Abstract

An experimental and genomic approach to the regulation of alternative pre-mRNA splicing in *Drosophila rnp-4f*

By Rebecca A. Fetherson

*rnp-4f*, a *Drosophila* gene containing nine spliceosomal introns, encodes a nuclear PRP protein that functions constitutively in spliceosome assembly. Nothing is known about molecular mechanisms regulating splicing decisions for this gene, or the factors that modulate the degree of mRNA translation for the encoded spliceosomal assembly factor during development. Here we report results of a systematic study of alternative pre-mRNA splicing in *rnp-4f*. Reverse transcription-polymerase chain reaction (RT-PCR) and hybridization analysis show that introns #I - IV and #VI - VIII are constitutively spliced, while five patterns of alternative splicing are observed in two other pre-mRNA regions. Intron V is infrequently spliced in all developmental stages, which results in generation of an in-frame stop codon and a predicted truncated protein lacking a nuclear localization signal, so that facultative splicing regulates the subcellular localization of the encoded protein. Intron 0 plus a portion of exon 2 (intron-0-alt) is alternatively spliced from the 5'-UTR, resulting in loss of an evolutionarily conserved stem-loop located in exon 2. Northern analysis of poly (A⁺) mRNAs reveals two differently sized *rnp-4f* mRNA isomers, one dominating during mid-embryo stages of CNS development. The size of one of these mRNAs corresponds to that predicted for the constitutively spliced *rnp-4f* transcript. The other mRNA corresponds to the predicted size of the alternatively spliced transcript in which intron 0 plus a portion of exon 2 is excised, a result confirmed by hybridization. Potential molecular mechanisms which could regulate the observed splicing patterns are discussed, specifically from the viewpoint of putative cis-regulatory elements within *rnp-4f* pre-mRNAs. The alternative mRNA isoforms observed within the 5’-UTR are predicted to affect regulatory processes, including translational efficiency. A model is proposed in which RNP-4F controls its own level of expression by feedback against evolutionarily conserved twin stem-loops located at the 3'-end of the intron-0-alt pre-mRNA tract.
An experimental and genomic approach to the regulation of alternative pre-mRNA splicing in *Drosophila rnp-4f*

A Thesis
Submitted to the
Faculty of Miami University
In partial fulfillment of
the requirements for the degree
Master of Science
Department of Zoology

By
Rebecca A. Fetherson
Miami University
Oxford, OH
2005

Advisor_____________________
Dr. Jack C. Vaughn

Reader_____________________
Dr. Susan M. Hoffman
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detailed Table of Contents</td>
<td>iii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vi</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>viii</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>14</td>
</tr>
<tr>
<td>Results</td>
<td>24</td>
</tr>
<tr>
<td>Discussion</td>
<td>32</td>
</tr>
<tr>
<td>Future Research Directions</td>
<td>44</td>
</tr>
<tr>
<td>Appendices</td>
<td>65</td>
</tr>
<tr>
<td>Literature Cited</td>
<td>69</td>
</tr>
</tbody>
</table>
## Detailed Table of Contents

### Introduction

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The <em>rnp-4f</em> Gene</td>
<td>1</td>
</tr>
<tr>
<td>Constitutive pre-mRNA Splicing and Its Regulation</td>
<td>3</td>
</tr>
<tr>
<td>Alternative pre-mRNA Splicing and Its Regulation</td>
<td>5</td>
</tr>
<tr>
<td>(A) SXL Regulation of Alternative Splicing: A Model <em>Drosophila</em> System</td>
<td>6</td>
</tr>
<tr>
<td>(B) Splicing Enhancers and Silencers</td>
<td>8</td>
</tr>
<tr>
<td>Translation Regulation by the 5’-Untranslated Region (UTR) in mRNAs</td>
<td>10</td>
</tr>
<tr>
<td>Surveillance Checkpoints for mRNA Nuclear Exit</td>
<td>12</td>
</tr>
<tr>
<td>Specific Goals of This Thesis Research</td>
<td>13</td>
</tr>
</tbody>
</table>

### Materials and Methods

*H. marina* Stocks and Propagation                                     | 14   |
| Total Cell RNA Isolation                                               | 14   |
| Poly(A⁺) RNA Isolation                                                 | 14   |
| Reverse Transcription (RT)                                             | 15   |
| Primers Utilized                                                       | 15   |
| Polymerase Chain Reaction (PCR)                                        | 17   |
| Determination of *rnp-4f* Isoform Relative Abundance                   | 17   |
| Agarose Gel Electrophoresis                                            | 18   |
| Southern Blot Analysis                                                 | 18   |
| Northern Blot Analysis                                                 | 18   |
| Membrane Hybridizations                                                | 19   |
| (A) End-Labeled Probes                                                 | 19   |
| (B) Random-Primed Probes                                               | 20   |
DNA Fragment Isolation ................................................................. 21
Fragment Ligation/Transformation/Colony Selection ......................... 21
Small Scale Plasmid Isolation ....................................................... 22
DNA Sequencing ........................................................................ 22
Computer-Assisted Sequence Analysis ............................................ 22

Results
Qualitative Analysis of *rnp-4f* mRNA isoforms Across Development .......... 24
(A) Constitutively-Spliced mRNA Tracts ........................................ 25
(B) Cryptic Folded mRNA Tracts .................................................. 25
(C) Facultatively and Alternatively Spliced mRNA Tracts ..................... 27
Conserved Putative Cis-Regulatory Elements Within Intron 0 .................. 28
Quantitative SYBR-Green Analysis of *rnp-4f* Primer Set “A” RT-PCR Products ............................................................................ 30
Putative *cis*-Regulatory Elements Within IntronV Isoforms ................. 30
Northern Blot Analysis of *rnp-4f* mRNA Transcription Across Development ...... 30

Discussion
*rnp-4f* Isoforms Identified Across Development ................................... 32
Regulation of *rnp-4f* Alternative Splicing ......................................... 33
(A) Facultative Splicing Within *rnp-4f* pre-mRNAs ............................ 33
(B) Alternative Splicing Within *rnp-4f* pre-mRNAs ............................ 37
*rnp-4f* Translation Regulation by the 5’-UTR ....................................... 40
A Model for Auto-Regulation of *rnp-4f* mRNA Translation .................... 42

Future Research Directions ............................................................. 44

Appendices
Appendix I: Conventional Reverse Transcription .................................. 65
Appendix II: Thermocycler Reverse Transcription .................................. 66
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td><em>rnp-4f</em> gene</td>
<td>47</td>
</tr>
<tr>
<td>2.</td>
<td>(A) The <em>Drosophila rnp-4f</em> gene is evolutionarily conserved</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>(B) A model for RNP-4F function in the yeast orthologue PRP-24</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Schematic representation of the spliceosome in terms of the role of</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>snRNP particles in pre-mRNA splicing including transitions that require</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a PRP protein</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Constitutive splicing</td>
<td>4</td>
</tr>
<tr>
<td>5.</td>
<td>Alternative splicing</td>
<td>6</td>
</tr>
<tr>
<td>6.</td>
<td>Expression of the <em>Drosophila</em> sex- lethal (SXL) protein</td>
<td>50</td>
</tr>
<tr>
<td>7.</td>
<td>SXL target genes <em>tra</em> (A) and <em>msl-2</em> (B)</td>
<td>51</td>
</tr>
<tr>
<td>8.</td>
<td><em>Drosophila melanogaster</em> life-cycle</td>
<td>52</td>
</tr>
<tr>
<td>9.</td>
<td>Experimental design for detection of alternatively spliced <em>rnp-4f</em></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>transcripts</td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td><em>rnp-4f</em> RT-PCR analysis across development</td>
<td>53</td>
</tr>
<tr>
<td>11.</td>
<td>Model for generation of cryptic folded cDNAs via RT-PCR</td>
<td>54</td>
</tr>
<tr>
<td>12.</td>
<td>Predicted RT-PCR induced single-stranded cDNA secondary structure</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>arising from use of primer set C</td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Predicted RT-PCR induced single-stranded cDNA secondary structure</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>arising from use of primer set D</td>
<td></td>
</tr>
</tbody>
</table>
14. *mp-4f* 5’-UTR sequence alignment for *D. melanogaster* and *D. yakuba* ................................................................. 57

15. Conserved intron 0 alternative intron 0 stem-loops ......................... 58

16. Quantification of *mp-4f* mRNAs across development ...................... 59

17. Predicted stem-loops surrounding intron V 5’- and 3’- splice sites ............ 60

18. *mp-4f* and *sas10* developmental mRNA Northern ....................... 61

19. Model for generation of unspliced (A) and facultatively spliced (B) intron V .................................................................................. 62

20. Model for generation of unspliced intron 0 (A), facultatively spliced intron 0 (B) and alternatively-spliced intron-0-alt (C) isoforms .......... 63

21. Model for control of RNP-4F expression by auto-regulation ............... 64
Acknowledgements

I would first and foremost like to thank my advisor Dr. Jack Vaughn for his endless patience, constant motivation and boundless energy. My interaction with Dr. Vaughn, as both an undergraduate and a graduate student in his laboratory, has not only made me a better scientist, but I believe a better person as well. Dr. Vaughn is my other “dad” who I will dearly miss upon graduation. I would also like to thank the members of my Thesis committee: Dr. Eileen Bridge, Dr. Joyce Fernandes, Dr. Susan Hoffman and Dr. David Pennock. Their comments, suggestions, and insights were greatly appreciated. In particular, I extend my utmost appreciation to Dr. Susan Hoffman for going out of her way to discuss the progress of my research and suggest several procedures that afforded not only completion of my project, but also elevated the implications of my results.

I would also like to thank undergraduate research assistants: Laura Armbruster, Steven Strock, Katie Niese and Kristen White. Their extensive work on many components of this research was invaluable. I’d especially like to thank Kristen White for not only being an excellent research assistant, but also a wonderful friend. I also thank Sunetra Bhatla, Malini Varadarajan and Rajkumar Rajeshwaran for always making me feel funnier than I truly am. Chris Wood generously provided invaluable assistance with sequence analysis. I thank former members of the Vaughn lab, Janaki Rangarajan and Nick Peters for the training they provided on many of the techniques I employed in my research.

Finally, I am grateful for my family, friends and Joe Krueger for their support and constant encouragement. This research work would not have been possible without them.
Graduate Assistantships from Miami University, an NIH grant awarded to Dr. Vaughn and several Miami University undergraduate research awards supported this work.
Introduction

The *rnp-4f* gene

*Drosophila rnp-4f* is a single-copy nuclear gene (Petschek et al. 1997) located at the distal tip of the X-chromosome (Hess et al. 1996). The evolutionarily conserved *rnp-4f* gene encodes a predicted 943 amino acid protein that contains: at least two single-stranded RNA binding motifs (RRM), several half-tetratricopeptide (TPR) repeats, a conserved C-terminal motif (CTM) and a predicted nuclear localization signal (NLS) that is located in the hinge region (Rader and Guthrie 2002). *rnp-4f* contains nine spliceosomal introns (Petschek et al. 1997; Fetherson and Vaughn, unpublished) plus one potential HAC1-like intron (Vaughn et al. 2001), and it transcribes mRNAs which encode corresponding proteins throughout fly development (Fig.1). It has recently been found that *rnp-4f* is an orthologue of the yeast gene *prp-24*, which encodes a protein playing a key role in the recycling of two factors required during intron splicing or removal (Bell et al. 2002; Rader and Guthrie 2002), PRP-24 orthologues in species ranging from yeast to humans have been identified (Fig. 2A). A model for the function of PRP-24 during spliceosome assembly shows that the entering U4/U6 snRNP dimer is stabilized by interactions with the RRM domains and the CTM. The CTM also interacts with Lsm proteins complexed to U6-snRNA (Fig. 2B). Interactions between spliceosomal assembly factors have been extensively studied, and are summarized in Figure 3.

No systematic study of *rnp-4f* alternative splicing has ever been carried out, so that what is currently known is based on the sequencing of a few cDNA clones obtained by screening cDNA libraries (Petschek et al., 1997). In the cited study, it was found that with the exception of intron V, all introns I-VIII are spliced in the one 0-4 h embryo and also in one non-edited adult head cDNAs sequenced, resulting in transcripts predicted to encode a 943 amino acid protein having an estimated mass of 124 kDa. This result was also reported for two virtually full-length cDNA clones isolated from 0-24 h embryos and from Schneider L2 tissue culture cells, following sequencing of the corresponding EST clones (GenBank accession numbers NM_078492 and NM_206630, respectively). Since the predicted proteins in these studies would contain all sequence domains
required for recycling of spliceosomal factors U4 and U6 (Rader and Guthrie 2002), this is probably the constitutive protein among the potential isoforms. One 5’-truncated cDNA clone obtained from pupae was found to splice intron V, resulting in a frameshift-generated stop codon and a predicted protein having only 639 amino acids with an estimated mass of 84 kDa (Petschek et al., 1997). Finally, in a preliminary study utilizing developmental Northernns (Hess et al., 1996) it was reported that two mp-4f mRNA isomers are present throughout development, their sizes being estimated at approximately 3,200- and 3,500-nt. The molecular basis behind these two mRNA isoforms was not determined, but these transcripts may have arisen via (a) alternative splicing, (b) use of alternative promoters, or (c) differential use of 3’-UTR termination sites. The relative frequencies of the two isomers were not reported in the cited study.

At least three different RNP-4F protein sizes have recently been detected in fly embryos based on a study utilizing developmental Westerns (Petschek and Concel 2003, Petschek and Concel 2004), giving reason to predict that different mRNA isoforms may arise from mp-4f pre-mRNA processing, and that protein isoforms may therefore exist. Whether or not these various proteins of apparently different sizes are the result of alternative splicing, and whether they localize to different tissues of the developing fly, await testing. It would be expected that if RNP-4F has the predicted role in spliceosome recycling, then the mRNA encoding this isomer would have a homogeneous distribution in the developing fly. However, Hess et al. (1996) utilized in situ localization and reported that mp-4f mRNAs concentrate in the dorsal roof of the ventral nerve cord, in addition to their predicted broad homogeneous tissue distribution, which may reflect presence of a variant mp-4f mRNA having a unique function at this CNS location. It is however not yet clear whether the mRNA at this location resides in midline glial cells, horizontal nerve tract glial cells, or within the horizontal nerve tract cells themselves.

**Constitutive pre-mRNA splicing and its regulation**

The recent total genome sequencings of higher organisms, including human (Lander et al. 2001) and the genetic model organism *Drosophila melanogaster* (Adams et al. 2000), are significant achievements in modern biology. As a result of this work, it
was found that the fruit fly *Drosophila* contains over 13,000 genes, and human surprisingly contains only about 35,000. The unexpectedly small gene numbers were explained when it was found that many genes can encode more than one protein. One way genes accomplish this is by way of a process termed alternative splicing. In the process of conventional splicing, each intron segment is cut by the spliceosome at both its left and right margins, followed by joining of the margins to form the mature mRNA (Fig. 4).

The spliceosome is a large nuclear macromolecular complex containing as many as 300 different proteins (reviewed in Jurica and Moore 2003), which directs the excision of introns from pre-mRNAs. The spliceosome also contains five small uracil-rich RNAs (snRNAs) designated U1, U2, U4, U5, U6. This complex assembles onto pre-mRNA exon/intron junctions, in part by utilizing information contained in specific sequences within the intron. The 5’-end of the intron contains a GU dinucleotide within a conserved *Drosophila* consensus sequence AG/GU (Mount et al. 1992). The 3’-end of the intron contains three conserved sequence elements including: a branch point with a consensus resembling CTAAAT located at least 15-nt from the 3’-intron end, a pyrimidine-rich tract located between the branch point and the 3’-splice site, and an AG dinucleotide at the 3’-end of the intron within a conserved *Drosophila* consensus sequence TTNCAG/RT (Mount et al. 1992). In the process of splicing (reviewed in Hastings and Krainer 2001; Black 2003), the U1-snRNP binds to the 5’-splice site of the intron by base pairing between the splice site and the RNA portion of U1-snRNP. The branch point is then bound by the SF1 binding protein (BBP in yeast) and the 65 kDa subunit of the U2 auxiliary factor (U2AF) binds the pyrimidine-rich tract located near the 3’-end of the intron. The binding of U1-snRNP and U2AF define the earliest spliceosome assembly steps, resulting in the E or commitment.
complex. Following the formation of this complex, U2-snRNP then binds to the branch point through base pairing of its snRNA. This defines what is called the A complex in spliceosome assembly. Next, the U4/U5/U6 tri-snRNP, recruited by the RNP-4F orthologue, joins the A complex, forming what is now termed the B complex. The B complex is then rearranged to facilitate replacement of U1-snRNP with U6-snRNP at the 5'-splice site to form the C complex. U1 and U4 are lost from the spliceosome in the process of C complex formation. This C spliceosomal complex then catalyzes two transesterification steps that result in intron excision and ligation of the adjacent exons.

It is largely accepted that intronic splice site consensus sequences are not sufficient to direct assembly of the spliceosome onto pre-mRNAs. Therefore, exon/intron boundaries must contain regulatory elements that help direct spliceosome assembly and function in splicing (reviewed in Ladd and Cooper 2002). Non-splice site cis-regulatory sequences have almost exclusively been identified through deletion and/or mutation analyses that enhance or inhibit spliceosome assembly (Wang et al. 2004). These auxiliary elements share several common features: they are small, have variable sequences, are individually weak and therefore usually present in multiple copies, and are evolutionarily conserved between species (Black 2003). Splicing
enhancers are found within both intronic (intron splicing enhancer or ISE) and exonic (exon splicing enhancer or ESE) regions and act to positively stimulate splicing. Even constitutively spliced genes commonly contain exonic splicing enhancers. Several groups of splicing enhancers are known, including purine-rich, AC-rich types and others with more complex composition (Zheng et al. 2004).

Splicing silencer sequences are also found in both introns (intron splicing silencer or ISS) and exons (exon splicing silencer or ESS), though not usually adjacent to constitutively spliced introns. These silencing elements act to inhibit spliceosome assembly, thereby abrogating the splicing of certain introns and/or exons. Compared to splicing enhancer elements, relatively few splicing silencer elements have been characterized by mutational/deletion analysis. The silencing elements that have been identified share little similarity, suggesting there are more yet to be discovered (Zheng et al. 2004). Some of these cis-regulatory elements have been shown to form a secondary structure that can block or enhance splice site selection (Colman et al. 1998); however, most seem to bind trans-acting regulatory proteins that influence splicing (Newton et al. 2003).

**Alternative pre-mRNA splicing and its regulation**

The process of alternative splicing is thought to be governed through a combination of cis-regulatory elements and the trans-acting proteins with which they bind. In the alternative splicing process, the spliceosomal machinery can for example skip the 3'- cutting margin of a particular intron and instead make an incision further downstream, at sites within an exon or at the 3'- margin of an adjacent intron, thus excluding part or all of an intervening exon(s). This is followed by the usual joining step and results in a different mature mRNA (Fig. 5) that will go on to encode a different protein isoform (reviewed in Schutt and Nothiger 2000). The mechanism of alternative splicing is highly important for the regulation of gene expression, as approximately 60% of all human genes are estimated to generate multiple mRNAs by alternative splicing (Lander et al. 2001), and the estimate is similar for *Drosophila* genes (K. White, Yale University personal communication).
Figure 5. Alternative splicing generates multiple mature mRNA transcripts from a single pre-mRNA molecule. All the possibilities are not shown.

(A) SXL regulation of alternative splicing: a model Drosophila system

The somatic sex determination pathway in Drosophila melanogaster (reviewed in Penalva and Sanchez 2003) is the best understood system of alternative splicing regulation, and may provide valuable insights into mechanisms for RNP-4F splice assembly factor control. The master regulatory gene of this pathway is sex lethal (sxl), which encodes the RNA binding protein SXL. A small amount of SXL is synthesized early in development of females by the early promoter (P_E) (Fig. 6A). This promoter is activated in response to the ratio of X chromosome number to autosome (A) chromosome number. The 2X:2A ratio in female flies activates this early sxl promoter whereas the X(Y):2A ratio found in males does not (Fig. 6B). Functional SXL protein is produced exclusively in female flies and subsequently acts to repress splicing patterns in genes that would otherwise lead to male development. SXL protein regulates splicing in the transformer (tra) and male specific lethal-2 (msl2) pre-mRNAs, as well as auto-regulating splicing of its own transcripts. SXL interaction with these transcripts leads to exclusive production of SXL and TRA protein in females and MSL-2 protein in males.
The *tra* pre-mRNA contains a stop codon in the 5'-region of exon 2. In males, conventional splicing through binding of U2AF spliceosomal splicing factor to the 3'-pyrimidine-rich tract and U1-snRNP to the 5'-splice site of intron 1 results in intron excision and retention of the stop codon in adjacent exon 2 (Fig. 7A). The inclusion of this stop codon results in formation of a truncated, non-functional TRA protein. In females, *tra* pre-mRNA is alternatively spliced to produce a functional TRA protein. SXL protein binds preferentially to the poly (U) sequence within the pyrimidine-rich tract near the 3'-splice site of intron 1, blocking interaction of the U2AF spliceosomal factor with the 3'-pyrimidine-rich tract. This results in use of an alternative 3'-splice site located within exon 2. The use of this alternative splice site results in loss of intron 1 and the 5' portion of exon 2 containing the stop codon. This leads to production of a functional TRA protein in females that has regulatory roles in other parts of female sex determination (Fig. 7A). *tra* pre-mRNA, spliced through SXL, encodes a splicing regulatory factor that subsequently directs splicing of the *fruitless* and *doublesex* pre-mRNAs. SXL protein regulation of the *tra* gene has been found to require at least two other Drosophila genes: *virilizer* and *female-specific lethal-2D*. Therefore, there is probably more to this mechanism than is presented here and/or currently known.

The second major gene regulated by the SXL protein is *msl-2*. The *msl-2* gene regulates X chromosome dosage compensation in male flies. Therefore, production of this protein is necessary for male development and deleterious to females. In female flies, the SXL protein binds to the poly (U) tracts adjacent to the 5'- and 3'-splice sites within intron 1 located in the 5'-untranslated region (UTR). This binding results in blockage of spliceosome U2AF factor association with the intronic pyrimidine-rich tract. The binding of U1-snRNP and a regulator protein TIA-1 to the 5'-intron splice site is also inhibited by SXL protein. As a result, intron 1 is retained and transported to the cytoplasm (Fig. 7B). How this intron-containing transcript exits the nucleus is not well understood. One could speculate that the nuclear surveillance machinery does not recognize the bound SXL protein as a spliceosomal derivative and has no way of detecting the unspliced intron. Although the transcript is successfully transported from the nucleus, it is not translated once it reaches the cytoplasm. It is thought that the
bound SXL protein inhibits translation by interfering with the association of the 40S ribosomal subunit and msl-2 mRNA (Gebauer et al. 2003).

In male flies, the absence of SXL leads to constitutive splicing of msl-2 pre-mRNA. The U2AF spliceosomal factor binds to the intronic 3'-pyrimidine-rich tract and the U1-snRNP and TIA-I regulatory factor binds the 5'-splice site. After removal of all introns, the transcript is transported to the cytoplasm and translated into functional MSL-2 protein (Fig. 7B).

In addition to regulating tra and msl-2 transcript splicing, SXL auto-regulates splicing of its own transcripts to maintain the female splicing phenotype. As mentioned previously, the early promoter of the sxl gene is activated in female embryos and produces a small amount of SXL protein. Later in development the sxl gene is transcribed in both males and females by means of a late promoter (PL). This late promoter produces transcripts with a stop codon in exon 3, which causes formation of a truncated SXL protein if not excised from the sxl transcript. SXL protein produced early in female development regulates the splicing of sxl exon 3 in females to produce more functional SXL protein (Fig. 6A). However, in males the SXL protein is not present and exon 3 with the stop codon is included in male sxl transcripts, leading to production of non-functional SXL protein (Fig 6B). In females the SXL protein binds to multiple sites flanking exon 3. SXL in conjunction with sans fille protein, SPF45 protein, and Drosophila homologues of U1A and U2B, splices exon 3 out of the female sxl transcripts through a currently unknown mechanism.

(B) Splicing enhancers and silencers

There is a growing body of evidence suggesting that many alternative splice sites are flanked by both splicing enhancers and silencers (Smith et al. 2000). Dynamic competition between the trans-acting factors that bind these cis-regulatory elements is thought to determine the alternative splicing pattern observed in different cell types and developmental stages. There is, essentially, no unregulated state for alternatively spliced transcripts, but rather a balance of positive and negative factors present in a given cell type or developmental stage which determines the ratio of alternative splice forms (Del Gatto et al. 1995).
The best-studied trans-acting factors that positively influence splice site selection through cis-regulatory elements are found in the serine/arginine-rich (SR) protein family, which includes the following: SRp20, SRp30c, SRp40, SRp55, SRp70, 9G8, ASF/SF2 and SC35 (Ladd and Cooper 2002, Black 2003). Each of these SR proteins has been implicated in splicing regulation, although the ASF/SF2- and SC35-SR proteins are among the best studied. SR proteins contain one or two RNA recognition motifs (RRM) as well as an RS domain, which contains a number of arginine/serine dipeptide repeats. SR proteins are believed to interact with pre-mRNAs through their RNP binding domain(s), and phosphorylation/ dephosphorylation of their RS domain is believed to regulate the ability of the SR protein to modulate splicing efficiency (Prasad et al. 1999). Other non-SR proteins, such as U2AF and the Drosophila sex determination pathway-splicing regulator, TRA2, which have been implicated in modulating splice site selection, also contain this RS domain (Barnard et al. 2002). Most experimental data show that SR proteins increase splicing efficiency by binding cis-splicing enhancers in pre-mRNAs; however, they have also been shown on occasion to bind specific cis-splicing silencers (Kanopka et al. 1996).

The best-characterized trans-acting factors that bind splicing silencers and inhibit splicing are the heterogeneous nuclear ribonucleoproteins (hnRNPs) (Krecic and Swanson 1999). The best studied of the hnRNPs is hnRNP A1, which has been implicated in numerous alternative splicing events, and several other processes (LaBranche et al. 1998). hnRNP A1 contains two RNA recognition motifs (RRMs) and a glycine-rich auxiliary domain. The RNA binding domains are believed to allow binding of multiple splicing silencer elements and loop out the RNA between the elements so that it is unavailable to the splicing machinery and is retained in the mature mRNA. Several mechanisms for hnRNP A1 splicing inhibition explaining how this occurs have been proposed. Some studies suggest that A1 interferes with spliceosome assembly by blocking attachment of certain spliceosome components to pre-mRNAs (Tange et al. 2001). Other studies suggest that hnRNP A1 blocks splicing by binding splicing enhancer elements, inhibiting SR protein function (Zhu et al. 2001).

hnRNP A1 was first implicated as a possible splicing modulator in an assay that showed A1 counteracted SR protein splicing efficiency (Mayeda et al. 1993). Many
subsequent studies have shown that the relative amount of A1 and SR proteins varies between cell type and developmental stage (Caceres et al. 1994, Hanamura et al. 1998). This would suggest that alternatively spliced introns, flanked by both splicing enhancers and silencers, are spliced according to the relative availability of positive or negative trans-acting factors. Thus, the interplay between cis-regulatory elements and the available trans-acting factors may ultimately determine transcript-splicing efficiency.

**Translation regulation by the 5’-untranslated region (UTR) in mRNAs**

The 5’-untranslated region (UTR) of many transcripts is used to regulate translation. Translational regulation by 5’-UTRs can occur via the following: presence of upstream open reading frame(s) (uORF), alternative initiation sites, secondary structure, and RNA-protein interactions (reviewed in Gray and Wickens 1998). Many of these regulatory factors are present in larger 5’-UTRs of lengths greater than 100-nt. The degree of regulation imparted by uORFs depends on the efficiency of the uORF start site and the proximity of the uORF stop site to the ORF start site. uORFs contain initiation sites that are encountered by the scanning ribosome upstream of the ORF initiation site. Therefore, the ribosome must re-initiate at the ORF start codon after terminating in the UTR in order to translate the transcript. If the distance between the stop codon of the uORF and the start codon for the ORF is small, translation efficiency is greatly decreased due to insufficient space and/or time for the initiation complex to reassemble (Vilela and McCarthy 2003). In addition, several mRNAs contain start sites within the 5’-UTR that do not terminate prior to the ORF start site (Descombes and Schibler 1991). Several mRNAs have been found to produce functional proteins from both the 5’-UTR and ORF start sites. The significance of these alternative proteins has not been extensively studied; however, they are expected to have a role in translation regulation.

Secondary structure in the 5’-UTR has been largely shown to inhibit translational efficiency of mRNAs. The stability and placement of the secondary structure within the 5’-UTR determine its inhibitory effect. Moderately stable, cap-proximal secondary structures can inhibit access of the ribosome 43S pre-initiation complex (Migone et al. 2002). However, when a moderately stable secondary structure is not close to the cap,
it usually is not capable of impeding access of the 43S complex, so translation proceeds. A more stable secondary structure of magnitude -50/-60 kcal/mol nearly always inhibits translation, irrespective of the cap. This was discovered through RNase protection assays, in which the 43S pre-initiation complex was repeatedly found 5’ of the stable secondary structure in the 5’-UTR (Kozak 1989a).

Secondary structure introduced in the 5’-UTR has also been shown to enhance translation by increasing the use of inefficient initiation sites (Kozak et al. 1990). The effect of enhanced translation requires that the secondary structure is placed 14-nt downstream from the initiation codon. The structure in this orientation seems to pause the 43S pre-initiation complex near the start codon and allow more time for its recognition. The complete 80S ribosomal complex that assembles there is not impeded by the structure, and translation proceeds (Kozak 1990).

The 5’-UTR can also regulate translation through the binding of regulatory proteins (Wilkie et al. 2003). Much of the knowledge we have regarding this type of regulation comes from the iron regulatory protein (IRP)-mediated regulation system. IRPs control several mRNAs that contain a secondary stem-loop structure known as the iron response element (IRE). In most mRNAs the IREs are located proximal to the cap complex and experimental evidence suggests that this position is important for their function. A cap proximal IRE bound by the IRP has been shown to inhibit binding of the 40S ribosomal subunit to pre-mRNAs. When the IRE-IRP complex is not located close to the cap, the 40S subunit association with pre-mRNA is not blocked, but has been found to inhibit the subsequent scanning by the subunit. Experimental results have shown that RNA-protein IRE-IRP-like complexes placed in the 5’-UTRs of pre-mRNAs, which do not normally modulate translational control, show IRE-IRP-like regulation (Gray and Wickens 1998). This suggests that IRE-IRP complex regulation of translation could provide a model for the function of other 5’-UTR repressor proteins and the pre-mRNAs they act upon.

**Surveillance checkpoints for mRNA nuclear exit**

Processed transcripts occasionally retain introns. Intron-containing transcripts are usually retained in the nucleus through association of spliceosome components and
the perinuclear export protein Mlp1 (Galy et al. 2004). Therefore, presence of retained introns presents a nuclear transport challenge, where by a nuclear checkpoint must be breached. Several non-spliced retrovirus transcripts overcome this challenge through use of factors encoded by the host (Pasquinelli et al. 1997). However, such elaborate mechanisms may not be utilized for transport of eukaryotic intron-containing mRNAs. For example, consider scenarios in which the spliceosome machinery is blocked from associating with an intron by trans-acting factors that bind silencing elements flanking the 5'- and 3'-splice sites or by steric hindrance of RNA secondary structures near splice sites. In both scenarios, the spliceosome components would not be present on this intron’s 5’- and 3’-splice sites for Mlp1 to bind and sequester the transcript in the nucleus. The intron-containing transcript would thus be efficiently transported to the cytoplasm for translation.

Intron retention and/or alternative splicing events can result in formation of a premature termination codon (PTC) in the resulting transcript. During the splicing process, a group of proteins called the exon junction complex (EJC) is deposited 20-25-nt upstream of each newly ligated exon-exon junction (Singh and Lykke-Andersen 2003). Important EJC proteins include Y14 and RNPS1, which recruit nonsense mediated decay (NMD) effector protein Upf3, followed by Upf2, to the EJC. The first round of translation removes the EJCs. If the processed transcript has a premature termination codon (PTC) located more than 50-55-nt upstream of the last exon-exon junction, the EJC downstream of the PTC is not removed from the transcript and the ribosome terminates translation prematurely. In this situation, Upf1 is then believed to be recruited by Upf2 or some other translation release factor, to bridge the terminated ribosome with the downstream EJC, forming an activated NMD complex that induces rapid decay of the mRNA. NMD ensures that only properly processed mRNAs capable of encoding functional proteins are translated (reviewed in Baker and Parker 2004).

**Specific goals of this Thesis research**

In the present study, our initial aim was to systematically identify all examples of alternative or facultative intron splicing of *rnp-4f* during *Drosophila* development. In addition to identifying mRNA isoforms, a further aim of this study was to reveal potential
protein isoforms, if any, encoded by the transcripts. This was worth doing, insofar as alternative functions for a spliceosomal assembly factor such as RNP-4F may include roles in regulation of the splicing process. We were also interested in identifying potential cis-regulatory elements, which could be important in understanding why facultative introns are sometimes spliced and sometimes not spliced. In the event that examples of alternative splicing were found, we were interested in identifying potential cis-acting elements that may be important here. Finally, we wanted to gain some insights into potential molecular regulatory elements, which may play a role in controlling the efficiency of rnp-4f mRNA translation in diverse tissues during fly development. The regulatory mechanisms behind alternative splicing are largely unknown for most genes. The best understood example comes from studies of the Drosophila somatic sex determination pathway (reviewed above), where many of the key regulatory players and how they interact are known.
Materials and Methods

**Drosophila stocks and propagation**

Wild-type *Drosophila* strain Oregon-R, originally obtained from the National *Drosophila* stock center, was used for this study. Flies were cultured on standard media consisting of cornmeal, agar, dark corn syrup, Brewer's yeast, propionic acid and mold inhibitor. Various early developmental stages were isolated using 60 mm diameter Petri dishes containing an egg-laying media consisting of dextrose, sucrose, apple juice, propionic acid, phosphoric acid and agar. The transparent apple juice fly media plates facilitated fly embryo collections at appropriate time increments. Later developmental stages (third instar, pupa, adult) were isolated using organisms grown on cornmeal and molasses media in half-pint bottles. All collected materials were stored at -70°C for future RNA isolation. The *Drosophila* life cycle is reviewed in Fig. 8.

**Total cell RNA isolation**

Total cell RNAs were isolated from developmental stages, usually in either four or six hour (h) developmental increments, as described by Nakagawa and Cohen (1967), with modifications including DNase I treatment (Koller et al. 1987). The quality of the RNA collected was verified via 1% agarose gel electrophoresis, using a brine shrimp RNA standard, and using intactness of rRNA bands as the criterion. Several developmental stages of total cell RNA had been isolated in previous endeavors in the lab and were also utilized in this study. Concentrations of RNAs were determined via absorbance (A_260nm).

**Poly(A⁺) RNA isolation**

*rnp-4f* transcripts are at very low abundance at some developmental stages (Peters et al. 2003), so that reverse transcription-PCR (RT-PCR) of total cell RNAs sometimes yielded little or no product. To solve this problem, mRNAs were separated from the total RNA isolated and utilized for subsequent RT-PCRs. Ten ml oligo-dT cellulose affinity columns (0.4 X 4 cm poly-prep columns, Bio-Rad) were prepared according to Chabat (1994). Approximately 3 mg total cell RNA was processed through each
column, containing 0.3 ml dry oligo-dT cellulose type 7 powder (Pharmacia). Post resuspension of column resin was in 10 ml RNA binding buffer (10 mM Tris-HCL pH 7.5, 400 mM LiCl, 1 mM disodium EDTA, 0.5% SDS), the column was washed with 10 ml of 0.1 M NaOH, and column neutralization was with 20 ml RNA binding buffer. RNAs were dissolved in 9 ml of RNA elution buffer (10 mM Tris-HCL pH 7.5, 1 mM disodium EDTA, 0.05% SDS) and denatured (10 min at 65°C), 0.9 ml 5 M LiCl was added and RNA was then loaded onto the columns. The mixture was run through the column four times. The column contents were then washed with 5 ml RNA binding buffer. The bound poly (A+) RNA was subsequently eluted with 1.5 ml RNA elution buffer. Poly (A+) RNA was precipitated by adding 75 ul 5 M NaCl and two volumes 100% EtOH. RNA yields were determined via A260nm absorption. Yields were typically about 2% of the input RNA. Poly (A+) RNAs were dissolved in appropriate volumes of diethylpyrocarbonate (DEPC)-treated water and stored at -70°C.

Reverse transcription (RT)
Reverse transcription of poly( A+) RNAs was performed at 42°C (Appendix I/II) using random hexamer primers and 400 U of M-MLV reverse transcriptase enzyme (Promega) to create complementary DNA (cDNA) strands (Kawasaki 1990). Reaction volumes were 60 ul and contained 1ug of RNA template. The quality of the RT products was then tested using rp49-specific primers (see below). The rp49 gene is constitutively expressed at comparable levels in each fly developmental stage (O’Connell and Rosbash 1984).

Primers utilized
Primers for RT-PCR analysis of alternative splicing were designed to bracket specific previously recognized intron regions (Petschek et al. 1997; FlyBase gene EST annotation at: http://flybase.bio.indiana.edu/annot/fbannquery.hform) of the rnp-4f gene (Fig. 9). Gradient-optimized PCR annealing temperatures were determined for each primer set and are indicated below.
(1) *rnp-4f* primer sets for alternative splicing detection

A, fwd (5' intron 0): 5'-[ATTCGCATATTATTCACACT]-3' (47°C)
A, rev (3' intron 1): 5'-[GATCAGATCATACTCGTC]-3' (47°C)

B, fwd (5' intron 1): 5'-[ATGGACGCAGACAACAG]-3' (60°C)
B, rev (3' intron 3): 5'-[TGACAACGTATTCGACAG]-3' (60°C)

C, fwd (5' intron 4): 5'-[CCGTGAGATTTTGCCGCC]-3' (56°C)
C, rev (3' intron 8): 5'-[GGTGGGTTTGAAATGGC]-3' (56°C)

D, fwd (5' intron 4): 5'-[CCGTGAGATTTTGCCGCC]-3' (54°C)
D, rev (3' intron 5): 5'-[GCTGCTGCGCGCCTT]-3' (54°C)

E, fwd (5' intron 6): 5'-[AGGAATCGAATTTTTAAGT]-3' (50°C)
E, rev (3' intron 8): 5'-[GAAATGGCCACAGAGAT]-3' (50°C)

F, fwd (5' intron HAC): 5'-[ATCTCTGTGGCCATTTC]-3' (50°C)
F, rev (3' intron HAC): 5'-[CACAGCAGCTGGCAGTC]-3' (50°C)

(2) *rnp-4f* primer sets for intron 0 characterization

A, fwd (5' intron 0): 5'-[ATTCGCATATTATTCACACT]-3' (49°C)
Int-0R-2, rev: 5'-[TGCAATTTTTCCCCCCC]-3' (49°C)

Alt-0-L, fwd: 5'-[AGTGCTGGAAAGTTTTTCC]-3' (53°C)
Alt-0-R, rev: 5'-[TGCAATTTTTCCCCC]-3' (53°C)

(3) *rp49* primer set

L, fwd (5' to intron): 5'-[CCAAGGACTTTCATCCGACC]-3' (57°C)
R, rev (3' to intron): 5'-[GCGGGTGCGCTTTGCATCC]-3' (57°C)
Figure 9. Experimental design for detection of alternatively spliced transcripts. The *rnp-4f* gene is diagrammed showing all known or suspected intron positions 0-VIII and HAC1, (upper line) and numbered predicted exons 1-11 (second line). Coding regions are shown in black and non-coding in white. Six location-specific primer pairs (A-F) were utilized during RT-PCR analysis of alternative splicing.

**Polymerase chain reaction (PCR)**

PCR amplifications (Mullis and Faloona, 1987) were carried out in 25-50 ul reaction volumes containing 10 mM Tris-HCL pH 8.3, 50 mM KCL, 1.5 mM MgCl₂, 0.05% NP-40, 200 uM each dNTP, 1 uM each primer, 1-1.5 ul cDNA product from RT reaction (see above), and 0.5-1 ul (2.5-5 U) Taq polymerase (Promega). Parameters used for the thermocycler are as follows: 2 min initial denaturation at 97°C, followed by 30-35 cycles of denaturation at 95°C for 30 sec, 1 min annealing at gradient-optimized temperature (see preceding sub-section), and 1 min extension at 72°C.

**Determination of rnp-4f mRNA isoform relative abundance**

Following electrophoresis, 4% agarose gels were submerged in a mixture containing 10 ul SYBR Green I (Molecular Probes) in 100 ml of 1 X agarose gel buffer (40 mM Tris-HCL pH 7.9, 5 mM sodium acetate, 1 mM disodium EDTA) for approximately 1 h with continuous agitation. The gel was then scanned in the Molecular Dynamics Storm 860 phosphorimager using the “fluorescence” mode, and band intensities were quantified using version 5.2 of the ImageQuant software program.
**Agarose gel electrophoresis**
RT-PCR samples were run in 100 ml, approximately 15 cm X 20 cm, 2% or 4% agarose gels according to Southern (1975). Agarose was dissolved in 1X buffer (40 mM Tris-HCL pH 7.9, 5 mM NaAc, 1 mM disodium EDTA) and buffer reservoirs contained 200 ml of 1X buffer. Wells 1.5-3 mm diameter were loaded with 8-25 ul of sample and run at 200 V for 1.5-2 h. pBR-322 DNA cut with Alu I was used as a size standard and 50% glycerol-XCFF or 50% glycerol-BPB dye solution was incorporated into each PCR sample. Band quality was then assessed on the “EagleEye” (Stratagene TM) UV imager.

**Southern blot analysis**
DNA fragments were transferred from agarose gels to nitrocellulose or nylon membranes, by capillary blotting. DNAs were immobilized by baking filters for 2 h at 80°C (nitrocellulose) or UV-crosslinking (nylon). Hybridizations to verify *mp-4f* gene fragments following RT-PCR were carried out using probes prepared from two different widely spaced DNA oligos that are internal to the PCR primer positions. Probes were 5’-end-labeled using γ-ATP32 incorporation, which is catalyzed by T4-polynucleotide kinase (Sambrook et al. 1989). Probe-positive bands were visualized using the phosphorimager.

**Northern blot analysis**
Northern analyses of *mp-4f* mRNAs were carried out in collaboration with Stephen Strock, a Miami University U.S.S. research intern. Heat-denatured poly (A+) RNA samples estimated by A260 nm to contain 5 ug were loaded into 1.2% denaturing formaldehyde gels 7 mm thick prepared in 1 X MOPS buffer (20 mM MOPS pH 7.0, 5 mM sodium acetate, 1 mM EDTA), and run for 1,275 V-h at 4°C (Ausubel et al. 1994-1997). Flanking lanes were loaded with bovine ribosomal RNA sizing standard (Sigma), containing RNAs of about 2,000 and 5,300-nt. In addition, *D. melanogaster* rRNAs were utilized as size standards, which produced bands of about 1,790-, 2000-, and 2,110-nt. After electrophoresis, RNAs were visualized by staining for 1 h with 10 ul of
1:10,000 diluted SYBR Green II (Molecular Probes) in 100 ml 1 X MOPS. Sizer RNAs were then visualized using a Molecular Dynamics Storm 860 phosphorimager using the blue fluorescence mode, and images stored in the computer for visualization.

RNA transfer to nylon membranes was carried out in 10 X SSPE (100 mM sodium phosphate buffer pH 7.4, 3 M NaCl, 20 mM EDTA), overnight (Sambrook et al. 1989). Membranes were then dried at room temperature, baked 20 min at 80°C, and UV-crosslinked to immobilize RNA bands.

Hybridization probe for \textit{rnp-4f} was prepared by template excision with \textit{Xho I} restriction enzyme from a 3.5 kb genomic DNA fragment previously cloned in pBluescript plasmid (Hess et al. 1996). This fragment contains the downstream half of the \textit{rnp-4f} gene and also the adjacent \textit{sas-10} gene, so that hybridizations would be expected to detect transcripts arising from both genes. \textit{rnp-4f} specific probes were also prepared following template excision with \textit{Eco RI/Hind III} of the 2,650-bp 0-4 h embryo cDNA clone E-17.2, available in our lab. \textit{rnp-4f} cDNA fragments were also obtained via RT-PCR and labeled to detect this constitutively-expressed gene’s transcripts, as a control.

Probes were labeled with \textsuperscript{32}P-dCTP using the random priming method employing a commercially available “Prime-a- Gene” kit (Promega), following the manufacturer’s directions. Typically, 50 ng of template DNA was labeled with 100-200 uCi \textit{\textsuperscript{\alpha}{-}32P-dCTP} of 5,000 Ci/m mole specific activity (MP Biomedical) in a 50 ul reaction volume to a probe specific activity of about 2 X 10\textsuperscript{9} cpm/ug.

**Membrane hybridizations**

(A) End-labeled probes

End-labeled oligonucleotide probes were prepared by labeling 1 ug DNA with 100 uCi of 4,500 Ci/m mole specific activity (MP Biomedical) \textit{\textsuperscript{\gamma}{-}32P-ATP in a 25 ul volume to a probe specific activity of about 2 X 10\textsuperscript{7} cpm/ug.}

Filters were subjected to initial pre-hybridization for 2 h at 37°C in 150 ml of solution prepared by mixing: 50 ml of “Aravin mix” (200 mM sodium phosphate buffer pH 7.4, 1 M NaCl, 50 mM EDTA), 3 ml 50 X Denhardt’s solution, 42 ml 25% SDS, 55 ml H\textsubscript{2}O, with continuous agitation. Filters were then placed in Seal-a-Meal bags for the
second pre-hybridization for 2 h at 37°C in 21 ml of solution prepared by mixing 6.7 ml “Aravin mix”, 400 ul 50 X Denhardt’s solution, 5.6 ml 25% SDS, 8 ml 100% deionized formamide and 4,000 ug tRNA. After the second pre-hybridization, the bag was opened and about 2.5 X 10^6 cpm rnp-4f specific probe was added. Hybridization was carried out overnight at 37°C.

Filters were twice washed at room temperature for 5 min each, in a solution prepared by mixing 20 ml 20 X SSC, 180 ml H₂O, with agitation. Filters were then twice washed at 42°C for 15 min each, in a pre-warmed solution prepared by mixing: 20 ml 20 X SSC, 4 ml 25% SDS, 176 ml H₂O, with agitation.

Filters were exposed to intensifying screens and then scanned in the Molecular Dynamics Storm 860 phosphorimager using the “storage phosphor” mode, and band intensities were quantified using version 5.2 of the ImageQuant software program.

(B) Random-primed probes
Filters were subjected to initial prehybridization for 2 h at 40°C in 150 ml of a solution prepared by mixing: 100 ml of “Aravin mix” (200 mM sodium phosphate buffer pH 7.4, 1 M NaCl, 50 mM EDTA), 15 ml 50 X Denhardt’s solution, 6 ml 25% SDS, 29 ml H₂O, with continuous agitation. Filters were then placed in Seal-a-Meal bags for the second prehybridization for 2 h at 40°C in 20 ml of hybridization solution (Aravin et al. 2001): 100 mM sodium phosphate buffer pH 7.4, 0.5 M NaCl, 25 mM EDTA, 1 X Denhardt’s solution, 0.1% SDS, 25% (v/v) deionized formamide. After the second prehybridization, the bag was opened and about 7 X 10^7 cpm rnp-4f specific probe plus about 5 X 10^6 cpm rp49 specific probe was added, following probe denaturation for 3 min at 100°C. Hybridization was carried out overnight at 40°C.

Filters were twice washed at room temperature for 15 min each, in a solution prepared by using 10 ml 20 X SSC, 2 ml 25% SDS, 188 ml H₂O, with agitation. Filters were then twice washed at room temperature for 15 min each, in a solution prepared by mixing: 2.5 ml 20 X SSC, 400 ul 25% SDS, 197 ml H₂O. Finally, filters were washed at 50°C for 30 min in a solution prepared by mixing: 1 ml 20 X SSCC, 800 ul 25% SDS, 198 ml H₂O.
Filters were exposed to intensifying screens and then scanned in the Molecular Dynamics Storm 860 phosphorimager using the “storage phosphor” mode, and band intensities were quantified using version 5.2 of the ImageQuant software program.

**DNA fragment isolation**

Low-melt agarose gels, 2-4%, were utilized to recover DNA bands produced via RT-PCR. Excised gel bands were combined with Agar-ACE enzyme (Promega), which was used according to the manufacturer’s directions to digest agar, and DNAs were ethanol-precipitated. Alternatively, excised gel slices were dissolved in QG buffer (Qiagen) and loaded onto a Qiaquick spin column according to the manufacturer’s directions. To bind DNA to the column and remove excess buffer, the column was microcentrifuged at 13,000 rpm for 1 min. Column-bound DNA was then washed by loading 0.75 ml PE buffer onto the column, followed by centrifugation at 13,000 rpm for 1 min. The DNA was eluted from the column with 30 ul double-distilled water (Appendix II).

**Fragment ligation/transformation/colony selection**

Gel-isolated PCR-generated DNA fragments were ligated into pGEM-T Easy plasmids (Promega) by incubation at 14°C overnight in the presence of DNA ligase, following the manufacturer’s directions. Heat shock method (N. Pace, personal communication) using Rb+/Mn++ treated competent DH5αF’ E. coli cells was used to transform recombinant plasmids into bacterial cells (Sambrook et al. 1989). These cells were plated onto ampicillin-drugged LB nutrient agar coated with 100 µl X-Gal (20 mg/ml stock solution). Colonies with the plasmid insert are white in color, while colonies without the insert appear blue. The plasmid insert with the DNA fragment disrupts the ability of the Lac-Z gene to produce B-galatosidase. B-galatosidase digests X-Gal to produce a blue pigmentation in cells where the Lac-Z gene is intact.

**Small scale plasmid isolation**

White colonies assumed to have the desired DNA fragment plasmid insert were picked off plates and grown in LB growth medium at 37°C overnight with agitation. A small scale plasmid isolation protocol (Davis et al. 1980) was utilized to isolate the plasmids.
After cell lysis and DNA precipitation, preparations were digested with 10 ug RNase A at 37°C for 1 h, extracted once with chloroform/isoamyl alcohol (24:1, v/v) and EtOH-precipitated. Plasmids were dissolved in 1 X EcoRI buffer and digested with 12-24 U EcoRI enzyme (Promega), cutting sites on either side of the DNA fragment insert. After cutting, fragments were run on a 2% agarose gel with pBR 322 plasmid DNA cut with Alu I size standard to determine if they were the predicted size. Successful transformations show one band that is the predicted size of the insert and another that is the size of the vector.

**DNA sequencing**

cDNA clones obtained from plasmid isolation or column purification were sequenced using either the ABI-3100 instrument or the ABI Prism 310 genetic analyzer located in the Miami University DNA Core Facility. Both sequencing machines utilize the dideoxynucleotide triphosphate (ddNTP) method (Sanger et al. 1977). PCR sequencing reactions were carried out following directions of the supplier (Amersham-Pharmacia), and contained ~0.1 ug cDNA template, 1 ul of 1uM forward or reverse primer, 2 ul ET sequencing mix, 6 ul 2.5 X sequencing buffer, and water to bring the total reaction volume to 20 ul. Parameters used for the thermocycler are as follows: 25 cycles of denaturation at 95 for 20 sec, 15 sec annealing at 50°C, and 1 min extension at 60°C. PCR products were then precipitated with ETOH, washed and stored at -20°C according to the supplier’s instructions. Products were re-suspended in 11 - 20 ul of megabase loading solution prior to sequencing.

**Computer-assisted sequence analysis**

Several software programs were utilized for various analyses of DNA and RNA sequences. Data resulting from the automated sequencing machines were interpreted using EditView software, and manually verified. Genomic DNA comparisons were performed using BLAST [http://www.ncbi.nlm.nih.gov/BLAST/]. Predictions for most probable RNA foldings were obtained using version 3.1 of the mfold program [http://www.bioinfo.rpi.edu/applications/mfold; Zuker (2003)].
Results

**Qualitative analysis of rnp-4f mRNA isoforms across development**

*rnp-4f* full-length transcripts are in excess of 3,000-nt, but contain introns ranging in size from about 60- to 80-nt (Petschek et al. 1997). Detection of subtle length differences in mRNA isoforms due to alternative splicing would not be possible using Northern blot analysis, owing to limitations in size resolution. To solve this problem, a series of short cDNA fragments were studied following amplification using a collection of RT-PCR primer sets. Detection of *rnp-4f* mRNAs across development (Fig. 10) was performed using six different primer sets (A-F). Primers were designed to bracket specific regions of the *rnp-4f* pre-mRNAs where alternative splicing was predicted to potentially occur, based on known intron positions. Primer set A brackets 5' of intron 0 through 3' of intron I, primer set B brackets 5' of intron I through 3' of intron III, primer set C brackets 5' of intron IV through 3' of intron VIII, primer set D brackets 5' of intron IV through 3' of intron V, primer set E brackets 5' of intron VI through 3' of intron VIII, and primer set F brackets the previously observed putative HAC1-like intron just downstream of intron VIII (Vaughn et al. 2001).

RT-PCR sometimes results in a mixture of cDNA product bands following gel electrophoresis, in which an unknown subset of product does not contain the expected gene sequences. To resolve this uncertainty, Southern blots were prepared to further characterize each band by hybridization using gene-specific probes. After electrophoresis of RT-PCR products from each primer set (B-F), cDNAs were transferred to nitrocellulose or nylon membranes. Membranes were then hybridized with two widely spaced end-labeled *rnp-4f* DNA oligos internal to the PCR primer positions, the probes being specific to each primer set RT-PCR product.

Gene-specific hybridizations can illuminate potential products of splicing or alternative splicing, but cannot reveal the exact sites of intron excision events. Sequence analysis was therefore carried out on all probe-positive bands detected. All *rnp-4f* RT-PCR bands detected by Southern blot hybridizations were excised from 2% or 4% low melt agarose gels. DNA gel slices were then either cloned or column-purified (QIAGen) in preparation for sequence analysis.
(A) Constitutively-spliced mRNA tracts

Several primer sets produced from one to three bands, only one of which reacted with the gene-specific probe. In those cases, the sole band to hybridize was found to be of an appropriate size for all introns within the fragment to have been spliced. RT-PCR analysis using primer set B revealed one band upon electrophoresis (Fig. 10D) of approximately 730-bp. Upon hybridization, this 730-bp fragment was found to react with the probe. This size is consistent with constitutive splicing of introns I - III, a prediction that was confirmed following sequencing of this DNA band. RT-PCR analysis using primer set E revealed only one band upon electrophoresis, approximately 270-bp in size (Fig. 10E). Hybridization showed that this fragment reacted with the probe. The observed size of this band is consistent with constitutive splicing of introns VI - VIII, which was confirmed by sequencing. RT-PCR analysis using primer set F revealed only one band upon electrophoresis, sized at approximately 380-bp (Fig. 10F). This band hybridized with the probe and is of a size expected for a genomic DNA transcript from which no intron has been removed. The putative HAC1-like intron is therefore not detected in developmental Southerns, although a very low frequency of splicing may have escaped detection.

(B) Cryptic folded mRNA tracts

RT-PCR analysis using both primer sets C and D led to some unexpected and initially confusing results. Resolution of these results demonstrated the importance of obtaining sequence verification of putative isoforms arising from alternative splicing. For primer set C, two bands were observed upon electrophoresis (Fig. 10I), a large band of approximately 730-bp and a small band of approximately 420-bp. Both bands reacted with the gene-specific probe following hybridization. RT-PCR analysis using primer set D revealed three bands upon electrophoresis (Fig. 10G), each of which hybridized with the appropriate probe. Using primer set D, a large band of approximately 335-bp, a medium size band of approximately 275-bp and a small band estimated to be 220-bp were observed.
Taken at face value, probe-positive bands arising from use of primer sets C and D were initially interpreted to reveal instances of alternative splicing. However, difficulties were encountered during the many attempts to clone some of the bands. For the two primer set C bands, the 730-bp fragment was readily cloned and sequenced, which showed that introns IV, VI, VII and VIII are spliced from the mRNA isoform. However, intron V is retained, a result previously observed for several cDNAs isolated by screening libraries (Petschek et al. 1997). The 420-bp fragment proved to be completely resistant to cloning. This problem was solved by preparation of this fragment for sequencing using Qiaquick spin columns (QIAGen). Analysis of RT-PCR products arising from primer set D gave a similar result in that the large 335-bp cDNA band was readily cloned and sequenced, showing constitutive splicing of intron IV but intron V retention. The 275-bp cDNA band was also readily cloned and sequenced revealing that both introns IV and V are spliced from this isoform. However, the 220-bp fragment was completely resistant to cloning, a problem again solved by using Qiaquick columns. Intron V is an example of a facultatively spliced intron, in that it is spliced from some mRNA isoforms but not from others.

Analysis of the small, predicted “420-bp” unclonable transcript fragment arising from primer set C surprisingly revealed that it is identical in sequence to the constitutively spliced form of *rnp-4f*. Given the appreciable difference in apparent band size, as estimated by agarose gel electrophoresis, it was hypothesized that a small fraction of the template fails to remain denatured for primer annealing/extension during the PCR step of RT-PCR (Fig. 11). This fraction could then fold back upon itself, forming a compact single-stranded molecule that would migrate fast for its size. Such a behavior could be encouraged in a fragment having a relatively high % GC content. To test this hypothesis, the computer assisted RNA folding program mFOLD (Zucker 2003) was employed. The results of this analysis (Fig. 12) show that a highly stable secondary structure would be predicted to form for this fragment. Furthermore, this folded cDNA tract has a 56% GC content, appreciably higher than the ~38% typical for most eukaryotic genes. Therefore, it is predicted that the small band running at approximately 420-bp contains many single-stranded constitutively spliced *rnp-4f* transcripts folded in a similar secondary structure. This unexpected discovery readily
explains why the “420-bp” fragment was completely resistant to cloning into a double-
stranded DNA vector, yet capable of hybridizing to the probes employed.

Analysis of the small band arising from primer set D, which was observed to run
at approximately “220-bp” during agarose gel electrophoresis, showed that it is identical
in sequence to the constitutively spliced form of mp-4f, and is yet another example of a
single-stranded molecule running at a deceptively rapid rate during electrophoresis.
Here again, the computer-assisted RNA folding program mFOLD (Zuker 2003) predicts
that single-stranded template molecules produced from this primer set during the PCR
step of RT-PCR have the ability to fold into a very stable secondary structure (Fig. 13).
Furthermore, this cDNA tract has a 57% GC content, appreciably higher than the ~38%
typical for most eukaryotic genes. Therefore, it is predicted that the small band running
at approximately “220-bp” also contains many single-stranded constitutively spliced mp-
4f transcripts folded in a similar secondary structure.

(C) Facultatively and alternatively spliced mRNA tracts

RT-PCR analysis using primer sets A and D revealed that both facultative and
alternative splicing occur in the mp-4f gene. RT-PCR analysis using primer set D
revealed three bands upon electrophoresis (Fig. 10G), each of which hybridized with the
gene-specific probe. As previously described, the smaller “220-bp” transcript fragment
is merely an unusual cryptic variant of the largest band, in which intron IV is spliced but
intron V is retained. The medium sized 275-bp band arising from primer set D results
from splicing of both intron IV and intron V. Facultative splicing of intron V results in the
creation of an in-frame termination codon just downstream of intron V (Petschek et al.
1997). This is predicted to result in production of a severely truncated form of the RNP-
4F protein. Several of the protein’s C-terminal sequence motifs would be excluded from
this truncated form of RNP-4F: the nuclear localization signal (NLS), two RNA
recognition motifs (RRMs) and the C-terminal motif. Loss of the NLS would thus result
in a location change of this protein from the nucleus to cytoplasm.

RT-PCR analysis using primer set A revealed three bands upon electrophoresis
(Fig. 10B), a large band of approximately 400-bp, a medium size band of approximately
320-bp and a small band estimated to be 220-bp. Sequence analysis revealed that the
large 400-bp band retains intron 0 and that intron I is spliced. The medium sized 320-bp band observed in this primer set is the result of splicing both intron 0 and intron I. Insofar as intron 0 is spliced from some transcript isoforms but not from others, this is a second example (along with intron V) of a facultatively spliced intron. The small 220-bp band is a result of constitutive splicing of intron I and the alternative splicing of intron 0 using the 5’-splice site for this intron and an alternative 3’-splice site within exon 2 designated intron-0-alt, located about 100-nt further downstream.

**Conserved Putative cis-regulatory elements within intron 0 isoforms**

Total genomic sequencing has recently been completed for several fly species which are closely related to the previously sequenced *D. melanogaster*, (Adams et al. 2000). The genomic sequences for *D. pseudoobscura* and *D. yakuba* are currently incorporated into the NCBI publicly-accessible BLAST website, and therefore most convenient to use for interspecific comparisons of *rnp-4f* regions of interest. A BLAST search of these two genomic sequences against *D. melanogaster mp-4f* (Petschek et al. 1997) surprisingly did not reveal this gene in the *D. pseudoobscura* sequence. However, the gene sequence was found by first locating the *sas-10* gene via a search of GenBank. *sas-10* is adjacent to *mp-4f* in *D. melanogaster*, and was expected to be similarly located in *D. pseudoobscura*. Study of the annotated scaffold containing the *sas-10* gene in *D. pseudoobscura* (accession number CH379063) then revealed the adjacent *mp-4f* gene. The latter gene’s sequence is relatively highly divergent. Extensive sequence similarity was found between the *D. melanogaster* and *D. yakuba* *mp-4f* sequences that stopped near the predicted start codon, necessitating manual alignment upstream of this point. This analysis (Fig.14) revealed a pattern of scattered nucleotide tracts showing extensive sequence identity, which was broken up by several insertion/deletion tracts. *D. melanogaster mp-4f* EST clone #SD-17673 (GenBank accession number NM_206630) shows the furthest upstream extension, and was used to fix the approximate predicted 5’-end for the *D. yakuba mp-4f* sequence.

This study had already provided experimental evidence for the precise 5’- and 3’-splice junctions of intron 0, in addition to the intron-0-alt isoform 3’-splice junction, 100-nt further downstream in *D. melanogaster*. The alignment was utilized in an effort to
predict the presence or absence of corresponding splice junctions for *D. yakuba*. Although not yet experimentally verified, acceptable corresponding putative splice junctions do appear at similar positions for both species, using the compilation in Mount et al. (1992) to assign these junctions. It is clear that both intron 0 and the alternatively spliced version of intron 0 are present in both species, and it appears likely that both regions may be spliced in *D. yakuba*. It now becomes interesting to ask if there are conserved putative cis-regulatory elements at the splice junctions of either splice variant that may play important functional roles in both species. Aside from the obvious extensive tracts of sequence identity, two structures have been identified that are of particular interest. One is a stem-loop, arising at corresponding positions in the two sequences and spanning the 3'-splice junction of intron 0 (Fig. 15). The region of sequence identity between these elements in the two species is not extensive. The other, an evolutionarily conserved stem-loop pair, includes the 3'-splice junction of intron-0-alt (Fig. 15). Both stem-loops utilize nucleotides which occur at corresponding positions in the two species. The intron-0-alt stem-loop pair exhibits considerable sequence identity between the two species, which may be important in recognition and binding to trans-regulatory elements. In addition, the similar conformations of the two corresponding conserved stem-loops could be utilized for such a function. The validity of the upstream proposed stem-loop is supported by the powerful comparative approach (Fox and Woese 1975), in which the existence and extent of suspected helices are proven by the demonstration of compensatory base changes in the stems (Noller et al. 1981; Vaughn et al. 1995). There are two examples of compensatory base changes in the proposed upstream intron-0-alt stem-loop and at least one unambiguous example at the very top of the proposed intron 0 stem-loop (Fig. 15). The potential functions of these structural elements will be addressed in the Discussion.

**Quantitative SYBR Green analysis of *rnp-4f* primer set “A” RT-PCR products**

Following electrophoresis of RT-PCR products amplified using primer set A, agarose gels were exposed to SYBR Green I DNA fluorescent tag to quantitate relative amounts of each isoform at each stage of development. The intensity of the large (intron 0 retained) band was compared to that of the small (intron-0-alt excised) band resulting
from the same primer set. Band intensities were expressed as ratios of large-to-small bands across development. It was found (Fig. 16) that the relative abundance of the two isoforms varies reproducibly across development, showing that splicing of the intron-o-alt isoform is developmentally regulated. In particular, the isoform retaining intron-0-alt becomes relatively abundant during the embryo stages of development, and then again during 2\textsuperscript{nd} and 3\textsuperscript{rd} instar stages.

**Putative cis-regulatory elements within intron V isoforms**

An alignment between *D. melanogaster* and *D. yakuba* in the vicinity of intron V revealed considerable sequence identity, even within the intron itself (Fig. 17). It is clear from this alignment that no nucleotide deletions, which could introduce frame-shift mutations leading to generation of stop codons, exist in the *D. yakuba* intron. It is also clear that the *D. yakuba* intron sequence begins with the sequence AU… rather than the canonical GU…, which very likely prevents splicing of this intron in this species. In a computer-assisted exercise to demonstrate existence of potential stem-loops spanning one or both splice junctions in *D. melanogaster*, structures were identified which span both junctions and which could function to prevent binding of snRNPs required for intron splicing in this species.

**Northern blot analysis of rnp-4f mRNA transcription across development**

Heat-denatured poly (A\textsuperscript{+}) RNAs were loaded into 1.2\% denaturing formaldehyde gels. After electrophoresis, RNA sizers were visualized by staining with SYBR Green II, and RNAs were then transferred to nylon membranes for hybridization with a general *rnp-4f*-specific probe following UV-crosslinking of RNAs to the membrane. Two bands estimated to be about 3,300- and 3,600-nt in size, were detected upon hybridization (Fig. 18). The smaller transcript isoform (S) is relatively more abundant in very early embryos and also in adults, but then there is an apparent switch in relative isoform abundance during embryo stages 6-12 h, when the larger 3,600-nt transcripts (L) become more abundant. This qualitative result was quantified using the phosphorimager software (Fig. 18). Among the various *rnp-4f* isoforms of differing lengths observed in this study, only one resulted in alternative fragment lengths of
around 300-nt. This is the expanded intron-0-alt splice variant, resulting in an ~180-nt mRNA excision event. Given the high degree of uncertainty in sizing long molecules in Northern gels, the two mRNAs observed by the different experimental methods may very well correspond. The relative levels of rnp-4f total mRNA across development were estimated by comparing the ratio of their band intensities to that of the constitutively-expressed rp49. The results (Fig. 16) show that the total rnp-4f mRNA levels dramatically decline during the mid-embryo stage of development, and then again increase during the pupal and adult stages. This result is in excellent agreement with that obtained for this gene using microarray-based technology (Arbeitman et al. 2002).

To test the hypothesis that the transcript length switch occurring at about 6-12 h of development may correlate with the intron-0-alt splice event, the blot was stripped and re-hybridized with a probe specific for the alternatively spliced ~180-nt intron 0 expanded isoform. The results (Fig. 18) show that this probe exclusively hybridized with the larger rnp-4f mRNA band, showing that this band represents the unspliced intron 0 isoform.
Discussion

We have previously reported characterization of the *D. melanogaster rnp-4f* gene and a few cDNA clones from the standpoint of their nucleotide sequences (Petschek et al. 1996, 1997). In those studies, eight spliceosomal introns were identified and named I-VIII. Subsequently, characterization of EST clones deposited in FlyBase (http://flybase.bio.indiana.edu) and then full length sequencing of the longest cDNAs (GenBank accession numbers NM_078492 and NM_206630) revealed a ninth intron located in the 5’-UTR, which had previously escaped our notice and which we have named intron 0. The cited studies were not concerned with introns beyond their identification, and only described examples of alternative splicing that had been found by accident.

**rnp-4f isoforms identified across *Drosophila* development**

Although many introns were found to be constitutively spliced, three alternative splice sites in the *rnp-4f* gene were observed in this study, one at intron V and two at intron 0. Intron V was found to be retained in a large percentage of the transcripts detected by RT-PCR. This was expected based on previous studies that defined inclusion of intron V in the constitutive *rnp-4f* isoform (Petschek et al. 1997). Secondary structure sequestering of the 5’- and 3’-splice sites, in addition to predicted cis-regulatory splicing silencer elements, is proposed to be a major factor regulating intron retention. Introns 0 and V were found to be facultatively spliced from only a small percentage of the total transcripts detected by RT-PCR across development, perhaps as a result of formation of alternative 5’- and 3’-splice site secondary structures that do not sequester the splice sites. In the case of intron 0, the lack of splicing success may also be due to a weak 3’-splice site consensus sequence.

Alternative splicing was also observed in the 5’-UTR, suggesting a mechanism for introducing or removing regulatory elements in *Drosophila rnp-4f* that control RNP-4F expression posttranscriptionally. Intron 0 was found to often be alternatively spliced through use of an alternative 3’-splice site located near the center of the adjacent, downstream exon 2. This alternative splicing event is proposed to occur due to the
weak, sequestered 3′-splice site of intron 0. The tract of exon 2 that is removed upon utilization of this secondary 3′-splice site contains one of two evolutionarily conserved twin stem-loops that are proposed to participate in alternative splice site selection and, if retained, regulation of translational efficiency.

**Regulation of rnp-4f alternative splicing**

(A) Facultative splicing within *rnp-4f* pre-mRNAs

The examples of alternative and facultative splicing observed across development during processing of *rnp-4f* pre-mRNAs suggest that these processes are regulated to express specific quantities of the different isoforms during development and in specific locations. Intron V is approximately the same size as the other constitutively spliced introns and has highly conserved 5′- and 3′-splice site consensus sequences, so that constitutive splicing of this intron would be expected. However, intron V is not spliced from the majority of pre-mRNAs. The molecular mechanism behind the retention of this intron is not currently known. A computer-assisted RNA folding program predicts that the 5′-splice site of intron V is sequestered in the base of a stem-loop structure and is adjacent to a second stem-loop positioned one nucleotide downstream, which is located in the intron (Fig. 17). This predicted double stem-loop structure spanning the exon 6/5′-intron V splice site junction has a -8.4 kcal/mol standard free energy value. The 3′-splice site of this intron is also sequestered in a -10.1 kcal/mole stem-loop that spans the 3′-intron V/exon 7 splice junction. The predicted secondary structures around the 5′- and 3′-splice sites of intron V strongly suggest that these splice sites are unavailable to the spliceosome and as a result, this intron is rarely excised. A model to explain retention of intron V is shown in Fig. 19A.

The 5′- and 3′-splice sites of intron V may also contain or be flanked by cis-regulatory exonic splicing silencer elements or intronic silencing elements. These silencing elements could bind hnRNP A1 or hnRNP-like trans proteins to further enhance blocking of spliceosome assembly, as reported for other experimental systems (Black 2003). This is the mechanism observed in the *Drosophila* sex determination pathway splicing of *msl-2*, where intron I is retained in female flies as a result of the SXL protein binding to poly (U) tracts adjacent to the 5′- and 3′-splice sites of *msl-2* (Penalva
and Sanchez 2003). As previously described, SXL blocks binding of the spliceosomal factors U1-snRNP to the 5’-splice site and U2AF to the 3’-pyrimidine-rich tract of the msl-2 transcript.

Perhaps the silencing of intron V splicing results from a combination of these two splicing inhibitory mechanisms. The secondary structures, which contain the splice sites, may also harbor splicing silencing elements that bind trans silencing proteins. An example of such an inhibitory mechanism was reported by Dirksen et al. (2003) in myosin phosphate targeting subunit 1 (mypt1). The mypt1 alternative exon is regulated by intronic cis elements and a bipartite exonic enhancer/silencer element. The alternative exon in chicken mypt1 is skipped in fast phasic muscle tissue found in the gizzard. However, the alternative mypt1 exon is included in slow phasic muscle tissue found in the aorta. Deletion and mutational analysis of MYPT1 mini-gene constructs was performed to identify any cis-regulators of splicing and/or exon retention. Splicing of the alternative exon was found to require a cis-enhancer complex located near the 5’-splice site of the exon which includes two UCUU motifs within the U-rich putative pyrimidine-rich binding tract and a 67-nt enhancer, both located within the flanking intron, along with a novel enhancer within the exon. Silencing of the alternative exon was found to require a 13-nt ESS imperfect palindrome sequence adjacent to the ESE perfect 11-nt palindrome sequence, as well as a distant intronic silencing element. RNA secondary structure analysis of the ESE/ESS region of the alternative exon showed that the 13-nt ESS forms part of a 20-nt loop, and that the ESE and 5’-splice site of the alternative exon base pair with one another. The authors hypothesize that this stem-loop structure prevents U1-snRNP from binding to the 5’-splice site and directing subsequent splicing, unless a trans-acting factor binds to the 11-nt ESE to disrupt the stem-loop. The ESS located in the loop would function to sterically interfere with binding of trans enhancer proteins to the ESE, possibly through binding trans silencing factors.

This model has also been proposed for regulation of splicing in human tau alternative exon 10 associated with inherited dementia (Grover et al. 1999) and chicken β-tropomyosin alternative exons. Such a system of alternative inclusion or exclusion of exons could be extrapolated to the facultative use of introns 0 and V in mp-4f. Given
the propensity of this gene’s transcripts to form secondary structures, it is possible that *rnp-4f* ESS and ESE sequences are held in a configuration similar to that observed in *mypt1*, *tau* and *β-tropomyosin* pre-mRNAs, and are regulated accordingly.

An inhibitory structure at the splice sites of intron V, such as the one described by Dirksen et al. (2003) would inhibit binding of spliceosome components at the 5’- and 3’-splice sites. Without the spliceosome components attached, the nuclear surveillance machinery would not detect the intron-containing transcript. These transcripts would therefore be transported out of the nucleus and enter the cytoplasm for translation. The inclusion of intron V does not result in creation of a premature stop codon in the transcript, which is therefore efficiently translated. This would explain why and how the transcript containing intron V is detected as the most abundant *rnp-4f* isoform (Fig. 10G) and therefore considered to be a component of the constitutive isoform.

Intron V is excised in a small percentage of *rnp-4f* transcripts. The molecular regulation mechanism behind this splicing decision is also not known. It is hypothesized that in a small percentage of the transcripts the conserved secondary structures sequestering the 5’- and 3’-splice sites are not correctly folded. As described by Estes et al. (1992), alternative secondary structures may form near the splice sites that do not sequester the 5’- and 3’-splice sites. This would allow binding of the spliceosome at these sites followed by intron V excision (Fig. 19B). Another possible explanation for this phenomenon could lie in the observation that most facultatively spliced introns and exons are flanked by both splicing silencers and enhancers. A study by Del Gatto et al. (1995) reported that multiple sequence elements influenced splicing of a fibroblast growth factor receptor 2 alternative exon (K-SAM). A short splicing silencer sequence (5’-TAGGGCAGGC-3’) found within the alternative exon, in addition to two splicing enhancers located in the flanking intron, repressed and activated splicing of this exon, respectively. The exon was efficiently spliced in SVK14 cells; however, it was frequently retained in HeLa cells. The authors suggest that the SVK14 cells may express more splicing enhancer (SR-like) proteins than HeLa cells, and bind the cis-splicing enhancer elements to promote splicing of this exon. Perhaps in a small minority of *rnp-4f* transcripts splicing enhancer proteins bind splicing enhancer elements...
near the 5'- and 3'-splice sites of intron V and block splicing silencing elements, enabling the spliceosome to recognize the splice sites and excise the intron.

Splicing of *rnp-4f* intron V creates a frameshift mutation, resulting in an in-frame stop codon just downstream of intron V (Petschek et al. 1997). As a result, the nuclear localization signal (NLS) encoded by the C-terminal portion of the transcript would be omitted from the predicted RNP-4F protein. A study by Paterno et al. (2002) shows that the NLS of fibroblast growth-inducible early response gene *mi-er1* is contained within a facultative intron in the C-terminal region. Regulated use of a C-terminal facultative intron resulted in production of two distinct isoforms: hMI-ER1α (intron removed) and hMI-ER1β (intron retained), where the hMI-ER1β transcript was found to be the predominant isoform in all tissues examined except testes. Transfection assays revealed that hMI-ERα is not localized to the nucleus and remains in the cytoplasm following translation. This study shows that facultative use of an intron can regulate subcellular localization of a protein and may have vital functional implications. Although the NLS of RNP-4F is not contained within facultative intron V, the consequences of loss of either intron are nonetheless comparable. If the *rnp-4f* transcript with the premature stop codon eludes the nuclear surveillance machinery and is translated, it too would remain in the cytoplasm to possibly perform a specific function. However, the likelihood of this occurring is relatively low, as nonsense-mediated decay (NMD) would be expected to detect the premature stop codon and destroy the transcripts that would encode the protein lacking a NLS before they could be translated.

Intron 0, much like intron V, is excised in a small percentage of *rnp-4f* transcripts (Fig. 10C). This intron contains a conserved 5'-splice site consensus sequence; however, its 3'-splice site consensus is predicted to be relatively weak. A computer-assisted RNA folding program predicts a stem-loop sequesters the 3'-splice site, making it potentially unavailable to spliceosome components. Therefore, this 3'-splice site is most likely not often detected by the spliceosome, and a downstream alternative 3'-splice site is utilized instead. Excision of intron 0 could take place, as it might for intron V, if an alternative secondary structure that does not sequester the 3'-splice site of this intron forms and the site becomes available for spliceosome assembly. A model to explain facultative splicing of intron 0 is shown in Fig. 20B.
Alternative isoform intron-0-alt results from utilization of an alternative 3'-splice site for intron 0 located within exon 2. A similar splicing pattern is observed in *Drosophila tra* pre-mRNA. Here, SXL binds to poly (U) sequences near the 3'-splice site, blocking spliceosomal factor U2AF binding. This results in the use of an alternative splice site within exon 2 in that system. In the case of *rnp-4f* intron 0 alternative splicing, evolutionarily conserved secondary structure is proposed to play a role in the alternative 3'-splice site choice, although conserved primary structure elements and/or involvement with trans-acting factors may also be important. As mentioned previously, a stem-loop structure is predicted to form at the 3'-splice site of intron 0. This stem-loop sequesters the 3'-splice site within the stem, which would cause it to be unavailable to the spliceosomal machinery, causing the spliceosome to utilize an alternative 3'-splice site downstream. A similar observation was made by Estes et al. (1992) in a study of two growth hormone isoforms generated through use of an alternative 3'-splice acceptor. A secondary structure containing the intronic 3'-splice site resulted in use of the downstream 3’ alternative splice site. When the secondary structure harboring the intronic 3'-splice site was abrogated, splicing took place exclusively at this upstream splice site. The 3'-splice site of intron 0 has a weak consensus sequence ending in AAG instead of CAG, which would also be expected to reduce spliceosome recognition at this site. Taken together, these observations support a mechanism in which the 3'-splice site of intron 0 is very often not recognized by the spliceosome, and therefore an alternative 3'-splice site is utilized. A model to explain alternative splicing of intron-0-alt is shown in Fig. 20C.

Interestingly, an evolutionarily conserved pair of stem-loops located at the 3’ alternative splice site for intron 0 was also detected during this study. This stem-loop pair was also found at approximately the same position in *D. yakuba*, the sequence of which is highly conserved (Fig. 14). The conserved twin stem-loops located at this position suggests that they may play a role in alternative splicing of the intron-0-alt isoform, since the structure would be predicted to interfere with binding of spliceosome factor U2AF. The twin stem-loops could act as a platform for both splicing silencing and
enhancing protein binding, as proposed in the Dirksen et al. (2003) model mentioned previously, to modulate splicing efficiency. For instance, if there were more splicing enhancer proteins than silencing proteins to bind the stem-loop, enhancer proteins could bind the stem to destroy the stem-loop steric hindrance and assist spliceosome assembly. In this scenario, splicing would result in generation of the intron-0-alt isoform (Fig. 20C). In addition, spliceosome components are known to have the potential to unfold pre-mRNA secondary structure (Hong-Xiang et al. 1995). The 3'-splice site, being further downstream than the 5'-splice site, is exposed to the spliceosome for a longer period of time. Thus, the spliceosome could have adequate time to unfold or “melt” secondary structure at this splice site. This could explain how, despite being sequestered in a conserved stem-loop, the alternative 3'-splice site is efficiently recognized by the spliceosome.

On the other hand, splicing silencer proteins could bind to the alternative 3'-splice site secondary structure and sterically inhibit splicing enhancers from binding to splicing enhancer elements. In this instance, splicing may not occur due to steric hindrance of the stem-loop bound by a trans-acting factor. This would subsequently inhibit splicing of the intron 0/exon 2 alternative splicing region altogether, as observed in the reported unspliced isoform. A model to explain intron 0 retention is shown in Fig. 20A.

Dirksen et al. (2003) report that trans-acting splicing silencers bound to the splicing silencer element in a stem-loop sequestering the splice site inhibited excision of the mypt1 exon. Therefore, the amount of splicing silencing proteins available could determine the percentage of the unspliced vs alternatively spliced isoforms observed in a particular developmental stage. In this scenario, the more splicing inhibitor protein, the more unspliced isoform observed; the more splicing enhancer protein, the more alternatively spliced isoform. Experimental results suggest that, in virtually all Drosophila developmental stages, the unspliced isoform is more abundant. This would suggest that there is a greater, although fluctuating, number of splicing silencers compared to splicing enhancers during Drosophila development. This result could also be interpreted as a indication that there are more splicing silencer elements in or flanking the 3’ alternative splice site structure that bind more trans-silencing factors.
Therefore, these silencing *trans* factors that bind silencing elements could have more influence on the splicing process.

Another possible role for the conserved twin stem-loops observed at the alternative 3′-splice site could be to exclusively enhance this splice site’s recognition by the spliceosome. Human and mouse fibronectin EDA exon inclusion or exclusion is regulated by a polypurine enhancer element (exon splicing enhancer or ESE) and a silencer element (exon splicing silencer or ESS). It was observed, in a study by Buratti et al. (2004), that the conserved secondary structure context contained by the ESS is important for its function. Rather than having any specific binding function of its own, the authors hypothesized that the function of the ESS element might be to maintain a secondary structure that allows efficient access of the ESE enhancer sequence to *trans*-acting factors. To test this hypothesis, mutations were introduced in the mouse and human ESS homologous regions. It was observed that the ESS mutations abrogated the secondary structure of the mouse EDA exon and induced the ESS element to behave as an ESE. This was not surprising, as previous experiments had determined that an ESS in the context of a hybrid exon with EDA and EDB exon sequences acquires the properties of an enhancer element (Muro et al. 1999). The human secondary structure was not compromised by mutations in the ESS and functioned normally in splicing enhancement, showing that the structure and not the ESS sequence is important for regulation of alternatively splicing the EDA exon.

Taken together, these results suggest a mechanism whereby the ESS maintains secondary structure of the EDA exon so that the ESE can act to enhance splicing through binding SR or SR-like *trans* factors. In the absence of SR factors, secondary structure is not sufficient to induce splicing and the EDA exon is retained. The alternative 3′-splice site stem-loop in *mp-4f* could also contain an ESS that maintains the structure to optimize enhancer protein binding to an ESE, which would be consistent with the conserved nucleotide sequence elements here. As mentioned previously, the relative amounts of silencing and enhancing proteins present at a particular developmental stage would determine whether or not splicing occurs.

In a similar study, Coleman et al. (1998) show that a secondary structure sequestering the 3′-splice site of rat calcitonin/CGRP exon 4 is necessary for
spliceosome recognition of this splice acceptor. Destabilization of the secondary structure resulted in retention of the exon, whereas compensatory mutations that did not destabilize the structure restored splicing to near wild-type levels. Therefore, it is the secondary structure, not the sequence of this structure that facilitates splicing. The 3’-splice site AG intron dinucleotide was found in the loop or bulge portion of the helix structure. Studies have demonstrated that bulges represent unique structures in overall folding of RNA molecules and that cellular factors involved in RNA maturation have high affinity for these bulges (Baker et al. 1994). The bulge containing the 3’-splice site AG of exon 4 is thought to be recognized by a specific trans-acting factor necessary for splice site recognition.

Intriguingly, the relative abundance of the two intron 0 isoforms differs across development (Fig. 15). It is interesting that the mRNA isoforms which have undergone splicing of the intron-o-alt transcript predominate during mid-embryo stages, during which time the CNS is beginning to develop and corresponding to stages during which an as yet uncharacterized isoform is relatively abundant along the dorsal roof of the ventral nerve cord (Hess et al. 1996). The relative abundances of intron-0-alt isoforms detected during RT-PCR is similar to those detected during analysis of full-length transcripts across development via Northerns (Fig.18), and it is likely that both sets of information are reflections of the same transcripts.

**rnp-4f translation regulation by the 5’-UTR**

The fact that alternative splicing occurs in the 5’-UTR of *rnp-4f* suggests that the alternatively spliced isoforms arising from this region could confer translational regulation. The frequently observed unspliced 5’-UTR isoform retains intron 0. The retention of intron 0 may interrupt a regulatory sequence activated upon its excision. As observed by Roberts et al. (2005), the inclusion of exon 1B in the 5’-UTR resulted in inhibition of translation of the ALAS1 transcript. The authors’ data suggested that inclusion of this alternative exon could interrupt a regulatory sequence necessary for translation initiation. As a corollary to this, a second 5’-UTR isoform generated by excision of intron 0 may result in production of an active regulatory element that controls translation efficiency.
The third 5′-UTR isoform, intron-0-alt, is generated as a result of utilization of an alternative 3′-splice site for intron 0. As mentioned previously, a conserved twin stem-loop structure exists at this alternative 3′-splice site. Excision of this ~180-nt tract within the 5′-UTR would eliminate the upstream member of this twin stem-loop structure. Stem-loop structures within the 5′-UTR have been implicated in translation repression by impeding 40S ribosome subunit scanning (Kozak, 1990). However, the stem-loop is required to be proximal to the cap or have stability greater than -50 kcal/mole to exert this effect (Wilkie et al. 2003). The 3′-alternative splice site stem-loop is not close to the cap and is only moderately stable at -8.2 kcal/mole in *D. melanogaster* and -4.3 kcal/mole for the corresponding structure in *D. yakuba* (Fig. 15). However, Newton et al. (2003) report that a -12.8 kcal/mole stem-loop structure in the 5′-UTR of neuronal nitric-oxide synthase mRNA was sufficient to inhibit translation. The stem-loop in that system was shown to interact with an unknown RNA-binding complex that likely confers the observed inhibition to translation. Therefore, it is possible that the observed *mp-4f* 5′-UTR twin stem-loops could also confer translational repression through binding trans-acting protein(s). In this regard, it is interesting that the conserved upstream member of this twin stem-loop structure contains a bulge-loop “A” in both species. As previously mentioned, such structures have often been implicated in the binding of trans-acting factors involved in RNA maturation (Baker et al. 1994).

The size and relative stability of the *mp-4f* 5′-UTR conserved stem-loop structure is similar to the IRE conserved stem-loop structure, which has a moderate stability at -9.2 kcal/mole in humans. The IRE is found within a number of 5′-UTRs, including ferritin mRNA, where it binds IRP (Hentze and Kuhn 1996). The IRE-IRP complex inhibits translation of ferritin in response to increased cellular iron concentrations. Such a mechanism could function in the *mp-4f* 5′-UTR to confer translation inhibition. A protein (X) could bind the stem-loop at the alternative 3′-splice site creating a structure like the IRE-IRP complex that impedes progression of the ribosome and inhibits translation initiation.
A model for auto-regulation of *rnp-4f* mRNA translation

A case can be made for RNP-4F protein auto-regulation of *rnp-4f* mRNA translation itself. *rnp-4f* encodes a spliceosomal assembly protein that is distributed homogeneously in *Drosophila* embryos, although more abundantly represented along the dorsal roof of the ventral nerve cord. Orthologues of this factor have been identified in many other organisms and are believed to have a similar function (Raider and Guthrie 2002). This conserved RNP-4F factor is therefore of considerable importance in promoting the splicing mechanism. It is hypothesized that when RNP-4F protein is abundant, it binds to the twin stem-loop elements at the 3'-alternative splice site located in the 5'-UTR. The binding of the protein is predicted to inhibit splicing at the 3'-alternative stem-loop and this complex would presumably remain in the processed transcript for transport to the cytoplasm. The 3'-alternative stem-loop bound by RNP-4F would then inhibit progression of the scanning 40S ribosomal subunit, resulting in failure of RNP-4F translation and subsequent degradation of this mRNA isoform. As increasingly less RNP-4F protein is produced, as a result of this inhibition, less RNP-4F will be available to bind *rnp-4f* pre-mRNAs. If RNP-4F is not bound to the alternative twin 3'-stem-loop, the structure is more easily accessible to spliceosomal factors. In this scenario, the upstream element within the alternative 3'-stem-loop is excised from the transcript and no longer impedes ribosomal scanning or translation. This mRNA isoform is then expected to increase in relative abundance. A model to explain RNP-4F auto-regulation is shown in Fig. 21.

There are several experimental observations which are interpreted in support of the proposed RNP-4F auto-regulatory model. If one begins with the premise that relative abundance of an mRNA isoform will be correlated with enhancement in protein levels derived from that isoform, then the relative abundance of the shorter (S) mRNA isoform in very early embryos and in adults, as observed in the Northern results (Fig. 18) and RT-PCR results (Fig. 10B), is correlated with the shorter intron-0-alt spliced mRNA isoform. The model predicts that RNP-4F will be extensively translated at these stages, and this expectation is supported by the finding that *rnp-4f* mRNA levels are elevated during these periods, as shown by measurement of the Northern-derived ratio of total rnp-4f to rp49 mRNAs (Fig. 16).
Conversely, the relative abundance of the longer (L) mRNA isoform in mid-embryo stages of development, as observed in the Northern results (Fig. 18) and RT-PCR results (Fig. 10B), is correlated with the intron 0 non-spliced mRNA isoform. The model predicts that RNP-4F will be translated at minimal levels at these stages, and this is supported by the finding that rnp-4f mRNA levels dramatically decline during this period, as shown by measurement of the Northern-derived ratio of total rnp-4f to rp49 mRNAs (Fig. 16). This prediction is further supported by the preliminary observation (Petschek and Concel 2003) that RNP-4F protein is expressed at relatively low levels during first, second and third larval instar stages, as indicated by immunocytochemistry. The molecular basis underlying the observed mRNA level decline remains to be determined, and may depend on an associated antisense RNA mechanism (Peters et al. 2003). The proposed model will undoubtedly be the subject of additional future research, as it is readily testable.
Future Directions

In this Thesis study, five distinct \textit{rnp-4f} pre-mRNA splicing variants have been identified: unspliced intron 0 (abundant), spliced intron 0 (infrequent), intron-0-alt (abundant), unspliced intron V (abundant), and spliced intron V (infrequent). These splicing events are hypothesized to be regulated, in part, by \textit{cis}-regulatory elements, visualized as stem-loops that are located flanking and sequestering the 3'-splice sites of intron 0 and intron-0-alt, in addition to similar elements at the 5'- and 3'-splice sites of intron V. In order to determine if these predicted stem-loop structures exist in \textit{rnp-4f} pre-mRNA transcripts, ribonuclease digestion analysis, as described in Colman et al. (1998), could be utilized to detect double-stranded RNA in the regions of \textit{rnp-4f} transcripts thought to contain secondary structure. In this approach, structure-specific ribonucleases, namely RNase CV\textsubscript{1}, which is specific for double-stranded RNA, and single-stranded specific RNases T\textsubscript{1} or T\textsubscript{2}, are used to probe the regions harboring proposed secondary structure. If CV\textsubscript{1} RNases are capable of cutting the RNA in the proposed stem-loop region and T\textsubscript{1} and T\textsubscript{2} are not, the presence of secondary RNA structure is confirmed.

If any of the probed regions of \textit{rnp-4f} indeed form the predicted stem-loop structures, the next logical step would be to make compensatory mutations in the stem-loops to determine if they have a role in \textit{rnp-4f} splicing regulation. Compensatory mutations, as described by Buratti et al. (2004), change a base or a set of bases at a particular site in the stem-loop so as not to disrupt the overall secondary structure, but only the primary sequence. If splicing regulation is not significantly disrupted upon compensatory mutation of the stem-loop, then it is concluded that the structure and not the sequence of the stem-loop is important for splicing regulation. However, if regulation of splicing is disrupted upon compensatory mutation, then the sequence contained within the stem-loop is important for splicing regulation and the secondary structure is thought to play little if any role. These specific sequences, termed \textit{cis}-regulatory sequences, can act to enhance or silence splicing. There are cases in which both the stem-loop and the sequence contained within the secondary structure have been found to be important for splicing regulation. This may be the case for \textit{rnp-4f} pre-
mRNA transcripts, as it is hypothesized that the alternative intron 0 3'-splice site twin stem-loops may contain both splicing enhancer elements and splicing silencing elements, oriented strategically to either enhance or silence splicing. A model for this type of regulation is described in Dirksen et al. (2003), where a 5'-splice site stem-loop contains a splicing silencer element in the loop region of the structure that if bound by a trans-acting factor blocks access of splicing enhancing trans-acting factors from binding the cis enhancer element located in the stem.

Further testing will be needed to determine the composition and function of each predicted stem-loop described in this Thesis study. Cis-regulatory sequences are almost exclusively identified by deletion or mutational analysis, as these sequences lack substantial consensus. Such deletion/mutation analysis could be carried out, as described by Buratti et al. (2004), in or flanking the predicted stem-loop regions to determine their composition and influence on splicing regulation of rnp-4f pre-mRNAs.

Another line of future research could be to determine whether or not trans-acting factors play a role in rnp-4f splicing regulation. It is proposed that trans-acting factors may bind the predicted stem-loops located at the alternative 3'-splice site of intron 0, at both splice sites of intron V, and at the intron-0-alt 3'-splice site. The RNA electrophoretic mobility shift assay technique, as described by Newton et al. (2002), could reveal whether or not transacting factors interact with these specific stem-loops. 32P-labeled stem-loop region-specific riboprobes could be commercially constructed and incubated with cytoplasmic protein extracts, followed by resolution during non-denaturing polyacrylamide gel electrophoresis (PAGE). To test the specificity of trans-acting factor interaction with the riboprobe, 32P-labeled riboprobe deletion mutants of the specific stem-loop being tested (also commercially constructed) could be incubated with the same cytoplasmic protein extracts and also resolved by non-denaturing PAGE. If the rnp-4f stem-loops in fact bind trans-acting factors, the RNA would be expected to migrate through the gel more slowly than observed when riboprobe alone is utilized. If this interaction were specific for rnp-4f stem-loop regions, the mutant riboprobe would not be capable of binding the trans-acting factors and would be expected to migrate through the gel at the predicted rate for its nucleotide sequence length.
In this thesis study, RNP-4F protein is hypothesized to auto-regulate its own translation through binding the alternative intron 0 3'-splice site twin stem-loops in the 5'-UTR. One way to determine if this auto-regulation occurs would be to transfer the alternative intron 0 3'-splice site stem-loop RNA bound by trans-acting factor(s), discussed in the previous section, by electroblotting onto a nylon membrane. Once transferred, RNP-4F radioactively labeled or fluorescently labeled antibody could be applied to the blot to determine if the protein is present.

In a further test of the auto-regulation model, it would be informative to quantify the levels of RNP-4F protein across development and compare the results to RT-PCR derived intron-0-alt transcript levels across development. It is proposed that this transcript, from which the intron 0 alternative 3'-splice site upstream stem-loop is excised, encodes nearly all RNP-4F protein across development. The unspliced version retains the intron 0 alternative 3'-splice site twin stem-loops, which is proposed to inhibit ribosome scanning and subsequent initiation of translation. If the level of intron-0-alt transcripts across development matches the actual RNP-4F protein levels across development, it would make a strong case for the intron-0-alt transcript being translationally active. The results could also make a case for the unspliced version, which retains the stem-loop, as being translationally inhibitory.

RNA interference (RNAi) could also be utilized to determine the effect of elimination of intron 0 unspliced transcripts. Short, commercially synthesized duplex 21-23 nt long RNAi fragments with sequences identical to those in the ~180 nt excised from the intron-0-alt splice variant could be microinjected into *Drosophila* embryos. After a predetermined incubation period, *in situ* hybridizations could be performed to determine the location of the *rnp-4f* transcripts. In wild-type (WT) embryo *in situ* hybridizations, there is a concentration of *rnp-4f* transcripts in the dorsal roof of the ventral nerve cord as well as a homogeneous distribution throughout the embryo. The knockout embryos could be compared to the WT embryos to determine if the unspliced *rnp-4f* variant has a specific localization.
Figure 1. rnp-4f gene. The *rnp-4f* gene is diagrammed, showing all known or suspected intron positions 0–VIII and HAC1, and numbered predicted exons 1–11. Coding regions are shown in black and non-coding in white. The encoded protein contains a C-terminal motif (CTM), at least two RNA- recognition motifs (RRM), a nuclear localization signal (NLS), and several half-tetracopeptide (TPR) repeats [modified from Rader and Guthrie (2002)].
Figure 2. (A) The Drosophila rnp-4f gene is evolutionarily conserved. In all species studied, the encoded proteins contains a C-terminal motif (CTM), at least two RNA-recognition motifs (RRM), a nuclear localization signal (NLS) within the hinge region, and several half-tetratricopeptide (TPR) repeats. (B) A model for RNP-4F function in the yeast homologue PRP-24. During spliceosome assembly, the entering U4/U6 snRNPs dimer is stabilized by interactions with the PRP-24 RRM domains, while the CTM interacts with Lsm proteins complexed to U6-snRNA [modified from Rader and Guthrie (2002)].
Figure 3. Schematic representation of the spliceosome in terms of the role of snRNP particles in pre-mRNA splicing, including transitions that require a PRP protein. Pre-mRNA containing two exons separated by an intron enters splicing complexes with snRNPs and exits as mRNA (bottom line) and excised lariat intron (left border). CC, A, B1, B2, C1, C2 and I represent complexes within the splicing pathway that have been distinguished biochemically, genetically or both. **Abbreviations:** 5' SS, 5'-splice site; 3' SS, 3'-splice site; bp, branch site; Py, pyrimidine-rich tract. The individual snRNPs are: U1, U2, U4, U5 and U6. [Sharp 1994].
Figure 6. Expression of the *Drosophila* sex-lethal (SXL) protein.

(A) In early embryogenesis, the early promoter (P_e) initiates production of a small amount of a SXL protein isoform in female but not male embryo. In the late female embryo, SXL produced earlier masks the 3'-splicing signal for the second intron, which induces skipping of exon 3 containing a termination codon, resulting in production of functional SXL protein. (B) The *sxl* early promoter is not initiated in males and functional SXL protein is not produced. Exon 3 containing termination codon is included in male *sxl* transcripts, due to lack of SXL protein to induce exon 3 skipping, and a non-functional SXL protein is produced [www.web-books.com/MoBio/Free/Ch5A4.htm].
Figure 7. SXL target genes *tra* and *msl-2*. (A) In males, U2AF binds to the 3'-pyrimidine-rich tract of intron 1, promoting splicing and resulting in retention of a stop codon in exon 2, leading to production of a non-functional TRA protein. In females, SXL binds the poly (U) tract of the 3'-splice site of intron 1, inhibiting U2AF from binding the intronic 3'-pyrimidine-rich tract, and U2AF is displaced to an alternative 3'-splice site within exon 2. As a result of this SXL-induced alternative splicing, the stop codon located in the upstream portion of exon 2 is skipped and a functional TRA protein is produced. (B) In males, U1 binds to the 5'-splice site of intron 1 and U2AF binds the 3'-pyrimidine-rich tract of intron 1 inducing splicing of this intron, and leads to production of a functional MSL-2 protein. In females, S XL binds to poly(U) tracts near the 5'- and 3'-splice sites within intron 1. This blocks U1 binding at the 5'-splice site and U2AF at the 3'-pyrimidine-rich tract, inducing intron retention. SXL remains bound to the *msl-2* transcript in the cytoplasm and inhibits its translation [modified from Penalva and Sanchez (2003)].
Figure 8. *Drosophila melanogaster* life-cycle.
Female *Drosophila* lay eggs that undergo 24 hours of embryonic development that are sectioned into hourly increments based on appearance of certain developmental markers (0-2 h, 0-4 h, 6-12 h, 12-18 h and 18-24 h). Following this 24-h period the embryo hatches into a 1st instar larva. The 1st instar larva grows into a 2nd and then 3rd instar larva, at which point the organism pupates. The pupa emerges as an adult following pupation. One complete cycle takes approximately 10 days [modified from Flagg (1998)].
Figure 10. rnp-4f RT-PCR analysis across development. (A) The rnp-4f gene is diagrammed, showing six location-specific primer pairs (A-F) utilized during RT-PCR analysis of alternative splicing. (B) Primer set A reaction products. From top to bottom, the three bands represent: intron 0 unspliced, intron I spliced, followed by (intron 0 spliced, intron I spliced) and (intron 0-alt spliced, intron I spliced). Upon sequencing the central band, it was found that two different molecules were present, the fastest migrating band being included. Therefore, the intensity of this central band is not indicative of the frequency of the intron 0 spliced event. (C) A second primer set was utilized to bracket only intron 0, which resulted in two bands: (intron 0 unspliced) and (intron 0 spliced). (D) Primer set B reaction products. Introns I—II are spliced. (E) Primer set E reaction products. Introns VI—VIII are spliced. (F) Primer set F reaction products. The single band detected is identical to genomic DNA, indicating absence of splicing. (G) Primer set D reaction products. From top to bottom, the three bands represent: (intron IV spliced, intron V unspliced), followed by (intron IV spliced, intron V spliced indicated by upper arrow), and another faint band (cryptic single-stranded: intron IV spliced, intron V unspliced, indicated by lower arrow). (H) Southern hybridization of primer set D reaction products (gel not identical to G) showing binding of probe to all three bands. (I) Primer set C reaction products. From top to bottom, the two bands represent: (intron IV spliced, intron V retained, introns VI—VIII spliced) and a faint band (cryptic single-stranded molecule, arrow, with identical intron content to the other band). (J) Southern hybridization of primer set C reaction products (gel not identical to I) showing binding of probe to both bands. (K, L) rnp-4f gene aligned with encoded RNP-4F proteins, showing protein domains. All observed 5' UTR splice variants are shown, as it is not currently known which variant(s) is/are part of transcripts that retain or splice intron V. However, information from two sequenced full-length cDNAs obtained from screening cDNA libraries (see text) shows that whereas one splices intron 0 and one does not, both retain intron V. (K) Retention of intron V results in generation of the full-length RNP-4F protein. (L) Facultative splicing of intron V results in a frameshift mutation that introduces a stop codon downstream, resulting in generation of a truncated RNP-4f protein from which the NLS is deleted.
Figure 11. Model for generation of cryptic folded cDNAs via RT-PCR.

**Expected:** During the RT-PCR reaction, *mp-4f* cDNA double helix denatures, forward and reverse primers anneal to the template, followed by extension of a new complementary DNA strand. One band would be expected upon electrophoresis (linear, double-stranded cDNA).

**Model:** Following RT-PCR denaturing step, a small percentage of the single-stranded cDNA templates fold upon themselves, forming highly coiled secondary structure. Two bands are then detected upon electrophoresis: one band migrating at the expected rate for its size (linear, double-stranded cDNA) and another faint band migrating substantially faster than its actual size (single-stranded secondary structure cDNA).
Figure 12. Predicted RT-PCR induced single-stranded cDNA secondary structure arising from use of primer set C. The indicated secondary structure was generated using the miFold program (Zuker 2003). $dG = -253$ kcal/mole.
Figure 13. Predicted RT-PCR induced single-stranded cDNA secondary structure arising from use of primer set D. Structure was generated using the mFold program (Zuker 2003). $\Delta G = -103$ kcal/mole.
Figure 14. *mp-4f* 5'-UTR sequence alignment for *D. melanogaster* and *D. yakuba.* (A) Orientation diagram, showing major 5'-UTR features and the three different mRNA isoforms arising from splicing around intron 0: a) unspliced intron 0, b) intron 0 facultatively spliced and c) intron-0-alt. (B) Alignment between *D. melanogaster* and *D. yakuba* showing extensive sequence identity, as indicated by vertical lines between nucleotide bases. Dashes have been inserted to improve alignment. Underlined regions represent the location of stem-loop surrounding 3'-splice site for intron 0. Wavy underline shows location of twin stem-loops surrounding alternative 3'-splice site for intron 0. The branch point for both the intron 0 and alternative putative intron 0 splicings is indicated with semi-circles.
Figure 15. Conserved intron 0 alternative intron 0 stem-loops. (A) Orientation diagram shows positions of intron 0 3′-splice site stem-loop and intron 0 3′-alternative splice site stem-loop, in addition to the three different mRNA isoforms arising from splicing around intron 0: a) unspliced intron 0, b) intron 0 facultatively spliced and c) intron-0-alt. (B) The stem-loop predicted at the intron 0 3′-splice site in D. melanogaster sequesters that splice site (indicated by bracket) near the base of the stem. The stem-loop at this site in D. yakuba also sequesters the 3′-splice site (indicated by bracket) near the base of the stem. (C) The twin stem-loops predicted at the intron 0 3′-alternative splice site for D. melanogaster and D. yakuba are shown as sequestering the (indicated by brackets). Compensatory base changes in both species are indicated by solid circles. Locations of nucleotides forming indicated stem-loops are shown in alignment (Fig. 14).
Figure 16. Quantification of mnp-4f mRNAs across development.

**Northern Analysis:** mnp-4f transcripts are very abundant in early embryos (0-2 h) relative to rp49 internal loading control. mnp-4f transcript abundance begins to decline during subsequent embryo development, reaching a low plateau level in 12-18 h embryos. Low plateau levels persist until pupal stage, when mRNA levels begin to increase. Indicated mnp-4f mRNA transcript ratio intensities were obtained by calculation following scanning of hybridization results into phosphorimager. **RT-PCR Analysis:** In early embryos, unspliced intron 0 transcripts are approximately twice as abundant as alternatively spliced intron-0-alt transcripts. Unspliced transcripts then increase to approximately 3 times the level of spliced intron-0-alt transcripts by 12-18 h, remaining at that level through 3rd instar larvae, then decline in pupa and adult to early embryo levels. Indicated mRNA isoform transcript ratio intensities were obtained by calculation following scanning of SYBRGreen I stained gel bands into phosphorimager. Standard deviation bars are indicated for the repeated analysis.
Figure 17. Predicted stem-loops surrounding intron V 5'- and 3'-splice sites. (A) The positions of the predicted 5' and 3'-splice site stem-loops are shown in the upper orientation diagram, in addition to the two different mRNA isoforms arising from splicing intron V: a) unspliced intron V, and b) facultatively spliced intron V. (B) Sequence alignment between D. melanogaster (D. MEL) D. yakuba (D. YAK) and D. pseudoobscura (D. PSE) shows sequence identity as indicated by vertical lines between nucleotide bases. Encoded amino acids are shown for D. MEL and D. PSE. Intron location is within brackets. Underlined regions represent the locations of stem-loops spanning the 5' and 3'-splice sites. The corresponding sequences spanning the 5' and 3'-splice sites in D. yakuba and D. pseudoobscura are not predicted to form stem-loops; however, these two species have one or more splice site consensus changes which are expected to mimic the effect of the secondary structure found in D. melanogaster at these locations. Arcs show potential branchpoint consensus sequence, predicted to be sequestered within the 5'-stem loop in D. melanogaster. (C) The 5'-splice site in D. melanogaster (bracket) is sequestered at the base of the twin stem-loops. The 3'-splice site (bracket) is sequestered within the indicated stem-loop.
Figure 18. mnp-4f and sas10 developmental mRNA Northern. (A) Orientation diagram (upper panel) shows structure for cloned genomic Xho I/Xho I DNA fragment used for initial probe preparation, containing entire sas10 gene and 3'-half of mnp-4f gene. (B) Hybridization results following poly (A') RNA electrophoresis and Northern transfer across development. Three sas10 mRNA isoforms (S, I, L) and two mnp-4f isoforms (S, L), sized at about 3,300- and 3,600-nt, are resolved. Locations of denatured fly and bovine rRNA size markers (not shown) are indicated. rp49 mRNA was probed for an internal loading control. (C) Inset showing results of re-hybridization using mnp-4f specific probe having higher specific activity. A switch in relative mRNA isoform abundance occurs during embryo stage of development (heavy arrow), so that the larger isoform predominates. (D) Inset showing results, following membrane stripping, of re-hybridization with a probe specific to the ~180-nt intron-0-alt 5'-UTR tract. Results show that only one mnp-4f mRNA isomer, the larger one, retains this tract.
Figure 19. Model for generation of unspliced and facultatively spliced intron V. (A) The stem-loop pair sequestering the 5′-splice site of intron V blocks U1-snRNP binding. The stem-loop sequestering the 3′-splice site of intron V blocks U2AF from binding the pyrimidine-rich tract. Intron V is retained due to the presence of 5′- and 3′-splice site secondary structure, resulting in a full-length protein. (B) In a small minority of mps4f transcripts the secondary structure at the 5′- and/or 3′-splice sites does not form, or a modified secondary structure develops that does not sequester the splice sites. U1-snRNP and U2AF bind the 5′- and 3′-splice sites, respectively, and intron V is spliced, resulting in a truncated protein resulting from a stop codon created via frame-shift. Splicing decisions here may also involve trans-acting factors that are not yet identified.
Figure 20. Model for generation of unspliced intron 0, facultatively spliced intron 0, and alternatively-spliced intron 0-alt isoforms. (A) The 5'-splice site of intron 0 is free to bind U1-snRNP. However, the 3'-splice site of this intron has a weak splice site consensus sequence and is also predicted to form a stem-loop structure that sequesters the splice site. U2AF is therefore unable to bind the pyrimidine-rich tract at this splice site. The alternative intron 0 3'-splice site is sequestered within a pair of stem-loops spanning this site. A trans-acting splicing silencer is proposed to bind this alternative splice site stem-loop pair and to inhibit U2AF from binding the pyrimidine-rich tract. Spliceosome assembly is incomplete and intron 0 is retained. (B) In a small minority of mp-4/7 transcripts, U1-snRNP binds the 5'-splice site of intron 0 and U2AF binds the 3'-splice site, due to lack of formation of the stem-loop sequestering this splice site and/or recognition of the weak splice site consensus sequence by U2AF. The spliceosome is assembled and intron 0 is spliced. (C) U1-snRNP binds the 5'-splice site of intron 0 and a proposed trans-acting splicing enhancer binds the stem-loop pair at the alternative intron 0 3'-splice site. The splicing enhancer factor abrogates the secondary structure effect, allowing U2AF to bind the pyrimidine-rich tract at the alternative 3'-splice site. The spliceosome is assembled and intron-0-alt is spliced.
Figure 21. Model for control of RNP-4F expression by auto-regulation. (A) RNP-4F abundance: When RNP-4F protein is in abundance it binds the alternative intron 0 3'-splice site stem-loop pair and functions as a trans-acting splicing silencer, the resulting complex inhibiting U2AF from binding the pyrimidine-rich tract. U2AF is also blocked from binding to the 3'-splice site of intron 0, due to its weak splice site consensus and/or presence of a stem-loop sequestering this splice site. The spliceosome is unable to assemble and splicing does not take place. RNP-4F protein remains bound to the alternative intron 0 3'-splice site stem-loop pair in the cytoplasm, where it impedes 40S ribosome scanning and subsequent translation initiation, resulting in inhibition of RNP-4F protein production from these transcripts. It is predicted that rnp-4f mRNA levels will then decline. (B) RNP-4F deficit: When RNP-4F protein is not abundant, trans-acting splicing enhancers bind the alternative intron 0 3'-splice site stem-loop pair, which abrogates this secondary structure effect and allows U2AF to bind the pyrimidine-rich tract. U1-snRNP binds the 5'-splice site of intron 0 and the spliceosome assembles to generate intron-0-alt transcripts via alternative splicing. The abrogation of the alternative intron 0 3'-splice site stem-loop pair, due to lack of RNP-4F protein available to bind and maintain this structure, enables 40S ribosomal subunit scanning and subsequent 80S translation of RNP-4F protein.
Appendix I

Conventional Reverse Transcription (60 ul reaction)

1. Combine 3 ul random hexamer primers, 1 ug poly (A+) RNA template and DEPC H₂O to a total volume of 25.5 ul for each RT.

2. Incubate tubes at 90°C for 2 min.

3. Transfer 300 ml of the 90°C water to a 1000 ml beaker.

4. Float tubes in the 1000 ml beaker and ramp to a temperature of 28-30°C (1-1 1/2 h).

5. Make a master mix by combining 12 ul DEPC H₂O, 12 ul 5 X MMLV reverse transcriptase buffer, 6 ul 10 mM dNTPs, and 1 ul (40 U) Rnasin for each RT.

6. Add 31.5 ul of the master mix to each RT tube.

7. Add 2 ul (400 U) MMLV reverse transcriptase enzyme to each tube.

8. Incubate tubes at 20°C for 15 min.

9. Incubate tubes at 42°C for 2 h.

10. Incubate tubes at 90°C for 2 min.

11. Store RTs at -70°C.
Appendix II

Thermocycler Reverse Transcription (20 ul reaction)
(up to 50 ul reactions can be performed with this method)

(1) In 0.2 ml microcentrifuge tube combine the following for each 20 ul reaction:
   1 ug appropriate RNA sample with enough DEPC water to total 11.5 ul
   4 ul 5 X MMLV reverse transcriptase buffer
   2 ul 10 mM dNTPs
   1 ul hexanucleotide random primer mix
   0.5 ul (20 U) RNasin
   1 ul (200 U) M-MLV reverse transcriptase

(2) Place all tubes in the thermocycler for a single cycle (25 min):
   42°C for 15 min
   99°C for 5 min
   5°C for 5 min

(3) When cycle complete, immediately transfer tubes to -70°C for storage.
Appendix III

QIAquick Gel Extraction Kit Protocol (using a microcentrifuge)

(1) Weigh excised DNA agarose gel slices in a 1.5 ml microcentrifuge tube. (The maximum amount of gel slice per QIAquick column is 400 mg)

< 2% agarose = add 3 volumes of Buffer QG to 1 volume of gel

> 2% agarose = add 6 volumes of Buffer QG to 1 volume of gel

(2) Incubate at 50°C for 10 min (or until gel is completely dissolved). To help dissolve gel, vortex every 2-3 min. Check that the color of the mixture is yellow (similar to the color of the QG buffer) after gel slice(s) has dissolved. If color is orange or violet add 10 ul 3 M NaAc pH 5.0 and mix. This will turn mixture yellow.

(3) For fragments < 500 bp and > 4 kb add 1 gel volume 100% isopropanol to the sample mix.

(4) To bind DNA, apply the sample to the QIAquick column with the 2 ml collection tube attached, and centrifuge for 1 min at 13,000 rpm. (The column reservoir holds a maximum of 800 ul. For sample volume more than 800 ul, reload the reservoir and spin again.)

(5) Discard the flow-through and place the QIAquick column back in the same collection tube.

(6) Add 0.5 ml Buffer QG to the QIAquick column and centrifuge at 13,000 rpm for 1 min.

(7) Add 0.75 ml of Buffer PE to the QIAquick column and centrifuge at 13,000 rpm for 1 min to wash the sample.

(8) Discard the flow-through and centrifuge the QIAquick column for an additional 1 min at 13,000 rpm to completely rid column of Buffer PE.

(9) Place column in clean, autoclaved 1.5 microcentrifuge tube.

(10) Let the column stand for at least 3 min to allow all ethanol to evaporate. Smell the column for any traces of ethanol before proceeding to the next step.

(11) To elute DNA from column, add 30 ul autoclaved double distilled H$_2$O to the center of the QIAquick column and let it stand at least 1 min.
(12) Centrifuge at 13,000 rpm for 1 min.
(13) Store DNA at -20°C.
Literature Cited


of Mason-Pfizer monkey virus (MPMV) accesses a cellular mRNA export pathway. EMBO J. 16: 7500-7510.


