YFP coding sequence (top panel) and to a loading control, ribosomal protein 49 (bottom panel). The identity of the slowly migrating band as partially-spliced pre-mRNA, was confirmed by hybridization to a probe specific for the middle intron (data not shown).

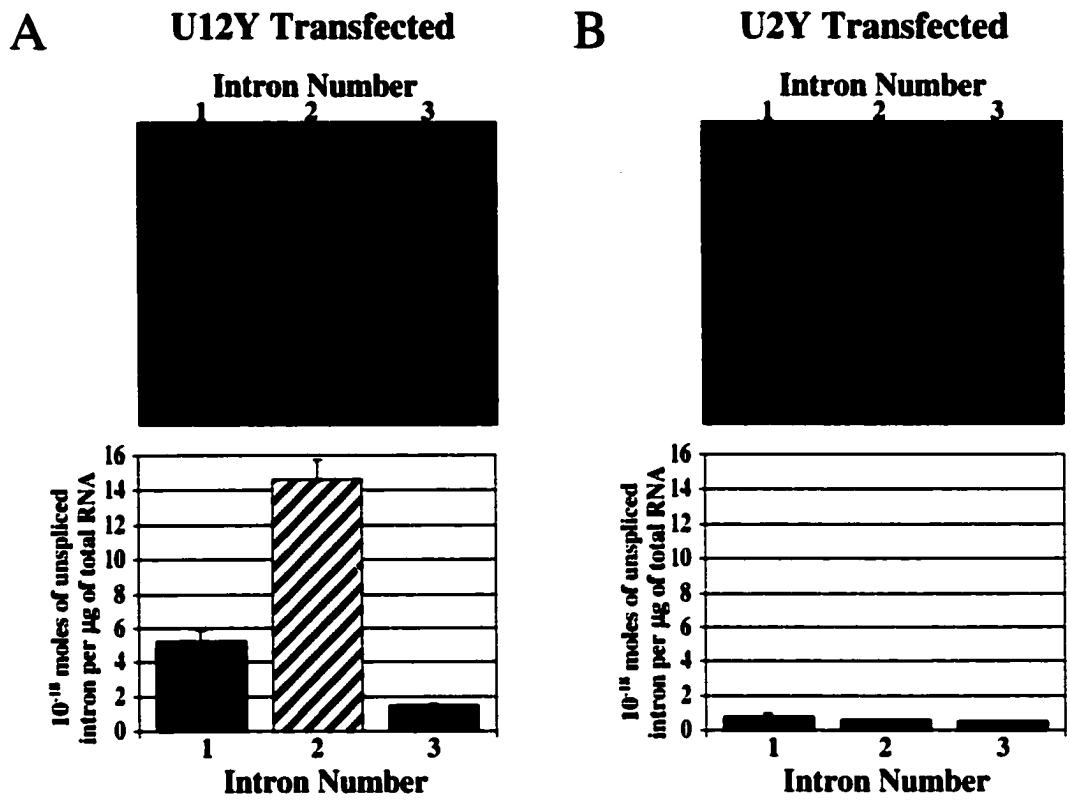


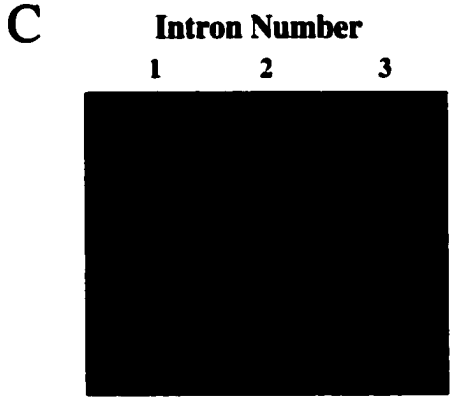
confirmed using a probe specific for the middle intron (data not shown). Lanes corresponding to the U12Y-transfected cells also reproducibly showed a smear below the full-length bands (Figure 13), suggesting that the pre-mRNA is prone to degradation if not efficiently spliced.

Finally, to examine differences in the splicing rate of the middle intron relative to its neighboring introns for both the U2Y and U12Y constructs, Matt McCarthy and I carried out quantitative RT-PCR experiments, as described for the endogenous human pre-mRNAs (Figure 5). S2 cellular RNA prepared 12 hours after induction of expression was RT-PCR amplified in the presence of a single *in vitro* transcribed RNA standard containing three separate 10-nucleotide deletions, allowing quantitation of all three introns. The levels of unspliced introns from the U12Y construct indicate that the unspliced U12-type middle intron was more abundant than its U2-type neighbors, confirming that it is spliced more slowly (Figure 14A). Conversion of the consensus sequences to U2-type dramatically increased the splicing of that intron as evidenced by the lower relative abundance of its unspliced form (Figure 14B). Interestingly, conversion to U2-type sequences also appeared to increase the splicing rate of the two adjacent introns, suggesting cooperativity in the excision of these introns. As before, to verify that the amplification efficiencies of the full-length and truncated RNAs were equal, equimolar amounts of full-length and truncated *in vitro* generated transcripts were co-amplified and observed to produce amplicons of equal intensity (Figure 14C).

Figure 14. Splicing Rate of the Middle Intron Increases upon Mutation of U12-type to U2-type Consensus Sequences

(A) Quantitative RT-PCR analysis of unspliced introns (similar to Figure 7) from total RNA prepared 12 hours after induction of U12Y-transfected cells and amplified in the presence of a single *in vitro* transcribed RNA standard containing three separate 10-nucleotide deletions, allowing quantitation of all three introns. $1.33x10^{-17}$ moles of control RNA standard was added per μg of total cellular RNA. Hatched bar = U12-type intron; solid bars = U2 type introns; error bars = standard deviation of two experiments. (B) Similar analysis of U2Y-transfected cells. Note that 10-fold less control RNA standard ($1.33x10^{-18}$ moles per μg of total RNA) was used. (C) Confirmation of equal amplification efficiency of *in vitro* transcribed full-length and 10-nucleotide truncated RNAs (amplicon ratio of 1.02±0.02) (as in Figure 2C). 10^{-19} moles of each RNA species were RT-PCR amplified. (These experiments were performed in collaboration with Matthew McCarthy, Yale College '02.)





Discussion

In this chapter of the thesis, I demonstrate that mutation of native U12-type intron consensus sequences to U2-type sequences within a *Drosophila melanogaster* minigene construct results in significantly increased levels of mRNA and protein expression. This result complements the previous chapter's observation that U12-type introns are spliced more slowly than U2-type introns from endogenous transcripts *in vivo*. The present result demonstrates that the slower splicing of U12-type introns can have a direct effect on protein expression.

My *Drosophila melanogaster* splicing constructs produced mRNAs differing by only nine codons for CFP versus YFP. It is possible that these changes affect either the efficiency of splicing, polyadenylation, export, or translation. Therefore, the reciprocal constructs with reversed color schemes were important to control for differences not only in fluorescence yield, but also expression. Use of these constructs confirmed that the differences in fluorescent protein expression are attributable to the changed intron type.

Because gene expression from the transfected constructs was driven by Gal4, I was able to induce expression at a well-defined time, and then follow accumulation of fluorescence over time. In fact, a microscope-mounted perfusion chamber was made to constantly replenish fresh growth media to *Drosophila* S2 cells expressing fluorescent proteins. This permitted me to observe and record the accumulation of fluorescent proteins within S2 cells over a period of 12 to 15 hours (after which the cells began to die). I expected that if there were any changes in U12-type versus U2-type splicing efficiency during the cell cycle (which is about 30 hours for S2 cells [Lengyel et al., 1975]), I might observe this as a change in the intensity ratio of the two fluorescent

colors. However, I did not detect any significant changes in the fluorescence ratio during the time the S2 cells were observed (data not shown). Although I was only able to observe viable cells for approximately half of the duration of a division cycle, I should have been able to follow cells at all different stages of the cell cycle within the observed population. Nevertheless, I cannot exclude the possibility that the cells stopped cycling under the non-ideal conditions of the perfusion chamber, causing me to miss changes in fluorescence that might otherwise have been observed under ideal growth conditions.

The cooperativity of splicing that was observed in Figure 14 raises some interesting questions for further investigation. I noticed that conversion of the middle intron from U12-type to U2-type appeared to increase the splicing rate of not only that intron, but also of the two neighboring introns. Would the same cooperativity be observed for non-adjacent introns, if the minigene construct had multiple introns? Does the presence of one unspliced intron within a gene have a global effect on the splicing of all introns within that gene? Alternatively, is the complete assembly of all early splicing components required for every intron within a transcript before catalysis can occur? How might deletion of the middle intron from the minigene construct affect the splicing rates of the neighboring introns, and the overall rate of protein production?

The increased degradation observed for the U12-type intron-containing construct (Figure 13) suggests that the transcripts are more prone to degradation if they are not properly spliced. This is consistent with previous findings (Bousquet-Antonelli et al., 2000). It would be informative to investigate the mechanism that mediates this degradation. Are the unspliced RNAs degraded by general nucleases within the cell or are they degraded via nonsense-mediated decay (the middle intron would have introduced

a premature in-frame stop codon if it were not excised)? Are the unspliced RNAs retained in the nucleus or are they still exported? Does the degradation take place within the nucleus or the cytoplasm or both?

My observations also raise the question of what is causing the U12-type introns to be spliced more slowly. *In vitro* splicing of U12-type introns has consistently been reported to be extremely inefficient by different groups using different splicing substrates (Wu and Krainer, 1996; Tarn and Steitz, 1996b). This could be attributed to the approximately 100-fold lower abundance of the minor-class snRNPs relative to the major-class snRNPs in mammalian cell nuclei and nuclear extracts (Montzka and Steitz, 1988; Tarn and Steitz, 1996b). U12-dependent splicing in these extracts might be stalled by the slow assembly of low-concentration splicing components. However, since splicing components are thought to be assembled co-transcriptionally *in vivo* (Misteli and Spector, 1999; reviewed in Hirose and Manley, 2000), the effects of snRNP concentrations on *in vivo* splicing rates are less predictable. It is therefore unclear whether the slower splicing of U12-type introns *in vivo* is caused by slower spliceosomal assembly or by inherently slower catalysis by the minor-class spliceosome.

Further insight might be gained by investigating whether minor-class splicing components are deposited onto nascent pre-mRNAs by RNA Polymerase II in a manner analogous to major-class components, or whether they assemble onto pre-mRNAs independently of transcription. Given the extreme rarity of U12-type introns, one might expect co-transcriptional deposition of minor-class splicing components to be less efficient than deposition of major-class components. Nevertheless, if minor-class components are required to be deposited co-transcriptionally, then might their low

abundance cause RNA polymerase II to stall at U12-type intron sequences until they are deposited?

The existence of a specific group of genes whose expression might be regulated by the presence of inefficiently spliced introns hints at a functional commonality among these genes. Burge et al. (1998) note that a large proportion of U12-type intron containing genes fall into the category of information processing, including genes involved in DNA replication/repair, transcription, RNA processing, and translation, while few fall into the category of genes performing operational functions such as energy metabolism or biosynthesis. Do these genes share similarities in their patterns of expression which might shed light on the specific regulatory role of U12-type introns? Since expression of these genes is probably upregulated in proliferating cells compared to non-proliferating cells, perhaps the activity of the U12-type spliceosome increases during proliferation. Perhaps U12-type splicing activity is modulated in a tissue-specific or a developmental-stage specific manner. I further speculate that U12-type introns may be acting as post-transcriptional bottlenecks, preventing overexpression of genes whose unchecked expression might be harmful to the organism. The fluorescent protein-coding constructs described here provide a means of investigating these questions in the developmentally well-studied organism, *Drosophila melanogaster.* Attempts to create transgenic flies in which these constructs were stably integrated into the organism's genome are described in Appendix A. Thus far, however, I have been unsuccessful at producing detectable fluorescence in these animals.

Chapter IV

Materials and Methods

Preparation of RNA Standards

Segments of genomic sequences from human E2F2, SmE, and INSIG1 were PCR-amplified from genomic clones RPCI1-15005 (Sanger Centre), RPCI11-397P13 (Sanger Centre), and HP-5 (Peng et al., 1997), respectively. The segments (between 301 and 853 nucleotides) included most of the upstream exon and >195nt of intronic sequence surrounding each 5' splice site. To create an RNA standard for the *Drosophila* constructs, a single 872-nucleotide segment including all three introns was PCR-amplified from the U12Y plasmid (construction of U12Y described below) using primers MATTF and MATTR (MATTF: 5'-GCG ACT CGA GAT GGT GGC CAT ATT CAA CGA TCT TA-3'; MATTR: 5'-GCG ACT GTG CGG CCG CGA AGA AGT CGT GCT GCT TCA TGT G-3'). XhoI and NotI restriction sites, introduced by the PCR primers, were used to insert the amplified segments into pBluescript SK+ vectors (Stratagene) downstream of a T7 promoter. 10-nucleotide deletions were introduced by either PCR mutagenesis or using the QuickChange Mutagenesis system (Stratagene). Oligonucleotides for PCR cloning and mutagenesis are listed in Tables 1, 2, and 3. RNA standards were prepared by *in vitro* transcription from the T7 promoter essentially as described by Tarn and Steitz (1996b), internally trace-labeling with α-^{32}P-UTP. Gel-purified transcripts were quantified by liquid scintillation counting and stored at -70°C in aqueous solution containing 20µg/mL yeast total carrier RNA.

Table 1. PCR Primers used to amplify segments of three human genes for insertion into XhoI and NotI sites of pBluescript SK+ (Stratagene)

Gene	Intron Number	Primer Direction	Primer Sequence (5'→3') (Bold and Underlined = XhoI or NotI site) (Bold only = added terminal base pairs)
SmE	1	Forward	**GCGACTCGAG**GTGACCTTCACACTTCCGCTTCC
SmE	1	Reverse	**GCGACTGTGCGGCCGC**GCATAGACTCTGCTTCCTGAAACAAC
SmE	2	Forward	**GCGACTCGAG**TGGTGCAGAACCTCATCTTCAGATAC
SmE	2	Reverse	**GCGACTGTGCGGCCGC**GGGTTTTTCTTCCAATGTTTCATCC
SmE	3	Forward	**GCGACTCGAG**TTATTTCAGAGATCGCGGATTCAGG
SmE	3	Reverse	**GCGACTGTGCGGCCGC**GATCCTAGCTACTTGGGAGGCTGAG
SmE	4	Forward	**GCGACTCGAG**TGAACCTTGTATTAGATGATGCAGAAGAG
SmE	4	Reverse	**GCGACTGTGCGGCCGC**TCTACTTGGGTTTTTCCTGCCACTC
E2F2	1	Forward	**GCGACTCGAG**GCTCTGCCCAGCTACTGCTACCTAC
E2F2	1	Reverse	**GCGACTGTGCGGCCGC**ACTACACTTTGGGCCCTCTCTGG
E2F2	2	Forward	**GCGACTCGAG**CTCTCTCTCCCCCTTTTCTCCAAG
E2F2	2	Reverse	**GCGACTGTGCGGCCGC**GCTGAGGTGAGAGAGTTGCTTGAAC
E2F2	3	Forward	**GCGACTCGAG**GTGACTGGGTCCCTTTCTTCTTCAG
E2F2	3	Reverse	**GCGACTGTGCGGCCGC**TTTATGCCAAGGGCTATAGGTCAGC
E2F2	4	Forward	**GCGACTCGAG**GAGTTTCACCCACAGACTGACACTG
E2F2	4	Reverse	**GCGACTGTGCGGCCGC**GTGGCAGCTCATCATTCTTTTTCTC
E2F2	5	Forward	**GCGACTCGAG**CACTTTTCTTGACTCTCCTGCCCTAC
E2F2	5	Reverse	**GCGACTGTGCGGCCGC**GAGAATCGCTTCAACCCAGGAG
E2F2	6	Forward	**GCGACTCGAG**AATGACCTGGCTTCTCTTCCTCCTC
E2F2	6	Reverse	**GCGACTGTGCGGCCGC**CTTCCCGAGTAGCTGGACTACAGG
INSIG1	1	Forward	**GCGACTCGAG**AGATTGCACGACCACTTCTGGAG
INSIG1	1	Reverse	**GCGACTGTGCGGCCGC**GTGACGGGTTTGTTTTCAGTTATCC
INSIG1	2	Forward	**GCGACTCGAG**AAATTCTGGAACTTGGAATCTGCTG
INSIG1	2	Reverse	**GCGACTGTGCGGCCGC**AGGATTTGTTGGATGCAGGTTACAG
INSIG1	3	Forward	**GCGACTCGAG**CTTTTATATCACCAGAAATTGGATTTTGC
INSIG1	3	Reverse	**GCGACTGTGCGGCCGC**CAGTCAATTTCACCTGGCTCAATTC
INSIG1	4	Forward	**GCGACTCGAG**CGTCCCCAGATTTCCTCTATATTCG
INSIG1	4	Reverse	**GCGACTGTGCGGCCGC**CATGTGCCTGTAATCCCAGCTACTC

Table 2. PCR primers used to create 10nt truncations within control RNA plasmid constructs of the three human genes

Gene	Intron Number	Primer Direction	Primer Sequence (5'→3')
SmE	1	Forward	TATGGTGCAGCCTACGCAGGATGTCAGGACTAGGAG
SmE	1	Reverse	TCCTGCGTAGGCTGCACCATAACCTTCTGCACT
SmE	2	Forward	CTTCAGATACTTACAAAATGTCTTGTTTCGTAACTACTTTTTA AATAAGAGG
SmE	2	Reverse	GTAGTTACGAAACAAGACATTTTGTAAGTATCTGAAGATGAGG TT
SmE	3	Forward	GATAGAAGGCTGTATATCCAGGCGATTTCATCTCATAGC
SmE	3	Reverse	GAAATCGCCTGGATATACAGCCTTCTATCCGCATATTCACTTG
SmE	4	Forward	CAAAGTCAAGAAAACAGATAGAAGTGGTCTTACAGAATTCTAG AAATATTTT
SmE	4	Reverse	GTAAGACCACTTCTATCTGTTTTCTTGACTTTGTTTTAGAATG AATCTCT
E2F2	1	Forward	CTGCCGGCAGGCCGGCGCTGGGGGCCCCCACCGCTT
E2F2	1	Reverse	CAGCGCCGGCCTGCCGGCAGGCAT
E2F2	2	Forward	CTCCCCAGCCCCAGTGCTGATGGGGCCTCTG
E2F2	2	Reverse	ATCAGCACTGGGGCTGGGGAGGCCATCCAC
E2F2	3	Forward	GAACAACATCCAGTGCCTGGGCGGCCAGTGGTA
E2F2	3	Reverse	CCCAGGCACTGGATGTTGTTCTTGGCCTTCTT
E2F2	4	Forward	ACAAGGCCAACCTCCTTGGTTGGGGGAAGGT
E2F2	4	Reverse	CAACCAAGGAGGTTGGCCTTGTCCTCAGTCAGG
E2F2	5	Forward	CTGGAAGTGCCCGACAGAGAGGGGAAATATTTGGTGGAGAG
E2F2	5	Reverse	CCTCTCTGTCGGGCACTTCCAGTCTCG
E2F2	6	Forward	CCACAGCATCTCCCTTGTCCAGAGGGACAGGA
E2F2	6	Reverse	CTGGACAAGGGAGATGCTGTGGGCTCCATGATG
INSIG1	1	Forward	GCGGGACAGCTACCCCTCCTGGTTCTTCTGGAAG
INSIG1	1	Reverse	AGGAGGGGTAGCTGTCCCGCAGCAGGGAGGGA
INSIG1	2	Forward	CATTAACCACGCCCCTTAATTTTCTGTGCTACGTCCAGAG
INSIG1	2	Reverse	CAGAAAATTAAGGGGCGTGGTTAATGCCACCAAA
INSIG1	3	Forward	CGTGTATAATGGTGTCCAGTAAGTGTGTGTTTTCAAATATTGG C
INSIG1	3	Reverse	CACACACTTACTGGACACCATTATACACGAGAAACTGCG
INSIG1	4	Forward	CATAGGACGACAGTTAGTGAAATGATCATATTATCTTCTAAAA CTTG
INSIG1	4	Reverse	GATAATATGATCATTTCACTAACTGTCGTCCTATGTTCCCCAC

Table 3. Oligonucleotides used to make 10nt internal deletions in the *Drosophila* constructs using the QuickChange mutagenesis system (Stratagene):

Intron Number	Primer Direction	Oligonucleotide Sequence (5'→3')
1	Forward	ACGATGCCGTGGCCATAGTCCTTTTTGCTACTTTTTTTTTCGTTGGGAA
1	Reverse	TTCCCAACGAAAAAAAAAGTAGCAAAAAGGACTATGGCCACGGCATCGT
2	Forward	CGCGGACATGGGATGCTTGACAGCTTTTCAAAGATTAATGATAAAT
2	Reverse	ATTTATCATTAATCTTTGAAAAGCTGTCAAGCATCCCATGTCCGCG
3	Forward	CAAGTTCAGCGTGTCCGGCGAGGTTATTTTCAGCATAGAAATACTT
3	Reverse	AAGTATTTCTATGCTGAAAATAACCTCGCCGGACACGCTGAACTTG

Quantitative RT-PCR Analysis

RNA for analysis of E2F2 and SmE was prepared, using Trizol (Gibco-BRL) according to the manufacturer's directions, from log-phase HeLa cells grown in suspension in RPMI (Gibco-BRL) containing 10% fetal bovine serum (FBS, Gibco-BRL). RNA for analysis of INSIG1 was similarly prepared from ~70% confluent SK Hep cells grown in 35mm dishes in DMEM (Gibco-BRL) with 10% FBS. RNA for analysis of U12Y and U2Y was isolated from transiently transfected *Drosophila* S2 cells 12 hours after induction of expression as described below. Contaminating genomic DNA was removed by treating with RQ1 DNase (Promega) in 5mM $MgCl_2$ and 50mM Tris pH 8.0 at 37°C for 1 hour, followed by heat-inactivation at 65°C for 10 minutes. Reverse transcription was then performed using Thermoscript RT (Invitrogen) according to manufacturer's directions, with 0.5μM gene-specific primers (Tables 4 and 5), 7.5ng/μL (for SmE, E2F2, and INSIG1) or 0.188ng/μL (for U12Y and U2Y) total cellular RNA, and various concentrations (between $3.36x10^{-20}$ and $7.60x10^{-22}$ moles/μL) of *in vitro* transcribed RNA standards. RT reactions were incubated at 37°C for 15 minutes, then 55°C for 30 minutes, then 65°C for 30 minutes, and finally 85°C for 10 minutes. 0.1units/μL of *E.coli* RNase H (Invitrogen) were added and the samples incubated at 37°C for 30 minutes. PCR was then performed using Platinum Taq polymerase (Invitrogen) in the buffer supplied with 1.5 mM $MgCl_2$, 0.4mM dNTPs, 250nM unlabelled primers (Tables 4 and 5) doped with 5'-radiolabeled forward primer (to label only one strand of amplicon), and 10% (by volume) of reverse-transcribed cDNA. Between 24 and 28 cycles of PCR were

Table 4. Primers used for quantitative RT-PCR of the three human genes:

Gene	Intron Number	Used for RT or PCR	Primer Direction	Primer Sequence (5'→3')
SmE	1	PCR	Forward	GGGTGTGCTCTTTGTGAAATTCC
SmE	1	PCR	Reverse	CTCCTAGTCCTGACATCCTGCGTAG
SmE	1	RT	Reverse	CTCCTAGTCCTGACAT
SmE	2	PCR	Forward	AACCTCATCTTCAGATACTTACAAAATGT
SmE	2	PCR	Reverse	TTACCCTCCAATCTCCAGTTACATC
SmE	2	RT	Reverse	TTACCCTCCAATCTCC
SmE	3	PCR	Forward	CGGATTCAGGTGTGGCTCTATG
SmE	3	PCR	Reverse	TTAAGTCTTTTTCTACGACCGAACTGC
SmE	3	RT	Reverse	TTAAGTCTTTTTCTAC
SmE	4	PCR	Forward	TGAACCTTGTATTAGATGATGCAGAAGAG
SmE	4	PCR	Reverse	GCAGGTCACTGAATTCTGCAAGTG
SmE	4	RT	Reverse	GCAGGTCACTGAATTC
E2F2	1	PCR	Forward	ACACACCGCTGTACCCGCAGAC
E2F2	1	PCR	Reverse	ACCTGGAAAAGCATAGGGGGAAG
E2F2	1	RT	Reverse	ACCTGGAAAAGCATAG
E2F2	2	PCR	Forward	AAGGGGAAGTGCATCAGAGTGG
E2F2	2	PCR	Reverse	TGGAAAGTGATCCAAATATGACTACCAG
E2F2	2	RT	Reverse	TGGAAAGTGATCCAAA
E2F2	3	PCR	Forward	GACATCACCAACGTGCTGGAAG
E2F2	3	PCR	Reverse	CCCACAGCACAAGTACCCATTG
E2F2	3	RT	Reverse	CCCACAGCACAAGTAC
E2F2	4	PCR	Forward	GGAGCTGAAGGAGCTGATGAACAC
E2F2	4	PCR	Reverse	GTTCCTGCTCCCTGACACCTTC
E2F2	4	RT	Reverse	GTTCCTGCTCCCTGAC
E2F2	5	PCR	Forward	ACTTACCAGGATATCCGTGCTGTTG
E2F2	5	PCR	Reverse	CTCTCCACCAAATATTTCCCCTCTC
E2F2	5	RT	Reverse	CTCTCCACCAAATATT
E2F2	6	PCR	Forward	GAGCCTCTCCCCTCTACCTCCAC
E2F2	6	PCR	Reverse	AGAGAAAACTAAGGCCGGTCTCTCC
E2F2	6	RT	Reverse	AGAGAAAACTAAGGCC
INSIG1	1	PCR	Forward	GATCCAGAGGAATGTCACTCTCTTCC
INSIG1	1	PCR	Reverse	CCGAAAAGACACCCTTTTTACTTCC
INSIG1	1	RT	Reverse	CCGAAAAGACACCCTT
INSIG1	2	PCR	Forward	AGAGAATGGGCCAGTGTCATGC
INSIG1	2	PCR	Reverse	GAGTCATTTGAGTATATAACAAAACAAAACG
INSIG1	2	RT	Reverse	GAGTCATTTGAGTATA
INSIG1	3	PCR	Forward	GTGGACATTTGATCGTTCCAGAAG
INSIG1	3	PCR	Reverse	GCTTTCCAAAGCCAATATTTGAAAAC
INSIG1	3	RT	Reverse	GCTTTCCAAAGCCAAT
INSIG1	4	PCR	Forward	GTATACGTCCCCAGATTTCCTCTAT
INSIG1	4	PCR	Reverse	TCAAGGTAAAGAGAAGCAAGTTTTAGA
INSIG1	4	RT	Reverse	TCAAGGTAAAGAGAAG

Table 5. Primers used in quantitative RT-PCR of U12Y and U2Y:

Intron Number	Used in RT or PCR Step	Primer Direction	Primer Sequence (5'→3')
1	PCR	Forward	TCGACGTAAACCTATATGCGCTAGTC
1	PCR	Reverse	GAAATGATGTCTTACTGATTCCCAACG
1	RT	Reverse	GAAATGATGTCTTACT
2	PCR	Forward	GTCGTCTCCAACTCCTCACTATCCA
2	PCR	Reverse	TTACCATTCACATGTATACTGCCCTTAAA
2	RT	Reverse	TTACCATTCACATGTA
3	PCR	Forward	GACGTAAACGGCCACAAGTTCAG
3	PCR	Reverse	AAGAAAAATAACTTTCTGTTTGGGAACG
3	RT	Reverse	AAGAAAAATAACTTTC

performed with 2 minute extension times and annealing temperatures of 62.5°C, 60°C, 61.5°C, and 61°C for the E2F2, SmE, INSIG1, and U12Y/U2Y samples, respectively (based on lowest predicted Tm). Amplicons were separated on a 10% polyacrylamide sequencing gel, and quantitated using a Molecular Dynamics phosphorimager.

Construction of *Drosophila* Splicing Reporter Plasmids

Reporter plasmids contained a *Drosophila* NHE3 gene segment fused to CFP or YFP coding sequences (destabilized with "PEST" sequences) (Clontech) with modifications as schematized in Figure 3 (sequences available upon request). The NHE3 segment spanning exon 4 through exon 6 was PCR-amplified from genomic clone DS07134 (Berkeley *Drosophila* Genome Database) with primers that introduced an upstream translation start site, a downstream linker sequence, and mutations to remove hydrophobic amino acids from the predicted transmembrane domain (Primer sequences included below). Alternative primers were used to mutate U12-type consensus sequences to U2-type sequences. PCR products were inserted between the BglII and AgeI sites of pd2-ECFP-N1 and pd2-EYFP-N1 plasmids (Clontech). Synthetic DNA including the entire NHE3 intron 6 sequence (see below) was inserted into CFP and YFP coding sequences via a BcgI site. A Gal4-responsive UAS promoter was excised from pUAST (Brand and Perrimon, 1993) using BamHI and BglII sites, and was inserted within a BamHI site upstream of the NHE3 gene segment. Finally, tandem constructs were created by inserting sequences from one plasmid, cut with XhoI and NgoMIV, into the XhoI and XmaI sites of the second plasmid.

Primers for amplification of NHE3 gene segment (and adding translation start and linker):

Forward: 5'-GCA GAC TCA GAT CTC TCG AGA TAC GAG TCC CGG GGG ATC CGC CGC CAC CAT GGT GGC CAT ATT CAA CGA TCT TAG GGT CG-3'

Reverse: 5'-GCA GAC TCA CCG GTC CAG CCG ATC CAG CCG ATC CAG CTC CGG AAG CTC CGT CTC GAA CCC GAG TAA ATT TGG TC-3'

Primers used to mutate hydrophobic residues in predicted transmembrane region:

Forward: 5'-TCC AAC TCC TCA CTA TCC ATG ATG AGT GGC GCG GAC ATG GGA TGC TTG ACA GCA TTG-3'

Reverse: 5'-GCC ACT CAT CAT GGA TAG TGA GGA GTT GGA GAC GAC ATC GCT CAA CGA GCG TAG AAA AG-3'

Oligonucleotide used to mutate U12-type intron to U2-type: (forward)

5'-TCC AAC TCC TCA CTA TCC ATG ATG AGT GGC GCG GAC ATG GGA TGC TTG ACA GCA TTG GTA AGT GCC TCA AAG ATT AAT GAT AAA TTA TTT TAT TTA AGG GCA GTA TAC ATG TGA ATG GTA AAA TTT GAG TTA ATT TTT GTT ACC TCA AAG TGG CAT TGA AGA TGT ATA TCT ATT CCG CAT TAA TTA GAA AAC ACC TAT GCC TGA CCC TGT CCC TTT TTT TTC CAC AGA TGA CCA AAT TTA CTC GGG TTC GAG AC-3'

Oligonucleotides used to insert NHE3 intron 6 sequence into BcgI site of YFP and CFP:

Forward: 5'-GTC CGG CGA GGG CGA GGT AAG TTA TTT TCA GCA TAG AAA TAC TTA AAA AGG CGT TCC CAA ACA GAA AGT TAT TTT TCT TTT TTT CAC AGG GCG ATG CCA CCT ACG GCA A-3'

Reverse: 5'-TTG CCG TAG GTG GCA TCG CCC TG-3'

Transfection of *Drosophila* S2 Cells

S2 cells were grown at 25°C in Shields and Sang M3 medium (Sigma) containing 12.5% fetal bovine serum (Gibco-BRL) in 35mm dishes and transiently transfected using 25μL of Lipofectin reagent (Gibco-BRL) per dish, according to the manufacturer's directions. For most experiments, 0.5μg of reporter plasmids were co-transfected with 1.5μg of either pmt-Gal4 (metallothionein expression, gift of S. Artavanis-Tsakonas) or heat-shock-Gal4 plasmid (gift of T. Xu). Experiments to investigate splicing saturation used between 1μg and 0.05μg reporter plasmid, supplementing with unrelated plasmid to keep the total transfected DNA constant. Expression was induced 24 hours after transfection either by adding 0.7mM cupric sulfate (Sigma) or by heat shock at 37°C for 30 minutes.

Fluorescence Microscopy

Cells were visualized 12-24 hours after induction using a Zeiss Axiophot II fluorescence microscope with CFP (Ex: D436/20; Em: D480/40; BS: 455dclp) and YFP (Ex: HQ500/20; Em: HQ535/30; BS: Q515lp) filter sets (Chroma). Images were captured

using a Quantix CCD camera (Photometrics) and analyzed with IP Lab software (Scanalytics).

FACS Analysis

Cells were analyzed 12-24 hours after induction with a FACS Vantage dual laser flow cytometer (Becton-Dickinson). CFP was excited at 457nm with a Spectra-Physics 2025 argon laser and the fluorescence emission was collected through a 480/30nm band pass filter. YFP was excited at 514nm using a Coherent Innova 70 argon-krypton laser and the fluorescence collected through a 550/30 nm band pass filter. 50,000 cells were observed for each sample. Statistical analysis was performed using WINMIDI software (Joseph Trotter, Scripps Clinic, CA).

Northern Blot Analysis

Growth medium was removed from S2 cells at various times after induction, and cells were disrupted in 1mL of Trizol (Gibco-BRL) for RNA isolation according to the manufacturer. Total RNA (~10μg) was run per lane on a formaldehyde-agarose gel for Northern blotting and was probed with internally ^{32}P-labeled DNA probes covering either the entire open reading frame of YFP, or ribosomal protein 49 (gift of L. Cooley), or the non-consensus internal sequences (nucleotides 319-436, Figure 3B) of NHE3 intron 5.

Appendix A

Creation of Transgenic *Drosophila melanogaster* Lines to Assess Developmental or Tissue-Specific Changes in U12-type Splicing

---

Summary

In order to monitor any changes in U12-type splicing activity during embryogenesis or in different tissues, I made transgenic flies containing stably integrated genomic copies of two of the fluorescent reporter constructs described in Chapter III. P-element-mediated integration was used to create several different fly lines with the constructs randomly inserted at different positions within the genome. In each fly, either the U2C-U12Y construct or the U12C-U2Y construct was integrated. Placement of both transcription units into a single P-element insertion segment was intended to minimize any variations of gene expression efficiency that might result from insertion at different sites within the genome (Spofford, 1976). I intended to follow the accumulation of cyan and yellow fluorescent protein in the entire fly, and to use the ratio of the two signals to monitor the activity of each type of spliceosome. The integration of the two constructs with reciprocal color schemes into different fly lines was intended to control for differences in cyan and yellow fluorescence signal intensities. Several fly lines (each with a different P-element integration site) were made to be sure that the fluorescence patterns would be reproducible. However, fluorescence was not detectable in these transgenic flies. Fluorescent protein production was also undetectable in the flies by Western blot, but paradoxically, substantial mRNA production was observed. These

observations suggested that fluorescent protein accumulation was hindered either because of inefficient translation or rapid degradation.

Results

Transgenic Drosophila melanogaster lines made to report U12-type and U2-type splicing efficiency

The fluorescent protein reporter constructs U2C-U12Y and U12C-U2Y described in Chapter III (Figure 9) were stably integrated into the genomes of several flies using P-element-mediated integration. The two constructs were separately inserted, via BglII and StuI restriction sites, into the compatible BamHI and StuI sites of a pW8 vector (gift of T. Xu), which contains a red-eye marker and a multiple cloning site flanked by P-element ends. The two resultant plasmids were separately microinjected (injections generously performed by Y.-Q. Tan in Carl Hashimoto's lab), along with a plasmid encoding the P-element transposase, into >200 pre-blastoderm *Drosophila* embryos from a white-eyed fly line (within 45 minutes of egg-laying) according to standard procedures (Spradling, 1986; Ashburner, 1989). Plasmids were injected into the posterior pole of the embryo to maximize the probability of P-element insertion into the DNA of germline progenitor cells. Embryos that survived to adulthood were mated in separate vials with white-eyed adults. The resultant offspring were screened for the presence of red eyes, indicating insertion of the P-element sequences. Four separate fly lines containing the U2C-U12Y insertion, and three lines containing the U12C-U2Y insertion were obtained. A standard set of crosses was then performed to identify which chromosome contained the transgene

insertion for each fly line, and lines were created with appropriate balancer chromosomes to ensure propagation of the transgenes.

Assessing the production of fluorescent proteins from the transgenic fly lines

Since transcription from the integrated transgene constructs was driven by Gal4 responsive promoters, the transgenic flies were mated with various "driver lines" expressing Gal4 from different tissue-specific or ubiquitous promoters. Flies expressing Gal4 from a heat-shock promoter (gift of T. Xu), a ubiquitous actin 5C promoter (gift of T. Xu), a neuron-specific *Nrv2* promoter (gift of P. Salvaterra [Sun et al., 1999]), or a muscle-specific *Nrv1* promoter (gift of P. Salvaterra [Sun et al., 1999]) were used as driver lines. These driver lines are known to express Gal4 in great abundance, which was confirmed by mating these lines with a UAS-GFP line (gift of T. Xu; UAS is the Gal4 responsive promoter or "upstream activating sequence") and observing strong GFP fluorescence. However, mating of these driver lines with the U2C-U12Y or U12C-U12Y lines produced no detectable fluorescence in any of the resulting offspring. This was surprising since abundant fluorescence signal was observed in *Drosophila* tissue culture cells transfected with the identical constructs.

I first assumed that the destabilizing "PEST" sequences may have been responsible for the undetectable fluorescent protein levels. Normal GFP protein is known to have a half-life of 24 hours (in mammalian cells), whereas the fusion proteins encoded by the constructs were destabilized to have two-hour half lives (Clontech product literature). Since this would be expected to dramatically diminish the fluorescence signal, I decided to create a second set of transgenic flies with the "PEST" sequences

removed (between the BsrGI and NotI sites). Thus, new clones were made for U2C-U12Y(-PEST) and U12C-U2Y(-PEST) in the pW8 vector and were injected into embryos to produce new transgenic lines. This time, four separate insertion lines were obtained for each of the two constructs. Unfortunately, mating of any of these new lines with any of the Gal4 driver lines mentioned above again produced offspring that showed no detectable fluorescence. To verify that the modified constructs were still capable of producing fluorescent protein, they were transfected into *Drosophila* S2 cells and expression was induced as before. They produced fluorescence that was 2- to 3-fold brighter than the destabilized constructs, but with similar yellow-to-cyan ratios as observed before.

Although the protein was undetectable by fluorescence microscopy, I hoped that it might be detectable by Western blot. Fly crosses were carried out between female virgins carrying an Actin 5C-Gal4 transgene and males carrying the following transgenes: 1) UAS-GFP, 2) U2C-U12Y(-PEST), 3) U12C-U2Y(-PEST), or 4) no transgene. Offspring from the first mating produced bright, green fluorescence within 24 hours of egg laying, while the other three matings produced none. Five adult flies were homogenized in protein loading buffer, boiled, centrifuged to remove debris, and 1/5th of the supernatant was run on a protein gel. The flies were chosen based on the lack of phenotypic markers for the balancer chromosomes to ensure that the Actin 5C-Gal4 and the fluorescent protein transgenes were present. A polyclonal rabbit anti-GFP antibody (gift of I. Mellman) was used to probe the blot for the presence of GFP, YFP, or CFP (rabbit polyclonal anti-GFP has been reported to cross-react well with YFP and CFP on Western blots [Clontech product literature]). The blot (Figure 15) showed an extremely

Figure 15. Fluorescent Protein Production Assessed by Western Blot

Total protein prepared from adult flies was probed with rabbit polyclonal α-GFP antibody, which is known to cross-react well with YFP and CFP. Offspring from the mating of Actin 5C-Gal4 transgene-containing flies with flies containing either UAS-GFP, U2C-U12Y(-PEST), U12C-U2Y(-PEST), or wild-type flies were used (lanes 1, 2, 3, and 4, respectively). The presence of both transgenes in the offspring was determined by choosing flies that lacked the phenotypic markers for balancer chromosomes. Protein samples were prepared by homogenizing five adult flies in protein loading buffer, boiling, and centrifuging to remove debris. One-fifth of the supernatant was loaded in each lane of the protein gel. On the right are schematics of the transgene reporter constructs (not drawn to scale) that are present in the fly genomes.

intense GFP band for the positive control UAS-GFP flies, and no similar band for the negative control wild-type flies. Since the U2C-U12Y(-PEST)- and U12C-U2Y(-PEST)-containing flies were expected to produce fusion proteins with YFP and CFP, the predicted protein product should have migrated more slowly on a gel than GFP alone. No such bands were observed on the blot, but I cannot exclude the possibility that they co-migrated with the slower-migrating nonspecific cross-reactive band that is present in all four lanes.

Assessing mRNA production from the transgenic fly lines

Since no fluorescent protein was detectable in the transgenic flies, I next asked whether the mRNA was being made in adequate amounts. RNA was isolated from the adult offspring of Actin 5C-Gal4 driver flies mated, as before, with flies containing either 1) UAS-GFP, 2) U2C-U12Y(-PEST), or 3) U12C-U2Y(-PEST). These RNA samples were run on a formaldehyde-agarose gel, alongside RNA (for size comparison) isolated from S2 cells transfected with either U2Y(-PEST) or the original Clontech YFP construct driven by a UAS promoter and with the sixth intron of NHE3 inserted. The blot was probed with internally-radiolabeled random-primed DNA complementary to the entire YFP coding sequence. Although I was initially surprised to see (Figure 16, lane 1) that the GFP producing flies showed no mRNA band on the Northern blot (since GFP, YFP, and CFP differ from each other by only a few codons), I realized that the YFP probes were made with humanized codon sequences (from the Clontech construct) whereas the UAS-GFP flies carry GFP with the original jellyfish codons. These sequences differ enough that cross-hybridization would not be expected. Lanes 2 and 3 showed mRNA

Figure 16. Fluorescent Protein-Coding mRNA Produced in Transgenic Flies Co-migrates with mRNA Produced in Tansfected S2 Cells

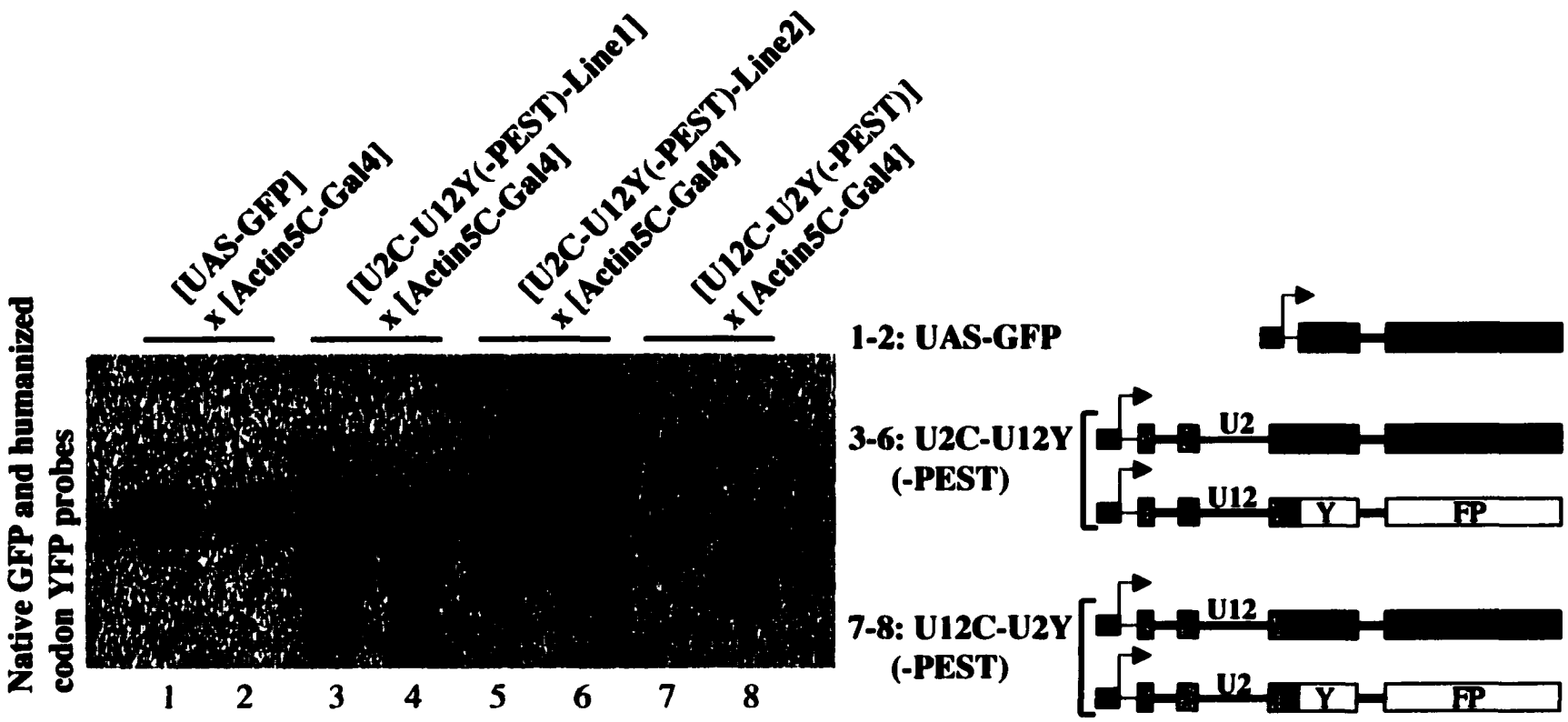
Northern blot showing RNA samples isolated from either transgenic flies or transfected *Drosophila* S2 cells hybridized with an internally-labeled random-primed DNA probes complementary to the entire yellow fluorescent protein coding sequence (with humanized codons). RNA was prepared from adult flies carrying both the Actin 5C-Gal4 transgene and either UAS-GFP, U2C-U12Y(-PEST), or U12C-U2Y(-PEST) (lanes 1, 2, and 3, respectively). Lanes 4 and 5 contain RNA prepared from S2 cells co-transfected with metallothionein-Gal4 and either UAS-YFP (with intron 6 of NHE3 inserted) or U2Y(-PEST), respectively. RNA was isolated 12 hours after induction with cupric sulfate. The lack of hybridization to RNA in lane 1 is due to dramatic differences in humanized codon YFP sequences and native jellyfish GFP that are expressed in the fly. The reporter constructs expressed in the different lanes are schematized on the right.

bands produced from the U2C-U12Y(-PEST) and U12C-U2Y(-PEST) flies respectively. As expected, they migrated at the same level as the mRNA band in lane 5, which corresponds to the spliced U2Y(-PEST) transcript. The YFP mRNA in lane 4, as expected, migrated more quickly. The difference in intensity between the mRNA signals in lanes 2 and 3 may be the result of differences in gel loading as well as differences in transcription levels from constructs inserted at different chromosomal locations.

Having confirmed that the transgenic flies do indeed produce an mRNA species of the predicted size, I next wanted to ask whether the amount of the mRNA was comparable to that produced by the UAS-GFP flies (which produce ample fluorescent protein). To probe for the GFP mRNA in these flies, internally labeled random-primed DNA complementary to the native (non-humanized) GFP coding sequence was made. A new Northern blot (Figure 17) was performed with RNA samples isolated from similar sets of flies as above. Each RNA preparation was made in duplicate from sibling flies to ensure reproducibility. The blot was probed simultaneously with equally radioactive native GFP and humanized-codon YFP probes. Although the two probe sequences might hybridize to their targets with differing affinities, it is reasonable to use the relative intensities of the GFP and YFP bands on the Northern blot to estimate the relative mRNA abundance. The U2C-U12Y(-PEST) and U12C-U2Y(-PEST) transgenic flies produced mRNA bands (lanes 3-8) that were both more and less intense than the GFP mRNA bands (lanes 1 and 2) produced by the UAS-GFP flies. Although this blot shows variation in mRNA levels among the different fly lines, we can estimate that at least some of the lines produce fluorescent protein-coding mRNA in quantities that are comparable to that produced by the UAS-GFP flies.

Figure 17. Similar Levels of Fluorescent Protein-Coding mRNAs Are Produced in Non-Fluorescent Transgenic Flies as Compared to Fluorescent GFP-Expressing Flies

Northern blot of RNA prepared from various transgenic *Drosophila* lines simultaneously probed with internally-labeled random-primed DNA complementary to humanized-codon YFP coding sequences and to native jellyfish GFP coding sequences. Five million CPM of each probe were used. RNA was prepared from adult flies with the same genotypes as in Figure 16. For the U2C-U12Y(-PEST) transgene, two different lines were used, each with the transgene inserted at a different chromosomal position. For each genotype, RNA was prepared in duplicate from sibling flies to assess reproducibility. The reporter transgenes are schematized on the right.



Discussion

It was perplexing that the transgenic flies produced no detectable fluorescent protein even though the mRNA appeared to be produced in adequate quantities. This was especially difficult to understand, given that the same constructs did produce ample fluorescent protein when transfected into *Drosophila* tissue culture cells.

I attempted to troubleshoot this problem by first creating a second set of transgenic flies with the "PEST" sequences removed, believing that the instability of the fluorescent proteins in the original flies may have been responsible for the reduced protein levels. However, this did not provide a sufficient increase in protein levels.

Early on, I also suspected that the production of large quantities of reporter protein may have been toxic to the flies, so that no flies would survive embryogenesis if they had both the Gal4 driver transgene and the reporter transgene in their genome. However, this idea was disproved by following segregation of the transgene-containing chromosomes by the absence of balancer chromosome marker traits in the offspring. I could thereby demonstrate that flies containing both transgenes were viable to adulthood. Furthermore, the fact that these flies produced the expected mRNA was a strong indication that both the driver and reporter transgenes were present.

I also suspected that perhaps the tandem arrangement of the two transcription units within the transgene may have caused gene silencing, as reported previously for other repeated genes in *Drosophila* (Dorer and Henikoff, 1994, 1997; Fanti et al., 1998; Birchler et al., 2000). Such silencing is believed to be mediated by association with a chromosomal heterochromatin complex, but has not yet been demonstrated to occur for fewer than three repeated units. Since the chromosomally integrated transgenes would be

susceptible to such silencing but transfected plasmids would not, I thought this might explain the discrepancy in protein expression between transgenic flies and transfected cells. However, the production of mRNA to levels that were comparable with GFP mRNA levels in fluorescent flies argued against transcriptional silencing.

The evidence therefore points to either inefficient protein translation or poor protein stability within the flies. It is possible that translation might be hindered by inefficient codon usage. Since the GFP producing flies carry the original jellyfish fluorescent protein codons, while the transgenic flies carry humanized codon versions of CFP and YFP, one might expect the translation efficiency to differ between the two types of mRNA. But upon closer examination, I found that codon usage frequencies are very similar between *Drosophila* and humans (Nakamura et al., 2000). Therefore, the humanized YFP and CFP codons should not hinder translation in *Drosophila*. It is possible, however, that the NHE3 coding sequences in the fusion constructs contain inefficient codons. Although the NHE3 gene is endogenously produced in *Drosophila*, it does have several suboptimal codons (Nakamura et al., 2000), perhaps because NHE3 production in large quantities is not normally required. It is possible that inefficient translation might not produce detectable fluorescent protein when the transgene is present in one copy per cell in transgenic flies, but might produce much more protein when multiple copies are transfected into each tissue culture cell. Unfortunately, testing this hypothesis by optimizing all of the inefficient NHE3 codons is not a trivial task.

It is also possible that poorly understood but important differences exist between cells in a fly and cells in tissue culture that allow expression of fusion proteins in the

latter but not in the former. Although the inefficient codon explanation is the best I am able to offer, other, less well-understood mechanisms may be involved.

References

Ashburner, M. (1989). Drosophila, A laboratory manual (Cold Spring Harbor, Cold Spring Harbor Press).

Bauren, G., and Wieslander, L. (1994). Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription, Cell *76*, 183-92.

Beyer, A. L., and Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts, Genes Dev *2*, 754-65.

Birchler, J. A., Bhadra, M. P., and Bhadra, U. (2000). Making noise about silence: repression of repeated genes in animals, Curr Opin Genet Dev *10*, 211-6.

Black, D. L. (2000). Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology, Cell *103*, 367-70.

Bousquet-Antonelli, C., Presutti, C., and Tollervey, D. (2000). Identification of a regulated pathway for nuclear pre-mRNA turnover, Cell *102*, 765-75.

Bozzoni, I., Fragapane, P., Annesi, F., Pierandrei-Amaldi, P., Amaldi, F., and Beccari, E. (1984). Expression of two Xenopus laevis ribosomal protein genes in injected frog oocytes. A specific splicing block interferes with the L1 RNA maturation, J Mol Biol *180*, 987-1005.

Brand, A. H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes, Development *118*, 401-15.

Burge, C. B., Padgett, R. A., and Sharp, P. A. (1998). Evolutionary fates and origins of U12-type introns, Mol Cell *2*, 773-85.

Burge, C. B., Tuschl, T., and Sharp, P. A. (1999). Splicing of Precursors to mRNAs by the Spliceosomes. In The RNA world : the nature of modern RNA suggests a prebiotic RNA, R. F. Gesteland, T. Cech, and J. F. Atkins, eds. (Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press), pp. 525-560.

Caffarelli, E., Arese, M., Santoro, B., Fragapane, P., and Bozzoni, I. (1994). In vitro study of processing of the intron-encoded U16 small nucleolar RNA in Xenopus laevis, Mol Cell Biol *14*, 2966-74.

Caffarelli, E., Fragapane, P., and Bozzoni, I. (1992). Inefficient in vitro splicing of the regulatory intron of the L1 ribosomal protein gene of X.laevis depends on suboptimal splice site sequences, Biochem Biophys Res Commun *183*, 680-7.

Caffarelli, E., Fragapane, P., Gehring, C., and Bozzoni, I. (1987). The accumulation of mature RNA for the Xenopus laevis ribosomal protein L1 is controlled at the level of splicing and turnover of the precursor RNA, Embo J *6*, 3493-8.

Datta, B., and Weiner, A. M. (1991). Genetic evidence for base pairing between U2 and U6 snRNA in mammalian mRNA splicing, Nature *352*, 821-4.

Dietrich, R. C., Incorvaia, R., and Padgett, R. A. (1997). Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns, Mol Cell *1*, 151-60.

Dorer, D. R., and Henikoff, S. (1994). Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila, Cell *77*, 993-1002.

Dorer, D. R., and Henikoff, S. (1997). Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans, Genetics *147*, 1181-90.

Fanti, L., Dorer, D. R., Berloco, M., Henikoff, S., and Pimpinelli, S. (1998). Heterochromatin protein 1 binds transgene arrays, Chromosoma *107*, 286-92.

Fragapane, P., Caffarelli, E., Lener, M., Prislei, S., Santoro, B., and Bozzoni, I. (1992). Identification of the sequences responsible for the splicing phenotype of the regulatory intron of the L1 ribosomal protein gene of Xenopus laevis, Mol Cell Biol *12*, 1117-25.

Freeman, W. M., Walker, S. J., and Vrana, K. E. (1999). Quantitative RT-PCR: pitfalls and potential, Biotechniques *26*, 112-22, 124-5.

Frilander, M. J., and Steitz, J. A. (1999). Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions, Genes Dev *13*, 851-63.

Gardner, D. G., Cathala, G., Lan, N. Y., David-Inouye, Y., and Baxter, J. D. (1988). Processing of the primary transcript for the rat growth hormone gene in vivo, DNA *7*, 537-44.

Grabowski, P. J., and Black, D. L. (2001). Alternative RNA splicing in the nervous system, Prog Neurobiol *65*, 289-308.

Gudas, J. M., Knight, G. B., and Pardee, A. B. (1990). Ordered splicing of thymidine kinase pre-mRNA during the S phase of the cell cycle, Mol Cell Biol *10*, 5591-5.

Hall, S. L., and Padgett, R. A. (1994). Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites, J Mol Biol *239*, 357-65.

Hall, S. L., and Padgett, R. A. (1996). Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns, Science *271*, 1716-8.

Hastings, M. L., and Krainer, A. R. (2001). Functions of SR proteins in the U12-dependent AT-AC pre-mRNA splicing pathway, Rna *7*, 471-82.

Hausner, T. P., Giglio, L. M., and Weiner, A. M. (1990). Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles, Genes Dev *4*, 2146-56.

Hirose, Y., and Manley, J. L. (2000). RNA polymerase II and the integration of nuclear events, Genes Dev *14*, 1415-29.

Ivey-Hoyle, M., Conroy, R., Huber, H. E., Goodhart, P. J., Oliff, A., and Heimbrook, D. C. (1993). Cloning and characterization of E2F-2, a novel protein with the biochemical properties of transcription factor E2F, Mol Cell Biol *13*, 7802-12.

Jackson, I. J. (1991). A reappraisal of non-consensus mRNA splice sites, Nucleic Acids Res *19*, 3795-8.

Kenan, D. J., Query, C. C., and Keene, J. D. (1991). RNA recognition: towards identifying determinants of specificity, Trends Biochem Sci *16*, 214-20.

Kessler, O., Jiang, Y., and Chasin, L. A. (1993). Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA, Mol Cell Biol *13*, 6211-22.

Kiseleva, E., Wurtz, T., Visa, N., and Daneholt, B. (1994). Assembly and disassembly of spliceosomes along a specific pre-messenger RNP fiber, Embo J *13*, 6052-61.

Kozak, M. (1987). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells, J Mol Biol *196*, 947-50.

Lengyel, J., Spradling, A., and Penman, S. (1975). Methods with insect cells in suspension culture. II. Drosophila melanogaster, Methods Cell Biol *10*, 195-208.

Levine, A., and Durbin, R. (2001). A computational scan for U12-dependent introns in the human genome sequence, Nucleic Acids Res *29*, 4006-13.

Madhani, H. D., and Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome, Cell *71*, 803-17.

Mayeda, A., Zahler, A. M., Krainer, A. R., and Roth, M. B. (1992). Two members of a conserved family of nuclear phosphoproteins are involved in pre-mRNA splicing, Proceedings of the National Academy of Sciences of the United States of America *89*, 1301-4.

Michaud, S., and Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway, Genes Dev *5*, 2534-46.

Miller, D. M., 3rd, Desai, N. S., Hardin, D. C., Piston, D. W., Patterson, G. H., Fleenor, J., Xu, S., and Fire, A. (1999). Two-color GFP expression system for C. elegans, Biotechniques *26*, 914-8, 920-1.

Misteli, T., and Spector, D. L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo, Mol Cell *3*, 697-705.

Montzka, K. A., and Steitz, J. A. (1988). Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. Proc Natl Acad Sci U S A *85*, 8885-9.

Montzka-Wassarman, K. M., and Steitz, J. A. (1992). The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex, Molecular and Cellular Biology *12*, 1276-1285.

Moore, M. J., Query, C. C., and Sharp, P. A. (1993). Splicing of Precursors to mRNA by the Spliceosome. In The RNA world : the nature of modern RNA suggests a prebiotica RNA world, R. F. Gesteland, and J. F. Atkins, eds. (Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press), pp. 303-357.

Moore, M. J., and Sharp, P. A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing, Nature *365*, 364-8.

Mount, S. M. (1982). A catalogue of splice junction sequences, Nucleic Acids Res *10*, 459-72.

Nakamura, Y., Gojobori, T., and Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000, Nucleic Acids Res *28*, 292.

Nelson, K. K., and Green, M. R. (1989). Mammalian U2 snRNP has a sequence-specific RNA-binding activity, Genes Dev *3*, 1562-71.

Newman, A. J. (1997). The role of U5 snRNP in pre-mRNA splicing, Embo J *16*, 5797-800.

Otake, L. R., Scamborova, P., Hashimoto, C., and Steitz, J. A. (2002). The Divergent U12-type Spliceosome is Required for Pre-mRNA Splicing and Is Essential for Development in Drosophila, Mol Cell *in press*.

Peng, Y., Schwarz, E. J., Lazar, M. A., Genin, A., Spinner, N. B., and Taub, R. (1997). Cloning, human chromosomal assignment, and adipose and hepatic expression of the CL-6/INSIG1 gene, Genomics *43*, 278-84.

Pierandrei-Amaldi, P., Bozzoni, I., and Cardinali, B. (1988). Expression of the gene for ribosomal protein L1 in Xenopus embryos: alteration of gene dosage by microinjection, Genes Dev *2*, 23-31.

Reed, R. (1996). Initial splice-site recognition and pairing during pre-mRNA splicing, Curr Opin Genet Dev *6*, 215-20.

Robberson, B. L., Cote, G. J., and Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons, Mol Cell Biol *10*, 84-94.

Sharp, P. A., and Burge, C. B. (1997). Classification of introns: U2-type or U12-type, Cell *91*, 875-9.

Smith, C. W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control, Trends Biochem Sci *25*, 381-8.

Solnick, D. (1985). Trans splicing of mRNA precursors, Cell *42*, 157-64.

Sontheimer, E. J., and Steitz, J. A. (1993). The U5 and U6 small nuclear RNAs as active site components of the spliceosome [see comments] [published erratum appears in Science 1994 Feb 11;263(5148):739], Science *262*, 1989-96.

Spafford, J. D., Spencer, A. N., and Gallin, W. J. (1998). A putative voltage-gated sodium channel alpha subunit (PpSCN1) from the hydrozoan jellyfish, Polyorchis penicillatus: structural comparisons and evolutionary considerations, Biochem Biophys Res Commun *244*, 772-80.

Spofford, J. B. (1976). Position-effect variegation in Drosophila. In Genetics and Biology of Drosophila, M. A. a. E. Novitski, ed. (Londor, Academic Press), pp. 955-1019.

Spradling, A. C. (1986). P-element mediated transformation. In Drosophila: A Practical Approach, D. B. Roberts, ed. (Oxford, IRL Press), pp. 175-197.

Staley, J. P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things, Cell *92*, 315-26.

Stanford, D. R., Perry, C. A., Holicky, E. L., Rohleder, A. M., and Wieben, E. D. (1988). The small nuclear ribonucleoprotein E protein gene contains four introns and has upstream similarities to genes for ribosomal proteins, J Biol Chem *263*, 17772-9.

Steitz, J. A., Black, D. L., Gerke, V., Parker, K. A., Kramer, A., Frendewey, D., and Keller, W. (1988). Functions of the Abundant U-snRNPs. In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, M. L. Birnstiel, ed. (Berlin, Springer-Verlag), pp. 115-154.

Steitz, T. A., and Steitz, J. A. (1993). A general two-metal-ion mechanism for catalytic RNA, Proceedings of the National Academy of Sciences of the United States of America *90*, 6498-502.

Sun, B., Xu, P., and Salvaterra, P. M. (1999). Dynamic visualization of nervous system in live Drosophila, Proc Natl Acad Sci U S A *96*, 10438-43.

Tarn, W. Y., and Steitz, J. A. (1996a). Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns, Science *273*, 1824-32.

Tarn, W. Y., and Steitz, J. A. (1996b). A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro, Cell *84*, 801-11.

Tarn, W. Y., and Steitz, J. A. (1997). Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge, Trends Biochem Sci *22*, 132-7.

Wassarman, D. A., and Steitz, J. A. (1992). Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing, Science *257*, 1918-25.

Wetterberg, I., Bauren, G., and Wieslander, L. (1996). The intranuclear site of excision of each intron in Balbiani ring 3 pre- mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns, Rna *2*, 641-51.

Wu, J., and Manley, J. L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing, Genes Dev *3*, 1553-61.

Wu, J. A., and Manley, J. L. (1991). Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA, Nature *352*, 818-21.

Wu, Q., and Krainer, A. R. (1996). U1-mediated exon definition interactions between AT-AC and GT-AG introns, Science *274*, 1005-8.

Zahler, A. M., Lane, W. S., Stolk, J. A., and Roth, M. B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors, Genes Dev *6*, 837-47.