IDENTIFICATION OF CLK/STY AS A
SERINE/ARGININE-RICH PROTEIN KINASE

by

Karen Colwill

A Thesis submitted in conformity with the requirements
for the Degree of Doctor of Philosophy in the
Department of Molecular and Medical Genetics
University of Toronto

© Copyright by Karen Colwill 1997
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
Karen Colwill
Identification of Clk/Sty as a serine/arginine-rich protein kinase
Doctor of Philosophy 1997
Graduate Department of Molecular and Medical Genetics
University of Toronto

Abstract

I have investigated the function of the protein kinase Clk/Sty. To discover proteins that may interact with Clk/Sty, I performed a two-hybrid screen in yeast using Clk/Sty as bait. Twenty six mouse cDNA clones were identified in the screen.

Four of these clones encode known transcription factors c-Myc, ZFP38, N-CoR, as well as a novel zinc finger protein. Other clones identified in the screen suggest a role for Clk/Sty in RNA processing. For instance, one of the clones, CARS-Cyp, encodes a novel arginine/serine-rich (RS) cyclophilin which co-localizes with splicing factors to nuclear speckles. Furthermore, six of the two-hybrid clones encode RNA-binding proteins. One of these clones is a mammalian homologue of the Drosophila splicing regulator tra-2. Three of the remaining five RNA-binding proteins are members of the SR family of essential splicing factors.

I focused my studies on one member of the SR family, ASF/SF2, and found that Clk/Sty phosphorylated ASF/SF2 in vitro on sites that are also phosphorylated in vivo. As well, a catalytically inactive Clk/Sty co-localized to nuclear speckles with SR family members. Interestingly, overexpression of wild-type Clk/Sty led to disassembly of nuclear speckles resulting in redistribution of SR proteins. Therefore, Clk/Sty regulates the intranuclear distribution of SR proteins, presumably by phosphorylation.
SRPK1 was the first kinase to be cloned and identified as an SR protein kinase. I found that SRPK1 phosphorylated RSRS repeats in ASF/SF2. Clk/Sty, on the other hand, phosphorylated SPR and KSKS repeats as well as RSRS repeats. Therefore, Clk/Sty displays a broader substrate specificity than SRPK1 in vitro, suggesting that Clk/Sty may phosphorylate more substrates than SRPK1 does in vivo.

Pre-mRNA splicing is a complex process that is executed with high fidelity. In order to choose splice sites appropriately, the splicing machinery must be responsive to its external environment. I propose that Clk/Sty functions to transduce signals to this machinery enabling proper splice site selection.
Acknowledgments

Now that the long daunting task of a PhD is at its end, I want to acknowledge those who helped me this far. First of all, I want to thank my supervisor Tony Pawson who provided me with the opportunity to work and develop as a scientist in his lab, encouraging me when I needed it but also allowing me to find my own way. I'd like to thank my supervisory committee members Brenda Andrews, Sam Benchimol, and John Dick for their help and support during my PhD. A special thanks goes out to Brenda in whose lab I started the two hybrid screen which is the cornerstone of my PhD thesis.

All members of the Pawson lab, past and present, have been invaluable to me by providing expertise and support over the past five years. In particular, I'd like to thank Lee Mizzen for showing me the ropes, Jerry Gish for his expertise and knowledge, Louise Larose for her guidance, and Sara Jacob and Venus Lai for help and assistance along the way.

I'd like to thank the co-authors on my thesis chapters both for their input into the Clk project and for giving me permission to reproduce their results in my thesis. My thanks also extends to EMBO Journal, Journal of Biological Chemistry, and GENE for allowing me to reproduce published papers in my thesis. I'd also like to thank Susan Taylor, Structure journal, Jack Keene, and Trends in Biochemical Sciences for allowing me to reproduce their crystal structures in my thesis. I'm indebted to the MRC for their support in providing me with a studentship.

My final thanks goes to members of my family. My husband, Darryl, has been a wonderful anchor supporting me in my work and listening to my latest developments. My father taught me the value of working hard and doing a job well. Finally, I'd like to acknowledge my mother who was always there for me when I needed her and whose love, courage, and strength I will carry with me always. I dedicate this thesis to her memory in the belief that the cure to cancer lies in research.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdML</td>
<td>Adenovirus major late</td>
</tr>
<tr>
<td>AFC</td>
<td><em>Arabidopsis fus3</em> complementing cDNA</td>
</tr>
<tr>
<td>ASF/SF2</td>
<td>alternative splicing factor/splicing factor 2</td>
</tr>
<tr>
<td>ASFΔRS</td>
<td>ASF/SF2 construct lacking the RS domain</td>
</tr>
<tr>
<td>ASLV</td>
<td>avian sarcoma-leukosis virus</td>
</tr>
<tr>
<td>bHLH-ZIP</td>
<td>basic helix-loop-helix leucine zipper</td>
</tr>
<tr>
<td>cAPK</td>
<td>cyclic AMP dependent protein kinase</td>
</tr>
<tr>
<td>CARS-Cyp</td>
<td>Clk associating RS-cyclophilin</td>
</tr>
<tr>
<td>CDK</td>
<td>cyclin dependent kinase</td>
</tr>
<tr>
<td>CKII</td>
<td>casein kinase II</td>
</tr>
<tr>
<td>Clk/Sty</td>
<td>CDC28 like kinase/serine-threonine-tyrosine</td>
</tr>
<tr>
<td>Doa</td>
<td><em>darkener of apricot</em></td>
</tr>
<tr>
<td>dsx</td>
<td><em>doublesex</em></td>
</tr>
<tr>
<td>DsxF</td>
<td>female-specific <em>doublesex</em> protein</td>
</tr>
<tr>
<td>DsxM</td>
<td>male-specific <em>doublesex</em> protein</td>
</tr>
<tr>
<td>dsxRE</td>
<td>repeat element in exon 4 of <em>doublesex</em></td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>ERK</td>
<td>extracellular regulated kinase</td>
</tr>
<tr>
<td>ERS</td>
<td>exon recognition sequence</td>
</tr>
<tr>
<td>EST</td>
<td>express sequence tag</td>
</tr>
<tr>
<td>GST</td>
<td>glutathione S transferase</td>
</tr>
<tr>
<td>HA</td>
<td>hemagglutinin</td>
</tr>
<tr>
<td>hnRNP</td>
<td>heterogenous nuclear ribonucleoprotein</td>
</tr>
<tr>
<td>IGC</td>
<td>interchromatin granule cluster</td>
</tr>
<tr>
<td>InsR</td>
<td>insulin receptor kinase</td>
</tr>
<tr>
<td>KD</td>
<td>lysine/aspartic acid-rich</td>
</tr>
<tr>
<td>KE</td>
<td>lysine/glutamic acid-rich</td>
</tr>
<tr>
<td>KNS1</td>
<td>kinase next to SPA2</td>
</tr>
<tr>
<td>LBR</td>
<td>lamin B receptor</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
</tr>
<tr>
<td>MBP</td>
<td>myelin basic protein</td>
</tr>
<tr>
<td>N-CoR</td>
<td>nuclear co-repressor</td>
</tr>
<tr>
<td>NK</td>
<td>natural killer</td>
</tr>
<tr>
<td>NK-TR1</td>
<td>natural killer tumour recognition 1</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localization signal</td>
</tr>
<tr>
<td>NOPP140</td>
<td>nucleolar phosphoprotein of 140 kDa</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PF</td>
<td>perichromatin fibril</td>
</tr>
<tr>
<td>PKC</td>
<td>protein kinase C</td>
</tr>
<tr>
<td>PKI</td>
<td>protein kinase inhibitor</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethysulphonylfluoride</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PP1</td>
<td>protein phosphatase 1</td>
</tr>
<tr>
<td>PP2A</td>
<td>protein phosphatase 2A</td>
</tr>
<tr>
<td>RBP1</td>
<td>RNA-binding protein 1</td>
</tr>
<tr>
<td>RD</td>
<td>arginine/aspartic acid</td>
</tr>
<tr>
<td>RE</td>
<td>arginine/glutamic acid-rich</td>
</tr>
<tr>
<td>RNP</td>
<td>ribonucleoprotein</td>
</tr>
<tr>
<td>RS</td>
<td>arginine/serine-rich</td>
</tr>
<tr>
<td>RRM</td>
<td>RNA recognition motif</td>
</tr>
<tr>
<td>SAF</td>
<td>scaffold attachment factor</td>
</tr>
<tr>
<td>SAR</td>
<td>scaffold attachment region</td>
</tr>
<tr>
<td>SC35</td>
<td>spliceosome component of 35 kDa</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SELEX</td>
<td>systematic evolution of ligands by exponential enrichment</td>
</tr>
<tr>
<td>SH2</td>
<td>Src homology 2</td>
</tr>
<tr>
<td>snRNA</td>
<td>small nuclear RNA</td>
</tr>
<tr>
<td>snRNP</td>
<td>small nuclear ribonucleoprotein</td>
</tr>
<tr>
<td>SR</td>
<td>serine/arginine-rich</td>
</tr>
<tr>
<td>SRPK1</td>
<td>SR protein kinase 1</td>
</tr>
<tr>
<td>su(w*)</td>
<td>suppressor of white apricot</td>
</tr>
<tr>
<td>sxl</td>
<td>sex-lethal</td>
</tr>
<tr>
<td>TLC</td>
<td>thin layer cellulose</td>
</tr>
<tr>
<td>tra</td>
<td>transformer</td>
</tr>
<tr>
<td>tra-2</td>
<td>transformer-2</td>
</tr>
<tr>
<td>U2AF</td>
<td>U2 auxilliary factor</td>
</tr>
<tr>
<td>w</td>
<td>white eye</td>
</tr>
<tr>
<td>w*</td>
<td>apricot allele at the white eye locus</td>
</tr>
<tr>
<td>WT</td>
<td>wild-type</td>
</tr>
<tr>
<td>ZFP</td>
<td>zinc finger protein</td>
</tr>
</tbody>
</table>
# Table of Contents

Abstract ................................................................................................................. i  
Acknowledgments ...................................................................................................... iii  
Abbreviations ............................................................................................................ iv  
Table of Contents ...................................................................................................... vi  
Figures ..................................................................................................................... x  
Tables ....................................................................................................................... xii

CHAPTER 1: General Introduction .................................................................................. 1  
Part I: Kinase Introduction ............................................................................................. 2  
Section I: Kinase structure and the catalytic reaction ................................................. 2  
  Residues involved in MgATP binding ........................................................................ 3  
  The catalytic reaction .............................................................................................. 7  
  Kinase activation ................................................................................................... 7  
Section II: Substrate specificity ...................................................................................... 8  
  Substrate consensus sites ....................................................................................... 8  
  Serine/threonine versus tyrosine specificity .......................................................... 9  
  Dual specificity ..................................................................................................... 10  
Section III: Physiological roles for dual specificity kinases ........................................ 11  
Section IV: The Clk/Sty protein kinase ......................................................................... 12  
  Dual specificity function of Clk/Sty ....................................................................... 13  
  Transcriptional regulation of Clk/Sty .................................................................... 13  
Section V: The LAMMER family of protein kinases .................................................... 14  
  Function of the LAMMER family kinases .............................................................. 15  
Part II: Pre-mRNA Splicing Introduction .................................................................... 18  
Section I: Cis-acting elements and the catalytic reaction ......................................... 18  
Section II: U snRNPs and the spliceosome cycle ..................................................... 20  
Section III: Factors involved in splice site selection .................................................. 22  
  U2 auxiliary factor (U2AF) .................................................................................... 22  
  SR family ........................................................................................................... 23  
  Differences between SR family members ........................................................... 26  
  hnRNP A1 ........................................................................................................... 27  
Section IV: Structural motifs of SR proteins ............................................................... 28  
  RNA recognition motifs (RRMs) ......................................................................... 28  
  RS domains ........................................................................................................ 29  
Section V: Spliceosome assembly .............................................................................. 32  
  Assembly at the 5' Splice Site .............................................................................. 33  
  Selection of the 3' Splice Site .............................................................................. 34  
Section VI: Somatic sex determination in *Drosophila* ............................................... 36  
  Initiation of the sex determination signal ............................................................ 36  
  Sex-lethal (Sxl) .................................................................................................. 38  
  Transformer (Tra) ............................................................................................... 39  

vi
Transformer 2 (Tra-2) .................................................. 39
Doublesex (Dsx) ..................................................... 40
Section VII: Mammalian purine-rich enhancers ..................... 41
Section VIII: Location of splicing in the nucleus .................. 43
  Co-transcriptional splicing ...................................... 43
  Nuclear speckles ................................................ 43
  Link between transcription and splicing .......................... 44
Section IX: Regulation of splicing by phosphorylation .......... 45
  Phosphorylation-dephosphorylation cycle of splicing ........ 45
  Targets of phosphorylation-dephosphorylation ............... 46
  SR protein kinases .............................................. 47

Part III: Thesis Overview .............................................. 49

CHAPTER 2: The Clk/Sty Protein Kinase Phosphorylates SR Splicing Factors and Regulates their Intranuclear Distribution .......... 50

Introduction .......................................................... 51
Materials and Methods .............................................. 51
  DNA subcloning .................................................. 51
  Two-hybrid screen ............................................... 52
  Sequencing of clones .......................................... 53
  Protein purification ............................................. 53
  In vitro kinase reaction ....................................... 54
  In vivo labelling ................................................ 54
  Phosphoamino acid analysis .................................. 54
  Tryptic mapping ................................................ 55
  Immunofluorescence assay ..................................... 55
  Triton X-100 extraction ...................................... 55

Results ................................................................. 56
  The Clk/Sty kinase interacts with splicing factors ............ 56
  Analysis of protein domains required for interactions of Clk/Sty with SR proteins ....................................... 59
  Phosphorylation of ASF/SF2 .................................... 60
  Specificity of Clk/Sty protein kinase activity .................. 62
  Phosphoamino acid analysis of ASF/SF2 ........................ 64
  Comparative tryptic phosphopeptide mapping of ASF/SF2 .... 64
  Subnuclear localization of Clk/Sty and SR proteins in vivo ... 67

Discussion ............................................................... 70
  Clk/Sty interacts with RNA-binding proteins ................. 70
  ASF/SF2 is a candidate physiological substrate of Clk/Sty ... 73
  Comparison of Clk/Sty and SRPK1 .............................. 74
  Regulation of SR splicing factors by Clk/Sty in vivo .......... 76

CHAPTER 3: SRPK1 and Clk/Sty Protein Kinases Show Distinct Substrate Specificity for SR Splicing Factors ........................................... 79
Isolation of CARS-Cyp cDNA clones .............................................. 114
Sequencing of CARS-Cyp .......................................................... 114
Northern analysis ........................................................................ 114
Results ......................................................................................... 114
Isolation of the CARS-Cyp cDNA .................................................... 114
Analysis of CARS-Cyp expression .................................................... 118
Discussion ..................................................................................... 120
A family of RS-Cyclophilins .............................................................. 120
RS-Cyclophilins are multi-domain proteins ........................................ 120
CARS-Cyp interacts with the Clk/Sty kinase ...................................... 122
Part III: Analysis of the Two-hybrid Clones ...................................... 123
Introduction ................................................................................... 123
Materials and Methods .................................................................. 123
Results .......................................................................................... 123
The two-hybrid library .................................................................... 123
Clk/Sty constructs in the two-hybrid screen ...................................... 123
Identity of the two-hybrid clones ...................................................... 125
Clones 3.9 and 5.3 are SAF-B ......................................................... 125
Clone 5.7 is KIAA0164 .................................................................... 126
Clone 6.12 is N-CoR ..................................................................... 128
Clone 8.10 is similar to an acidic 82 kDa protein .............................. 129
Clone 2.8 is c-Myc ........................................................................ 129
Clone 4.11 is ZFP38 ...................................................................... 130
Clone 5.12 is a novel zinc finger protein ......................................... 130
Discussion ...................................................................................... 130
Clk/Sty interaction sequences ......................................................... 130
Clk/Sty interaction sequences ......................................................... 132

Chapter 5: General Discussion and Future Directions ...................... 134
The two-hybrid screen ................................................................... 135
Identification of interaction domains ................................................. 136
Relevance of the two-hybrid clones .................................................. 137
Specificity of SR protein phosphorylation ....................................... 141
The functional role of Clk/Sty dual specific kinase activity ................. 143
Future Directions .......................................................................... 143
Two-hybrid clones ......................................................................... 143
Endogenous Clk/Sty ...................................................................... 144
References ..................................................................................... 146
Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Ribbon diagram of the catalytic subunit of cAPK.</td>
<td>4</td>
</tr>
<tr>
<td>1.2</td>
<td>Alignment of the catalytic domains of cAPK, CDK2, ERK2, InsR, Clk/Sty, and MEK.</td>
<td>5</td>
</tr>
<tr>
<td>1.3</td>
<td>Schematic representation of the kinase active site.</td>
<td>6</td>
</tr>
<tr>
<td>1.4</td>
<td>Consensus sites and pre-mRNA splicing.</td>
<td>19</td>
</tr>
<tr>
<td>1.5</td>
<td>The spliceosome cycle.</td>
<td>21</td>
</tr>
<tr>
<td>1.6</td>
<td>Structure of the RRM1 of U1A.</td>
<td>30</td>
</tr>
<tr>
<td>1.7</td>
<td>SR proteins mediate a network of interactions in the assembling spliceosome.</td>
<td>35</td>
</tr>
<tr>
<td>1.8</td>
<td>Somatic sex determination pathway of <em>Drosophila</em>.</td>
<td>37</td>
</tr>
<tr>
<td>2.1</td>
<td>A schematic of Clk/Sty and five interacting RNA-binding proteins.</td>
<td>58</td>
</tr>
<tr>
<td>2.2</td>
<td>Phosphorylation of ASF/SF2.</td>
<td>61</td>
</tr>
<tr>
<td>2.3</td>
<td>Specificity of the phosphorylation of ASF/SF2 by Clk/Sty.</td>
<td>63</td>
</tr>
<tr>
<td>2.4</td>
<td>Phosphoamino acid analysis of ASF/SF2.</td>
<td>65</td>
</tr>
<tr>
<td>2.5</td>
<td>Tryptic phosphopeptide maps of ASF/SF2.</td>
<td>66</td>
</tr>
<tr>
<td>2.6</td>
<td>Nuclear localization of Clk/Sty and its effects on the distribution of SR proteins.</td>
<td>68</td>
</tr>
<tr>
<td>2.7</td>
<td>Active and inactive Clk/Sty kinases display differential solubility in non-ionic detergent.</td>
<td>71</td>
</tr>
<tr>
<td>2.8</td>
<td>Alignment of the LAMMER motif.</td>
<td>75</td>
</tr>
<tr>
<td>3.1</td>
<td>Phosphorylation of arginine/serine-rich proteins by Clk/Sty and SRPK1.</td>
<td>84</td>
</tr>
<tr>
<td>3.2</td>
<td>Relative specific activities of Clk/Sty and SRPK1.</td>
<td>86</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Phosphorylation of ASF/SF2 by Clk/Sty generated the mAb104 phosphoepitope.</td>
<td>89</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Tryptic phosphopeptide maps of ASF/SF2.</td>
<td>90</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Tryptic phosphopeptide maps of ASF/SF2, RSRS, and SPRY.</td>
<td>94</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Comparison of SRPK1 and Clk/Sty phosphorylation of ASF/SF2 mutants.</td>
<td>96</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Sequence of H11792 and clone 6.9.</td>
<td>106</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>A schematic of Tra-2 and mammalian cDNAs of clone 6.9.</td>
<td>107</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Alignment of the RRM of Tra-2 with the RRM from the composite of H11792 and clone 6.9.</td>
<td>109</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Isolation of CARS-Cyp cDNA clones.</td>
<td>116</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Sequence of CARS-Cyp.</td>
<td>117</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Northern blot of human peripheral blood mRNA.</td>
<td>119</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Schematic representation of the clones isolated in the two-hybrid screen.</td>
<td>127</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>Sequence of the 5' end of clone 5.12.</td>
<td>131</td>
</tr>
</tbody>
</table>
Tables

Table 3.1  Kinetic parameters for ASF/SF2, RSRS, and SPRY after phosphorylation by Clk/Sty or SRPK1. ....................... 87

Table 4.1  Characteristics of the two-hybrid clones. ....................... 124
CHAPTER 1: General Introduction

Part I: Kinase Introduction

Part II: Pre-mRNA Splicing Introduction
Part I: Kinase Introduction

The development and viability of multicellular organisms depends on complex regulatory pathways that control cell-cell communication. Cells communicate either by physical association of receptors on their surfaces, or by releasing factors that are transported to the receiving cell that bears receptors for that ligand. Once a cell receives a signal, it must be relayed to the nucleus where it induces cellular response.

Frequently, signals are transduced by phosphorylation of downstream proteins. Protein kinases are responsible for this modification and are divided into several classes depending on the amino acid that they phosphorylate (Hunter, 1991; Lindberg et al., 1992). Three classes of protein kinases phosphorylate hydroxyamino acids: i) serine/threonine kinases phosphorylate the hydroxyl group on the β carbon of serine and threonine. ii) tyrosine kinases phosphorylate the hydroxyl group on the phenolic ring of tyrosine and iii) dual specificity kinases phosphorylate all three hydroxyamino acids.

Section I: Kinase structure and the catalytic reaction

Kinases recognize the amino acid to be phosphorylated in the context of surrounding amino acids. To understand how a kinase works and why it phosphorylates a particular residue, a detailed analysis of kinase structure is required. An alignment of kinase domains has identified amino acids that are conserved among protein kinases (Hanks et al., 1988). The first crystal structure of a kinase domain from cAMP dependent protein kinase (cAPK) demonstrated key roles in kinase function for these amino acids (Knighton et al., 1991a; Knighton et al., 1991b; Bossemeyer et al., 1993). Since then, crystal structures of other kinase domains have been
solved including those of CDK2, ERK2, and the insulin receptor kinase (InsR) (De Bondt et al., 1993; Zhang et al., 1994; Hubbard et al., 1994).

As expected from the high degree of similarity among kinases, the overall structure of the catalytic domain is conserved. The kinase domain is a bi-lobed structure with the small amino (N) terminal lobe separated from the large carboxy (C) terminal lobe by a linker region (figure 1.1). The cleft between the two lobes forms the catalytic site.

**Residues involved in MgATP binding**

The γ-phosphate of ATP is added to the substrate being phosphorylated. Residues found primarily in the N-terminal lobe anchor the nucleotide so that the γ-phosphate is in correct position for phosphotransfer (Bossemeyer et al., 1993). The adenine base and the ribose ring of ATP are buried in a hydrophobic pocket at the base of the cleft. The position of the base and ring is further stabilized by hydrogen bonds with residues in the kinase domain (i.e., E127 and E170 with the ribose ring, see figure 1.2 for cAPK sequence).

The conserved nucleotide binding motif, GXGXXG (amino acids 50 to 55), lies between β strands 1 and 2 in the phosphate anchor loop (Hanks et al., 1988; Bossemeyer et al., 1993). As the name implies, this β-turn β structure acts as a clamp to anchor the non-transferable α- and β-phosphates (Bossemeyer et al., 1993). K72, which resides in β strand 3, aids in orientation of the γ-phosphate by hydrogen bonding with the α- and β-phosphates (figure 1.3) (Bossemeyer et al., 1993). This interaction is stabilized by a salt bridge formed between K72 and E91 of helix C. D184, of the conserved DFG sequence, chelates the Mg2+. It also helps orient the γ-phosphate for correct transfer, perhaps through interactions with K72 and E91.
Figure 1.1. **Ribbon diagram of the catalytic subunit of cAPK.** The structure is bi-lobed with the N-terminal lobe at the top, and the C-terminal lobe at the bottom. β strands (coloured blue and pale green) are numbered 1 to 9, and helices (coloured pink, red, and purple) are labelled A to H. The side chains of D166, D184, E208, and R280 are shown in black. The Mg$^{2+}$ ion that binds the β- and γ-phosphates is depicted as a black dot. ATP is shown in black. Phosphorylated T197 is indicated by a yellow arrow. The peptide inhibitor PKI is shown in dark green. Functional loops are labelled and shown in yellow. This figure was reproduced from (Taylor and Radzio-Andzelm, 1994).
Figure 1.2. Alignment of the catalytic domains of cAPK, CDK2, ERK2, InsR, Clk/Sty, and MEK. At the top, the secondary structure of cAPK to which the amino acids correspond is noted (Knighton et al., 1991a). The various subdomains as outlined by Hanks et al. (1988) are shown at the bottom. Residues conserved among all six kinases are highlighted in bold reverse. Residues that are 80 to 99% conserved are bolded. GenBank accession numbers - cAPK - A25125. CDK2 - X61622, ERK2 - X58712, InsR - A37348, Clk/Sty - X57186, MEK - L02526. Abbreviations: α - alpha helix, β - beta strand, Phos. - phosphate anchor loop, Cat. - catalytic loop.
Figure 1.3. **Schematic representation of the kinase active site.** The probable mechanism for phosphotransfer is illustrated. K72, E91, and D184 are critical for the proper position of the $\gamma$ phosphate of ATP. D166, as the catalytic base, abstracts the proton from the serine to trigger the phosphotransfer. Also shown is the interaction of phosphorylated T197 with K189 and R165. This figure is reproduced from (Johnson et al., 1996).
The catalytic reaction

The residues involved in catalysis help orient the γ-phosphate and the hydroxyamino acid so phosphotransfer can occur. The catalytic loop (residues D₁₆₆LKPEN) is located between β strands 6 and 7. The two invariant residues in this loop are D₁₆₆ and N₁₇₁. D₁₆₆ is the catalytic base, and is positioned near the γ-phosphate and the hydroxyamino acid that will receive the phosphate. The side chain of N₁₇₁ hydrogen bonds to the backbone carbonyl of D₁₆₆ to stabilize the loop. D₁₈₄, that is near K₇₂ and E₉₁ in the tertiary structure of cAPK, also interacts with N₁₇₁ and D₁₆₆.

Interactions occur at the active site between the kinase and MgATP, the kinase and substrate, and the substrate and MgATP (Bossemeyer et al., 1993). These interactions refine the orientation of the γ-phosphate and the amino acid that will receive the phosphate group (amino acid at the P position). The reaction likely proceeds as an in-line transfer mechanism with D₁₆₆ as the catalytic base (figure 1.3) (Ho et al., 1988; Yoon and Cook, 1987). The catalytic base accepts the proton from the hydroxyl group of the P amino acid. At the same time, the γ-phosphate of ATP undergoes nucleophilic attack by the oxygen on the P amino acid. K₁₆₈, of the catalytic loop, neutralizes the γ-phosphate during the transfer (Bossemeyer et al., 1993).

Product release is likely driven by repulsion of the phosphoserine by D₁₆₆ and the β-phosphate.

Kinase activation

cAPK is fully activated upon phosphorylation of T₁₉₇ in the activation loop (Shoji et al., 1979, Adams et al., 1995). This loop follows β sheet 9, and forms part of the outer edge of the cleft on the C-terminal lobe (Knighton et al., 1991a). Once phosphorylated, T₁₉₇ resides in a pocket formed by R₁₆₅, K₁₈₉, H₈₇, and N₉₀ (Knighton et al., 1991b; Taylor and
Radzio-Andzelm, 1994) (figure 1.3). The interaction of T197 with this pocket helps stabilize the kinase in several ways. First, binding of phosphorylated T197 to R165 stabilizes the P+1 (amino acid directly C-terminal to the P position) binding site. Second, the interaction of T197 with H87 and N90 of the N-terminal lobe provides an additional link between the two lobes. Finally, K189 is near D184 that chelates Mg$^{2+}$. The binding of T197 to K189 may alter the conformation of D184.

Kinase activation by phosphorylation within the activation loop is a common phenomenon. Kinases that contain an arginine preceding the catalytic base, aspartate, are referred to as RD kinases (Johnson et al., 1996). RD kinases require ionic interactions of the arginine with a phosphate or a carboxylate group. The phosphate group is usually provided by a hydroxyamino acid within the activation loop. RD kinases CDK2, ERK2, and InsR are phosphorylated in the activation loop (Morgan, 1995; Anderson et al., 1990; Hubbard et al., 1994). For the above three kinases, phosphorylation of this loop removes it from the active site (De Bondt et al., 1993; Zhang et al., 1994; Hubbard et al., 1994). CDK2, ERK2, and InsR contain residues analogous R165, K189, H87, and N90 of cAPK that form a binding pocket for phosphorylated T167 (figure 1.2). Thus, cAPK, CDK2, ERK2, InsR, and likely the majority of RD kinases require phosphorylation within their activation loop to become fully active.

Section II: Substrate specificity

Substrate consensus sites

The consensus site for phosphorylation by cAPK is RRXS/TY where S/T is the site of phosphorylation, and Y is a large hydrophobic amino acid (Taskén et al., 1995). cAPK was crystallized with a 20 amino acid peptide inhibitor, PKI, that contains the consensus sequence
where alanine substitutes for serine at the P position (Knighton et al., 1991b).

The residues in cAPK involved in substrate binding are spread across the large lobe (Knighton et al., 1991b). Interactions between cAPK and PKI involve ionic and hydrophobic interactions. For example, ionic interactions include ion pairs formed between the P-2 and P-3 arginines, and kinase residues E127 and E170. A prominent hydrophobic interaction occurs at the P+1 loop. Residues within this loop, L198, C199, P202, and L205 form a pocket for the isoleucine at the P+1 position. Invariant residues E208 and R280 form an ion pair underneath the P+1 binding loop to stabilize this loop, and the large lobe in general.

Differences between kinase sequences in the P+1 loop may reflect substrate specificity. For example, both CDK2 and ERK2 are proline-directed kinases, i.e., they recognize proline at position P+1 (Harper and Elledge, 1995; Cobb, 1995). In the P+1 loop of both CDK2 and ERK2, the residue analogous to L205 of cAPK is replaced with arginine (figure 1.2). As well, G200 and P202 are replaced with alanine and arginine (ERK2), or valine and leucine (CDK2). These bulkier amino acids occlude the P+1 binding site of these kinases, allowing only small residues such as proline to bind (Zhang et al., 1994).

Serine/threonine versus tyrosine specificity

The solution of the cAPK and InsR structures facilitated comparisons between serine/threonine and tyrosine kinases. Three apparent differences between these two kinases may contribute to substrate specificity. First, the InsR displays a more open conformation than cAPK (Hubbard et al., 1994; Knighton et al., 1991a). As a result, K1030 (K72 of cAPK) is farther away from the catalytic base in the insulin receptor (see figure 1.2 for InsR versus cAPK sequence). Second, these kinases differ in the amino acid composition of the catalytic loop.
KPE (amino acids 168 to 170 of cAPK) is found in serine/threonine kinases, and AAR or RAA is present in tyrosine kinases (Hanks et al., 1988). In cAPK, K168 neutralizes the γ-phosphate (Bossemeyer et al., 1993). For InsR, R1136 of the AAR motif provides the neutralizing charge (Hubbard et al., 1994). This difference within the active site of InsR is likely to be important in accommodating tyrosine at this site.

Lastly, the final difference is the position of the P+1 loop in serine/threonine versus tyrosine kinases. In the InsR, the P+1 loop is linked to the catalytic loop by a direct interaction between W1175 and R1136, and an indirect interaction between W1175, E1201 and A1135 (Hubbard et al., 1994). E1201 and W1175 are well conserved in the tyrosine kinase family (Hanks et al., 1988). In serine/threonine kinases, E1201 is less conserved, and W1175 is usually replaced with tyrosine. In cAPK, the link between the P+1 loop and the catalytic loop is provided by the interaction of T201 with D166 and K168 (Knighton et al., 1991a). In the tyrosine kinase family, T201 is replaced by proline (Hanks et al., 1988). As a result of the different links employed, the P+1 loop is further away from the catalytic loop in InsR than in cAPK. The more open conformation of InsR, and the looser link between the P+1 and the catalytic loops create a larger active site that is conducive to tyrosine binding.

Dual specificity

The differences between the active sites of cAPK and InsR suggest that serine/threonine and tyrosine phosphorylation are mutually exclusive. Yet, dual specificity kinases exist. Phosphorylation of tyrosine by these kinases may be an aberrant reaction that is kinetically and sterically favoured in autophosphorylation (Lindberg et al., 1992). Many dual specificity kinases phosphorylate the exogenous substrate poly(glut) (Lindberg et al., 1992). However, cAPK and
calmodulin-dependent kinase also phosphorylate poly(glu,tyr) indicating that poly(glu,tyr) is not a good substrate to verify dual specificity (Stern et al., 1991).

Whether catalyzing autophosphorylation or trans-phosphorylation, the active site of dual specificity kinases must be flexible enough to allow phosphotransfer to both serine/threonine and tyrosine. The kinase domains of dual specificity kinases most closely resemble serine/threonine kinases (see MEK, ERK2, and Clk/Sty in figure 1.2). In many dual specificity kinases, the last residue of the KPE motif of the catalytic loop is replaced by a small hydrophobic amino acid (Lindberg et al., 1992). In cAPK, the glutamic acid in the KPE motif interacts with the substrate and with ATP (Knighton et al., 1991b; Bossemeyer et al., 1993). Perhaps the loss of contacts mediated by the glutamic acid in dual specificity kinases expands the active site to allow for tyrosine at the P position. Apart from this glutamic acid, an alignment of eleven putative dual specificity kinases did not reveal any amino acids or short motifs unique to dual specificity kinases (Lindberg et al., 1992). Therefore, there is no obvious universal mechanism that accounts for dual specificity. Crystal structures of dual specificity kinases may reveal important residues that are not obvious from primary structure alignment.

Section III: Physiological roles for dual specificity kinases

Dual specificity kinases autophosphorylate serine/threonine and tyrosine residues, but very few of these kinases phosphorylate exogenous substrates on tyrosine (Lindberg et al., 1992). MEK is the best example of a true dual specificity kinase. MEK activates ERK1 or ERK2 by phosphorylating them on both threonine and tyrosine (Crews et al., 1992; Nakielny et al., 1992a; Nakielny et al., 1992b). A few dual specificity kinases phosphorylate exogenous substrates on tyrosine, but not on serine or threonine. For example, Wee1 phosphorylates p34<sup>csk</sup> on tyrosine.
but not on threonine (Featherstone and Russell, 1991; Parker et al., 1992).

Other dual specificity kinases may phosphorylate substrates on tyrosine, but these substrates have not been identified yet. Alternately, an ability to autophosphorylate on tyrosine may be the reason for dual specificity. For example, GSK3\(\beta\) is a dual specificity kinase that autophosphorylates on serine/threonine and tyrosine (Wang et al., 1994). In this case, serine/threonine phosphorylation inactivates the kinase, and tyrosine phosphorylation results in increased activity. As well, autophosphorylation on tyrosine may be required for protein-protein interactions. Phosphotyrosines are docking sites for SH2 domains (Koch et al., 1991; Pawson and Gish, 1992). Dual specificity kinases may interact with SH2-containing proteins via their phosphotyrosine motif (Douville et al., 1994).

**Section IV: The Clk/Sty protein kinase**

Murine Clk/Sty (CDC28 Like Kinase/Serine-Threonine-Tyrosine) was isolated in an anti-phosphotyrosine antibody screen of a Friend erythroleukemia cDNA expression library as well as a P19 embryonal carcinoma cDNA expression library (Ben-David et al., 1991; Howell et al., 1991). In this screen, phage-infected bacteria that contain kinases that are autophosphorylated on tyrosine are identified using anti-phosphotyrosine antibodies (Letwin et al., 1988). Since this screen is not dependent on primary structure, it allows identification of tyrosine kinases that differ in sequence from known tyrosine kinases. Indeed, Clk/Sty is more related to serine/threonine kinases than to tyrosine kinases (figure i.2).

Clk/Sty is a 483 amino acid protein with an estimated molecular weight of 57 kDa. The first 156 amino acids of Clk/Sty are non-catalytic. This N-terminal region has no significant homology to other proteins but it has two important features. First, this region contains a
potential SV40 nuclear localization signal that is likely responsible for the presence of Clk/Sty in the nucleus (Howell et al., 1991; Duncan et al., 1995). Second, this domain is rich in arginine, lysine, and serine.

**Dual specificity function of Clk/Sty**

Since Clk/Sty was isolated in an anti-phosphotyrosine screen, it was presumed to be a tyrosine kinase; however, its resemblance to serine/threonine kinases suggested that it phosphorylated serine and threonine. In actuality, Clk/Sty phosphorylates all three amino acids. Clk/Sty, purified from bacteria or mammalian cells, auto phosphorylates on serine, tyrosine, and to a lesser extent on threonine (Ben-David et al., 1991; Howell et al., 1991; Duncan et al., 1995). Whether this autophosphorylation on tyrosine regulates the kinase activity of Clk/Sty or controls interactions with other proteins is not known. Thus far, the only exogenous substrate that has been shown to be phosphorylated on tyrosine by Clk/Sty is poly(glu,tyr) (Ben-David et al., 1991).

**Transcriptional regulation of Clk/Sty**

Northern blot analysis revealed that the Clk/Sty transcript is expressed in every mouse tissue examined as well as in malignant cell lines (Ben-David et al., 1991; Howell et al., 1991). There are four Clk/Sty transcripts of sizes 5.6, 3.2, 1.8 and 1.7 kb (Duncan et al., 1995). The 5.6 and 3.2 kb transcripts correspond to immature transcripts that are retained in the nucleus. The 1.8 and 1.7 kb transcripts lack introns and are present in the cytoplasm, where presumably they are translated. The difference between the two small transcripts is the presence of exon B. The 1.8 kb RNA retains exon B, and codes for the full-length protein of 483 amino acids. Exon B is absent from the 1.7 kb transcript. The joining of exon A to exon C in the 1.7 kb transcript causes a frameshift in the coding sequence. Translation of this mRNA results in a protein that
lacks the catalytic domain.

In mouse tissues, the larger transcripts display higher abundance than the smaller transcripts (Ben-David et al., 1991; Howell et al., 1991). In contrast, the smaller transcripts predominate in rapidly dividing erythroleukemia cell lines. A different pattern of transcripts also occurs upon differentiation of P19 cells when more 3.2 and 5.6 kb transcripts appear (Howell et al., 1991). Therefore, Clk/Sty is regulated at the level of pre-mRNA processing such that rapidly dividing cells produce relatively more mature transcripts than differentiating cells.

Section V: The LAMMER family of protein kinases

Clk/Sty is a member of the LAMMER family of protein kinases (Yun et al., 1994). Presently, there are nine members of the LAMMER family. These kinases are: Clk (mouse and human) (Ben-David et al., 1991; Howell et al., 1991; Johnson and Smith, 1991), Clk2 (human) (Hanks et al., 1991; Hanes et al., 1994), Clk3 (human) (Hanes et al., 1994), Doa (Drosophila melanogaster) (Yun et al., 1994), AFC1, AFC2, AFC3 (Arabidopsis thaliana) (Bender and Fink, 1994), and KNS1 (Saccharomyces cerevisiae) (Padmanabha et al., 1991). The N-terminal non-catalytic region of these kinases is not conserved, although in the three Clks and Doa the region is rich in arginine and serine. In contrast, the kinase domain at the C-terminus is well conserved among family members. Compared to murine Clk/Sty, the percentage identities across the LAMMER kinase domain are: human Clk (92\%), Clk2 (66\%), Clk3 (60\%), Doa (62\%), AFC1 (44\%), AFC2 (47\%), AFC3 (48\%), and KNS1 (44\%).

The LAMMER kinases likely constitute a functionally relevant kinase subfamily given the similarity of their kinase domains and the conservation of key residues within this domain. LAMMER family members are not RD kinases, as threonine is substituted for the well conserved
arginine (R165 of cAPK) in the catalytic loop. Members of the LAMMER family also lack the equivalent of K189, H87 and N90 that together with R165 form a pocket in cAPK for phosphorylated T197 (Knighton et al., 1991a). Members of the LAMMER family do contain hydroxyamino acids in their activation loop. If the LAMMER family is regulated by phosphorylation in this loop, then the phosphorylated residue must fit into a different pocket than the one used for RD kinases. This pocket may influence substrate specificity through alignment of the active site.

The P+1 loop is well conserved in LAMMER kinases and has the sequence LVSTRHYRAPE. In ERK2, the arginine preceding APE occludes the P+1 pocket so that only small amino acids fit in that site (Zhang et al., 1994; Songyang et al., 1994). Therefore, the LAMMER family members may bind proline, or a small residue at the P+1 site. Finally, the kinases received their name from a unique EHLAMMERILGPLP motif found in subdomain X (Yun et al., 1994). This motif maps to helix G of cAPK (figure 1.1 and 1.2). This helix is found below the binding cleft and may contact the substrate (Knighton et al., 1991b).

**Function of the LAMMER family kinases**

Little is known about the LAMMER family. Yeast KNS1 (kinase next to SPA2) was isolated as an open reading frame that lies near the SPA2 locus (Padmanabha et al., 1991). Yeast lacking KNS1 are viable with no obvious phenotype. Therefore, KNS1 may be redundant with another kinase, or may have a specific function not tested for in its initial characterization.

AFC1 (*Arabidopsis fus3* complementing cDNA) was isolated in a screen to identify MAP kinases in *Arabidopsis* (Bender and Fink, 1994). Fus3 and Kss1 are MAP kinases involved in the response to pheromone stimulation in budding yeast *S. cerevisiae* (Elion et al., 1990);
Courchesne et al., 1989). Yeast strains deleted for these two kinases are sterile (Elion et al., 1991). AFC1 was identified as a clone from an Arabidopsis cDNA library that complemented the sterile phenotype of a fus3/kss1 double mutant (Bender and Fink, 1994). Further analysis revealed that AFC1 did not rescue the fus3/kss1 phenotype, but constitutively activated the downstream Ste12 transcription factor. AFC2 and AFC3 were isolated by degenerate PCR using primers for AFC1 and homologous kinases from other organisms. Since AFC1 activates Ste12 in yeast, it is possible that AFC kinases regulate transcription in Arabidopsis.

Drosophila Doa (Darkener of apricot) is the best characterized LAMMER family member from a genetic perspective. Flies homozygous for a Doa loss-of-function mutation die at the early larval stage indicating that Doa is essential for development (Rabinow and Birchler, 1989). Doa is a second site modifier of the apricot allele (w+) at the white eye locus (w) (Rabinow and Birchler, 1989). Eye pigment is reduced in w+ mutants because of a copia retrotransposon insertion in the second intron (Bingham and Judd, 1981). Few wild-type w transcripts are produced because most of the transcripts terminate within the long-terminal repeats (LTRs) of the copia insert (Levis et al., 1984; Pirrotta and Brockl, 1984; Zachar et al., 1985). One model for Doa function is that Doa regulates the transcription complex that assembles on the copia LTRs (Rabinow et al., 1993). If Doa is absent, the complex does not bind to the copia LTRs, and RNA polymerase II can read through this insertion to generate more wild-type transcripts. Therefore, Doa, like AFC1, may be linked to transcriptional regulation.

Human Clk was isolated with a degenerate oligonucleotide probe derived from a semi-conserved QSNVWSIGV sequence in subdomain IX (Johnson and Smith, 1991). Human Clk is 87% identical to murine Clk/Sty across the entire protein, and 92% identical across the kinase
domain. Ctk2 was first isolated in a PCR screen used to identify novel kinases (Hanks et al., 1991) and named PSK-G1. Later, Ctk2 and Ctk3 were isolated fortuitously in a 5' RACE experiment designed to complete the human APLP2 cDNA sequence (Hanes et al., 1994). Like murine Ctk/Sty, all three human Ctk genes encode two protein isoforms, one of which lacks the catalytic domain (Hanes et al., 1994).

A role for Ctk/Sty in the differentiation of rat PC12 cells has been suggested, particularly after cells are stimulated with nerve growth factor (Myers et al., 1994). In these cells, expression of Ctk/Sty lead to increased activity of ERK1, ERK2, and pp90RSK. Since Ctk/Sty is located in the nucleus (Duncan et al., 1995), Ctk/Sty may activate the small fractions of ERK1, ERK2, and pp90RSK that are also present in the nucleus (Chen et al., 1992). In this thesis, I demonstrate that Ctk/Sty lies upstream of the SR family of splicing factors and functions in regulation of their cellular localization. Therefore, activation of the above kinases by Ctk/Sty may occur at the level of pre-mRNA splicing rather than by phosphorylation.
Part Two: Pre-mRNA Splicing Introduction

“Split genes” that contain protein-coding sequences (exons) and non-protein coding sequence (introns) allow for protein diversity at the level of pre-mRNA splicing (Berget et al., 1977; Chow et al., 1977). Retention of introns, exon skipping, and differential use of exon/intron boundaries (known as splice sites) are just a few examples of how alternative pre-mRNA splicing is utilized to generate multiple protein isoforms (Smith et al., 1989a).

Section I: Cis-acting elements and the catalytic reaction

Pre-mRNA splicing must occur faithfully to ensure that transcripts are processed accurately. Correct splice site selection requires cis-acting elements on the pre-mRNA, and trans-acting factors that recognize these elements. Together, the trans-acting elements and the pre-mRNA make up the spliceosome (Brody and Abelson, 1985).

Pre-mRNA intron sequences have diverged over evolution with the exception of sequences flanking exon/intron junctions (Moore et al., 1993). The 5' splice site, or splice donor, is found at the end of the upstream exon. The 3' splice site, or splice acceptor, of the downstream exon is divided into three sections: the actual 3' splice site, the polypyrimidine tract, and the branchpoint (figure 1.4).

Catalysis of pre-mRNA splicing is a two step transesterification reaction that employs an in-line transfer mechanism (figure 1.4) (Padgett et al., 1984; Ruskin et al., 1984; Moore et al., 1993). In step 1, the phosphate of the guanylate at position +1 of the intron undergoes nucleophilic attack by the 2' hydroxyl group of the adenylate at the branchpoint. A free exon one and a lariat intermediate are formed. In step 2, the free hydroxyl group of the guanylate in
Figure 1.4. Consensus sites and pre-mRNA splicing. a) The mammalian consensus sequences for the 5' and 3' splice sites are indicated. Underlined nucleotides are highly conserved. The first AG after the branchpoint/polypyrimidine tract is usually chosen as the 3' splice site (Smith et al., 1989b; Reed and Maniatis, 1985). b) The catalysis of pre-mRNA splicing is depicted. In the first step, the 5' phosphate on the +1 guanylate at the 5' splice site undergoes nucleophilic attack by the 2' hydroxyl of the invariant adenosine at the branchpoint. This reaction results in a free 5' exon and a lariat intermediate. The lariat is formed by a 2'5' phosphodiester bond between the adenosine and guanylate. Next, the free 3' hydroxyl on the guanosine of the upstream exon attacks the 5' phosphate of the guanylate in the downstream exon. A 3'5' phosphodiester bond is reformed with the intervening intron excluded. Abbreviations: R - purine, Y - pyrimidine, N - any nucleotide, n - number of repetitive pyrimidines within a tract, pY - polypyrimidine tract.
a) 5' splice site

\[
\text{AGURAGU} \quad \text{YNYURAC} \quad \text{(Y)}_n \quad \text{YAGG}
\]

Exon 1

\[ +1 \]

3' splice site

Branchpoint

pY tract

Exon 2

b) Step 1

\[
\text{AGURAGU} \quad \text{YNYURAC} \quad \text{(Y)}_n \quad \text{YAGG}
\]

\[
\text{UGARUG} \quad \text{YNYURAC} \quad \text{(Y)}_n \quad \text{YAGG}
\]

Step 2

\[
\text{mRNA} \quad + \quad \text{UGARUG} \quad \text{YNYURAC} \quad \text{(Y)}_n \quad \text{YAG}
\]

Lariat Product
exon one attacks the phosphate group of the first guanylate in exon two. This reaction results in exon ligation and removal of a lariat product.

Section II: U snRNPs and the spliceosome cycle

The spliceosome is composed of three different components: the pre-mRNA, the small nuclear ribonucleoprotein particles (snRNPs) and other trans-acting factors (Smith et al., 1989a). There are 5 so-called 'U' snRNPs (U1, U2, U4, U5, and U6) involved in pre-mRNA splicing. They are composed of U-rich RNA and proteins. U4 and U6 snRNA assemble into a single U4/6 snRNP, and join the assembling spliceosome as a tri-snRNP with the U5 snRNP (Bringmann et al., 1984; Hashimoto and Steitz, 1984; Konarska and Sharp, 1987).

Spliceosome assembly, catalysis, and recycling of splicing components are all steps of the spliceosome cycle (figure 1.5) (for review, see Moore et al., 1993). The first complex to assemble on nascent pre-mRNA is known as the H complex, and is composed of heterogeneous ribonucleoprotein (hnRNP) particles (Bennett et al., 1992b; Dreyfuss et al., 1993). H complexes do not develop into active spliceosomes (Staknis and Reed, 1994a).

The first pre-spliceosome complex to assemble is called E and requires 5' and 3' splice sites (Michaud and Reed, 1991; Staknis and Reed, 1994a). After this complex assembles on the pre-mRNA, the pre-mRNA is committed to the splicing pathway (Seraphin and Rosbash, 1989). U1 snRNP, which binds the 5' splice site, is the only snRNP to assemble at this point (Zhuang and Weiner, 1986; Michaud and Reed, 1993). In the next complex, A, U2 snRNP binds to the branchpoint site bulging out the nucleophilic adenosine (Parker et al., 1987; Liao et al., 1992; Michaud and Reed, 1993; Konarska and Sharp, 1987).

The U4/6.U5 tri-snRNP enters next to form the B1 complex (Konarska and Sharp, 1987;
Figure 1.5. The spliceosome cycle. U1 snRNP binds in E complex in a step that requires both 5' and 3' splice sites. U2 snRNP binds in A complex followed by the tri-snRNP in B1 complex. Rearrangements of the U snRNPs occur between B1 and B2 complexes. In B2 complex, U6 snRNA is associated with U2 snRNA, and the U4 snRNP is displaced. The U6 snRNP displaces U1 snRNP from the intron sequences at the 5' splice site. C1 complex is formed after the first catalytic step, and C2 complex is formed after the second catalytic step. After catalysis, the ligated exons are released, and the U snRNPs remain bound to the lariat product to form I complex. Subsequently, the lariat mRNA is degraded, and the U snRNPs are recycled for another round of splicing. This figure is a modified version of figure 4 from (Moore et al., 1993).
Abmayr et al., 1988). During B2 complex formation, RNA:RNA interactions are switched such that U6 snRNA is released from U4 snRNA and binds U2 snRNA (Blencowe et al., 1989; Madhani and Guthrie, 1992). The interaction of U6 snRNA with U2 snRNA places the U6 snRNA near the branchpoint and the 5' splice site (Madhani and Guthrie, 1992). U5 snRNP associates with the exon sequences flanking the 5' and 3' splice sites (Newman and Norman, 1992). Interactions of U6 snRNA with the 5' splice site displaces the U1 snRNP, consistent with the concentration of U1 snRNP in the spliceosome decreasing at this step (Michaud and Reed, 1993). In the C1 complex, the first step of splicing has occurred, and splicing intermediates are present (Reed et al., 1988). The C2 complex represents the spliceosome after the final step in catalysis. After catalysis is completed, the exons are released. The U snRNPs remain on the lariat product to form the I complex (Konarska and Sharp, 1987). The lariat mRNA is rapidly degraded, and the U snRNPs are recycled for the next splicing reaction.

Section III: Factors involved in splice site selection

The U snRNPs bind specifically to their respective sites with the aid of non-snRNP proteins. Prominent factors in splice site selection include U2AF, the SR family proteins, and hnRNP A1.

U2 auxiliary factor (U2AF)

U2AF was first identified as a factor required for U2 snRNP to bind the branchpoint site (Ruskin et al., 1988; Zamore and Green, 1989). U2AF binds the polypyrimidine tract in the E complex, and guides the U2 snRNP to the branchpoint in the A complex (Ruskin et al., 1988; Zamore and Green, 1989; Bennett et al., 1992a). U2AF binds with high affinity to polypyrimidines, and the strength of U2AF binding increases with the length of the
polypyrimidine tract (Zamore et al., 1992).

U2AF is composed of two polypeptides of 65 kDa (U2AF65) and 35 kDa (U2AF35). U2AF65 contains three RNA-binding domains (RRMs) and an effector domain that is rich in arginine and serine (RS domain) (see section IV) (Zamore et al., 1992). All three RRM$s are required for the sequence specific, high affinity binding of U2AF to the polypyrimidine tract (Zamore et al., 1992). U2AF35 contains an RS domain but lacks an RNA-binding domain (Zhang et al., 1992). It functions in protein-protein interactions in the assembling spliceosome (Wu and Maniatis, 1993, Zuo et al., 1996).

**SR family**

The SR (serine/arginine-rich) proteins are a family of essential splicing factors that function in splice site selection. Five criteria must be met in order for a protein to be included in the SR family (Zahler et al., 1992); i) The protein must be recognized by a specific monoclonal antibody mAb104. mAb104 recognizes SR family members from species throughout the animal kingdom (Roth et al., 1990) as well as from the plant Arabidopsis (Lazar et al., 1995). ii) the protein must be purified in a two-step procedure of 65-90% ammonium sulphate precipitation followed by precipitation in 20 mM Mg$^{2+}$. iii) the protein must be related to other SR proteins at the primary sequence level. All SR family members contain one or two RRM$s at their N-terminus and an RS domain at their C-terminus (see section IV). The SR proteins are divided into two classes depending on whether they have one or two RRM$s (Zahler et al., 1993b), iv) the size of a protein must be conserved through the animal kingdom. In the initial characterization of the SR family, SR proteins were identified and named based on their size: SRp20, SRp30, SRp40, SRp55, and SRp75, and v) the protein must complement an S100 extract
for splicing. S100 extracts contain all essential splicing components except for SR proteins. Other proteins are rich in arginine and serine, but do not meet the above criteria; they are grouped together as SR-related polypeptides (Fu, 1995).

The smallest member of the SR family, SRp20, is identical to mouse X16 (Zahler et al., 1992). X16 was isolated in a differential hybridization screen to identify genes that are transcribed at a higher level in pre-B cells than in mature B cells (Ayane et al., 1991). The *Drosophila* counterpart, RBP1 (RNA-binding protein 1), was isolated in a degenerate PCR screen to identify RRM containing proteins (Kim et al., 1992). SRp75 is the largest member of the SR family. It was isolated by the two-step purification method described above (Zahler et al., 1993b).

SRp30 is composed of several proteins. SRp30a corresponds to ASF/SF2 (Zahler et al., 1992). ASF/SF2 (alternative splicing factor/splicing factor 2) was isolated by two different methods. First, it was isolated as a constitutive splicing factor that complemented S100 extracts for splicing of β-globin pre-mRNA (Krainer and Maniatis, 1985; Krainer et al., 1990; Krainer et al., 1991). Second, ASF/SF2 was identified as an alternative splicing factor that altered 5' splice site selection of the SV40 DNA tumour early (SV40 T antigen) pre-mRNA when added to HeLa cell nuclear extracts that already contained endogenous SR proteins (Ge and Manley, 1990; Ge et al., 1991).

SRp30b is also known as SC35 (Zahler et al., 1992). SC35 (spliceosome component of 35 kDa) was identified using a monoclonal antibody raised against a partially purified spliceosome (Fu and Maniatis, 1990; Fu and Maniatis, 1992a). SC35 was also cloned as PR264, a gene on the antisense strand of a c-myb exon (Vellard et al., 1992). Extracts depleted for SC35
are not functional for splicing suggesting that SC35 is essential for spliceosome activity (Fu and Maniatis, 1990; Fu and Maniatis, 1992a).

SRp40 was isolated from conserved sequences in other SR proteins using degenerate PCR (Screaton et al., 1995). SRp40 was first cloned as rat HRS which was identified in a differential screen of insulin treated liver-like H35 cells (Diamond et al., 1993). This screen identified HRS because its mRNA levels increase following mitogen and insulin stimulation.

SRp55 was first cloned from *Drosophila* using the two-step purification method (Roth et al., 1991). It was subsequently cloned from humans in the same screen that identified SRp40 (Screaton et al., 1995). B52 is identical to *Drosophila* SRp55 except that B52 has a 21 amino acid insert near its C-terminus (Champlin et al., 1991). B52 was isolated using a monoclonal antibody that recognizes puffs on polytene chromosomes from *Drosophila* (Champlin et al., 1991). The proper expression of B52 is crucial for fly development. Both overexpression of B52 and loss-of-function of B52 are lethal (Kraus and Lis, 1994; Ring and Lis, 1994).

Three other proteins have been classified as SR proteins since the original characterization of SR proteins (Zahler et al., 1992; Fu, 1995). SRp30c was isolated in the same screen as SRp40 and SRp55 (Screaton et al., 1995). 9G8 was cloned using a monoclonal antibody that inhibits splicing *in vitro* (Cavaloc et al., 1994). Unique to this protein is the presence of a zinc knuckle motif between the RRM and the RS domains. p54 was isolated using an anti-nuclear monoclonal antibody from a patient with autoimmune disease (Chaudhary et al., 1991). This protein contains an atypical RRM (Birney et al., 1993), an RS domain, and a region rich in basic, acidic, and hydroxyamino acids. X16, SC35, 9G8, and p54 fall into the first class of SR proteins that contain only one RRM.
Differences among SR family members

At first glance, the SR family members seem indistinguishable in function from one another as any SR protein is competent to rescue splicing of an S100 extract (Ge et al., 1991; Krainer et al., 1991; Zahler et al., 1992; Fu et al., 1992; Zahler et al., 1993b). However, flies deleted for B52 are inviable implying that B52 function is not completely redundant with other SR proteins (Ring and Lis, 1994). Although B52 is involved in the splicing of w" and dsx transcripts, it is not required for the splicing of ddc. sxl. and ubx pre-mRNA (Peng and Mount, 1995; Ring and Lis, 1994). Therefore, there must be specificity among SR proteins for particular transcripts such that SR proteins function in a manner distinct from one another.

A major difference between SR proteins is their specificity in splice site selection. Like ASF/SF2, other family members are capable of influencing 5' splice site selection (Fu et al., 1992; Zahler et al., 1992; Zahler et al., 1993b). SR family members show distinct specificities, in the sense that they recognize different sites on the same pre-mRNA. Adenovirus E1A pre-mRNA transcript contains three choices for 5' splice site: a proximal (closest to the 3' splice site) 13S site, and two distal sites of 12S and 9S (Berk and Sharp, 1978; Chow et al., 1979). ASF/SF2 and SC35 increase use of the 13S site (Krainer et al., 1990; Zahler et al., 1993a). SRp40. SRp55. and SRp75 promote both 13S and 12S splicing (Zahler et al., 1993a). In contrast. SRp20 decreases use of the 13S and 9S sites while increasing use of the 12S site (Screaton et al., 1995). These differences among SR proteins are also seen in vivo (Cáceres et al., 1994; Wang and Manley, 1995). In transfection experiments, overexpression of ASF/SF2 increases proximal 13S and decreases distal 12S splicing of the E1A pre-mRNA transcript. SC35. on the other hand. decreases splicing of the two distal sites. 12S and 9S. but does not
increase proximal 13S splicing (Wang and Manley, 1995).

Like B52, other SR proteins may act on a subset of pre-mRNAs. The SR proteins are required early in spliceosome assembly (Fu and Maniatis, 1990; Krainer et al., 1990; Staknis and Reed, 1994b). In vitro, β-globin pre-mRNA is committed to the splicing pathway if it is pre-incubated with SC35 (Fu, 1993). ASF/SF2, SRp55, but not SRp20, mimic this effect to a lesser extent. In contrast, of these four SR proteins, only ASF/SF2 commits tat pre-mRNA to splicing (Fu, 1993).

The different effects on pre-mRNA splicing produced by SR family members likely reflects a difference in their RNA-binding specificity. In an in vitro selection system, SELEX, that identifies RNA bound with high affinity (Tuerk and Gold, 1990). ASF/SF2 and SC35 selected different purine-rich sequences (Tacke and Manley, 1995). ASF/SF2 selected the sequence RGAAGAAC (R is purine) whereas SC35 preferred the sequence AGSAGAGTA (S is C or G). Therefore, ASF/SF2 may function differently from SC35 because it binds different pre-mRNAs, or binds the same pre-mRNA at a different site. The RRM3s of the SR proteins are responsible for sequence specific binding (Cáceres and Krainer, 1993; Zuo and Manley, 1993; Tacke and Manley, 1995).

hnRNP A1

hnRNP A1 is an antagonist of the SR family that promotes distal 5' splice site selection (Mayeda and Krainer, 1992; Fu et al., 1992). Neither hnRNP A1 nor SR proteins are indiscriminate. The normal 5' splice site is mutated at the β-thalessemia allele of β-globin allowing use of three cryptic sites (Treisman et al., 1983; Krainer et al., 1984). In an in vitro splicing reaction, ASF/SF2 promotes use of the proximal site, whereas hnRNP A1 promotes the
distal site (Krainer et al., 1991; Mayeda and Krainer, 1992). However, in the presence of the wild-type 5' splice site, none of the cryptic sites are activated by excess SR or hnRNP A1 proteins (Mayeda and Krainer, 1992).

The levels of SR proteins compared to hnRNP A1 also influence exon inclusion (Mayeda et al., 1993). In DUP33, a pre-mRNA construct with three exons, the internal exon is normally skipped. ASF/SF2 promotes inclusion of this exon in S100 extracts. Adding increased amounts of hnRNP A1 counteracts this effect. Strengthening the polypyrimidine tract preceding the internal exon of DUP33 prevents exon skipping induced by hnRNP A1. As well, hnRNP A1 is unable to promote exon skipping of the constitutively spliced rabbit β-globin mRNA.

Therefore, it seems that constitutive exons have evolved to be independent of the relative levels of SR protein versus hnRNP A1. In contrast, alternatively spliced exons rely on the relative levels of these factors to dictate 5’ splice site selection and exon inclusion. The levels of hnRNP A1 and ASF/SF2 vary from tissue to tissue and the ratio between these two proteins is likely to be important for tissue-specific pre-mRNA splicing (Cáceres et al., 1994).

Section IV: Structural motifs of SR proteins

RNA recognition motifs (RRMs)

RRMs are the most common RNA-binding domains (Burd and Dreyfuss, 1994). There are two hallmark motifs in RRs: RNP1 and RNP2 (Adam et al., 1986; Dreyfuss et al., 1988). RNP1 is an octamer with the consensus sequence (K/R)G(F/Y)(G/A)FVX(F/Y) (Bandziulis et al., 1989). RNP2 is a hexamer with the consensus sequence (L/I)(F/Y)(V/I)(G/K)(N/G)(L/M) (Bandziulis et al., 1989). The second RRM of SR proteins is atypical because it has degenerate RNP1 and RNP2 sequences (Birney et al., 1993; Zahler et al., 1992).
The RRM structure is an anti-parallel four stranded $\beta$ sheet flanked by two $\alpha$ helices (figure 1.6) (Nagai et al., 1990; Hoffman et al., 1991). Many of the conserved residues form a hydrophobic core that is crucial for positioning the $\alpha$ helices relative to each other and to the $\beta$ sheet. RNP1 and RNP2 make up the two central $\beta$ strands. RNA binds to the RRM in an open conformation extending across the $\beta$ sheet and has three places of contact: the $\beta$ sheet, the loops, and both the N and C termini of the RRM (Görlach et al., 1992; Oubridge et al., 1994).

RNA-binding proteins often contain more than one RRM. For ASF/SF2, both RRM$s$ are required for optimal sequence specific binding of ASF/SF2 to the same pre-mRNA in vitro (Cáceres and Krainer, 1993; Zuo and Manley, 1993; Tacke and Manley, 1995). Mutating the first RRM1 of ASF/SF2 slightly reduces RNA-binding but eliminates constitutive splicing (complementing S100 extract) which can be interpreted as loss of specific RNA-binding (Cáceres and Krainer, 1993).

**RS domains**

RS domains are arginine/serine-rich sequences that contain repeating dipeptides of SR or RS. Members of the SR family have statistically significant RS or SR repeats (Birney et al., 1993). RS domains were first identified in the *Drosophila* splicing regulators su(w$d$), Tra, and Tra-2 (Chou et al., 1987; Amrein et al., 1988; Goralski et al., 1989; Boggs et al., 1987). These proteins had statistically significant Rx or Sx repeats (x is any amino acid) but no significant RS repeats (Birney et al., 1993). Therefore, it is unclear what makes up a minimal RS domain. U1 70K is a component of the U1 snRNP (Wooley et al., 1983). The RS domain of U1 70K contains RD and RE repeats as well as RS repeats (Spritz et al., 1987; Theissen et al., 1986). RD/RE repeats are common to splicing factors, as approximately half of the proteins found in
Figure 1.6. Structure of the RRM1 of U1A. The structure of the RRM1 of U1A is a four stranded anti-parallel β sheet with flanking α helices. RNP1 forms β3, and RNP2 forms β1. Conserved aromatic residues within RNP1 and RNP2 are shown. Loops 1 and 3, the β sheets, and the N and C termini (not shown here) contribute to RNA-binding. This figure is reproduced from (Kenan et al., 1991).
E or B complexes contain these repeats to some extent (Neugebauer et al., 1995).

RS domains are phosphorylated in vivo (see section IX). Phosphorylated RS domains mimic the + - + - alternating charge of RD/RE repeats. It is postulated that RD repeats form a polar zipper of β strands in which oppositely charged residues interact (Perutz, 1994). Alternately, RD repeats may form an α helix that is stabilized by charge compensation between ridges of arginines and grooves of aspartates. Two such helices could interact if the positive ridges of one helix fills the negative grooves of the other.

Many proteins that contain RS domains are found in discrete subnuclear locations known as speckles (see section VIII). The RS domain of su(w²) is sufficient to target a β-galactosidase fusion protein to speckles implying that RS domains are important in this localization (Li and Bingham, 1991). One speckle localization signal is linked to the bipartite nuclear localization signal (NLS) and is characterized by RS repeats surrounding the NLS (Hedley et al., 1995).

The RS domains of SR proteins are involved in protein-protein interactions (Wu and Maniatis, 1993; Kohtz et al., 1994; Amrein et al., 1994). SC35 lacking an RS domain is unable to bind Tra or Tra-2 (Wu and Maniatis, 1993). As well, the RS domains of U1 70K and ASF/SF2 are both essential for U1 70K and ASF/SF2 interaction (Kohtz et al., 1994). It is unclear whether RS domains are sufficient for this interaction, or if additional sequences outside of the RS domains are required.

Not all RS domain-containing proteins interact with each other (Wu and Maniatis, 1993). U1 70K does not interact with U2AF35 even though both proteins have RS domains. Similarly, U2AF65 does not bind ASF/SF2 or SC35. Whether the specificity of these interactions resides in the RS domains or in external sequences has not been determined.
Both the arginine and serine residues are important for RS domain function. ASF/SF2 is unable to complement an S100 extract in a constitutive splicing assay if all the RS dipeptide repeats in the RS domain are deleted (Cáceres and Krainer, 1993). A similar result is observed if just the arginines or just the serines are mutated.

For ASF/SF2, both RRM{s} are required for constitutive and alternative splicing (5' splice site switch in nuclear extracts) (Cáceres and Krainer, 1993; Zuo and Manley, 1993). These results indicate that, for both assays, specific binding of ASF/SF2 to pre-mRNA is required. In contrast, the RS domain is required for constitutive splicing but not for alternative splicing. In this case, endogenous SR proteins may provide the protein-protein interactions in the assembling spliceosome.

Section V: Spliceosome assembly

In vitro, assays using a substrate containing only one intron do not approach the complexity of pre-mRNA in a mammalian cell. The diversity of the 5' splice site consensi in mammalian cells suggests that candidates for splice sites are found at approximately 0.5 to 1% of positions in random RNA (Eperon et al., 1993). Yet the correct sites are chosen with high fidelity. Therefore, there must be a way to distinguish authentic sites.

Splice sites are likely to be defined across the exon as opposed to the intron (Robberson et al., 1990). The average length of a vertebrate internal exon is 137 nucleotides whereas the average length of a vertebrate internal intron is 1127 nucleotides (Hawkins, 1989). By defining sites across the shorter exon, the chances of picking cryptic splice sites are reduced. In cases where the intron is smaller than the exon, then definition may occur across the intron (Talerico and Berget, 1994).
Assembly at the 5' splice site

The 5' splice site is the first to be bound by a U1 snRNP in E complex (Michaud and Reed, 1991). SR proteins are believed to bind first for the following reasons: i) SR proteins are present in the earliest spliceosome complex and are required for E complex formation (Fu and Maniatis, 1990; Krainer et al., 1990; Staknis and Reed, 1994b). ii) SR proteins are sufficient to commit a pre-mRNA to the splicing pathway (Fu, 1993). iii) Addition of ASF/SF2 to nuclear extracts increases the affinity of U1 snRNP for 5' splice sites (Eperon et al., 1993), and iv) stable binding of U1 snRNP to Adenovirus major late (AdML) pre-mRNA 5' splice site requires ASF/SF2 to bind first (Kohtz et al., 1994). The interaction between ASF/SF2 and U1 snRNP is likely to be mediated by the interaction of ASF/SF2 with U1 70K.

Extracts depleted for U1 snRNP can be rescued by ten-fold overexpression of SR proteins (Crispino et al., 1994). This rescue is only partial as SR proteins do not compensate for 5' splice site fidelity (Tarn and Steitz, 1994). It is believed that SR proteins rescue U1 snRNP by promoting binding of U6 snRNA to the 5' splice site. U6 snRNA can rescue the splicing reaction, but it can not compensate for U1 snRNP in accurate splice site selection.

For alternative 5' splicing, there must be at least two sites that can be employed during the splicing reaction. In the case of two sites with strong consensus sequences, U1 snRNP binds both sites (Eperon et al., 1993). When both sites are bound, the proximal is chosen (Eperon et al., 1993; Reed and Maniatis, 1986). For weaker sites, U1 snRNP may be bound to only one site at a time due to decreased affinity. In this case, the site bound at the time of spliceosome assembly is used.

SR proteins usually increase the affinity of U1 snRNP for both sites, and as a
consequence the proximal site is chosen (Eperon et al., 1993). For SV40 T antigen splicing, SC35 promotes small t splicing, and SRp40 or SRp55 promote large T antigen splicing (Zahler et al., 1993a). In this case, SC35 promotes binding of U1 snRNP to both sites, and the proximal site is chosen (Zahler and Roth, 1995). SRp40 or SRp55 enhance U1 snRNP binding only to the distal 5' splice site which is chosen as a consequence.

Selection of the 3' splice site

During E complex formation, U1 snRNP and SR proteins bind to the 5' splice site, and U2AF binds to the polypyrimidine tract (Staknis and Reed, 1994b). If the polypyrimidine tract is strong, U2AF binding is independent of a downstream 5' splice site (Hoffman and Grabowski, 1992). Binding of U2AF to a weak polypyrimidine tract is dependent on a 5' splice site. In the exon definition model, the important 5' splice site is downstream of U2AF. Binding of U2AF to the polypyrimidine tract of exon 4 of preprotachykinin is dependent on U1 snRNP binding the downstream 5' splice site (Hoffman and Grabowski, 1992). For inclusion of exon 18 of neural cell adhesion molecule, the strength of the upstream 5' splice site is important (Côté et al., 1995). Therefore, both upstream and downstream 5' splice sites can promote binding of U2AF to the 3' splice site. Enhancers within exons are also important for this binding (see sections VI and VII).

An interaction between the 5' and 3' splice sites is established in E complex, and retained throughout the splicing reaction (Michaud and Reed, 1993). SR proteins are implicated in protein-protein interactions that bring together the splice sites as ASF/SF2 and SC35 bind to both U2AF35 and U1 70K (figure 1.7) (Kohtz et al., 1994; Wu and Maniatis, 1993). Depletion of SC35 and other SR proteins from nuclear extracts results in reduced interaction between the U1 and U2 snRNPs consistent with SR proteins mediating 5'-3' splice site interactions (Fu and
Figure 1.7. SR proteins mediate a network of interactions in the assembling spliceosome.

First, SR proteins promote U1 snRNP binding to the 5' splice site. If the polypyrimidine tract is weak, SR proteins recruit U2AF to this site. Next, SR proteins mediate interactions between U2AF35 at the 5' splice site and U1 70K at the 5' splice site. These interactions may take place across the exon or the intron. Finally, SR proteins are required to escort the tri-snRNP into the spliceosome complex.
SR proteins are also involved in escorting the U4/6.U5 tri-snRNP into the splicing complex (Roscigno and Garcia-Blanco, 1995). To chase pre-spliceosomes (E and A complexes) into spliceosomes (B and C complexes) in vitro, the tri-snRNP, SR proteins, and additional factors are required. ASF/SF2, SC35, SRp40, SRp55, and SRp75, but not SRp20, are all capable of promoting B complex assembly. Therefore, SR proteins appear to have roles in all three U snRNP assembly steps (figure 1.7).

**Section VI: Somatic sex determination in *Drosophila***

The best example of alternative splicing regulating a developmental process is the sex determination pathway in somatic tissues of *Drosophila* (figure 1.8) (for review, see MacDougall *et al.*, 1995). The primary sex determination signal in *Drosophila* is the ratio of X chromosomes to autosomal chromosomes (Bridges, 1921). If the ratio is 1, the fly follows the female differentiation pathway. If the ratio is 0.5, the fly develops into a male. The default pathway is male, and the switch to female differentiation requires a cascade of alternative splicing beginning with the RNA-binding protein Sex-lethal (Sxl) and ending with the Doublesex transcription factor (Dsx) (Baker and Belote, 1983).

**Initiation of the sex determination signal**

The ratio of X chromosomes to autosomes is determined by counting elements (Cline, 1993). Transcription factors on X chromosomes make up the numerator whereas transcription factors on autosomes make up the denominator. Numerator elements activate female-specific transcription from the early promoter of the sxl locus whereas denominator elements repress this
Figure 1.8. Somatic sex determination pathway of *Drosophila*. In females, the X:A ratio is 1, and female specific transcription from the *sxl* early promoter is initiated. The transcript is processed to exclude exon 3 which contains a stop codon. The Sxl protein autoregulates its own activity by blocking exon 3 inclusion in subsequent *sxl* mRNAs. Sxl also binds to the non-sex-specific 3' splice site of exon 2 of *tra*, thereby preventing the incorporation of a stop codon in the final transcript. Tra and Tra-2, along with SR proteins, bind the *dsxRE* (striped region) of exon 4 of *dsx*. Here, they recruit U2AF to the weak polypyrimidine tract of exon 4. As a result, exon 4 is included in the mRNA and DsxF is made. In males, transcription from *sxl* early promoter is inhibited. Therefore, no Sxl protein is produced from late transcription. As a result, no Tra protein is produced, and exon 4 of *dsx* is skipped. DsxM is produced by default. Proteins that are specific to the female splicing pathway are coloured in grey.
X:A Signal

Ratio = 1

Ratio = 0.5

Late

Early

sxl DNA

Late

Early

Early sx1 pre-mRNA

UGA

Early Sxl Protein

late sx1 pre-mRNA

UGA

tra pre-mRNA

UAG

Tra protein

SR

Tra-2

UGA

Tra protein

UGA

DsxF

DsM

Ds protein

DsxA

DsxB
transcription. In females, there are twice as many numerator elements than in males, therefore, transcription of sxl from the early promoter is restricted to females.

**Sex-lethal (Sxl)**

The RNA-binding protein Sxl is the ON/OFF switch that determines sex differentiation in *Drosophila* somatic tissues (Bell *et al.*, 1988). Loss of sxl function in XX somatic tissue leads to male differentiation and gain of sxl function in XY somatic tissue leads to female differentiation (Cline, 1979; Maine *et al.*, 1985).

Both males and females undergo transcription from the late promoter of sxl (Maine *et al.*, 1985). The male transcript is 190 base pairs longer than the female transcript as it retains exon 3 (Bell *et al.*, 1988). Due to stop codons in exon 3, no functional protein is produced in males. The early female-specific transcript is spliced so that exon E1 is joined to exon 4, bypassing the stop codons present in exon 3 (Keyes *et al.*, 1992). The Sxl protein then prevents exon 3 inclusion in later transcripts (Bell *et al.*, 1988; Bell *et al.*, 1991).

Once the signal has been initiated, Sxl autoregulates its own activity and becomes independent of counting elements (Cline, 1984; Bell *et al.*, 1991). The maintenance of this autoregulatory loop is required throughout female development and for dosage compensation (Bell *et al.*, 1991).

*In vitro*. Sxl binds to the GU₈C sequence in the polypyrimidine tract of exon 3 (Kanaar *et al.*, 1995). In addition to the polypyrimidine tract, runs of poly(U)s are also present in both the upstream and downstream introns of exon 3 (Bell *et al.*, 1988). These poly(U) motifs are important for proper female splicing (Horabin and Schedl, 1993), and Sxl binds these motifs in a co-operative manner *in vitro* (Wang and Bell, 1994). Binding of Sxl to multiple sites may
block access of splicing factors to both the 5' and 3' splice sites of exon 3, thereby preventing its inclusion in the final product. The co-operative nature of Sxl binding to RNA may serve to set a threshold that prevents accidental activation of Sxl in males, and may render female splicing less sensitive to Sxl protein fluctuations.

**Transformer (Tra)**

*Tra* functions downstream of *sxl* (Steinmann-Zwicky and Nöthiger, 1985; Nagoshi *et al.*, 1988). Loss-of-function *tra* mutants display no phenotype in XY flies, but transform XX flies into pseudomales that are morphologically, physiologically, and behaviourally male, but are sterile (Sturtevant, 1945). Splicing of the *tra* transcript is regulated by Sxl such that a 1.2 kb transcript is present in both sexes, and a 1 kb transcript is female-specific (McKeown *et al.*, 1987; Nagoshi *et al.*, 1988). The 175 additional basepairs in the larger transcript contain stop codons in all three frames so that no functional protein is produced (Boggs *et al.*, 1987). In females, the 1 kb transcripts encodes a 22 kDa protein which lacks an RRM but contains an RS domain.

The difference in transcript size is due to alternative splicing at the 3' splice site of the first intron (Boggs *et al.*, 1987). The non sex-specific transcript is spliced at the proximal site. In females, Sxl binds to the polypyrrimidine tract of the non sex-specific splice site to prevent subsequent binding of U2AF (Valcárcel *et al.*, 1993). Addition of the RS domain of U2AF to Sxl promotes usage of the non sex-specific site, indicating that Sxl blocks this site because it lacks an effector domain (Sosnowski *et al.*, 1989; Valcárcel *et al.*, 1993).

**Transformer 2 (Tra-2)**

*Tra-2* functions downstream of *sxl*, and in parallel to *tra* (Watanabe, 1975; Cline, 1979;
Nagoshi et al., 1988). Loss of tra-2 function is indistinguishable from loss of tra function (Watanabe, 1975). Although, functional Tra-2 is found in both sexes (Amrein et al., 1988; Goralski et al., 1989), Tra-2 does not function in males because it requires Tra for its activity (McKeown et al., 1988).

**Doublesex (Dsx)**

Dsx lies at the bottom of the hierarchy of alternatively spliced mRNAs, downstream of tra and tra-2 (Baker and Ridge, 1980; Steinmann-Zwicky and Nöthiger, 1985). Flies homozygous for dsx display an intersexual phenotype regardless of whether they are XX or XY (Hildreth, 1965). This result implies that the action of Dsx is required in both sexes, and varies depending on the chromosome balance. Male and female dsx transcripts are different sizes (Baker and Wolfner, 1988). In females, exon 3 is spliced to exon 4 (Burtis and Baker, 1989). In males, exon 3 is spliced to exons 5 and 6. The female Dsx (DsxF) is 427 amino acids, and the male Dsx (DsxFM) is 549 amino acids. The two proteins have in common the first 397 amino acids in which a zinc finger domain is present (Burtis and Baker, 1989; Erdman and Burtis, 1993). DsxF activates female-specific genes and represses male-specific genes, whereas DsxM functions in the opposite manner (Coschigano and Wensink, 1993; MacDougall et al., 1995).

The female-specific exon 4 of dsx contains a weaker polypyrimidine tract than that of the male-specific exon 5 (Burtis and Baker, 1989). Therefore, the stronger male 3' splice site is chosen as a default. Tra and Tra-2 are required in females to switch splicing to exon 4. If the strength of the polypyrimidine tract of exon 4 is increased, female splicing occurs independently of Tra and Tra-2 (Hoshijima et al., 1991). Exon 4 contains the dsx repeat element (dsxRE) that is composed of six conserved copies of a 13 nucleotide repeat (Burtis and Baker, 1989).
Deletion of these repeats renders dsx splicing insensitive to the presence of Tra and Tra-2 (Hoshijima et al., 1991; Hedley et al., 1991; Ryner and Baker, 1991). Tra-2 binds specifically to these repeats (Inoue et al., 1992; Tian and Maniatis, 1992) whereas Tra binds these repeats indiscriminately (Tian and Maniatis, 1992).

SR proteins also bind the dsxRE (Tian and Maniatis, 1993; Tian and Maniatis, 1994). SR proteins, along with Tra and Tra-2, recruit U2AF to the weak polypyrmidine tract of exon 4 (Zuo et al., 1996). Tra-2 contains two RS domains separated by an RRM (Amrein et al., 1988; Goralski et al., 1989). The RRM is required for site-specific binding to the dsxRE, and the second RS domain is required for interactions with Tra, ASF/SF2, SC35, and U2AF35 (Amrein et al., 1994; Wu and Maniatis, 1993). The first RS domain likely stabilizes interaction of Tra-2 with RNA and/or proteins.

Section VII: Mammalian purine-rich enhancers

Sequences within exons that enhance splicing are also found in mammalian pre-mRNAs. A subset of these enhancers is referred to as exon-recognition sequences (ERSs) (Watakabe et al., 1993). ERSs are composed of GAR repeats where R represents A or G (Watakabe et al., 1993; Tanaka et al., 1994; Dirksen et al., 1994; Dominski and Kole, 1994). ERSs are sufficient for stimulation of splicing, as ERSs placed into foreign exons can stimulate splicing of the upstream introns (Watakabe et al., 1993; Lavigneur et al., 1993; Xu et al., 1993; Tanaka et al., 1994). ERSs are position-dependent to a certain extent. The ERS from exon ED1 of fibronectin does not function if it is more than 293 nucleotides away from the 3' splice site (Lavigneur et al., 1993). This result is in striking parallel with the dsxRE that acts as a constitutive enhancer when it is placed closer than 300 nucleotides away from the 3' splice site (Tian and Maniatis,
Splicing of sites that conform well to the consensus is independent of exonic enhancers (Lavigueur et al., 1993; Dirksen et al., 1994). Thus, an ERS is required only if the upstream splice site is weak. In this case, the 3' splice site, the ERS, and the downstream 5' splice site may operate together to increase use of the 3' splice site (Watakabe et al., 1993).

Both U1 snRNP and SR proteins bind ERSs (Watakabe et al., 1993; Sun et al., 1993b; Lavigueur et al., 1993; Ramchatesingh et al., 1995). The sequence of the enhancer dictates the type of SR protein that binds. For instance, the avian sarcoma-leukosis virus (ASLV) ERS binds SRp40 strongly, and SRp30 and SRp20 weakly (Staknis and Reed, 1994b). If six uracils are added to the ASLV ERS, SRp30 still binds but the amount of crosslinked SRp40 is greatly reduced.

The ERSs are required early on in spliceosome assembly (Sun et al., 1993a; Lavigueur et al., 1993; Staknis and Reed, 1994b). U1 snRNP and SR proteins are present in E complex, in enhancer complexes, and in complexes containing only the 5' or 3' splice sites (Staknis and Reed, 1994b). The following model helps explain the multiple roles of SR proteins in spliceosome assembly (Staknis and Reed, 1994b). When splice sites are strong, SR proteins promote U1 snRNP and U2AF binding to their respective sites independently of external sequences. When the 3' splice site is weak, the U1 snRNP/SR protein complex bound at the downstream 5' splice site and/or the ERS recruit U2AF to the 3' splice site (Hoffman and Grabowski, 1992; Staknis and Reed, 1994b). Normally, the upstream 5' splice site is also involved in recruiting U2AF to the downstream 3' splice site (Michaud and Reed, 1993; Côté et al., 1995). In cases where this 5' splice site is weak, U1 snRNP/SR proteins bound at the ERS
can compensate by promoting U2AF binding. Once U2AF is bound, it may help stabilize U1 snRNP/SR protein complex formation at the weak upstream 5' splice site (Eperon et al., 1993). Therefore, a variety of protein-protein interactions mediated by SR proteins can be employed to compensate for weak splice sites.

Section VIII: Location of splicing in the nucleus

The framework of an intact nucleus provides a level of complexity that is lacking in \textit{in vitro} splicing assays. Here, the spatial arrangement of splicing components both to each other and to pre-mRNA factor into the splicing reaction.

Co-transcriptional splicing

Pre-mRNA splicing takes place co-transcriptionally (Beyer et al., 1980; Osheim et al., 1985; Beyer and Osheim, 1988). Consistent with the co-transcriptional occurrence of splicing, snRNPs and SR proteins are present on active lampbrush chromosome loops, and along the RNA fibres extending from these chromosomes (Gall and Callan, 1989; Fakan et al., 1986; Roth et al., 1990). \textit{Drosophila} SR proteins, B52 and RBP1, co-localize with RNA polymerase II at heat shock induced puffs on \textit{Drosophila} polytene chromosomes (Champlin et al., 1991; Kim et al., 1992). In this case, B52 is localized to the border between decondensed regions of transcribed chromatin and flanking inactive condensed chromatin (Champlin and Lis, 1994). This location may allow B52 to ‘track’ the splice sites as they appear in the nascent transcript.

Nuclear speckles

Although snRNPs and SR proteins are localized to active chromatin, they are concentrated in speckles that are discrete subnuclear structures in the nucleus (Fu and Maniatis, 1990; Spector, 1993). Speckles overlap two nuclear substructures: interchromatin granule clusters (IGCs) and
perichromatin fibrils (PFs) (Fakan et al., 1984; Puvion et al., 1984; Spector et al., 1991). PFs form a reticular network between speckles, and are thought to be composed of pre-mRNA (Spector, 1983; Fakan and Puvion, 1980; Fakan et al., 1976; Spector, 1993).

One model is that speckles are storage sites for splicing factors and splicing factors are recruited from speckles to sites of splicing on PFs (Jiménez-Garcia and Spector, 1993). Support for this model is based on the fact that speckle structure is altered when either transcription or splicing is inhibited. When either process is blocked, PFs disappear, and speckles display an enlarged rounder appearance (Spector, 1993; O'Keefe et al., 1994). The enlarged speckles are likely due to splicing factors (i.e., SR proteins, and some snRNPs) returning to their storage site when RNA processing is inhibited and nascent pre-mRNAs (PFs) are degraded.

**Link between transcription and splicing**

Co-regulation of transcription and RNA processing makes sense since it might ensure that a cell does not waste energy making transcripts that cannot be spliced or priming splicing factors for a reaction that can not take place due to lack of substrate. The redistribution of splicing factors during inhibition of transcription or splicing confirms this link. As well, when antisense U snRNA oligonucleotides block splicing, transcription is also blocked as measured by lack of \[^{1}H\] uracil incorporation (O'Keefe et al., 1994).

Certain transcription factors are also localized to speckles providing a further link between transcription and splicing. The phosphorylated form of the largest subunit of RNA polymerase II is found in speckles (Bregman et al., 1995). This localization is most pronounced when transcription is inhibited, suggesting that transcription factors, like splicing factors, may use speckles as storage sites. Transcription factors Rb and WT1 also localize to speckles (Durfee
Section IX: Regulation of splicing by phosphorylation

In order for transcripts to be spliced in a tissue-specific manner, the spliceosome must be programmed in a cell-specific manner. One method to transduce signals to the spliceosome is by phosphorylation of splicing components. As well, phosphorylation of components seems to play a role in regulating the correct assembly of the spliceosome.

Phosphorylation-dephosphorylation cycle of splicing

Phosphorylation of a subset of spliceosome components appears to be required for spliceosome assembly. When protein phosphatase 1γ (PP1γ) is added to nuclear extracts, spliceosome assembly is blocked before E complex formation (Mermoud et al., 1994). When there is a choice of two 5' splice sites, addition of PP1γ inhibits use of the distal site and increases use of the proximal site (Cardinali et al., 1994).

If PP1γ is added after spliceosome assembly but before catalysis, splicing products are formed (Mermoud et al., 1994). This result suggests that dephosphorylation of components does not block catalysis. In fact, dephosphorylation is required for catalysis to proceed. Inhibitors of PP1 (tautomycin, microcystin LR) block the first step of splicing, and inhibitors of PP2A (okadaic acid) block the second step of splicing (Mermoud et al., 1992; Tazi et al., 1992). Substrates thiophosphorylated with ATPγS are resistant to dephosphorylation by phosphatases (Li et al., 1988). When ATPγS is added to splicing reactions, the second step of splicing is blocked (Tazi et al., 1992). Neither ATPγS nor phosphatase inhibitors interfere with splicing assembly which is consistent with a requirement for phosphorylation for these steps (Mermoud et al., 1992; Tazi et al., 1992).
Targets of phosphorylation-dephosphorylation

One target for phosphorylation is U1 70K (Wooley et al., 1983). U1 70K is phosphorylated on serine in its RD/RE/RS domain (Wooley et al., 1983; Woppmann et al., 1993). U1 snRNPs containing thiophosphorylated U1 70K are unable to rescue splicing of U1 snRNP-depleted nuclear extracts suggesting that U1 70K is a key target for protein phosphatases during catalysis (Tazi et al., 1993).

SR proteins are also targets for kinases as they are phosphorylated in vivo (Roth et al., 1990; Gui et al., 1994a). mAb104, which binds SR proteins in vivo, recognizes a phosphoepitope (Roth et al., 1990). SR proteins are likely phosphorylated on multiple sites as they migrate slower than expected on SDS polyacrylamide gels (Zahler et al., 1993b; Roth et al., 1991). This slow migration does not occur when SR proteins are dephosphorylated by alkaline phosphatase.

SR proteins purified from HeLa cells rescue extracts blocked in spliceosome assembly by PP1γ (Mermoud et al., 1994), suggesting that SR proteins are the targets of PP1γ in these reactions. Consistent with this idea, SR proteins are substrates for PP1γ in vitro (Mermoud et al., 1994). Phosphorylated SR proteins are also required to chase pre-spliceosomes into spliceosomes, as SR proteins treated with PP1 are unable to promote B complex formation in vitro (Roscigno and Garcia-Blanco, 1995). These results suggest that phosphorylated SR proteins are required at several steps in spliceosome assembly.

hnRNP A1, the antagonist of SR proteins, is also a phosphoprotein in vivo (Cobianchi et al., 1993). In contrast to SR proteins, phosphorylation of hnRNP A1 inhibits its function (Municio et al., 1995; Cobianchi et al., 1993).
SR protein kinases

Several kinases have been identified that phosphorylate RS domain-containing proteins. A U1 snRNP-associated kinase phosphorylates the RD/RE/RS region of U1 70K as well as the RS domain of ASF/SF2 (Woppmann et al., 1993). The cDNA for this kinase has not been cloned.

SRPK1 (SR protein kinase 1) was cloned from HeLa cells on the basis of its ability to phosphorylate SC35 (Gui et al., 1994a). SRPK1 phosphorylates all RS domain-containing proteins that have been tested. SRPK1 does not phosphorylate ASF/SF2 lacking an RS domain. As well, if the arginines and serines in RS repeats are deleted, phosphorylation of ASF/SF2 by SRPK1 is inhibited (Gui et al., 1994b). These results suggest that RS repeats are the main targets of SRPK1.

SRPK1 is believed to phosphorylate SR proteins in vivo based on the following observations. First, SRPK1 is most active in mitosis when SR proteins are highly phosphorylated (Gui et al., 1994a; Roth et al., 1990). Second, ASF/SF2 is recognized by mAb104 after phosphorylation by SRPK1 implying SRPK1 phosphorylates ASF/SF2 on sites phosphorylated in vivo (Gui et al., 1994b).

SRPK1 has at least two effects on splicing. First, high concentrations of SRPK1 inhibit splicing in vitro (Gui et al., 1994a). This inhibition likely occurs in catalysis where protein phosphatases are overwhelmed with phosphorylated targets. Second, addition of SRPK1 to permeabilized cells results in disassembly of nuclear speckles. Interestingly, overexpression of either U1 70K or Tra causes speckle aggregation in vivo (Romac and Keene, 1995; Hedley et al., 1995). In these aggregates, U1 70K is unphosphorylated (Romac and Keene, 1995). Together,
these results suggest that speckles contain unphosphorylated splicing factors that are released from speckles upon phosphorylation. This release of factors may allow them to move to sites of splicing.

A third SR protein kinase, LBR kinase, is part of the lamin B receptor (LBR) complex (Simos and Georgatos, 1992; Nikolakaki et al., 1996). LBR kinase phosphorylates the RS domain of LBR (Nikolakaki et al., 1996). The cDNA for LBR kinase has not been identified.

A fourth SR protein kinase is DNA topoisomerase I (Rossi et al., 1996). This protein was isolated from the same HeLa cell extracts that contain the U1 70K kinase, and may be responsible for that activity. When DNA topoisomerase I is inhibited by camptothecin in vivo, the phosphorylation levels of SR proteins are reduced. The fact that phosphorylation is not abolished when DNA topoisomerase I is inhibited is testament to other SR protein kinases operating in vivo.
Part III: Thesis Overview

In my thesis work, I sought to discover the function of Clk/Sty in cellular regulation. In chapter two, I describe a two-hybrid screen used to isolate Clk/Sty binding proteins. In this screen, I identified 26 proteins that bound specifically to Clk/Sty, three of which are SR proteins X16, ASF/SF2, and SRp55. I went on to show that Clk/Sty phosphorylates ASF/SF2 on sites that are also phosphorylated in vivo. Furthermore, overexpression of Clk/Sty disassembles nuclear speckles in vivo. From these results, I suggest that Clk/Sty regulates SR proteins in vivo.

In chapter three, I compared the activities of SRPK1 and Clk/Sty towards SR proteins in vitro. I propose that SRPK1 is specific for RS domain-containing proteins whereas Clk/Sty is a more general kinase that may regulate other proteins in addition to RS domain-containing proteins.

In chapter four, I describe my work with the remaining two-hybrid clones that I identified in my screen. In the first part, I describe the initial characterization of a mammalian counterpart of the Drosophila tra-2 gene. In the second part, I describe the cloning of CARS-Cyp. Finally, I include a table of all the putative Clk/Sty interacting two-hybrid clones, and discuss the function of the two-hybrid clones whose identities are known.

In chapter five, I summarize what my studies suggest about the function of Clk/Sty. I suggest future directions for the study of Clk/Sty emphasizing the need to identify proteins that regulate Clk/Sty, and by extension, pre-mRNA splicing.
CHAPTER 2: The Clk/Sty Protein Kinase Phosphorylates SR Splicing Factors and

Regulates their Intranuclear Distribution

I did the work reported in this chapter with the following exceptions:

Sara Jacob assisted me in sequencing of the two-hybrid clones.

Peter Duncan in John Bell's lab made the GST-Clk/Sty and M-Clk/Sty constructs and performed the immunofluorescence work.

Chapter 2 is a modified version of the following publication:


© 1996 by EMBO/Oxford University Press
Introduction

As described in chapter one, mammalian Clk/Sty is the prototype for a family of dual specificity kinases (termed LAMMER kinases) that have been conserved in evolution, but whose physiological substrates are unknown. In this chapter, I describe a yeast two-hybrid screen (Fields and Song, 1989) I performed to identify proteins that interact with the murine Clk/Sty kinase. In this screen, the Clk/Sty kinase specifically interacted with RNA-binding proteins, particularly members of the serine/arginine-rich (SR) family of splicing factors. Clk/Sty itself has an SR-rich non-catalytic N-terminal region which is important for its association with SR splicing factors. In vitro. I showed that Clk/Sty efficiently phosphorylated the SR family member ASF/SF2 on serine residues located within its SR-rich region (the RS domain). Using tryptic phosphopeptide mapping, I demonstrated that the sites on ASF/SF2 phosphorylated in vitro overlap with those phosphorylated in vivo. Immunofluorescence studies showed that a catalytically inactive form of Clk/Sty co-localized with SR proteins in nuclear speckles. Overexpression of the active Clk/Sty kinase caused a redistribution of SR proteins within the nucleus. I propose that SR splicing factors are physiologically relevant substrates of Clk/Sty and that the Clk/Sty kinase directly regulates the activity and compartmentalization of SR splicing factors.

Material and Methods

DNA subcloning - The following Clk/Sty constructs were amplified by PCR from plasmid pE3.1 (Ben-David et al., 1991) using BamHI-tagged primers and inserted into vector pAS1 (gift of S. Elledge) (Durfee et al., 1993) in frame with the DNA-binding domain: pAS-Clk/Sty (corresponding to amino acids (aa) 1 to 483), pAS-Clk/Sty^{RS} (aa 1 to 156), and pAS-Clk/Sty^{KIN}.
To generate pAS-Clk/Sty<sup>K190R</sup>, the plasmid pE3-2KR-6 (Ben-David <i>et al.</i>, 1991), with lys 190 mutated to arg, was cut with HindIII and EcoRI to release fragment 1 (nucleotides (nt) 265 to 1701). Fragment 2 (nt 70 to 504) was generated by PCR from pE3.1 with EcoRI-tagged primers and cut with HindIII to generate a smaller fragment (nt 70 to 264). The two fragments were ligated into the EcoR1 site of the yeast plasmid pRD54 (gift of R. Deschenes). A SalI fragment (nt 277 to 1521) was released from pAS-Clk/Sty and a similar SalI fragment from pRD54-Clk/Sty<sup>K190R</sup> was inserted in its place. To construct pACT-X16 RS, X16 (aa 86 to 164) was amplified from pX16.12S5 (Ayane <i>et al.</i>, 1991) by PCR with NcoI/EcoRI-tagged primers and cloned into NcoI/EcoRI-digested pACTII vector (gift of S. Elledge). pACT-X16 RNA-binding domain (aa 1 to 85) was constructed similarly. To construct pACT-ASF RS, ASF/SF2 (aa 198 to 248) was amplified by PCR from pDS-H6F1 (Ge <i>et al.</i>, 1991) with EcoRI/SalI-tagged primers and cloned into the EcoRI/XhoI sites of pACTII. To generate GST-ASF RS, the RS domain of ASF/SF2 was amplified by PCR using BamHI/SalI-tagged primers and cloned into the BamHI/SalI-digested pGEX-4T-2 vector (from Pharmacia). To construct RC-ASF, pDS-H6F1 was digested with BamHI releasing full-length ASF/SF2 (aa 1 to 248) that was then ligated into BamHI-digested pET-28a (Novagen) to create plasmid pET-ASF that coded for a T7 Tag epitope at the N-terminus of ASF/SF2. The ASF/SF2 insert was then digested with EcoRI, blunt ended with klenow, digested with XbaI, and inserted into XbaI/Blunt End sites of pRC/CMV (Invitrogen). All PCR products were sequenced to confirm their sequence integrity.

**Two-hybrid screen** - Yeast strain Y153 (Durfee <i>et al.</i>, 1993) was transformed with pAS-Clk/Sty. Expression of the construct was confirmed by Western analysis using the monoclonal antibody 12CA5F (BABCO) which recognizes an HA-epitope tag preceding the N-terminus of Clk/Sty.
The resulting strain was co-transformed with an unstimulated murine T cell cDNA library fused to the GAL4 activation domain in the vector pACT (library was gift from S. Elledge) (Durfee et al., 1993). Putative interacting clones were identified on selective plates containing 50 mM 3-amino triazole and screened for β-galactosidase activity (Durfee et al., 1993). Positive clones were purified in bacteria selecting for their LEU marker and were re-introduced into yeast strain Y153. Any clone that activated the reporter on its own was eliminated. To eliminate false positives, Y153 transformants containing the cDNA clones were mated with yeast strain Y187 that had been transformed with one of the following plasmids: pAS1, pAS-Clk/Sty, pAS-CDK2, pAS-SNF1 (Harper et al., 1993). To define the domains of interaction, Y153 was transformed with pACT-ASF RS and pACT-X16 RS. These transformants along with the library clone transformants were mated to yeast strain Y187 containing one of the following: pAS-Clk/Sty, pAS-Clk/StyK190R, pAS-Clk/StyKI, and pAS-Clk/StyRS and assayed for activation of reporters.

**Sequencing of clones** - Clones that showed a positive interaction with Clk/Sty were partially sequenced with [³⁵S] ATP or fluorescent primers using the Automated Laser Fluorescence DNA sequencer.

**Protein purification** - GST-Clk/Sty (Duncan et al., 1995) was induced in bacteria using 1 mM IPTG, harvested after 3 hours of induction in buffer A (50 mM NaPi pH 7.5, 150 mM NaCl, 0.1 mM EDTA, 1% Triton X-100, 2 mM dithiothreitol (DTT), 2 mM benzamidine), and bound to glutathione agarose. GST-Clk/Sty was eluted from the glutathione agarose using 30 mM glutathione, and concentrated into Buffer B (Buffer A without 1% Triton). GST and GST-ASF RS were induced in bacteria, harvested in PBS with 1% Triton and 1% Tween, eluted from glutathione agarose with 10 mM glutathione, and concentrated into 40 mM Hepes pH 7.5.
ASF/SF2 and ASFΔRS (Zuo and Manley, 1993) were purified by nickel chromatography as described previously (Ge et al., 1991).

Histone H1 and enolase were purchased from Boehringer Mannheim and myelin basic protein (MBP) and β-casein were purchased from Sigma. GST-c-Jun 5-89 was a gift of Jim Woodgett. Enolase was acid-treated in 25 mM acetic acid at 30°C for 15 minutes.

**In vitro kinase reaction** - GST-Clk/Sty (500 ng) was incubated in 50 μL of 40 mM Hepes pH 7.5, 10 mM MgCl₂, 20 μM ATP, 2 mM DTT, and 2 μCi of [γ³²P]ATP for 30 minutes at room temperature. The reaction was terminated by boiling in 50 μL of 2 X SDS sample buffer. Where indicated, 1 μg of exogenous substrate was added. For the kinase specificity assay, 150 ng of Clk/Sty, cAPK, Protein Kinase C (PKC), Casein Kinase II (CKII), and ERK1/ERK2 were used to phosphorylate 1 μg of ASF/SF2 or an exogenous substrate as indicated. PKC, ERK1/ERK2 and CKII were a gift of Jim Woodgett, and cAPK was purchased from Sigma.

**In vivo labelling** - COS-1 cells were transfected with RC-ASF (tagged with T7 tag epitope) using lipofectin. Approximately 68 hours post-transfection, the medium was replaced with 5 mCi of [³²P] orthophosphate in DMEM - phosphate medium plus 10% fetal bovine serum. The cells were incubated for 4 hours, and then harvested in NP40-RIPA buffer (150 mM NaCl, 50 mM Tris HCl pH 7.5, 1% sodium deoxycholate, 1% NP40, 0.1% SDS, 10 mM NaF, 1 mM DTT, 2 mM EDTA, 2 μg/mL aprotinin, 5 μg/mL leupeptin, 200 μg/mL phenylmethylsulphonyl fluoride (PMSF)). The supernatant was incubated with an α-T7 Tag antibody (from Novagen) bound to mouse protein A sepharose for 1.5 hours. The immunoprecipitates were washed four times with NP40-RIPA buffer, and boiled in 2 X SDS sample buffer.

**Phosphoamino acid analysis** - The samples were resolved by SDS-PAGE, and subsequently
transferred to PVDF membrane. The [32P]-labelled bands were located using autoradiography, excised, hydrolysed, and subjected to two-dimensional electrophoresis (Boyle et al., 1991) in the presence of non-radioactive markers.

**Tryptic mapping** - Pertinent bands were excised from the polyacrylamide gel. The proteins were then eluted, precipitated, and trypsin digested (Boyle et al., 1991). After digestion, the samples were spotted on a thin-layer cellulose plate, and subjected to electrophoresis for 60 minutes at 1000 V in pH 1.9 buffer followed by chromatography using phosphochromatography buffer (Boyle et al., 1991). For the mixes, equal amounts of radioactivity of each constituent were mixed and spotted on the same plate.

**Immunofluorescence assay** - COS-1 cells were transfected with plasmids encoding M-Clk/Sty or M-Clk/StyK190R, both of which contain the Myc epitope (Duncan et al., 1995). Cells were fixed, and probed as described (Duncan et al., 1995). For double staining by indirect immunofluorescence, fixed cells were incubated with mAB9E10 (α-Myc, 25 μl of culture supernatant) (Evan et al., 1985), followed by secondary fluorescein-conjugated goat α-mouse IgG (1:100, Jackson Immunoresearch Laboratories) specific for mouse IgG and subsequently incubated with mAb104 (α-SR, 25 μl culture supernatant) (Roth et al., 1990) followed by secondary biotin-labelled α-mouse IgM (1:100, Amersham) and streptavidin coupled to Texas Red (1:100, Amersham). Cells were viewed by confocal microscopy using an upright Leica Confocal Laser Scanning Microscope equipped with a 55 mW Krypton/Argon air-cooled laser and a 25X or 63X Plan Apo oil immersion lens.

**Triton X-100 extraction** - Transfected cells were extracted on coverslips (for immunofluorescence) or in microcentrifuge tubes (for immunoblotting) with extraction buffer (20
mM Tris HCl pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 2 mM NaF, 2mM sodium pyrophosphate, 500 μM sodium vanadate, 200 μg/ml PMSF, 2 μg/ml aprotinin, 5 μg/ml leupeptin) for 30 minutes at 4°C. For immunofluorescence, coverslips were washed with PBS, and fixed and processed as above. For immunoblotting, lysates were cleared by centrifugation. Supernatants (Triton-extracted fractions) and pellets (Triton-insoluble fraction) were boiled in SDS buffer and samples resolved by 10% SDS-PAGE and transferred to nitrocellulose membrane. Membranes were probed with anti-Myc mAb as described above.

Results

The Clk/Sty kinase interacts with splicing factors - To investigate Clk/Sty function, I performed a yeast two-hybrid screen to isolate proteins that interact with the mouse Clk/Sty kinase. The full-length Clk/Sty cDNA was fused to a cDNA encoding the GAL4 DNA-binding domain in order to generate the pAS-Clk/Sty construct. Yeast strain Y153 carrying the pAS-Clk/Sty vector was transformed with an unstimulated murine T cell cDNA plasmid library in which cDNAs were fused to the coding sequence for the GAL4 DNA activation domain (Durfee et al., 1993). I screened 2.5 million colonies containing both Clk/Sty and cDNA plasmids for stimulation of two reporter genes, β-galactosidase and HIS3. Eighty-six colonies were positive in the initial screen, indicating in each case a potential interaction between Clk/Sty and a protein from the cDNA library. I then purified the positive cDNA plasmids from yeast and selected in E. coli.

To eliminate false positives, yeast strain Y153 was transformed with the purified clones to check whether the encoded proteins stimulated the reporters on their own. In addition, Y153 carrying the clones were mated to Y187 carrying one of the following plasmids: pAS1, pAS-
Clk/Sty, pAS-SNF1, and pAS-CDK2 (Harper et al., 1993). Thirty three clones that interacted specifically with Clk/Sty were analyzed further.

I sequenced at least 100 basepairs of every clone and the resulting sequences were used to search GenBank. Due to redundancy in the library, the actual number of different clones was 26. Of these cDNAs, five encode RNA-binding proteins: hnRNP G (Soulard et al., 1993), RNP S1 (Schmidt and Werner, 1993), ASF/SF2 (also known as SRp30a) (Ge et al., 1991; Krainer et al., 1991), X16 (also known as SRp20) (Ayane et al., 1991), and SRp55 (Screaton et al., 1995) (figure 2.1). The last three clones are members of the serine/arginine-rich (SR) family of splicing factors (Zahler et al., 1992). Clone 9.11 is 84% identical at the amino acid level to human SRp55 over the region sequenced. In comparison, the human and mouse forms of both X16 and ASF/SF2 are 100% identical to one another at the amino acid level (Zahler et al., 1992; Tacke et al., 1992). Therefore, SRp55 is a less conserved member of the SR family.

Three independent clones of X16 and one clone of the other four RNA-binding proteins were isolated. With a single exception, the fusion proteins encoded by these cloned cDNAs begin within the first 45 residues of the corresponding full-length proteins (figure 2.1) and extend to their C-terminal ends. Clone 9.11, in contrast, terminates within the RS domain of SRp55.

All five of these cloned proteins contain RNA-binding domains known as RNA Recognition Motifs (RRM) at their N-termini (figure 2.1) (for review, see chapter one). The RNA-binding regions of ASF/SF2 and SRp55 contain a second RRM. The two domains of ASF/SF2 act synergistically to bind RNA (Cáceres and Krainer, 1993; Zuo and Manley, 1993).

A common feature of the five isolated RNA-binding proteins is the presence of a serine/arginine-rich domain (the RS domain) characterized by stretches of alternating serine and
Percentage of Arginine and Serine in the RS region

<table>
<thead>
<tr>
<th>Protein</th>
<th>R</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLK/STY</td>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td>hnRNP G</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>RNP S1</td>
<td>21</td>
<td>11</td>
</tr>
<tr>
<td>X16 (SRp20)</td>
<td>40</td>
<td>29</td>
</tr>
<tr>
<td>ASF/SF2 (SRp30a)</td>
<td>36</td>
<td>40</td>
</tr>
<tr>
<td>SRp55</td>
<td>28</td>
<td>38</td>
</tr>
</tbody>
</table>

Figure 2.1. A schematic of Clk/Sty and five interacting RNA-binding proteins. Each protein has an RS domain, and the percentage of arginine and serine residues is indicated. The five RNA-binding proteins have an RNA-binding domain known as the RNA recognition motif (RRM), whereas Clk/Sty has a kinase domain. The RNA-binding regions of ASF/SF2 and SRp55 contain an additional RRM. The bent arrows indicate where the cloned fusion proteins start in relation to the intact proteins. The clones are numbered after the arrows. The vertical arrows in Clk/Sty and ASF/SF2 denote the corresponding amino acid position of alternative splice sites. The horizontal arrow on SRp55 indicates where clone 9.11 overlaps. Abbreviations: RS - arginine/serine-rich; RRM - RNA recognition motif; RSG - arginine/serine/glycine-rich; L - leucine-rich; P - proline-rich; G - glycine-rich.
arginine residues. ASF/SF2 is alternatively spliced to produce three isoforms, ASF1, ASF2, and ASF3 of which ASF2 and ASF3 lack the RS domain (Ge et al., 1991). Sequence analysis confirmed that the ASF/SF2 isoform cloned through its interaction with Clk/Sty was ASF1, the splice variant which retains the RS domain.

**Analysis of protein domains required for interactions of Clk/Sty with SR proteins** - To investigate whether the interaction of Clk/Sty with RNA-binding proteins is dependent on kinase activity, I tested a mutant form of Clk/Sty lacking catalytic activity for its binding properties in the two-hybrid system. In this mutant (pAS-Clk/Sty$^{K190R}$) the invariant lysine (lys 190) of the Clk/Sty kinase domain, which is required for kinase activity, was replaced with arginine (Ben-David et al., 1991). In the two-hybrid assay, kinase-inactive Clk/Sty bound to the RNA-binding proteins with similar efficiency to wild-type (data not shown).

In addition to the C-terminal kinase domain, Clk/Sty possesses an N-terminal non-catalytic region that is itself serine/arginine-rich (figure 2.1). To further delineate the sites of interaction between Clk/Sty and the RNA-binding proteins, I divided Clk/Sty into its kinase domain and its N-terminal RS domain. The kinase domain alone (pAS-Clk/Sty$^{KIN}$) did not interact detectably with the RNA-binding proteins (data not shown). In contrast, the N-terminal RS region of Clk/Sty (pAS-Clk/Sty$^{RS}$) when expressed in the two-hybrid system associated with RNP S1, ASF/SF2, and X16, but not SRp55 or hnRNP G (data not shown).

RS domains have been implicated in protein-protein interactions (Amrein et al., 1994; Wu and Maniatis, 1993; Kohtz et al., 1994). I tested the RS domains of X16 and ASF/SF2 for binding to the four pAS-Clk/Sty polypeptides in the two-hybrid screen. The X16 RS domain bound weakly to pAS-Clk/Sty and pAS-Clk/Sty$^{K190R}$ but not to pAS-Clk/Sty$^{KIN}$ or pAS-Clk/Sty$^{RS}$
The RS domain of ASF/SF2 did not by itself bind to any of the four Clk/Sty proteins (data not shown). Since the RS domains of X16 and ASF/SF2 were not sufficient to bind Clk/Sty, the RNA-binding domain of X16 was tested in the two-hybrid system. Like the RS domains, the RNA-binding domain of X16 was unable to bind any of the four Clk/Sty constructs. These results indicate that although the RS domain of Clk/Sty is sufficient to bind a subset of the RNA-binding proteins, more extensive contacts along both binding partners are in general required for efficient association of Clk/Sty with SR splicing factors.

**Phosphorylation of ASF/SF2** - The interaction between Clk/Sty and SR proteins suggests that these RNA-binding proteins might be substrates for Clk/Sty phosphorylation. I, therefore, tested the ability of Clk/Sty to phosphorylate the SR protein ASF/SF2 in vitro. A bacterially expressed GST-Clk/Sty fusion protein (Duncan *et al.*, 1995) phosphorylated recombinant ASF/SF2 in vitro (figure 2.2A, lane 3). To localize the regions of ASF/SF2 which were phosphorylated by Clk/Sty, I separated ASF/SF2 into its RNA-binding and RS domains. For this purpose, two polypeptides were expressed in bacteria: ASFARS which retains the RNA-binding domain but lacks the RS domain (Zuo and Manley, 1993) and GST-ASF RS, which contains only the RS domain. Phosphorylation of ASFARS by Clk/Sty was markedly reduced compared to intact ASF/SF2 (figure 2.2A, lane 4). In contrast, Clk/Sty phosphorylated GST-ASF RS, which possesses the isolated RS domain, to a similar level as intact ASF/SF2 protein (figure 2.2A, lane 5). Phosphorylation of the GST-ASF RS polypeptide was likely to be specific to the RS domain, as GST alone was not phosphorylated by Clk/Sty (figure 2.2A, lane 6). These results indicated that the ASF/SF2 RS domain contains the major sites of in vitro phosphorylation by Clk/Sty.
Figure 2.2. Phosphorylation of ASF/SF2. A) CLK/Sty phosphorylation of ASF/SF2 in an *in vitro* kinase assay. GST-CLK/Sty (500 ng) was incubated alone (lane 2) or with 1 μg of ASF/SF2 (lane 3), ASFΔRS (lane 4), GST-ASF RS (lane 5), or GST (lane 6). As a negative control, GST alone was incubated with ASF/SF2 (lane 1). The samples were resolved on an SDS 12% polyacrylamide gel and exposed to autoradiography. B) Phosphorylation of ASF/SF2 *in vivo*. ASF/SF2 was transfected into COS-1 cells, labelled with $^{32}$P orthophosphate, immunoprecipitated with an α-T7 Tag antibody, purified by SDS-PAGE, and exposed to autoradiography.
To confirm that ASF/SF2 is a phosphoprotein in vivo, COS-1 cells were transfected with an expression vector encoding ASF/SF2 with an N-terminal T7 epitope tag, and labelled with \[^{32}\text{P}]\text{orthophosphate}. ASF/SF2 was immunoprecipitated with an \(\alpha\)-T7 Tag antibody and the isolated protein was further purified by SDS gel electrophoresis. A single band of phosphorylated ASF/SF2 was detected (figure 2.2B).

**Specificity of Clk/Sty protein kinase activity** - To determine if ASF/SF2 is a preferred substrate of Clk/Sty, I tested other exogenous substrates for phosphorylation by Clk/Sty in an in vitro kinase assay. GST-Clk/Sty was unable to phosphorylate \(\beta\)-casein, enolase, or a GST-c-Jun N-terminal fusion protein (containing c-Jun residues 5 to 89) (figure 2.3A). Clk/Sty was able to phosphorylate the basic proteins H1 histone and myelin basic protein (MBP) although not to the same extent as ASF/SF2 (figure 2.3A).

Previous work has shown that ASF/SF2 is a substrate for SRPK1 but not p34\(^{cdc2}\) (Gui \textit{et al.}, 1994a). Given the key role of ASF/SF2 in alternative splicing, it is likely to be regulated by multiple kinases. I also tested the ability of other protein kinases, including ERK1/ERK2, protein kinase C (PKC), casein kinase II (CKII), and cAPK to phosphorylate ASF/SF2 in vitro. ASF/SF2 has potential phosphorylation motifs for each of these kinases (Pearson and Kemp, 1991) with the exception of CKII, which has an acidic recognition site. Neither ERK1/ERK2 nor CKII were able to phosphorylate ASF/SF2. However, both PKA and PKC induced ASF/SF2 phosphorylation in vitro. Taken together, these results suggest that of the substrates tested, ASF/SF2 is a preferred substrate for Clk/Sty. It is apparent from these and previously published data that ASF/SF2 is potentially phosphorylated by several kinases (Gui \textit{et al.}, 1994a; Woppmann \textit{et al.}, 1993).
Figure 2.3. Specificity of the phosphorylation of ASF/SF2 by Clk/Sty. A) 150 ng of GST-Clk/Sty was incubated alone (lane 1) or with ASF/SF2 (lane 2), β-casein (lane 3), acid-treated enolase (lane 4), H1 histone (lane 5), GST-c-Jun 5-89 (lane 6), and myelin basic protein (MBP, lane 7). B) 150 ng of kinase was tested for its ability to phosphorylate ASF/SF2 or a known substrate of the respective kinase. Lanes 1 and 2: Clk/Sty incubation with ASF/SF2 and H1 histone, lanes 3 and 4: ERK1/ERK2 incubation with ASF/SF2 and MBP, lanes 5 and 6: PKC incubation with ASF/SF2 and MBP, lanes 7 and 8: CKII incubation with ASF/SF2 and β-casein, lanes 9 and 10: cAPK incubation with ASF/SF2 and H1 histone. Kinase assays were terminated after 30 minutes by addition of 2 X SDS sample buffer, and the samples were loaded onto a SDS polyacrylamide gel. Abbreviations: Clk - Clk/Sty, ERK - ERK1/ERK2, A - ASF/SF2, H - H1 histone, M - MBP, β - β-casein.
**Phosphoamino acid analysis of ASF/SF2** - Phosphoamino acid analysis showed that both full-length ASF/SF2 phosphorylated *in vitro* by Clk/Sty and ASF/SF2 isolated from $[^{32}P]$-orthophosphate-labelled cells contained predominantly phosphoserine (figure 2.4A,B, (Gui et al., 1994a)). A trace amount of phosphothreonine was also seen in ASF/SF2 phosphorylated *in vivo* and *in vitro* by Clk/Sty. No phosphorylation on tyrosine was detected. ASFARS, which was phosphorylated *in vitro* to a very low level by Clk/Sty, contained phosphoserine and a lesser amount of phosphothreonine (figure 2.4C). The GST-ASF RS polypeptide was phosphorylated *in vitro* exclusively on serine residues (figure 2.4D).

**Comparative tryptic phosphopeptide mapping of ASF/SF2** - To investigate ASF/SF2 phosphorylation sites in more detail, full-length ASF/SF2 was phosphorylated *in vitro* by Clk/Sty and then digested with trypsin. *In vitro* phosphorylated ASF/SF2 generated twelve phosphorylated tryptic peptides (figure 2.5A). A subset of these peptides likely represent partial tryptic digestion products of ASF/SF2 due to the presence of adjacent arginine residues (in tandem), as well as the inability of trypsin to cut after an arginine in arginine-X-phosphoserine motifs (Hardie et al., 1993). Indeed, the presence of five of the twelve phosphopeptides (peptides A to D, and peptide 2, figure 2.5F) was variable from one experiment to another (compare figure 2.5A with 2.5B). I investigated the location of the phosphopeptides identified in full-length ASF/SF2 phosphorylated by Clk/Sty by preparing a tryptic phosphopeptide map of GST-ASF RS which had been phosphorylated *in vitro* by Clk/Sty (figure 2.5C). All of the phosphopeptides detected in full-length ASF/SF2 co-migrated with phosphopeptides from GST-ASF RS which only contains the RS domain. There are two extra peptides in the map of GST-ASF RS that may represent phosphorylation of an RS dipeptide generated by the fusion of GST to the RS domain.
Figure 2.4. Phosphoamino acid analysis of ASF/SF2. Samples were hydrolysed in 5.7 M HCl for 1 hour at 100°C, and the amino acids were separated by two-dimensional electrophoresis (Boyle et al., 1991). A) ASF/SF2 phosphorylated in vitro by Clk/Sty. B) In vivo phosphorylated ASF/SF2. C) ASFΔRS phosphorylated in vitro by Clk/Sty. D) GST-ASF RS phosphorylated in vitro by Clk/Sty. The circles indicate the migration of non-radioactive standards. Abbreviations: pS - phosphoserine. pT - phosphothreonine. pY - phosphotyrosine.
Figure 2.5. Tryptic phosphopeptide maps of ASF/SF2. Protein samples were eluted from SDS polyacrylamide gels, precipitated, and digested with trypsin. Individual peptides were resolved on thin layer cellulose plates as described (Boyle et al., 1991). A) and B) ASF/SF2 phosphorylated in vitro by Clk/Sty, C) GST-ASF RS phosphorylated in vitro by Clk/Sty, D) In vivo phospholabelled ASF/SF2, and E) A mix of ASF/SF2 phospholabelled in vitro and in vivo. F) A schematic representation of the phosphopeptide map. Open circles correspond to peptides found only in ASF/SF2 phosphorylated in vivo. Closed circles represent peptides phosphorylated both in vivo, and in vitro by Clk/Sty. Grey circles indicate peptides found only in ASF/SF2 phosphorylated in vitro by Clk/Sty. The origin, also coloured in grey, is shown at the left.
These results confirm that ASF/SF2 is phosphorylated primarily within its RS domain by Clk/Sty.

A tryptic phosphopeptide map of ASF/SF2 labelled in vivo (figure 2.5D) showed several peptides (1, 2, 5, 6, 7, 8, and 9) in common with ASF/SF2 phosphorylated in vitro by Clk/Sty. In vivo labelled ASF/SF2 lacked peptides E and F found in the in vitro phosphorylated ASF/SF2 and contained 3 additional peptides (3, 4, and 10). The lack of peptides E and F in the in vivo tryptic map may be due to inaccessibility of these sites owing to the formation of protein complexes in vivo, phosphatase activity, or to reduced phosphorylation due to ASF/SF2 overexpression in COS-1 cells (Romac and Keene, 1995). A mix of tryptic digests of [32P] labelled ASF/SF2 phosphorylated either in vivo and in vitro (figure 2.5E) illustrates the presence of the co-migrating phosphopeptides. In aggregate, these results suggest that Clk/Sty phosphorylates ASF/SF2 at physiological sites within the RS domain.

**Subnuclear localization of Clk/Sty and SR proteins in vivo** - If Clk/Sty interacts with SR proteins in mammalian cells then these polypeptides might be anticipated to have a common subcellular localization. During interphase, pre-mRNA splicing factors, including the SR proteins, are found localized in nuclear structures termed speckles (Verheijen et al., 1986; Spector et al., 1983; Fu and Maniatis, 1990). To investigate this point, COS-1 cells were transiently transfected with plasmids encoding either full-length Clk/Sty (M-Clk/Sty) or a kinase-inactive Clk/Sty (M-Clk/StyK190R) (Duncan et al., 1995). These proteins were fused to the myc epitope to allow detection by mAb9E10 (Evan et al., 1985). Cells were subsequently probed with mAb104 which recognizes SR proteins (Roth et al., 1990) and mAb9E10 to detect transfected Clk/Sty proteins, and were analyzed by indirect immunofluorescence followed by confocal microscopy imaging. Representative fields are shown in figure 2.6. M-Clk/StyK190R was found
Figure 2.6. Nuclear localization of Clk/Sty and its effects on the distribution of SR proteins.

Indirect immunofluorescent staining of transfected COS-1 cells using the anti-Myc mAb9E10 (α-Myc, a,d,g,j,m), anti-SR mAb104 (α-SR, b,e,h,k,n), or an overlay of the two signals (α-Myc + α-SR, c,f,i,l,o). Alignment of green and red signals appears yellow. Both low power magnification (25X) of M-Clk/StyK190R transfected COS-1 cells (d-f) and M-Clk/Sty transfected cells (j-l), and high power magnification (63X) of untransfected COS-1 cells (a-c), M-Clk/StyK190R (g-i) and M-Clk/Sty (m-o) transfected cells is shown.
<table>
<thead>
<tr>
<th></th>
<th>$\alpha$-Myc</th>
<th>$\alpha$-SR</th>
<th>$\alpha$-Myc + $\alpha$-SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>l</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
predominantly in nuclear speckles, where it co-localized with endogenous SR proteins (figure 2.6d-i). In contrast, cells expressing catalytically active M-Clk/Sty displayed a diffuse nucleoplasmic staining pattern for both the kinase and SR proteins (figure 2.6j-o). Untransfected cells adjacent to those expressing M-Clk/Sty displayed a normal distribution of SR proteins in nuclear speckles (figure 2.6a-c). As a control, the experiment was repeated using an HA-tagged Clk/Sty to demonstrate that Clk/Sty is responsible for the localization and not the Myc epitope tag (data not shown). Identical results were also obtained using an antibody specific for a single SR protein family member, SC35 (data not shown) (Fu and Maniatis, 1990).

Since anti-SC35 and mAB104 recognize phosphoepitopes, it is conceivable that the dispersion of the immunofluorescence signal representing the SR family seen when Clk/Sty is overexpressed may be due to phosphorylation of SR members outside of the speckle once Clk/Sty is released from the speckles. To ensure that overexpression of Clk/Sty caused disassembly of speckles and redistribution of SR family members rather than phosphorylation of SR members outside the speckles, the anti-Sm antibody which recognizes components of the snRNPs was used (Spector et al., 1983). When M-Clk/Sty

$^{\text{K190R}}$

is overexpressed, the Sm proteins are co-localized with M-Clk/Sty

$^{\text{K190R}}$

in speckles (data not shown). Overexpression of M-Clk/Sty causes a redistribution of the Sm proteins similar to the SR family indicating that Clk/Sty does indeed cause disassembly of nuclear speckles and subsequent redistribution of speckle proteins (data not shown).

The apparently contrasting subnuclear localization of M-Clk/Sty and M-Clk/Sty

$^{\text{K190R}}$

proteins as detected by immunofluorescent staining was supported by the observation that these two proteins display differential solubility in a non-ionic detergent. COS-1 cells transfected with
either M-Clk/Sty or M-Clk/Sty\textsuperscript{K190R} were separated into Triton X-100-soluble and -insoluble fractions as described in Materials and Methods and analyzed by immunoblotting (figure 2.7A). While essentially all of the M-Clk/Sty protein was soluble in Triton X-100, a significant portion of the M-Clk/Sty\textsuperscript{K190R} protein remained insoluble. It is of interest to note that the active kinase, M-Clk/Sty, migrates as a broad band while the inactive point-mutant, M-Clk/Sty\textsuperscript{K190R}, migrates as a discrete band following SDS-PAGE. This banding pattern suggests multiple phosphorylation states of the active kinase as has been observed for other kinases such as ERK1/ERK2 and Wee1 (Gotoh \textit{et al.}, 1991; Tang \textit{et al.}, 1993).

When Triton X-100-treated COS-1 cells were subjected to analysis by immunofluorescence, the speckled pattern of M-Clk/Sty\textsuperscript{K190R} remained intact (figure 2.7B) while the wild-type protein staining pattern was lost (data not shown). Therefore, the catalytic activity of M-Clk/Sty may be involved in the regulation of its subnuclear localization.

\textbf{Discussion}

\textit{Clk/Sty interacts with RNA-binding proteins} - I have isolated potential binding partners for the Clk/Sty protein kinase by screening a T cell cDNA library using the yeast two-hybrid system. Five of the clones isolated are RNA-binding proteins, and three are members of the SR family of essential splicing factors. Additional clones may represent novel serine/arginine-rich proteins (see chapter 4). The function of hnRNP G or RNP S1 has not been ascertained. hnRNP G is a member of the heterogeneous nuclear ribonucleoprotein family that is involved in pre-mRNA processing (for review, see Dreyfuss \textit{et al.}, 1993). SR proteins, such as ASF/SF2, X16, and SRp55 are required for splicing, and play a role in selection at the 5\textsuperscript{'} splice site (Zahler \textit{et al.}, 1992; Zahler \textit{et al.}, 1993b).
Figure 2.7. Active and inactive Clk/Sty kinases display differential solubility in non-ionic detergent. A) COS-1 cells transfected with M-Clk/Sty or a catalytic mutant M-Clk/Sty<sup>K190R</sup> were lysed in Triton X-100-containing buffer to yield a Triton-soluble fraction (S) and a Triton-insoluble fraction (I). Extracts were analysed by SDS-PAGE, and immunoblotting with the anti-Myc antibody. Positions of molecular mass markers (kDa) are indicated to the left. B) the Triton X-100-insoluble M-Clk/Sty<sup>K190R</sup> co-localized with SR splicing proteins. COS-1 cells transfected with M-Clk/Sty<sup>K190R</sup> and grown on coverslips were extracted with Triton X-100-containing buffer. Following extractions, cells were fixed and processed for indirect immunofluorescence by confocal microscopy using the anti-Myc antibody (α-Myc) and the anti-SR antibody (α-Sr). a) α-Myc, b) α-Sr, c) overlay of the α-Myc and α-SR signals. Alignment of the green and red signals appear yellow.
While the Clk/Sty kinase domain was not sufficient for a positive interaction with these RNA-binding proteins, the isolated N-terminal RS region of Clk/Sty associated with ASF/SF2, X16, and RNP S1. This result is consistent with previous data implicating RS domains in protein-protein interactions (Amrein et al., 1994; Wu and Maniatis, 1993; Kohtz et al., 1994). A characteristic of previously described RS domains is the presence of repeating RS/SR dipeptides. Although Clk/Sty does not contain long stretches of RSRS repeats, it contains ten RS/SR dipeptides and one RSRS motif. However, the N-terminal domain of Clk/Sty possesses one of the main features of known RS domains, which is an ability to interact with SR proteins. The RS domain of Clk/Sty corresponds to an alternatively spliced truncated Clk/Sty polypeptide (Clk/Sty') whose function is unknown (Duncan et al., 1995). The tight association of the Clk/Sty RS domain with SR proteins suggests that this splice variant could be involved in the regulation of SR protein compartmentalization. This could occur by the formation of inactive Clk/Sty:Clk/Sty' heterodimers or by the sequestering of SR proteins by Clk/Sty' (Thomis and Samuel, 1993; Reith et al., 1991).

Although the Clk/Sty kinase domain on its own could not detectably interact with any of the RNA-binding proteins, it probably participates in binding as not all of the RNA-binding proteins could interact with the isolated Clk/Sty RS domain. The finding that an inactive Clk/Sty protein was able to bind the RNA-binding proteins is not surprising, as contacts between a kinase and its substrate are not limited to the catalytic site but instead extend throughout the kinase domain (Knighton et al., 1991b).

In reciprocal experiments, the RS domains of X16 and ASF/SF2 as well as the RNA-binding domain of X16 were not sufficient for the efficient interaction of either protein with
Clk/Sty, indicating that both domains are required for recognition by Clk/Sty. Taken together, these results suggest that the RS region of Clk/Sty is important for its recognition of SR splicing factors, but that extensive contacts are likely made between the kinase and its binding partners. **ASF/SF2 is a candidate physiological substrate of Clk/Sty** - Phosphorylation-dephosphorylation of splicing factors may be an important aspect of splicing regulation (Mermoud et al., 1992; Tazi et al., 1992), since SR protein function *in vitro* can be affected by phosphorylation (Gui et al., 1994a; Mermoud et al., 1994). *In vitro*, Clk/Sty phosphorylated ASF/SF2, predominantly in the ASF/SF2 RS domain. This conclusion is based on the observation that the isolated ASF/SF2 RS domain was phosphorylated to a level comparable to the full-length protein whereas ASFARS, which lacks the RS domain, was only very weakly phosphorylated. In addition, all of the tryptic phosphopeptides observed in full-length ASF/SF2 phosphorylated by Clk/Sty can be accounted for by the RS domain alone. However, ASF/SF2 phosphorylated *in vivo* or *in vitro* by Clk/Sty displayed a trace amount of phosphothreonine. Since, phosphothreonine was seen only in the RNA-binding domain and not in the RS domain, it is likely that the RNA-binding domain was phosphorylated to a minor extent by Clk/Sty.

Other potential protein substrates were used to test the specificity of Clk/Sty. Clk/Sty phosphorylated the basic proteins H1 histone and MBP although more weakly than ASF/SF2, but not more acidic substrates such as β-casein, enolase, or a GST fusion protein containing the c-Jun amino-terminal region. It seems likely that Clk/Sty recognizes the site to be phosphorylated in the context of basic amino acids as found in the RSRS stretch of ASF/SF2. These results indicate that ASF/SF2 is a preferential substrate for Clk/Sty.

The finding that the tryptic phosphopeptide maps of ASF/SF2 labelled *in vivo* or
phosphorylated in vitro by Clk/Sty are similar to one another is consistent with the suggestion that Clk/Sty phosphorylates ASF/SF2 in intact cells, and may therefore regulate its activity. Additional sites are phosphorylated in vivo that cannot be accounted for by Clk/Sty phosphorylation, suggesting that Clk/Sty is not the only kinase that phosphorylates SR proteins in vivo. Indeed, three other kinases have been identified that can phosphorylate ASF/SF2 in vitro, SRPK1 (Gui et al., 1994a), an uncloned U1 70K kinase (Woppmann et al., 1993), and DNA topoisomerase I (Rossi et al., 1996). As well, we have shown here that PKC and PKA are able to phosphorylate ASF/SF2 in vitro. It remains to be seen if either of these latter two kinases, like Clk/Sty, is able to phosphorylate ASF/SF2 on sites that are also phosphorylated in vivo. SRPK1, the first SR family kinase to be identified, phosphorylates ASF/SF2 on sites in vitro that comprise the phosphoepitope recognized by mAB104, indicating that SRPK1 is likely to regulate ASF/SF2 function in vivo (Gui et al., 1994b).

Comparison of Clk/Sty and SRPK1 - SRPK1 is most closely related to a Schizosaccharomyces pombe kinase DSK1 and a Caenorhabditis elegans Clk/Sty-like hypothetical kinase (CEHK) (Takeuchi and Yanagida, 1993; Wilson et al., 1994). Interestingly, DSK1 autophosphorylates on all three amino acids identifying it as a dual specificity kinase similar to Clk/Sty (Takeuchi and Yanagida, 1993).

Across the kinase domain, there are interesting relationships between Clk/Sty and the SRPK1 family. The entire EHLAMMERILGPLP sequence is not conserved, but there are two invariant amino acids that correspond to His 386 and Pro 398 of Clk/Sty (figure 2.8). Clk3 has a valine at position 388, which is a conservative change from the alanine found in the other family members. These three amino acids are not conserved among other kinases. Two other
Figure 2.8. Alignment of the LAMMER motif. The LAMMER motif of the LAMMER family kinases is aligned with the analogous regions in SRPK1, DSK1, the C. elegans kinase, CDK2, and SNF1. The last amino acid in the motif is indicated on the right. Amino acids that are conserved among 80% of the aligned kinases are highlighted in black. Amino acids that are identical in 60% to 80% of these kinases are bolded. Abbreviations: mClk - mouse Clk/Sty, huCLK - human CLK, CEHK - C. elegans hypothetical kinase.
protein kinases, CDK2 and SNF1, did not interact with RNA-binding proteins in the two-hybrid assay (data not shown) and their equivalent sequences are shown for comparison.

Besides the residues conserved in the LAMMER motif, there are other key amino acids that are identical between Clk/Sty and SRPK1. In the catalytic loop (subdomain 6), an arginine which is otherwise highly conserved among protein kinases is replaced by a threonine in both Clk/Sty and SRPK1 (Thr 286 of Clk/Sty). In the cAMP dependent kinase structure, this arginine residue interacts with phosphorylated threonine 197 to form a pocket for the amino acid (P+1) following the hydroxyamino acid being phosphorylated (Knighton et al., 1991b). The substitution of this arginine for threonine may affect the pocket for the P+1 amino acid and possibly the orientation of a peptide substrate in the active site.

The Clk/Sty and SRPK1 kinase families also have a conserved arginine corresponding to amino acid 345 of Clk/Sty. In the ERK2 structure, this arginine has been shown to form part of the P+1 binding site (Zhang et al., 1994). The conserved regions in these two kinase families may help explain why the two kinases phosphorylate the same protein.

**Regulation of SR splicing factors by Clk/Sty in vivo** - It has been shown previously that Clk/Sty is localized to the nucleus (Duncan et al., 1995). In this study, a kinase-inactive mutant of Clk/Sty was localized to discrete speckles within the nucleus where SR proteins are located. In contrast to the Clk/Sty kinase-inactive mutant, wild-type Clk/Sty displayed a more diffuse nuclear distribution. *In vivo*, kinase-inactive Clk/Sty appeared to be hypophosphorylated, whereas the wild-type protein was hyperphosphorylated, presumably as a consequence of autophosphorylation (Duncan et al., unpublished results, and figure 2.7). This result indicates that phosphorylation may oppose retention in the speckles. This hypothesis is further borne out
by the loss of SR proteins from the speckles when wild-type Clk/Sty was overexpressed. I propose that the SR proteins and Clk/Sty are targeted to the speckles through their RS domain (Li and Bingham, 1991), but are released from these sites upon phosphorylation. The redistribution of proteins from speckles after phosphorylation was first reported in permeabilized cells that overexpress the SRPK1 protein kinase, indicating that speckle disassembly may be a common phenomenon after SR family phosphorylation (Gui et al., 1994a).

Nuclear speckles have been proposed to act as storage sites for splicing components (Zhang et al., 1994; Huang et al., 1994; Spector, 1993; Fakan et al., 1984). There is no evidence that speckles are disassembled prior to the onset of splicing (Zhang et al., 1994), and this may occur only when the kinase responsible for the release of factors is overexpressed. One possibility is that at physiological levels of Clk/Sty, only a subset of SR factors are phosphorylated and released from the speckles. The availability of a subset of SR factors may be a significant regulatory mechanism, given the influence of the relative concentration of individual splicing factors on pre-mRNA processing (Ge and Manley, 1990; Krainer et al., 1990; Zahler et al., 1993a; Fu, 1993). Protein phosphatase 1 has been shown to affect 5' splice site selection in vitro, indicating that phosphorylation plays a role in alternative splicing (Cardinali et al., 1994). Indeed, preliminary results indicate that overexpression of Clk/Sty in 293 cells can alter 5' splice site selection in vivo (J. Prasad, K. Colwill, T. Pawson, J.L. Manley, unpublished results).

Due to the complexity of factors involved in pre-mRNA processing, it seems likely that multiple kinases, perhaps including other members of the LAMMER family, regulate the many facets of splicing. I propose that Clk/Sty and SRPK1 are two members of a family of kinases
that regulate SR protein trafficking and activity.
CHAPTER 3: SRPK1 and Clk/Sty Protein Kinases Show Distinct Substrate Specificities for SR Splicing Factors

I did the work reported in this chapter with the following exceptions:

Lana Feng, Joanne Yeakley, and Xiang-Dong Fu made the Baculovirus SRPK1 construct, tested various RS proteins and ASF/SF2 mutants as substrates for Clk/Sty and SRPK1, performed the kinetics of phosphorylation of ASF/SF2 by SRPK1 and Clk/Sty, and examined the relative activity of the two kinases isolated from rabbit reticulocyte lysates.

Gerald Gish synthesized the RSRS and SPRY peptides.

Chapter 3 is a modified version of the following publication:


© 1996 by Journal of Biological Chemistry/American Society for Biochemistry and Molecular Biology
Introduction

Serine/arginine-rich (SR) proteins are essential for pre-mRNA splicing, and modify the choice of splice site during alternative splicing in a process regulated in part by protein phosphorylation (see chapter 1). Several protein kinases phosphorylate SR proteins in vitro including SRPK1 and Clk/Sty (see chapters 1 and 2). In this chapter, it is shown that SRPK1 and Clk/Sty phosphorylate the same SR proteins in vitro, but that SRPK1 has the higher specific activity towards ASF/SF2. I show that SRPK1, like Clk/Sty, phosphorylates ASF/SF2 in vitro on sites that are also phosphorylated in vivo. Tryptic peptide mapping of ASF/SF2 revealed that three of the phosphopeptides from full-length ASF/SF2 phosphorylated in vitro contain consecutive phosphoserine-arginine residues or phosphoserine-proline residues. In vitro, the Clk/Sty kinase phosphorylated Ser-Arg, Ser-Lys, or Ser-Pro sites whereas SRPK1 had a strong preference for Ser-Arg sites. From these results, I suggest that SRPK1 and Clk/Sty may play different roles in regulating SR splicing factors, and suggest that Clk/Sty has a broader substrate specificity than SRPK1.

Materials and Methods

Protein purification and antibody production - GST-ASF RS (chapter 2) and GST-Clk/Sty (Duncan et al., 1995) were purified as described previously (chapter 2) except GST-Clk/Sty was eluted in 100 mM Tris pH 7.5, 120 mM NaCl, 20 mM glutathione, 2 mM dithiothreitol (DTT), 2 mM benzamidine, and used directly in kinase assays. ASF/SF2, ASFΔRS, and the RS domain mutants of ASF/SF2 were purified from bacteria by nickel chromatography (Ge et al., 1991; Zuo and Manley, 1993; Cáceres and Krainer, 1993). SRPK1 was cloned into a baculovirus expression vector pAcSecG2T (PharMingen) and expressed as a GST-fusion protein, which was purified the
same way as GST-Clk/Sty. For substrate comparison between SRPK1 and Clk/Sty, ASF/SF2, SC35, SRp55, and U2AF65 were prepared from Baculovirus as described previously (Gui et al., 1994a). Histone H1 and myelin basic protein (MBP) were purchased from Boehringer and Sigma, respectively. Anti-ASF/SF2 and Anti-GST-ASF RS antibodies were prepared by immunizing New Zealand White rabbits individually with purified ASF/SF2 or GST-ASF RS.

**Peptide synthesis and purification** - Synthetic peptides (RSRS peptide - GRSRSRSRSR, SPRY peptide - RGSPRYSPRHS) were prepared using conventional Fmoc chemistry with HBTU activation on an Applied Biosystems 3H Peptide Synthesizer. Products were deprotected as outlined by the manufacturer, and the crude material was purified by reverse phase high-pressure liquid chromatography. The authenticity of the peptides was confirmed by mass spectrometry and amino acid analysis.

**In vitro kinase assays** - 100 nM GST-Clk/Sty (measured by Coomassie-stained gel using bovine serum albumin as a standard) and an equivalent kinase activity of GST-SRPK1 (determined by titration using ASF/SF2 as standard substrate) were incubated in kinase reaction buffer (40 mM Hepes pH 7.5, 120 mM NaCl, 10 mM MgCl₂, 20 μM ATP, 2 mM DTT, 2 mM benzamidine, and 2 μCi of [γ³²P] ATP) for 30 minutes. The kinase reaction was terminated by boiling in equal volume of 2 X SDS sample buffer. Samples were separated on an SDS 10% polyacrylamide gel, and the gel was exposed to autoradiography. For tryptic phosphopeptide maps, 600 nM ASF/SF2 or 500 μM peptide were added to the reaction. For comparison of substrates between GST-Clk/Sty and GST-SRPK1, 1 μM exogenous substrate was added. To test if ASF/SF2 is recognized by mAb104 after phosphorylation, 500 nM ASF/SF2, ASFARS, and GST-ASF RS were incubated plus or minus 100 nM GST-Clk/Sty in kinase reaction buffer that contained 1
mM ATP and no [$\gamma^{32}$P] ATP. These samples were resolved on an SDS 10% polyacrylamide gel, transferred to nitrocellulose, and probed with 5 ml of mAb104 conditioned medium (Roth et al., 1990) or 1:250 dilution of either anti-ASF/SF2 or anti-GST-ASF RS antibodies. Immunoblots were developed by using ECL (Amersham).

**Determination of relative specific activities of Clk/Sty and SRPK1** - The coding sequences of Clk/Sty and SRPK1 were fused with a FLAG tag sequence at their N-termini followed by cloning into the pSP73 vector from which the tagged kinases can be transcribed *in vitro* by T7 RNA polymerase (Promega). 0.5 µg of each plasmid was transcribed and translated in the TNT system (Promega) in a 25 µl reaction according to the manufacturer's protocol. 2 µl of the reaction mixture was loaded on an SDS 12.5% polyacrylamide gel followed by autoradiography for detecting *in vitro* translated product. The remaining reaction mixture was immunoprecipitated using 5 µg of an anti-FLAG monoclonal antibody (M2 from IBI) bound to protein G beads (Pharmacia) in 0.2 ml of NP40 buffer (20 mM Tris-HCL pH 7.5, 150 mM NaCl, 0.5% NP40). The kinase reaction, on beads, was initiated by adding 20 µl of kinase reaction buffer containing 1 µM ASF/SF2. The reaction was incubated at 30°C for 30 minutes. [$^{35}$S]-labelled *in vitro* translated kinases and [$^{32}$P]-labelled ASF/SF2 bands were quantitated by phosphoimaging. Relative specific activity was calculated by dividing relative kinase activity with relative protein concentration.

**ASF/SF2 and peptide kinetics** - For kinetic comparison between Clk/Sty and SRPK1 in phosphorylating ASF/SF2, 100 nM GST-Clk/Sty and an equivalent kinase activity of GST-SRPK1 were used to phosphorylate five concentrations of ASF/SF2 (0.22 µM, 0.44 µM, 0.9 µM, 1.85 µM, and 3.7 µM). For peptide kinetics, set concentrations of peptides (25 µM to 1 mM) were
phosphorylated by GST-SRPK1 or GST-Clk/Sty in peptide reaction buffer (60 mM Hepes pH 7.5, 10 mM NaF, 10 mM MgCl₂, 100 μM ATP, 2 mM DTT, 2 mM benzamidine, and 4 μCi of [γ³²P]ATP). Aliquots of the reaction were spotted on to P81 paper, washed in 75 mM orthophosphoric acid, and counted in a liquid scintillation counter as described (Casnellie, 1991). The rate of phosphorylation was determined as pmol of phosphate transferred/minute. Km and Vmax were determined by linear regression using a Lineweaver-Burk plot.

**Tryptic mapping** - Phosphopeptide mapping was performed as described in chapter 2. The SPRY and RSRS peptides, after phosphorylation by GST-Clk/Sty, were purified by electrophoresis on thin layer cellulose (TLC) plates for 25 minutes at 1000 V. The peptides were then isolated from the TLC plates as described (van der Geer et al., 1993) and digested with trypsin. For mixes, equal amounts of radioactivity of each constituent were mixed and spotted on the same plate.

**Results**

**Comparison between SRPK1 and Clk/Sty in phosphorylating RS domain-containing splicing factors** - SRPK1 phosphorylates several members of the SR family (Gui et al., 1994a). Only one member of the SR family, ASF/SF2, has been confirmed to be a substrate of Clk/Sty (chapter 2). To determine if Clk/Sty, like SRPK1, phosphorylates other SR proteins, several SR and SR-related proteins were tested as substrates for the two kinases. Using quantities of Clk/Sty and SRPK1 normalized to give equal level of ASF/SF2 phosphorylation, we found that both kinases phosphorylated the RS domain-containing splicing factors, SC35 (Fu and Maniatis, 1992a), SRp55 (Roth et al., 1991), and U2AF65 (Zamore et al., 1992) to similar levels (figure 3.1). Clk/Sty also phosphorylated histone H1 and myelin basic protein (figure 3.1 and (chapter 2)).
Figure 3.1. Phosphorylation of arginine-serine-rich proteins by Clk/Sty and SRPK1.

GST-Clk/Sty (100 nM) or GST-SRPK1 (an equivalent activity, normalized to ASF/SF2) were incubated with 1 μM substrate as indicated. The samples were resolved on an SDS 12.5% polyacrylamide gel followed by autoradiography. GST-SRPK1 is present in lanes 1, 3, 5, 7, 9, and 11 while GST-Clk/Sty is present in lanes 2, 4, 6, 8, 10, and 12. Molecular weight markers are indicated on the right.
In contrast, SRPK1 had little activity on these substrates (figure 3.1 and (Gui et al., 1994b)). This result suggests that Clk/Sty has a broader substrate specificity than SRPK1.

Another difference between these two kinases was noted while normalizing their activities to give equal phosphorylation of ASF/SF2. SRPK1 appeared much more active for ASF/SF2 than Clk/Sty. Because SRPK1 and Clk/Sty were isolated from different sources, a direct comparison between the two kinases was sought. SRPK1 and Clk/Sty were in vitro translated as FLAG-tagged proteins (figure 3.2A). The translated proteins were immunoprecipitated using an anti-FLAG monoclonal antibody, and their kinase activities on ASF/SF2 were determined (figure 3.2B). Quantitative analysis showed that under these conditions, SRPK1 was 150-fold more active than Clk/Sty. Since SRPK1 and Clk/Sty had similar Km values for ASF/SF2 (table 3.1), I postulated that the greater activity of SRPK1 was due to an increase in Kcat. This result indicates that SRPK1 turns over ASF/SF2 faster than Clk/Sty. Perhaps the interaction of the RS domain of Clk/Sty with ASF/SF2 results in the slower release of ASF/SF2 from Clk/Sty (chapter 2).

**ASF/SF2 phosphorylated by Clk/Sty is recognized by mAb104** - Although SRPK1 and Clk/Sty phosphorylate the same SR proteins, these polypeptides have many potential phosphorylation sites. The two kinases might, therefore, have identical, overlapping, or distinct substrate specificities. The monoclonal antibody mAb104 recognizes a phosphorylated epitope found on SR proteins (Roth et al., 1991). Bacterially expressed ASF/SF2 is recognized by mAb104 after phosphorylation in vitro by SRPK1 (Gui et al., 1994b). To determine if bacterially expressed ASF/SF2 reacts with mAb104 after phosphorylation by Clk/Sty, I subjected purified ASF/SF2 to an in vitro kinase assay in the presence or absence of Clk/Sty. Clk/Sty generated
Figure 3.2. Relative specific activities of Clk/Sty and SRPK1. A) Wild-type (wt) and mutant (M) Clk/Sty and SRPK1 were in vitro translated as FLAG-tagged proteins. Two µl of this reaction was loaded onto an SDS 12.5% polyacrylamide gel. The mutant Clk/Sty contains a K-to-R change at position 190 (Ben-David et al., 1991; Duncan et al., 1995), and the mutant SRPK1 contains a K-to-M change at position 109. Both mutations target the ATP binding site within each kinase. B) In vitro translated products were immunoprecipitated by an anti-FLAG antibody, and subjected to a kinase assay using ASF/SF2 as substrate. The immunoprecipitated kinases ([35S]-methionine labelled) and in vitro phosphorylated ASF/SF2 ([32P]-labelled) were quantitated by phosphoimaging. Clk/Sty was immunoprecipitated 5 times more efficiently than SRPK1 in this experiment. Because the signal of wild-type Clk/Sty is stronger than the mutant, possibly due to autophosphorylation, the relative protein level of immunoprecipitated Clk/Sty was based on the signal detected with the mutant kinase. The relative specific activities were determined by dividing the relative kinase activities by relative protein levels.
A. In vitro translation

<table>
<thead>
<tr>
<th></th>
<th>Cik/Sty</th>
<th>SRPK1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>wt</td>
</tr>
</tbody>
</table>

B. Immunoprecipitation/kinase assay

<table>
<thead>
<tr>
<th></th>
<th>Cik/Sty</th>
<th>SRPK1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>wt</td>
</tr>
</tbody>
</table>

MW (kD)
- 101
- 99
- 50.6
- 35.5
- 29.1
- 20.9

Rel. Protein Level: 1 0.2
Rel. Kinase Act.: 1 30
Rel. Sp. Act.: 1 150
<table>
<thead>
<tr>
<th></th>
<th>Km (μM)</th>
<th>Vmax (pmol PO₄)</th>
<th>Vmax/Km</th>
<th>RSRS:SPRY</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASF/SF2 - Clk/Sty</td>
<td>0.40</td>
<td>0.20</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>ASF/SF2 - SRPK1</td>
<td>0.28</td>
<td>0.20</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>RSRS - Clk/Sty #1</td>
<td>146.23</td>
<td>342.14</td>
<td>2.34</td>
<td></td>
</tr>
<tr>
<td>RSRS - Clk/Sty #2</td>
<td>95.68</td>
<td>45.51</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>RSRS - Clk/Sty #3</td>
<td>64.09</td>
<td>17.90</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>RSRS - Clk/Sty Final</td>
<td>102.00 +/- 41.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPRY - Clk/Sty #1</td>
<td>124.87</td>
<td>62.80</td>
<td>0.51</td>
<td>4.60</td>
</tr>
<tr>
<td>SPRY - Clk/Sty #2</td>
<td>236.90</td>
<td>22.46</td>
<td>0.01</td>
<td>5.02</td>
</tr>
<tr>
<td>SPRY - Clk/Sty #3</td>
<td>61.83</td>
<td>3.35</td>
<td>0.54</td>
<td>5.15</td>
</tr>
<tr>
<td>SPRY - Clk/Sty Final</td>
<td>141.20 +/- 88.67</td>
<td></td>
<td></td>
<td>5.06 +/- 0.48</td>
</tr>
<tr>
<td>RSRS - SRPK1 #1</td>
<td>25.00</td>
<td>45.40</td>
<td>1.82</td>
<td></td>
</tr>
<tr>
<td>RSRS - SRPK1 #2</td>
<td>61.03</td>
<td>4.10</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>RSRS - SRPK1 #3</td>
<td>78.30</td>
<td>3.86</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>RSRS - SRPK1 Final</td>
<td>54.78 +/- 27.19</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1. Kinetic parameters for ASF/SF2, RSRS, and SPRY after phosphorylation by Clk/Sty or SRPK1. RSRS:SPRY represents the ratio of Vmax/Km for RSRS over SPRY. Vmax is expressed in pmol of phosphate transferred per minute.
the phosphoepitope recognized by mAb104 on full-length ASF/SF2 (figure 3.3A). Besides full-length ASF/SF2, I also tested two derivatives of ASF/SF2: ASFΔRS, that has the arginine-serine-rich (RS) domain deleted, and GST-ASF RS, that contains only the RS domain. Clk/Sty phosphorylates the RS domain of ASF/SF2 to similar levels as full-length ASF/SF2, but only weakly phosphorylates ASFΔRS (chapter 2). GST-ASF RS was recognized by mAb104 after phosphorylation by Clk/Sty indicating that the ASF/SF2 sites phosphorylated by Clk/Sty and subsequently recognized by mAb104 lie in the C-terminal RS domain (figure 3.3A). ASFΔRS, on the other hand, was not recognized by mAb104 after incubation with Clk/Sty. This result indicates that this region does not contain the correct phosphoepitope or that Clk/Sty does not phosphorylate ASFΔRS to an extent where it is recognized by mAb104 (figure 3.3A). In vivo, SR proteins show altered migration on SDS gels due to multiple sites of phosphorylation (Zahler et al., 1993b). ASF/SF2 also displays reduced mobility after in vitro phosphorylation by SRPK1 (Gui et al., 1994b). Purified bacterial ASF/SF2 and GST-ASF RS, but not ASFΔRS, exhibited a similar mobility shift after in vitro phosphorylation by Clk/Sty (figure 3.3B). Therefore, both kinases can phosphorylate ASF/SF2 on multiple sites, including the mAb104 phosphoepitope.

**SRPK1 and Clk/Sty phosphorylate overlapping sites on ASF/SF2** - Tryptic phosphopeptide mapping has shown previously that sites on ASF/SF2 phosphorylated in vitro by Clk/Sty overlap sites phosphorylated on ASF/SF2 in vivo, indicating that Clk/Sty may regulate ASF/SF2 in vivo (chapter 2). To test whether SRPK1 phosphorylates ASF/SF2 at similar sites as Clk/Sty, I compared tryptic phosphopeptide maps of purified ASF/SF2 phosphorylated in vitro by SRPK1 or Clk/Sty. ASF/SF2 phosphorylated in vitro by either SRPK1 or Clk/Sty generated the phosphopeptides 1, 2, C, 5, 6, E, 8, and F (figure 3.4A, B, D). The numbers refer to peptides
Figure 3.3. Phosphorylation of ASF/SF2 by Clk/Sty generated the mAb104 phosphoepitope. A) 500 nM ASF/SF2, ASFARS, or GST-ASF RS were incubated in an in vitro kinase assay plus or minus 100 nM GST-Clk/Sty. Lanes 1, 3, and 5 do not contain GST-Clk/Sty. The samples were separated on an SDS 10% polyacrylamide gel, transferred to nitrocellulose, and probed with mAb104. B) The same samples as in A) were probed with α-ASF/SF2 (lanes 1 to 4) or α-GST-ASF RS (lanes 5 and 6).
Figure 3.4. Tryptic phosphopeptide maps of ASF/SF2. Samples were eluted from SDS-polyacrylamide, precipitated, and digested with trypsin. Individual peptides were resolved on thin layer cellulose plates as described (chapter 2). A) ASF/SF2 phosphorylated in vitro by GST-SRPK1. B) ASF/SF2 phosphorylated in vitro by GST-C1k/Sty. C) A mix of ASF/SF2 phosphorylated in vitro by GST-SRPK1 or GST-C1k/Sty. D) A schematic of the phosphopeptide map. Numbers refer to peptides phosphorylated in vivo (chapter 2). Letters refer to peptides present only in ASF/SF2 phosphorylated in vitro by GST-C1k/Sty or GST-SRPK1. Closed circles represent phosphopeptides seen in ASF/SF2 phosphorylated in vivo, and in vitro by C1k/Sty. The origin is denoted by a striped circle.
present in ASF/SF2 phosphorylated *in vivo* in COS-1 cells, and the letters refer to peptides thus
far seen only in ASF/SF2 phosphorylated *in vitro* by Clk/Sty (chapter 2) or SRPK1. The tryptic
phosphopeptides maps of *in vitro* phosphorylated ASF/SF2 also revealed differences between
SRPK1 and Clk/Sty mediated phosphorylation. Peptides 7 and 9 were not seen following SRPK1
phosphorylation of ASF/SF2, and peptide G was not present in ASF/SF2 following
phosphorylation by Clk/Sty (figure 3.4A, B). Mixing of tryptic digests of ASF/SF2 phosphorylated
*in vitro* by SRPK1 or Clk/Sty confirmed these results (figure 3.4C).

The extent of phosphorylation of two specific peptides in ASF/SF2 by SRPK1 or Clk/Sty
also differed. There was a decrease in peptide 5 phosphorylation and an increase in peptide 6
phosphorylation in ASF/SF2 phosphorylated by SRPK1 as compared to Clk/Sty (figure 3.4A, B).
It is possible that some spots (i.e., 6, E, and 7) may represent differentially phosphorylated states
of the same peptide (Boyle *et al.*, 1991). If so, the lack of peptide 7 in ASF/SF2 phosphorylated
by SRPK1 may represent a different stoichiometry of phosphorylation of a peptide rather than
lack of that phosphopeptide. Neither kinase phosphorylated peptides 3, 4, and 10 that were
present in ASF/SF2 phosphorylated *in vivo* suggesting that other kinases exist that also
phosphorylate ASF/SF2 (chapter 2).

**SRPK1 and Clk/Sty have distinct substrate specificities for peptides derived from ASF/SF2
sequence** - Both SRPK1 and Clk/Sty phosphorylate ASF/SF2 within its RS domain (chapter 2;
Gui *et al.*, 1994a). The RS domain of ASF/SF2 is 50 amino acids long and contains 20 serines
(Ge *et al.*, 1991; Krainer *et al.*, 1991). Of these 20 serines, 13 have arginine at the P+1 position
(amino acid following the serine), 4 have proline, 2 have tyrosine, and 1 has asparagine. To
determine if SRPK1 or Clk/Sty prefer arginine or proline at the P+1 position, two peptides
derived from the ASF/SF2 RS domain were synthesized. The first peptide (termed RSRS) has the sequence GRSRSRSRSR, and the second peptide (termed SPRY) has the sequence RGSPRYSPRHS. Clk/Sty phosphorylated both peptides (table 3.1). SRPK1 phosphorylated RSRS but not SPRY, even though concentrations of up to 1 mM SPRY were tested (table 3.1 and data not shown). To determine which of the two peptides was a better substrate for Clk/Sty, I undertook a kinetic analysis of SPRY and RSRS phosphorylation by Clk/Sty. Clk/Sty had a similar affinity for either peptide as measured by Km values (compare the Km of RSRS of 102 μM +/- 41 with the Km of SPRY of 147 μM +/- 81). The Vmax for RSRS was consistently higher than that for SPRY. When comparative catalytic efficiency (Vmax/Km) was compared between the two peptides, RSRS was shown to be a better substrate for Clk/Sty by a ratio of 5 +/- 0.5 (table 3.1). Based on the Km and Vmax values, it appeared that Clk/Sty bound RSRS and SPRY equally well but turned over the RSRS peptide at a faster rate. This observation correlated with competition assays in which SPRY inhibited Clk/Sty phosphorylation of full-length ASF/SF2 whereas RSRS was a weak inhibitor at best (data not shown).

This peptide analysis revealed one major difference between the two kinases. Clk/Sty phosphorylated serines with proline or arginine at the P+1 position although arginine was preferred. In contrast, SRPK1 was able to only phosphorylate serines followed by arginine.

The Km for RSRS phosphorylated by SRPK1 (55 μM +/- 27) was similar to that seen for Clk/Sty. (table 3.1).

**Tryptic mapping of SPRY and RSRS phosphopeptides** - To determine if RSRS or SPRY sequences are phosphorylated in full-length ASF/SF2, I subjected the two peptides, after phosphorylation by Clk/Sty, to trypsin digestion and subsequent peptide mapping. The RSRS
peptide produced two phosphopeptides (figure 3.5A) that appeared similar to peptides 8 and F of full-length ASF/SF2 phosphorylated by Clk/Sty (figure 3.5B). Further comparative analysis of RSRS and ASF/SF2 demonstrated that the phosphorylated RSRS tryptic fragments did co-migrate with peptides 8 and F from full-length ASF/SF2 (figure 3.5C, 3.5D). Both peptides were present in tryptic digests of ASF/SF2 phosphorylated by SRPK1 (figure 3.4A) consistent with the ability of SRPK1 to phosphorylate the RSRS peptide.

The SPRY peptide, phosphorylated by Clk/Sty, also produced two smaller phosphopeptides upon tryptic digestion (figure 3.5E), one of which co-migrated with peptide 5 of full-length ASF/SF2 (figure 3.5D, 3.5F). This result suggests that phosphopeptide 5 of full-length ASF/SF2 is derived from the SPRY sequence. SRPK1, which could not phosphorylate the SPRY peptide, was able to phosphorylate peptide 5 in the context of full-length ASF/SF2 albeit less efficiently than Clk/Sty (figure 3.4A). Therefore, it appears that in full-length ASF/SF2, SRPK1 may phosphorylate serines followed by proline (SP), although SP is not a good site in the context of a peptide. SRPK1 has higher affinity for full-length ASF/SF2 than for the RSRS peptide (table 3.1), which is consistent with SRPK1's ability to recognize more sites within full-length ASF/SF2.

Both SPRY and RSRS peptides produced tryptic fragments that overlapped peptide F of full-length ASF/SF2 (figure 3.5D). In a mix of RSRS and SPRY tryptic fragments, only three peptides were resolved, which is consistent with the SPRY and RSRS peptides overlapping at fragment F (figure 3.5G). Therefore, peptide F is likely a mix of two different phosphopeptides that share similar electrophoretic properties.
Figure 3.5. Tryptic phosphopeptide maps of ASF/SF2, RSRS and SPRY. RSRS and SPRY peptides, after phosphorylation by GST-Clk/Sty, were purified on thin layer cellulose plates, isolated as described (van der Geer et al., 1993), and subjected to trypsin digestion. ASF/SF2, after phosphorylation by GST-Clk/Sty, was isolated from SDS-polyacrylamide, precipitated, and digested by trypsin. Samples were run for 1 hour in the electrophoretic dimension and overnight in the chromatographic dimension unless otherwise noted. A) A tryptic map of RSRS peptide reveals two tryptic phosphopeptides. B) A tryptic map of ASF/SF2. This sample was run for 1.5 hours in the electrophoretic dimension. C) A mix of ASF/SF2 and RSRS peptides reveals the co-migration of RSRS and ASF/SF2 tryptic phosphopeptides. D) ASF/SF2, RSRS, and SPRY phosphopeptides were spotted 1 cm apart on the same plate, and run for 1.5 hours in the electrophoretic dimension only. Note that RSRS tryptic phosphopeptides co-migrated with peptides 8 and F of full-length ASF/SF2 and SPRY tryptic phosphopeptides co-migrated with peptides 5 and F. E) A tryptic map of SPRY peptide shows two tryptic phosphopeptides. F) A mix of SPRY and ASF/SF2 phosphopeptides demonstrates the co-migration of SPRY and ASF/SF2 phosphopeptides. G) A mix of RSRS and SPRY tryptic phosphopeptides indicates that a phosphopeptide from RSRS and another from SPRY co-migrate at the position of peptide F. This sample was run for 1.5 hours in the electrophoretic dimension.
SRPK1 and Clk/Sty differ in their ability to phosphorylate mutants of ASF/SF2 - To further investigate the catalytic properties of SRPK1 and Clk/Sty, the abilities of SRPK1 and Clk/Sty to phosphorylate ASF/SF2 mutants that have substitutions in their RS repeats were tested (Cáceres and Krainer, 1993). SRPK1 phosphorylation of ASF/SF2 markedly decreased when all the serines in RS dipeptides were mutated to glycines (RG) or threonines (RT) consistent with the above observations that SRPK1 phosphorylated ASF/SF2 on serines in RS dipeptides (figure 3.6 and (Gui et al., 1994b)). The residual phosphorylation must have occurred on one or more of the four serines not mutated in the RG or RT variants, two of which are found in the SPRY peptide. Mutation of the arginines in RS dipeptides to lysines (KS) or glycines (GS), resulted in loss of phosphorylation of ASF/SF2 by SRPK1. This observation demonstrates that phosphorylation of any of the 20 serines in the RS domain of ASF/SF2 by SRPK1 requires the presence of arginine in the RS dipeptides. In the ΔSR mutant, the RS dipeptides were deleted but the other amino acids in the RS domain were retained. Similar to the KS and GS mutants, SRPK1 was unable to phosphorylate the ΔSR variant.

Phosphorylation of the various RS domain mutants by Clk/Sty revealed a different pattern than that observed for SRPK1. Phosphorylation by Clk/Sty also decreased when the RS dipeptides were mutated to RG, but not to the same extent as SRPK1, consistent with Clk/Sty phosphorylating the SPRY peptide. When the arginines in RS dipeptides were mutated to lysines in the KS mutant, ASF/SF2 phosphorylation by Clk/Sty showed only a slight decrease in level of phosphorylation. This result indicates that Clk/Sty can accommodate lysine in place of arginine in its substrate binding site, and is consistent with its broader substrate specificity. Unlike SRPK1, Clk/Sty phosphorylated both the GS and ΔSR mutant proteins at a low level.
Figure 3.6. Comparison of SRPK1 and CLK/Sty phosphorylation of ASF/SF2 mutants. 1 μM indicated wild-type or mutant ASF/SF2 were incubated with 100 nM GST-CLK/Sty or GST-SRPK1 with equivalent activity. The samples were separated on an SDS 12.5% polyacrylamide gel followed by autoradiography. A) Amino acid sequence of the RS domain of ASF/SF2 is indicated as WT. The mutants are indicated at the left with their corresponding mutations shown below the WT sequence (Cáceres et al., 1993). For ΔSR, only the amino acids that are retained are shown. Amino acid numbering is shown above the sequence. B) Lanes 1 to 6 - the mutants, as noted, were incubated with GST-SRPK1. Lanes 7 to 12 - the mutants, as indicated, were incubated with GST-CLK/Sty.
In the ΔSR construct, only the serines within the RNA-binding domain, and the four serines outside of the RS dipeptide repeats remain. From phosphopeptide mapping, the sites of phosphorylation of ASF/SF2 by Clk/Sty are limited to the RS domain (chapter 2). Of the four serines in the RS domain, two of the arginines flanking serines 201, 227, and 234 were deleted. In contrast, the arginines flanking serine 238 were retained. Therefore, it is possible that the minor level of phosphorylation of the GS and ΔSR mutants by Clk/Sty occurs on serine 238 within the SPRY peptide sequence. Note that the RT mutant was not phosphorylated to the same level as wild-type by either SRPK1 or Clk/Sty, indicating that threonine cannot substitute for serine as the phosphoacceptor.

Discussion

Phosphorylation of SR proteins by SRPK1 and Clk/Sty - Both SRPK1 and Clk/Sty phosphorylated several SR proteins in vitro, and neither kinase preferred one SR protein over another. SRPK1 weakly phosphorylated myelin basic protein (Gui et al., 1994b), and was unable to phosphorylate ASFΔRS, or histone H1 (figure 3.1 and (Gui et al., 1994a)). Although Clk/Sty preferred SR proteins as substrates, it also phosphorylated basic proteins such as histone H1 and myelin basic protein and weakly phosphorylated ASFΔRS (chapter 2 and figure 3.1). Therefore, in vivo, Clk/Sty may phosphorylate proteins outside of the SR family, and play a more general role in cellular regulation than SRPK1.

Consistent with SRPK1 being a more specific SR protein kinase, SRPK1 displayed greater specific activity towards ASF/SF2 as compared to Clk/Sty. This suggests that, in vivo, SRPK1 may be the more active SR protein kinase. However, Clk/Sty may be expressed at higher levels in cells or may require modification, such as phosphorylation, to become fully active.
SRPK1 and Clk/Sty phosphorylate ASF/SF2 on sites phosphorylated *in vivo* - I used two methods to determine if SRPK1 or Clk/Sty phosphorylate SR proteins on sites phosphorylated *in vivo*. One approach involved the use of mAb104 that recognizes a phosphoepitope found on SR proteins *in vivo* (Roth *et al.*, 1991). Bacterially expressed ASF/SF2 was specifically recognized by mAb104 after phosphorylation by Clk/Sty, indicating that Clk/Sty phosphorylates this site(s) *in vitro*. Furthermore, mAb104 recognized the GST-ASF RS fusion protein following incubation with Clk/Sty, suggesting that the phosphorylated site recognized by mAb104 lies within the RS domain. Previous work has shown that SRPK1 phosphorylates ASF/SF2 *in vitro* at a site recognized by mAb104 (Gui *et al.*, 1994b). The finding that both SRPK1 and Clk/Sty generate a mAb104 phosphoepitope on ASF/SF2 suggests that they can phosphorylate similar sites.

The second method of phosphopeptide mapping confirmed the above observation. A tryptic map of ASF/SF2 phosphorylated *in vitro* by Clk/Sty generated twelve phosphopeptides of which seven were also found in ASF/SF2 isolated from [*32P*]-labelled cells (chapter 2). A phosphopeptide map of ASF/SF2 phosphorylated *in vitro* by SRPK1 was similar to the tryptic map of ASF/SF2 phosphorylated *in vitro* by Clk/Sty except that it lacked two peptides (seen *in vivo*) and contained an additional one (not seen *in vivo*). The position of peptide 7, which ASF/SF2 phosphorylated by SRPK1 lacks, is consistent with it being a lower phosphorylation state of peptide 6 that SRPK1 phosphorylates. Three conclusions can be drawn from these data. First, SRPK1 phosphorylated ASF/SF2 on a set of peptides that overlapped with those phosphorylated *in vivo*, consistent with its being an SR protein kinase *in vivo*. Second, SRPK1 did not phosphorylate peptide 9 that was seen *in vivo*, indicating that Clk/Sty or another kinase...
is likely to be responsible for phosphorylation of this site in vivo. Finally, peptides 3, 4, and 10 were present in ASF/SF2 phosphorylated in vivo, but not in ASF/SF2 phosphorylated in vitro by either kinase (chapter 2). Therefore, another SR protein kinase is probably responsible for these latter phosphorylation events. Possibilities for this kinase include the U1 70K kinase (Woppmann et al., 1993), cAPK or PKC (chapter 2), LBR kinase (Nikolakaki et al., 1996), DNA topoisomerase I (Rossi et al., 1996), or another unidentified SR protein kinase.

Clk/Sty and SRPK1 display distinct substrate specificities - Two types of serines predominate in the RS domain of ASF/SF2: serines followed by arginine at the P+1 site or serines followed by proline. In the CDK2 and ERK2 crystal structures, an arginine forms the base of the P+1 binding pocket (Zhang et al., 1994; De Bondt et al., 1993). The bulkiness of the arginine side chain fills the P+1 binding pocket leaving room for only small amino acids, such as proline, at the P+1 site (Zhang et al., 1994; De Bondt et al., 1993; Songyang et al., 1994). This arginine is conserved in SRPK1 and Clk/Sty (chapter 2). To test if SRPK1 or Clk/Sty phosphorylate substrates with arginine or proline at the P+1 site, two peptides were synthesized: RSRS (GRSRSRSRSR) and SPRY (RGSPRYSPRHS). Both kinases preferred the RSRS peptide suggesting that their P+1 binding pockets can accommodate arginine in contrast to CDK2 or ERK2. The preference for arginine at P+1 was more dramatic for SRPK1 as it was unable to phosphorylate the SPRY peptide. Although Clk/Sty preferred the RSRS peptide, it was competent in phosphorylation of the SPRY peptide, suggesting that Clk/Sty has fewer constraints than SRPK1 on the type of amino acid it can accommodate at the P+1 position.

To determine if residues contained within the RSRS and SPRY peptides are phosphorylated in full-length ASF/SF2, I subjected these peptides to trypsin digestion and peptide
mapping. Both peptides generated two phosphorylated fragments that co-migrated with tryptic peptides contained within the full-length ASF/SF2. Based on these data, ASF/SF2 peptide 8 is apparently derived from an RSRS sequence, and peptide 5 from a SPRY sequence. Both phosphorylated peptides generated a second fragment that in each case migrated to a similar position as peptide F in full-length ASF/SF2. Since these two fragments must have distinct sequences, despite their similar migration properties, I cannot tell whether peptide F phosphorylated in vitro by SRPK1 or Clk/Sty contains one or both fragments.

The data discussed above indicate that both the RSRS and SPRY peptides contain sites phosphorylated in vivo, and in vitro by Clk/Sty. RSRS peptide 8 was present in ASF/SF2 phosphorylated by SRPK1 consistent with SRPK1 phosphorylating the RSRS peptide. Surprisingly, phosphopeptide 5 was present in ASF/SF2 phosphorylated by SRPK1 even though peptide 5 contained sequence from the SPRY peptide. However, this site was less prominent in full-length ASF/SF2 phosphorylated by SRPK1 than by Clk/Sty, indicating that it is a less preferred site for SRPK1. Since the SPRY peptide is not phosphorylated at all by SRPK1, I suggest that sites in full-length ASF/SF2, perhaps the RSRS stretch, are required for SRPK1 to phosphorylate the SP motifs of peptide 5. An alternate explanation is that peptide 5 contains distinct co-migrating phosphopeptides that do not correspond to the SPRY peptide and SRPK1 does not phosphorylate SPRY sequence at all.

**Clk/Sty has a broader substrate specificity than SRPK1** - When the RS dipeptides of ASF/SF2 were replaced by RG, phosphorylation of ASF/SF2 by SRPK1 and Clk/Sty was greatly reduced but not eliminated, indicating that either kinase can phosphorylate the four serines not found in an RS repeat. The higher level of phosphorylation of the RG ASF/SF2 mutant by Clk/Sty is
consistent with the ability of Clk/Sty to phosphorylate the SPRY peptide. Deletion of all the RS dipeptides or mutation of the arginine residues to glycines, resulted in loss of phosphorylation by SRPK1, and a decrease in phosphorylation by Clk/Sty, indicating that the arginines in these repeats are important for recognition of serines inside and outside of the repeats. The amino acids flanking serine 238 were barely affected by the GS and ΔSR mutations and the low level of phosphorylation of these mutants by Clk/Sty probably occurred on serine 238. Threonine cannot replace serine as a substrate for phosphorylation by either kinase. Of interest, Clk/Sty, but not SRPK1, phosphorylated the KS mutant to almost wild-type levels. This result indicates that Clk/Sty requires only a basic charge for substrate recognition whereas SRPK1 specifically requires arginine.

The function of SR proteins is thought to be controlled by phosphorylation. Indirect evidence suggests that the location of SR proteins and their function in spliceosome assembly is altered by phosphorylation (Mermoud et al., 1994; Roscigno and Garcia-Blanco, 1995; Gui et al., 1994a; chapter 2). In vitro, Clk/Sty phosphorylated more sites in ASF/SF2 than SRPK1 did. The last 24 amino acids that include all of the SPRY sequence, but not the stretch of 8 RS repeats, are dispensable for ASF/SF2 function in vitro (Zuo and Manley, 1993). In contrast, deletion of the 16 amino acid RS stretch, or substitution of all the arginines or serines in this stretch to glycines, blocks the constitutive splicing function of ASF/SF2 (Zuo and Manley, 1993; Cáceres and Krainer, 1993). Since the deletion of the SPRY sequence does not affect splicing in vitro, the significance of Clk/Sty phosphorylation of this sequence is unknown.

Another difference between the two kinases is the presence of an RS domain at the N-terminus of Clk/Sty. This RS domain may provide Clk/Sty with a function in protein-protein
interaction that SRPK1 lacks (chapter 2; Amrein et al., 1994; Kohtz et al., 1994; Wu and Maniatis, 1993). To this end, it should be noted that there is an isoform of Clk/Sty that contains only the RS domain (Duncan et al., 1995).

In this study, I have compared the activities of SRPK1 and Clk/Sty in vitro. In vivo, SRPK1 and Clk/Sty may be found in different cell types or in distinct locations in the cell. SRPK1 is most active during mitosis when splicing is inactivated, and SRPK1 inhibits splicing in vitro at high concentrations (Gui et al., 1994a). It is not known whether Clk/Sty activity is regulated during the cell cycle. In vivo, Clk/Sty alters 5' splice site selection when overexpressed (J. Prasad, K. Colwill, T. Pawson, and J.L. Manley, unpublished results). Whether the main role of SRPK1 is to turn SR proteins off during mitosis and the function of Clk/Sty is to control the activity of the SR family during alternative splice site selection remains to be determined. Further analysis is necessary to understand the functions of these kinases with respect to the regulatory role they play in pre-mRNA splicing and other processes.
CHAPTER 4: Analysis of the Two-hybrid Clones that Interact with Clk/Sty

Part I: Cloning of a Mammalian tra-2 Homologue

Part II: RS Cyclophilins: Identification of a NK-TR1 Related Cyclophilin

Associated with the Clk/Sty Kinase

Part III: Analysis of the Two-hybrid Clones

I did the work reported in this chapter with the following exceptions:

Elke Aippersbach assisted me in the sequencing of clone 6.9.

Frederick Nestel and Sarah Harper in Steve Anderson’s lab cloned and sequenced the full-length human CARS-Cyp cDNA and performed the Northern Analysis.

Parts I and III represent unpublished material

Part II is a modified version of the following publication:


© 1996 by Gene/Elsevier Science
I. Cloning of a Mammalian tra-2 Homologue.

Introduction

In Drosophila melanogaster, the Tra-2 protein is essential for female somatic cell differentiation, and male germ cell development (see chapter 1). In female somatic cell differentiation, Tra-2 and another splicing factor, Tra, are required for the female-specific splicing of the dsx pre-mRNA transcript (Tian and Maniatis, 1992; Hedley et al., 1991; Hoshijima et al., 1991, Ryner and Baker, 1991). Tra-2 contains an RNA recognition motif (RRM) that binds the dsxRE in exon 4 of the dsx transcript (Amrein et al., 1998; Goralski et al., 1989; Amrein et al., 1994. In addition, Tra-2 contains two RS domains that are involved in protein-protein interactions.

Here, I describe the cloning of a tra-2 like cDNA, clone 6.9, from murine T cells. This clone was isolated because its protein product interacted with the Clk/Sty protein kinase in a two-hybrid screen. In clone 6.9, the domain structure of Tra-2 is conserved. The RRM$s of clone 6.9 and Tra-2 are 58% identical at the amino acid level. I propose that clone 6.9 is a mammalian counterpart of tra-2, and suggest that clone 6.9 may recruit SR proteins to exonic enhancers in mammalian cells.

In vitro, Clk/Sty phosphorylates several members of the SR family, and apparently regulates their function in vivo (chapter 2; chapter 3). The isolation of clone 6.9 by Clk/Sty suggests that Clk/Sty may also regulate clone 6.9 function in vivo.
Materials and Methods

Two-hybrid screen - I described the screen used to isolate clone 6.9 in chapter 2.

Sequencing of clone 6.9 - Clone 6.9 was sequenced using the Automated Laser Fluorescence DNA sequencer with fluorescent primers. A blast search of the NCBI express sequence tag (EST) database was used to reveal other cDNAs for clone 6.9.

Results

Sequencing of clone 6.9 - I used a two-hybrid screen to detect proteins that bind to the Clk/Sty protein kinase (chapter 2). This screen isolated several RNA-binding proteins including three members of the SR family of splicing factors. Clone 6.9 was isolated in this screen.

Clone 6.9 is a novel cDNA that has similarity to the Drosophila tra-2 gene (Amrein et al., 1988; Goralski et al., 1989). Tra-2 codes for a 264 amino acid protein that has an RRM flanked by RS domains on either side. Clone 6.9 is about 1500 nucleotides in length, and codes for an open reading frame of 141 amino acids (figure 4.1). This open reading frame corresponds to amino acids 127 to 259 of Tra-2 that encompasses all of the C-terminal RS domain, most of the RRM but none of the first RS domain (figure 4.2). Clone 6.9 is 41% identical (53% similar) to Tra-2 in the C-terminal RS domain, and 62% identical (76% similar) to Tra-2 in the RRM.

I performed a blast search of the NCBI expressed sequence tag (EST) database to search for cDNAs that correspond to clone 6.9. Six cDNAs for clone 6.9 were found (figure 4.2). Five of these cDNAs encode the C-terminal RS domain of clone 6.9. Of these five, four are from human cDNA libraries, and one is from a Rat PC12 cell cDNA library. At the amino acid level, these clones are 100% identical to clone 6.9, indicating that this domain is well-conserved between mammalian species. One human cDNA, H11792, contains 5' sequence that clone 6.9
Figure 4.1. Sequence of H11792 and clone 6.9. a) H11792 is 2102 basepairs in length. The sequence of H11792 that is novel from clone 6.9 plus a 21 base pair overlap is shown. The nucleotide number is marked at the right. The corresponding amino acids are shown beneath the sequence. b) Clone 6.9 is 1500 nucleotides in length. The nucleotide sequence that corresponds to the coding sequence is shown. The amino acid sequence of clone 6.9 is shown beneath the nucleotide sequence.
A) CGTTCTGCTTCCAGAAGTGGAAGTGCTCACGGATCGGCGCCCT 61
RSASRSGSASHGSGKSARHRTP
GCAAGGTCTCGTCCAAAGAGAAGTTCCAGGCGTTCTCAGATCAAAGTGTCGGTGCCATCT 121
ARSSRSKEDSRSSRSSRSSRSRSSRRRHYTRSR
GAATCTAGGTCATAGATCCAGAAGACTCTCCGAAAGCATTATAACCCGCTACATTTCTCGC 181
ESRSRSSRSSRRRHYTRSR
TCCCGCTCCCATAGACGATCATCAGTGGTCTATACGAGATTATCTCTAGAGACGAC 241
SRRSSRSRSRSRSRSRSRSRRRRHYTRSR
AGCCACAGGCATTCTCCCATGTCTACTGCAAGGCATGTGGGAAATGGGGCCTATCTCCT 301
SHSSHSPMSTRRRHHGNARNP
GATCTAAGTTGTTCTCTGAGATATTGCTGGACTTTGTTACCCACAGAAGAGATCTA 361
DPNCCLGLGFGLSLYTTERDL
AGAGAAGTGTTCTCTAAATATGGCTCCATTTGGCAGTGTCTATTTGCTATATTTGAA 421
REVFSKYGPIADVSIIVYDQQ

B) TCTATTGTATATGACCAGCAGAATCTAGACGGTTGCTTTTGTTATATTTGAA 60
SIIVYDQQQSRSSRSFAFWYFEE
AATGTAGACGATGCAAAGAAGCTAAAAGAACGTGCGGCAATGGGAGCTGAGTTGGAGCT 120
NVDDAIAKEAKERANGMELDGR
CGAAATTAGAGTCATTGTTCTCTCTATAACACAAAAGGGCCCATACCCCCACACAGGAAATTAT 180
RIRVDFSISTKRPHTPTPGIY
ATGGGGGAAGGACCATCTTTAGTGGCTCCGCGCCAGACTATTATGACAGAGGTTACGAT 240
MGRTYPGSSRRRDRYYDYDRGYD
CGGGGTATGATGAGCAGGGGACTATTACGAGCAGCTCACAGAGAGGCTGTTGGAGAA 300
RGYDGRDYYSSRSYRGGGGGGG
GGTGAGATGGAGACGCTGAGTCAGAGAGCTGATTTACAGAAGAGCCTGATCTTCTCTCT 360
GGWRRAAQDRDQIYRRRSSPSP
TACTACGTGCTGGAGGACATCGCTCGATCACTCAGCTACTGACCTACTACCCCTCTGTCGC 420
YYSSRGGYKRSRSRSSRSYSRR
TACTAA 426
Y *
Figure 4.2. A schematic of Tra-2 and mammalian cDNAs of clone 6.9. Tra-2 contains an RRM divided by two RS domains. Clone 6.9 contains the C-terminal RS domain and part of the RRM. H11792 contains most of the first RS domain and the rest of the RRM that is missing from clone 6.9. The portion of H11792 that is unique from clone 6.9 is diagrammed. The amino acid numbering of each protein is shown at the bottom. For clone 6.9 and H11792, the amino acid numbers above the protein indicate the analogous position in Tra-2. Five cDNAs from the expressed sequence tag database that correspond to the second RS domain are shown. They are identified by their GenBank accession numbers and the libraries from which they were isolated. The percentage identity between these various clones is indicated.
lacks. H11792 cDNA is 2102 basepairs in length, of which 514 nucleotides are sequenced. Of the sequenced nucleotides, 54 overlap clone 6.9, and the other 460 represent novel upstream sequence (figure 4.1). This novel upstream sequence corresponds to amino acids 3 to 133 of Tra-2 (figure 4.2). This sequence includes the remainder of the RRM, and the majority of the first RS domain of Tra-2. H11792 lacks a good candidate for an initiating methionine. Therefore, the H11792 cDNA does not contain all of the 5' coding sequence for clone 6.9. Although clone 6.9 sequence is not complete, comparison of the human H11792 cDNA and the mouse clone 6.9 cDNA suggests that there is a mammalian counterpart to Tra-2 that contains two RS domains divided by an RRM.

A composite sequence of H11792 combined with clone 6.9 is 42% identical (58% similar) to amino acids 3 to 259 of Tra-2. The composite RRM contains all the conserved residues of an RRM including the RNP1 and RNP2 consensus motifs (Birney et al., 1993; Bandziulis et al., 1989). This RRM is 58% identical (73% similar) to the RRM of Tra-2. Besides the high degree of conservation, there are specific regions of identity that suggest Tra-2 and the composite of clone 6.9 have similar RNA-binding specificity. First, over loop 5, β4, and the C-terminal extension, twelve amino acids are conserved between the two RRMś (figure 4.3). In U1A, this region contacts U1 snRNA (Oubridge et al., 1994). Second, in order to bind RNA, many RRMś require residues directly outside of the RRM to make additional contacts with RNA (Kenan et al., 1991). Clone 6.9 and Tra-2 have nine identical residues and three conservative changes in the twelve amino acids that follow the RRM.

Expression pattern of clone 6.9 - The EST cDNAs of clone 6.9 were isolated from human lung,
Figure 4.3. Alignment of the RRM of Tra-2 with the RRM from the composite of H11792 and clone 6.9. Identical residues between the two sequences are highlighted in bold reverse. The corresponding secondary structure of the RRM of U1A is shown above the alignment (Nagai et al., 1990). The RNP1 and RNP2 motifs are shown on the bottom. Abbreviations: α - alpha helix, β - beta strand, CTE - C terminal extension. 6.9comp - composite sequence of H11792 and clone 6.9.
white blood cells, prostate, and fetal brain. I isolated clone 6.9 from murine T cells. Therefore, clone 6.9 is probably expressed in a variety of tissues. None of the cDNAs isolated so far correspond to the alternatively spliced variants of *tra-2* seen in *Drosophila* (Amrein et al., 1990; Mattox et al., 1990). It will be interesting to determine if these splice variants are conserved in mammalian cells.

**Discussion**

**Clone 6.9 is a mammalian counterpart of *Tra-2*** - *Tra-2* is necessary in *Drosophila* for the female specific splicing of the *dsx* pre-mRNA transcript. In addition, *Tra-2* regulates its own pre-mRNA splicing, and the splicing of the *exu* transcript in male germ cells (Mattox and Baker, 1991; Hazelrigg and Tu, 1994). We have cloned a *tra-2* like cDNA, clone 6.9, from murine T cells. Clone 6.9 encodes a C-terminal RS domain and an incomplete RRM but lacks further 5' sequence. H11792, a human cDNA of clone 6.9, contains additional 5' sequences that code for the remainder of the RRM and part of an N-terminal RS domain, but lacks an initiating methionine. A composite protein that combines clone 6.9 with H11792 is 42% identical (59% similar) to *Tra-2*. Over the RRM, this identity is increased to 58% (73% similarity). As well, this composite protein contains the same protein structure of *Tra-2* with two RS domains divided by an RRM. Given the high degree of similarity across the two proteins, we propose that clone 6.9 is the first member of the *tra-2* family in mammalian cells.

Clone 6.9 cDNAs are present in libraries generated from several human tissues suggesting that the transcript is widely expressed. None of these cDNAs code for alternatively spliced transcripts, although it seems likely that such mammalian *tra-2* transcripts exist. Work is in progress to clone the entire mouse cDNA for clone 6.9, and determine if clone 6.9 is alternatively
A Possible role for clone 6.9 - Tra-2 and SR proteins bind the dsxRE in Drosophila (Tian and Maniatis, 1993; Tian and Maniatis, 1994). In mammalian cells, SR proteins bind purine-rich enhancers, known as exon recognition sequences (ERSs) (Tanaka et al., 1994; Sun et al., 1993b; Ramchatesingh et al., 1995; Lavigne et al., 1993). There are several reasons why SR proteins are unlikely to be sufficient for the enhancer activity of these sequences. First, SR proteins cannot bind directly to a subset of GAA purine-rich enhancers (Fu, 1995). Second, SR proteins are general splicing factors. In addition to these factors, a transcript that is alternatively spliced may require specific factors. In this case, the specific factor may recruit SR proteins to the ERS of that transcript. Clone 6.9 is a good candidate to recruit SR proteins based on its similarity to Tra-2.

Clone 6.9 interacts with the Clk/Sty kinase - Clone 6.9 was isolated because it interacted with Clk/Sty in the two-hybrid system. Clk/Sty interacts with several SR proteins, and phosphorylates these proteins in vitro (chapter 2; chapter 3). In mammalian cells, there is mounting evidence that splicing is regulated by phosphorylation (Mermoud et al., 1994; Cardinali et al., 1994; Tazi et al., 1992; Tazi et al., 1993; Mermoud et al., 1992; Gui et al., 1994a; chapter 2). It is likely that Tra-2 in Drosophila and clone 6.9 in mammalian cells are also regulated by phosphorylation. Phosphorylation of clone 6.9, like the SR proteins, may regulate spliceosome assembly (Mermoud et al., 1994; Roscigno and Garcia-Blanco, 1995). Clk/Sty is a good candidate to regulate clone 6.9 during spliceosome assembly, because it phosphorylates clone 6.9 in vitro (E. Aippersbach, K. Colwill, T. Pawson, unpublished results). Besides Clk/Sty, there are four other mammalian kinases that phosphorylate RS domain-containing proteins: SRPK1, U1 70K kinase, LBR kinase.
and DNA topoisomerase I (Gui et al., 1994a; Woppmann et al., 1993; Nikolakaki et al., 1996; Rossi et al., 1996). One or more of these kinases may also regulate clone 6.9 in vivo.

Given the importance of Tra-2 in fly development, I expect that clone 6.9 will be important for development in the mouse by regulating key developmental decisions. It is likely that clone 6.9 will operate via splicing enhancers in a manner analogous to that of Tra-2. I suggest that the function of clone 6.9 in mammalian cells is modulated by the Clk/Sty protein kinase.
II. RS cyclophilins: Identification of a NK-TR₁ related cyclophilin associated with the Clk/Sty kinase.

Introduction

I describe the isolation of a cDNA coding for an arginine/serine-rich (RS) cyclophilin protein from a yeast two-hybrid screen using the Clk/Sty protein kinase as a probe. This Clk/Sty associating RS-cyclophilin (CARS-Cyp) possesses 39% identity to NK-TR₁ (Natural Killer Tumor Recognition protein-1) that plays a role in the lytic function of natural killer (NK) cells (Frey et al., 1991; Anderson et al., 1993; Chambers et al., 1994; Giardina et al., 1995). NK-TR₁ contains a cyclophilin domain, three Nopp140-related domains, and three RS domains. The cyclophilin domain of NK-TR₁ has a histidine residue in place of the characteristic tryptophan found in the CsA-binding pocket of Cyclophilin-A. This domain possesses the protein folding activity demonstrated for several other cyclophilin family members (Fischer et al., 1989; Freskgård et al., 1992; Rinfret et al., 1994). Nopp140 domains are acidic/basic regions of approximately 40 amino acids (Meier and Blobel, 1992). The Nopp140 protein acts as a chaperone for protein import into the nucleolus by binding to nuclear localization signals (NLSs) on these proteins. The ten Nopp140 domains in this protein are the assumed binding sites for NLSs.

Based on the similarity of CARS-Cyp and NK-TR₁, we have defined a new class of proteins termed RS-cyclophilins. I suggest that this family may be regulated by Clk/Sty as CARS-Cyp and Clk/Sty interact in the two-hybrid screen and CARS-Cyp contains RS dipeptide repeats that Clk/Sty may phosphorylate (chapter 3). Based on the presence of an RS domain, and CARS-Cyp association with the Clk/Sty kinase, I propose that RS-cyclophilins play an important role in the regulation of pre-mRNA splicing.
Materials and Methods

Isolation of CARS-Cyp cDNA clones - I described the two-hybrid screen in chapter 2. The cDNA insert of mouse T cell clone 8.9 was used to screen a λgt10 human thymus cDNA library (Clontech #HL1074a) using standard methods (Maniatis et al., 1982).

Sequencing of CARS-Cyp - cDNA inserts of individual λ clones cleaved with EcoRI or the XhoI inserts from the pACT two-hybrid vector were purified, digested with HaeIII, AluI, or MboI, and cloned into the appropriate restriction sites of M13mp18. Single-stranded template DNA was purified and sequenced with Sequenase (US Biochemical Corp., Cleveland, OH). Both strands of all possible fragments were sequenced. A blast search of the GenBank database of expressed sequence tagged cDNAs was performed to find cDNAs corresponding to CARS-Cyp.

Northern analysis - RNA was separated on an 0.8% agarose gel containing formaldehyde, and blotted onto nylon membranes following standard techniques (Maniatis et al., 1982). The blot was probed with a purified CARS-Cyp cDNA insert (HT-21) that was labelled with \([\alpha^{32}\text{P}]\) dCTP by the random primed method (BRL, Gaithersburg, MD). Leukocytes were separated from buffy coats of human peripheral blood by centrifugation on Ficoll-Hypaque gradients. Monocytes were adhered to plastic petri dishes, washed extensively with PBS, and collected by treatment with Versene solution and scraping. B cells were collected on nylon wool columns, and the remaining lymphocytes were fractionated on Percoll gradients as previously described (Frey et al., 1991). T cell preparations were >95% CD3<sup>+</sup> as determined by flow cytometry. NK cells were >90% pure and <5% CD3<sup>+</sup>.

Results

Isolation of the CARS-Cyp cDNA - I isolated two CARS-Cyp cDNA clones, 8.9 and 9.9, from
a mouse T cell cDNA library in a yeast two-hybrid screen using the Clk/Sty kinase as a probe (chapter 2). The binding of CARS-Cyp was specific to Clk/Sty, as neither clone activated the reporter gene directly, or interacted with CDK2 or SNF1 (data not shown). The mouse partial cDNA was used to screen a human thymus cDNA library. Overlapping human cDNAs spanning the entire coding sequence of CARS-Cyp were isolated (figure 4.4). Also shown in figure 4.4 are the relative locations of the mouse CARS-Cyp clones, and twenty four sequences from GenBank that represent fragments of the human CARS-Cyp cDNA. The nucleotide sequence of the composite human cDNA, and of the partial mouse CARS-Cyp cDNA are shown in figure 4.5. The mouse cDNAs possess 90% nucleotide and 95% amino acid identity to the human sequence.

The human cDNA has an open reading frame coding for 754 amino acids (predicted mass 89 kDa) preceded by a 500 nucleotide 5' untranslated region. The amino terminus of the predicted protein is homologous to cyclophilin. The remainder of the protein is highly charged, and contains two Nopp140-related domains and a large RS domain. Amino acids that are identical between NK-TR, and CARS-Cyp are underlined in figure 4.5. The cyclophilin domain is highly conserved between the two proteins (70% identity) including the substitution of histidine for tryptophan in the cyclosporin A (CsA) binding pocket (amino acid 133 of CARS-Cyp). This substitution results in a decreased affinity for CsA in the NK-TR, cyclophilin domain (Rinfret et al., 1994).

Although the similarity of the non-cyclophilin portion of these molecules is fairly low (28% identity), the type of domains and their relative locations are conserved. The CARS-Cyp protein appears to be equivalent to the amino-terminal half of the NK-TR, protein. In this region, both proteins contain two Nopp140-related domains at similar locations, followed by an RS
Figure 4.4. Isolation of CARS-Cyp cDNA clones. A schematic representation of the CARS-Cyp protein is shown above the composite human cDNA. The length and position of cDNA clones isolated from either mouse T cells or human thymus is shown. The lower portion indicates the location of 20 sequences from GenBank that are identical to the CARS-Cyp cDNA. These clones are identified by accession number and tissue source.
Figure 4.5. Sequence of CARS-Cyp. The nucleotide sequence of the human CARS-Cyp cDNA is shown, with the predicted amino acid sequence listed above it. The nucleotides or predicted amino acids of the partial mouse cDNA that differ from the human CARS-Cyp are shown below the human sequence. The 5' and 3' ends of the mouse cDNA sequences are indicated by vertical lines. The ends of clone 8.9 are shown as thin lines, and the ends of clone 9.9 are shown as thick lines. The nucleotide numbering of the human cDNA is shown on the right side of each line. Amino acid numbers are in brackets. The amino acids that are identical between CARS-Cyp and NK-TR, are underlined. Nopp140-related domains are boxed. The CARS-Cyp sequence is available from GenBank (accession number U40763).
domain. Nopp140 domains have a stretch of acidic amino acids followed by a stretch of basic amino acids (Meier and Blobel, 1992). The first Nopp140-related domain of CARS-Cyp has the same structure. However, the second Nopp140-related domain lacks the acidic portion of the repeat, and may not be a real Nopp140-related domain.

Since Clk/Sty binds proteins containing RS domains (chapter 2), it was surprising that neither of the Clk/Sty interacting clones 8.9 or 9.9 contain the large C-terminal RS domain. These clones do contain four separate RS dipeptides, and one stretch of "SRSRSR" that may be responsible for the ability of the two clones to interact with Clk/Sty. Alternately, Clk/Sty may bind to the first Nopp140-related domain that is present in both clones.

**Analysis of CARS-Cyp expression** - Northern blot analysis of CARS-Cyp mRNA from several human tissues showed similar levels of a 4 kb transcript in lung, liver, kidney, small intestine, testis, and brain (data not shown). Clones corresponding to fragments of the CARS-Cyp cDNA have also been isolated from fetal brain, cochlea, liver and spleen, infant brain, adult brain, placenta, T cells, heart, olfactory epithelium, colorectal cancer, and WI-38 fibroblasts (figure 4.4). It, therefore, appears that CARS-Cyp transcripts are widely expressed. However, since the NK-TR$_1$ gene is expressed only in NK cells, CARS-Cyp mRNA levels in human T, macrophage, NK, and B cells were examined (figure 4.6). Besides the 4 kb transcript, a minor band around 10 kb was detected in B cells. Interestingly, CARS-Cyp expression was barely detectable in NK cells, and the low level of expression can be attributed to the 10% contamination of B and T cells in the NK cell preparation. The CARS-Cyp mRNA was not detected in highly purified mouse NK cells (data not shown). The lack of expression of the CARS-Cyp gene in NK cells may be due to the presence of NK-TR$_1$ in this cell type. The function of CARS-Cyp may be performed by
Figure 4.6. Northern blot of human peripheral blood mRNA. A) 5 µg of total cellular RNA from human T cells (lane 1), macrophages (lane 2), NK cells (lane 3), and human B cells (lane 4) were hybridized with the CARS-Cyp cDNA. Panel B shows the ethidium bromide stained gel to control for differences in RNA loading.
NK-TR₁ in NK cells, since the proteins are closely related.

Discussion

A family of RS-Cyclophilins - I describe the cloning of CARS-Cyp, a novel cDNA similar to NK-TR₁. The similarity between these two proteins suggests they belong to a family of related proteins. Indeed, a similar protein, BmCyp-1, was cloned from the human filarial parasite Brugia malayi (Page et al., 1995). BmCyp-1 is a 93 kDa molecule that contains a cyclophilin domain closely related to the NK-TR₁ and CARS-Cyp cyclophilin domains including the characteristic histidine residue in the CsA binding site. BmCyp-1 also contains two Nopp140-related domains and a long carboxyl-terminal RS domain. These proteins are defined as RS-cyclophilins.

The CARS-Cyp transcript is present in every tissue and cell type examined except for NK cells. In contrast, the NK-TR₁ transcript is found only in NK cells (Anderson et al., 1993). From the differential expression of CARS-Cyp and NK-TR₁ transcripts, I suggest that NK-TR₁ functions in place of CARS-Cyp in NK cells.

RS-Cyclophilins are multi-domain proteins - CARS-Cyp, NK-TR₁, and BmCyp-1 contain three distinct domains: cyclophilin, Nopp140-related, and RS. The presence of the cyclophilin domain suggests that these proteins function as chaperone proteins to assist in protein folding. In fact, the cyclophilin domain of NK-TR₁ is an active peptidyl-prolyl cis-trans-isomerase and displays chaperone activity similar to the original cyclophilin (Rinfret et al., 1994; Freskgård et al., 1992; Fischer et al., 1989). Nup358 is a nuclear transport protein that contains a cyclophilin domain with the histidine substitution characteristic of the RS-cyclophilins (Wu et al., 1995). This protein contains binding sites for soluble transport factors, and likely functions as a docking site.
for transport factors carrying proteins through the nuclear pore. The cyclophilin domain of Nup358 may facilitate the refolding of proteins after they have been transported through the nuclear membrane.

The presence of Nopp140-related domains suggest that the RS-cyclophilins are involved in protein-protein interactions by binding the NLSs of other proteins (Meier and Blobel, 1992; Meier and Blobel, 1990). RS domains are also involved in protein-protein interactions (Amrein et al., 1994; Kohtz et al., 1994; Wu and Maniatis, 1993). RS-cyclophilins may utilize these domains to form multi-protein complexes.

One model for CARS-Cyp function is to transport SR proteins from the cytoplasm to nuclear speckles. In this model, CARS-Cyp binds SR proteins in the cytoplasm through their RS domains, or through the interaction of the Nopp140-related domains with the NLSs of the SR proteins. This complex is transported into the nucleus, where it is targeted to speckles via the RS domains of CARS-Cyp or the SR proteins. Once in the speckle, the cyclophilin domain of CARS-Cyp may refold the SR proteins and/or assist in the assembly of SR proteins and other proteins into the spliceosome. A simpler model for CARS-Cyp function is that CARS-Cyp is involved only in the folding of RS domain-containing proteins once they reach the speckle and are not involved in their transport.

A rat RS-cyclophilin, matrin Cyp, is 93% identical to CARS-Cyp at the amino acid level (M. Mortillaro and R. Berezney, unpublished results). Matrin Cyp is a nuclear matrix protein that is localized to nuclear speckles. This localization of a homologue of CARS-Cyp in rat cells provides further evidence for the above two models. Further analysis is necessary to determine if CARS-Cyp tracks RS domain-containing proteins to speckles, or if CARS-Cyp is retained in
CARS-Cyp interacts with the Clk/Sty kinase - Although I have not tested CARS-Cyp as a substrate for Clk/Sty, evidence suggests that CARS-Cyp is a substrate. First, Clk/Sty and CARS-Cyp interact in the two-hybrid system. Second, matrin Cyp, the rat homologue of CARS-Cyp is phosphorylated in vivo implying that CARS-Cyp is also a phosphoprotein. Third, CARS-Cyp contains both RSRS and KSKS repeats that can be phosphorylated by Clk/Sty (chapter 3). Lastly, matrin Cyp is localized to speckles suggesting that CARS-Cyp is present in speckles. Clk/Sty is also present in speckles (chapter 2) so it is possible for CARS-Cyp and Clk/Sty to interact in vivo. Phosphorylation by Clk/Sty within RS repeats may help neutralize the arginines in the RS domains of CARS-Cyp so it can interact with SR splicing factors.

CARS-Cyp is a novel member of the RS-cyclophilin family which includes three members to date. Given the domain structure of these proteins, it is possible that these proteins operate to fold proteins involved in pre-mRNA splicing perhaps during spliceosome assembly.
III. Analysis of the Two-hybrid Clones

Introduction

In the two-hybrid screen, Clk/Sty interacted with twenty six different proteins. Of these proteins, five are discussed in chapter two, and two are discussed within the first two parts of this chapter. In this section, I discuss aspects of the screen and the other nineteen cDNAs identified in my two-hybrid screen.

Materials and Methods

The two-hybrid screen and generation of Clk/Sty constructs have been described in chapter 2.

Results

The two-hybrid library - Clones identified in the Clk/Sty two-hybrid screen are summarized in table 4.1. I sequenced at least 100 basepairs from each clone. Two general points about the cDNA library can be noted from this table. First, 34 clones were isolated in total but due to redundancy in the library, there are only 26 different clones. Second, this library was size selected to be a minimum of 600 base pairs. In many cases, the resulting clones coded for almost full-length proteins. In the screen with Clk/Sty, this size likely proved helpful. For instance, when X16 was divided into its RRM and RS domains, an interaction between X16 and Clk/Sty was not detected for the RRM, and only a weak interaction was seen with the RS domain (chapter 2). Therefore, a smaller sized library may have missed the interaction between X16 and Clk/Sty.

Clk/Sty constructs in the two-hybrid system - In the initial screen, I used full-length Clk/Sty
<table>
<thead>
<tr>
<th>Clone Number</th>
<th>Redundant Clones</th>
<th>Clone Size (bp)</th>
<th>Binds to Clk RS Domain</th>
<th>Identity</th>
<th>GenBank Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.6</td>
<td>9.5, and 10.4</td>
<td>1100,1100,1150</td>
<td>yes (9.5 does not)</td>
<td>X16</td>
<td>X53824</td>
</tr>
<tr>
<td>6.3</td>
<td>None</td>
<td>1700</td>
<td>yes</td>
<td>ASP/SF2</td>
<td>M72709</td>
</tr>
<tr>
<td>9.11</td>
<td>None</td>
<td>1100</td>
<td>no</td>
<td>SRp55</td>
<td>U30883</td>
</tr>
<tr>
<td>6.8</td>
<td>None</td>
<td>1500</td>
<td>no</td>
<td>hnRNPG</td>
<td>Z23064</td>
</tr>
<tr>
<td>7.2</td>
<td>None</td>
<td>1600</td>
<td>yes</td>
<td>RNP S1</td>
<td>X70067</td>
</tr>
<tr>
<td>6.9</td>
<td>None</td>
<td>1500</td>
<td>not tested</td>
<td>mammalian tra-2</td>
<td></td>
</tr>
<tr>
<td>8.9</td>
<td>9.9</td>
<td>800,700</td>
<td>yes (8.9 does not)</td>
<td>CARS-Cyp</td>
<td>U40763</td>
</tr>
<tr>
<td>2.8</td>
<td>None</td>
<td>1300</td>
<td>no</td>
<td>c-Myc</td>
<td>X01023</td>
</tr>
<tr>
<td>3.9</td>
<td>5.3</td>
<td>2000,1500</td>
<td>yes</td>
<td>SAF-B</td>
<td>L43631</td>
</tr>
<tr>
<td>4.11</td>
<td>None</td>
<td>1600</td>
<td>yes</td>
<td>ZFP38</td>
<td>D10630</td>
</tr>
<tr>
<td>5.7</td>
<td>None</td>
<td>2400</td>
<td>no</td>
<td>KIAA0164</td>
<td>D79986</td>
</tr>
<tr>
<td>5.12</td>
<td>None</td>
<td>1500</td>
<td>yes</td>
<td>novel zinc finger protein</td>
<td></td>
</tr>
<tr>
<td>6.12</td>
<td>None</td>
<td>1200</td>
<td>no</td>
<td>N-CoR</td>
<td>U35312</td>
</tr>
<tr>
<td>8.10</td>
<td>None</td>
<td>1000</td>
<td>no</td>
<td>similar to 82 kDa protein</td>
<td>U15552</td>
</tr>
<tr>
<td>1.3</td>
<td>None</td>
<td>1100</td>
<td>no</td>
<td>Unique</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>None</td>
<td>900</td>
<td>yes</td>
<td>Unique</td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>None</td>
<td>700</td>
<td>no</td>
<td>Unique</td>
<td></td>
</tr>
<tr>
<td>2.9</td>
<td>None</td>
<td>2100</td>
<td>no</td>
<td>Unique</td>
<td></td>
</tr>
<tr>
<td>3.4</td>
<td>None</td>
<td>2100</td>
<td>yes</td>
<td>Unique</td>
<td></td>
</tr>
<tr>
<td>3.8</td>
<td>None</td>
<td>3000</td>
<td>yes</td>
<td>Unique</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>None</td>
<td>1300</td>
<td>no</td>
<td>Unique</td>
<td></td>
</tr>
<tr>
<td>5.8</td>
<td>None</td>
<td>1000</td>
<td>yes</td>
<td>Unique</td>
<td></td>
</tr>
<tr>
<td>6.4</td>
<td>None</td>
<td>800</td>
<td>yes</td>
<td>Unique</td>
<td></td>
</tr>
<tr>
<td>6.11</td>
<td>None</td>
<td>800</td>
<td>yes</td>
<td>Unique</td>
<td></td>
</tr>
<tr>
<td>9.1</td>
<td>3.10,9.2,9.7,10.6</td>
<td>700,900,750,900,800</td>
<td>no</td>
<td>Unique</td>
<td></td>
</tr>
<tr>
<td>10.8</td>
<td>None</td>
<td>1200</td>
<td>no</td>
<td>Unique</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.1. Characteristics of the two-hybrid clones.** The clones are divided into three sections where the first five clones are described in chapter 2, the next nine clones are described in chapter 4, and the remaining clones are not discussed at all. Abbreviations: bp - base pairs, Ref. - reference.
as the probe. In order to better define the interaction between Clk/Sty and the two-hybrid clones, I tested three constructs of Clk/Sty. First, I used a kinase inactive Clk/Sty. All clones, with the exception of clone 10.8, bound to the inactive Clk/Sty. Therefore, for the products of most clones, binding to Clk/Sty does not require Clk/Sty kinase activity. As well, these clones bind Clk/Sty independently of the presence of autophosphorylated sites within Clk/Sty.

Second, I divided Clk/Sty into its RS domain (amino acids 1 to 156) and its kinase domain (amino acids 157 to 483). None of the clones bound to the kinase domain by itself. Approximately half of the clones bound to the RS domain alone (table 4.1). Therefore, the RS domain of Clk/Sty is necessary for interaction with the two-hybrid clones, and in some cases, it is sufficient for these interactions. For Clk/Sty, half of the signals were weaker for the RS domain than for the full-length protein. Therefore, it is likely that both domains of Clk/Sty are required for optimal binding.

Identity of the two-hybrid clones - To identify the clones or to look for related proteins, I performed Fasta or Blast searches of GenBank. I described the first five proteins in table 4.1 in chapter 2. I discussed mammalian tra-2 and CARS-Cyp in parts I and II of this chapter. Twelve cDNAs are novel and do not show any homology to sequences within the database. I describe the remaining seven clones below.

Clones 3.9 and 5.3 are SAF-B Scaffold attachment factors (SAFs) anchor DNA on the scaffold nuclear substructure by binding to scaffold attachment regions (SARs) on DNA (Romig et al., 1992). SAF-B specifically binds to SARs and is associated with chromatin in vivo (Renz and Fackelmayer, 1996). The sequenced portion of clone 3.9 is 92% identical (97% similar) to SAF-B at the amino acid level. Therefore, clone 3.9 is likely the mouse SAF-B. For clone 5.3, the
identity to SAF-B is only 52% identical (69% similarity) over 29 amino acids. A further analysis of clone 5.3 will determine if this lack of identity is due to inaccurate sequencing, or if clone 5.3 represents a novel protein related to SAF-B. Clone 3.9 starts at amino acid 337, and clone 5.3 starts at amino acid 450 of SAF-B (figure 4.7a). Both clones contain the rest of the coding sequence of SAF-B.

SAF-B is an acidic protein with a pI of 5.21. It is enriched with arginine (11%), glutamic acid (13%), and serine (11%). Of particular interest is an RD/RE-rich region between amino acids 560 and 780 that contains RD and RE dipeptide repeats. RD/RE repeats may interact with each other in a polar zipper structure (Perutz, 1994). Phosphorylated RS domains mimic the alternating charge of RD/RE repeats. Perhaps the interaction of Clk/Sty with SAF-B is occurring through an interaction of the RS domain of Clk/Sty with the RD/RE region.

**Clone 5.7 is KIAA0164** - KIAA0164 cDNA was cloned in a screen that identified expressed mRNAs from the human myeloblast cell line KG1 (Nagase et al., 1995). This cDNA codes for a protein of 920 amino acids. Over 93 amino acids, clone 5.7 is 100% identical to human KIAA0164 except KIAA0164 has an additional serine at position 15. Clone 5.7 begins at amino acid 192 of KIAA0164, and extends through the rest of the protein (figure 4.7b).

The first 120 amino acids of KIAA0164 make up an RS domain (31% R, 30% S). Residues 670 to 920 are charged (13% E, 13.5% K, 12% R, 11% S). Clone 5.7 lacks the RS domain, but contains the charged region. Perhaps Clk/Sty binds the region that contains several RE repeats and a few RS repeats. If so, this interaction is not simply an RS domain/charged region interaction because the RS domain of Clk/Sty is not sufficient for this interaction (table 4.1).
Figure 4.7. Schematic representation of the clones isolated in the two-hybrid screen.

Proteins encoded by the individual clones are diagrammed (a) to (e). The length of each protein in amino acids is indicated on the bottom. Individual domains are noted. The arrow indicates the beginning of the clone in the context of the full-length protein. All clones except 6.12, code for the C-terminal end of the protein. The sequence of N-CoR that is represented in clone 6.12 is indicated by the arrow. Abbreviations: RD/RE - arginine/aspartic acid/glutamic acid-rich. RS - arginine/serine-rich. RepD - repressor domain. KD/KE - lysine/aspartic acid/glutamic acid-rich. ID - interaction domain. TD - transcription activation domain, bHZ - basic helix-loop-helix leucine zipper domain, Zn Fin - zinc finger domain.
Clone 6.12 is N-CoR - Thyroid-hormone and retinoic acid receptors are ligand-regulated transcription factors that modulate specific developmental programs of gene expression (Chambon, 1994). These receptors bind DNA as heterodimers with retinoid-X receptors (Yu et al., 1991; Leid et al., 1992; Kliwer et al., 1992; Marks et al., 1992; Zhang et al., 1992). If the ligand is absent, these heterodimers repress transcription of the genes that they activate when the ligand is present (Sap et al., 1989; Baniahrmad et al., 1990; Graupner et al., 1989). This repression is due to the Nuclear receptor co-repressor (N-CoR) that binds the heterodimer (Hörlein et al., 1995).

N-CoR is 2454 amino acids long. Over 185 amino acids, the C-terminus of 6.12 is 100% identical in amino acid sequence to N-CoR. However, the N-terminal end of 6-12 is not homologous to N-CoR. It is possible that the sequence of the 5' end of clone 6.12 is not accurate. Alternately, this clone may represent an alternatively spliced variant of N-CoR, or may represent a recombination event within the cDNA library. More sequence analysis is required to clarify the discrepancy between the 5' and 3' ends of clone 6.12. Clone 6.12 ends at amino acid 521 of N-CoR. Judging by the length of the cDNA, the clone 6.12 protein should begin around amino acid 100 of N-CoR (figure 4.7c).

N-CoR contains two repressor domains at its N-terminus, and a domain at its C-terminus that interacts with the thyroid hormone receptor or the retinoic-acid receptor. Two charged regions are found within N-CoR. The first one (amino acids 511 to 543) has the sequence EEEKVEEKEEDEKKEKDEEKDDKED. The second one (amino acids 1718 to 1744) has the sequence ERERERERERERERERERERERERERER. Clone 6.12 contains the first 13 amino acids of the first repeat, and lacks the second repeat. Perhaps Clk/Sty binds to the 13
amino acids of this motif present in clone 6.12. *In vivo*, these charged regions are prime candidates for interaction of N-CoR with SR or RD/RE proteins.

**Clone 8.10 is similar to an acidic 82 kDa protein** - Over the 52 amino acids predicted from the sequenced portion of clone 8.10, there is 54% identity and 63% similarity to a human acidic 82 kDa protein cDNA (P. Carlsson, unpublished results). Therefore, clone 8.10 is related to this protein. More sequence of clone 8.10 is needed to determine if the conservation extends past these 52 amino acids.

The pI of this 82 kDa protein is 5.17. This protein contains several charged runs: EEEDK (amino acids 506 to 510), EEEKEEEEDEKSEEED (amino acids 517 to 541), and DEED (amino acids 544 to 547). If these regions are conserved in clone 8.10, then they may be possible interaction sites with the positively charged Clk/Sty protein.

**Clone 2.8 is c-Myc** - c-Myc is a transcription factor involved in transformation, mitogenesis, and apoptosis (for reviews see Kato and Dang, 1992; Evan and Littlewood, 1993). The sequenced portion of clone 2.8 codes for 64 amino acids that are 100% identical to c-Myc. From this sequence, I conclude that clone 2.8 is c-Myc. Mouse c-Myc is 439 amino acids long, and is divided into several domains (figure 4.7d) (Stanton et al., 1984; Kato and Dang, 1992). Clone 2.8 protein begins after the transactivation domain, and extends to the end of the protein. Within this region is the HLH-ZIP motif that dimerizes with a similar motif in Max (Blackwood and Eisenman, 1991).

The portion of c-Myc that clone 2.8 overlaps contains 3 SR dipeptides, 1 SPK motif, and 1 SPR motif that may serve as consensus sites for Clk/Sty phosphorylation (chapter 3). As well, there is one acidic motif of EEEQEDEEE. There are no obvious RD/RE/RS motifs in this
Clone 4.11 is ZFP38 - ZFP38 is a zinc finger protein whose expression is highest in meiotic cells in both testes and ovaries (Noce et al., 1992; Chowdhury et al., 1992). Thus, ZFP38 may regulate transcription during gametogenesis. Over 42 amino acids, clone 4.11 is 100% identical to ZFP38 with one ambiguity in the coding sequence of clone 4.11. Clone 4.11 begins at amino acid 152 of ZFP38, and extends through the rest of the protein (figure 4.7e).

ZFP38’s seven zinc fingers reside in the C-terminal end of the protein. The transactivation domain is found at the N-terminus. Within this latter region, there are three acidic repeats. Clone 4.11 begins after the transactivation domain. In the two-hybrid system, Clk/Sty binds to the zinc fingers but not to the transactivation domain. There are no obvious RS or RD/RE regions to which Clk/Sty may bind.

Clone 5.12 is a novel zinc finger protein - Clone 5.12 is 1500 basepairs in length. I sequenced the first 545 nucleotides in the forward direction (figure 4.8). A stop codon appears at nucleotide 455 that likely represents the start of unreliable sequence data. The open reading frame so far is 161 amino acids. There are two zinc fingers within this region that fit within the consensus \( \text{X}_5\text{CX}_{2-4}\text{CX}_{12}\text{HX}_{3-4}\text{HX}_4 \) (Jacobs and Michaels, 1990). There are two other putative zinc fingers that lack either the last histidine, or the second cysteine and the last histidine. Sequencing of both strands will determine if the lack of these residues is due to a sequencing error. The four putative zinc fingers run from amino acids 31 to 143. Since Clk/Sty binds the zinc finger domain of ZFP38, it is possible that Clk/Sty binds this region of clone 5.12.

Discussion

I employed the two-hybrid screen to find binding partners for the Clk/Sty kinase. Before
Figure 4.8. **Sequence of the 5' end of clone 5.12.** The known nucleotide sequence of clone 5.12 is shown. The numbering of this sequence is shown on the right. The corresponding amino acid sequence is shown below the nucleotide sequence. The zinc fingers are divided by lines.
this screen, the function of Clk/Sty was unclear. The discovery of SR proteins in this screen allowed me to show that Clk/Sty is upstream of these splicing factors (chapter 2). However, RNA-binding proteins account for only six out of the twenty-six clones isolated unless some of the novel proteins turn out to be RNA-binding proteins. The other identified clones suggest that Clk/Sty may have other roles besides regulating splicing factors. One of these roles may be in transcription. In this screen, Clk/Sty interacted with the transcription factor c-Myc, two putative transcription factors, ZFP38 and the novel zinc finger protein, as well as a transcription repressor N-CoR. Further analysis of the interaction of Clk/Sty with these proteins is required before we can state that Clk/Sty does regulate transcription.

The function of SAF-B or KIAA0164 has not been determined so it is difficult to speculate on the interaction of Clk/Sty with these proteins. SAFs bind SAR regions in DNA to anchor DNA to the nuclear scaffold. In many cases, SAR elements overlap boundaries of actively transcribed chromatin (Mirkovitch et al., 1984; Bode and Maass, 1988). An interaction between SAF-B and Clk/Sty may place Clk/Sty near the sites of RNA processing.

**Clk/Sty interaction sequences** - The RS domain of Clk/Sty is sufficient for interaction with about half of the two-hybrid clones. However, the RS domain of X16 weakly interacts with Clk/Sty, and the RS domain of ASF/SF2 does not interact at all with Clk/Sty. Therefore, the interaction is not a simple RS-RS interaction. Other sequences are required. Still, RS (or KS) repeats and RD/RE (or KD/KE) repeats appear in at least ten of the twenty-six proteins isolated. This number will likely increase as the remaining clones are sequenced in their entirety. Therefore, it is likely that Clk/Sty recognizes these sequences to some extent. Clk/Sty phosphorylates RS and KS repeats providing further evidence of an interaction of Clk/Sty with
these domains (chapter 3).

The isolation of the three transcription factors that do not contain any obvious charged domains suggests that Clk/Sty can interact with proteins that lack charged repeats. For ZFP38, the interaction occurs through the zinc finger domain. Further analysis of all the clones is required to determine Clk/Sty interaction sites, as well as to determine if these clones interact with Clk/Sty outside of the two-hybrid system.
CHAPTER 5: General Discussion and Future Directions
The goal of my PhD research was to determine a role for Clk/Sty in cellular function. The discovery of Clk/Sty in an anti-phosphotyrosine screen gave no clue as to a possible function for this kinase. I reasoned that if I could identify proteins that interact with Clk/Sty, I could begin to understand the role of Clk/Sty in vivo. After considerable effort using more conventional biochemical and genetic approaches, the two-hybrid screen was used to identify interacting proteins.

**The two-hybrid screen**

I screened 2.5 million colonies in the two-hybrid system using Clk/Sty as bait. In order to ensure saturation of the library, I needed to screen 6 million colonies (Brent et al., 1987). Therefore, this screen may not have identified some poorly expressed mRNAs. In addition, I used a library from T cells that may not contain sequences for proteins that Clk/Sty interacts with in other cell types. As well, any protein that is lethal when over-expressed in yeast is not represented in this screen. Finally, during the initial transformation, more than one cDNA may have entered the same cell. Therefore, a positive colony may contain a 'negative' cDNA whose protein product does not interact with Clk/Sty. After purification of cDNAs from positive yeast cells, and re-introduction back into the two-hybrid system, several positive signals were lost. This loss of signal may have been due to isolating a 'negative' cDNA during purification. For all of the above reasons, it can be stated that lack of a cDNA in this screen does not indicate a lack of Clk/Sty interaction with its protein product. For instance, the two-hybrid screen identified only three members of the SR family as Clk/Sty-interacting proteins. SC35 was not isolated in this screen. Yet, Clk/Sty phosphorylates this protein in vitro, suggesting that it can interact with SC35 and may do so in the two-hybrid system.
Identification of interacting domains

I used a full-length Clk/Sty cDNA in my two-hybrid screen. To narrow down the site of binding, I divided Clk/Sty into its N-terminal RS domain and its C-terminal kinase domain. The kinase domain on its own was insufficient to interact with any of the two-hybrid clones. Therefore, the RS domain is essential for interaction with all of the two-hybrid clones. Approximately half of the clones interacted with the RS domain of Clk/Sty by itself. This result is consistent with the proposed role of RS domains in mediating protein-protein interactions (Wu and Maniatis, 1993; Kohtz et al., 1994; Amrein et al., 1994). Of the clones that interacted with the RS domain of Clk/Sty, half of them showed decreased binding to the RS domain as compared to the full-length protein. ASF/SF2 was one of these clones. Thus for ASF/SF2, both domains are required for optimal binding of Clk/Sty. This observation fits with my model that initial interaction between Clk/Sty and ASF/SF2 takes place through the RS domain of Clk/Sty. Once bound, the kinase domain recognizes its consensus site on ASF/SF2 to phosphorylate the protein.

There is no obvious domain or motif on the two-hybrid clones to which Clk/Sty binds. Many of the clones contain highly charged residues. These residues are often found in RS domains or RD/RE (also KD/KE) repeats. These repeats may interact with the phosphorylated RS domain of Clk/Sty in a polar zipper where all the negative charges are balanced by positive ones (Perutz, 1994). Eight of the clones contain RS domains and this number may increase after the clones are fully sequenced. RS domains are probably not sufficient for interaction with Clk/Sty. For example, the RS domain of ASF/SF2, by itself, does not interact with Clk/Sty. The RS domain of X16 displays a weak interaction with Clk/Sty, whereas the RRM of X16 does not interact with Clk/Sty at all. Therefore, both domains are required for the interaction of X16 or
ASF/SF2 with Clk/Sty. Presumably, other RS domain-containing proteins also make contact with Clk/Sty outside their RS domains.

Not all of the two-hybrid clones contain a high number of RS or RD/RE repeats. Neither c-Myc nor ZFP38 contain appreciable number of these repeats. The binding site for Clk/Sty on c-Myc has not been determined. For ZFP38, Clk/Sty interacts with the zinc finger domain. Therefore, besides charged repeats, Clk/Sty may interact with zinc fingers in the two-hybrid system.

**Relevance of the two-hybrid clones**

The two-hybrid screen identified proteins that interact with Clk/Sty. However, a positive interaction in this screen does not indicate that the same interaction occurs in vivo. I used four criteria to establish a biologically relevant interaction. First, Clk/Sty must interact with the protein. Second, Clk/Sty must co-localize with the protein in vivo. Third, the interaction of the two proteins should modify the function of at least one of them. Lastly, if the protein is a substrate for Clk/Sty, Clk/Sty must phosphorylate this protein on sites in vitro that are also phosphorylated in vivo.

These four criteria have been met to some degree by the SR family of splicing factors. The first criterion is met for X16, ASF/SF2, and SRp55 as they interact with Clk/Sty in the two-hybrid screen. Furthermore, Clk/Sty must have the potential to interact with ASF/SF2, SC35, and SRp55 because it phosphorylates them in vitro. With respect to the second criterion, SR proteins and an overexpressed kinase inactive Clk/Sty co-localize in speckles. The endogenous Clk/Sty kinase also co-localizes with SR proteins to speckles (X.-D. Fu, K. Colwill, T. Pawson, PharMingen, unpublished results). For the third criterion, a link between Clk/Sty and pre-mRNA
splicing has been established since overexpression of Clk/Sty in vivo releases SR proteins and snRNPs from nuclear speckles. As well, overexpression of Clk/Sty modulates 5' splice site switch, a process that is known to be regulated by SR proteins (J. Prasad, K. Colwill, T. Pawson, J.L. Manley, unpublished results). However, there is no direct evidence that Clk/Sty produces these effects by interacting with SR proteins. The last criterion has been tested only with ASF/SF2. In this case, the pattern of ASF/SF2 phosphorylation by Clk/Sty in vitro overlaps the pattern of ASF/SF2 phosphorylation in vivo, suggesting that Clk/Sty likely regulates ASF/SF2 in vivo.

The prominent role of SR proteins in alternative splicing suggests that SR proteins are targets for regulation by the external environment. Evidence to date suggests that SR proteins must be phosphorylated in order to function in spliceosome assembly (Mermoud et al., 1994; Roscigno and Garcia-Blanco, 1995). One reason for this phosphorylation may be to induce protein-protein interactions. Unphosphorylated RS domains may be too positively charged to interact with each other. As well, phosphorylation may reduce non-specific interaction with RNA that may impede spliceosome assembly. Therefore, Clk/Sty phosphorylation of SR proteins may promote protein-protein interactions.

Besides playing a role in protein-protein interactions, SR proteins also influence splice site selection (Ge et al., 1991; Krainer et al., 1991; Zahler et al., 1992). A role for Clk/Sty in this process has also been demonstrated (J. Prasad, K. Colwill, T. Pawson, J.L. Manley, unpublished results). Overexpression of Clk/Sty in vivo modulates 5' splice site selection of both E1A and SV40 T antigen. For E1A, Clk/Sty increases use of the distal 12S and 9S sites, and decreases use of the proximal 13S site. In vitro splicing assays to test the function of Clk/Sty
are still at a preliminary stage. Due to the number of proteins involved in splicing, it is not obvious if the effect of Clk/Sty on splicing is through phosphorylation of SR proteins.

If Clk/Sty acts through SR proteins, then there are several possibilities for how this phosphorylation may influence the role of SR proteins in alternative splicing. The specificity of splice site selection lies in RRM domains (Cáceres and Krainer, 1993; Zuo and Manley, 1993; Tacke and Manley, 1995). Clk/Sty phosphorylates the RRM of ASF/SF2 to a low level. Perhaps this phosphorylation of the RRM is sufficient to alter their substrate specificity. Alternately, Clk/Sty phosphorylation of the RS domain may reduce its non-specific interaction with RNA which, in turn, may influence the specificity of RRM. Most likely, phosphorylation of an SR protein by Clk/Sty may increase the ability of this protein to interact with the U1 snRNP, resulting in more efficient recruitment of the U1 snRNP to potential 5' splice sites.

Finally, Clk/Sty may be instrumental in regulating the availability of SR proteins for splicing by regulating their release from nuclear speckles. Overexpression of Clk/Sty disassembles nuclear speckles in vivo. In normal cells, endogenous Clk/Sty may only release a subset of SR proteins from the speckles. Splice site selection could be dependent on the type of SR protein released. For instance, in vitro, SRp20 promotes 12S splicing of E1A, whereas ASF/SF2 promotes 13S splicing (Screaton et al., 1995; Zahler et al., 1993a). If Clk/Sty released more SRp20 than ASF/SF2, then E1A may be spliced at the 12S site.

In addition to post-translational modification, SR proteins may also be regulated by tissue-specific factors. For instance, SR regulation of female dsx splicing is dependent on Tra and Tra-2 (Tian and Maniatis, 1993; Tian and Maniatis, 1994). Clk/Sty interacted with clone 6.9, a mammalian counterpart to Tra-2, in the two-hybrid screen. If clone 6.9 acts in a manner
analogous to *Drosophila* Tra-2, then it may operate on only a subset of exon enhancers. In this case, Clk/Sty may regulate both general splicing factors such as SR proteins, as well as specific splicing factors such as clone 6.9.

I focused on the interaction of Clk/Sty with SR proteins for three reasons. First, six different clones coded for proteins that contained RRM and RS domains. Therefore, these proteins may define the type of protein with which Clk/Sty interacts. Second, analysis of the N-terminal region of Clk/Sty revealed that it was similar to the RS domains isolated in the screen. Lastly, the function of the SR proteins is well-defined thus we could study the effect of Clk/Sty-SR protein interaction. My focus on the SR proteins does not imply that the other clones are less important. The link between Clk/Sty and transcription is intriguing and implies that Clk/Sty may be involved in regulating both RNA transcription and processing. Clk/Sty may stimulate transcription of a specific gene, then regulate the way it is spliced. To pursue the theory that Clk/Sty regulates transcription, I would first establish that Clk/Sty does interact with c-Myc, ZFP38, N-CoR, or the novel zinc finger *in vivo* using the four criteria described above.

Most of the other clones that interact with Clk/Sty have no assigned function. In the case of CARS-Cyp, the domain structure is intriguing. There are several reasons why Clk/Sty and CARS-Cyp may interact *in vivo*. First, CARS-Cyp may transport Clk/Sty to nuclear speckles. Second, Clk/Sty may be a substrate for the cyclophilin domain of CARS-Cyp. However, there is no evidence to suggest that Clk/Sty requires prolyl isomerization to function. Lastly, CARS-Cyp may be a substrate for Clk/Sty. This last point is probable as CARS-Cyp contains multiple RSRS repeats that correspond to Clk/Sty phosphorylation sites. Phosphorylation by Clk/Sty may increase the ability of CARS-Cyp to interact with other RS domain-containing proteins.
Specificity of SR protein phosphorylation

In vitro, I demonstrated that Clk/Sty generated more phosphorylated ASF/SF2 than phosphorylated MBP or H1 histone. From these results, I concluded that ASF/SF2 is a preferred substrate of Clk/Sty. The true test of specificity lies in the cell, where Clk/Sty is confronted with multiple substrates, and chooses its desired substrate. Normally, a kinase displays high catalytic efficiency (Kcat/Km) towards a preferred substrate. For ASF/SF2, the Km value for phosphorylation by Clk/Sty is 400 nM. The Kcat was not determined.

I tested two peptides, derived from sequences within ASF/SF2, as substrates for Clk/Sty: RSRS (GRSRSRSRSR) and SPRY (RGSPRYSPRHS). Both of these peptides displayed a similar Km value for phosphorylation by Clk/Sty. However, a comparison of relative catalytic efficiency indicated that RSRS is a better substrate. In this case, RSRS had a higher Vmax value, which means that RSRS is turned over faster than SPRY. This result emphasizes the importance of describing a kinase-substrate interaction by both Kcat and Km values and not just Km values.

Besides the specificity of Clk/Sty for SR proteins, one can consider the specificity of SR proteins for Clk/Sty. In other words, do several kinases phosphorylate SR proteins? In chapter 2, I showed that Clk/Sty, cAPK, and PKC phosphorylate ASF/SF2 in vitro. Besides these three kinases, SRPK1, a U1 70K kinase, LBR kinase, and DNA topoisomerase I also phosphorylate RS domain-containing proteins (Gui et al., 1994a; Woppmann et al., 1993; Nikolakaki et al., 1996; Rossi et al., 1996). In order to ascertain if these are SR protein kinases in vivo, the four criteria described above must be met. For SRPK1, these criteria have been met to some degree. SRPK1 phosphorylates several SR proteins thus it must interact with them in vitro (Gui et al., 1994a). A portion of SRPK1 co-localizes with SR proteins in speckles (X.-D. Fu, unpublished...
results). SRPK1 disassembles nuclear speckles when added to permeabilized cells, and also inhibits splicing in vitro. As well, in chapter three I show that SRPK1 phosphorylates ASF/SF2 in vitro on sites that are also phosphorylated in vivo. The remaining putative SR protein kinases have not been tested as SR protein kinases using the criteria that I have set out.

When a kinase has been established as an SR protein kinase, differences between the kinases should be determined. The kinases may be expressed in different cells, or active at different times in the cell cycle. In addition, they may phosphorylate different sites on SR proteins and, therefore, may influence SR protein activity in distinct ways. For SRPK1 and Clk/Sty, the pattern of phosphorylation of ASF/SF2 overlaps. However, Clk/Sty phosphorylates more sites than SRPK1, and this may prove to be important for ASF/SF2 function.

I also tested the RSRS and SPRY peptides to determine if they were substrates for SRPK1. One of the initial reasons for testing these peptides was to determine if SRPK1 or Clk/Sty are proline- or arginine-directed kinases. Based on the sequence of the P+1 loop of both kinases, it was predicted that arginine could not be accommodated in the P+1 binding pocket (Songyang et al., 1994). In contrast to Clk/Sty, SRPK1 could phosphorylate only the RSRS peptide and not the SPRY peptide. Since both of these kinases phosphorylate serines with an arginine at the P+1 position, then the P+1 binding pocket of these kinases must differ from those of ERK2 and CDK2.

Two main differences between Clk/Sty and SRPK1 are apparent. First, Clk/Sty is a better kinase for SPR and KSKS repeats than SRPK1. Therefore, Clk/Sty recognizes a broader substrate consensus motif than does SRPK1 that allows Clk/Sty to phosphorylate more sites within ASF/SF2. Thus, SRPK1, and not Clk/Sty, appears restricted to RS domain-containing
Second, SRPK1 generates more phosphorylated ASF/SF2 than Clk/Sty. Since the two kinases have similar affinity for ASF/SF2, the difference must lie within their K_cat values. Whether SRPK1 is more active than Clk/Sty *in vivo* has not been determined.

**The functional role of Clk/Sty dual specific kinase activity**

Clk/Sty phosphorylates ASF/SF2 on serine and to a lesser degree on threonine. Perhaps, Clk/Sty phosphorylates some of the other two-hybrid clones on tyrosine. If Clk/Sty does not phosphorylate exogenous substrates on tyrosine, then autophosphorylation on tyrosine may be important for Clk/Sty activity or for interaction of Clk/Sty with other proteins. Interestingly, clone 10.8 did not interact with a kinase inactive Clk/Sty that presumably is unphosphorylated. It is worth determining if the interaction of clone 10.8 with Clk/Sty requires Clk/Sty to be phosphorylated on tyrosine or serine/threonine residues.

**Future directions**

The study of Clk/Sty is still at a preliminary stage. My suggestions for future directions for the Clk/Sty project are in two parts. First, I would propose to continue the analysis of the two-hybrid clones. Second, antibodies could be used to study the function of endogenous Clk/Sty in mammalian cells. A key point that may be addressed by both methods is the identities of possible upstream regulators of Clk/Sty.

**Two-hybrid clones**

To complete characterization of clones identified in the Clk/Sty two-hybrid screen, the two-hybrid clones should be sequenced in their entirety. For clones that appear promising (i.e., possible nuclear localization, RS domains, kinase domains), the full-length cDNA should be obtained. The next step would involve determining which of these interesting clones meet the
four criteria that I have established to identify *in vivo* partners for Clk/Sty. As well, the clones from chapter 4 could be tested to determine if they meet the four criteria. Further study of c-Myc, ZFP38, and N-CoR will determine if Clk/Sty regulates transcription through these factors.

A second part of this analysis would be to continue the study of SR proteins. In chapters two and three, I started mapping phosphorylation sites on ASF/SF2. Defining the actual sites of phosphorylation may aid in assigning a consensus site for Clk/Sty phosphorylation. This consensus sequence could be used to look for other putative substrates for Clk/Sty. More importantly, once the sites are known, mutagenesis of these sites could be used to uncover the effect of Clk/Sty phosphorylation on ASF/SF2.

Throughout this thesis, I have hypothesized that phosphorylation increases the interaction of RS domain-containing proteins. This theory could be tested using bacterially expressed ASF/SF2, clone 6.9, and Clk/Sty. The interaction of ASF/SF2 with itself or with clone 6.9 could be examined plus or minus Clk/Sty phosphorylation.

**Endogenous Clk/Sty**

The second part of my proposed future directions would be to study the endogenous Clk/Sty kinase. Up to this point, I have been unable to study this kinase *in vivo* because I have lacked antibodies to this protein. Recently, I have collaborated with Xiang-Dong Fu and PharMingen to make a monoclonal antibody against Clk/Sty. Preliminary use of this antibody has demonstrated that endogenous Clk/Sty co-localizes with SC35 to nuclear speckles. With this antibody, several experiments could be performed. First, the type of cells that express Clk/Sty could be identified. Northern results suggest Clk/Sty is widely expressed. Protein-blot analysis would confirm this expression pattern. In humans, there are three Clk kinases. By the above
studies. It may be possible to differentiate the function of Clk/Sty from Clk2 and Clk3 by determining their expression and activation patterns.

Most importantly, I do not know what regulators lie upstream of Clk/Sty. Examination of endogenous Clk/Sty may reveal if it is activated at certain times during the cell cycle or upon external stimulation. These results may suggest potential upstream regulators. As well, co-immunoprecipitations could be used to look for associating proteins in vivo. It may be that some of the uncharacterized two-hybrid clones are upstream regulators. Further analysis of these clones, as suggested above, may uncover upstream regulators. If we can determine what regulates Clk/Sty, then we can, by extension, determine what regulates pre-mRNA splicing.
References


Bingham, P.M. and Judd, B.H. (1981) *Cell*, 25, 705-711. A copy of the copia transposable element is very tightly linked to the w allele at the white locus of *D. melanogaster*. 

148


Gotoh, Y., Moriyama, K., Matsuda, S., et al. (1991) EMBO J., 10, 2661-2668. Xenopus M phase MAP kinase: isolation of its cDNA and activation by MPF.


Wang, J. and Bell, L.R. (1994) *Genes Dev.*, 8, 2072-2085. The Sex-lethal amino acid terminus mediates cooperative interactions in RNA-binding and is essential for splicing regulation.


168


