

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600





Université d'Ottawa • University of Ottawa



Gene targeted disruption of murine Clk2 in ES cells

Alan Cheng

Thesis submitted to
the School of Graduate Studies and Research
in partial fulfilment of the requirements for the degree of
Master of Science

Department of Biochemistry
University of Ottawa
Ottawa, Ontario
CANADA



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

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.

0-612-28407-7

Abstract

The Clk (cdc2-like kinase) family of kinases has been well conserved throughout evolution and includes members from yeast to humans. In mammalian systems, four members have been identified: Clk1, Clk2, Clk3, and Clk4. Although Clk kinases have been implicated in the regulation of alternative splicing, their exact biological function is unknown. Our lab has used a gene targeted approach to ascertain loss-of-gene function of murine Clk kinases.

The work presented in this thesis describes the cloning and gene targeted disruption of murine Clk2 in J1 embryonic stem cells. Two lambda phage clones were isolated from a D3 genomic library, using full length human Clk2 cDNA as a probe. One clone is roughly 13 kb and spans the 5' end of the gene; the other spans the 3' end of the gene and is at least 20 kb in size. Partial sequence analysis reveals at least two coding exons in each clone, corresponding to amino acids 1-133 and 225-300 in human Clk2. Each exon-intron boundary possesses the consensus splice acceptor/donor sequences.

Employing primers derived from the genomic sequence, a 581 bp Clk2 partial cDNA fragment was cloned from P19 embryonal carcinoma cells by RT-PCR (reverse transcriptase coupled with polymerase chain reaction). Sequence analysis reveals that it shares 97% identity to human Clk2 at the amino acid level. When this partial cDNA was used to probe Northern blots of murine adult tissues and embryonic cell lines, at least 3 distinct transcripts were detected in the tissues, whereas only one transcript was detected in J1 embryonic stem cells and P19 embryonal carcinoma cells.

Using DNA fragments isolated from the genomic clones, a promoter trap targeting vector was constructed. The 5' arm of homology is 2.4 kb in length while the 3' arm of homology is 1.5 kb in length; a homologous recombination event should result in the deletion of a region of the gene that includes 149 bp of exon corresponding to amino acids 8-56 of Clk2. Transfection of the targeting vector into J1 embryonic stem cells and subsequent screening led to the isolation of a targeted ES clone, denoted Clk2 6B.

The generation of Clk2 targeted embryonic stem cells, provides a resource for the biochemical analysis of Clk2 as well as the means for the future generation of Clk2 deficient mice to determine the consequences of its loss of function in a mouse model.

Dedication

I would like to dedicate this thesis to my parents and my sister Jenny, for their devotion to my life.

- iv -

I know someday you'll have a beautiful life, I know you'll be a star

In somebody else's sky, but why

Why, why can't it be, why can't it be mine?

Eddie Vedder (Pearl Jam)

Acknowledgements

During my quest to become a professional scientist and a decent human being, I have encountered countless souls that have guided me in the right direction. I would especially like to thank my supervisor, Dr. John Bell, who took a great risk when he accepted me into his lab. His vote of confidence and unorthodox motivational techniques were an inspiration to me in the lab.

I am also indebted to Dr. Douglas Gray for his advice and help on more than one occasion. One day I will repay him by treating him to a Tragically Hip concert.

My sincere gratitude extends to all my friends who have supported me during my years at the Cancer Research Group. Unfortunately, that is a lengthy list. However, I would especially like to thank:

Pete for his unending optimism towards me, and his unending cynicism towards himself. His continual guidance, humour, and pranks are always appreciated.

Ninan, for being a model of integrity and thoroughness. Fundamentally, he's essentially and basically a natural scientist.

JosEe, for being the best benchmate available; thanks for the coffee, the sisterly advice, the Sephadex, the lysis buffer...

Dave, for his support, one way exchange of ideas, and tequila. An unselfish friend with whom I've had the privilege to work and drink (and almost sing) with.

Mike, for his moral support and intellectual conversations of a philosophical nature. Thanks for always sticking up for me and for watching my back.

Helen, for being a good listener to my constant whining. Oddly enough, she never remembers what I whine about.

Ricardo for providing interesting conversations concerning sports. None of my other friends will argue for the Bruins or the Redskins.

And finally, special thanks to:

Sheryl, for being understanding, kind, and helpful. Her willingness to help others is unparalleled. Her companionship has given me wonderful and happy moments to cherish.

Debbie, for her emotional support during the dark ages. Her friendship provided me with great times to remember that probably prevented me from going insane.

Table of Contents

Abstract	i
Dedication	iii
Acknowledgements	v
List of Abbreviations	xi
1. Introduction	- 1 -
1.1 mRNA Splicing	- 2 -
1.2 Alternative Splicing	- 5 -
1.3 Drosophila Sex Determination	- 6 -
1.4 SR Proteins	- 7 -
1.5 Phosphorylation of SR Proteins Regulates their Activity and Subnuclear Localization	- 9 -
1.6 Clk1 Kinase	- 11 -
1.7 Clk Family of Kinases	- 16 -
1.8 Area of Investigation:	- 18 -
2. Materials and Methods	- 19 -
2.1 Isolation of genomic clones:	- 19 -
2.2. Isolation of Phage DNA	- 20 -

2.3 Slot Blot Analysis of Phage Clones:	- 21 -
2.4 Subcloning Restriction Fragments from Phage Clones:	- 21 -
2.5 Sequencing of Subclones:	- 22 -
2.6 Isolation of a Partial cDNA Clone	- 22 -
2.7 RNA isolation and Northern blot analysis:	- 23 -
2.8 Construction of Gene Targeting Vectors / Knockout Strategy	- 24 -
2.9. J1 ES Cell Culture	- 25 -
2.10 Electroporation of J1 ES Cells and Isolation of Clones	- 25 -
2.11. Genomic DNA Isolation and Southern Blot Analysis:	- 26 -
2.12 Preparation of Feeder Layers from Embryonic Fibroblasts	- 27 -
3. Results	- 28 -
3.1 Delineation of the genomic structure of murine <i>Clk2</i>	- 28 -
3.2 Comparison of murine and human <i>Clk2</i> cDNAs	- 40 -
3.3 Expression of <i>Clk2</i> transcripts	- 43 -
3.4 Isolation of <i>Clk2</i> single knockout clones	- 43 -

4. Discussion - 49 -

5. Conclusion - 54 -

References - 56 -

List of Figures

Figure 1: Spliceosome dependent pre-mRNA splicing. 4

Figure 2: Alternative Splicing of Clk1 pre-mRNA generates two protein isoforms 14

Figure 3: Schematic Representation of Clk1 Protein Isoforms 15

Figure 4: Slot Blot Analysis of genomic clones C216 and H26 29

Figure 5: Southern Blot Analysis of Restriction Fragments of Clone C216 30

Figure 6: Southern Blot Analysis of Restriction Fragments of Clone H26 31

Figure 7: Partial Sequence of Subclone pH26E1 33

Figure 8: Partial Sequence of Subclone pC216H1 34

Figure 9: Partial Sequence of Subclone pC216X1 35

Figure 10: Alignment of lambda clone C216 with subclones pC216H1
and pC216X1 37

Figure 11: Subcloning Fragments Upstream of the First Coding Exon 38

Figure 12: Alignment of the 3 subclones relative to C216 39

Figure 13: Sequence of the 581 bp partial Clk2 cDNA 41

Figure 14: Partial Alignment of Human and Mouse Clk2 Proteins 42

Figure 15: Northern blot analysis of Clk2 expression in adult mouse tissues 44

Figure 16: Northern blot analysis of Clk2 in ES and EC cells 45

Figure 17: Schematic of the Clk2 Region Targeted. 47

Figure 18: Identification of a Clk2 targeted clone 48

List of Abbreviations

<i>AFC</i>	<i>Arabidopsis thaliana fus3</i> -complementing gene
ASF	Alternative splicing factor
bp	Base pair
cDNA	Complementary DNA
Clk	CDC28/ <i>cdc2</i> +-like kinase
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
<i>Doa</i>	<i>Darkener of apricot</i>
EDTA	Ethylenediaminetetraacetic acid
GST	Glutathione-S-transferase
hr	Hour
hClk	Human Clk
hnRNP	Heterogenous nuclear RNP
kb	kilobase pair
kDa	kiloDalton
<i>KNS1</i>	<i>Kinase Next to Spa2</i>
mAb	Monoclonal antibody
mClk	Murine Clk
MEM	Minimal essential medium
mRNA	Messenger RNA

NLS	Nuclear localization signal
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pre-mRNA	Precursor mRNA
RBD	RNA binding domain
RNA	Ribonucleic acid
RNP	Ribonucleoprotein particle
RRM	RNA recognition motif
RS domain	Arginine/serine-rich domain
RT-PCR	Reverse transcriptase coupled with PCR
SDS	Sodium dodecyl sulfate
SF2	Splicing factor 2
snRNP	Small nuclear ribonucleoprotein particle
SR protein	Serine/arginine-rich protein
Sty kinase	Serine, threonine and tyrosine kinase
U	Unit
U2AF	U2 snRNP auxiliary factor
UTR	Untranslated region

1. Introduction

Alternative splicing plays an important role in the regulation of eukaryotic gene expression at the post-transcriptional level. Stringent control is necessary to ensure the proper functioning of the cell, and in some cases, the growth and development of multicellular organisms.

Recently, serine/arginine-rich (SR) proteins have been identified to be crucial for both constitutive and alternative splicing (for review, see Valcarcel and Green, 1996). In addition, recent evidence suggests that phosphorylation of SR proteins modulates their activity (Colwill *et al.*, 1996; Tacke *et al.*, 1997). To date, several mammalian kinases that phosphorylate SR proteins have been cloned, a subgroup of which belong to the Clk family of kinases. Emerging evidence now suggests that Clk kinases regulate alternative splicing through phosphorylation of SR proteins. Members of the Clk family of kinases have been identified in organisms ranging from yeast to humans, and their conservation throughout evolution suggests that they confer a selective advantage to the organism.

The work in this thesis concerns itself with the cloning and characterization of murine Clk2, a family member of the Clk kinases. The objective was to generate Clk2 targeted embryonic stem cells for the future production of mice deficient in Clk2, to determine the consequences of its loss of function.

A review of several aspects of mRNA splicing will be provided, including recent models of the mechanism and the role of SR proteins. This is followed by a summary of the evidence supporting the role of mammalian Clk kinases in alternative splicing.

Finally, the last part of the introduction will review the information known about all the Clk kinases identified to date.

1.1 mRNA Splicing

In eukaryotes, the primary RNA transcript produced in the nucleus is extensively processed prior to its transport to the cytoplasm, where it is used to program the translation machinery. Typically, processing includes capping at the 5' end, polyadenylation at the 3' end, and splicing of the RNA transcript.

Capping of the 5' end is achieved through a covalent modification consisting of a 7-methyl guanosine residue linked to the 5' end of the transcript by a triphosphate bond. This 5' cap is an important part in the initiation of protein synthesis, and may also protect the growing RNA transcript from degradation (for review, see Lewis and Izaurralde, 1997). At the 3' end, stretches of 150-200 adenosine residues are added; the exact function of these "poly (A)" tails is unclear, but it has been suggested that they may play a role in protection against RNAses and/or aid in its export from the nucleus (Huang and Carmichael, 1996). Capping and poly (A) addition to the primary transcript results in an RNA molecule often referred to as the precursor mRNA (pre-mRNA).

In eukaryotes, the DNA segments that code for the structure of the protein (exons) are interrupted by intervening, non-coding sequences (introns). Before the pre-mRNA transcript is exported from the nucleus to be translated by the ribosomal machinery, splicing reactions are required to remove the introns and bring together the exons, thus forming mature messenger RNAs (mRNAs). Splicing is a highly complex process, which

requires the locating of exons and their precise ligation together in the correct order and without the loss of the correct reading frame. Significant progress has been achieved in understanding the mechanisms behind splicing reactions. The cellular machinery which catalyzes splicing reactions has been termed the spliceosome; this is composed of five small nuclear ribonucleoprotein (snRNP) complexes and non-snRNP proteins. The snRNP complexes include U1, U2, U4, U5, and U6. Splicing also involves four basic sequence elements on the pre-mRNA (Figure 1): 5' and 3' splice sites which define exon-intron boundaries; the branch point adenosine residue that forms a 2' -5' bond with the excised 5' end of the intron; and a polypyrimidine tract that binds to the non-snRNP splicing factor U2AF⁶⁵ (U2 auxiliary factor, 65kDa subunit).

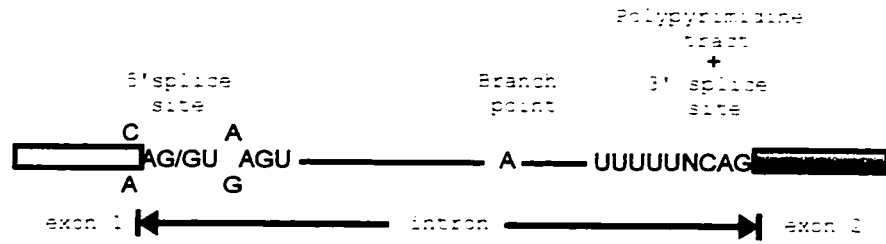
Initial events that lead to spliceosome assembly include 5' and 3' splice site recognition by the U1 snRNP and U2AF⁶⁵ respectively (Figure 1). This forms the "commitment complex", originally defined in yeast pre-mRNA splicing studies, where *in vitro* processing of pre-mRNA assembled in a pre-spliceosomal complex was found to be resistant to an excess of competitor RNA (Seraphin and Robash, 1989). U2AF⁶⁵ binding may then recruit the U2 snRNP to bind to the branch point near the 3' splice site (Valcarcel *et al.*, 1996). Subsequently, the intron is folded such that U1 associates with U2AF⁶⁵, probably via bridging proteins (Wu and Maniatis, 1993). The formation of a [U4/U6]U5 snRNP triplet then associates with this pre-spliceosome complex; U6 binding to the 5' splice site destabilizes U1 binding, U4 is released, and the spliceosome core is now catalytically active to excise the intron. Both spliced mRNA and intron are released and disassembly of the spliceosome allows recycling of splicing factors.

Figure 1: Spliceosome dependent pre-mRNA splicing.

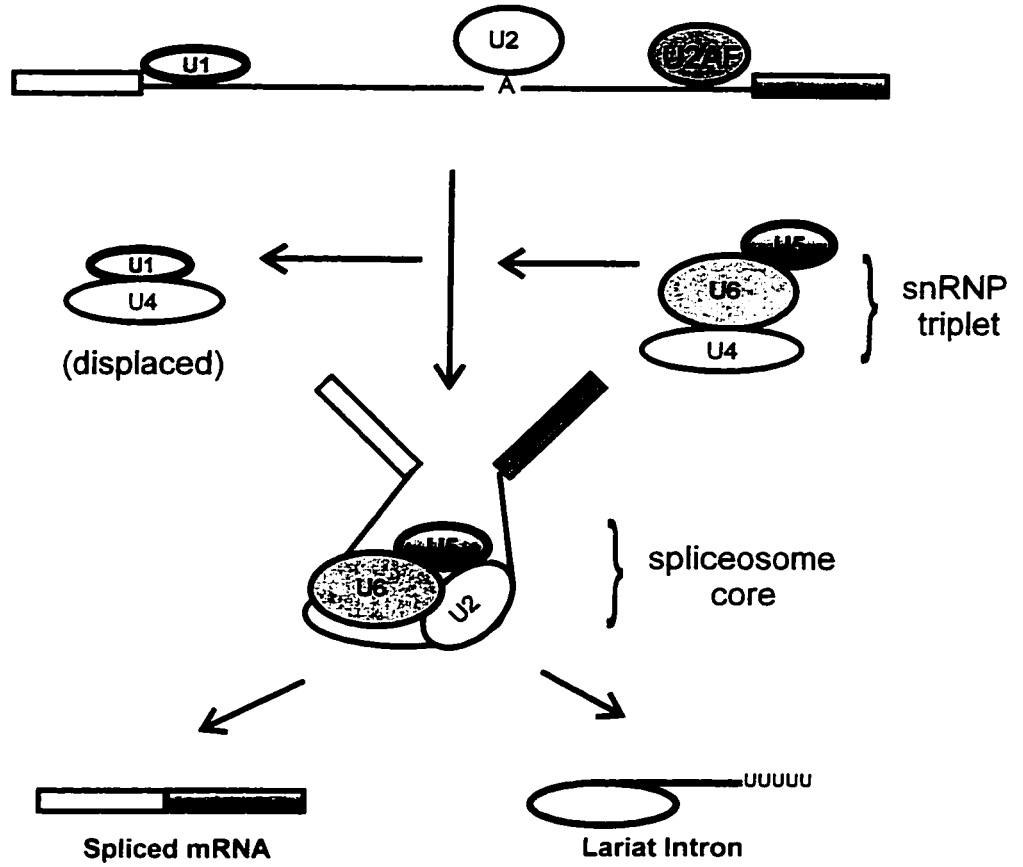
(A) Sequence elements crucial for splicing includes 5' and 3' splice sites, polypyrimidine tract, and the invariant branch point adenosine residue.

(B) Splicing involves the assembly of multiple proteins onto the pre-mRNA. Initial events include the recognition of the 5' and 3' splice sites by U1 snRNP and U2AF respectively. U2AF is thought to recruit U2 snRNP to bind to the branch point A residue. Formation of catalytic spliceosomes results from the involvement of the U4/U6-U5 tri-snRNP; exons are ligated together via transesterification reactions, and the lariat intron is removed in the process.

A



B



1.2 Alternative Splicing

Within a pre-mRNA, multiple combinations of 5' and 3' splice site usage exists. Alternative splicing describes the process of selecting different combinations of splice sites to generate multiple mRNA transcripts from a single pre-mRNA. Alternative splicing allows the generation of isoform diversity, to produce proteins of similar or radically different functions, and is a major mechanism involved in regulating gene expression.

Sometimes the number of alternative mRNAs is limited, as in the case of *bcl-x*, which produces mRNAs to generate only two isoforms, Bcl-xL and Bcl-xS (Boise et al., 1993); interestingly, though, they seem to have opposing effects when transfected into cells. Bcl-xL enhances cell survival in an IL-3 (interleukin-3) dependent cell line upon growth factor removal; Bcl-xS, however, opposes this effect.

Patterns of alternative splicing can also be very complex, as in the case for the fibronectin gene. A maximum of twenty alternatively spliced variants is possible in humans, and some forms are tissue, cell type, or developmentally specific (for review, see Kornblihtt *et al.*, 1996). Although the overall structure of each protein is similar, the cell type specific exons may code for domains that interact with cell-type specific regulators.

Several developmental processes have been shown to be regulated by alternative splicing. In fact, approximately 15% of genetic diseases are a consequence of defects in splicing (Krawczak et al., 1992). Common examples include Wilms' tumour (Gunning *et al.*, 1996; Simms *et al.*, 1995) and Alzheimer's disease (Hartmann *et al.*, 1996). However, probably the most well-understood and striking example involved in development is the

Drosophila sex-determination pathway.

1.3 *Drosophila* Sex Determination

In *Drosophila*, sexual differentiation is controlled by a cascade of changes in gene expression that is regulated by splicing events (for review, see Burtis, 1993). Each of the genes that determine whether a fly is male or female is transcribed into RNA in cells of both sexes, but RNA is spliced differently in males and females.

During early embryogenesis, the primary sex determining signal is the X:A (X chromosome to autosome ratio); this ratio is measured by a mechanism involving the relative concentrations of proteins that are encoded by X-linked and autosomal genes. Most of these proteins belong to the class of basic helix-loop-helix (bHLH) proteins (Parkhurst et al., 1990), and their ratio in females only, allows the formation of transcriptional complexes to activate the early promoter of Sex-lethal (*Sxl*), the master switch in the sex determination cascade. However, soon afterwards, the early promoter shuts off and transcription is resumed off the late promoter, which is active in both sexes.

In males, exon 3 in the *Sxl* pre-mRNA is included during splicing; as exon 3 contains an in-frame stop codon, no functional Sxl is produced. In females, however, exon 3 is spliced out to produce functional Sxl. Skipping of exon 3 in females is dependent on the initial presence of Sxl produced by the early promoter. Once activated, the Sxl product controls the splicing of its own pre-mRNA to generate a positive autoregulatory loop. In males, however, the absence of pre-existing Sxl results in exon 3 inclusion that produces

non-functional Sxl, and hence no autoregulatory loop is initiated.

Sxl then regulates alternative splicing of transformer (*tra*), the next gene in the cascade. Exon 2 of *tra*, which also contains an in-frame stop codon, is alternatively spliced in the two sexes. In females, Sxl competes out U2AF binding to the 3' splice site of exon 2 and thus causes exon skipping (Valcarcel et al., 1993). Tra, in combination with other proteins, causes a female specific splicing of doublesex (*dsx*), by promoting exon 4 inclusion in *dsx* pre-mRNA splicing (Hoshijima et al., 1991). In males, however, without *tra*, *dsx* pre-mRNA follows a default splicing pathway that skips over exon 4 (Hoshijima et al., 1991). Although, male and female *dsx* proteins act as transcriptional regulators on the same target genes, they have opposing effects, thus producing sex-specific transcriptional regulation (Coschigano and Wensink, 1993).

Since alternative splicing programs can be cell, tissue, or developmentally specific, the mechanism of regulation must be highly stringent. An increasing amount of evidence is now supporting the role of a family of proteins (SR proteins) in splicing regulation.

1.4 SR Proteins

In addition to the snRNPs required for spliceosome assembly, a growing number of non-snRNP proteins have been identified to be essential for splicing to occur. A prominent number of these proteins possess arginine-serine rich (RS) domains, and now have been classified into a family known as SR (serine/arginine) proteins (for review, see Valcarcel and Green, 1996). All SR proteins are structurally similar, possessing at least

one RNA recognition motif (RRM, Burd and Dreyfuss, 1994) and a RS domain. A variety of other splicing factors and regulators also contain RRM and/or RS domains. These include the 70 kD protein component of U1 snRNP (U1 snRNP 70K), U2AF, and the *Drosophila* regulators of sex determination. However, the term SR protein is often restricted to those proteins which are recognized by monoclonal antibody 104 (Roth *et al.*, 1990), and can restore splicing activity to cytoplasmic S100 extracts (Zahler *et al.*, 1992), which contain all splicing components except SR proteins (Krainer *et al.*, 1990).

At least eight human SR proteins have been cloned and sequenced: SRp20, 9G8, ASF/SF2, SC35, SRp30c, SRp40, SRp55, and SRp75. SR proteins are involved during early spliceosome assembly and are found in the commitment complex. Yeast two-hybrid experiments have shown that SR proteins can interact with U1 snRNP and U2AF through their RS domains (Wu and Maniatis, 1993), suggesting that SR proteins may act as a bridge to link both splice sites for subsequent intron removal.

The first SR protein identified was SF2 (Splicing Factor 2); it was originally isolated as a splicing factor in fractionation experiments that was necessary for pre-mRNA splicing *in vitro* (Krainer and Maniatis, 1985). Independently, Ge and Manley (1990) isolated a factor called ASF (Alternative Splicing Factor) from 293 cell extracts that was able to control the alternative splicing of SV40 early pre-mRNA *in vitro*, and subsequent cloning of ASF revealed that it was SF2 (Ge *et al.*, 1991). Furthermore, ASF/SF2 was later shown to, by overexpression studies, to activate 5' proximal splice sites in a number of reporter genes *in vivo* (Caceres *et al.*, 1994). Thus, it seemed that ASF/SF2 was a splicing factor involved in both constitutive and alternative splicing. Since then, other

SR proteins have also been shown to modulate 5' splice site selection; SC35, like ASF/SF2, promotes the 5' proximal splice site (Fu *et al.*, 1992), whereas SRp40 and SRp55 promote the use of distal sites (Zahler *et al.*, 1995).

It is also known now that SR proteins can regulate splicing in conjunction with sequences on the pre-mRNA known as exon enhancers. Sun *et al.* (1993) showed that ASF/SF2 directed alternative splicing of bovine growth hormone through the binding to a purine rich sequence in the last exon. SC35, in contrast, did not bind this enhancer. Thus, sequence specific exon enhancers might exist for individual SR proteins.

1.5 Phosphorylation of SR Proteins Regulates their Activity and Subnuclear Localization

Recent evidence suggests that phosphorylation and dephosphorylation cycles play crucial roles in splicing reactions. Splicing is thought to occur in two steps, 5' splice site cleavage, followed by 3' splice site cleavage along with exon ligation. Studies using inhibitors demonstrate that PP1 (Protein phosphatase 1) and PP2A (Protein phosphatase 2A) like activities are involved in both steps of splicing (Mermoud *et al.*, 1992; Tazi *et al.*, 1992). In contrast, spliceosome assembly requires phosphorylation steps to proceed. Spliceosome assembly has been shown to be blocked by PP1 at an early step (Mermoud *et al.*, 1994). Moreover, the addition of phosphorylated SR proteins overcomes this block, suggesting that SR protein phosphorylation regulates spliceosome assembly. An SR protein kinase activity was first identified by Wopmann *et al.* (1993), and was shown to

phosphorylate the serines in the RS domains of ASF/SF2 and the 70K protein component of U1 snRNP. Currently, four proteins have been cloned and shown to phosphorylate SR proteins: SRPK1, Clk1, DNA topoisomerase I, and Prp4.

SRPK1 (SR-protein-specific kinase) was first isolated from HeLa cell extracts based upon its ability to phosphorylate SC35 (Gui *et al.*, 1994), and has also been shown to phosphorylate ASF/SF2 *in vitro*. In addition, purified SRPK1 activity was able to inhibit splicing, *in vitro*, of human β -globin pre-mRNA in a dose-dependent manner. Cloning of SRPK1 from a HeLa cell cDNA library revealed 30% identity with fission yeast kinase Dsk1, which is believed to be involved in chromosome segregation during mitosis (Takeuchi and Yanagida, 1993).

Clk1 (cdc-like kinase, see section 1.5) was identified to be an SR protein kinase when it was used to isolate several members of the SR protein family in a yeast-two hybrid assay (Colwill *et al.*, 1996). In addition, bacterially expressed Clk1 phosphorylated recombinant ASF/SF2 on serine residues. SRPK1 and Clk1 are 32% identical across their catalytic domain, suggesting that they may represent members of a family of kinases involved in phosphorylating SR proteins.

Rossi *et al.* (1996) also isolated an SR kinase activity in HeLa cell nuclear extracts which was revealed to be DNA topoisomerase I upon peptide sequencing. In addition, specific inhibitors of DNA topoisomerase I were able to reduce its kinase activity for SR proteins. However, a role for DNA topoisomerase I in splicing has yet to be shown.

Most recently, mammalian homologs for the fission yeast Prp4 (pre-mRNA processing) protein kinase was cloned and shown to phosphorylate ASF/SF2 *in vitro* in

its RS domain (Gross et al., 1997). The fission yeast *prp4* gene was isolated in a temperature-sensitive genetic screen of mutants that accumulated pre-mRNA at the restrictive temperature (Rosenberg et al., 1991). However, the involvement of the mammalian homologs in pre-mRNA splicing has yet to be shown.

Immunofluorescence studies have shown that SR proteins concentrate in speckles within the nucleus (Spector, 1993). It has been proposed that these nuclear speckles act as storage sites for splicing factors and that these factors are released upon demand (Jimenez-Garcia and Spector, 1993). Thus it would seem that subnuclear localization, as well as phosphorylation state, regulates the activity of SR proteins, and hence pre-mRNA splicing.

Of the above SR protein kinases, only SRPK1 and Clk1 have been shown to affect subnuclear localization of SR proteins. When added exogenously or over expressed, both kinases disperse SR proteins from a speckled pattern into a diffuse subnuclear distribution (Gui et al., 1994; Colwill et al., 1996).

1.6 Clk1 Kinase

Murine Clk1/Sty was previously isolated through an anti-phosphotyrosine antibody screen of P19 embryonal carcinoma cell and mouse erythroleukemia cDNA expression libraries (Howell et al., 1991; Ben-David et al., 1991). Although the initial attempt was to identify novel mammalian protein tyrosine kinases, analysis of the amino acid sequence showed that Clk1 resembled serine/threonine kinases. When expressed in bacteria as *trpE*

fusion proteins, Clk1 demonstrated autophosphorylation on all three phosphoamino acids (Howell et al., 1991). This dual specificity for both Ser/Thr and Tyr residues was confirmed when Clk1 was shown to autophosphorylate on all three residues *in vivo* (Duncan et al., 1995).

A single exon in the primary Clk1 mRNA is alternatively spliced to give rise to two transcripts (Duncan et al., 1995). As shown in figure 2, inclusion of this exon results in a transcript to code for a full length protein that has a putative N-terminal regulatory domain and a C-terminal kinase domain. Exclusion of the alternatively spliced exon results in a frame shift as well as a termination codon, producing a truncated protein lacking the kinase domain (Clk1^T).

Full length Clk1 is depicted in figure 3. The protein can be thought to be divided into three regions: (1) an amino terminal portion that contains a putative nuclear localization signal, and an RS domain; (2) the domain encoded by the alternatively spliced exon; and (3) the catalytic domain.

The first clue for an *in vivo* role of Clk1 came from studies using PC12 cell lines transfected with an inducible Clk1 expression system (Myers et al., 1994). Expression of Clk1 resulted in neurite outgrowth in the cells, and this effect was synergistically enhanced by treatment with NGF (Nerve Growth Factor).

Colwill et al. (1995) attempted to identify *in vivo* substrates for Clk1, by employing a yeast two-hybrid screen (Fields and Song, 1989). Clk1 was found to interact with several RNA binding proteins, most notably three members of the SR protein family: X16/SRp20, ASF/SF2, and a clone with 95% identity to human SRp55. Clk1 was also

able to phosphorylate ASF *in vitro*, and tryptic phosphopeptide mapping revealed that these phosphorylated sites overlapped with those phosphorylated *in vivo* (Colwill et al., 1995).

In another yeast two-hybrid screen, Clk1 was shown to interact with a cyclophilin protein containing RS repeats (Nestel et al., 1996). This Clk associating RS-cyclophilin (CARS-Cyp) possesses 39% homology to the natural killer tumour recognition protein-1 (NK-TR1, Anderson *et al.*, 1993), contains a large RS domain, and two Nopp140 domains. The biological significance of this finding has yet to be elucidated.

Most importantly, Clk1 proteins have been shown *in vivo*, to regulate the splicing of at least two pre-mRNAs: one derived from an adenovirus E1A minigene, and the other derived from a Clk1 minigene (Duncan *et al.*, 1997). The model presented suggests that SR proteins are phosphorylated by Clk1, and that the phosphorylation state determines the splice site selection. In addition, it is hypothesized that Clk1^T opposes full length Clk1, possibly through inactivation by oligomerization with Clk1. Whether Clk1 affects general global splicing or a specific subset of genes remains to be determined.

Figure 2: Alternative Splicing of Clk1 pre-mRNA generates two protein isoforms

For simplicity, Clk1 pre-mRNA is represented by three exons (boxes) separated by two introns (lines). Alternative splicing of the second exon generates two mRNA species, one which encodes for full-length protein (Clk1); the other results in a change in the open reading frame, as well as a non-sense mutation, giving rise to truncated protein (Clk1^T).

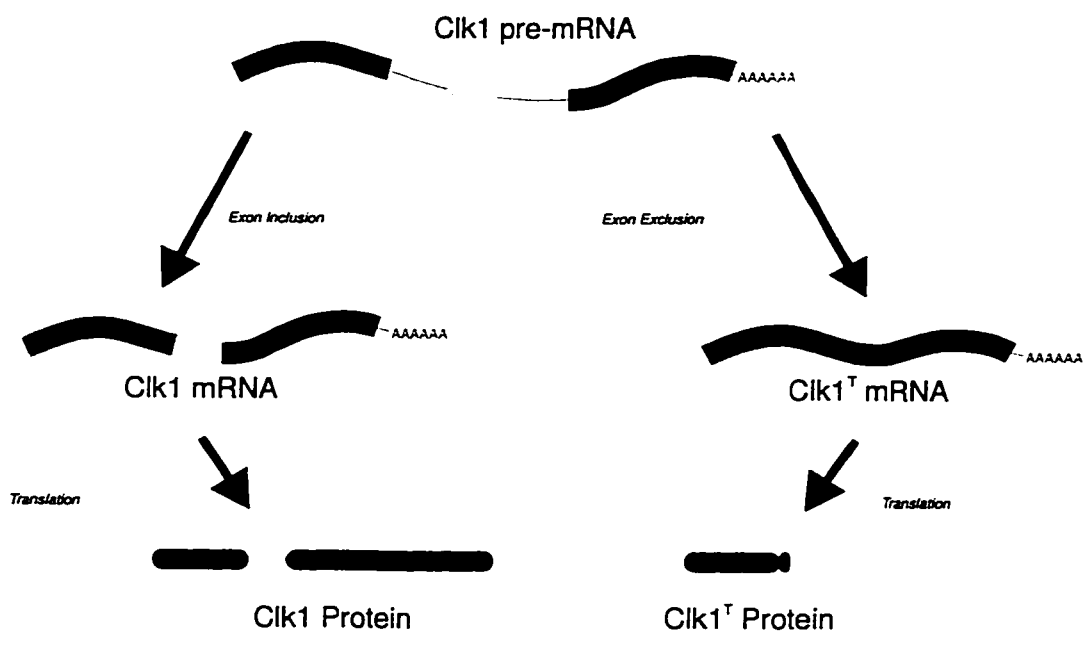


Figure 3: Schematic Representation of Clk1 Protein Isoforms

Full length and truncated isoforms of Clk1 are shown in this schematic. Full length Clk1 is roughly divided into three regions. The N-terminal regulatory domain is rich in serine/arginine residues, as well as having a nuclear localization signal. The C-terminal domain represents the kinase domain that resembles serine/threonine kinases, yet Clk1 possesses tyrosine kinase activity as well. The N-terminal and C-terminal domains is separated by the region encoded by the alternatively spliced exon.

1.7 Clk Family of Kinases

To date, several kinases have been identified which show sequence similarity to Clk1. These kinases have been grouped together in a subfamily known as "LAMMER" kinases based on a common amino acid motif found in subdomain X. Normally, this region is variable among other kinases, and thus may represent a novel motif involved in directing substrate specificity.

While studying the genomic organization of the *SPA2* gene in *Saccharomyces cerevisiae*, sequencing of the 3' flanking region revealed an open reading frame that codes for the putative kinase, KNS1 for *Kinase Next to Spa2* (Padmanabha et al., 1991). Deletion mutants derived by homologous recombination exhibited no differences compared to wild type cells in terms of growth rates and mating efficiencies. To date, no biological function has been assigned to KNS1. It may code for a non-essential protein, or there may be other genes with overlapping functions.

AFC1 (*Arabidopsis fus3*-complementing cDNA 1) was isolated from *Arabidopsis thaliana* in a genetic screen used to identify suppressors of *fus3 kss1* mating defects in *S. cerevisiae* (Bender and Fink, 1994). *FUS3* is known to activate *STE12*, a transcription factor required for the expression of mating-specific genes. *AFC1* complemented *FUS3* defects by restoring *STE12* dependent functions. Using degenerate primers to conserved regions in *AFC1* and its homologs, two related genes, *AFC2* and *AFC3*, were also identified. Functionally, *AFC2* does not activate *STE12* while for *AFC3*, it is unknown.

Using a differential display approach, Sessa et al. (1996) identified *PK12* from tobacco leaves as a protein kinase whose levels are induced upon stimulation with the

plant hormone ethylene.

Much of the genome of *C. elegans* has been sequenced, and sequence similarity searching has identified a hypothetical serine/threonine protein kinase. Our lab is currently in the process of isolating the gene for further analysis.

To date, the best characterized family member is *Doa* (*Darkener-of apricot*). The locus was first identified as a second site modifier of the *apricot* allele of the *white* locus in *Drosophila melanogaster* (Rabinow and Birchler, 1989). *Doa* mutations often result in recessive lethality at larval stages, during a time when an abundance of cell division and differentiation is occurring. *Doa* function seems to be critical for proper segmentation, development of the nervous system, and eye development (Yun et al., 1994). These findings strongly implicate *Doa* in multiple stages of development and differentiation.

In mammalian systems, several family members have been identified. Human Clk1 was isolated from a human T cell cDNA library using degenerate oligonucleotide probes to a conserved region of protein kinases related to p34^{cdc2} (Johnson and Smith, 1991). Human Clk2/PskG1 was also identified using this method (Hanks and Quinn, 1991). Hanes et al. (1994) independently isolated Clk2 and Clk3 from a human ovarian follicle cDNA library. Recently, rat Clk3 was isolated from a rat brain cDNA library (Becker et al., 1996). In murine systems, evidence supports the existence of at least 4 family members. Cloned genomic fragments suggest the existence of murine Clk2 and Clk3 (data not shown). A novel member, Clk4 was recently identified from linkage and genetic mapping of mouse chromosome 11 (Sally Camper, personal communication).

The evidence supporting the notion that Clk1 is implicated in alternative splicing

prompted our lab to investigate the possibility that other family members might also be involved. Indeed, it was recently shown that hClk2 and hClk3 were able to redistribute SR proteins in vivo (Duncan et al., submitted), as well as regulate the alternative splicing of reporter transcripts in a manner similar to mClk1 (Duncan et al., submitted).

Although implicated in alternative splicing, the biological functions of the Clk kinases have not yet been assigned. To address this question, Clk1 $-/-$ mice were generated by gene targeting experiments (Peter Duncan, personal communication). To date, no gross developmental defects have been identified in these mice. One possibility is that Clk1 may not be required for survival and that the null allele phenotype may be subtle. Yet there is the possibility that there is functional redundancy among the Clk family of kinases and that gross defects are not revealed until multiple family members have been disrupted for gene function.

1.8 Area of Investigation:

Although the Clk family of kinases have been implicated as regulators of alternative splicing, their biological roles have not yet been defined. *Clk1* $-/-$ mice generated by gene targeted disruption reveal no overt developmental defects, suggesting that either *Clk1* is a non-essential gene for viability, or that loss of Clk1 function is rescued by family members. The work of this thesis is directed at addressing these possibilities, through the cloning, characterization and gene targeting of murine Clk2.

2. Materials and Methods

2.1 Isolation of genomic clones:

Clk2 genomic clones were isolated from a D3 embryonic stem cell genomic library (a generous gift from Dr. D.A. Gray) that was cloned into the lambda DASH II vector (Stratagene) via the Bam HI site. Approximately 5×10^5 plaques were plated by infecting XL1 MRA bacterial strain (Stratagene) with bacteriophage and incubating at 37°C overnight. DNA from plaques were transferred to Hybond-N membranes (Amersham), heated to 100°C for 1 minute, and then UV cross-linked using the GS Gene Linker (Bio-Rad) at a setting of 100 mJ. Membranes were prehybridized for at least 2 hours at 42°C in hybridization buffer consisting of 6 X SSPE, 50% formamide, 5 X Denhardt's solution, 0.5% SDS, and 100 ug/ml sheared salmon sperm DNA. Hybridization was carried out with a ³²P-labelled hclk2 cDNA (a generous gift from Dr. K.H. Scheit) probe generated using the Random Primed DNA Labelling Kit (Boehringer Mannheim); hybridization was carried out at 42°C for 18-24 hours. Membranes were washed in 2 X SSC, 0.1% SDS at room temperature for 5 min, and once in 2 X SSC, 0.1% SDS at 50°C for 1 hour, then exposed to Reflection NEF-496 autoradiography film (NEN). For the positive clones, the plaque was isolated as an agarose plug and stored in 1 ml SM buffer (5.8g/L NaCl, 2g/L MgSO₄·7H₂O, 50 mM Tris-HCl pH 7.5) with 40 ul chloroform. Positive clones were purified by re-screening for 2-3 more rounds.

2.2. Isolation of Phage DNA

For the primary infection, 10 μ l of undiluted phage in SM buffer was pre-adsorbed with 100 μ l *E. coli* strain XL1 MRA for 10 minutes at 37°C. Next, 2ml of LBM broth (10g/L tryptone, 5g/L yeast extract, 10g/L NaCl, 2 mM MgSO₄·7H₂O) was added and incubated at 37°C until the infected culture became clear (4-6 hours). Primary lysate was obtained by mixing with 40 μ l chloroform, followed by centrifugation at 5000 x g for 5 min to remove cellular debris.

The secondary infection was performed by pre-adsorbing 15 μ l of the primary lysate to 1 ml of overnight XL1 MRA culture at 37°C for 10 min, followed by inoculation into 50 ml LBM broth and a 37°C overnight incubation.

To extract phage DNA, 2 ml of chloroform and 1.5 g of NaCl was added to the overnight culture, and swirled to dissolve the NaCl. Debris was pelleted by centrifuging at 4200 x g for 15 min. The phage solution was mixed with 5g of PEG (polyethylene glycol) and placed on ice for 1 hour. The phage was pelleted at 4200 x g for 30 min. at 4°C, resuspended in 0.6 ml SM buffer, and incubated with 25 μ g of DNase I for 1 hour at 37°C. DNase I was then inactivated by addition of 2.5 μ l DEPC (diethylpyrocarbonate), and the sample was mixed with 1.2 ml of a solution containing 1M Tris-HCl pH 8.5, 1% SDS, 0.1 M EDTA. After incubation at 65°C for 5 min., 0.6 ml 5M potassium acetate was added, and the sample was incubated on ice for 30 min. Debris was pelleted at 20800 x g for 15 min at 4°C, and the supernatant was mixed with 1 volume (about 1.4 ml) of isopropanol. DNA was pelleted at 20800 x g for 10 min., and resuspended in 400 μ l TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0). The sample was extracted once with

phenol/chloroform/isoamyl alcohol (25:24:1 v:v), and then DNA was precipitated in 2.5 volumes of ethanol. DNA was pelleted as above, and the pellet was washed once with 70% ethanol, followed by resuspension in 200 ul TE.

2.3 Slot Blot Analysis of Phage Clones:

Phage DNA was applied onto Hybond N membranes (Amersham) using a Minifold II Slot Blotter (Scheicher & Schuell). The membrane was immersed in denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 5 min, and then in neutralizing solution (1.5 M NaCl, 1 M Tris-HCl, pH 8.0) for 5 min. This was followed by UV crosslinking at a setting of 100 mJ using the GS Gene Linker (Bio-Rad). The membrane was then prehybridized (see section 2.1) and probed using various ³²P-labelled hclk2 cDNA restriction fragments. Membranes were washed in 2 X SSC, 0.1% SDS at room temperature for 5 min, and once in 2 X SSC, 0.1% SDS at 50°C for 1 hour, then analyzed using the PhosphorImager SI system (Molecular Dynamics).

2.4 Subcloning Restriction Fragments from Phage Clones:

Both phage clones H26 and C216 were digested with a variety of restriction enzymes and electrophoresed on agarose gels. DNA was transferred to Hybond-N membranes and probed with the hClk2 cDNA as described in section 2.1.

For clone H26, a 3kb Eco RI fragment was cloned into the Eco RI site of the pLITMUS 28 vector (NEB) (construct pH26E1). For clone C216, four fragments were subcloned: a 4kb Hind III fragment (construct pC216H1), and a 3.5 kb Xba I fragment

(construct pC216X1) were subcloned into the same sites as pLITMUS 28; a 3.5 kb Pst I-Kpn I fragment (construct pC216PK2), and a 3.5 kb Sst I-Kpn I fragment (construct pC216SK2), were all subcloned into the same sites in pBluescript II KS+.

2.5 Sequencing of Subclones:

All constructs were sequenced by the ABI Prism 377 DNA Sequencer (Perkin Elmer) using the M13 reverse and universal primers, in addition to the following:

Construct pH26E1 was also sequenced with primers AC1 (5'-cctcagctcaatgacctc-3'), AC2 (5'-ggagagatggctcatcc-3'), AC3 (5'-gcacttggatgggg-3'), and AC4 (5'-aggccagcttggctctgc-3').

Construct pC216H1 was also sequenced with primers AC5 (5'-attctgcctctcctcc-3'), AC11 (5'-ggtttggggttcactcc-3'), and Clk2/ATG (5'-atgccccatccccgaag-3').

Constructs pC216PK2 and pC216SK2 were also sequenced with primers AC7 (5'-gaactgaaatgcacataccc-3') and AC10 (5'-cagttcatctctgtgcc-3').

2.6 Isolation of a Partial cDNA Clone

A partial cDNA clone was isolated by PCR (polymerase chain reaction) with cDNA generated from P19 EC cells using primers derived from the sequencing of the clk2 genomic subclones. The forward and reverse primers used were clk2/235S (5'-tcaagtagcagtgaccgg-3') and clk2/818A (5'-aacatctggacacagagg-3') respectively. The PCR reaction was carried out using 300ng of cDNA in 10 mM Tris-HCl pH 8.3, 50 mM KCl, 2.5 mM MgCl₂, 0.2 mM dNTPs, 20 pmoles of each primer, and 1 unit of Taq polymerase

(Gibco). The reaction was carried out in the GeneAmp PCR System 2400 programmable thermal cycler (Perkin Elmer) through 25 cycles of denaturation at 94°C for 30 sec, annealing at 58°C for 30 sec, and primer extension at 72°C for 1 min.

The PCR product produced was electrophoresed on a 0.8% agarose gel and purified using the GeneClean kit (BIO 101, Inc.). The purified DNA was cloned into both the TA vector (Invitrogen) and pGEM-T vector (Promega) by ligation at 16°C overnight. One-tenth of the ligation reaction was used to transform competent *E. coli* DH5 bacteria by heat shock.

Two clones were sequenced using the M13 reverse and universal primers, as well as primers clk2/235S and clk2/818A.

2.7 RNA isolation and Northern blot analysis:

Total RNA was prepared from cells or tissues using the Total RNeasy Kit (Qiagen). All mouse tissues were isolated from female CD1 mice except heart (female and male) and testis (male). Selection of polyA⁺ mRNA was performed using the polyA Spin mRNA Isolation kit (NEB). Samples of either 10 ug of total RNA or 5ug of poly A⁺ selected mRNA were resolved on 0.7% agarose gels containing 2.2 M formaldehyde, 20 mM 3-[N-Morpholino]propane-sulfonic acid (pH 6.8), 1 mM EDTA, and 5 mM sodium acetate. Electrophoresed RNA was then transferred to Hybond-N membrane (Amersham), and UV cross-linked using the GS Gene Linker (Bio-Rad) at a setting of 125mJ. Membranes were prehybridized for at least 2 hours at 42°C in 6 X SSPE, 50% formamide, 5 X Denhardt's solution, 0.5% SDS, and 100 ug/ml sheared salmon sperm

DNA. Hybridization was carried out with a ³²P-labelled clk2 partial cDNA probe generated using the Random Primed Labelling Kit (Boehringer Mannheim): hybridization was carried out at 42°C for 18-24 hours. Membranes were washed in 2 X SSC, 0.1% SDS at room temperature for 5 min, and once in 0.2 X SSC, 0.1% SDS at 65°C for 1 hour, then analyzed using the PhosphorImager SI system (Molecular Dynamics). The positions of 18S and 28S rRNA were determined by ethidium bromide staining.

2.8 Construction of Gene Targeting Vectors / Knockout Strategy

The IRESbgeo vector (Mountford et al., 1994) was used as the backbone for construction of targeting vectors pDB1 and pDB2. IRESbgeo is a promoter trap type vector that relies on the activity of the endogenous (Clk2) promoter. The internal ribosome entry site (IRES) allows cap-independent translation of the bgeo fusion transcript derived from Lac Z and neo^R genes.

pDB1 was generated by first subcloning the 3' arm of homology, followed by the 5' arm. For the 3' arm, a 1.5 kb Xho I fragment from pC216X1 was subcloned into the Sal I site of IRESbgeo (the SalI/Xho sites are destroyed in this process). To generate the 5' arm, the Kpn I site of pC216PK2 was destroyed by digesting with Kpn I, followed by blunt-ending with T4 Polymerase; Not I linkers (NEB cat) were ligated onto the blunt-ended DNA, followed by digestion with Not I. The 3kb Not I fragment was then subcloned into the Not I site of the IRESbgeo vector to generate the 5' arm of homology.

pDB2 was generated by subcloning the 9.5 kb Xho I fragment from pDB1 into the Xho I site of Bluescript II KS+ (Stratagene).

Both constructs were sequenced at the junctions with M13 reverse and forward primers to verify that no major rearrangements had occurred.

2.9. J1 ES Cell Culture

Embryonic stem cells were grown on gamma-irradiated feeders (see section 2.12) at 37°C in 10% CO₂ in ES cell qualified DMEM (Dulbecco's Modified Eagle Medium, high glucose) media supplemented with 0.1 mM non-essential amino acids (Gibco), 0.1 mM β-mercaptoethanol, 15% fetal calf serum (HyClone), 500 units/ml leukaemia inhibitory factor (LIF, Gibco), and 50 ug/ml penicillin/streptomycin (Gibco). In the absence of feeders, 1000 units/ml LIF were added.

2.10 Electroporation of J1 ES Cells and Isolation of Clones

J1 ES cells (passage day 13) were seeded onto plates with 3×10^6 gamma irradiated feeders, and expanded until approximately 4×10^7 to 1×10^8 ES cells were obtained. Cells were monodispersed with T-EDTA (1mM EDTA, 0.025% Trypsin in PBS) at 37°C for 5-7 min., pelleted at 1000 rpm for 5 min., and resuspended in 5 ml of electroporation buffer (20 mM HEPES, 137 mM NaCl, 5 mM KCl, 0.7 mM Na₂HPO₄, 6 mM D-Glucose, and 0.1 mM β-mercaptoethanol). Cells were pelleted again and resuspended at a concentration of about 5×10^7 cells/ 0.7 ml total volume. Linearized targeting vector was added (100 ug) and the entire contents were transferred to an electroporation cuvette with a 0.4 cm gap (BioRad). Electroporation was carried out at 400 volts and 25 uF using the BioRad Gene Pulsar. The electroporated cells were then distributed to 10 cm tissue

culture plates (about 1×10^7 cells per plate), in ES cell media, each plate also containing 3×10^6 G418 resistant feeders. Approximately 24 hours post-transfection, G418/Geneticin (Gibco) was added to a final active concentration of 200 ug/ml. Medium was changed every 3rd day to remove debris and to replenish G418 levels. After 8 to 10 days of G418 selection, colonies were picked, half being transferred to 96 flat well plates (for later storage at -80°C) and half to 24 well plates (for isolation of DNA and analysis).

2.11. Genomic DNA Isolation and Southern Blot Analysis:

ES cell clones grown in 24 well plates were washed once with PBS, resuspended into lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM EDTA, 10 mM NaCl, 0.5% SDS, and 100 ug/ml Proteinase K), and incubated overnight at 37°C . The cell lysate was extracted twice with Tris-HCl (pH 8.0) equilibrated phenol/chloroform/isoamyl alcohol (24:24:1). Sodium acetate was added to a final concentration of 0.3 M, 2.5 volumes of 100% ethanol were added, and the solution was stored at -20°C overnight. Centrifugation was carried out at $20800 \times g$, the supernatant was discarded, and the pellets were washed with 70% ethanol. Pellets were air-dried for about 2 min, before being redissolved in 100-400 ul of TE.

Approximately 10 ug of DNA was digested to completion with Eco RI, electrophoresed on 0.8% agarose gels, and transferred to Hybond-N membranes. After cross-linking and prehybridization (as above), the membrane was incubated with a genomic fragment generated by digesting pC216PK1 with Pst I and Xho I to isolate a 0.6 kb probe (PK600).

2.12 Preparation of Feeder Layers from Embryonic Fibroblasts

The G418 resistant embryonic fibroblasts (EMFIs) were originally obtained as a kind gift from Dr. M. Rudniki, and their isolation has been well described (Joyner, 1993). EMFIs were grown in alpha-MEM (Gibco) supplemented with 10% calf serum, and expanded until approximately $2 \times 10^7 - 10^8$ cells were obtained (approximately 36 X 10 cm plates). Cells were trypsinized and resuspended in serum/media at a concentration of 10^7 cells/ml. Cells were then growth arrested by gamma-irradiation treatment at a dose of 3000 rads.

3. Results

3.1 Delineation of the genomic structure of murine *Clk2*

Prior to developing a gene targeting strategy for *mClk2*, some knowledge on the genomic structure is necessary. An embryonic stem cell line (D3) genomic library (generously provided by Dr. D.A. Gray), was screened with the human *Clk2* (hClk2) cDNA. Two clones were identified: C216, which spans 13kb and H26, which spans roughly 24kb. Using restriction fragments of the hClk2 cDNA as probes, slot blot analysis (Figure 4) revealed that C216 extends across the 5' end of the gene, containing exons with sequences homologous to the first half of the hClk2 cDNA (Bam HI-HinD III fragment). In contrast, H26 extends across the 3' end, containing exons with sequences homologous to the second half of the cDNA (HinD III-Eco RV fragment).

Using hClk2 cDNA as a probe, trapped exons were delineated by Southern blot analysis of DNA restriction fragments of the clones. Analysis of clone C216 revealed that the exons were encompassed in a 4kb HinD III and a 3.5kb Xba I fragment (Figure 5), whereas for clone H26, exons were present in a 3kb Eco RI fragment (Figure 6). These fragments were subcloned into pLITMUS 28 and designated pC216H1, pC216X1, and pH26E1 respectively.

Clones and subclones were further characterized by sequence analysis and restriction mapping. From the sequence information obtained to date, computer-assisted searches of GenBank, EMBL, NBRF/PIR, and Swiss data banks identified putative exons in the three

Figure 4: Slot Blot Analysis of genomic clones C216 and H26

For each lambda phage sample, 10 ng (slot 1) or 1 ng (slot 2) was applied to each slot. As a positive control full length hClk2 cDNA was used (1ng or 100 pg). Each slot blot was hybridized with ³²P-labelled restriction fragments of the hClk2 cDNA as indicated.

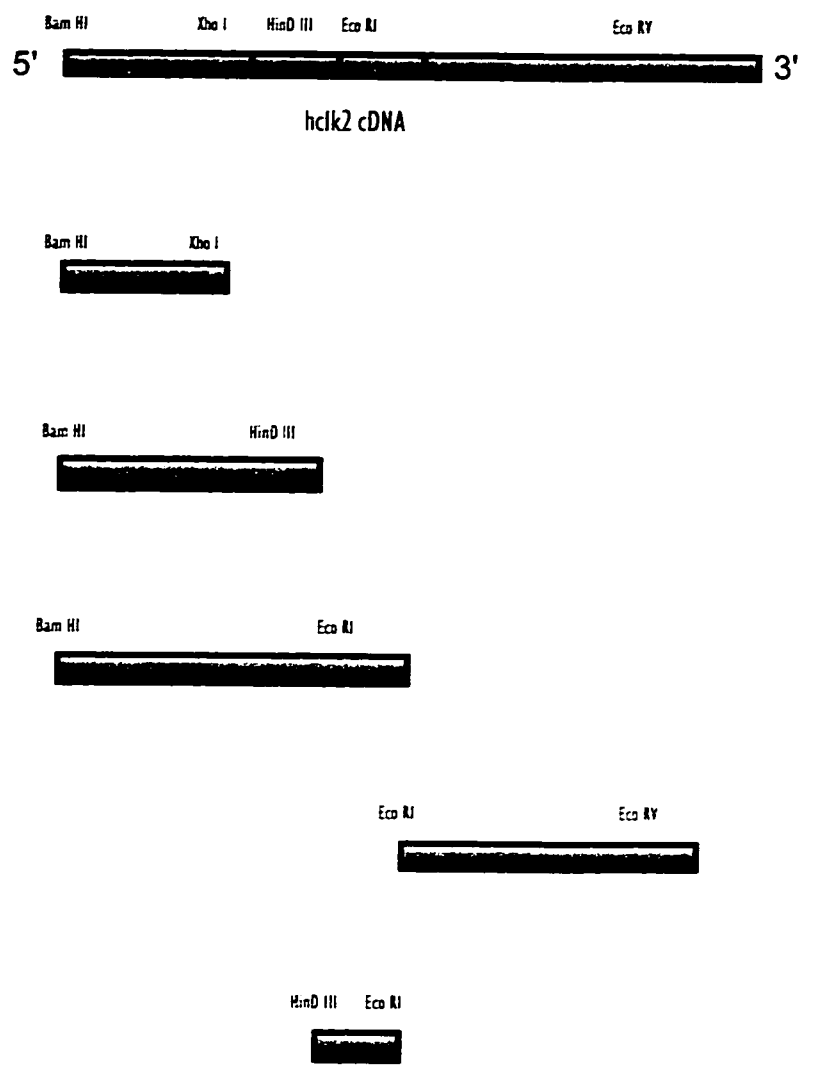
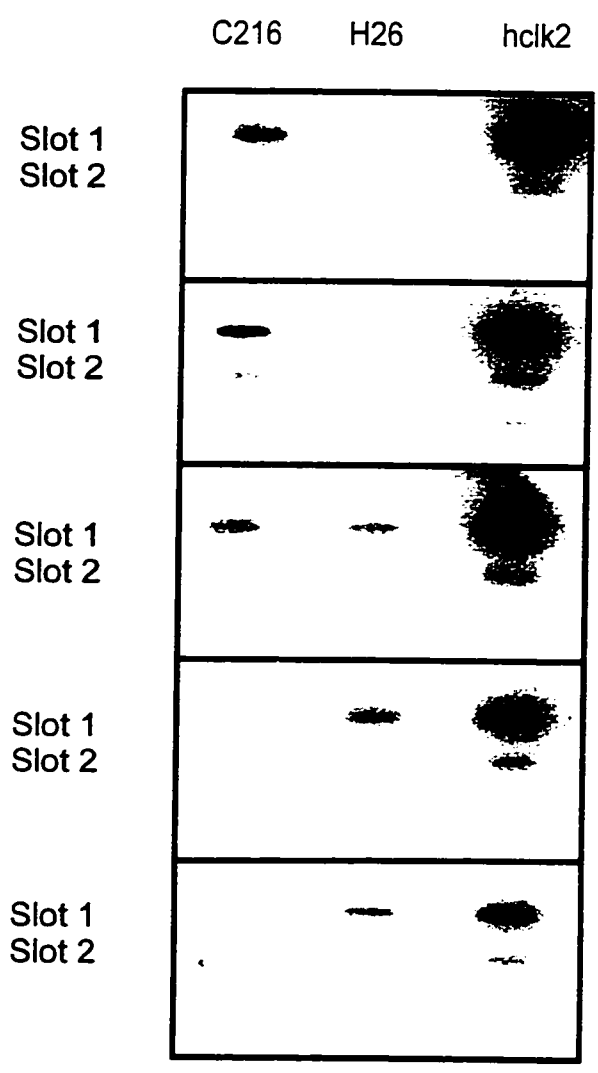


Figure 5: Southern Blot Analysis of Restriction Fragments of Clone C216

Each lane contains approximately 0.5 ug of phage DNA digested with the indicated restriction enzymes. The blot was hybridized with ³²P-labelled hclk2 cDNA probe. Note that M denotes the 1 kb ladder lane.

Eco RI Eco RV HinD III Kpn I Nco I Xba I Xho I M

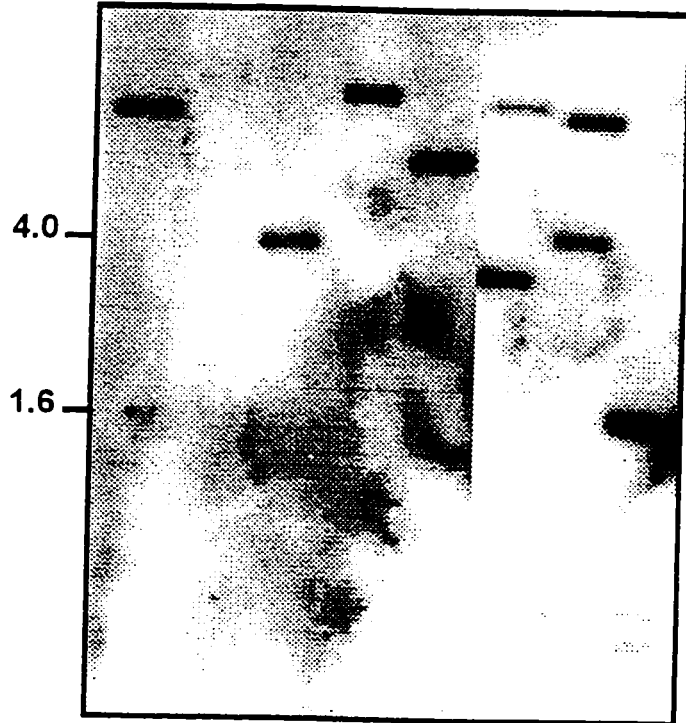


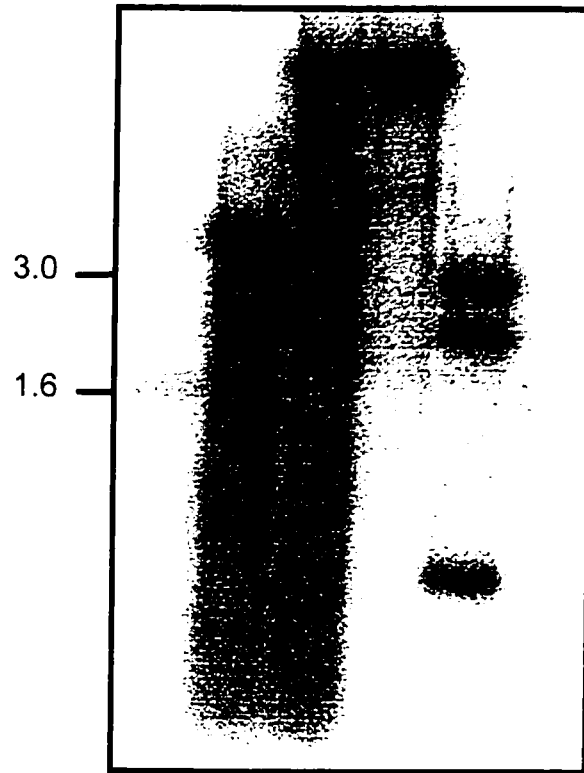
Figure 6: Southern Blot Analysis of Restriction Fragments of Clone H26

Each lane contains approximately 5 ug of phage DNA digested with the indicated restriction enzymes. The blot was hybridized with ³²P-labelled hck2 cDNA.

H26

C216

Eco RI Bgl II Eco RI Bgl II



subclones that shared high homology to hClk2 at the nucleotide level.

In the region sequenced for pH26E1, two exons were identified, or is shown in figure 7. One of the exons is incomplete as it ends at the site where the restriction fragment was cloned (Eco RI site). Both 3' splice sites conform to the consensus sequences, containing both the polypyrimidine tract and the NCAG sequence at the intron-exon boundary. At the nucleotide level, the combined 231 (167 + 64) nucleotides of the two exons exhibit 96% homology to *hClk2*; splicing together the two exons generates a continuous reading frame with a predicted amino acid stretch that is 97% identical (99% similar) to hClk2 protein. In addition, this stretch spans subdomains V to VI in the catalytic domain of hClk2.

Sequence analysis of pC216H1 and pC216X1 resulted in the identification of two more exons, shown in figures 8 and 9. The exon found in pC216H1 is 172 nucleotides in length and displays 85% homology to *hClk2*. The predicted amino acid sequence begins with the putative initiating methionine, and is 98% identical to hClk2 protein.

The exon located in pC216X1 is 229 nucleotides long and exhibits 87% homologous to *hClk2*. Both intron-exon boundaries conform to the GT-AG rule, however, the polypyrimidine tract at the 3' splice site is almost non-existent. This may be due to sequencing errors, since this region has only been sequenced once on one strand. The predicted amino acid sequence is 92% identical (96% similar) to hClk2 protein (Hanes *et al.*, 1994).

Predicted splicing of the exons identified in pC216H1 and pC216X1 essentially results in sequences coding for the first 133 amino acids for the putative mClk2 protein.

Figure 7: Partial Sequence of Subclone pH26E1

Exons and introns are represented by uppercase and lowercase letters respectively. The predicted amino acid sequences of the exons are given as well. The polypyrimidine tracts are underlined.

tttgtcaggtagcgctttttgggtccttctctagatgacctgaattttgttgatgatttta	60
attgtttttgtttgtttgtttgtttgggtgtttgtttgttttttttaagtgacagggta	120
tcatgtagccaggcaaacatcttcaaacttactatgtagtcgaggataatcttgaattt	180
atgattctcctgcctccaactcccaaagtctgagattacagattcatgccatcataattg	240
ttatatggaatgctgcagattgaacttgggccttctgcatgctaagtaggcactctgcca	300
acttacctacattcctagtccttatcagagacattagctagcatgtttttgtacgactgt	360
agaatcaacaaccatacagactttcattttgggtgggggtgctgggttttgggaacttatta	420
gacatccaacttcccctgtctctcccagCCTCTGTGTCCAGATGTTTGACTGGTTTGAC	480
L C V Q M F D W F D	
TACCATGGCCACATGTGTATCTCCTTTGAGCTTCTGGGCCTTAGCACCTTCGATTTCCCTC	540
Y H G H M C I S F E L L G L S T F D F L	
AAAGACAACAATTACCTGCCNTACCCATCCACCAAGTGCGCCACATGGCCTTCCAGTTC	600
K D N N Y L P Y P I H Q V R H M A F Q F	
TGCCAGGCCGTCAAGTgtgagtggggtgggctgaggcgaantctggggcagccctttcct	660
C Q A V K F	
ccatcaggcctcttgtgtgtacttgtgagctaggcatgcccctacgcagtcagctcactc	720
tcacacctgctgctgtcatggcagctcctcctccgacttaccacctgactgacatcttt	780
aatatcaccagttgctcttactgcaaaccatagcacttagtgtcttttctcctcctcttt	840
tttggaaaagtactgtacaacaagcacagaggtgtttcttctgctagagtcttaggtttgtc	900
taatgagtgcttctttatacagcagatgcaagaaccctgggtcattgagctgaggtacta	960
gcttcagaatctagagtagttgtttntctgcatcattaggcagtcacatagatcaaccg	1020
taagaccaaaaaaaaaactcccacttaggtacactgctgcttttctatcagatgaactata	1080
tgtgcattttatattactaagggcagtgcttaggccagccattttatttgattggagtgc	1140
agaaagtagccccagtgtagtctgggcctgcctcctcaccagctttgtcttttagtattca	1200
gtacgggggtcagttgctggcagttagttctagccacctggatcaagcctgatctag <u>ctt</u>	1260
<u>ctcccctgcagTCCTCCATGATAACAAGTTGACACATACGGACCTCAAACCTGAAAATAT</u>	1320
L H D N K L T H T D L K P E N I	
TCTGTTTGTGAATTC	1335
L F V N	

Figure 8: Partial Sequence of Subclone pC216H1

The exon and introns are represented by uppercase and lowercase letters respectively. The predicted amino acid sequence of the exons is given as well. The polypyrimidine tracts are underlined.

tctgagttcaaattctgccagata	60
agaacnttgtctcaataaaggaaagagagaaagaaaaggactggggcagcggaaacaccct	120
ttgaatctcagtggttgggaaggcataggcaggtggatctctgtgagttccaggccagcct	180
ggtctacaaaatcaatttccaaacagccagggcgacatagagaaaaccctgcctcaaacc	240
aaaagaaaaggaaaagtgcagttatctcctgcttaataaatccagtggtgctggggacat	300
acagatggacacacttcccttgtgctgtttacaactgtgtgaaaaatttggccagggcat	360
gtgagagacatttgtgggaccacatcctcgtatctctggaacttgttggtgtttgctgccag	420
ttatctgtaaaatcctgggagggtgaatgactgcctgtgacactcytagtctagtgggg	480
gagagaagatgagctgttaataactaaaacaacctgtaaactaaagctttatttgttt	540
gcttgtgaagcactcaggatggcacagctcgcatggttagatggcagcagagatgaactg	600
tggcaatatttatcagccacacacttgtcactgtcgtgaactgtggctcagaccagaag	660
gtcttcatggagccttgtttgaggtagagcaactgaatttgactttcttgagtttccagc	720
caaaatttgtaggtggctgtatagtccatgaagggaagactattaccaggaaaagaagcg	780
aggcgtttgtgtgctgaagggcgtagggacagaaagcagctttgcttcttagaagaggc	840
attccacgagaacggttatgaaaactcagtggtcctaccattctgcctctccttccctctc	900
cagATGCCCCATCCCCGAAGGTACCATTCTCAGAGCGAGGTAGCCGGGGGAGTTACCAC	960
M P H P R R Y H S S E R G S R G S Y H	
GAACACTATAGAAGCCGAAAGCATAAGCGAAGAAGAAGTCGCTCCTGGTCAAGTAGCAGT	1020
E H Y R S R K H K R R R S R S W S S S S	
GACCGGACAAGGCGGGCGGGAGGGAGGACAGCTACCACGTTCCGGTCCCGAAGgtaanaa	1080
D R T R R R R R E D S Y H V R S R	
agagactgattgcaggtcactggttacrttacatgatgaggccagtggtcaccagctg	1140
gttgaaggccaarggcctasgtatggacraatgtgcctgacatggatttggtttgggggtt	1200
cacttccctgccattgcaaaacaaagggaragggtgttgggtgggtggacacctgggatgggag	1260
gggaggggggattatganctcaaggctacttgggcttcatactgantttgaggtcagcct	1320
gntctacataatgagactctcaaagggggaaaaacatgttttgtttctttttttgcctaa	1380
atgttttttttttaataactttttattangtattttcctcatttacatttccaatgcta	1440
ttccactggaaaaacatgttttttgaagggaactttctagaggaagtacttcgggattaaa	1500
gctgtaccaccaggcatgttggaatggtcttgatataatagttcasggaaaacaac	1560
ttttaaaaaaatttttatgtgtataggtgttttaccacatgtataccttgattatgt	1620
catccatgggtgcccttgtccattanaanagaatatgggagtcctggaactggagttacag	1680
gcantgtgagctacnatgtgggtcatggaaatggaactcaggttactgaaagagcagt	1740
cagtgtcttaaccagtggaactatctctagtcattttccttttgttcttttagacaaagtg	1800
ttactgtgttaccctggcctggcctagaactttctgtgtt	1840

Figure 9: Partial Sequence of Subclone pC216X1

The exon and intron are represented by uppercase and lowercase letters respectively. The predicted amino acid sequence of the exon is given as well.

taagaaggacccattgcctgctgtggcagggaacgtggcctggactgaccttgctagtg 60
ggtatattgttatccaacagCAGCTATGATGACCATTCGTCCGATCGGCGGCTGTACGATC 120
S Y D D H S S D R R L Y D
GGCGGTACTGTGGCAGCTACAGGCGCAATGACTACAGCCGGGACAGAGGGGAGGCTTAYT 180
R R Y C G S Y R R N D Y S R D R G E A Y
ACGACACAGACTTTCGGCAGTCCTATGAATACCATCGAGAGAACAGCAGTTACCGAAGCC 240
Y D T D F R Q S Y E Y H R E N S S Y R S
AGCGCAGCAGCCGAAGGAAACACAGAAGGCGGAGGAGACGGAGCCGGACATTCAGCCGCT 300
Q R S S R R K H R R R R R R S R T F S R
CATCTTCAGtgagtgccagcccagccttttactctccccattcttcttcagtccctcagg 360
S S S
gaccctgatgagtaaaatgtccccttgttctcagacatgctttggagcatgaacattgga 420
gtgggcactaagtatgcctaccctcttgggagctctctgactctcgacctcgag 474

This region is very rich in arginine and serine residues (53%), making it a strong RS domain. In addition, a potential bipartite nuclear localization signal is located within amino acids 28 to 47.

Having both regions of the gene at my disposal allowed me to choose which region to be targeted for disruption. As deletion of the 5' end of the gene is less likely to generate functional polypeptides, this region was chosen, and thus C216 was more extensively characterized. Digestion with Eco RI revealed only one site in C216, and partial restriction mapping (Figure 10) defined its position.

Further sequence analysis of subclone pC216X1 showed that the insert included sequences from the phage arm of the λ DASH II vector. Since no Eco RI sites are present in pC216X1 (data not shown), the position of pC216X1 was also defined. In addition, sequencing data also revealed that subclones pC216X1 and pC216H1 overlap, thus permitting their alignment with clone C216. Their relative positions and sizes are depicted in figure 10.

The decision to disrupt the exon containing the putative initiating methionine necessitated the subcloning of DNA fragments upstream of pC216H1. Clone C216 was digested with a multitude of enzymes, and subjected to Southern analysis with probe HK465 (Figure 11). Sst I-Kpn I and Pst I-Kpn I fragments of about 3kb in size were found to hybridize to the probe and were subcloned into pBluescript II KS+. These were designated pC216SK2 and pC216PK2 respectively. Sequencing permitted their alignment along C216 (Figure 12).

Figure 10: Alignment of lambda clone C216 with subclones pC216H1 and pC216X1

Figure is drawn to scale. Exons are denoted by thin black boxes. Restriction sites: H, Hind III; Xb, Xba I; E, Eco RI.

| 2.9 kb E |

10 kb

| C216

H
| |

4 kb

H
|

PC216H1

Xb

3.5 kb

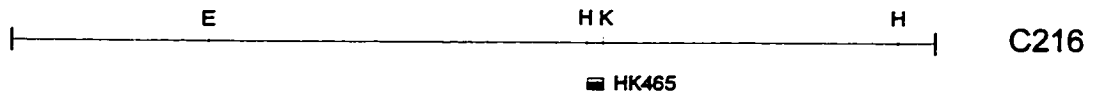
Xb

PC216X1

Figure 11: Subcloning Fragments Upstream of the First Coding Exon

- (A) The position of probe HK465 relative to clone C216 is shown.
- (B) Each lane contains approximately 1 ug of phage DNA restricted with the indicated enzymes. The blot was hybridized with the HK465 probe. M denotes 1 kb ladder lane.

A



B

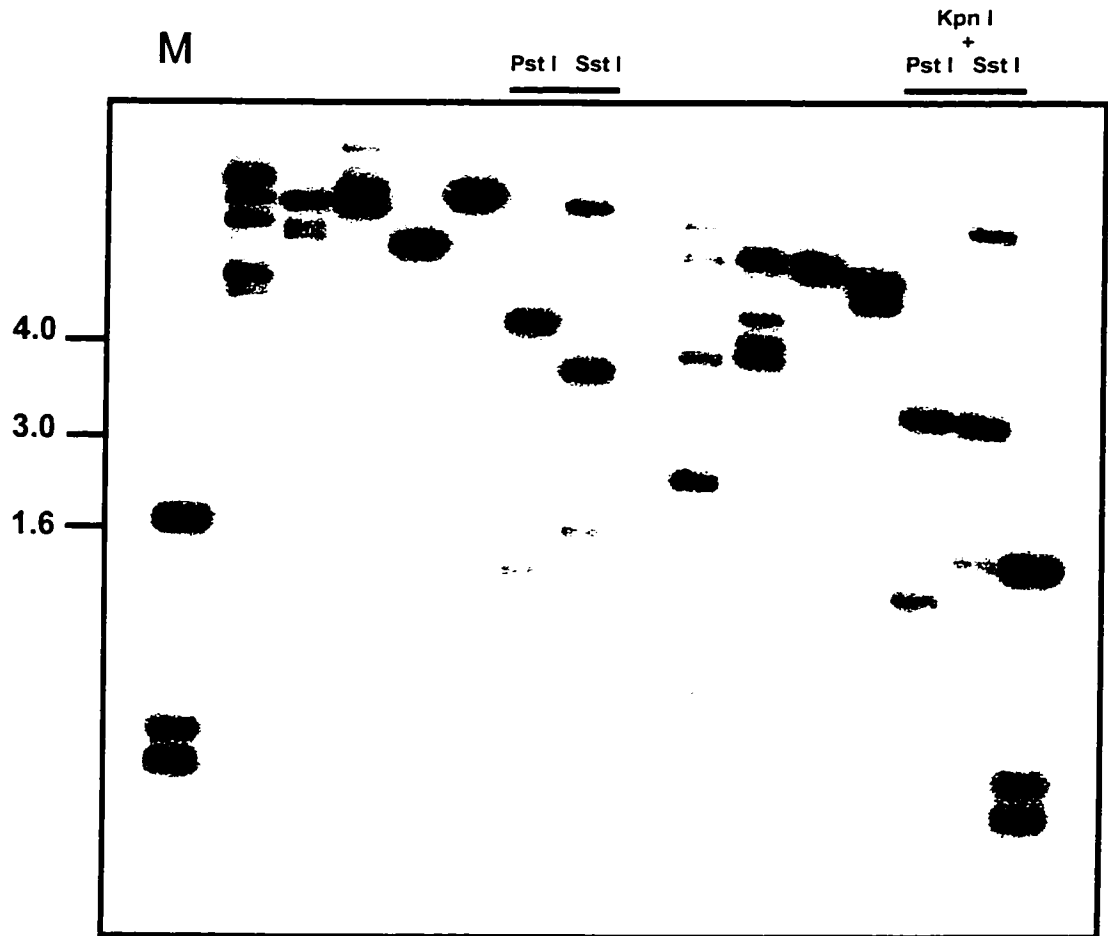


Figure 12: Alignment of 4 subclones relative to C216

Figure is drawn to scale. Exons are denoted by thin black boxes. Restriction sites: H, Hind III; Xb, Xba I; E, Eco RI; P, Pst I; K, Kpn I.

I 2.9 kb E I

10 kb

I C216

H I I

4kb

H I

pc216H1

Xb I

3.5kb I

Xb I

pc216X1

P I 3.5kb

K I

pc216PK2

S I 3.5kb

K I

pc216SK2

3.2 Comparison of murine and human Clk2 cDNAs

Identification of exons in the genomic DNA provided an advantage in the design of primers to isolate murine Clk2 cDNA by RT-PCR. Using poly A⁺ selected RNA from P19 murine embryonal carcinoma cells as a source, a 581 bp cDNA fragment was isolated. Two independent clones were sequenced on both strands and no discrepancies were found. The partial cDNA sequence is shown in figure 13. At the nucleotide level, the murine cDNA fragment exhibits 90% homology compared to hClk2 sequence. The predicted amino acid sequence, which corresponds to amino acids 35 to 229 in hClk2, is 99% identical to hClk2 over a stretch of 194 amino acids. Partial alignment of the human and putative mouse proteins is shown in figure 14. This represents most of the RS N-terminal putative regulatory domain, which is 54% rich in serine/arginine residues and contains 7 RS/SR dipeptides; the putative alternatively spliced exon encoding region (a.a. 134 to 163); and subdomains I and II of the catalytic kinase region, including the highly conserved YEIV and GXGXXG motifs, as well as the invariant lysine found in protein serine kinases (Hardie and Hanks, 1995).

When compared to the amino acid sequence predicted by the exons in the genomic sequence, the overlapping regions exhibited no divergence.

Figure 13: Sequence of the 581 bp partial Clk2 cDNA

Primers derived from the genomic sequence were used to isolate a partial Clk2 cDNA from P19 cells by RT-PCR. Numbering system for the nucleotide and amino acid sequence is arbitrary. The putative kinase domain begins with the leucine at amino acid 133. The GXGXXG motif conserved in serine kinases and the conserved YEIV motif among Clk kinases have been underlined. The invariant lysine found in kinases is amino acid 157.

TCAAGTAGCAGTGACCGGACAAGGCGGCGGGAGGGAGGACAGcTACCACGTTCCGGTCC	60
S S S S D R T R R R R R E D S Y H V R S	20
CGAAGCAGCTATGATGACCATTCGTCCGATCGGGCGGCTGTACGATCGGGCGGTACTGTGGC	120
R S S Y D D H S S D R R L Y D R R Y C G	40
AGCTACAGGCGCAATGACTACAGCCGGGACAGAGGGGAGGCTTACTACGACACAGACTTT	180
S Y R R N D Y S R D R G E A Y Y D T D F	60
CGGCAGTCCTATGAATACCATCGAGAGAACAGCAGTTACCGAAGCCAGCGCAGCAGCCGA	240
R Q S Y E Y H R E N S S Y R S Q R S S R	80
AGGAAACACAGAAGGCGGAGGAGACGGAGCCGGACATTCAGCCGCTCATCTTCACACAGC	300
R K H R R R R R R R S R T F S R S S S H S	100
AGCCGGAGAGCCAAGAGTGTAGAGGACGACGCTGAGGGCCACCTCATCTACCACGTCGGG	360
S R R A K S V E D D A E G H L I Y H V G	120
GACTGGCTACAAGAGCGATATGAAATTGTAAGCACCTTAGGAGAAGGGACCTTCGGCCGA	420
D W L Q E R <u>Y E I V S T L G E G T F G</u> R	140
GTTGTGCAGTGTGTGGACCATCGCAGGGGCGGAACACGAGTTGCCCTGAAGATCATTAAAG	480
V V Q C V D H R R G G T R V A L K I I K	160
AATGTGGAGAAGTACAAGGAAGCAGCCCGACTAGAAATCAACGTGCTGGAGAAAATCAAT	540
N V E K Y K E A A R L E I N V L E K I N	180
GAGAAAGATCCTGACAACAAGAACCTCTGTGTCCAGATGTT	581
E K D P D N K N L C V Q M	193

Figure 14: Partial Alignment of Human and Mouse Clk2 Proteins

Numbering system is only shown for the human sequence. Amino acid divergences are indicated by a (:) whereas conservative substitutions are shown by a (+).

mC1k2	SSSSDRTRRRRRREDSYHVRSRSSYDDHSSDRRLYDRRYCG	
hC1k2	SSSSDRTRRRRRREDSYHVRSRSSYDDRSSDRRVYDRRYCG	75
mC1k2	SYRRNDYSRDRGEAYYDTDFRQSYEYHRENSSYRSQRSSR	
hC1k2	SYRRNDYSRDRGDAYYDTDYRHSYEYQRENSSYRSQRSSR	115
mC1k2	RKHRRRRRRSRTFSRSSS-HSSRRAKSVEDDAEGHLIYHV	
hC1k2	RKHRRRRRRSRTFSRSSSQHSSRRAKSVEDDAEGHLIYHV	155
mC1k2	GDWLQERYEIVSTLGEETFGRVVQCVDHRRGGTRVALKII	
hC1k2	GDWLQERYEIVSTLGEETFGRVVQCVDHRRGGARVALKII	195
mC1k2	KNVEKYKEAARLEINVLEKINEKDPDNKNLCVQM	
hC1k2	KNVEKYKEAARLEINVLEKINEKDPDNKNLCVQM	229

3.3 Expression of Clk2 transcripts

To investigate the expression pattern of Clk2, northern analysis of poly A+ samples (Figure 15) from adult mouse tissues, E13 embryos, and P19 cells was performed with the mClk2 partial cDNA as a probe. Clk2 appears as a ubiquitous 2.0 kb message. Furthermore, in differentiated adult tissues, additional transcripts of 2.2 kb and 3.0 kb in size were detected. Interestingly, the 2.2 kb transcript appears to be testis specific. Semi quantification (normalizing against GAPDH) would suggest that expression of Clk2 is highest in testis, ovary, lung and brain.

Since I intended to use a promoterless vector in the knockout strategy, I examined the expression level of Clk2 in ES cells. Northern analysis of total RNA isolated from J1 and B7 ES cells also showed the expression of the 2.0 kb transcript (Figure 16). When normalized to GAPDH expression, Clk2 expression in ES cells is roughly equivalent to P19 EC cells. B7 is a J1 ES cell line that is homozygous-null for Clk1 (our unpublished data). The finding that the probe hybridizes to B7 RNA suggests that the probe is not cross-hybridizing to Clk1 mRNA in J1 cells.

3.4 Isolation of Clk2 single knockout clones

Previous studies using the bgeo vector for gene-targeting experiments with Clk1 resulted in high homologous recombination frequencies (around 50% homologous recombinants/resistant clones; Peter Duncan, personal communication). Although this strategy presents an attractive enrichment scheme, its feasibility depends on the

Figure 15: Northern blot analysis of Clk2 expression in adult mouse tissues

Approximately 5 ug of poly A+ selected RNA isolated from adult mouse tissues (T, testis; O, ovary; Lu, lung; Li, liver; K, kidney; H, heart; B, brain), embryonic day 13 embryos (E), and P19 cells (P19), were subjected to Northern blot analysis using the 581 bp mClk2 partial cDNA as a probe. The same transfer was reprobbed with a GAPDH probe as a loading control.

Note: exposure time for Clk2 expression in testis was considerably less.

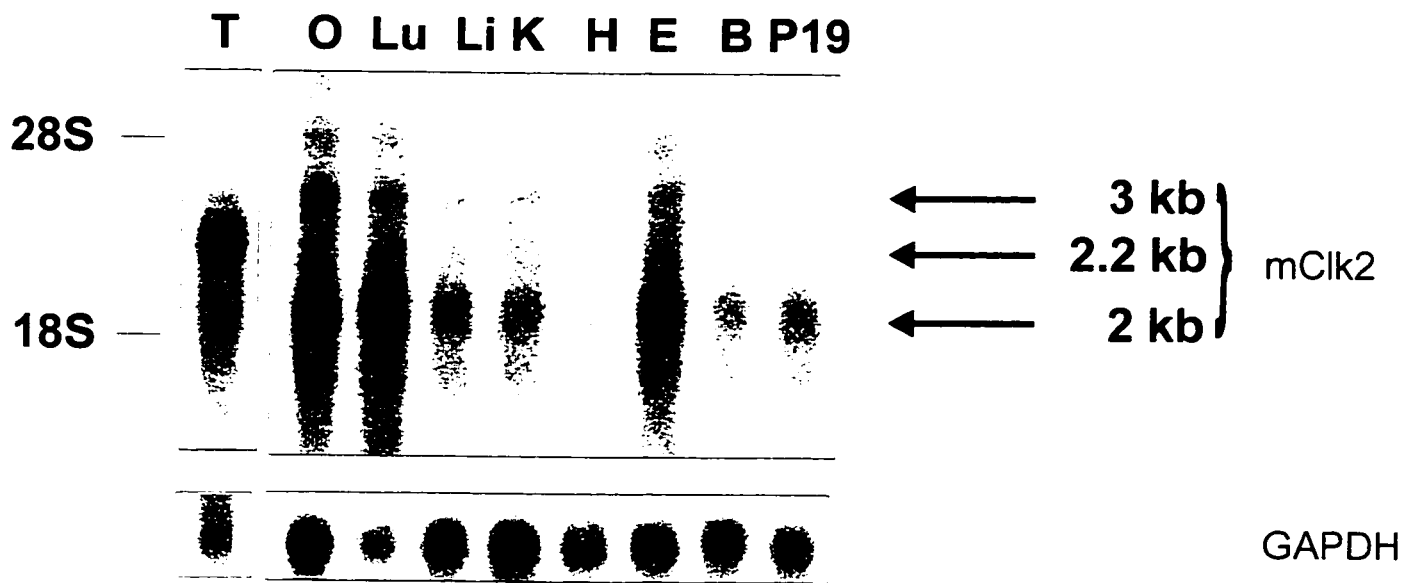
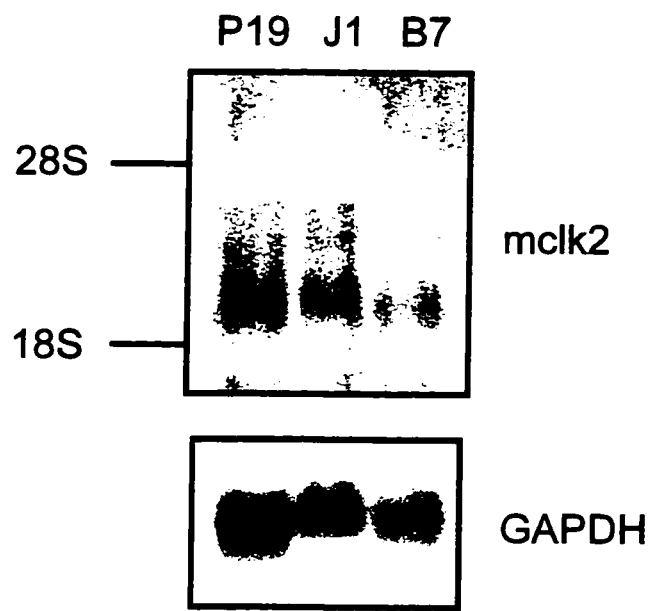


Figure 16: Northern blot analysis of Clk2 in ES and EC cells

Approximately 10ug of total RNA isolated from P19 cells, J1 and B7 (mClk1 $-/-$) ES cells were subjected to Northern blot analysis using the 581 bp mClk2 partial cDNA as a probe. The same transfer was reprobbed with a GAPDH probe as a loading control.



transcriptional activity of the gene in question in embryonic stem cells. Northern analysis of total RNA isolated from J1 ES cells showed detectable expression of *Clk2* transcript; thus, this approach seemed feasible.

Approximately 5×10^7 J1 ES cells were transfected with the knockout vector pDB1, and after selection in G418, 96 resistant clones were picked. Southern blot analysis was then performed to screen for clones that had undergone homologous recombination. After Eco RI digestion, using the probe depicted in figure 17, hybridization to the wild type *mClk2* allele would result in a band of at least 10 kb. However, in the case of a homologous recombination event, a 7kb band should be expected. In this experiment, southern blot analysis of 48 clones yielded one potential clone (clone 6B) that had undergone homologous recombination (Figure 18).

Using the knockout vector pDB2, 128 G418 resistant clone were picked from the transfection of approximately 4×10^7 J1 ES cells. Using the same screening strategy, Southern blot analysis of 42 clones has not identified any homologous recombinants to date.

Figure 17: Schematic of the Clk2 Region Targeted.

Diagram drawn to scale (except for exons). Restriction sites: Hind III; Xb, Xba I; E, Eco RI; P, Pst I; K, Kpn I; Xh, Xho I. Exons are denoted as black boxes. Upon a homologous targeting event between the vector and the locus, 149 bp of the first coding exon corresponding to amino acids 8-56 of Clk2 are deleted. Homologous events are screened by hybridization with probe PK600 to Eco RI digested DNA. A targeted event will result in the hybridization of a 7kb fragment to the probe.

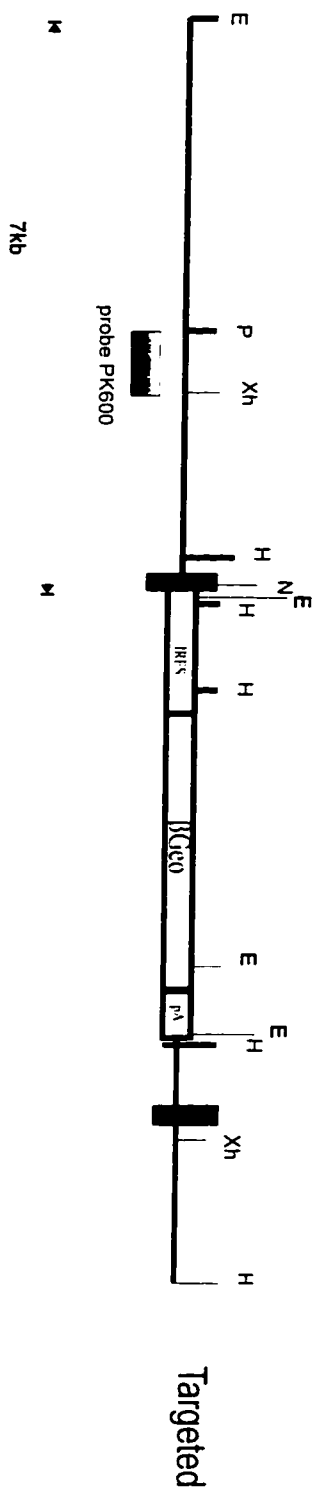
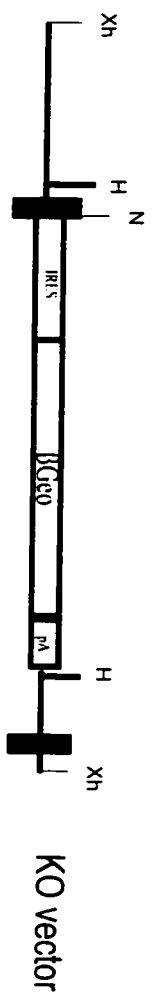
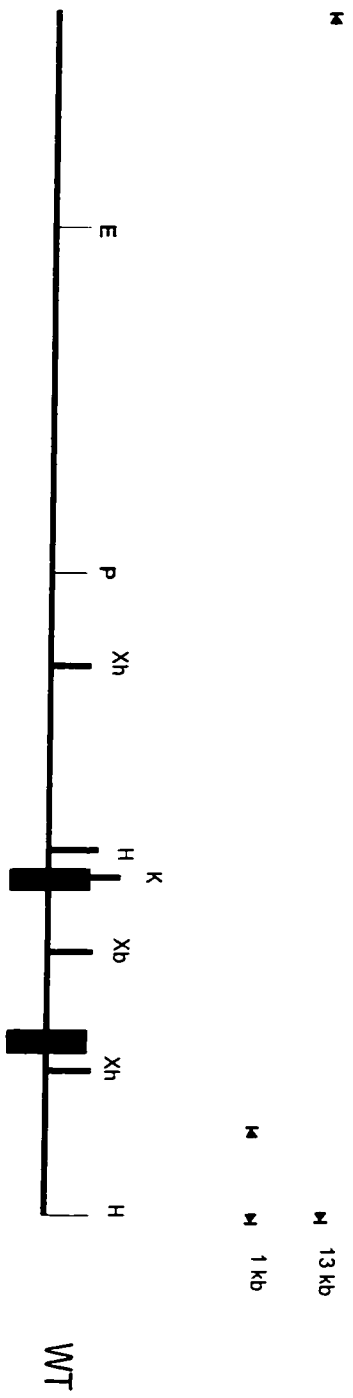
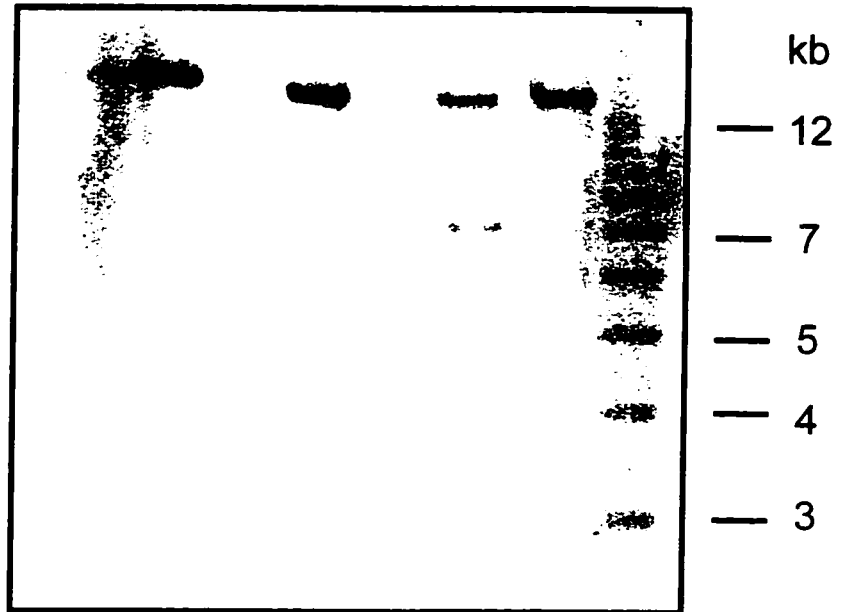


Figure 18: Identification of a Clk2 targeted clone

Approximately 10 ug of DNA from G418 resistant clones were isolated and digested to completion with Eco RI. Southern Blot analysis was performed using the probe PK600. Wild type band is expected to be greater than 10kb whereas a targeted allele would give rise to a 7kb band. Clone 6B is identified as a potential targeted clone. M denotes 1 kb ladder lane.

3C 4D 5A 6B J1 M



4. Discussion

The Clk family of kinases is well conserved from yeast to humans, arguing that these kinases confer a selective advantage in the organism. Within the family, the *Drosophila* homolog Doa is the best characterized. Doa mutants result in lethality at the larval stage, suggesting a developmental role for Doa. In an attempt to identify a developmental role for mClk1, mice null for *clk1* were generated; no developmental phenotype has been observed, raising the possibility that Clk1 loss of function is rescued by family members (Clk2, Clk3, and Clk4). The ultimate goal of this work was to derive Clk2-null mice through gene targeting experiments. Clk2 was chosen for this study over Clk3 and Clk4, primarily because it possesses the highest identity to Doa.

Most embryonic stem lines (including the J1 line that was used) were produced from embryos derived from 129 mice substrains (Evans and Kaufman, 1981). It is well known that the degree of homology increases the success of homologous recombination (Joyner, 1993). Thus, an isogenic genomic library (D3 genomic library) was screened to obtain fragments for the construction of targeting vectors.

Two genomic clones were isolated, and upon sequencing and slot blot analysis, they were found to span different ends of the gene. One clone, H26, spans the 3' end of the gene, and contains at least two exons that code for subdomains V to VI of the catalytic kinase region. Clone C216 spans the 5' end, and so far two exons have been identified that code for the first 77 amino acids of Clk2.

Using primers derived from the exons identified, a 581 bp cDNA fragment was

isolated from P19 cells by RT-PCR. Computer assisted analysis revealed that this partial cDNA exhibited 87% homology to hClk2 at the nucleotide level. The predicted amino acid sequence is 99% identical to hClk2. Although the human Clk kinases are alternatively spliced to generate the truncated isoforms (Hanes et al., 1994), a murine Clk2 truncated version has yet to be identified. It will be interesting to clone the putative alternatively spliced exon in the genomic sequence to see if exon exclusion would result in (1) a change in the open reading frame, and (2) a non-sense mutation resulting in truncation.

To determine the expression pattern of mClk2, northern analysis was performed on various tissues and cell lines. Clk2 message seems to be expressed in all adult tissues tested, and is also present in E13 mouse embryos, as well as in embryonic stem cells and in embryonal carcinoma cells. In all samples, a 2kb transcript was detected: this is consistent with the known size of hClk2 cDNA (1995 nt) and the approximate size of the human mRNA species (Hanes et al., 1994). In addition, differentiated tissues express two additional transcripts of unknown function, one of which is testis-specific. Multiple transcripts of Clk1 are also expressed in differentiated tissues, and it seems that these additional transcripts are partially spliced mRNAs and are restricted to the nucleus (Duncan *et al.*, 1995). Similar experiments for Clk2 need to be done to verify whether this is the case as well. However, this notion seems likely as the accumulation of these partially spliced mRNAs could be explained by the lack of a strong polypyrimidine tract upstream of the second coding exon (Figure 9), resulting in inefficient splicing.

Although the Clk2 gene is transcribed ubiquitously, it is unclear whether Clk2 protein

is found in any tissues. Consistent with this notion, is the finding that Clk1 message is found in all tissues examined, however, preliminary studies suggest that the protein is absent in adult tissues (David Stojdl, personal communication). Recently, an antibody generated against human Clk2 has been obtained (generously from Dr. Rabinow, University of Nebraska Medical Center), and it will be interesting to determine where Clk2 protein is found.

Having identified genomic clones spanning both ends of the Clk2 locus gave me the luxury of choosing the region for targeted disruption of Clk2. The decision was made to disrupt the 5' end in order to minimize the possibility of generating truncated polypeptides with residual activity. This is significant in view of the fact that truncated isoforms of Clk kinases exist, and that they oppose the functions of full-length Clk kinases in alternative splicing (Colwill *et al.*, 1996; Duncan *et al.*, in press).

I employed a promoterless knockout vector as a positive enrichment scheme for obtaining targeted clones (for brief review, see Joyner, 1993). Essentially, the strategy relies on the transcriptional activity of genes to be targeted: as Clk2 RNA was detected in ES cells, this approach seemed feasible. In addition, transcription of the positive selection marker (*neo*) is rendered under the control of the endogenous promoter due to the presence of the internal ribosome entry site (IRES), which allows cap-independent translation (Mountford *et al.*, 1994).

The design of the knockout vector (pDB1 or pDB2), results in insertion into the exon containing the putative initiating ATG, and causes the replacement of 50 codons by the vector, leaving only the first 7 amino acids of Clk2. As the sequence of the upstream

exon is unknown, it is difficult to predict the consequences of splicing over (skipping) this exon.

So far, attempts at gene targeting experiments in ES cells has only yielded one potential single knockout clone (clone Clk2 6B). Compared to the previous success rate for *Clk1* (Peter Duncan, personal communication), this seems quite low. Perhaps the arms of homology used for the knockout vector are suboptimal, or perhaps the *mClk2* locus is less amenable to homologous recombination. In any instance, less than 100 clones have been screened, so more potential clones will have to be investigated. In addition, clone Clk2 6B will also have to be screened by 3' and internal probes to stringently verify that it represents a true homologous recombination event (a true knockout).

Once the generation of multiple Clk2 targeted ES cell lines have been achieved, Clk2 ^{-/-} mice will be produced by injection into blastocysts or by aggregation with early stage embryos. Since the exact biological function of Clk2 is unknown, it is difficult to predict the consequences of loss of function. However, among the mammalian Clk kinases identified to date, Clk2 is the closest to the *Drosophila* homolog *Doa*. Since *Doa* mutants result in lethality, it is likely that a Clk2 ^{-/-} genotype will result in severe developmental defects as well. Analysis of Clk2 ^{-/-} mice should provide further insight to the biological function of Clk2. In the event that a *Clk2* null phenotype is lethal, analysis of the expression pattern (e.g. by *in situ* hybridization) during embryonic development may provide an insight to its biological function. It is also possible that functional redundancy within the family rescues Clk2 loss-of-function. In that case, Clk2 deficient mice will be bred to Clk1 deficient mice, to produce mice which are homozygous null for both

kinases, and perhaps this will overcome the functional redundancy. As another point of attack, cell lines can be generated from these mice; RT-PCR analysis can then be used on wild type and targeted cells to identify genes which may be regulated by these kinases. Thus these cell lines may provide a useful tool for the biochemical analysis of the kinases.

5. Conclusion

Recent evidence suggests a role for murine Clk kinases in the regulation of pre-mRNA alternative splicing. Since alternative splicing is a major mechanism regulating gene expression, the question was asked whether Clk kinases would be involved in regulating the expression of genes involved in development and differentiation. Clk1-null mice have been generated by gene-targeting experiments in an attempt to answer this question; however, no developmental defects have been identified to date. Perhaps Clk1 is non-essential in development; or loss of Clk1 function can be rescued by family members of Clk kinases. The work presented in this thesis is directed towards the generation of Clk2 $-/-$ mice by gene targeting experiments.

Murine Clk2 was isolated from a D3 genomic library and characterized by sequencing and mapping. Four exons have been identified, two coding for the first 133 amino acids of Clk2, and two which code for part of subdomains V and VI of the kinase region (a.a. 225-300). The organization of the Clk2 locus will be defined through further mapping and sequencing of genomic phage clones.

A 581 bp partial cDNA for Clk2 was also isolated from P19 cells, and the predicted amino acid sequence exhibits 99% identity to hClk2. Ongoing experiments are being done to isolate full-length mClk2 cDNA for biochemical and functional analysis.

Gene targeted disruption of Clk2 was attempted by insertion into the exon containing the putative initiating codon. To date screening of about 100 clones by Southern analysis has revealed only one potential targeted clone (Clk2 6B). This will have to be verified by

Southern analysis with 3' and internal probes as well. In the meantime, more clones will be selected and screened. Once targeted clones have been identified, they will be used to generate chimeras for the long term goal of producing Clk2 ^{-/-} mice to delineate the consequences of loss-of-function of Clk2 in the mouse model.

References

- Anderson, S.K., Gallinger, S., Roder, J., Frey, J., Young, H.A., and Ortaldo, J.R. (1993) A cyclophilin-related protein involved in the function of natural killer cells. Proc. Natl. Acad. Sci. USA, 90: 542-546
- Becker, W., Kentrup, H., Heukelbach, J., and Joost, H-G. (1996) cDNA cloning and characterization of rat Clk3, a LAMMER kinase predominately expressed in testis. Biochim. Biophys. Acta., 1312: 63-67
- Ben-David, Y., Letwin, K., Tannock, L., Bernstein, A., and Pawson, T. (1991) A mammalian protein kinase with potential for serine/threonine and tyrosine phosphorylation is related to cell cycle regulators. EMBO J., 10: 317-325
- Bender, J. and Fink, G.R. (1994) AFC1, a LAMMER kinase from *Arabidopsis thaliana*, activates STE12-dependent processes in yeast. Proc. Natl. Acad. Sci. USA, 91: 12105-12109
- Boise, L.H., Gonzalaz-Garcia, M., Postema, C.E., Ding, L., Lindsten, T., Turka, L.A., Mao, X., Nunez, G., and Thompson, C.B. (1993) *bcl-x*, a *bcl-2*-related gene that functions as a dominant regulator of apoptotic cell death. Cell, 74: 597-608
- Burd, C.G., and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. Science, 265: 615-621
- Burtis, K.C. (1993) The regulation of sex determination and sexually dimorphic differentiation in *Drosophila*. Curr. Opin. Cell Biol., 5: 1006-1014
- Caceres, J.F., Stamm, S., Helfman, D.M., and Krainer, A.R. (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. Science. 265: 1706-1709
- Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C., and Duncan, P.I. (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. EMBO J., 15: 265-275
- Coschigano, K.T., and Wensink, P.C. (1993) Sex-specific transcriptional regulation by the male and female doublesex proteins of *Drosophila*. Genes Dev., 7: 42-54
- Duncan, P.I., Stojdl, D.F., Marius, R.M., and Bell, J.C. (1997) In vivo regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. Mol. Cell. Biol., 17: 5996-6001

- Duncan, P.I., Howell, B.W., Marius, R.M., Drmanic, S., Douville, E.M.J., and Bell, J.C. (1995) Alternative splicing of STY, a nuclear dual specificity kinase. J. Biol. Chem., 270: 21524-21531
- Evans, M.J., and Kaufman, M.H. (1981) Establishment in culture of pluripotential cells from mouse embryos. Nature, 292: 154
- Fields, S., and Song, O-K. (1989) A novel genetic system to detect protein-protein interactions. Nature, 340: 245-246
- Fu, X.D., Mayeda, A., Maniatis, T., and Krainer, A.R. (1992) General splicing factors SF2 and SC35 have equivalent activities in vitro, and both affect alternative 5' and 3' splice site selection. Proc. Natl. Acad. Sci. U S A, 89: 11224-11228
- Ge, H., Zuo, P., and Manley, J.L. (1991) Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. Cell, 66: 373-382
- Ge, H., and Manley, J.L. (1990) A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. Cell, 13: 25-34
- Gross, T., Lutzberger, M., Wiegmann, H., Klingenhoff, A., Shenoy, S., and Kaufer, F. (1997) Functional analysis of the fission yeast Prp4 protein kinase involved in pre-mRNA splicing and isolation of a putative mammalian homologue. Nucleic Acids Res., 25: 1028-1035
- Gui, J-F., Lane, W.S., and Fu, X-D. (1994a) A serine kinase regulates intracellular localization of splicing factors in the cell cycle Nature, 369: 678-682
- Gui, J-F., Tronchere, H., Chandler, S.D., and Fu, X-D. (1994b) Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. Proc. Natl. Acad. Sci. USA, 91: 10824-10828
- Gunning, K.B., Cohn, S.L., Tomlinson, G.E., Strong, L.C., and Huff, V. (1996) Analysis of possible WT1 RNA processing in primary Wilms tumors. Oncogene, 13:1179-1185
- Hanes, J., von der Kammer, H., Kludiny, J., and Scheit, K.H. (1994) Characterization by cDNA cloning of two new human protein kinases. Evidence by sequence comparison of a new family of mammalian protein kinases. J. Mol. Biol., 244: 665-672
- Hanks, S.K., and Quinn, A.M. (1991) Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. Methods Enzymol 200: 38-62

Hardie, G., and Hanks, S. (ed.) (1995) The Protein Kinase Facts Book: Protein-Serine Kinases (Academic Press Limited, London)

Hartmann, T., Bergsdorf, C., Sandbrink, R., Tienari, P.J., Multhaup, G., Ida, N., Bieger, S., Dyrks, T., Weidemann, A., Masters, C.L., and Beyreuther, K. (1996) Alzheimer's disease betaA4 protein release and amyloid precursor protein sorting are regulated by alternative splicing. J Biol Chem 271:13208-13214

Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H., and Shimura, Y. (1991) Control of doublesex alternative splicing by transformer and transformer-2 in *Drosophila*. Science, 252: 833-836

Howell, B.W., Afar, D.E., Lew, J., Douville, E.M., Icely, P.L.E, Gray, D.A., and Bell, J.C. (1991) STY, a tyrosine-phosphorylating enzyme with sequence homology to serine/threonine kinases. Mol. Cell. Biol., 11: 568-572

Huang, Y., and Carmichael, G.C. (1996) Role of polyadenylation in nucleocytoplasmic transport of mRNA. Mol. Cell. Biol. 16:1534-1542

Jimenez-Garcia, L.F., and Spector, D.L. (1993) In vivo evidence that transcription and splicing are coordinated by a recruiting mechanism. Cell, 73, 47-59

Johnson, K.W., and Smith, K.A. (1991) Molecular cloning of a novel human cdc2/CDC28-like protein kinase J. Biol. Chem., 266: 3402-3407

Joyner, A.L. (ed.) (1993) Gene Targeting: A Practicable Approach (Oxford University Press, New York)

Kornblihtt, A.R., Pesce, C.G., Alonso, C.R., Cramer, P., Srebrow, A., Werbajh, S., and Muro, A.F. (1996) The fibronectin gene as a model for splicing and transcription studies. FASEB J. 10: 248-257

Krainer, A.R., Mayeda, A., Kozak, D., and Binns, G. (1991) Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. Cell, 66: 383-394

Krainer, A.R., Conway, G.C., and Kozak, D. (1990) Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. Genes Dev., 4: 1158-1171

Krainer, A.R., and Maniatis, T. (1985) Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. Cell, 42: 725-736

- Krawczak, M., Reiss, J., and Cooper D.N. (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum. Genet., 90: 41-54
- Lewis, J.D., and Izaurralde, E. (1997) The role of the cap structure in RNA processing and nuclear export. Eur. J. Biochem. 247: 461-469
- Mermoud, J.E., Cohen, P.T.W., and Lamond, A.I. (1994) Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. EMBO J., 13: 5679-5688
- Mermoud, J.E., Cohen, P.T.W., and Lamond, A.I. (1992) Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. Nucleic Acids Res., 20: 5263-5269
- Misteli, T., and Spector, D.L. (1996) Serine/threonine phosphatase 1 modulates the subnuclear distribution of pre-mRNA splicing factors. Mol. Biol. Cell., 7:1559-1572
- Mountford, P., Zevnik, B., Duwel, A., Nichols, J., Li, M., Dani, C., Robertson, M., Chambers, I., and Smith, A. (1994) Dicistronic targeting constructs: reporters and modifiers of mammalian gene expression. Proc. Natl. Acad. Sci. USA., 91: 4303-4307
- Myers, M.P., Murphy, M.B., and Landreth, G. (1994) The dual-specificity CLK kinase induces neuronal differentiation of PC12 cells. Mol. Cell. Biol., 14: 6954-6961
- Nestel, F.P., Colwill, K., Harper, S., Pawson, T., and Anderson, S.K. (1996) RS cyclophilins: identification of an NK-TR1-related cyclophilin. Gene., 180: 151-155
- Padmanabha, R., Gehrung, S., and Snyder, M. (1991) The KNS1 gene of *Saccharomyces cerevisiae* encodes a nonessential protein kinase homologue that is distantly related to members of the CDC28/cdc2 gene family. Mol. Gen. Genet., 229: 1-9
- Parkhurst, S.M., Bopp, D., and Ish-Horowicz, D. (1990) X:A ratio, the primary sex-determining signal in *Drosophila*, is transduced by helix-loop-helix proteins Cell., 63: 1179-1191
- Rabinow, L. and Birchler, J.A. (1989) A dosage-sensitive modifier of retrotransposon-induced alleles of the *Drosophila* white locus. EMBO J., 8: 879-889
- Rosenberg, G.H., Alahari, S.K., and Kaufer, N.F. (1991) prp4 from *Schizosaccharomyces pombe*, a mutant deficient in pre-mRNA splicing isolated using genes containing artificial introns. Mol. Gen. Genet., 226: 305-309
- Rossi, F., Labourier, E., Forne, T., Divita, G., Derancourt, J., Riou, J.F., Antoine, E.,

- Cathala, G., Brunel, C., and Tazi, J. (1996) Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. Nature, 381: 80-82
- Roth, M.B., Murphy, C., and Gall, J.G. (1990) A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. J. Cell Biol. 111: 2217-2223
- Seraphin, B., and Rosbash, M. (1989) Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. Cell, 59: 349-358
- Sessa, G., Raz, V., Savaldi, S., and Fluhr, R. (1996) PK12, a plant dual-specificity protein kinase of the LAMMER family, is regulated by the hormone ethylene. Plant Cell, 8: 2223-2234
- Simms, L.A., Algar, E.M., and Smith, P.J. (1995) Splicing of exon 5 in the WT1 gene is disrupted in Wilms' tumour. Eur. J. Cancer, 31A:2270-2276
- Spector D.L. (1993) Nuclear organization of pre-mRNA processing. Curr. Opin. Cell Biol. 5: 442-447
- Sun, Q., Mayeda, A., Hampson, R.K., Krainer, A.R., and Rottman, F.M. (1993) General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. Genes Dev., 7: 2598-2608
- Tacke, R., Chen, Y., and Manley, J.L. (1997) Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. Proc. Natl. Acad. Sci. USA, 94: 1148-1153
- Takeuchi, M., and Yanagida, M. (1993) A mitotic role for a novel fission yeast protein kinase *dsk1* with cell cycle stage dependent phosphorylation and localization. Molec. Biol. Cell, 4: 247-260
- Tazi, J., Daugeron, M-C., Cathala, G., Brunel, C., and Jeanteur, P. (1992) Adenosine phosphorothioates (ATP alpha S and ATP tau S) differentially affect the two steps of mammalian pre-mRNA splicing. J. Biol. Chem. 267: 4322-4326
- Valcarcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA Science 273: 1706-1709
- Valcarcel, J., and Green, M.R. (1996) The SR protein family: pleiotropic functions in pre-mRNA splicing. TIBS, 21: 296-301

Valcarcel, J., Singh, R., Zamore, P.D., and Green, M.R. (1993) The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. Nature, 362: 171-175

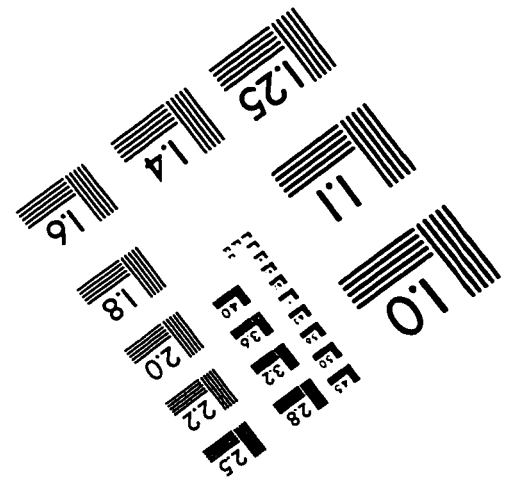
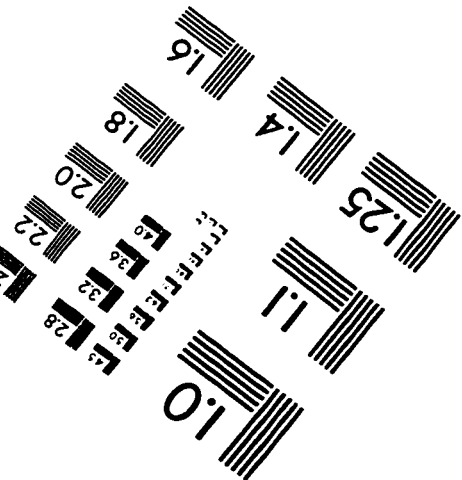
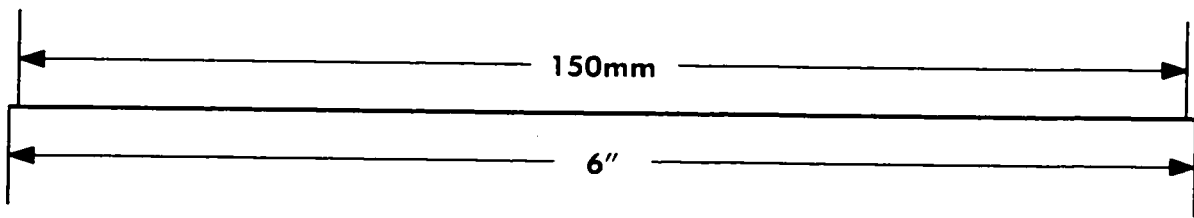
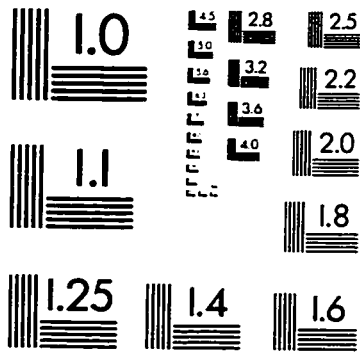
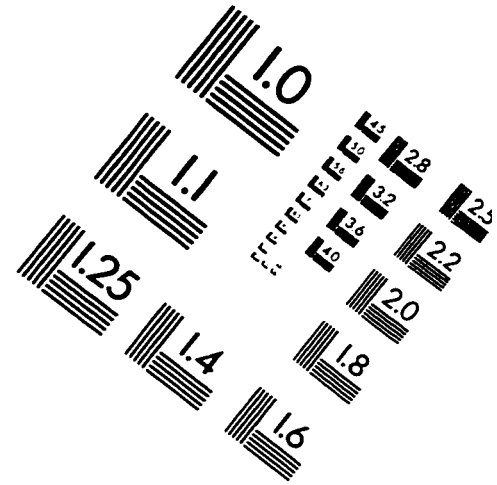
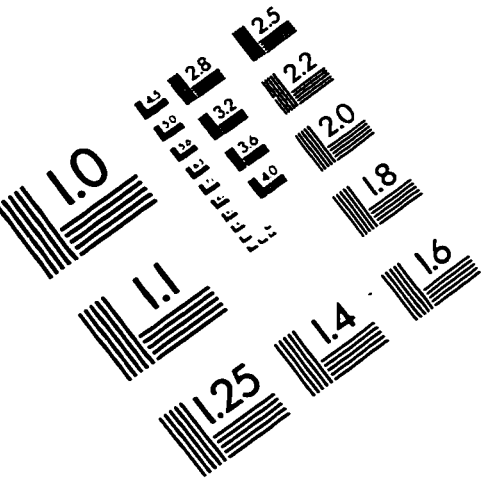
Wu, J.Y., and Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell, 75: 1061-1070

Yun, B., Farkas, R., Lee, K., and Rabinow, L. (1994) The Doa locus encodes a member of a new protein kinase family and is essential for eye and embryonic development in *Drosophila melanogaster*. Genes Dev., 8: 1160-1173

Zahler, A.M., and Roth, M.B. (1995) Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. Proc. Natl. Acad. Sci. USA, 92: 2642-2646

Zahler, A.M., Lane, W.S., Stolk, J.A., and Roth, M.B. (1992) SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev., 6: 837-847

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE, Inc
 1653 East Main Street
 Rochester, NY 14609 USA
 Phone: 716/482-0300
 Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved