Regulation of Alternative Splicing In the Rat Fibronectin Gene

by

Gene Simon Huh

B. Sc., Biochemistry
Dalhousie University, 1987

Submitted to the Department of Biology
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy in Biology

at the

Massachusetts Institute of Technology

May 1994

Signature of Author .....

Gene Huh
Department of Biology
May, 1994

Certified by .....

Richard O. Hynes
Professor, Department of Biology
Thesis Supervisor

Accepted by .....

Frank Solomon
Professor, Department of Biology
Chairman, Biology Department Graduate Committee

Regulation of Alternative Splicing In the Rat Fibronectin Gene

by

Gene Simon Huh

Submitted to the Department of Biology
on May, 1994 in partial fulfillment of the requirements for the
Degree of Doctor of Philosophy in Biology

ABSTRACT

The pre-mRNA of the rat fibronectin gene is alternatively spliced in a cell-type-specific fashion. The EIIIB exon in this gene is spliced into fibronectin mRNAs expressed in early embryos but is excluded from those expressed in adult liver. In these studies, the regulated inclusion of this exon was examined in detail. Using minigenes derived from the EIIIB region of the fibronectin gene, a number of sequence elements were found to alter the alternative splicing of EIIIB. Initial results indicated that EIIIB is an inefficiently recognized exon and that regulated inclusion requires a balance in 5' splice site strength between EIIIB and the exon upstream. In addition, intron sequences downstream of EIIIB were required for cell-type-specific EIIIB inclusion. Repeated copies of a hexanucleotide sequence (TGCATG) were present in this EIIIB-activating element. Further experiments established that repeated hexamers alone could account for most if not all intron element activity; furthermore, these hexamers exerted cell-type-dependent effects when placed in heterologous alternatively spliced genes. Therefore this hexamer sequence represents a major determinant of cell-type-specificity in the regulation of EIIIB alternative splicing. Repeated copies of this hexamer sequence were capable of activating usage of an upstream 5' splice site and could also suppress downstream 3' splice site usage. This hexamer was also found within elements implicated in the regulated splicing of the *c-src* and calcitonin/CGRP genes. In fact, repeated hexamers were found to influence calcitonin/CGRP alternative RNA processing in a manner consistent with a normal role of these repeats in the regulation of that gene. Using a UV-crosslinking assay, a number of factors in mammalian nuclear extracts were found to interact specifically with RNAs containing repeated hexamers. Taken together, these findings may provide insights regarding the regulation of mammalian alternative splicing in general.

Thesis Supervisor: Dr. Richard O. Hynes

Title: Professor of Biology

CURRICULUM VITAE

GENE SIMON HUH

Birth date: July 27, 1965

Place of Birth: Halifax, Nova Scotia, Canada

Education: Bachelor of Science, Biochemistry, first class honours, Dalhousie University, Halifax, Canada, 1987.

Ph.D., Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, May 1994.

Honors: Howard Hughes Medical Institute Predoctoral Fellowship, June 1, 1988 to June 1, 1993.

Natural Sciences and Engineering Research Council of Canada (NSERC) Undergraduate Research Awards, summers of 1985 and 1986.

University Undergraduate Academic Scholarships:
Lockward Memorial Scholarship, 1983-1984
Sir William Young Scholarship, 1984-1985
Alan Pollock Scholarship, 1985-1986
Ross Stewart Smith Scholarship, 1986-1987

Employment: Graduate student in the laboratory of Dr. Richard O. Hynes, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1988- present. Research on the alternative splicing of EIIIB in the rat fibronectin gene.

Exchange student/technician, Division of Technology for Society, Netherlands Organization for Applied Scientific Research (TNO), Zeist, The Netherlands, summer of 1987. Biotechnological research.

NSERC Undergraduate Research, Department of Biochemistry, Dalhousie University, summers of 1985, 1986. Research on ribosomal RNA gene structure in plants.

Computer programming and software development/computer assisted instruction: Department of Chemistry, Dalhousie University, summer, 1984.

Publications:

Coulthart, M. B., Huh, G. S., and Gray, M. W. (1990). Physical organization of the 18S and 5S ribosomal RNA genes in the mitochondrial genome of rye (Secale cereale L.). Curr. Genet. 17: 339-346.

Huh, G. S., and Hynes, R. O. (1993). Elements regulating an alternatively spliced exon of the rat fibronectin gene. Mol. Cell. Biol. 13: 5301-5314

Coulthart, M. B., Spencer, D. F., Huh, G. S., and Gray, M. W. (1994). Polymorphism for ribosomal RNA gene arrangement in the mitochondrial genome of fall rye (Secale cereale L.). Submitted.

Huh, G. S., and Hynes, R. O. (1994). Regulation of alternative splicing by a novel repeated hexanucleotide element. Submitted.

DEDICATION

To my family,

Dad,

Mom,

Mia,

Rena,

and

Peggy.

ACKNOWLEDGEMENTS

I am truly amazed that grad school went by so quickly. I mean, I don't really FEEL like I learned anything new; I mean, I feel just as mentally attenuated now as I did six years ago. Every now and then, however, I realize that I think, feel, say and do things now that would have just NEVER occurred to me as a first year. I guess it was mostly an osmotic thing.

Oh yeah, and I learned some science too. No, really.

A great deal of thanks go to my advisor Richard Hynes. His reason, succor, patience and subtle (yet effective) motivational techniques helped me get over many a grad student hurdle. His decision to take on a student whose primary interest was splicing (a field outside his own) has led to the development of the story that you now hold in your hands and are only beginning to read, a story that I feel was well worth the effort. I also thank Richard for fostering a great lab atmosphere that has been really fun to work in during these past years. Part of this has been due to his eye for superb lab people, of which more later.

I also wish to thank the members of my thesis committee (Phillip Sharp, David Housman, Uttam RajBhandary and Michael Green) for contributing comments and insights to this project. In particular, I thank Phil Sharp for not only providing valuably constructive criticism and advice but also for supplying a lab of local splicing aficionados (Charles Query, Melissa Moore, Anna Gil, John Crispino, Ben Blencowe, Patrick McCaw and Sandy Gilbert) with whom I have had many enjoyable discussions. Occasionally, we even talked about splicing, too. In addition, as a former TA for David Housman and Richard Mulligan's project lab course, I have had the pleasure of experiencing the Tao of the multipart ligation; some of the constructs in this thesis might otherwise have been quite off-putting.

The lab has nearly completely turned over since I arrived; the only person left from the old days is the lab manager Jane Trevithick, who runs the lab with an efficiency that constantly amazes. I owe a lot to her patience, good nature and seemingly infinite tolerance for guys like me ('nuff said). The same can be said for the lab's administrative maven, Colleen Leslie, who always seems to have the time to help you out without dropping anything else in her massive workload. I am also thankful for the intellectual and social company of past and present baymates Pamela Norton, Elisabeth Georges and Bernhard Bader. In addition to these good souls, much of what has made the lab enjoyable is due in part to lunchtime conversations, lab outings and seemingly innumerable baby showers with the likes of past (Doug DeSimone, Charles ffrench-Constant, Betsy George, Jun-Lin Guan, Mary Jalenak, Paul Johnson,

Tanya Mayadas-Norton, Ramila Patel-King, John Peters, Susan Zusman) and present (Laird Bloom, Steve Cornwall, Mike DiPersio, Phil Gotwals, Steph Paine-Saunders, Glenn Radice, Stephen Robinson, Karen Stark, Nabil Tawil, Grant Wheeler, Joy Yang) lab dudes. Many of these people have had the misfortune of reading my early attempts at scientific writing; to these people I extend both my apologies and gratitude for their golden grace and constructive comments. And for letting me sleep in the coffee room.

And, of course, there are those who toiled with me in the pursuit of similar academic goals. Looking back, I will fondly remember the companionship of (and occasional commiseration with) fellow grad students Gene Yee, whose nighttime/early morning lifestyle I often shared, and Mark Borowsky, whose flair for the practical (and the joke) has both lightened and enlightened many a moment. I will also miss the company of floormates from the Solomon and Robbins labs. In addition, I have drawn considerable support from (and hope to sustain) friendships developed over the years with Keith Ireton, Richard Josephson, Catherine and David Kirkpatrick, John LeDeaux, Scott Lowe, Ken Mungan, Chris Siebel, Bettina Winckler, Jim Yeadon and Tau-Mu Yi. Moreover, I owe a certain measure of sanity to the Charlie's Tap Beer Club, the Acid Blobs and the Big Sticks (y'all know who you are). I will always treasure the friendship of my high school buddy Stefan Jürgens (with whom much time was spent while convalescing in Halifax).

The computer on which the lion's share of this thesis was written is a Macintosh LC575; it belongs to a certain Peggy Kolm, whose noble contributions I can't even begin to describe. Her love, patience, understanding and immeasurable support, in addition to keeping me humble these past few years, has by some miracle persisted throughout the writing of this document. I think it's safe to say that this task would have been exceedingly more arduous without her. I can only hope that I can return this favor in kind when her turn comes.

It can be extraordinarily comforting to know that one has a haven to which one can occasionally retreat from the rigors of the scientific method. For me, this takes the form of a certain white house (with black shutters) in Halifax, Nova Scotia. Mom and Dad put up with a fair amount from me during my upbringing. To dedicate this thesis to them can only begin to repay their unconditional devotion, concern and love.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter One

General Introduction

GENERAL INTRODUCTION

A metazoan organism generates an amazing diversity of cell types during its development and lifetime. Generation and maintenance of these distinct cell types requires the selective expression of a common genetic repertoire. Cell-type-specific gene expression is controlled at several levels; of these, regulation at the post-transcriptional level via alternative mRNA splicing is particularly potent. Differential splicing represents a powerful means of augmenting a single gene's coding capacity, since it permits the manufacture of several related yet distinct proteins from a single pre-mRNA transcript via excision of different intervening sequence combinations. In higher eukaryotes, the catalogue of known alternatively spliced genes approaches encyclopedic proportions; splicing in many of these genes is regulated in a cell-type-specific or developmentally-dependent manner. The regulation of alternative splicing is therefore a fundamental process in the biology of the eukaryotic nucleus.

This phenomenon suggests that under appropriate circumstances the splicing mechanism exhibits a pliancy that can be manipulated in order to vary a gene's output. The ways in which this is accomplished are varied, plentiful and relatively unclear despite recent advances in the understanding of nuclear pre-mRNA splicing. The first section of this introduction briefly outlines current knowledge regarding the general mechanism of nuclear pre-mRNA splicing. The second part addresses the nature of alternative splicing mechanisms with a consideration of how *cis*- and *trans*-acting parameters can influence splice site selection. These two sections will by nature overlap, since studies of splicing mechanisms can provide insights into alternative splicing and vice versa. The third section focuses upon the fibronectin gene, which is

alternatively spliced in a cell-type-specific fashion and is the primary gene of focus in this thesis.

The mechanism of nuclear pre-mRNA splicing

An understanding of the basic splicing reaction has been achieved through a combination of biochemical and genetic analyses. These topics have been extensively reviewed (Moore et al., 1993; Green, 1991; Guthrie, 1991). Nuclear pre-mRNA splicing occurs in two chemical transesterification steps. In the first step, the 2' hydroxyl of a nucleotide some tens of bases upstream from the 3' splice site (the "branch" nucleotide, often an adenosine) attacks the phosphate group at the 5' splice site, generating a free 5' exon and a "lariat" intermediate. This lariat, in which the 5' end of the intron is covalently linked to the branch residue via a 2'-5' bond, contains the intron and 3' exon. The 3' hydroxyl of the 5' exon intermediate then attacks the phosphate group at the 3' splice site in the second step, generating the ligated exons and a lariat intron as products.

The importance of sequences at the intron boundaries for splicing has been demonstrated both by surveys of intron sequences and by analyses of directed and natural mutations. The metazoan consensus at the 5' end of the intron (the 5' splice site or 5'SS) is MAG/GURAGU (slash denotes the exon-intron junction); the GU at the intron junction is conserved in nearly all introns. The 3' end of the intron (the 3' splice site or 3'SS) contains at least three ordered elements: a very highly conserved 3' terminal dinucleotide (AG), the branch site (loose consensus YNYURAY; the underlined A is the branch) and a pyrimidine-rich sequence between the branchpoint and 3' AG. In yeast (and to some extent in *Drosophila*), a pyrimidine tract is dispensable; in addition the 5'SS and branchpoint sequences in yeast are more stringently conserved than in

vertebrates (G/GUAUGU and UACUAAC respectively). It is clear that these sequences are important determinants for splicing; however, they are not sufficient to ensure accurate splice site identification in metazoans.

The development of cell extracts that accurately splice exogenous pre-mRNA substrates have permitted a detailed scrutiny of splicing biochemistry. Such studies have revealed that splicing takes place in a large, multicomponent complex termed the "spliceosome", which is formed by the ordered association of the pre-mRNA with small nuclear ribonucleoprotein particles (snRNPs) U1, U2, U4, U5 and U6, as well as with many non-snRNP proteins. Spliceosome assembly *in vitro* (and probably *in vivo*) occurs on an RNA coated by RNA-binding proteins, many of which normally comprise heterogeneous nuclear ribonucleoprotein particles (hnRNPs; for review see Dreyfuss et al., 1993). Prior to the first step in splicing, U1 associates with both the 5'SS and branchpoint sequences in an ATP-independent fashion. An ATP-dependent U2-branchpoint interaction then occurs, followed by ATP-dependent association of U4/U6 and U5. Spliceosome assembly involves a complex network of pre-mRNA-snRNP and snRNP-snRNP interactions; many of these have been established via genetic suppression studies in yeast coupled with chemical crosslinking experiments carried out *in vitro*. These interactions probably reflect dynamic RNA structural rearrangements that are crucial for catalysis, an event which may in fact be RNA-mediated (for reviews, see Moore et al., 1993; Guthrie, 1991; Steitz, 1992; Wise, 1993).

In addition to the snRNPs, numerous non-snRNP proteins participate at various stages in spliceosome assembly. A number of splicing factors have been identified based on their ability to activate splicing in an extract rendered incompetent via immunodepletion or biochemical fractionation. These include U2 auxiliary factor (U2AF; Ruskin et al., 1988; Zamore and Green, 1991), SC35

(Fu and Maniatis, 1992a), SF2/ASF (Krainer et al., 1990) and a number of other factors that remain to be characterized (Krainer and Maniatis, 1985; Krämer and Utans, 1991). U2AF is required for the U2-branchpoint interaction and binds to the pyrimidine tract at the 3'SS (Ruskin et al., 1988); it exists as a 65 kDa/35 kDa heterodimer, the 65kDa subunit being sufficient for splicing *in vitro* (Zamore et al., 1991). The U2AF65 protein contains three copies of an RNA-recognition motif (RRM), a domain common to a large number of RNA-binding proteins (Kenan et al., 1991), as well as an N-terminal domain that is rich in arginine and serine residues. The RRMs were found to be required for RNA-binding; the RS domain was found to be dispensable for binding RNA but required for splicing *in vitro* (Zamore et al., 1992).

SC35 and SF2/ASF belong to a larger class of splicing factors termed SR proteins. These proteins contain one RRM (and sometimes a second RRM-like domain) and a C-terminal region rich in arginine-serine dipeptides (Fu and Maniatis, 1992a; Zahler et al., 1992, 1993b). SR proteins are highly phosphorylated as isolated from several sources; any of the SR proteins can complement a splicing-defective S100 extract for splicing (Zahler et al., 1992, 1993b; Mayeda et al., 1992); several can promote proximal 5'SS selection *in vitro* (Mayeda et al., 1992; Zahler et al., 1993a); in fact SF2/ASF was independently identified based on this latter activity (Ge et al., 1990). SR proteins can play a role very early in splice site selection and commitment (Fu, 1993). SF2/ASF can cooperate with U1 snRNP in binding to a 5'SS *in vitro* (Kohtz et al., 1994), possibly via SR domain-dependent interactions with the U1-specific 70K protein (Wu and Maniatis, 1993). SC35 has also been shown to interact with both the 5'SS and 3'SS (Fu and Maniatis, 1992b), probably by interacting with both U1-70K and the 35 kDa subunit of U2AF (Wu and Maniatis, 1993),

suggesting that SC35 (and perhaps SR proteins in general) may participate in early splice site pairing.

Other proteins have been identified as candidate splicing factors by virtue of their affinity for splice site-like sequences. Notably, a number of hnRNP proteins (A1, C, I and D) can bind to pyrimidine-rich or other 3'SS-associated sequences (Swanson and Dreyfuss, 1988b; Garcia-Blanco et al., 1989; Patton et al., 1991). Antibodies to hnRNPs can inhibit splicing *in vitro* (Choi et al., 1986; Sierakowska et al., 1986); however, reconstitution of splicing by readdition of pure hnRNP to depleted extracts has not been possible to date, possibly because removal of these abundant proteins may co-deplete other essential factors. In any event, the role of hnRNPs in splicing remain unclear. Many hnRNPs can be stripped off preformed spliceosomes by high salt (Bennett et al., 1992a); however, this does not preclude an early role for hnRNPs in splicing. Two-dimensional electrophoretic analyses of purified spliceosomes have identified an assortment of spliceosome-associated proteins (SAPs; Bennett et al., 1992a). A number of these have been found to correspond to PRP genes in yeast (Brosi et al., 1993; Bennett and Reed, 1993), thus validating this type of biochemical approach in the identification of splicing components.

Factors important for splicing have also been identified by genetic approaches: a number of pre-RNA processing genes in *S. cerevisiae* (PRP genes; reviewed by Ruby and Abelson, 1991) have been characterized in detail. A number of these genes, interestingly, encode proteins with motifs characteristic of ATP-dependent RNA helicases (reviewed by Guthrie, 1991).

A number of splicing components have been localized to discrete regions in the mammalian nucleus. Staining using snRNP or SC35 antisera results in punctate nuclear staining, within a set of 30-50 "speckles" (by

immunofluorescence) or to regions containing interchromatin granules (by electron microscopy; Spector, 1990; Spector et al., 1991). U2AF and snRNPs have also been detected within "foci" or coiled bodies, of which only a few exist per nucleus (Zamore and Green, 1991; Carmo-Fonseca et al., 1991a, 1991b). These structures may represent sites in which splicing, splicing factor storage or snRNP assembly occurs; this is not yet clear, although in certain cases transcription and splicing have been observed to occur near speckled structures (reviewed by Xing and Lawrence, 1993). Mammalian subnuclear organization could play a role in events such as alternative splicing, since splicing rates could potentially be a function of pre-mRNA nuclear location.

Given the multicomponent nature of the splicing process, it would not be surprising if regulation in any given situation were to occur at any of a number of steps in splicing. In principle, alterations in relative splicing factor levels could provide the basis for cell-type-specific alternative splicing (discussed later). Alterations in *cis* (e.g., mutations of pre-mRNA splice site sequences) or in *trans* (e.g., differing levels of RNA-binding proteins*in vitro*) have been found to affect differential splice site usage. The following section details some of the parameters affecting splice site choice.

Patterns of alternative splicing

Patterns of differential splice site selection are varied (reviewed by Smith et al., 1989a; McKeown, 1992). Studies of alternative splicing in numerous systems have provided diverse (and occasionally overlapping) contributions to an understanding of how sequences within the pre-mRNA can influence splice site selection.

To splice or not to splice an intron: perhaps this represents the simplest form of alternative splicing (Figure 1-1A). Examples that have been studied in

detail include the third intron in the *Drosophila melanogaster* P element transposase gene; this intron is spliced in germline but not somatic cells (Laski et al., 1986). The 3' terminal intron in the bovine growth hormone gene is retained at a low level (Sun et al, 1993a, 1993b). In budding yeast, splicing in the MER2 gene is dependent upon MER1 expression during meiosis (Engebrecht et al., 1991). Splicing in the yeast rpL32 ribosomal protein genes is autoregulated in an apparent feedback loop (Eng and Warner, 1991).

In certain situations, a single splice site may have a number of potential pairing partners. These are also relatively simple situations, since the identity of one splice site is fixed (Figure 1-1B, C). One extensively studied system is that of the SV40 early transcript, in which two donors compete for a shared acceptor; splicing to the upstream 5'SS generates the large T antigen mRNA whereas splicing to the downstream 5'SS generates an mRNA encoding the small t antigen.

Certain "cassette" exons are either included into or excluded (skipped) from the mature mRNA (Figure 1-1D); 5'SS and 3' SS selection thus occur simultaneously. Cassette exons that have been studied include the exon 4 in the rat preprotachykinin gene, the N1 exon in the murine *c-src* gene, exon 18 in the murine N-CAM gene, exon 5 in the chicken cardiac troponin gene and the male-specific exon in the *Drosophila Sex-lethal* gene. In more complex scenarios, multiple exons in a row can exhibit cassette-type behavior (Figure 1-1E); this is the case for the human leukocyte common antigen/murine Ly-5 (CD45) genes, in which exons 4 through 6 can be coordinately included or skipped in B or T lymphocytes respectively (Streuli and Saito, 1989; Saga et al., 1990).

Sometimes only one of a pair of adjacent exons are included into mRNA, a situation termed mutually exclusive splicing (Figure 11F). This type of splicing has been studied primarily in the tropomyosin genes: examples

include exon pairs 2/3 in the rat α-tropomyosin gene, $5^{NM}/5^{SK}$ in the human α-tropomyosin gene and also in the β-tropomyosin gene (pairs 6/7 and 6A/6B in rat and chicken genes respectively). For mutually exclusive exons, coincident exon skipping or inclusion is apparently forbidden; patterns of splice site selection are evidently restricted, since the inclusion of one exon must preclude inclusion of the other. Since translational frameshifts resulting from incorrect splicing may generate aberrant products that interfere with normal cell function, mechanisms must operate in order to allow the cytoplasmic appearance of only correctly spliced mRNAs. Several observations have suggested that upstream nonsense codons result in reduced steady state mRNA levels (Urlaub et al., 1989; Barker and Beemon, 1991; Cheng and Maquat, 1993; Pulak and Anderson, 1993); this phenomenon may in some cases explain why only correctly spliced mutually exclusive mRNAs are detected. For exons 2/3 in the α-tropomyosin gene, dual exon inclusion is prevented by the prohibitively close juxtaposition of the 5'SS and branchpoint between these two exons (Smith and Nadal-Ginard, 1989).

In cases where alternative splicing occurs at a gene's 5' terminus, differential splicing can often occur in combination with differential promoter usage. The basis for regulation in these situations is more complex, since splicing may not necessarily represent the determining event. Alternative splicing in the rat myosin alkali light chain gene (MLC1/3) results in inclusion of either exon 4 (making MLC1) or 3 (making MLC3), depending upon whether transcription begins in exons 1 or 2 respectively (Figure 1-1G). In this instance, differential promoter usage (resulting in different pre-mRNA sequences in *cis*) appears to be sufficient for determining exon choice (Gallego and Nadal-Ginard, 1990).

Figure 1-1.

Patterns of alternative splicing

Numbered horizontal rectangles and lines represent exon and intron sequences, respectively. Dashed diagonal lines above and below each exon/intron diagram show possible splicing combinations.

A. Intron retention. (5'ss, 5' splice site; 3'ss, 3' splice site)

B. 3' splice site selection.

C. 5' splice site selection.

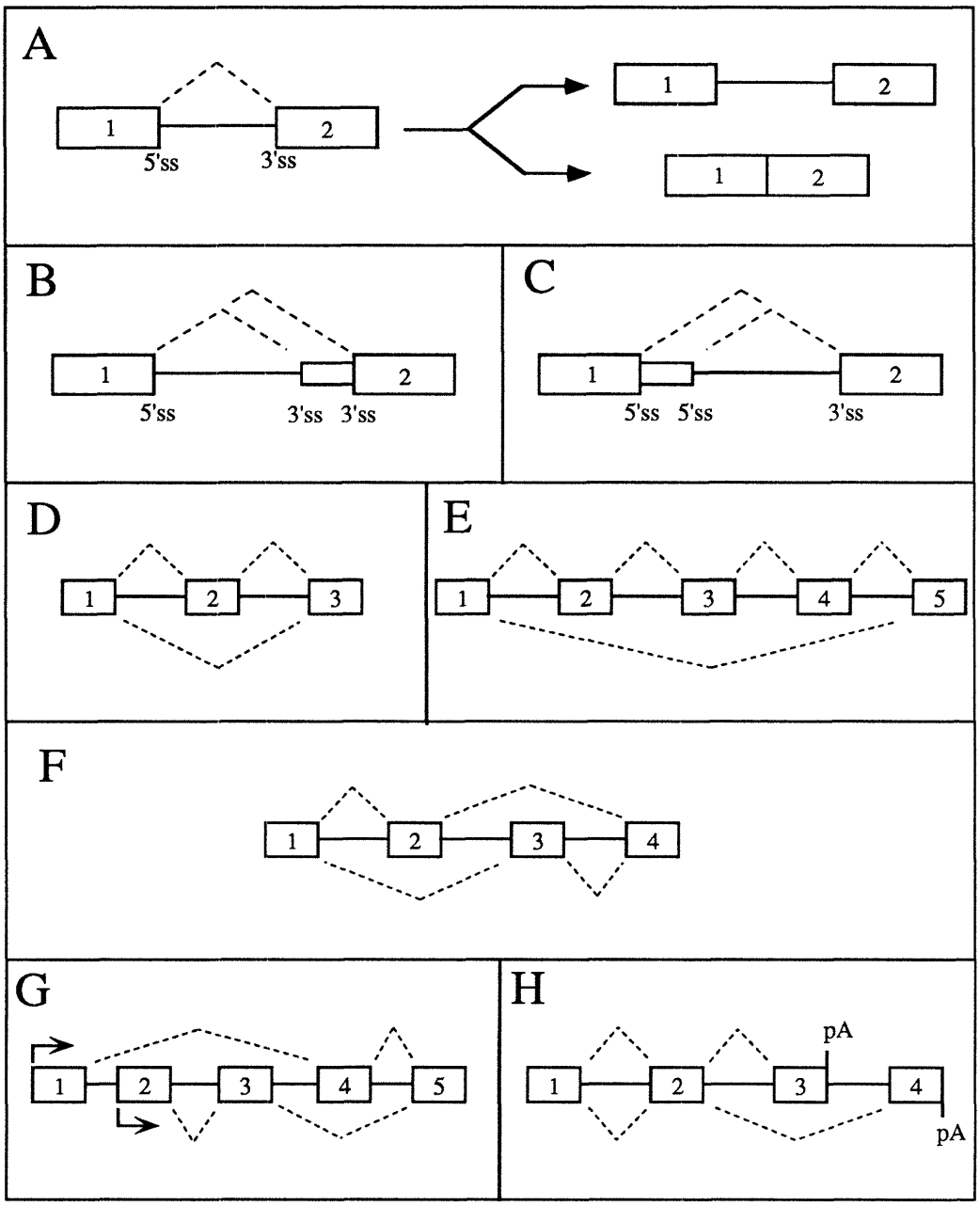
D. Optional "cassette" exons (exon skipping).

E. Multiple exon skipping.

F. Mutually exclusive exon pairs.

G. Alternative exon selection coupled with differential promoter selection (arrows represent transcription start sites).

H. Differential 3' terminal exon usage coupled with 3' end formation and polyadenylation (pA, polyadenylation site).


A
1
2
5'ss
3'ss
1
2
1
2
B
1
2
5'ss
3'ss
3'ss
C
1
2
5'ss
5'ss
3'ss
D
1
2
3
E
1
2
3
4
5
F
1
2
3
4
G
1
2
3
4
5
H
pA
1
2
3
4
pA

Alternative splicing also occurs in association with alternative 3' end formation in a number of genes (Figure 1-1H): systems that have been studied in detail include the immunoglobulin IgM gene, the rat and human calcitonin/CGRP genes and the *Drosophila* sex-determination gene *doublesex*. To study RNA processing in these systems, it is often necessary to clarify whether 3' end formation or splicing is the determining event. For example, in the immunoglobulin μ heavy chain (IgM) transcription unit, usage of a polyadenylation site competes with usage of an upstream 5'SS. Whether regulation occurs at the level of splicing or transcriptional termination/3' end formation has been a matter of debate, since evidence has been presented for both modes of regulation (Galli et al., 1987; Peterson and Perry, 1986; Tsurushita and Korn, 1988). Similar questions have been considered regarding the regulation of the calcitonin/CGRP transcription unit; these will be discussed in more detail later (Chapter Four).

Sequences that affect splice site selection

Many questions remain regarding the fidelity of the splicing process. Since 5' and 3'-splice site consenses are relatively degenerate in higher eukaryotes, a long primary RNA transcript can contain many splice site-like sequences, only a small fraction of which are actually used for splicing. Therefore the cell must identify correct splice sites among the many potential candidates; mechanisms for such discrimination remain unclear. This question becomes particularly relevant when considering alternative splicing, since the rules that govern splice site discrimination apparently can vary between cell types or developmental stages.

Splice site proximity has been proposed as a simple way to avoid exon skipping. Given two competing splice sites with sufficient exonic context, the

more proximal of the two is often selected (Reed and Maniatis, 1986). However, some unexplained exceptions to this proximity rule occur, where distal selection is preferred regardless of 5'SS order (Kedes and Steitz, 1988). In addition, proximity alone cannot account for accurate splicing *in vivo*, since usage of cryptic sites within long introns must also be avoided.

The 5'SS is recognized primarily via base pairing with the 5' end of U1 snRNA; this has been established by a number of elegant genetic suppression studies in mammalian cells and yeast, where the effects of mutations at a 5'SS could be reversed by the introduction of U1 snRNA genes containing compensatory mutations (Zhuang and Weiner, 1986; Siliciano et al., 1988). However, the 5'SS sequence alone is not sufficient to dictate its selection, since even a site that binds U1 efficiently *in vitro* may not be used if present in an inappropriate context (Eperon et al., 1986; Lear et al., 1990; Nelson and Green, 1988). In certain circumstances, the mutual inhibition of closely juxtaposed donors has been observed *in vitro* , even though both sites appeared to bind U1 snRNP (Nelson and Green, 1988); therefore additional steric requirements may exist for efficient 5'SS usage. Mutual inhibition may in fact underlie the soma-specific repression of the third intron (IVS3) in the Drosophila P element transposase pre-mRNA. A number of 5'SS-like sequences occur closely upstream of the authentic 5'SS of this intron; mutation of one of these "pseudo" sites was found to activate IVS3 splicing, suggesting that this "pseudo" site participated in the somatic repression of the authentic IVS3 5'SS (Siebel and Rio, 1990; Siebel et al., 1992).

A number of additional snRNP-pre-mRNA interactions may also influence 5'SS recognition. Genetic suppression studies in yeast and crosslinking studies in mammalian extracts have suggested that U5 snRNA participates in splice donor recognition by base pairing with exon sequences

immediately upstream of the 5'SS (Newman and Norman, 1991; Wyatt et al., 1992). Similar experiments have suggested an interaction between U6 snRNA and sequences downstream of the 5'SS, an interaction which may deterimne the 5'SS cleavage point (Sawa and Shimura, 1992; Lesser and Guthrie, 1993; Sontheimer and Steitz, 1993; Kandels-Lewis and Seraphin, 1993). However, since the yeast studies assayed the use of "aberrant" 5' cleavage events located at a single region of U1 complementarity, it remains to be seen whether U5 and U6 can contribute to selection between separate donors. In any event, the precise nature of sequence context effects on 5'SS selection remains relatively ill-defined.

The importance of a suboptimal 5′ splice site has been shown for a number of natural systems; mutation of an alternative donor site toward the consensus often elevates its usage, often to the complete exclusion of its competitor (Nasim et al., 1991; Tacke and Goridis, 1991; Black, 1991; Dominski and Kole, 1992). The 5'SS sequence of exon 18 in the murine N-CAM gene exhibits some interesting properties. Its efficiency in a heterologous gene appears to correlate with regulated exon 18 inclusion, indicating that the exon 18 5'SS sequence may coincide with cell-specific elements (Tacke and Goridis, 1991). The meiosis-specific intron in the *S. cerevisiae* gene MER2 has a 5'SS sequence that deviates from consensus; the variant nature of this splice site is important for meiosis-specific regulation, since mutations that improved this site (or U1 mutations that improved base-pairing with the MER2 donor) caused MER1-independent splicing (Nandebalan et al., 1993). 5'SS quality is not rate limiting in every situation, since 5'SS up-mutations in other alternative exons have little effect (Mullen et al., 1991; Graham et al., 1992). Therefore the presence of suboptimal 5'SS sequences appears to be important for some but not all cases of differential splicing.

The relative arrangement and quality of the 3' splice site components (the branchpoint, pyrimidine-rich tract and 3' AG) can affect 3' splice site selection. The length of the pyrimidine tract, along with its proximity to the branchpoint, can play an important role in 3'SS usage *in vitro*, particularly if the branchpoint is distant from the 3'AG (Reed, 1989; Smith et al., 1989b). The quality of the branchpoint sequence can also affect splicing if no nearby "cryptic" branches are available, as this region interacts with U2 snRNP via base pairing (Zhuang and Weiner, 1989; Wu and Manley, 1989). The importance of these elements in 3'SS selection has been supported by studies of the mutually exclusive rat α-tropomyosin exons 2 and 3. Exon 2 is selected in smooth muscle whereas exon 3 is selected in other cell types. If both exons were made to compete for splicing to exon 1, exon 3 was preferred. The 3'SS of exon 3 contains a very long pyrimidine tract immediately preceded by a consensus branchpoint. The intron upstream of exon 2 contains a poor pyrimidine tract and branch sequence. 3'SS preference was found to correlate with the pyrimidine tracts preceding exons 2 and 3 and, to a lesser extent, with branch site sequences (Mullen et al, 1991). Therefore the quality of the pyrimidine tract/branchpoint region can play an important role in 3' SS selection. The distance between the branch site and the 3' splice junction of exon 3 is unusually long (176 bases), a feature shared by other alternative exons in the β-tropomyosin and fibronectin genes (Helfman and Ricci, 1989; Norton and Hynes, 1990). The significance of this unusual organization is unclear at present.

Another feature unique to some alternatively spliced exons is the usage of multiple branch sites. This is seen in SV40 early pre-mRNA alternative splicing, where selection among a set of branchpoint residues can determine alternative 5'SS usage, partly because the proximal small t 5'SS is only 66 bases

from the common 3'SS (Fu and Manley, 1987). The selection of upstream or downstream branch sites prohibits or permits small t 5'SS usage respectively (Noble and Manley, 1988). Multiple branchpoint usage *in vitro* is also exhibited by exons in the rat fibronectin (Norton and Hynes, 1990) and β-tropomyosin genes (Helfman and Ricci, 1989). This *in vitro* phenomenon may reflect the lack of any suitable branchpoint sequences at these accceptors. The kinetics of recognition and usage of these acceptors (and possibly of acceptors with lengthy pyrimidine tracts) may therefore differ significantly from those of constitutive 3' splice sites.

Coordinate recognition of splicing signals: exon definition

From the above, it is clear that that splice site usage can be influenced in a multiple ways. However, the actual mechanisms that discriminate authentic splice sites from "fake" ones still remain somewhat mysterious. A number of systems have supported a model which may supply some of these mechanisms. This model, termed "exon definition", posits a functional interaction between the RNA processing signals that border an exon. Selection of an internal exon would then proceed as a result of coordinate recognition of a 3'SS plus a 5'SS located closely downstream.

Support for this model comes from *in vitro* studies using substrates containing an intron plus a downstream 5'SS. Many of these studies suggested that the splicing of an intron was facilitated if a nearby downstream 5'SS was present (Robberson et al., 1990). In three-exon substrates, mutation of the central exon 5'SS resulted in exon skipping rather than downstream intron retention *in vitro* (Talerico and Berget, 1990). Naturally occurring 5'SS mutations also result in exon skipping rather than the generation of partially or cryptically spliced mRNA (Talerico and Berget, 1990, and references therein).

Taken together, these results suggested that the 3'SS and 5'SS of an exon are functionally coupled. These findings were supported further by studies of exon E4 in the rat preprotachykinin gene. Improvement of the E4 splice donor was found to enhance splicing of the upstream (E3-E4) intron *in vitro* (Nasim et al., 1990). Additional mutations suggested a correlation between upstream intron stimulation and the affinity of this donor for U1 snRNP (Kuo et al., 1991; Grabowski et al., 1991). Moreover, improved U1 binding to the E4 donor was found to enhance U2AF65 binding at the E4 3'SS (Hoffman and Grabowski, 1992). These biochemical studies suggested that U1 and U2AF participate in a mechanism that recognizes exons as functional units in pre-mRNAs. Some *in vivo* evidence has provided findings consistent with this model, in that the presence of one intron in a transcript may facilitate the splicing of introns adjacent (Neel et al., 1993).

Vertebrate internal exons average only 137 nt in length and are infrequently longer than 300 nt (Hawkins, 1988), a finding which may reflect exon definition requirements. However, recognition of an exon may also be hampered if the exon is too short (Dominski and Kole, 1991), possibly due to steric considerations; this has been proposed as the basis for the inefficient recognition of the 18 nucleotide *c-src* N1 exon in non-neural cells, since this exon can be activated for splicing simply by increasing the length of the N1 exon (Black, 1991). Recognition of short exons can be facilitated if relative splice site strengths are altered or if another exon is placed nearby, suggesting that additional mechanisms associated with exon definition may aid the recognition of very short exons (Dominski and Kole, 1992; Black, 1991; Sterner and Berget, 1993). It is therefore apparent that the recognition of an exon may involve multiple interactions between splice site sequences.

The effect of exon sequences

Many studies of alternatively spliced systems have described the contribution of exon sequences to splice site selection in ways that are somewhat unpredictable. In many cases, mutations or deletions of alternative exon sequences can result in a dramatic shift in splice site usage patterns. In some cases these exon sequences appear to have a positive effect upon exon selection *in vivo* whereas sequences in other alternative exons have negative effects (for examples of the former, see Mardon et al., 1986; Cooper and Ordahl, 1989; Cote et al., 1992; van Oers et al., 1994; for examples of the latter, see Streuli and Saito, 1989; Tsai et al., 1989; Graham et al., 1992). Such effects have also been observed *in vitro* (Reed and Maniatis, 1986).

A number of alternative and constitutive exons have been shown to contain elements that assist exon selection. A particularly well characterized example of these is a purine-rich sequence originally defined *in vitro* in the murine IgM M2 exon (Watakabe et al., 1993). This exon recognition sequence (ERS) was found to stimulate splicing of the upstream intron and could also activate splicing when placed in a heterologous context. Similar elements have been identified in a number of other exons including those in genes encoding troponin T (Xu et al., 1993), bovine growth hormone (Sun et al., 1993a, 1993b), fibronectin (Lavigueur et al., 1993) and the calcitonin/CGRP gene (Yeakley et al., 1993; van Oers et al., 1994). Many of these are likely to represent degenerate forms of the same element, since even synthetic purine-rich inserts can also confer ERS-like activity (Xu et al., 1993; Tanaka et al., 1994). The ERS in the M2 exon can be specifically UV-crosslinked to U1 snRNA *in vitro*, suggesting that ERS function involves U1 snRNP (Watakabe et al., 1993). In addition, studies using the bovine growth hormone terminal exon have suggested that this element can specifically bind the splicing factor SF2/ASF

(Sun et al., 1993b). Likewise, an ERS-like element in the fibronectin ED1 (EIIIA) exon has been shown to interact with purified SR proteins (Lavigueur et al., 1993). These findings are consistent with a model in which SR proteins and/or U1 bind downstream of a 3'SS and activate its usage, possibly via exon definition-related mechanisms.

The ERS, although found in many exons, is not sufficient to account for all examples of exon sequence effects. Some exon sequences appear to inhibit splicing (for examples, see Streuli and Saito, 1989; Tsai et al., 1989; Graham et al., 1992), whereas other elements have a positive effect on splicing yet lack any discernible purine-rich sequences (for examples, see van Oers et al., 1994; Cote et al., 1992). Furthermore, exon sequences that affect 5'SS selection are known (Reed and Maniatis, 1986) but have not been extensively characterized. ERS function is apparently specific for upstream 3' splice sites, since it has little effect when placed upstream of a 5'SS *in vivo* (Xu et al, 1993) or upstream of a 3'SS *in vitro* (Lavigueur et al., 1993). Therefore the effects of exon sequences are probably due to a number of factors that include ERS elements and also RNA secondary structure (see below). The role of ERS-like elements in cell-type-specific regulation is not currently clear, since this element can also be found in at least one constitutively spliced exon (Yeakley et al., 1993) and is dispensable for regulation of cardiac troponin T exon 5 (Xu et al., 1993). It is possible that ERS-like elements occur in either constitutive or cell-type-specific forms, although this has not been demonstrated.

The role of RNA secondary structure

Splice site selection in yeast can be affected by relatively short hairpins *in vivo*, indicating that basepairing interactions may play a role in splicing regulation in yeast (Deshler and Rossi, 1991; Goguel et al., 1993).

Autoregulation of RPL32 splicing may involve a hairpin that sequesters the RPL32 5'SS (Eng and Warner, 1991). A long-range base pairing interaction may aid the pairing of the 5'SS and branch region of the unusully long yeast RP51B intron (Goguel and Rosbash, 1993). Such findings suggest potential mechanisms for preferential splice site pairing in metazoans; however, although splice site selection can indeed be modulated by altering the source of flanking sites in the rat MLC1/3 gene (Gallego and Nadal-Ginard, 1990) and in adenovirus (Ulfendahl et al., 1989), the role of secondary structure in these systems has not been addressed. Whether secondary structure plays an extensive role in vertebrate splicing is still a matter of question. Although hairpins that sequester splice sites can cause exon-skipping *in vitro*, these effects were much reduced when tested *in vivo* (Solnick, 1985; Eperon et al., 1988), therefore the formation of RNA secondary structure may often be at a disadvantage relative to other processes such as hnRNP association.

One notable exception is represented by the chicken β-tropomyosin gene. Mutally exclusive selection of exons 6A and 6B occurs in nonmuscle cells and skeletal muscle respectively. Mutational analyses and structural determination *in vitro* were consistent with the presence of a secondary structure that sequestered the exon 6B 3'SS (Goux-Pelletan et al., 1990). Mutations disrupting this potential RNA secondary structure were found to activate splicing of the exon 6B 3'SS, whereas additional mutations predicted to restore basepairing were found to reestablish exon 6B repression (D'Orval et al., 1991). The influence of part of this structure was confirmed *in vivo* by transfection (Libri et al., 1991). It was proposed that this RNA structure was stabilized in non-muscle cells and/or was disrupted in muscle cells by cell-type-specific factors (D'Orval et al., 1991; Libri et al., 1991; Guo et al., 1991).

However, direct evidence for the cell-type-specificity of this structure is lacking at present.

Other sequences that affect splicing

Deletion analyses of a number of alternatively spliced genes has uncovered numerous *cis*-sequences that are capable of affecting exon or splice site selection in either a positive or negative fashion. For example, positively-acting sequences can be defined downstream of the neural-specific *c-src* N1 exon that activate neural-specific splicing *in vivo* and *in vitro* (Black, 1991, 1992). In addition to the secondary structure that surrounds the exon 6B 3'SS, there also appear to be sequences upstream of this exon that function to inhibit usage of this exon; these sequences have been identified *in vitro* in the chicken gene (Gallego et al, 1992) and also in the rat gene by transfection analysis (Guo et al., 1991). In the rat calcitonin/CGRP gene, intron sequences in the vicinity of the calcitonin-specific exon 4 acceptor are thought to mediate repression of this exon in cells that produce CGRP (Emeson et al., 1989). An element that represses Rous sarcoma virus splicing has been shown to be functional in the context of a heterologous intron (the *nrs* or negative regulator of splicing; Arrigo and Beemon, 1988; McNally et al., 1991) and may specifically interact with U11 snRNP via base pairing (Gontarek et al., 1993), raising the possibility that *nrs* function may involve a snRNP whose function had not been previously established.

Many of the sequences described in previous sections have been proposed to participate in the cell-type-specific regulation of alternative splicing. However, it is important to note that cell-type-specificity has not yet been established for these elements. It is equally plausible that these elements simply maintain a regulatable status and that specific regulation is mediated

via other elements. In order to establish that a given element is cell-type-specific, it may be necessary to demonstrate that an element displays greater activity in one cell type than in another.

Regulated splicing in *Drosophila melanogaster* and other genetic systems

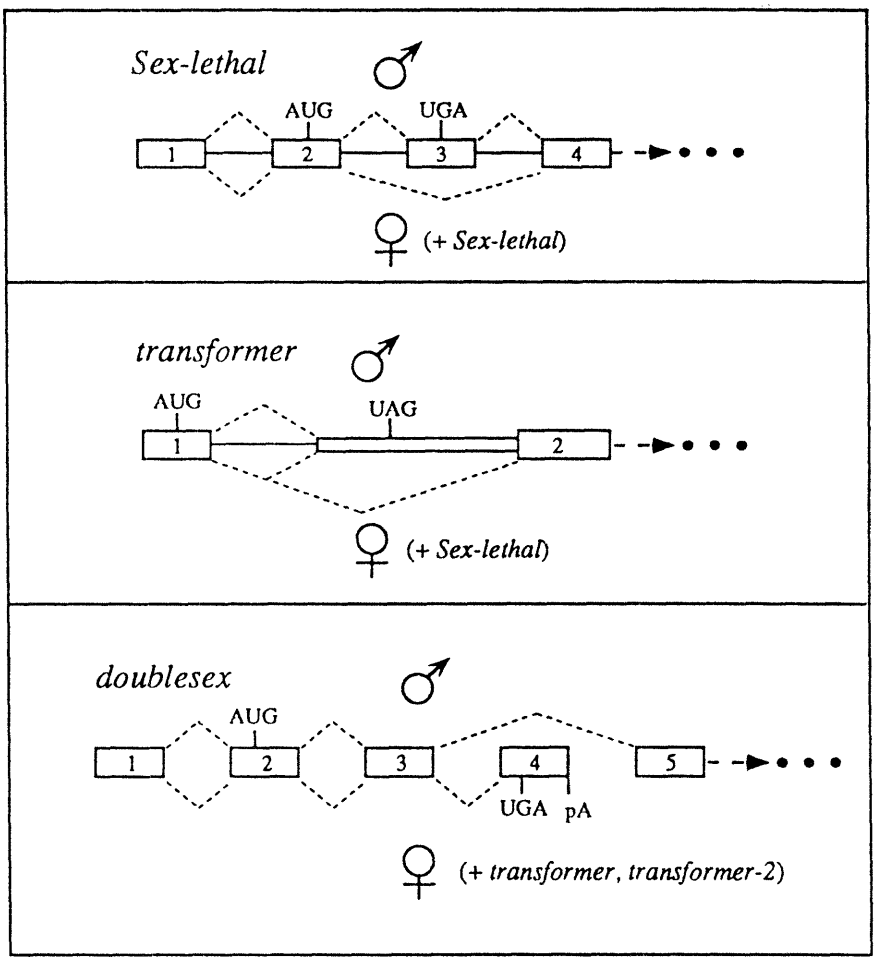
Detailed insights regarding the nature of *trans*-acting splicing regulators have been gained from studies in *Drosophila melanogaster*. An understanding of regulated splicing has been achieved due to the genetic identification of *cis* and *trans* lesions that specifically affect a particular splicing phenotype. The discovery that the regulation of a number of fly sex determination genes occurs via differential splicing has resulted in the identification of sex-specific splicing regulators and have provided information about the structure of a *bona fide* splicing regulator and the nature of the *cis*-elements through which these regulators function (for reviews, see Baker, 1989; Mattox et al., 1992; McKeown, 1992).

The first known alternatively spliced gene in the sex determination cascade is *Sex-lethal* (*Sxl*). This gene expresses distinct male- and female-specific mRNAs. Male-specific mRNAs contain an exon (exon 3) carrying a premature stop codon; female-specific exon 3 skipping results in full-length functional Sxl protein (Figure 1-2). The Sxl protein promotes female-specific splicing of the *transformer* (*tra*) gene as well as of its own pre-mRNA. The *tra* gene exhibits 3' splice site competition, a type of alternative splicing different from that found in *Sxl*. Usage of the proximal non-sex-specific 3' splice site results in premature frameshift termination; functional Tra protein is produced from female-specific distal 3' splice site usage (Figure 1-2). Tra acts together with the product of the *tra-2* gene to regulate splicing of *doublesex* (*dsx*) pre-mRNA. In the absence of Tra (males), *dsx* splicing produces an mRNA

Figure 1-2.

Examples of alternative splicing in *Drosophila melanogaster*

Partial exon/intron structures of genes involved in sex determination. Conventions are as in Figure 1-1; in addition, positions of initiator and termination codons are also indicated. The pattern above and below the exon-intron diagrams correspond to male- and female-specific splice patterns respectively. Female-specific patterns are observed in the presence of functional gene products indicated in parentheses.


Sex-lethal
AUG
UGA
1
2
3
4
(+ Sex-lethal)
transformer
AUG
UAG
1
2
(+ Sex-lethal)
doublesex
AUG
1
2
3
4
5
UGA
pA
(+ transformer, transformer-2)

containing exons 5 and 6 but not 4. In females, Tra and Tra-2 promote the splicing (and polyadenylation) of exon 4 (Figure 1-2). The male and female-specific forms of *dsx* perform distinct functions in sexual differentiation (reviewed by Baker, 1989; Mattox et al., 1992).

The regulation of *tra* splicing by the female-specific *Sxl* gene product appears to involve sequence-specific repression of the non-sex-specific (proximal) *tra* 3' splice site, as this site was used poorly in females even in the absence of the competing female-specific 3'SS. The site of Sxl action appears to be a sequence (U_8C) within the polypyrimidine tract of the non-sex-specific 3' splice site, since mutations at this sequence constitutively activated this site (Sosnowski et al., 1989; Inoue et al., 1990). Therefore female-specific Sxl-mediated blockage of the proximal *tra* acceptor results in distal acceptor usage. The U_8C sequence at the *tra* non-sex-specific 3' splice site is shared by the acceptor of the male-specific exon in *Sxl* , further suggesting that this sequence is the target of Sxl. The U_8C sequence can be specifically bound by recombinant female Sxl protein *in vitro* (Inoue et al., 1990; Valcárcel et al., 1993). *In vitro* splicing experiments have suggested that the affinity of Sxl for the pyrimidine tract of the non-sex-specific 3'SS is considerably higher than that of the general splicing factor U2AF, supporting a model wherein Sxl antagonizes U2AF action via occlusion at this site (Valcárcel et al., 1993).

Although *Sex-lethal* autoregulation appears in some respects to be similar to that of Sxl-mediated control of *tra*, regulation of the former is more complex, possibly because the pattern of *Sxl* differential splicing differs from that of *tra*. Sex-specific regulation was still observed if the 3' splice site of the *Sxl* male-specific exon was deleted (thereby creating a 5'SS competition situation; Horabin and Schedl, 1993). Additionally, a number of polyuridine sequences both upstream and downstream of this exon were found to

contribute to sex-specific regulation (Sakamoto et al., 1992; Horabin and Schedl, 1992, 1993). Therefore, in contrast to the single U_8C element that regulates *tra* splicing, *Sex-lethal* autoregulation may involve a number of functionally redundant *cis*-sequences and possibly additional regulatory mechanisms.

In contrast to negative regulation by *Sxl*, the products of the *transformer* (*tra*) and *transformer-2* (*tra*-2) genes act positively on female-specific splicing of the doublesex (*dsx*) gene. In males, usage of the female-specific *dsx* exon (exon 4) is prevented in part by the presence of a suboptimal polypyrimidine tract at its 3'SS; mutations increasing the pyrimidine content of this tract were found to activate exon 4 usage even in the absence of the Tra and Tra-2 gene products (Hoshijima et al., 1991). Conversely, expression of Tra and Tra-2 was found to activate exon 4 acceptor usage (Ryner and Baker, 1991) and may also stimulate polyadenylation (Hedley and Maniatis, 1991). A number of triskadecameric repeats within the female specific were required for female-specific exon 4 usage (Nagoshi and Baker, 1990; Hoshijima et al., 1991; Ryner and Baker, 1991; Inoue et al., 1992; Hedley and Maniatis, 1991). These 13-mer repeats have been found to specifically bind Tra-2 and possibly also Tra protein (Hedley and Maniatis, 1991; Inoue et al., 1992; Tian and Maniatis, 1992). Therefore the repeated 13-mer element represents a sex-specific regulatory element and the products of the tra and tra-2 genes represent sex-specific regulators that act via these 13-mer repeats.

In vitro, the activation of exon 4 splicing in HeLa nuclear extracts was dependent upon the presence of 13-mer repeats in *cis* and also required the addition of either recombinant Tra or Tra2 protein. Tra and Tra-2 were also capable of activating splicing of a mutant globin 3'SS in a repeat-depndent fashion, illustrating autonomous function of the 13-mer element (Tian and

Manatis, 1992). Tra and Tra-2 were found to form a stable complex on the repeat sequences in nuclear extract; this complex was found to contain SR proteins (Tian and Maniatis, 1993). Using a yeast two-hybrid interaction assay, Tra and Tra-2 proteins were also found to interact specifically with the 35 kDa subunit of U2AF as well as with SC35 and SF2/ASF (Wu and Maniatis, 1993). Taken together, these data indicated that the Tra and Tra-2 gene products act by recruiting general splicing factors (such as U2AF and SR proteins) to a suboptimal splice acceptor.

The third intron in the *Drosophila* P element transposase pre-mRNA (IVS3) is removed only in the cells of the germline (Laski et al., 1986). An *in vivo* analysis of germline-specific IVS3 splicing has suggested that 5' exon sequences contained a soma-specific negatively regulating element (Laski and Rubin, 1989; Chain et al., 1991). Splicing of IVS3 in Drosophila Kc cell nuclear extracts was activated by addition of competitor RNAs containing 5' exon sequences, suggesting that 5' exon-specific factors in Kc extract repressed IVS3 removal (Siebel and Rio, 1990). Binding of U1 snRNP to a "pseudo" 5'SS near the authentic site appears to be involved in the mechanism of somatic repression, since the removal of this "pseudo"-site activated IVS3 splicing *in vitro* (Siebel et al., 1992). UV-crosslinking and gel retardation assays have characterized a complex that assembles on this 5' exon sequence that contains several proteins (M_r 97K, 65K and 40K; Siebel and Rio, 1990; Siebel et al., 1992). The identification of these may provide further insights into the mechanism of this regulated splicing event. Yet another example of regulated splicing in Drosophila involves the regulation by the product of the *suppressor of white apricot* gene of its own pre-mRNA, apparently at the level of splicing (Zachar et al., 1987). This regulated event remains to be characterized in more detail.

Many of the genes that regulate the above splicing events have been cloned and characterized. The structures of these genes contain features shared by a number of RNA-binding proteins and known splicing factors. The products of the *tra*, *tra-2*, and *su(w^a)* genes contain domains that are arginine- and serine-rich, a feature shared with members of the SR protein family and U2AF. Tra-2 and Sxl also contain RRMs. Therefore these structural motifs are common among alternative splicing factors in *Drosophila* and constitutive splicing factors in mammalian systems.

The study of other genetic organisms have also identified regulators of splicing. In *S. cerevisiae*, splicing of the RPL32 pre-mRNA intron is negatively autoregulated by RPL32 ribosomal protein, presumably as part of a mechanism to prevent RPL32 overproduction. The RPL32 gene product represses the splicing of its own pre-mRNA via binding to a evolutionarily conserved base-paired structure in the 5' exon of the pre-mRNA (Eng and Warner, 1991). Rpl32 was found to inhibit splicing of this pre-mRNA by blocking splicing complex assembly at a step after U1 binding but before U2 association (Vilardell and Warner, 1994). Inhibition of splicing thus occurs at an intermediate step in splicing after the recognition of the 5'SS.

Only one case of developmentally regulated splicing in yeast has been discovered to date. The MER1 and MER2 genes are important for the meiotic process in yeast. Expression of the MER1 gene product is meiosis-specific; this gene product appears to positively act upon the splicing of the MER2 pre-mRNA (Engebrecht et al., 1991). The exact mechansm of MER1 function is unknown at present; it may act to facilitate recognition of the MER2 5'SS, the sequence of which differs from the yeast consensus (G/GUUCGU versus the consensus G/GUAUGU; Nandabalan et al., 1993).

Regulators of alternative RNA processing in mammalian genes

In contrast to systems in *Drosophila* , few if any authentic splicing regulators have been definitively identified as such in mammals. One potential regulator is the human immunodeficiency virus Rev gene product, which binds to a target (called the RRE) in its own pre-mRNA and thus promotes the cytoplasmic appearance of unspliced RNA. The mechanism of Rev action has been a matter of debate, since two different models have been proposed: (i) the active promotion of nucleocytoplasmic transport of unspliced RNA (Malim et al., 1989), and (ii) a Rev-mediated inhibition of splicing, thus allowing unspliced RNA to escape the nucleus by default (Chang and Sharp, 1989).

Since genetic tools that can identify upstream splicing regulators are not readily available for the analysis of mammalian splicing, biochemical complementation approaches have most often been adopted to identify potential regulators of splicing. Such an approach is relatively difficult, given that alternative splice site selection is sensitive to parameters such as extract dilution (Reed and Maniatis, 1986) and also variations in monovalent or divalent cation concentrations (Schmitt et al., 1987; Mayeda and Ohshima, 1988; Helfman et al., 1988). Therefore it is difficult to assess whether *in vitro* conditions could represent physiologically realistic situations. Nevertheless, such approaches have been attempted, yielding results which have differed in some respects from those obtained by studies in genetic systems.

The *c-src* gene contains a number of neural-specific exons, one of which has been studied as a model system for regulated alternative splicing. A number of sequence elements lie in the intron downstream of this exon that are required for neuron-specific N1 inclusion. At least two elements can be distinguished on the basis of these studies; these elements appear to exhibit some functional redundancy, since the deletion of either element alone reduced

N1 inclusion partially whereas deletion of both sequences abolished inclusion (Black, 1992). Neurally-derived splicing extracts were able to display N1 inclusion splicing in an element-dependent fashion, whereas HeLa extracts did not (Black, 1992). These experiments suggested that sequence-specific neural factors promoted the N1 inclusion splice *in vitro*. These experiments lay a promising foundation for biochemical complementation experiments that may identify neural-specific splicing regulators.

As mentioned previously, the splicing factor SF2/ASF was isolated from human 293 cell nuclear extracts by its ability to promote SV40 small t splicing (proximal 5'SS usage) over large T splicing in nuclear extracts (Ge et al., 1990). SF2/ASF also promoted proximal selection on a number of globin-derived pre-mRNA substrates containing duplicated splice sites (Krainer et al., 1990). In addition, it was found that hnRNP A1 could antagonize the effect of SF2 (Mayeda and Krainer, 1992; Fu et al., 1992). These findings suggested that differential ratios of ubiquitous RNA-binding proteins could modulate splice site preference *in vitro* (Maniatis, 1991).

As discussed previously, SF2/ASF is an SR protein (Zahler et al., 1992, 1993a, 1993b). Several of the SR proteins can promote proximal donor selection (Mayeda et al., 1992); in addition, SF2/ASF and SC35 can also promote proximal 3'SS selection (Fu et al., 1992). SF2/ASF and hnRNP A1 have also been shown to have effects *in vitro* upon alternative splicing of substrates derived from natural alternatively spliced pre-mRNAs (Mayeda et al., 1993). Taken together, these data suggest that SR proteins possess overlapping functions in splicing and may operate on selection of 5' splice sites via similar mechanisms.

To date, none of the mammalian SR proteins have been shown to directly regulate alternative splicing *in vivo*; such proof may have to await the

overexpression or knockout of these genes. It is interesting to note, however, that SR family members do exhibit some differential substrate preference, both with respect to committing different pre-mRNA substrates to splicing *in vitro* (Fu, 1993) and with respect to promoting proximal donor selection (Zahler et al., 1993a). Therefore each of the SR proteins may regulate distinct sets of alternatively spliced genes. In this respect, the SF2/ASF pre-mRNA is alternatively spliced to yield forms lacking the SR domain, which can promote proximal donor selection but do not complement an S100 extract for splicing (Caceres and Krainer, 1993; Zuo and Manley, 1993); therefore differential ASF/SF2 splicing could concievably have functional consequences *in vivo*.

The fact that specific splicing regulators have been identified in *Drosophila* and yeast makes it unlikely that all alternatively spliced genes are regulated by ratios of common splicing factors. However, these genetic screens are inherently biased in favor of specific regulator genes over essential factors. Since SR protein functions are apparently redundant, it is possible that a subset of these are actually true regulators for some genes. In this light, it will be interesting to learn the basis underlying the distinct substrate specificities that certain SR proteins exhibit. The recent description of SR homologues in *Drosophila* may shed some light on these questions (Kim et al., 1992; Mayeda et al, 1992), since it is possible to ask whether the mutation of SR protein-like genes in *Drosophila* will produce phenotypes related to differential splicing.

The rat fibronectin gene as a model for splicing regulation

It is therefore clearly evident that numerous questions must be examined when considering the regulation of an alternatively spliced gene *in vivo* or *in vitro*. It also seems likely that the study of new systems will reveal novel insights into the nature of regulation at the level of RNA splicing. In an

effort to gain further insights regarding how differential splicing is regulated in a cell-type-specific manner, experiments were carried out that characterized in detail the alternative splicing of one regulated exon in the rat fibronectin gene.

Fibronectin (FN) is a large multifunctional extracellular glycoprotein that is found in many basement membranes and is an important extracellular matrix component in vertebrates. This protein is a major substrate for the adhesion and migration of cells during development and also plays a role in maintaining cellular morphology, interacting also with other extracellular matrix molecules such as collagen and heparin. In addition to playing essential roles in development, FN appears also to participate in processes in the adult that involve cell migration and adhesion, such as the healing of wounds and blood clotting (for review, see Hynes, 1990).

The FN protein exhibits a repeated modular structure and contains a considerable degree of internal homology (Figure 1-3A). Three types of homology were detected based upon partial protein sequencing of bovine FN (Petersen et al., 1983) and also by sequencing of rat and human FN cDNA clones (Schwarzbauer, et al., 1983; Kornblihtt et al., 1984). Twelve type I repeats, two type II repeats and at least fifteen type III repeats are present in the protein. Various portions of FN function in cell adhesion and binding other extracellular matrix components (Figure 1-3A). In particular, the tenth type III repeat in FN (repeat III-10) contains a site particularly important for cell adhesion, the essential core being a tripeptide segment (RGD) within this repeat. This segment comprises the binding site for the $\alpha_5\beta_1$ integrin, a member of a family of cell surface receptors important for cell-matrix and cell-cell interactions (reviewed by Hynes, 1990, 1992).

The rat gene that encodes FN is approximately 70 kilobases in length and consists of at least 48 exons. The FN genomic structure was found to

consist of exons that corresponded to each of the three repeat types. Type I and II repeats are each encoded by single exons; for the most part, type III repeats are each encoded by a pair of exons. Expression of this pre-mRNA features a single transcriptional start and a single polyadenylation site (Patel et al., 1987). In addition, the rat FN pre-mRNA is alternatively spliced in at least three regions (Figure 1-3B). Two "cassette" exons (EIIIA and EIIIB) are optionally spliced into the mRNA; each of these exons encodes a single type III repeat. The third region, termed the V region, exhibits differential selection among three 3' splice sites: differential splicing here produces segments of 120, 95 or 0 amino acids in this region of the FN molecule.

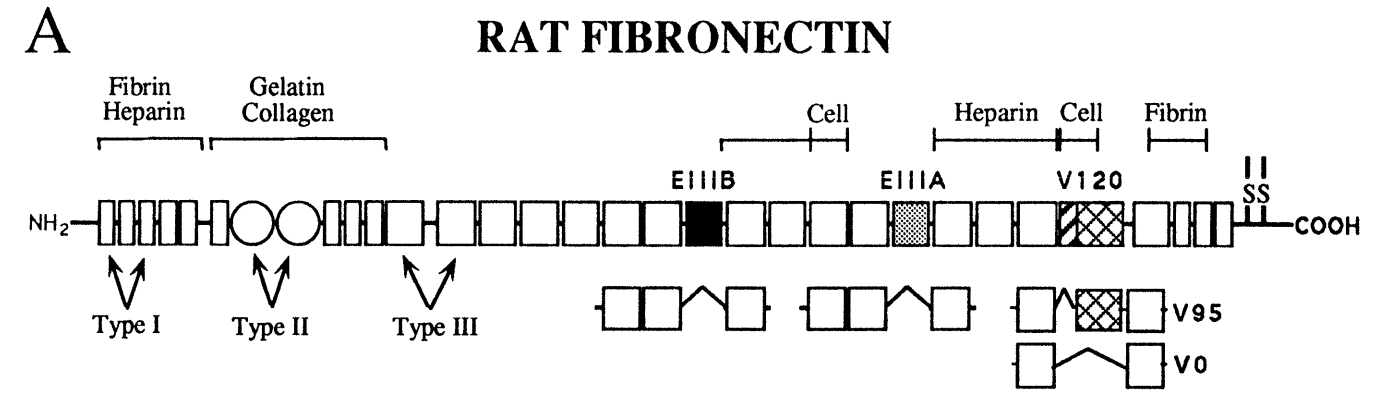
Alternative splicing of the fibronectin gene was first documented in the V region based upon sequences of FN cDNAs isolated from rat liver and S1 analyses of liver mRNA; the existence of at least three fibronectin mRNAs was inferred from these data. Isolation of rat genomic clones confirmed that these mRNAs were generated from a single gene (Schwarzbauer et al., 1983; Tamkun et al., 1984). This region has been termed the IIICS region in the human gene; interestingly, the splicing of the human IIICS region generates forms of FN distinct from those known to exist in rat; these novel forms arise from usage of an extra 5' splice site within the IIICS segment (Kornblihtt et al., 1985). The chicken and Xenopus FN genes also exhibit patterns of differential V splicing that vary slightly from those exhibited by the rat and human genes. Splicing of this region in the chicken gene results in partial or complete V segment inclusion; however, complete omission of V has not been observed (Norton and Hynes, 1987). Splicing in the *Xenopus* gene results in complete inclusion or omission of V, but not partial inclusion (DeSimone et al., 1992). Therefore, although the presence of V region differential splicing is conserved among

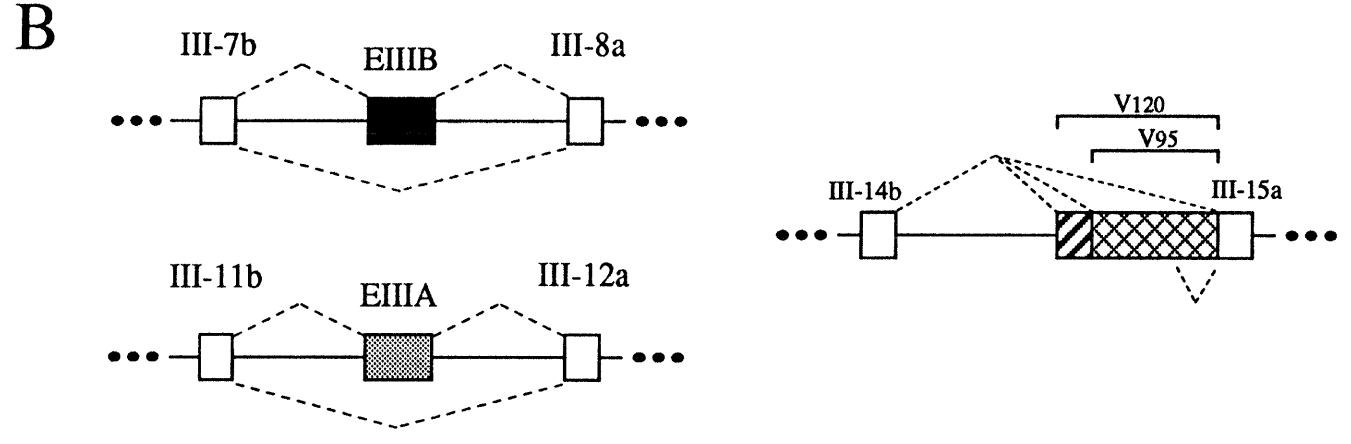
Figure 1-3.

Alternative splicing in the rat fibronectin gene

A. Schematic of the domain structure of full length rat fibronectin, showing the three types of homology and alternatively spliced segments. Regions of the protein that interact with components of the extracellular matrix or plasma are shown above the diagram; segments below the main diagram indicate the heterogeneity within that segment generated via alternative splicing (carats between domains indicate omitted segments; adapted from Hynes, 1990).

B. Alternative splicing patterns in the fibronectin gene. Exon-intron structure of the three alternatively spliced regions are schematically illustrated; drawings are not to scale. Upstream and downstream exons are labeled; other conventions are as in Figure 1-1. For EIIIB and EIIIA, possible patterns are shown both above and below each diagram. For the V region, possible patterns generated in the rat gene are indicated above the exon/intron structure, while an extra splice that occurs in the human gene is shown beneath.


A
RAT FIBRONECTIN
Fibrin
Heparin
Gelatin
Collagen
Cell
Heparin
Cell
Fibrin
EIIIB
EIIIA
V120
SS
NH2
COOH
Type I
Type II
Type III
V95
V0


B
III-7b
EIIIB
III-8a
III-11b
EIIIA
III-12a
V120
V95
III-14b
III-15a

species, the precise patterns exhibit a degree of variability among the organisms studied.

The V segment contains a segment that is important for adhesion to some cell types. Addition of peptides containing parts of the V segment have been shown to inhibit the adhesion of certain cells, presumably by competing for cell surface receptors that bind these segments. One such peptide occurs within the N-terminal portion of the V segment (Humphries et al., 1987). This segment (also called CS-1) has been shown to be the binding site for the $\alpha_4\beta_1$ integrin, which is found on cells of lymphoid origin (Guan and Hynes, 1990; Wayner et al., 1990). Another cell binding site, CS-5, appears to be present in the C-terminal portion of the V segment (Humphries et al., 1986, 1987). Therefore differential V splicing has a functional consequence with respect to FN-mediated cell adhesion. In rat liver, the FN mRNAs encoding the V95 and V120 forms comprise about 50% of the FN mRNA; the remainder encodes the V0 form (Schwarzbauer et al., 1983, 1985). In other tissues, V120 and V95 forms predominate (Tamkun and Hynes, 1983; Paul et al., 1986; Pagani et al., 1991). Thus V region alternative splicing is cell-type-specific. These findings suggest that the optional inclusion of the CS-1 and CS- 5 segments via cell-type-specific alternative splicing is important for the proper physiological function of FN.

The EIIIA and EIIIB segments each encode a single structural type III structural repeat unit that is either included into or excluded from FN mRNAs. With the exception of the III-9 repeat, all the other type III homologies in the rat FN gene are encoded by two exons each. The differential inclusion of these exons was inferred from the sequence comparison of different cDNA isolates (Kornblihtt et al., 1984; Schwarzbauer et al., 1987; Zardi et al., 1987).

Alternative splicing of EIIIA and EIIIB is also cell-type-specific. Neither of these exons are present in liver FN mRNAs from rat, chicken or *Xenopus*, whereas these segments are found in a high percentage of FN mRNAs from early embryos (Schwarzbauer et al., 1987; Norton and Hynes, 1987; DeSimone et al., 1992). In fact, differences in biochemical behaviour between soluble plasma fibronectin (which is synthesized by hepatocytes) and cellular fibronectin (an insoluble fibrillar form found in the extracellular matrix and on cells in tissue culture) can in part be accounted for by the differential inclusion of these exons as well as of the V segment (Paul et al., 1986). In many instances, inclusion of EIIIA and EIIIB have been found to correlate, with EIIIA often being present at higher levels than EIIIB (ffrench-Constant et al., 1989; Magnuson et al., 1991; Pagani et al., 1991), suggesting that these exons are co-regulated. However, certain exceptions suggest that EIIIA and EIIIB regulation can occur independently, as cartilage contains EIIIA$^-$/EIIIB$^+$ FN whereas some blood vessel walls contain FN that is EIIIA$^+$/EIIIB$^-$ (Bennett et al., 1991; ffrench-Constant et al., 1989).

In contrast to the V segment, potential functions for the EIIIA and EIIIB exons have not yet been clearly identified. The EIIIB segment has been thought to be associated with cellular transformation (Carnemolla et al., 1989) but other studies have failed to corroborate this finding (Schwarzbauer et al., 1987; Norton et al., 1987; Magnuson et al., 1991). As mentioned above, EIIIA and EIIIB are abundant in early embryos; in addition, in situ hybridization studies using segment-specific probes have indicated that EIIIB expression is elevated at locations of active cell migration and proliferation in the chicken embryo (ffrench-Constant and Hynes, 1989) and that both EIIIA and EIIIB levels rise at sites of wound-healing in rats (ffrench-Constant et al., 1989). In addition, EIIIA and EIIIB are responsive to growth factors applied to cells in

tissue culture (Magnuson et al., 1991; Borsi et al., 1990). Therefore EIIIA and EIIIB may play a role in modulating FN-mediated cell migration and/or proliferation.

However, it is apparent from studies of recombinantly expressed FN isoforms that all forms of FN are capable of promoting cell adhesion, spreading and migration (Guan et al., 1990). EIIIA and EIIIB have been thought to play a role in fibrillogenesis; however, only minor differences have been detected among fibronectin isoforms in their ability to be retained by cells via incorporation into matrix (Guan et al., 1990; Mardon et al., 1992). It should be noted that differences among FN isoforms may have been masked by (i) the presence of the cell-binding site in the III-10 segment, which is a major determinant for adhesion for many tissue culture cell lines and (ii) the other domains in FN that interact with extracellular matrix components such as collagen or heparin. In addition, differences among recombinant FN isoforms might have gone unnoticed if EIIIA- or EIIIB- specific posttranslational modifications were not present in the recombinant FNs. The amino acid sequence of the EIIIB exon is very highly conserved among the warmblooded animals (96% versus 81% overall between rat and chicken; Norton and Hynes, 1987), suggesting some importance of the EIIIB sequence in function and/or regulation. To date, however, EIIIA and EIIIB functions remain unknown.

The control of alternative splicing in the fibronectin pre-mRNA

The regulation of fibronectin alternative splicing has been investigated to some extent by a number of groups. A minigene system has been established for studying the splicing of the human IIICS region; in this systems, a minigene containing the IIICS region and flanking regions was found to exhibit all the expected spliced patterns when transfected into HeLa cells

(Mardon and Sebastio, 1992). Th *cis*-acting sequences that determine splice site selection have not yet been determined in this system.

Regulation of the EIIIA exon (also known as ED-A, ED1 and EDIIIA) has proceeded via *in vivo* and *in vitro* studies. A minigene containing the human EIIIA region was found exhibit both splice choices (inclusion and skipping) when transfected into HeLa cells (Vibe-Pederson et al., 1984). One particular requirement for the recognition of the EIIIA exon in this system appears to reside within the exon itself, as the removal or inversion of a sequence within EIIIA resulted in the failure to splice this exon into mRNA (exon-skipping). Therefore exon sequences contributed in a positive manner for EIIIA inclusion splicing (Mardon et al., 1987). *In vitro* splicing studies that used EIIIA-containing substrates have identified a purine-rich sequence within this exon that was important for utilization of the human EIIIA 3'SS in HeLa cell nuclear extracts (Lavigueur et al., 1993). The elements that regulate the cell-specific expression of EIIIA have not yet been characterized; although it is possible that the purine-rich element within EIIIA acts in a cell-specific manner, it is not known whether this element possesses differential activity among cell types.

Studies have also been carried out using minigenes carrying the human EIIIB exon (also known as ED-B, ED2 and EDIIIB). These studies have indicated that a genomic segment carrying only EIIIB and its immediately flanking exons was capable of being alternatively spliced when transfected into HeLa cells (Paolella et al., 1988). The regulated inclusion of EIIIB in this genomic segment appeared to correlate with that of endogenously expressed FN in various human cell strains and lines. These studies also suggested that low EIIIB levels was not determined by preferential cytoplasmic degradation of EIIIB-containing mRNAs, as the relative proportions of EIIIB$^+$ and EIIIB$^-$

mRNAs did not change following actinomycin D treatment (Barone et al., 1989).

The EIIIB exon 3'SS exhibits a number of interesting features. Using rat derived substrates, the splicing of the intron upstream of EIIIB in HeLa nuclear extracts was found to be inefficient under standard (<3 mM $MgCl_2$) conditions but was detectable at high divalent cation concentrations (>5 mM $MgCl_2$). Splicing was found to utilize three branchpoint residues located 62, 70 and 76 bases upstream of the 3'AG. In addition, the pyrimidine tract separating the 3'AG and these branchpoints is also unusually long (55-60 bases). Therefore this 3'SS shares a number of features with a number of other alternatively spliced exons (Norton and Hynes, 1990). UV-crosslinking assays (using a shortened III-7b-EIIIB intron as a substrate) revealed that different nuclear extract proteins bound to EIIIB 3'SS RNA in a Mg^{2+}-dependent fashion. At low $MgCl_2$ concentrations (conditions restrictive for splicing), a 56/58K doublet appeared to bind this intron, whereas a 62K protein bound at higher (permissive) $MgCl_2$ concentrations. Biochemical fractionation and immunoprecipitation analyses suggested that the 56/58K doublet corresponded to hnRNP I (PTB) and that the 62K protein was the splicing factor U2AF. Therefore the splicing of the intron preceding EIIIB may depend upon a competition between U2AF and PTB for sites in the extended pyrimidine tract of the EIIIB 3'SS, a model analogous to that proposed for the regulation of *tra* pre-mRNA splicing by *Sex-lethal* (Norton and Hynes, 1993; Valcárcel et al., 1993). This model remains to be supported *in vivo*.

It would appear that a large number of parameters may govern alternative splicing in a particular gene. It is also apparent that some of the current issues regarding the nature of vertebrate alternative splicing regulation may require extensive *in vivo* studies, since it is not clear whether any *in vitro*

effects can be extended to *in vivo* situations. The EIIIB exon provides a number of attractive features for study as a model system for the study of regulated splicing. The EIIIB exon is tightly regulated *in vivo*; therefore an analysis of EIIIB splicing may reveal helpful clues regarding the exact nature of cell-type-specific regulation in vertebrates. In addition, given that this exon has some features similar to those exhibited by other alternative exons, an examination of these features and how they participate in the regulation of EIIIB splicing may prove informative. Furthermore, this exon is very highly conserved among homeotherms, suggesting an important function for this segment. Given that EIIIB may participate in cell migration or proliferation, the characterization of EIIIB regulation may contribute some valuable insights regarding how cell migration and proliferation might be controlled at the posttranscriptional level.

This thesis describes an extensive analysis of the alternative splicing of the EIIIB exon in the rat fibronectin gene. It is evident from many studies that a number of sequence parameters must collaborate to create a situation in which differential splicing is possible. The experiments in Chapter Two address this issue with regard to EIIIB specifically. Another question concerns how the cell-type-specific regulation of EIIIB splicing is mediated. The regulation of EIIIB presumably occurs through the action of *trans*-acting factors that act in a cell-type-specific fashion, probably by interacting with specific *cis*-elements that are present in the EIIIB region. The experiments described in Chapter Three identify a repeated hexanucleotide element that exhibits cell-type-specific characteristics. This element has features that distinguish it from other vertebrate elements identified to date and furthermore may regulate splicing in genes other than that encoding FN, an issue that is addressed in part in Chapter Four. This element may therefore be of significance with respect to

the regulation of alternative splicing in general. Finally, experiments described in Chapter Five comprise a biochemical analysis that identifies factors that may be involved in the regulation of alternative splicing. These factors are sequence-specific RNA-binding proteins that will hopefully provide mechanistic insights into the regulation of alternative splicing in the FN gene and possibly in other mammalian genes.

Chapter Two:

Elements Regulating An Alternatively Spliced Exon In The Rat Fibronectin Gene

ABSTRACT

We have investigated the regulation of splicing of one of the alternatively spliced exons in the rat fibronectin gene, the EIIIB exon. This 273 nucleotide exon is excluded by some cells and included to varying degrees by others. We find that EIIIB is intrinsically poorly spliced and that both its exon sequences and splice sites contribute to its poor recognition. Therefore, cells which recognize the EIIIB exon must have mechanisms for improving its splicing. Furthermore, in order for EIIIB to be regulated, a balance must exist between the EIIIB 5' splice site and that of the exon upstream. Although the intron upstream of EIIIB does not appear to play an essential role in the recognition of EIIIB for splicing, the intron downstream contains sequence elements which can promote EIIIB recognition in a cell-type-specific fashion. These elements are located an unusually long distance from the exon they regulate, more than 518 nucleotides downstream from EIIIB, and may represent a novel mode of exon regulation.

INTRODUCTION

It is evident, from many studies of alternative splice site selection, that a wide variety of potential mechanisms must be considered in analyzing alternative splicing (for reviews, see Smith et al., 1989a; McKeown, 1992, and references therein). In addition to the intrinsic quality of the splice sites themselves (for reviews, see Moore et al., 1993; Green, 1991), splice site selection may also depend upon factors such as exon sequences, pre-mRNA secondary structure, relative proximity of alternative splice sites and possibly also exon size. Therefore a multiplicity of determinants can influence splice site selection in profound and often unpredictable ways.

Considerable insight into the regulation of splicing has been gained from studies of genes identified in *Drosophila melanogaster*, where examples of both positive and negative regulation can be found (for reviews, see Mattox et al., 1992; Maniatis, 1992; McKeown, 1992). In contrast, the characterization of mammalian regulated alternative splicing has been more elusive. A directed mutational analysis of regulation can be a complex task, particularly since splice site choice can be affected by numerous contextual parameters; mutations which affect any of these can abolish regulated splicing without necessarily affecting specific regulatory elements. Nevertheless, candidate cell-type-specific elements have been identified in genes such as those encoding calcitonin/CGRP (Emeson et al., 1989), *c-src* (Black, 1992) and N-CAM (Tacke and Goridis, 1991). The trans-acting factors regulating these alternative splice choices have not yet been identified. Interestingly, recent biochemical studies have implicated general splicing factors in the regulation of alternative splicing (Ge and Manley, 1990; Mayeda and Krainer, 1992; Zahler et al., 1992); however,

it is not known yet whether these *in vitro* effects can be extended to *in vivo* situations.

In the FN gene, the two alternatively spliced exons EIIIA and EIIIB each encode a type III structural repeat in FN and exhibit exon-skipping; a third region in this gene displays species-specific patterns of differential splice site selection. EIIIA and EIIIB are regulated in a cell type-specific and developmental fashion; for example, adult liver FN mRNAs lack both EIIIA and EIIIB, whereas all early embryonic FN mRNAs contain both exons. The appearance of EIIIB often correlates with that of EIIIA, with some exceptions (for review, See Hynes, 1990). Of the FN type III repeats, EIIIB (also called EDB) is one of the most highly conserved among species and appears in fewer tissues than EIIIA, suggesting that it is more strictly regulated. Previous studies (Norton and Hynes, 1990) have shown that: (i) EIIIB is mostly skipped in vitro, (ii) in vitro splicing of the intron preceding EIIIB requires unusually high Mg^{2+} concentrations, (iii) this intron utilizes three lariat branchpoints 62, 70 and 76 nucleotides (nt) upstream from the 3' splice site, and (iv) the pyrimidine-rich stretch preceding this 3' splice site is relatively long (55-60 nt). Either of these latter two features are shared by other alternative 3' splice sites (Helfman and Ricci, 1989; Gattoni et al., 1988; Smith and Nadal-Ginard, 1989; Goux-Pelletan et al., 1990). A human minigene containing EIIIB and its flanking exons is alternatively spliced in a cell-type-specific fashion (Barone et al., 1989), indicating that regulatory elements lie in this three-exon region.

In order to study EIIIB regulation, we have transfected rat EIIIB-containing minigenes into cell lines which recognize EIIIB to varying extents. Our results indicate that EIIIB is a poor exon as a consequence of suboptimal splice sites and exon sequences. Proper regulation requires a balance in 5' splice site strength between EIIIB and the upstream exon. The unusual EIIIB 3'

splice site was important for exon repression but was not essential for cell-type-specific EIIIB regulation. In contrast, a region in the intron downstream of EIIIB is required for EIIIB recognition; a subsection of this region is particularly important for cell-type-specific activation of EIIIB. This latter element, which we term the ICR (intronic control region), lies at least 519 nt from EIIIB. The distant location of the ICR relative to EIIIB raises interesting questions regarding the possible function of this novel cis-element.

RESULTS

A minigene containing EIIIB exhibits regulated alternative splicing

We used three-exon minigenes to study regulated EIIIB alternative splicing. The primary fibronectin minigene in this study is called 7iBi89 (Figure 2-1A). This 3.1 kb minigene is derived from genomic and cDNA clones of the rat FN gene. The first and second exons of 7iBi89 correspond to exons III-7b and EIIIB, respectively; the third exon is a cDNA composite of exons III-8a, III-8b and part of III-9. The introns upstream (IVS1) and downstream (IVS2) of EIIIB are 1296 and 1071 nucleotides (nt) long respectively. Transcription was driven by a modified human β-actin promoter and was terminated by the human growth hormone polyA signal. Two vectors with these expression signals were used (Figure 2-1B). The vector for transient transfections (pBAGH.Sv) contains an SV40 origin of replication which facilitated expression in COS cells. The vector for generating stably expressing cell populations (pBAGH) did not contain an origin, since the SV40 origin/enhancer downregulated expression in F9 teratocarcinoma cells (Gorman et al., 1985).

We carried out RNase protection analyses to demonstrate correct splicing of the transfected 7iBi89 minigene. A uniformly labeled antisense RNA probe was transcribed from a form of 7iBi89 with intronic deletions (Figure 2-1A). Protection of this probe by correctly spliced RNA results in bands corresponding to spliced exons. Constitutive exons are represented at stoichiometric intensities (after normalizing for radiolabel content); in contrast, alternative exons are represented in substoichiometric proportions. Since this probe also detects endogenous FN mRNA, this RNase assay was used only on RNA from transfections of COS and 293 cells, which did not express significant

amounts of endogenous FN message. As Figure 2-2 illustrates, spliced EIIIB$^-$ and EIIIB$^+$ minigene RNAs represent the major bands in the RNase protection analysis. 293 cells exhibit a low but significant amount of EIIIB inclusion; COS cells skip EIIIB completely. Therefore the 7iBi89 transcript is spliced correctly and efficiently in cell types which exhibit one or both EIIIB splice choices.

In order to establish that 7iBi89 exhibited regulated alternative splicing, transfections of a variety of mammalian cell lines were analyzed by S1 analysis, using 5'-end-labelled cDNA probes which overlapped EIIIB. In order to compare endogenous and minigene EIIIB splicing, we used two probes: (i) E, for analysis of endogenous rat (or mouse) FN mRNA from untransfected rodent cell lines, and (ii) B-89, an S1 probe which specifically detects minigene RNA. Analyses of total RNA from untransfected, transiently- or stably-transfected cells are compared in Figures 2-3 and 2-4 and summarized in Table 2-1.

Using probe E (Figure 2-3), rodent lines were found to exhibit varying degrees of endogenous EIIIB inclusion; of these, Rat-1 cells exhibited very little inclusion (lane 1, EIIIB$^+$/EIIIB$^-$ ratio of ca. 0.1), whereas undifferentiated F9 teratocarcinoma cells exhibited the most (lane 3, ratio of ca. 8). NIH3T3 and differentiated F9 cells (lanes 2 and 4) showed intermediate EIIIB inclusion. Analyses of 7iBi89 expression (using probe B-89; Figure 2-4) showed that minigene EIIIB alternative splicing paralleled endogenous FN splicing. Overall, minigene EIIIB inclusion was less compared to endogenous FN, particularly when transient transfections were carried out; however, the overall correlation with endogenous EIIIB splicing was still evident (Table 2-1).

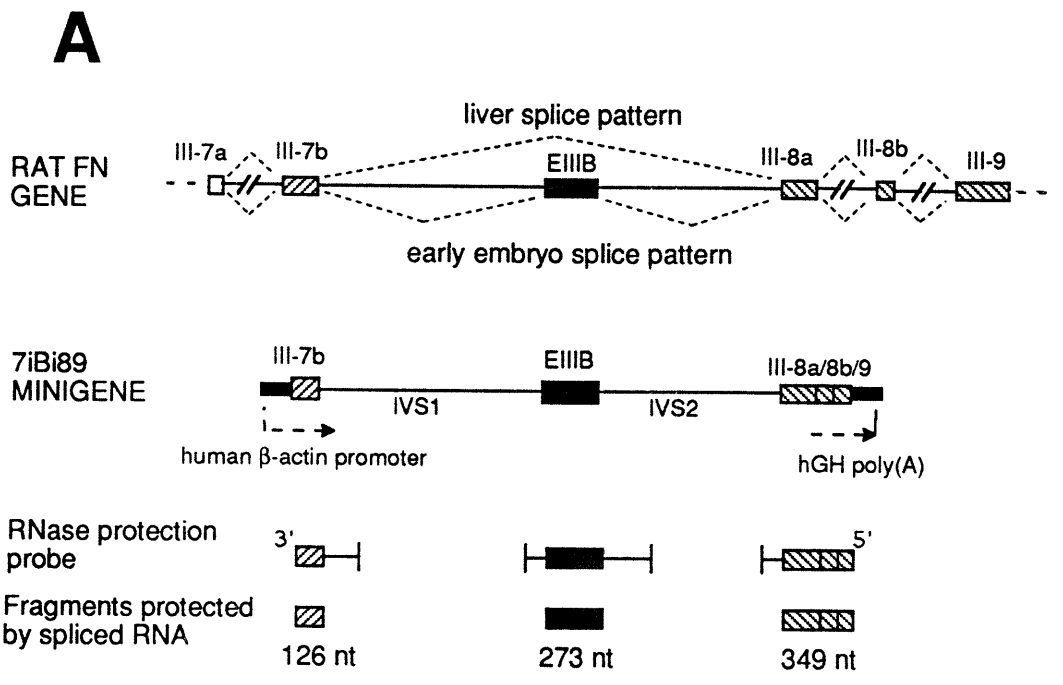
Of the non-rodent lines, COS cells expressed the lowest proportion of minigene EIIIB inclusion (Figure 2-4; lane 8), whereas 293 cells spliced EIIIB into mRNA at the highest relative level (lanes 11 and 12; Table 2-1). Other

Figure 2-1.

Construction of a three-exon rat fibronectin EIIIB minigene.

A. The EIIIB region of the rat FN gene and 7iBi89, the 3.1 kb minigene derived from this region. Exons and introns are denoted by labelled rectangles and lines respectively. Alternative splice patterns are denoted by dashed lines. Promoter and 3' polyA signals of 7iBi89 are shown. Also shown is the RNase protection probe used in Figure 2-2, as well as the sizes of anticipated products arising from protection by spliced exons.

B. Expression vectors pBAGH and pBAGH.Sv used in this study. Minigene inserts were placed at the BamHI cloning site in these plasmids (Arrow, start of transcription; "pA", poly(A) site).


A
liver splice pattern
RAT FN GENE
III-7a
III-7b
EIIIB
III-8a
III-8b
III-9
early embryo splice pattern
7iBi89 MINIGENE
III-7b
EIIIB
III-8a/8b/9
IVS1
IVS2
human β-actin promoter
hGH poly(A)
RNase protection probe
3'
5'
Fragments protected by spliced RNA
126 nt
273 nt
349 nt

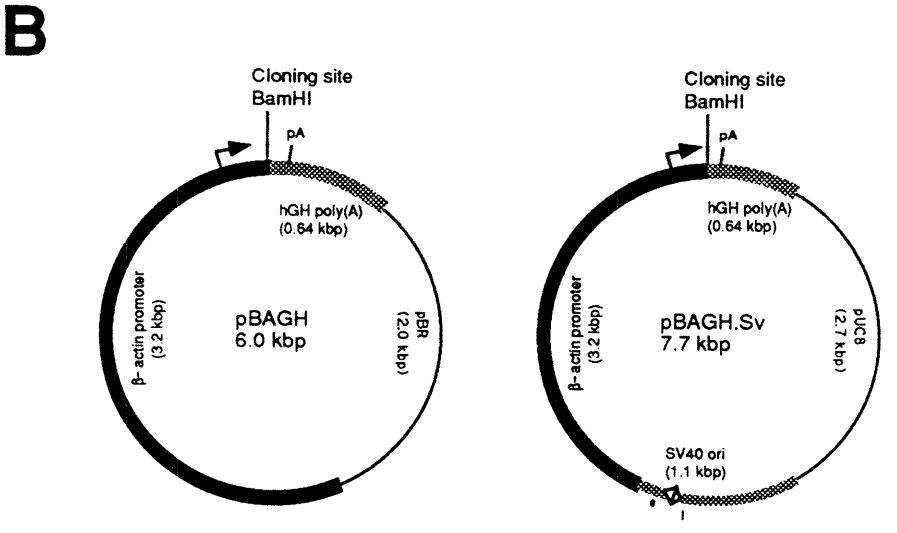

B
Cloning site
BamHI
pA
hGH poly(A)
(0.64 kbp)
pBAGH
6.0 kbp
β- actin promoter
(3.2 kbp)
pBR
(2.0 kbp)
Cloning site
BamHI
pA
hGH poly(A)
(0.64 kbp)
pBAGH.Sv
7.7 kbp
β- actin promoter
(3.2 kbp)
pUC8
(2.7 kbp)
SV40 ori
(1.1 kbp)

Figure 2-2.

A minigene containing EIIIB exhibits alternative splicing.

RNase protection analysis of total RNA from COS and 293 cells transiently transfected with 7iBi89 in the expression vector pBAGH.Sv. RNAs from untransfected (-) or transfected (+) cells were analyzed using a uniformly labeled, antisense RNase protection probe (shown in Figure 2-1). An aliquot of probe was run alongside this analysis (lane "Probe"); the probe band is only faintly visible at this exposure. Molecular weight markers (lane M) were end-labeled pBR322/MspI. Labeled arrowheads show positions of undigested probe ("probe") and products from protection by correctly spliced 7iBi89 exons. Asterisks show products protected by unspliced RNAs. One of these latter products (triangle) is overrepresented; this probably reflects the infrequent use (<3%) of a cryptic 5' splice site within IVS1.

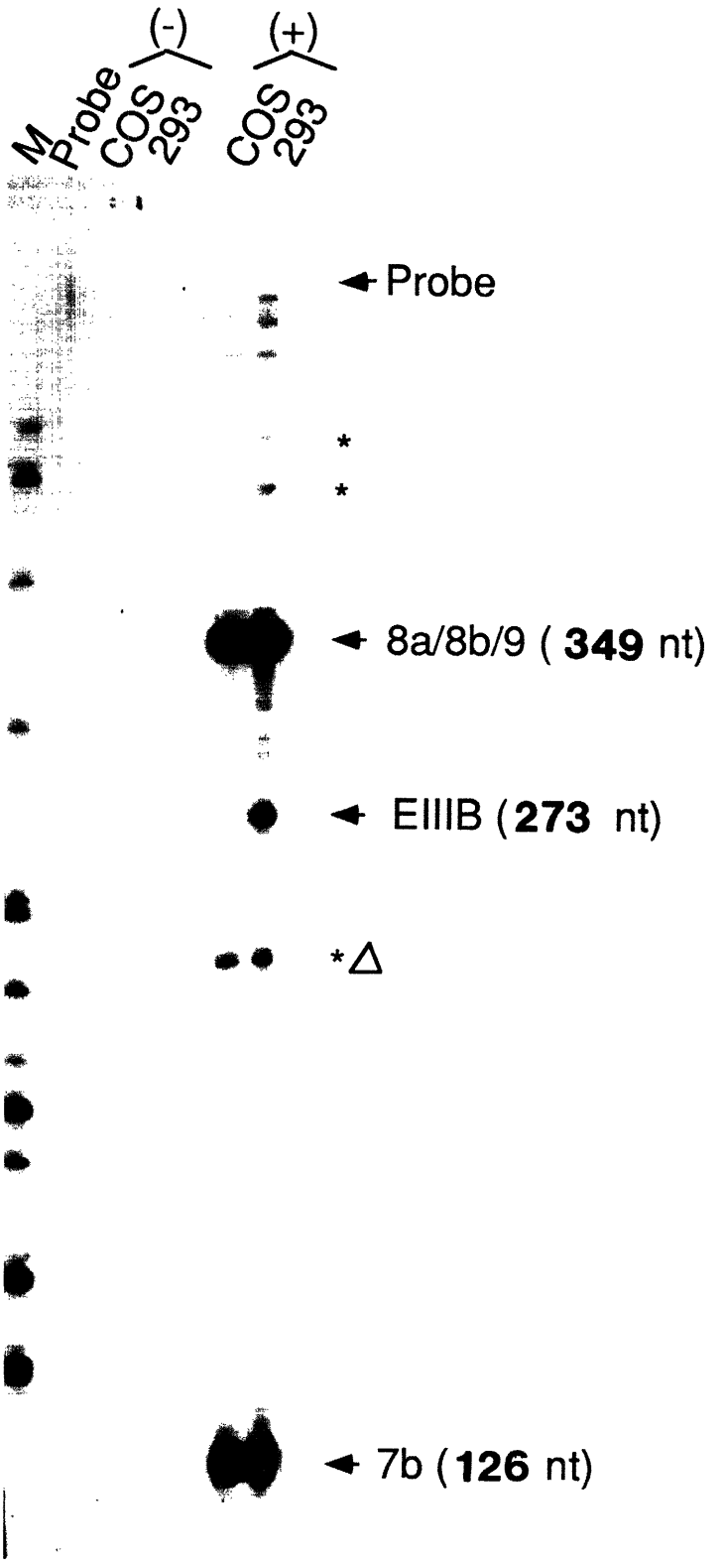

M
Probe
COS
293
(-)
COS
293
(+)
Probe
*
*
8a/8b/9 (349 nt)
EIIIB (273 nt)
*Δ
7b (126 nt)

Figure 2-3.

Endogenous alternative splicing is cell-type-specific.

S1 analysis of EIIIB alternative splicing of endogenous FN mRNA in rodent cell lines. The structure and size of 5' end-labeled probe E (which detects endogenous rodent FN mRNA) is shown below the gel (wavy line, vector sequence; black and stippled rectangles, exons EIIIB and III-8a/8b/9 respectively). S1 products are labeled by arrowheads alongside the gel (undigested probe, $EIIIB^+$, $EIIIB^-$). Up to 35 micrograms of total RNA (from cell lines indicated above each lane) were analyzed. yRNA, control yeast RNA; Rat-1, a rat fibroblast line; 3T3, murine NIH3T3 cells; F9, undifferentiated murine F9 cells; F9diff, F9 cells treated with retinoic acid and dibutyryl cyclic AMP for 96 hr; M, pBR322/MspI markers).

S1 Probe E (endogenous FN mRNA)

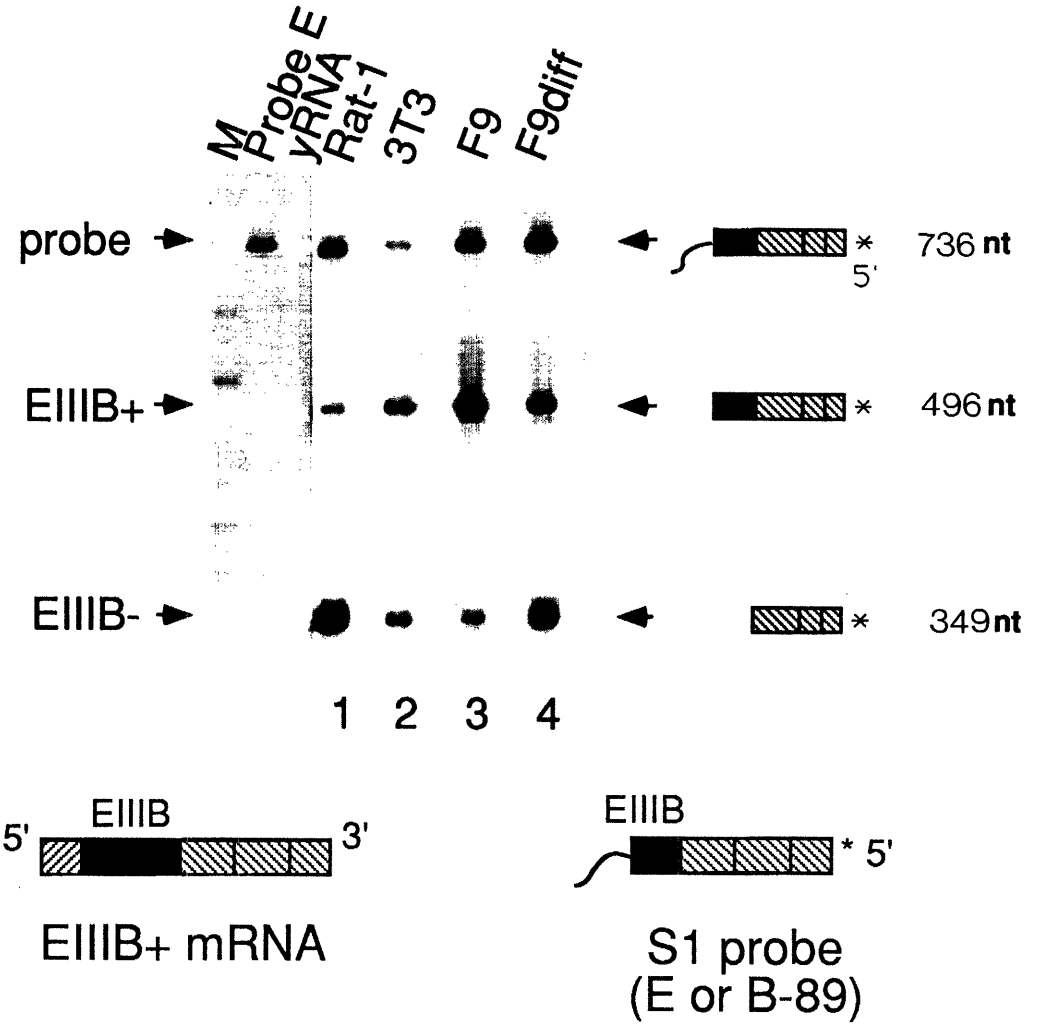


Figure 2-4.

Minigene alternative splicing is cell-type-specific.

S1 nuclease protection analyses of EIIIB splicing of the 7iBi89 minigene. Probe B89 contains the same cDNA sequences as does probe E (see Figure 2-3); its labeled 5' end lies in expression vector sequences and therefore does not detect endogenous FN mRNA. Results from transfections of rodent and non-rodent cell lines are shown. Cell lines and transfection methods are indicated above each lane (Ctrl, control untransfected cells; Trans, transiently transfected cells using expression vector pBAGH.Sv; Stab, stably transfected cell populations using vector pBAGH; other conventions are as in Figure 2-3).

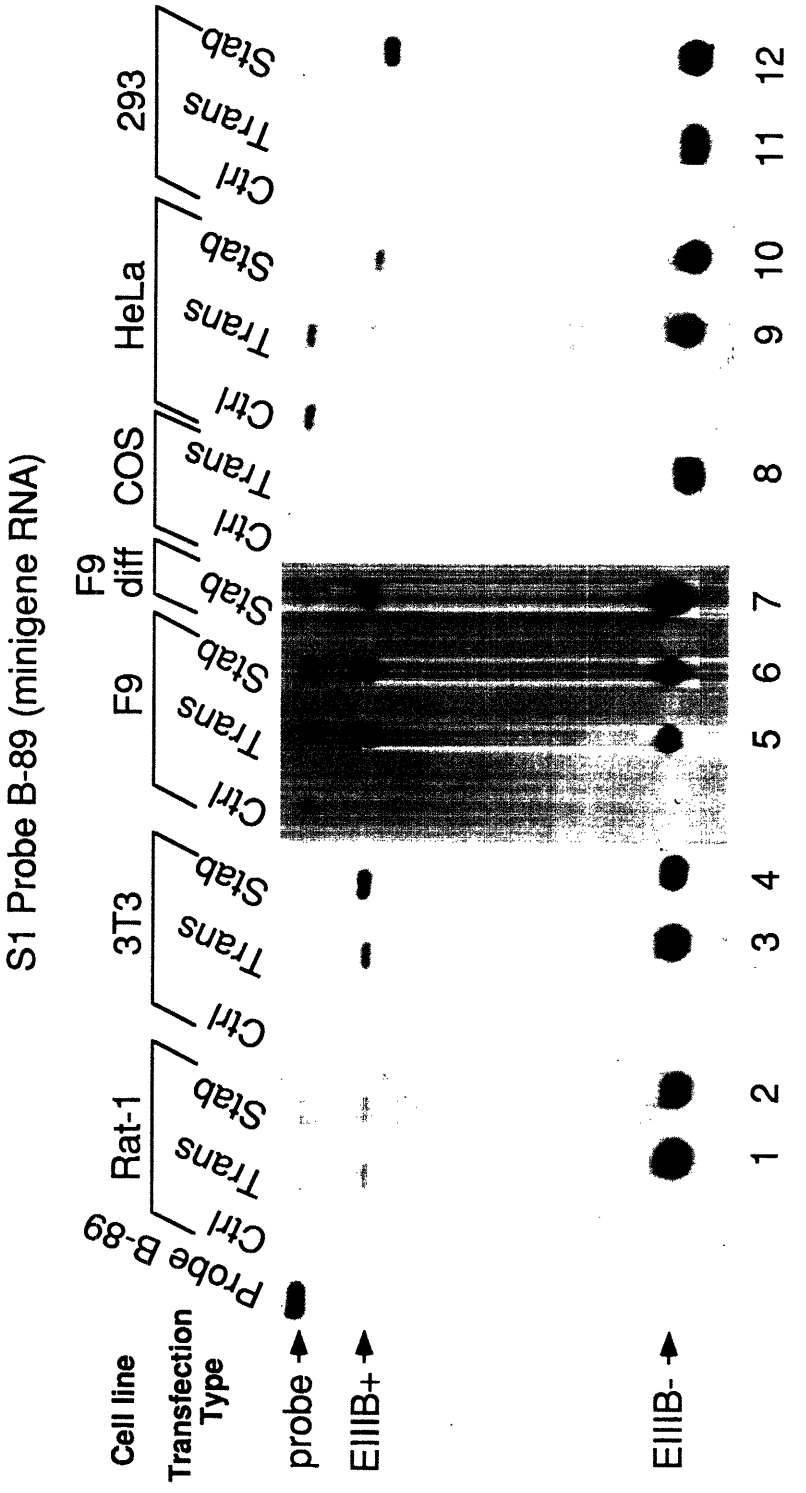

S1 Probe B-89 (minigene RNA)
Cell line
Transfection Type
Probe B-89
Ctrl
Trans
Stab
Rat-1
3T3
F9
F9 diff
COS
HeLa
293
probe
EIIIB+
EIIIB-
1 2 3 4 5 6 7 8 9 10 11 12

Table 2-1.
Endogenous and minigene alternative splicing of EIIIB: a comparison

Values represent $EIIIB^+/EIIIB^-$ ratios and varied by 10% between experiments.

(n.t. = not tested)

Cell line	Endogenous FN mRNA	7iBi89 (Transient)	7iBi89 (Stable)
Rat-1	0.1	0.03	0.03
NIH3T3	1	0.05	0.2
F9 (diff)	0.9	0.1-0.2	0.1
F9	8	0.4	0.4
COS	n.t.	0.008	n.t.
HeLa	n.t.	0.03	0.03
293	n.t.	0.08	0.2

studies have shown that human EIIIB (EDB) is found in 6-10% of total FN message in HeLa cells (Barone et al., 1989); this value compares well with 7iBi89 EIIIB inclusion in HeLa cells (about 3%), given that minigene values were lower than endogenous. We conclude that the 2.8 kb which surrounds EIIIB in 7iBi89 is sufficient to direct cell-type-specific EIIIB inclusion in a relatively faithful manner.

EIIIB is skipped in a heterologous context

We initially looked for elements which prevented constitutive EIIIB inclusion. Segments of 7iBi89 were replaced with analogous regions from a separate FN minigene with a constitutive internal exon, 8i9i10 (Figure 2-5), derived from a region of the rat FN gene containing the exon III-9. As the only constitutive FN type III repeat encoded by a single 270 nt exon (Schwarzbauer et al., 1987), III-9 seemed an appropriate non-alternatively spliced counterpart to EIIIB. When 8i9i10 was expressed in these cell lines, III-9 was spliced into more than 98% of the mRNA (Figure 2-7, lanes 19-21).

Three-exon chimeric minigenes were constructed as shown in Figure 2-5. With the exceptions of SW1 and SW2, these were named by chimera junction position (L, M,or R for left, middle or right of the central exon as pictured) and were numbered 1 or 2 depending upon whether the 5' segment of the chimera was derived from 7iBi89 (e.g., L1) or 8i9i10 (e.g., L2). Chimeras were made as perfect reciprocal pairs except for L1 and L2: a 123 bp BstEII-BstEII segment of IVS1 in 7iBi89 (between the two BstEII sites in Figure 3A) was not present in either L1 or L2; however, deletion of this fragment (minigene LC1) had no effect on EIIIB splicing (see below). Chimeras were transiently transfected into COS, 293 and F9 cells; these lines exhibited zero, moderate and high EIIIB inclusion respectively. To analyze transfectant RNAs,

cDNAs containing the second and third exons of each chimera were used as 5' end-labelled S1 probes similar to those shown in Figure 2-4. Figures 2-6 and 2-7 shows the S1 results; Figure 2-5 provides a summary.

Since 7iBi89 exhibited complete EIIIB-skipping in COS cells, we used data from these cells to identify regions of 7iBi89 sufficient to direct exon-skipping. All minigenes which exhibited complete exon-skipping contained a segment of 7iBi89 encompassing the EIIIB exon; in contrast, minigenes containing exon III-9 exhibited essentially complete exon inclusion. Chimera M2 (where the 5' half of 7iBi89 was replaced by 8i9i10 sequences) exhibited intermediate inclusion of its 259 nt hybrid exon. The reciprocal chimera (M1) exhibited complete inclusion of its composite 285 nt exon. Therefore, the sequences which are responsible for making EIIIB a non-constitutive exon are within and/or immediately local to EIIIB in a 590 nt segment; the 3' half of this segment contributes more than the 5' half in this regard.

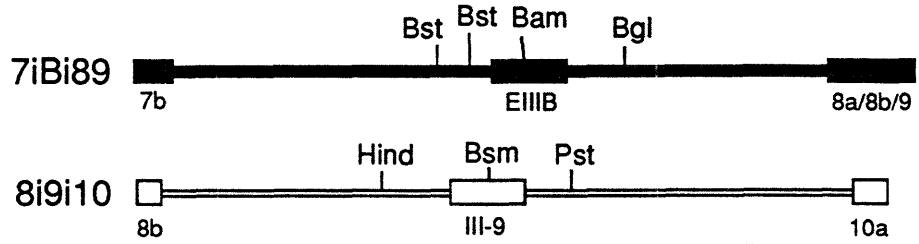
EIIIB has suboptimal splice sites and exon sequences

Based on the above results, we hypothesized that poor splice sites and/or exon sequences incompatible with splicing contributed to EIIIB insufficiency. This was tested by mutation of the EIIIB splice sites or EIIIB exon sequences (Figure 2-8A). The mutation Ad3 replaced the last 78 nt of IVS1 with the last 39 nt of the adenovirus L1-L2 intron. The 5AG mutation changed the fifth base of IVS2 from A to G (CGG/GTGAAT to CGG/GTGAGT) , improving the fit of the EIIIB 5' splice site to the consensus. The B:9 mutation precisely replaced EIIIB (273 nt) exon with III-9 (270 nt), leaving intron and splice site sequences unchanged. Mutations were made in 7iBi89dBs, a version of 7iBi89 modified to simplify mutagenesis. 7iBi89dBs behaved exactly like 7iBi89 in all the cell lines that were tested.

Figure 2-5.

Chimera analysis of EIIIB splicing: contructs, probes and summary.

Schematic representation of minigene chimeras containing parts from 7iBi89 (black rectangles) and a minigene containing the constitutive III-9 exon, 8i9i10 (white rectangles). Exons and introns are denoted respectively by wide and narrow rectangles. Restriction sites used for making chimera junctions are indicated (Bst, BstEII; Bam, BamHI; Bgl, BglII; Hind, HindIII; Bsm, BsmI; Pst, PstI). Also shown is a variant of 7iBi89, LC1, which is deleted for a 123 nucleotide BstEII-BstEII segment missing from chimeras L1 and L2. Each minigene is shown along with its S1 probe (probes consisted of the second and third exons of the minigene being analyzed, analogous to those used in Figure 2-4). A summary of the results from Figures 2-6 and 2-7 is shown; numerical values represent $exon^+/exon^-$ ratios (IN, ratio>50; OUT, ratio<0.02).



			Central Exon phenotype in:		
Minigene/ Chimera	Structure	S1 probe	COS cells	293 cells	F9 cells
7iBi89		B-89	OUT	0.08	0.4
8i9i10		9-10	IN	IN	IN
LC1		B-89	OUT	0.08	0.4
L1		9-10	IN	IN	IN
M1		9-10	IN	IN	IN
R1		B-10	OUT	OUT	OUT
SW1		B-10	OUT	OUT	OUT
L2		B-89	OUT	OUT	OUT
M2		B-89	0.39	1.8	4.4
R2		9-89	IN	IN	IN
SW2		9-89	IN	IN	IN

Figure 2-6.

Chimera analysis of EIIIB splicing, Part A:

Analysis of minigenes containing the 3' exon of 7iBi89.

S1 analyses of total RNA from untransfected cells (Ctrl) or from cells expressing each of the indicated constructs. Arrows indicate bands corresponding to undigested probe (probe) and to protection by spliced RNAs either containing ($Exon^+$) or missing ($Exon^-$) the middle exon. Data from transiently-expressing COS, 293 and F9 cells are shown (indicated above each lane) for minigenes 7iBi89, LC1, L2, M2, R2 and SW2. When normalized to co-transfected pSV2neo plasmid, minigenes containing exon III-7b expressed about five-fold lower steady-state levels of spliced RNA than those which did not; however, transfection of the highly-expressing minigenes under conditions where expression was ten-fold lower yielded identical results (see Experimental Procedures; data not shown). Exposures for minigenes 7iBi89, LC1 and SW2 were approximately five times longer than those for minigenes L2, M2 and R2 (except for F9 exposures for 7iBi89, LC1, L2, M2 and R2, which were for similar durations).

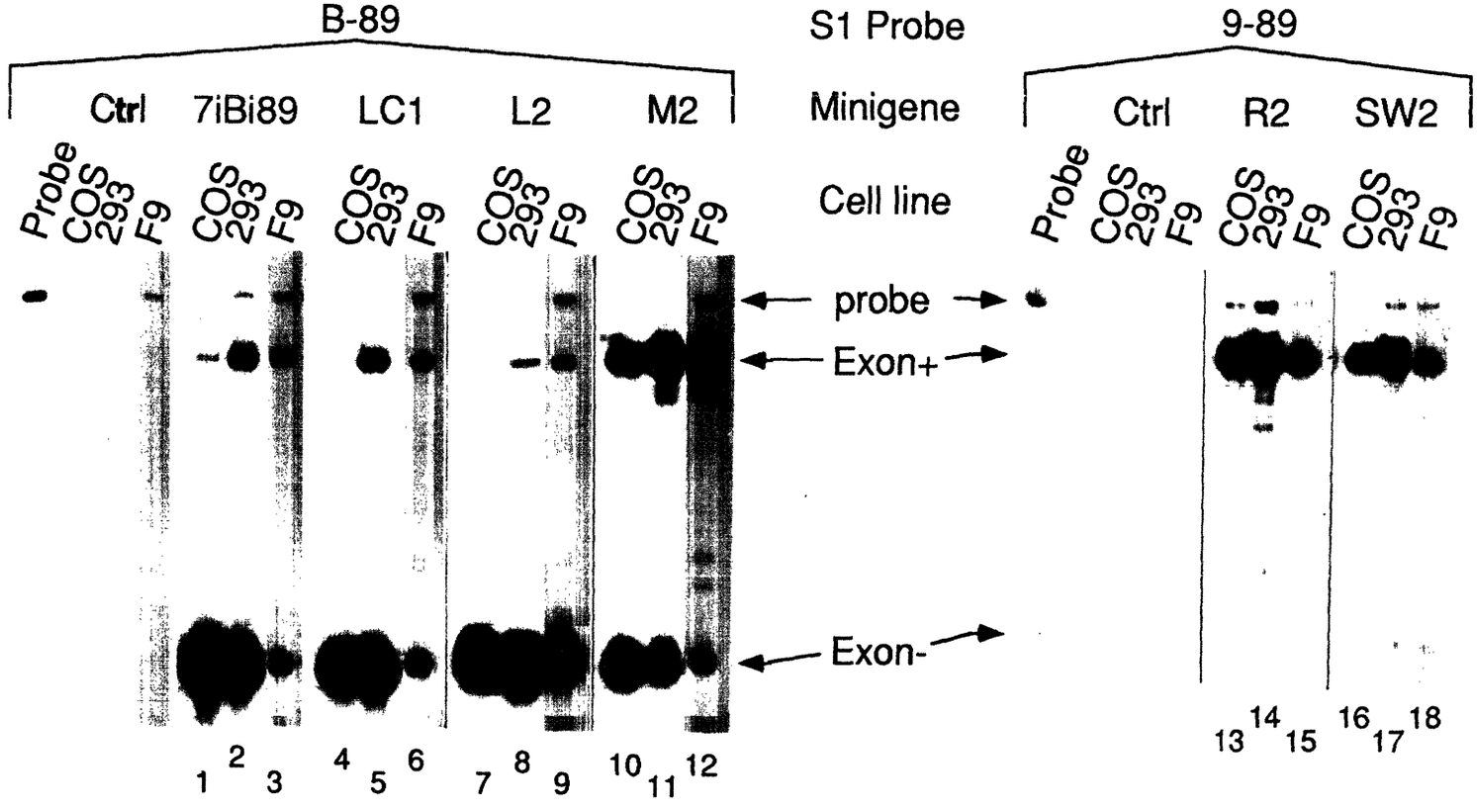

B-89
S1 Probe
9-89
Ctrl
7iBi89
LC1
L2
M2
Minigene
Ctrl
R2
SW2
Probe
COS
293
F9
Cell line
probe
Exon+
Exon-
1 2 3 4 5 6 7 8 9 10 11 12
13 14 15 16 17 18

Figure 2-7.

Chimera analysis of EIIIB splicing, Part B:

S1 analysis of minigenes containing the 3' exon of 8i9i10.

S1 analyses of total RNA from untransfected cells (Ctrl) or from cells expressing each of the indicated constructs. Conventions are as in Figure 2-6. Data from transiently-expressing COS, 293 and F9 cells are shown (indicated above each lane) for minigenes 8i9i10, L1, M1, R1 and SW1. Exposures for minigenes L1, M1 and R1 were approximately five times longer than those for 8i9i10 and SW1. Numbering of these lanes continues from Figure 2-6.

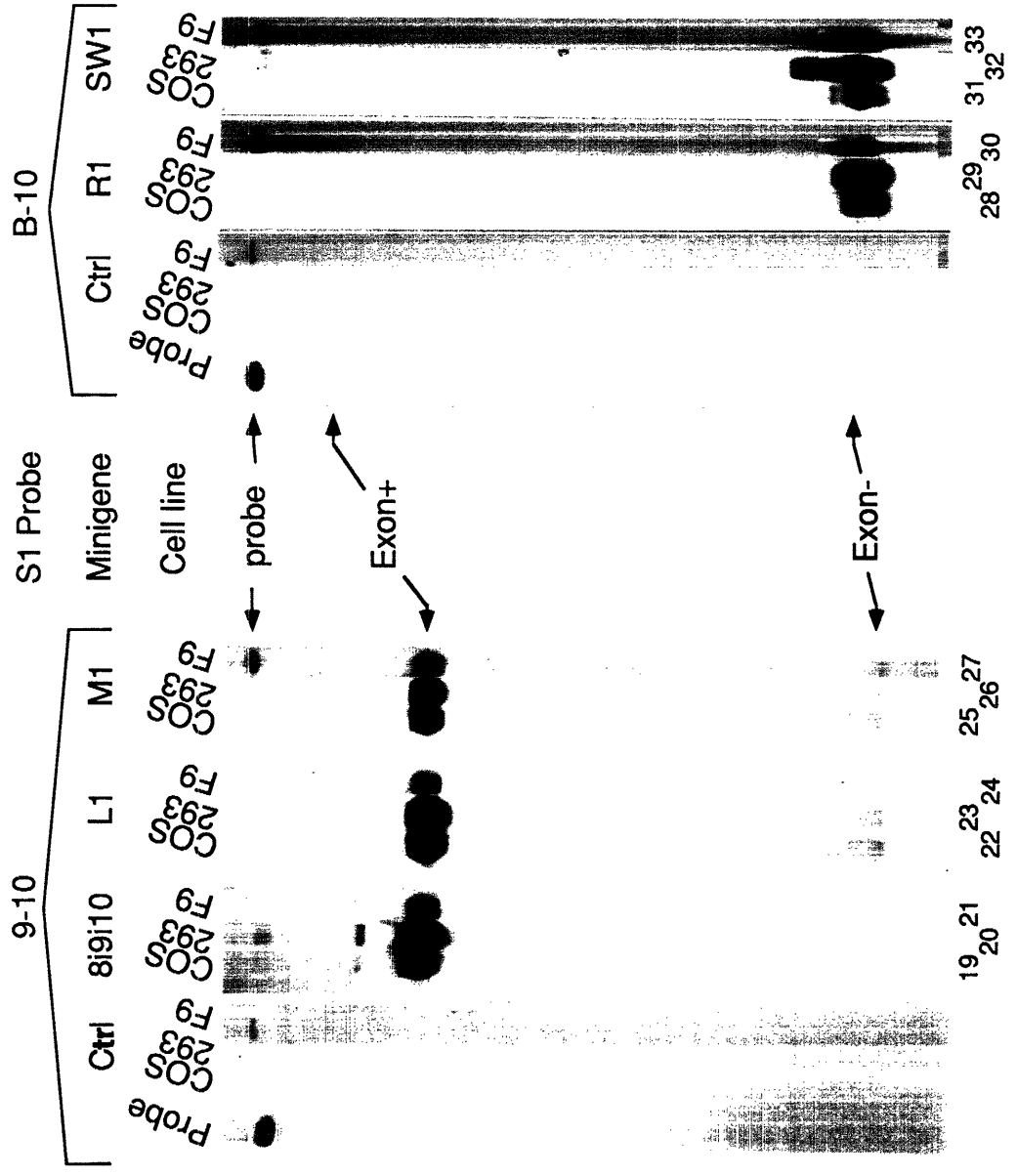
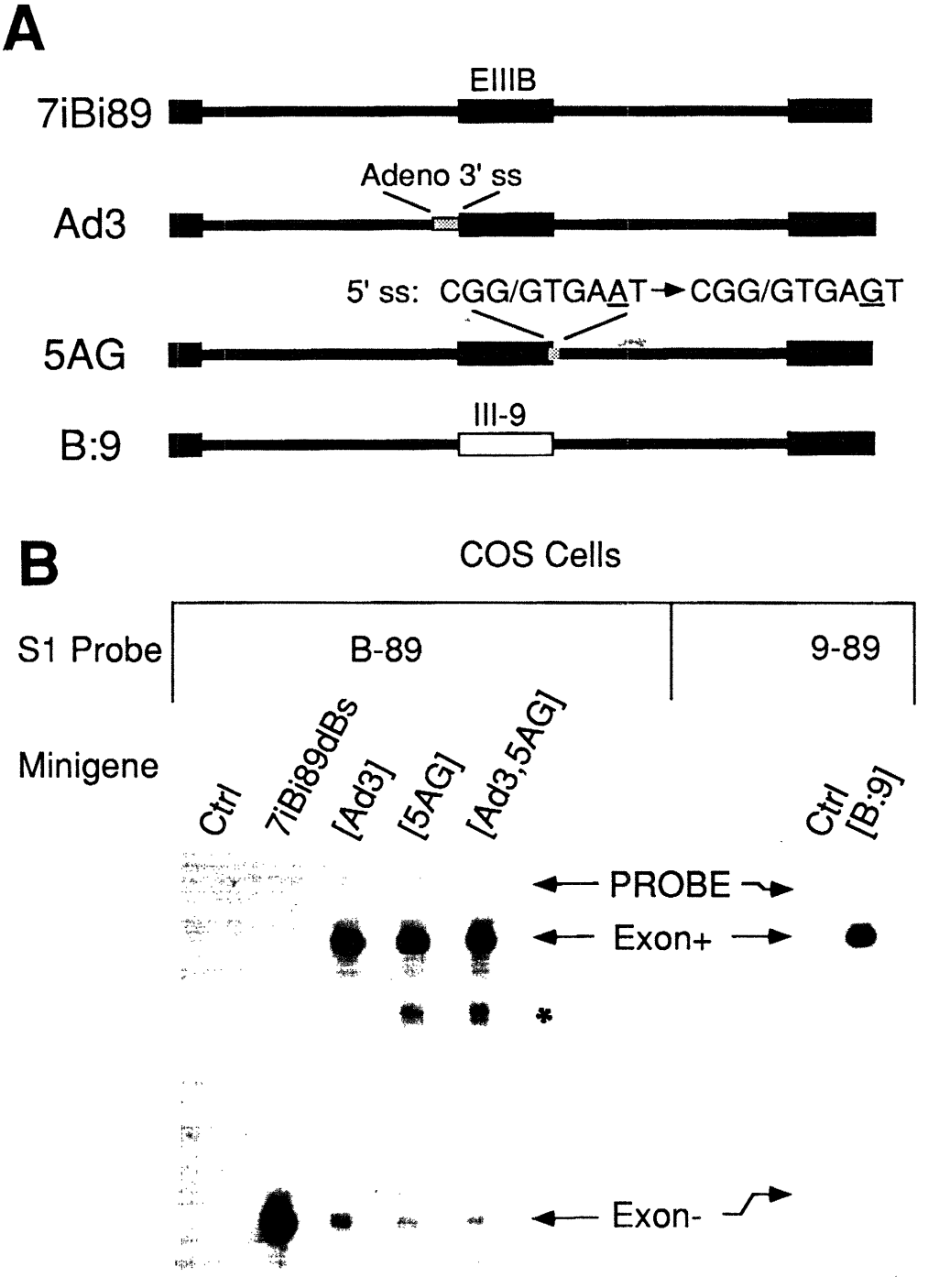

9-10
B-10
S1 Probe
Minigene
Cell line
Ctrl
8i9i10
L1
M1
Ctrl
R1
SW1
Probe
T9
A2C93S
probe
Exon+
Exon-
19 20 21
22 23 24
25 26 27
28 29 30
31 32 33

Figure 2-8.

Mutations made at the EIIIB splice sites and within exon sequences.

A. The 7iBi89dBs minigene is shown (conventions as in Figure 2-5) along with the splice site and exon changes made. 7iBi89dBs is a version of 7iBi89 in which one BstEII site was destroyed; this change did not affect regulated EIIIB splicing but facilitated mutant construction. In Ad3, the 78 nucleotides (nt) immediately upstream of EIIIB were deleted and replaced by 39 nt from the 3' splice site of the first intron from the adenovirus late leader transcript. In 5AG, the EIIIB 5' splice site was mutated from CGG/GTGAAT to CGG/GTGAGT (slash indicates exon-intron boundary), improving this site's fit to consensus. The B:9 mutation precisely replaces the entire 273 nt EIIIB exon with the 270 nt III-9 exon; B:9 does not alter intron, 3' or 5' splice site sequences. The analyses of these minigenes are shown in Figure 2-9.

B. S1 analysis of minigenes in which EIIIB splice site mutations or exon alterations were made (diagrammed in Figure 2-8). 7iBi89dBs and derived mutant minigenes (in pBAGH.Sv) were transfected transiently into COS cells; total RNA was analyzed using probes B-89 (for 7iBi89dBs, Ad3' and 5AG) and 9-89 (for B:9). Labeling conventions are as in Figure 2-6. The faint bands marked by an asterisk in the [5AG] and [Ad3, 5AG] lanes most probably represent cryptic 3' splice sites within EIIIB; these sites are used at a very low level in the unmutated 7iBi89 minigene but are more visible in constructs containing 5AG.


A
EIIIB
7iBi89
Adeno 3' ss
Ad3
5' ss: CGG/GTGAAT→CGG/GTGAGT
5AG
III-9
B:9
B
COS Cells
S1 Probe
B-89
9-89
Minigene
Ctrl
7iBi89dBs
[Ad3]
[5AG]
[Ad3,5AG]
Ctrl
[B:9]
PROBE
Exon+
*
Exon-

As shown by S1 analyses of COS transient transfections (Figure 2-8B), each of the test mutations was found to increase EIIIB inclusion (normally less than 1%) to near maximal levels (Ad3, 85%; 5AG, 90%; Ad3,5AG double mutant, ca. 95%; B:9 substitution, >98%). The 5AG and B:9 mutations had similar effects in other cell lines (not shown). We conclude that the inefficiency of EIIIB usage in COS cells is due to a combination of poor splice sites and exonic sequences which either repress, or are insufficient for, EIIIB recognition.

The EIIIB exon is not regulated in a heterologous context

If the sequences sufficient for EIIIB inclusion were local to EIIIB, then one would expect that chimeras L2, R1 and SW1 (which skipped EIIIB in COS cells) would exhibit some EIIIB inclusion in 293 and F9 cells. Surprisingly, this was not the case, as L2, R1 and SW1 expressed predominantly EIIIB-skipping patterns in all three cell types (Figure 2-5). Versions of chimeras L2 and R1 which contained more 7iBi89 intron sequences also skipped EIIIB in all cell types (data not shown); therefore additional intron sequences alone were insufficient for EIIIB recognition in EIIIB-positive cells. We wished to address more specifically the requirements for appropriate EIIIB regulation.

A balance between flanking and internal 5' splice sites is necessary for regulated EIIIB inclusion

We focused upon the region surrounding the exon upstream of EIIIB (exon III-7b; region A in Figure 2-9). Substitution of region A with analogous sequences from 8i9i10 (chimera L2) essentially abolished EIIIB inclusion in all cell types (Figure 2-10, lanes 1-3 and 7-10). In contrast, cell-type-specific EIIIB inclusion was not abolished when region A was replaced by the 5' splice site region from the human immunodeficiency virus *tat* intron (TatB; Figure 2-10,

lanes 4-6). EIIIB$^+$/EIIIB$^-$ ratios for TatB correlated well with those for 7iBi89, with EIIIB usage being even higher than in 7iBi89 in EIIIB-positive cells.

It was surprising that sequences from HIV allowed regulated EIIIB splicing whereas those from 8i9i10 did not. It was possible that the III-8b 5' splice site was more efficiently used than those of either III-7b or *tat*. If this were the case, then in chimera L2 the 5' splice site of EIIIB could have been outcompeted by a stronger one from III-8b, resulting in EIIIB-skipping in all cells even if regulatory elements were left intact. Consistent with this hypothesis, the III-8b site matches the consensus better than does III-7b (III-8b, CAG/GUGAGC; III-7b, CAG/GUAAUA; consensus, MAG/GURAGU).

If this hypothesis were true, then improving EIIIB recognition should restore regulated EIIIB inclusion in L2. To test this, we put the mutation 5AG into the L2 chimera (L2[5AG], illustrated in Figure 2-9). Transiently expressed L2[5AG] was found to display enhanced EIIIB inclusion which correlated with that of 7iBi89. Like 7iBi89, EIIIB inclusion in L2[5AG] was lowest in COS cells and highest in F9 cells, although the correlation between L2[5AG] and 7iBi89 was not complete (e.g., 7iBi89 EIIIB inclusion was higher in 293 than in HeLa cells, whereas L2[5AG] did not reflect this difference). This correlation was evident within as well as between species (e.g., murine NIH3T3 and F9 cells). These data suggest that a regulated state of EIIIB splicing is maintained in part by a balanced competition between the exon III-7b and EIIIB 5' splice sites for splicing to the exon III-8a 3' splice site.

Alteration of downstream sequences disrupts regulated EIIIB splicing

Chimera M2 appeared to exhibit exon inclusion patterns which also correlated appropriately with those of 7iBi89 (Figures 2-5 and 2-9). In M2, only the 3' 145 nucleotide portion of EIIIB and sequences downstream of EIIIB are

derived from 7iBi89. The central exon of M2 is a 259 nucleotide hybrid exon with a 3' splice site from the constitutive III-9 exon. The M2 exon, like the improved EIIIB exon in L2[5AG], competed effectively for inclusion at least part of the time. Analyses of M2 transient expression is shown in Figure 2-10 and are compared with 7iBi89 and L2[5AG] in Figure 2-9. If the fold difference between the F9 and COS exon$^+$/exon$^-$ ratios is an index of the range of differences among cell-types (i.e., the cell type-differential), then the F9/COS cell-type-differentials for M2 and L2[5AG] are similar (11-fold each). Therefore the upstream half of 7iBi89 is largely dispensable for cell-type-specific exon regulation; this half includes the EIIIB 3' splice site and the 5' portion of the EIIIB exon. Therefore regulatory elements are likely to reside in the 3' half of 7iBi89.

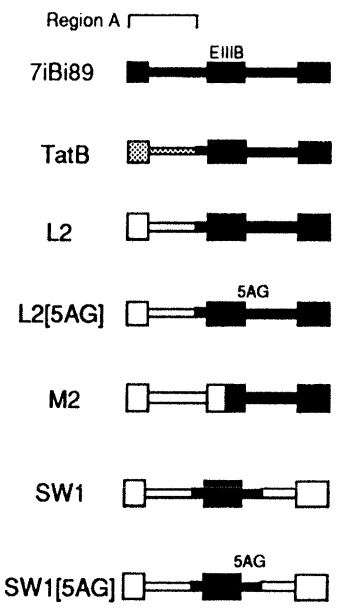
To test the contribution of sequences downstream of EIIIB, we made chimera SW1[5AG], a derivative of SW1 that contains the 5AG mutation. SW1[5AG] resembles L2[5AG] but diverges at a point 234 nucleotides 3' of EIIIB (Figure 2-9). Like L2[5AG] or M2, SW1[5AG] exhibited detectable levels of EIIIB inclusion in all cell types (Figure 2-11, lanes 23-26). However, unlike L2[5AG] and M2, a clear trend in EIIIB inclusion among cell lines was not apparent for SW1[5AG]. Although EIIIB inclusion was still lowest in COS cells, the F9/COS cell-type differential was significantly lower than those for L2[5AG] or M2; moreover, EIIIB inclusion in F9 cells did not differ from the other cell lines in a way which indicated properly regulated EIIIB splicing (Figure 2-9). The contrast between L2[5AG] and SW1[5AG] patterns suggested that the disruption of sequences downstream of EIIIB curtails the cell-type-specificity of EIIIB splicing even if EIIIB is competent for inclusion.

Figure 2-9.

Cell-specific EIIIB inclusion requires balanced splice site competition

Part A: Minigene diagrams and summary.

Diagrams and transfection summary of minigenes which exhibited different degrees of cell-specific central exon splicing. Conventions and chimeras L2, M2 and SW1 are described in Figure 2-5. Region A of 7iBi89 was replaced by the 5' splice site region from a construct containing the HIV-1 tat intron (shaded) , producing TatB. L2[5AG] and SW1[5AG] are versions of L2 and SW1 that contain the 5AG mutation (described in Figure 2-8). All values (except those denoted by asterisks) are derived from transient transfection experiments (n.t., not tested). Asterisked values represent data from stable transfections (for minigenes L2 and SW1, all values were less than 0.02 regardless of transfection protocol). Transient transfections were performed at least in duplicate; standard deviations are shown in parentheses. Also shown are fold differences in $exon^+/exon^-$ ratios between F9 and COS cells; the large deviation in the 7iBi89 F9/COS differential is due to the high relative error of the extremely low COS value.



Minigene	Exon inclusion (Exon+/Exon-)					Differential (F9/COS)
	COS	HeLa	293	3T3	F9	
7iBi89	0.008 (± 0.004)	0.03 (± 0.01)	0.078 (± 0.003)	0.05 (± 0.01)	0.41 (± 0.03)	51 (± 25)
TatB	*0.06 (± 0.01)	*0.059 (± 0.001)	*0.51 (± 0.03)	n.t.	*1.7 (± 0.1)	28 (± 5)
L2	<0.02	<0.02	<0.02	n.t.	<0.02	unknown
L2[5AG]	0.46 (± 0.02)	3.1 (± 0.6)	2.7 (± 1.3)	2.5 (± 0.5)	5.2 (± 0.9)	11 (± 2)
M2	0.39 (± 0.02)	1.8 (± 0.1)	1.8 (± 0.1)	2.5 (± 0.2)	4.4 (± 0.6)	11 (± 2)
SW1	<0.02	<0.02	<0.02	n.t.	<0.02	unknown
SW1[5AG]	0.74 (± 0.04)	3.1 (± 0.1)	2.6 (± 0.5)	n.t.	2.5 (± 0.3)	3.4 (± 0.4)

Figure 2-10.

Cell-specific EIIIB inclusion requires balanced splice site competition

Part B: S1 analysis of minigenes TatB, L2, L2[5AG] and M2.

S1 analyses, using probe B-89. Figure conventions are as in Figure 2-6. Indicated above each panel is the chimera being tested and the S1 probe for that chimera. Cell lines are indicated above each lane; exposures were adjusted to facilitate comparison between cell lines. For technical reasons, transient expression of TatB was not feasible; therefore both stable and transient transfections of L2 are shown in order to allow comparison between L2 and TatB (stable) and between L2 and L2[5AG] (transient). Asterisks mark bands representing cryptic 3' splice sites within EIIIB (seen also in Figure 2-8); intensities of these bands tended to parallel the amount of EIIIB$^+$ RNA present.

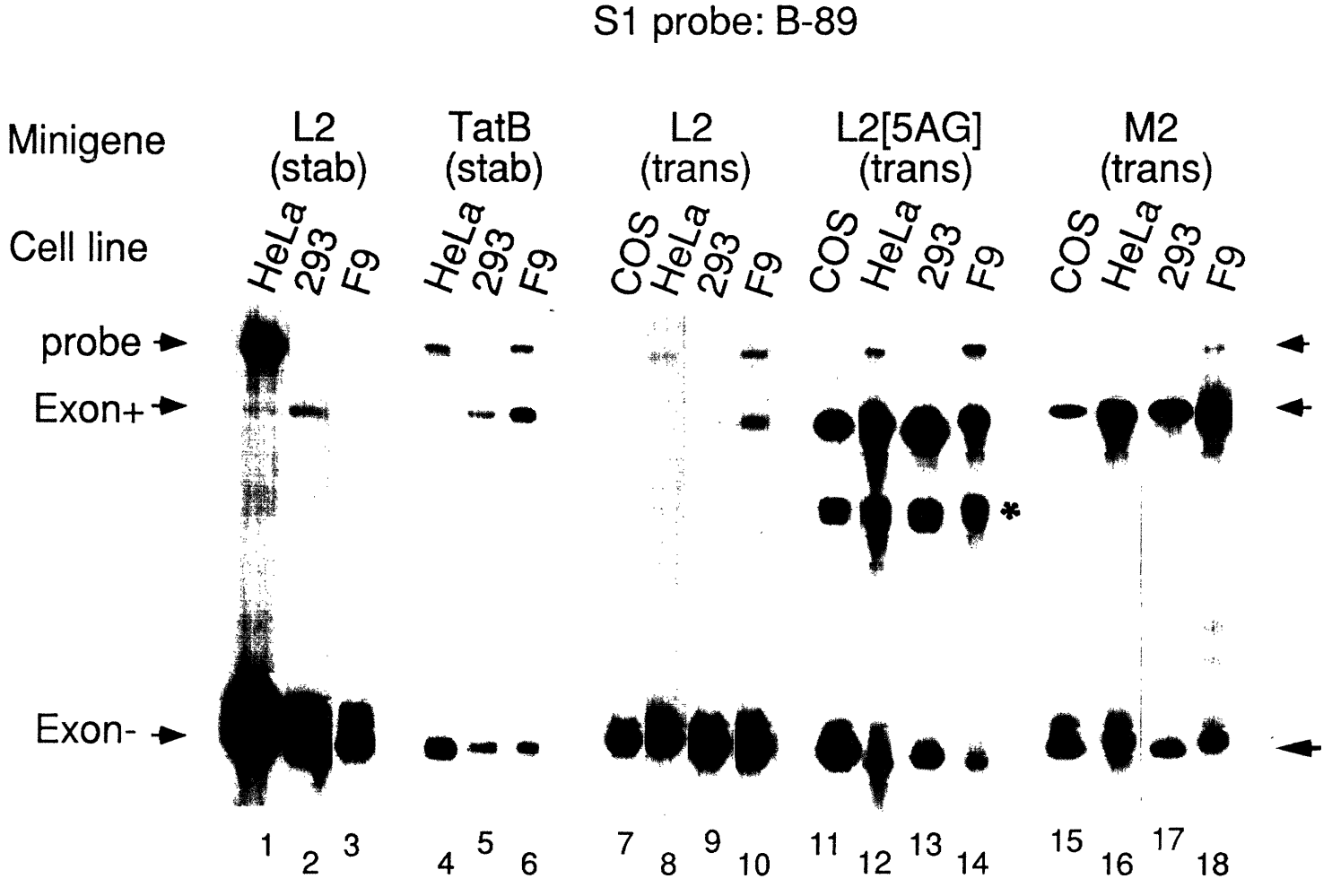
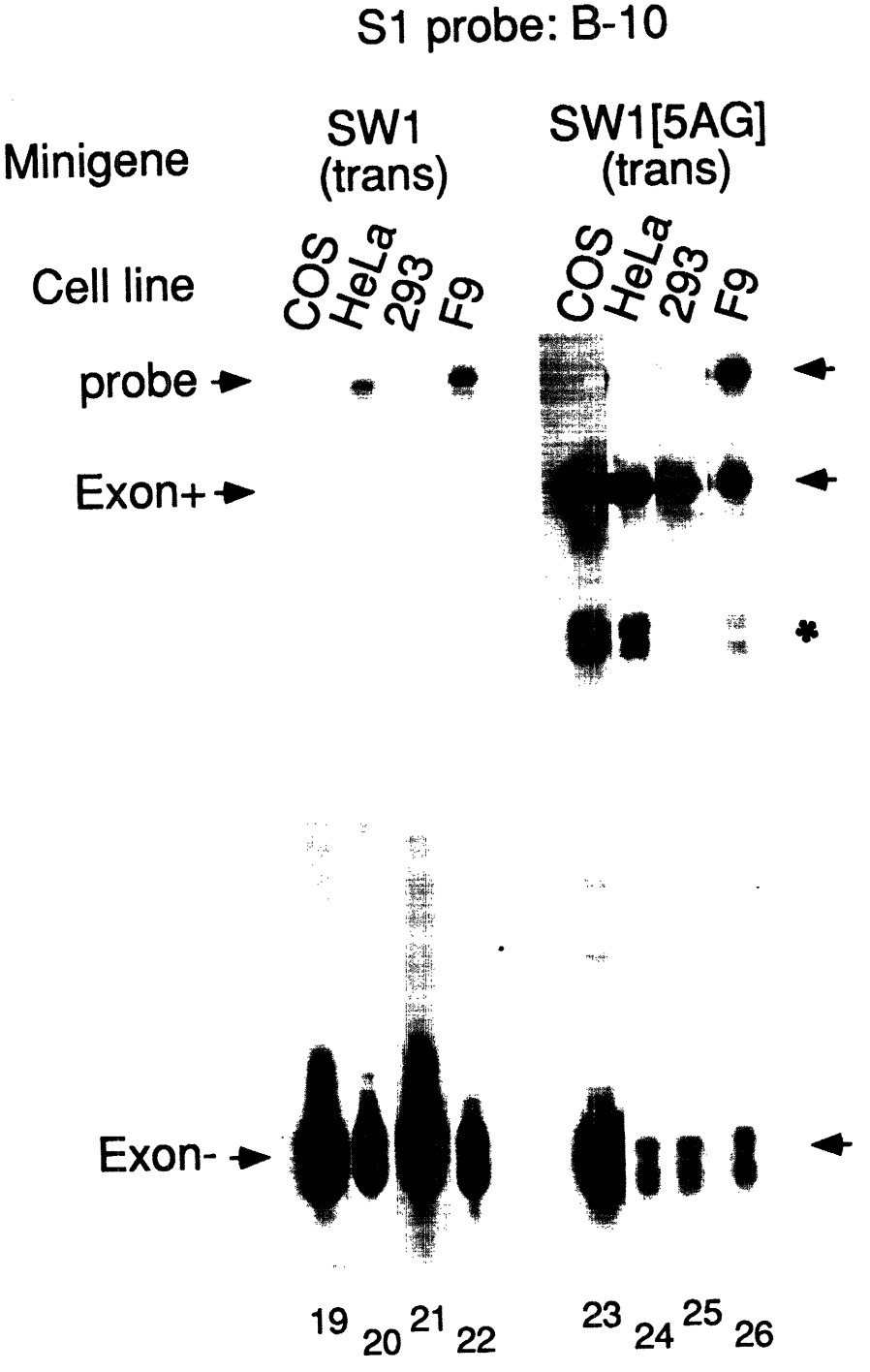

S1 probe: B-89
Minigene
L2 (stab)
TatB (stab)
L2 (trans)
L2[5AG] (trans)
M2 (trans)
Cell line
HeLa
293
F9
COS
probe
Exon+
Exon-
*
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Figure 2-11.

Cell-specific EIIIB inclusion requires balanced splice site competition

Part C: S1 analysis of minigenes SW1 and SW1[5AG].

S1 nuclease analyses, using probe B-10. Figure conventions are as in Figures 2-7 and 2-10; lane numbering continues from Figure 2-10.


S1 probe: B-10
Minigene
SW1 (trans)
SW1[5AG] (trans)
Cell line
COS
HeLa
293
F9
COS
HeLa
293
F9
probe
Exon+
*
Exon-
19
20
21
22
23
24
25
26

Downstream intron sequences are required for EIIIB inclusion

To examine further the participation of downstream sequences in differential cell type-specific exon regulation, various deletions in 7iBi89 were examined for their effects upon cell-type-specific EIIIB splicing. Figure 2-12 illustrates the 7iBi89 deletions and a summary; Figure 2-13 and 2-14 show S1 analyses of stable transfections of these variant minigenes in HeLa, 293 and F9 cells.

An extensive deletion in IVS1 (XE; Figure 2-12) did not alter EIIIB regulation significantly. Although EIIIB inclusion is slightly reduced in 293 cells, inclusion in F9 cells is as high as for unaltered 7iBi89. A 100 nt deletion within EIIIB (HB) activated EIIIB inclusion somewhat in all cell types; however, differences among cell lines were unaffected. Therefore these sequences in IVS1 and EIIIB have little influence upon EIIIB regulation. Similarly, deletions within the 5' half of IVS2 (GL1, GR1, GR2) did not significantly affect EIIIB splicing, although they did slightly elevate EIIIB usage in 293 and F9 cells.

In contrast, deletions within or extending into the 3' portion of IVS2 (GR3, AL1, AL2, AL3, AR1, AR2a) displayed a markedly negative effect upon EIIIB usage (Figure 2-12). This effect was observed only in cell types that normally exhibited EIIIB inclusion (293 and F9 cells); the EIIIB-skipping pattern expressed by HeLa cells was unaltered. Therefore these deletions disrupted a region required for EIIIB inclusion but not for EIIIB skipping. More extensive deletions in this region appeared to have greater effects than smaller ones. Deletion AL1 exhibits a moderate reduction in relative EIIIB$^+$ levels in 293 and F9 cells; longer deletions in this region (GA, GR3, AL2, AL3) almost completely abolished EIIIB inclusion, as did deletions extending toward the 3' splice site of exon III-8a (AR1, AR2a). A point insertion of four nucleotides at an AflII site in this region (dAfl) did not affect EIIIB inclusion, suggesting that the region

required for EIIIB$^+$ splicing is not continuous. Therefore EIIIB recognition appears to be sensitive to gross deletions of certain sequences in IVS2. The effects upon EIIIB inclusion are not due simply to shortening of IVS2, since (i) IVS2 deletions of similar size have different effects (compare GR2 versus AL1 and AR1) and (ii) restoration or extension of IVS2 length with heterologous intron sequences did not restore EIIIB inclusion (compare GA, GA+SmAc and GA+EH). These deletions are intronic; therefore it is unlikely that this effect on EIIIB inclusion is at the level of cytoplasmic RNA stability or transport. These deletions define a region of IVS2 required for EIIIB recognition in EIIIB-positive cells (Figure 2-12). The 5' endpoint of the region containing these elements lies between 445 and 689 nucleotides downstream of EIIIB; the 3' endpoint of this region is undefined.

Disruption of IVS2 sequences attenuates cell-type-specific differences in exon recognition

We focused attention upon this region in IVS2, as it represented potential *cis*-sequences which promote EIIIB usage. If this region acts nonspecifically, then its disruption should reduce EIIIB inclusion uniformly in all cell types. Alternatively, if this region activates EIIIB usage in a cell-type-specific fashion (e.g., if it contains targets for cellular factors that activate EIIIB) then removal of this region should reduce EIIIB recognition specifically in EIIIB-positive cells. It was impossible to distinguish between these two possibilities by using the 7iBi89 minigene. Since this minigene already exhibited undetectable levels of EIIIB inclusion in HeLa cells, any nonspecific effects of IVS2 deletions would go unnoticed.

To distinguish between nonspecific and cell-type-specific EIIIB activation, some of the IVS2 deletions shown in Figure 2-12 were transferred

into the regulated chimeric minigene M2. Since M2 exhibits detectable exon inclusion even in COS cells, any effect of IVS2 deletions upon differential inclusion among cell types could be quantitated. We also tested additional deletions in M2, designed to localize putative *cis*-elements further. Figure 2-15 shows the deletions made in M2 and histograms of the results; Figure 2-16 shows S1 analyses of some of these M2 variants after transient transfection into COS, HeLa and F9 cells.

As mentioned previously, the IVS2 deletion GR2 slightly activated EIIIB inclusion in the minigene 7iBi89 (Figure 2-12); this deletion in M2 (M2GR2) increased M2 exon recognition in all three cell lines (Figure 2-15). Relative to M2, M2GR2 exhibits uniformly raised exon$^+$/exon$^-$ ratios; however, cell-type-differentials for M2 and M2GR2 (as assessed by the fold difference between the F9 and COS exon$^+$/exon$^-$ ratios) were similar, with M2GR2 exhibiting only a modest increase in cell type-differential compared with M2. In contrast, other IVS2 deletions reduced the F9/COS differential significantly. Deletions GR2a, GR3b, GR3a, GR3, GA, GA+SmAc and AL2 reduced the F9/COS differential by at least two-fold and, in one case, by as much as four-fold (M2AL2). The reduction in cell type-differential was primarily due to a reduction of M2 exon inclusion in F9 cells; inclusion in COS cells was relatively unaffected. Deletions that reduced the F9/COS differential also attenuated the HeLa/COS differential. In contrast, deletions AL1, AR1 and AR2a had little effect upon M2 exon recognition, despite their marked effect in 7iBi89. None of these deletions completely eliminated the F9/COS differential.

The deletions that reduced the F9/COS differential defined a 122 nt region in IVS2, between 519 and 640 nt downstream of EIIIB (the 3' endpoint of deletion GR2b and 5' endpoint of deletion AL1; Figure 2-15) which was important for F9-specific (and, to a lesser extent, HeLa-specific) stimulation of

M2 exon usage. We have designated this element the ICR (intronic control region).

Figure 2-12.

Diagrams of 7iBi89 minigene deletions and summary

The 7iBi89 minigene and its deletion variants are shown (conventions as in Figure 2-5). Deleted areas are indicated by blank regions within each minigene. Vertical dotted lines indicate sites used for generating IVS2 deletions: BglII (labeled G: GL1, GR1, GR2, GR3) and AflII (labeled A: AL1, AL2, AL3, AR1 and AR2a). Deletion GA deletes the 495 bp DNA segment between the BglII and AflII sites; GA+SmAc and GA+EH are variants of GA into which intronic fragments from other regions of the rat FN gene were inserted (shown as stippled bars numbered as to their length); GA+SmAc restores IVS2 to 1096 nucleotides; GA+EH expands IVS2 to 1704 nucleotides. Deletions were made in 7iBi89/pBAGH and are drawn to scale; the 3' endpoint of deletion AR2a lies 32 nt upstream of exon III-8a. Beside each deletion is a semiquantitative indication of the $EIIIB^+/EIIIB^-$ ratio: -, ratio<0.05; +, 0.05<ratio<0.15; ++, 0.15<ratio<0.3; +++, 0.3<ratio<0.5; ++++, 0.5 <ratio<0.8; +++++, ratio >0.8. The region of IVS2 required for $EIIIB^+$ splicing is indicated below this diagram (black rectangle with undefined 3' endpoint).

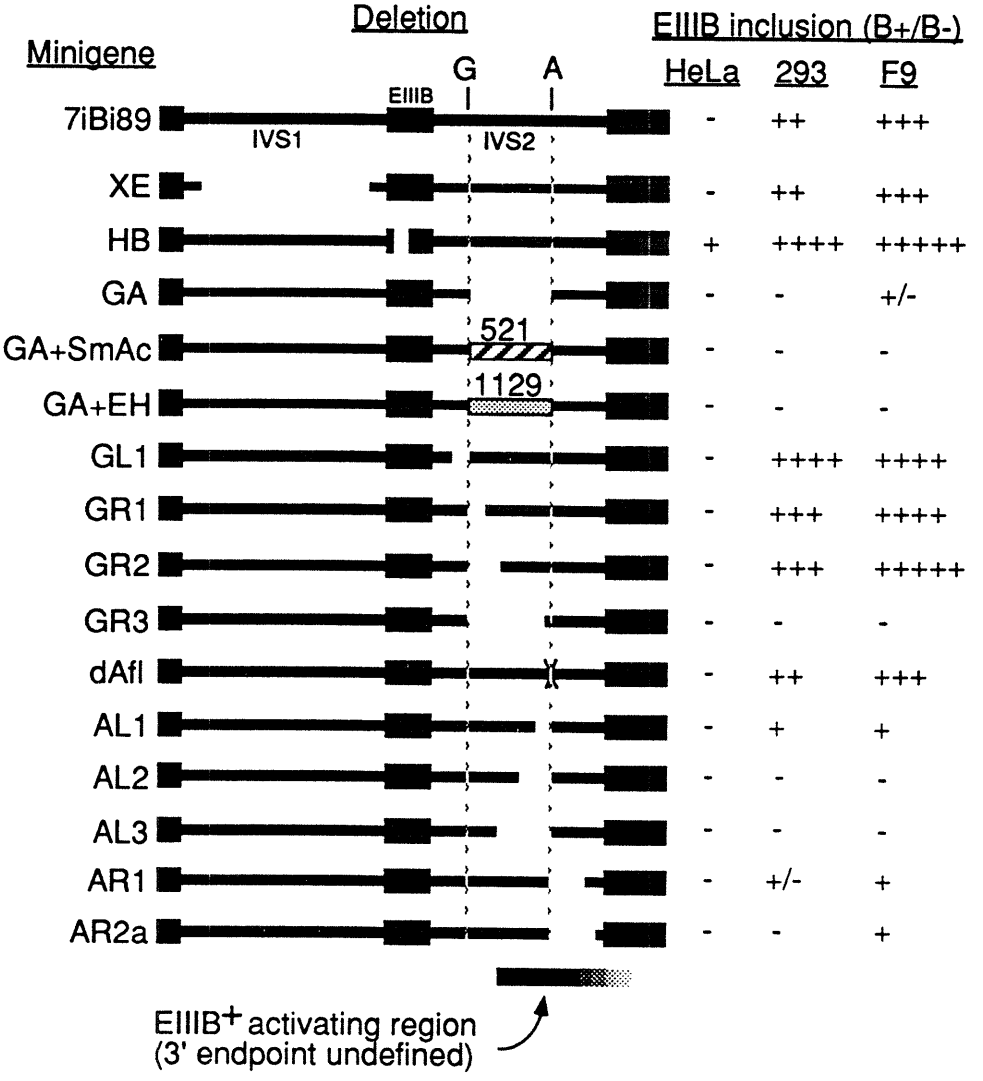

Minigene
Deletion
EIIIB inclusion (B+/B-)
G
A
EIIIB
HeLa
293
F9
7iBi89
IVS1
IVS2
XE
HB
GA
GA+SmAc
521
GA+EH
1129
GL1
GR1
GR2
GR3
dAfl
AL1
AL2
AL3
AR1
AR2a
EIIIB+ activating region
(3' endpoint undefined)

Figure 2-13.

Deletion analysis of 7iBi89: HeLa and 293 cells.

S1 nuclease analyses of stably-transfected 7iBi89 deletion mutants. 35 μg cytoplasmic RNA (HeLa) or 5 μg total RNA (293 cells) were analyzed with the S1 probe B-89. Constructs (described in Figure 2-13) are indicated above each lane; other figure conventions are as in Figure 2-4. Each cell line is represented by a separate panel. Transient and stable transfections of 293 cells yielded similar results; in addition, normalization to cotransfected pSV2neo showed no gross changes in steady state RNA levels between constructs (data not shown but see Chapter Three, Figure 3-5).

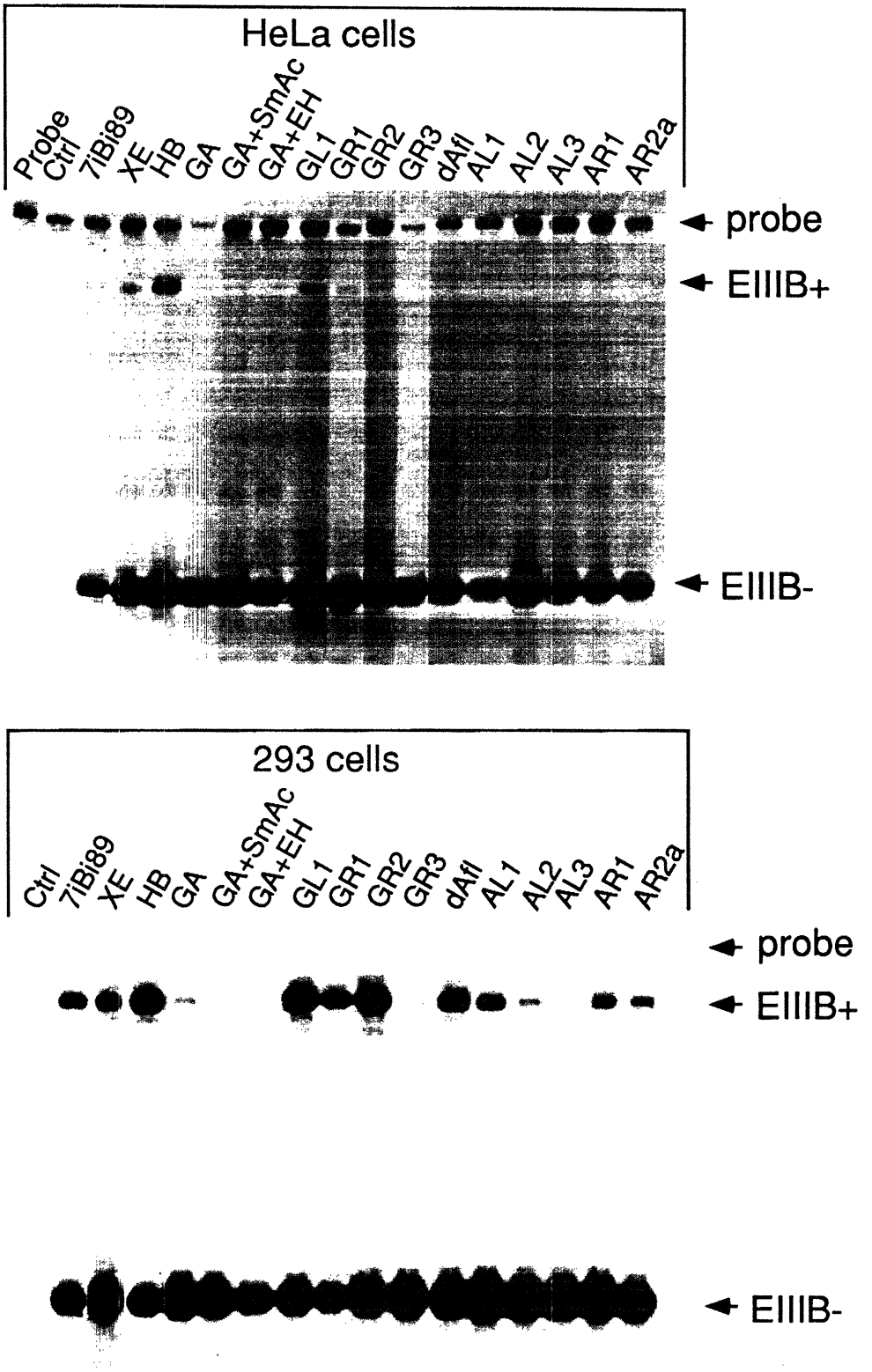

HeLa cells
Probe
Ctrl
7iBi89
XE
HB
GA
GA+SmAc
GA+EH
GL1
GR1
GR2
GR3
dAf1
AL1
AL2
AL3
AR1
AR2a
probe
EIIIB+
EIIIB-
293 cells
Ctrl
7iBi89
XE
HB
GA
GA+SmAc
GA+EH
GL1
GR1
GR2
GR3
dAf1
AL1
AL2
AL3
AR1
AR2a
probe
EIIIB+
EIIIB-

Figure 2-14.

Deletion analysis of 7iBi89: F9 cells.

S1 nuclease analyses of stably-transfected 7iBi89 deletion mutants. 35 μg of F9 cytoplasmic RNA from each stable transfectant was analyzed with the S1 probe B-89. Constructs (Figure 2-12) are indicated above each lane; other figure conventions are as in Figure 2-4.

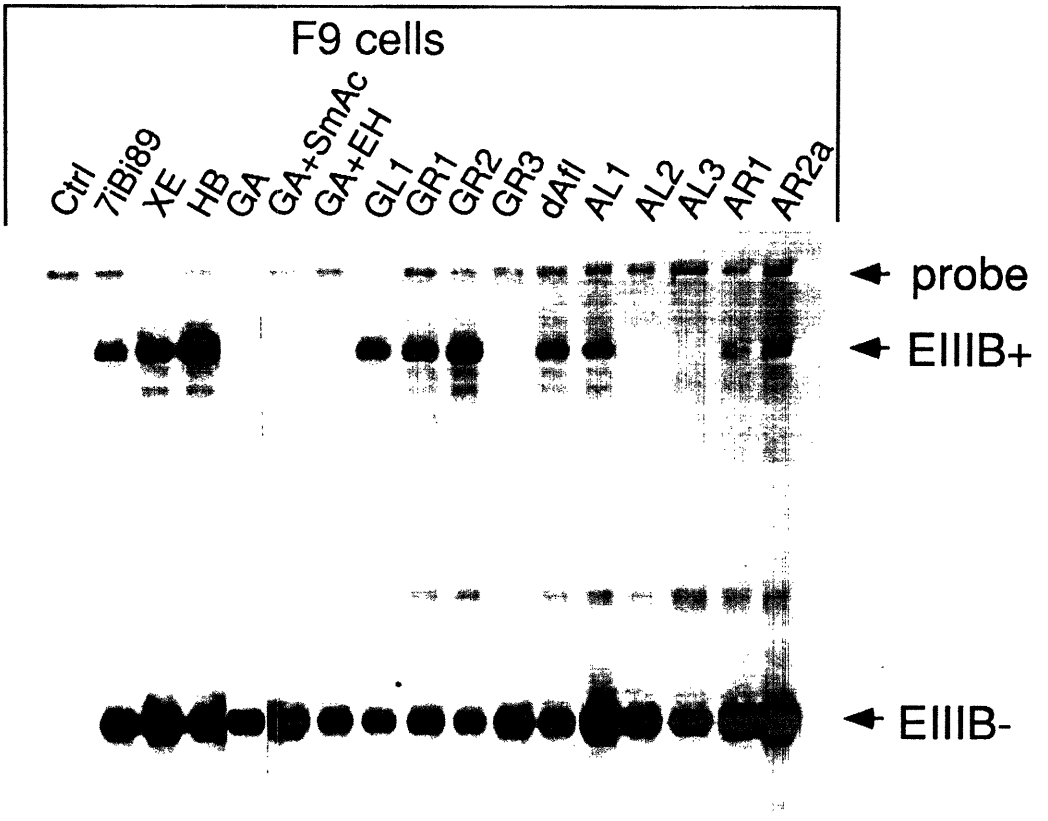

F9 cells
Ctrl
7iBi89
XE
HB
GA
GA+SmAc
GA+EH
GL1
GR1
GR2
GR3
dAf1
AL1
AL2
AL3
AR1
AR2a
probe
EIIIB+
EIIIB-

Figure 2-15.

Deletion analysis of the M2 minigene: constructs and summary

Diagram of the M2 minigene and its deletion variants; diagram conventions are as in Figure 2-12. Also shown are histogram representations of M2 exon splicing (as $exon^+/exon^-$ ratios) in COS (white bars), HeLa (shaded bars) and F9 cells (black bars). Values shown represent at least two independent transient transfections; error bars indicate standard deviations. F9/COS differentials for each M2 variant are also shown. The segment of IVS2 most important for the F9/COS differential (the intronic control region or ICR), is indicated below the minigene diagrams (shaded rectangle).

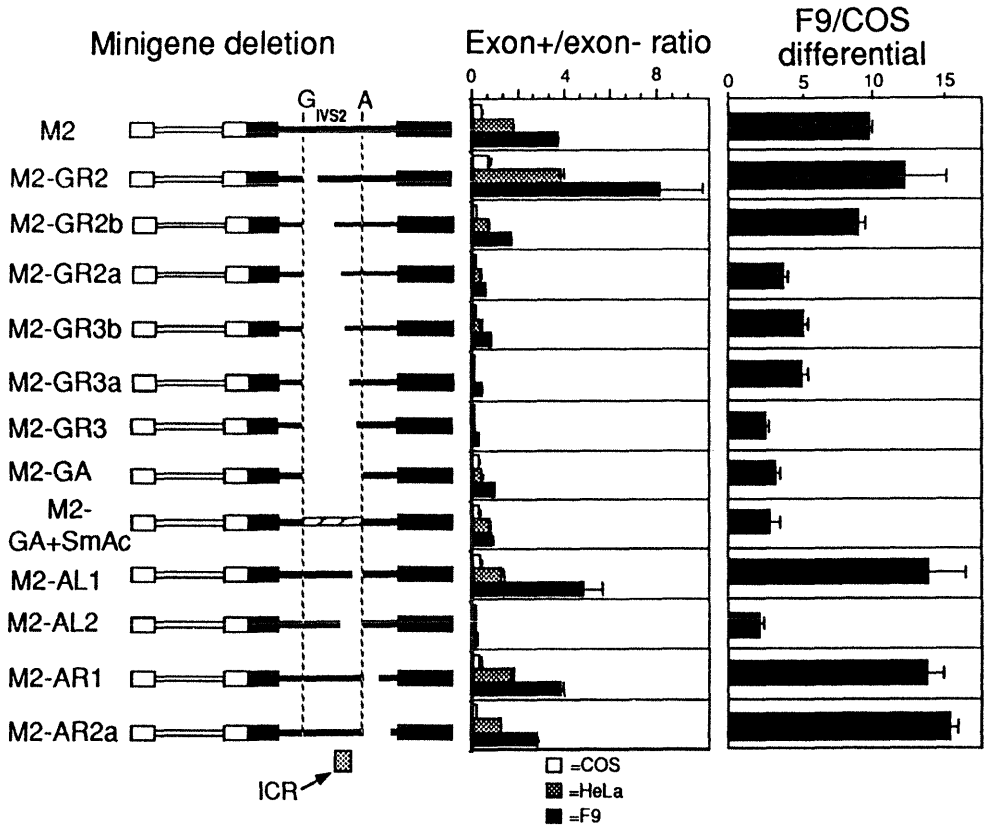
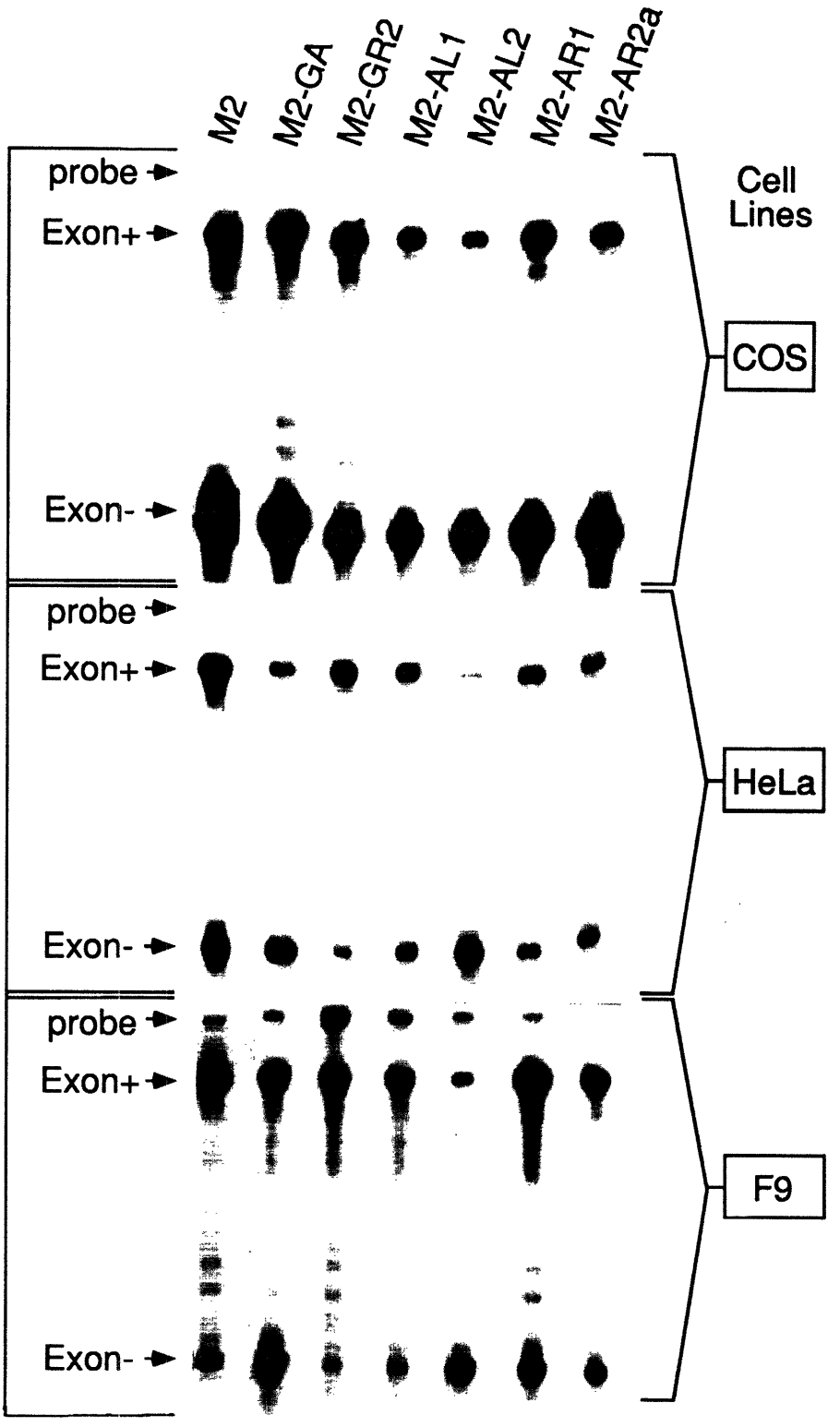

Minigene deletion
Exon+/exon- ratio
F9/COS differential
G
IVS2
A
M2
M2-GR2
M2-GR2b
M2-GR2a
M2-GR3b
M2-GR3a
M2-GR3
M2-GA
M2-GA+SmAc
M2-AL1
M2-AL2
M2-AR1
M2-AR2a
ICR
0
4
8
0
5
10
15
=COS
=HeLa
=F9

Figure 2-16.

Deletion analysis of the M2 minigene: S1 analysis

S1 analyses of M2 deletion mutants. Probe B-89 was used for this assay. M2 variants were made in the vector pBAGH.Sv. Total RNA from transiently transfected COS (5 μg), HeLa (35 μg) and F9 cells (35 μg) were analyzed as indicated. Other labeling conventions are as in Figures 2-13 and 2-14.


M2
M2-GA
M2-GR2
M2-AL1
M2-AL2
M2-AR1
M2-AR2a
Cell Lines
probe
Exon+
Exon-
COS
probe
Exon+
Exon-
HeLa
probe
Exon+
Exon-
F9

DISCUSSION

Transfection of minigenes into different cell lines has enabled the identification of certain *cis*-acting influences upon cell-type-specific alternative splicing of the EIIIB exon in the rat FN gene. A rat FN-derived EIIIB minigene expressed spliced EIIIB$^+$ and EIIIB$^-$ transcripts in proportions which correlated with endogenous FN EIIIB splicing. We asked a number of questions regarding the requirements for EIIIB regulation in different cell types.

EIIIB is a poorly recognized exon

By replacing regions of the EIIIB minigene with analogous regions from a constitutive exon-containing minigene, we were able to localize sequences around EIIIB that are important in the exon-skipping event. Our results suggest that EIIIB is an intrinsically poor splicing substrate. The EIIIB exon is skipped even when placed between two heterologous exons; therefore EIIIB, within a 590-nucleotide segment, is not recognized in a heterologous context. Inefficient EIIIB recognition is due to exonic sequences and suboptimal splice sites, since improvement or alteration of any of these elements switched EIIIB splicing patterns from exclusively EIIIB$^-$ to almost completely EIIIB$^+$ in COS cells. The EIIIB exon sequences may harbor information in *cis* that represses use of nearby splice sites; alternatively, EIIIB may simply lack RNA sequences or higher order structures that normally permit efficient splice site usage. The influence of sequence context upon alternative splice site usage has been documented by many studies (for examples, see Mardon et al., 1987; Eperon et al., 1986; Cooper and Ordahl, 1989; Somasekhar and Mertz, 1985; Streuli and Saito, 1989). Therefore these findings were not overly surprising. The nature of these effects upon splice site selection is at present relatively ill-defined.

Extended pyrimidine-rich tracts, like that of the EIIIB 3' splice site, occur in a number of alternative 3' splice sites in genes such as rat α-tropomyosin (Smith and Nadal-Ginard, 1989) and rat/chicken β-tropomyosin (Helfman and Ricci, 1989; Goux-Pelletan et al., 1990). Despite its longer pyrimidine tract, the EIIIB 3' splice site is weaker than those of either adeno L2 or the III-9 exon (Figure 2-8 and compare chimeras L2 and M2 in Figure 2-5). The EIIIB 3' splice site is therefore suboptimal; however, it is not solely responsible for repression of EIIIB, since other mutations in EIIIB can independently promote EIIIB inclusion. These findings agree with studies of exons 7/6B in the rat/chicken β-tropomyosin genes. In these studies, a negative effect on 3' splice site efficiency was attributed to sequences in and around the extended pyrimidine tract, since deletions or mutations in these regions activated use of exon 7 or 6B (Goux-Pelletan et al., 1990; Libri et al., 1990; Helfman et al., 1990). In contrast, the long pyrimidine tract preceding exon 3 in the α-tropomyosin gene had a positive effect in determining default selection of exon 3 over exon 2 in vitro (Mullen et al., 1991). It is unclear why long pyrimidine tracts have different effects in different contexts, although the repression of EIIIB or of β-tropomyosin exons 6B/7 probably involves additional sequences not found in α-tropomyosin exon 3.

The downstream half of 7iBi89 is sufficient for regulation of splicing

An EIIIB-containing 590-base segment was sufficient for EIIIB-skipping in COS cells. However, EIIIB was not recognized when this segment was placed between two heterologous exons, even in EIIIB-positive cells. Therefore information beyond this segment was required for EIIIB inclusion. One chimera, M2, exhibited regulated inclusion of its hybrid exon which correlated with that of 7iBi89. Therefore the downstream half of 7iBi89 is sufficient to

direct regulation of exon inclusion, although we cannot rule out the presence of other nonessential regulatory elements upstream of EIIIB.

Cell-type-specific EIIIB regulation does not absolutely require the EIIIB 3' splice site with its associated extended pyrimidine tract. This contrasts sharply with work done with the rat and chicken β-tropomyosin genes, in which extended pyrimidine tracts precede exon 7 in rat (Helfman et al., 1990) or its chicken homologue 6B (Goux-Pelletan et al., 1990). Since mutations in these tracts derepressed exon 6B/7 usage, sequences in these tracts were proposed to inhibit in a cell-type-specific manner, either by recruiting repressor factors (Guo et al., 1991; Helfman et al., 1990) and/or by forming RNA secondary structures which block exon usage in nonmuscle cells (Libri et al., 1991; D'Orval et al., 1991). Our results suggest that the EIIIB 3' splice site is dispensable for exon regulation, although it is possible that this splice site sequence supplies an auxiliary regulatory component.

Regulated EIIIB inclusion requires 5' splice site balance

We have concluded that EIIIB regulation is dependent upon a balance between internal and flanking exon donor strengths. The suboptimal nature of the exons flanking EIIIB in the FN gene plays an important role in establishing a state where competing splice sites are balanced with respect to the weak EIIIB exon; without this balance, regulation is lost even if EIIIB-promoting cellular regulatory factors are present, as discussed below.

The replacement of the 5' exon region in 7iBi89 with two other 5' exon regions gave strikingly contrasting results. A 5' exon from HIV-1 *tat* preserved cell line-specific EIIIB regulation (TatB), whereas that from 8i9i10 abolished EIIIB inclusion in all cell types (L2). In the latter case, improvement of EIIIB recognition (L2[5AG]) had a compensatory effect, restoring EIIIB inclusion and

regulation (Figures 2-9, 2-10). We already knew that the entire 5' half of 7iBi89 was not essential for cell-type-specific regulation (from minigene M2). However, in order for regulated EIIIB inclusion to be seen, the exon upstream of EIIIB must have a suboptimal 5' splice site comparable in strength to that of EIIIB. The 5' splice site derived from the *tat* intron probably met this requirement, since it is intrinsically inefficient (Chang and Sharp, 1989).

The fact that the hybrid M2 central exon is regulated suggests that the M2 minigene also maintains a balance between internal and terminal exon strengths. The M2 exon is always detectably included, even though its flanking exons are identical to those of L2; this is probably because the M2 central exon's 3' splice site is normally constitutive. The cell type differentials for M2 and L2[5AG] are similar, suggesting again that the EIIIB 3' splice site (which is not present in M2) does not participate directly as a cell-type-specific regulatory component.

Suboptimal splice sites are commonplace in alternative splicing. Proper retroviral replication requires inefficient splicing, since both spliced and unspliced retroviral RNAs encode essential gene products (Katz and Skalka, 1990). The suboptimal polypyrimidine tract of the *Drosophila doublesex* female-specific 3' splice site is required for sex-specific regulation (Hoshijima et al., 1991). The neural-specific *c-src* N1 exon still exhibits neural specificity when placed between two adenovirus exons; however, inclusion of N1 is considerably lower than that of N1 in its normal *c-src* context, indicating that flanking exons in the *c-src* gene contribute to normal N1 exon regulation (Black,1991). Similar findings have been obtained in studies of the *Drosophila Sex-lethal* male-specific exon (Horabin and Schedl, 1992). Although other studies have altered the sequences flanking a regulated exon without effect (Streuli and Saito, 1989; Tacke and Goridis, 1991), the effect of context may not

be as critical in these systems as it is for EIIIB. These results imply that even a slight alteration in splice site strength might have sizeable effects on EIIIB splicing; in addition, our findings suggest that the regulation of an alternative exon can potentially be regulated by elements influencing the splice sites of flanking exons, a mechanism which has not been seriously considered by many to date.

IVS2 sequences are essential for regulated EIIIB recognition

Since replacement of splice site sequences could potentially interfere with any interpretation that we could easily make, we adopted a different approach by carrying out a deletion analysis that removed regions other than the splice sites themselves. Surprisingly, deletions within or extending into the 3' half of IVS2 were found to dramatically reduce EIIIB inclusion (Figures 2-12 through 2-14). Therefore the integrity of IVS2 was essential for EIIIB recognition. The critical region extends over a large portion of IVS2; many of the deletions which abolished EIIIB inclusion were located more than 200 nucleotides from either IVS2 splice site. This region is distinct from EIIIB and begins at least 445 nucleotides into IVS2 ; it is not continuous, since a small 4 nucleotide insertion within this region does not alter EIIIB regulation.

To investigate the role of this region of IVS2 in cell-type-specific EIIIB regulation, we monitored the effect of IVS2 deletions in the context of the regulated M2 minigene. The results (Figure 2-15, 2-16) confirmed that certain IVS2 deletions reduced exon recognition in all the cell-types we tested. Moreover, the effect of these deletions was greater in F9 cells than in COS cells; as a consequence, the cell-type differential between F9 and COS cells (our index of cell-type-specificity) was decreased. The region most critical for cell-type-specific M2 exon regulation is at most 122 nucleotides long (the intronic

control region or ICR; Figure 2-15). Interestingly, the ICR is much smaller than the region required for EIIIB inclusion in the 7iBi89 minigene. These results are consistent with the presence of cellular factors in F9 cells that promote EIIIB inclusion in a manner contingent upon the ICR. ICR deletions also reduced the HeLa/COS differential, suggesting that EIIIB-promoting factors may also be present at some low level in HeLa cells and probably at very low or zero levels in COS cells.

The ICR is at least 519 nucleotides downstream from EIIIB; it is unusually distant from the exon it regulates. In fact, the region required for 7iBi89 EIIIB inclusion is considerably closer to exon III-8a than to EIIIB. As such, our results contrast with studies of exons 4 and 6 in the leukocyte common antigen gene (Streuli and Saito, 1989; Saga et al., 1990), exon E18 in the murine N-CAM gene (Tacke and Goridis, 1991) and exon N1 in the murine *c-src* gene (Black, 1991, 1992). In these systems the elements that are sufficient for cell-type-specific exon regulation lie closely proximal to the regulated exon; in the case of EIIIB, it is apparent that a *cis*-element can act on an alternative exon from a relatively long distance either directly or indirectly, possibly via some novel regulatory mechanism .

Do redundant elements regulate EIIIB?

None of the ICR deletions completely abolishes the M2 exon cell-type-differential, indicating that the ICR is not the sole determinant of cell-type-specific exon regulation. In addition, not all of the deletions that abolish EIIIB inclusion (in 7iBi89) affect M2 exon regulation. It is possible that the deletions that abolish exon inclusion in 7iBi89 but not in M2 (AR1, AL1, AR2a) confer nonspecific effects on splice site selection. Alternatively, multiple EIIIB-activating elements may exist in IVS2, only a subset of which affect the M2

minigene. The ICR may simply represent the most potent (and hence most easily detected) of these EIIIB-activating sequences. Redundancy would explain why the M2 cell-type differential was not completely abolished by ICR deletions, since ICR removal may unmask minor elements located elsewhere. The regulation of *Drosophila dsx* alternative splicing involves six 13-nucleotide repeats located downstream of the female-specific 3' splice site; the Tra and Tra-2 proteins activate this splice site in a repeat-dependent fashion in vivo and in vitro (Ryner and Baker, 1991; Hedley and Maniatis, 1991; Hoshijima et al., 1991; Tian and Maniatis, 1992). Neuron-specific N1 splicing in *c-src* involves at least two sequence elements, either of which was sufficient for N1 exon inclusion in vivo and in vitro; like EIIIB, these lie in the intron downstream of the regulated exon (Black, 1992). Therefore the cell-type-specific control of EIIIB splicing by several sequence signals within IVS2 would not be without precedent.

Models for EIIIB regulation

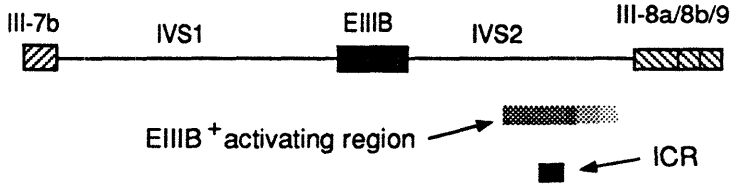
These tissue culture studies have given us a better understanding of how cis-acting components in the alternatively spliced EIIIB region play a role in splice site selection. The components we have identified fall into two classes: those in the first class are important for maintaining a state of splice site competition upon which cell-type-specific regulatory influences can be exerted. The second, more interesting class consists of sequences which influence splice site selection in a cell-type-specific fashion; these may represent targets for interaction with trans-acting factors specific to EIIIB$^+$ cells.

Possible models for regulation of EIIIB splicing are outlined in Figure 2-17. In the default state (e.g., COS or adult hepatocytes), the EIIIB exon is a poor splicing substrate. This exon has suboptimal splice sites; in particular, its

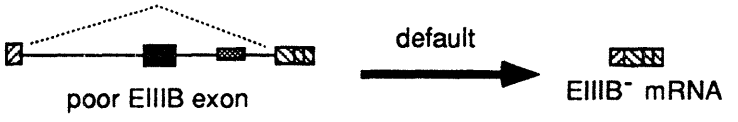
Figure 2-17.

EIIIB regulation: splicing models

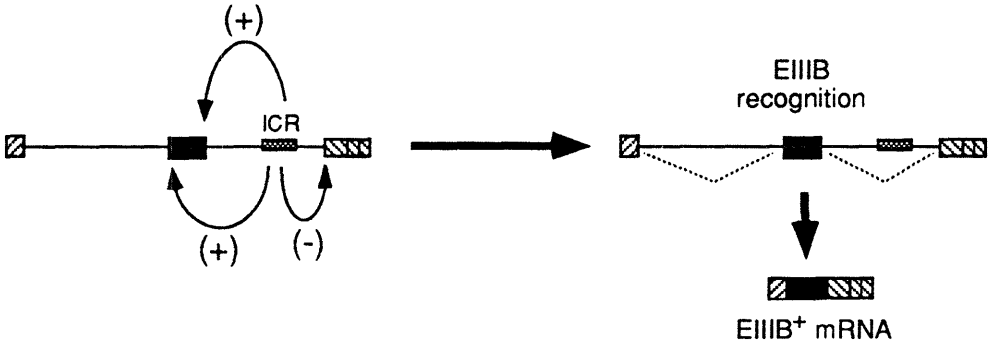
Shown is the EIIIB minigene 7iBi89 and potential mechanisms of EIIIB-skipping (in COS, HeLa or liver cells) and EIIIB inclusion (in early embryos or F9 cells). Bars below the 7iBi89 map indicate the positions of (i) the IVS2 region required for EIIIB inclusion (mapped in 7iBi89) and (ii) the putative intronic control region or ICR (mapped in M2). Curved arrows indicate mechanisms by which the ICR (or ICR-bound cellular factors) might affect splice site selection in order to enhance EIIIB inclusion ([+] or [-] indicate stimulatory or inhibitory effects, respectively).



EIIIB-skipping (adult liver, COS/HeLa cells):



EIIIB inclusion (early embryos, F9/293 cells):



3' splice site has an unusual structure which may curtail its efficiency. These properties, in conjunction with unfavorable exonic context, create a poorly recognized exon that is skipped. In an EIIIB-positive cell (e.g., F9 or early embryo), EIIIB is somehow relieved from its repressed state and can compete more effectively with its flanking exons, resulting in inclusion. Cell-type-specific derepression depends upon sequences in the ICR. The splice sites of flanking exons III-7b and III-8a appear to be of adequate strength to prevent their own inappropriate skipping, but are not so strong as to overwhelm an enhanced (but still weak) EIIIB exon. Cell-type-specific factors may enhance EIIIB inclusion via binding to the ICR. Alternatively, the ICR may render the FN pre-mRNA transcript more sensitive to differential concentrations of general splicing factors such as ASF/SF2 (Ge and Manley, 1990; Krainer et al., 1990). Possible splicing models for EIIIB recognition include, but are not limited to: (i) EIIIB 5' splice site activation, possibly by facilitating binding of U1 (Kuo et al., 1991) or other accessory splicing components; (ii) EIIIB 3' splice site activation (although a heterologous 3' splice site must be accomodated) or (iii) negative effects leading to a delay in exon III-8a 3' splice site usage, creating a kinetic window that permits EIIIB recognition before a commitment to EIIIB-skipping can occur. The third of these mechanisms is intriguing, particularly since the region of IVS2 required for EIIIB inclusion is closer to III-8a than to EIIIB. In some systems the alteration of a common 3' splice site is capable of changing 5' splice site preference; these phenomena could arguably be explained by alterations in the order or kinetics of splice site recognition (Ulfendahl et al., 1989; Fu et al., 1988). The above models are not mutually exclusive; in fact, multiple mechanisms may collaborate to regulate EIIIB more tightly in different cell types.

MATERIALS AND METHODS

Expression Vectors

Enzymes for DNA work were obtained from New England Biolabs and Boehringer Mannheim. Molecular biological manipulations were carried out as described by Ausubel et al (1987) and Sambrook et al (1989). Two expression vectors were used in this study: pBAGH and pBAGH.Sv for stable and transient transient expression respectively (Figure 2-1B). Both contained a modified human β-actin promoter and a short 3' untranslated region from the human growth hormone gene. pBAGH.Sv contained in addition the SV40 origin of replication for amplification in COS cells. The vector backbone and human β-actin promoter of pBAGH were from BAPGAL, a plasmid containing the human β-actin promoter (Leavitt et al., 1984), the lacZ gene and SV40 late 3' polyA signal (a gift from Urban Lendahl, Karolinska Institute). The human growth hormone 3' untranslated region came from pLENX, a variant of pLEN with a XhoI linker inserted at the BamHI cloning site (a gift from Ty White, Calbiotech; J.-L. Guan, pers. comm.); pLENX contains the SV40 enhancer and origin of replication, human metallothionein-II promoter and human growth hormone 3' untranslated region. A 6.2 kb SalI-EcoRI fragment containing the BAPGAL backbone and β–actin promoter was ligated to the 635 bp XhoI-EcoRI 3' polyA-containing segment from pLENX (destroying the SalI and XhoI sites), generating BH, an expression vector with a BamHI cloning site. The promoter fragment in BH contains the first exon and intron from the human β–actin gene. This intron and four nucleotides of the first exon were deleted via exoIII digestion, generating pBAGH. To make pBAGH.Sv, the 835 bp HindIII-XhoI promoter fragment in pLENX was replaced by a 4 kb EcoRI-SalI fragment with the β–actin promoter from BAPGAL, making LB (destroying all sites involved).

A XhoI-BamHI fragment of LB was then replaced with the XhoI-BamHI fragment from vector pBAGH, in effect deleting the β–actin intron and generating pBAGH.Sv.

Minigenes and Probes for Analysis

The primary 3.1 kb minigene of this study, 7iBi89, was derived from genomic clone λrFN3 and cDNA clones BXB32.5 and BdP (Schwarzbauer et al, 1987). A 2.8 kb XmnI-ApaI λrFN3 fragment (with 125 nucleotides [nt] of exon III-7b, the 1296 nt intron following III-7b [IVS1], EIIIB, the 1071 nt intron following EIIIB [IVS2] and 68 nt of exon III-8a) was ligated to an ApaI-AvaI cDNA fragment containing the remaining 115 nt of III-8a, the 90 nt exon III-8b and 73 nt of exon III-9. The presence of additional cDNA sequences in the third exon of 7iBi89 was necessary for efficient splicing. 7iBi89 was constructed between the XbaI and SmaI sites of pGEM3 (Promega). The minigene 8i9i10 contains a 2.9 kb BglII-BglII fragment from λrFN3 (with 91 nt of exon III-8b, the 1140 nt intron following III-8b, the 270 nt exon III-9, the 1332 nt intron following III-9 and 17 nt of exon III-10a) ligated to an BglII-EcoRI fragment containing the remaining 100 nt of III-10a and 21 nt of exon III-10b. 8i9i10 was made between the BamHI and SmaI sites of pGEM3. Minigene inserts (SalI-SacI fragments) were subcloned into the BamHI cloning sites of pBAGH and pBAGH.Sv via blunt-end ligation; orientations were checked by restriction digests.

Chimeric minigenes were made in pGEM3 and subsequently transferred into the appropriate expression vectors as above. With respect to 7iBi89, chimera junctions were: (i) 207 or 84 nucleotides 5' of EIIIB for chimeras L1 or L2 respectively (upstream of the pyrimidine tract and three mapped branch points of the EIIIB 3' splice site; Norton and Hynes, 1990), (ii) 234 nucleotides 3'

of EIIIB (chimeras R1/R2), as well as (iii) within EIIIB (127 nucleotides from the 5' end of EIIIB; chimeras M1/M2). With respect to 8i9i10, junctions were: 286 nt upstream of III-9 (L1/L2), 114 nt downstream from the 5' end of III-9 (M1/M2) and 161 nt downstream of III-9 (R1/R2). Chimeras SW1 and SW2 share junctions with pairs L1/R2 and L2/R1 respectively. L1 and L2 were made by swapping the BstEII-SacI fragment of 7iBi89 with the HindIII-SacI fragment of 8i9i10 (in Figure 2-5, the SacI site is at the 3' terminus of both 7iBi89 and 8i9i10 inserts). LC1 was made by cutting 7iBi89 with BstEII and religating, deleting 123 bp. M1 and M2 were made by exchanging the BamHI-SacI fragment of 7iBi89 with the BsmI-SacI fragment of 8i9i10. R1 and R2 were made by exchanging the BglII-SacI fragment of 7iBi89 with the PstI-SacI fragment of 8i9i10. SW1 was made by cloning the EIIIB-containing 590 bp BstEII-BglII fragment of 7iBi89 into a HindIII/PstI-cut 8i9i10 in a version of pGEM3 without polylinker HindIII and PstI sites; SW2 was made by cloning the III-9-containing 720 bp HindIII-PstI fragment of 8i9i10 into BstEII/BglII-cut 7iBi89. Incompatible restriction enzyme ends were blunted with Klenow fragment before ligation; orientations were checked by restriction digests.

TatB (Figure 2-9) was made by blunt-end cloning the BstEII-EcoRI fragment from 7iBi89/pBAGH (84 nt of IVS1, EIIIB, IVS2, exon 8a/8b/9 and the 3' region from human growth hormone) into the BamHI and PstI sites of pSPL1, a vector with the SV40 early promoter and an intron-containing fragment of the HIV-1 *tat* gene (Buckler et al, 1991). This created a SV40 promoter-driven minigene in which the first exon of pSPL1 (rabbit β-globin exons 1 and 2 fused to 240 nucleotides of the *tat* 5' exon) and first 305 nucleotides of the *tat* intron were ligated to the last 84 nt of IVS1 in 7iBi89.

In general, S1 probes (containing the third exon and part of the central exon from each minigene as cDNA; Figure 2-5) were subcloned from rat FN

cDNA clones into pGEM3; S1 probes were made minigene-specific by kinasing at a BamHI site within vector sequences. S1 probes 9-89 and B-10 (for which natural cDNA clones were not available) were made by inverse PCR using primers complementary to exon termini; primers for the III-9/III-8a junction (for probe 9-89) were 5'-CCGTGGATTGCTGGCCAATC-3' and 5'-CCGTCCCTCCTCCCACGGAT-3'. The inverse PCR product of minigene R2 was blunt-ended, religated and transformed; the appropriate segment was subcloned to make 9-89. B-10 was made similarly to 9-89, except that the primers for the EIIIB/III-10a junction were 5'-CCGTTTGCTGTGTCAGTGTA-3' and 5'-TTTCCGATGTCCCGAGAGAT-3'; minigene R1 was the PCR substrate. The above probes were kinased at a BamHI site (Sambrook et al., 1989) and then recleaved with PvuII. Probe E was kinased at an AvaI site and then recut with NheI. S1 probes were purified via preparative 5% nondenaturing polyacrylamide gel electrophoresis.

RNase protection probes were transcribed (with [^{32}P]-UTP [NEN] and T7 RNA polymerase [Stratagene]) from a SalI-linearized, shortened version of 7iBi89 in pGEM3. For probing 7iBi89 (Figures 2-1A, 2-2), the XE deletion in IVS1 (Figure 2-12) was combined with a BglII-DraI deletion in IVS2. These riboprobes were purified via preparative electrophoresis on a 4% or 5% denaturing polyacrylamide gel.

Mutagenesis and Deletions

Oligonucleotide-directed mutagenesis was carried out using uracil-containing ssDNA templates generated by E. coli strain CJ236 (Kunkel et al, 1987; Ausubel et al, 1987). The 590 bp EIIIB-containing BstEII-BglII segment of 7iBi89 was subcloned into Bluescript SK$^-$ (Stratagene) to make BsG. Single-stranded, uracil-containing template contained the strand antisense to EIIIB.

The mutagenesis primer used for making the 5AG mutation had the sense sequence 5'-AAACGGGTGAGTCTTGAAGTC-3' (the underlined G is the mutated base). The B:9 mutation was made using the "sticky-feet"-directed mutagenesis procedure described by Clackson and Winter (1989); the "forward" and "reverse" primers used were
5' TTGCCTCCCCTTTGCTTCATAACTCAATAGGTCTGGACTCCCCAACTG GT 3' and
5' GTCTCAAACACAGAAGACTTCAAGATTCACCCGTGGATTGCTGGCCA ATC 3'. The Ad3 mutation was made by PCR of BsG using a primer containing 39 nucleotides of the adenovirus L2 3' splice site followed by 18 nucleotides at the start of EIIIB (primer sequence
5' CGATGATGTCATACTTATCCTGTCCCTTTTTTTTCCACAGAGGTGCCCC AGCTCACTG 3') and a reverse primer within EIIIB
(5' GTGATGCGGTACCCAATAATG 3'); the 145 bp PCR fragment was blunted with Klenow, cut with HindIII and ligated to 7iBi89dBs (see below) which had been BstEII-cut, blunted and then recut with HindIII; this procedure regenerated the BstEII site. All mutants were sequenced in pBluescript SK- before insertion into a modified 7iBi89 minigene (7iBi89dBs, in which the 5' BstEII site was destroyed to facilitate reinsertion of BsG into 7iBi89 with inimal sequence alteration). Mutant minigenes were then cloned into pBAGH.Sv as described above.

Deletion XE removed a 1116 bp BstXI-BstEII fragment from IVS1. Deletion HB removed nucleotides 28 to 127 of EIIIB (a 100 bp HindIII-BamHI segment). Deletions within IVS2 were made by exoIII deletion of the 7iBi89 insert in pBAGH. BglII- or AflII-cleaved plasmid was treated at 30°C with exonuclease III according to Ausubel et al (1987). Aliquots of the exo III digest (at 15-second intervals) were purified and cleaved with XhoI; appropriate

fragments were gel-purified and ligated with 7iBi89/pBAGH (cut with AflII or BglII, blunted and cut with XhoI). The resulting libraries contained unidirectional deletions extending in either direction from the AflII or BglII sites. Clones were characterized by sequencing and restriction mapping. IVS2 deletions in this study (in 7iBi89 or in M2) removed the following nucleotides from the 1071 base IVS2 (numbering from the first nucleotide in IVS2): GR2, 235-445; GR2b, 235-518; GR2a, 235-573; GR3b, 235-642; GR3a, 235-669; GA, bases 235-729; AL1, 641-729; AL2, 417-729; AR1 734-969; AR2a, 734-1039. Deletions within IVS2 of the M2 minigene were made by subcloning the BglII-EcoRI fragment of 7iBi89 IVS2 deletions into BglII/EcoRI-cut M2/pBAGH.Sv.

Cell Culture and Transfections

COS cells (from Benjamin Geiger, Weizmann Institute) and Rat-1 cells (clone F2408) were cultured in DMEM (Gibco) with 5% fetal bovine serum (FBS; Hazleton). HeLa cells (from Phillip Sharp, MIT) were cultured in DMEM with 10% FBS. 293 cells (from Earl Ruley, MIT) and NIH3T3 cells (ATCC) were cultured in DME with 10% calf serum (CS; Sigma, Hazleton). F9 teratocarcinoma cells (from Lorraine Gudas, Harvard Medical School) were cultured in MEMa with 7.5% CS and 2.5% FBS (Robertson, 1987); differentiation of F9 cells into parietal endoderm was carried out with 0.1 μM retinoic acid and 1 mM dibutyryl cyclic AMP for 72-96 hours (Strickland et al, 1980)

COS cells were transfected using DEAE-dextran. Half-confluent COS cells in 10 cm dishes were transfected in 2 ml Tris-saline (25 mM Tris-HCl pH 7.4, 0.14 M NaCl, 3 mM KCl, 1 mM $CaCl_2$, 0.9 mM Na_2HPO_4) for 40 minutes at 37°C, followed by a 3.5 to 4 hour treatment with 100 μg/ml chloroquine (in DMEM+5% FBS); the cells were then washed and refed. Calcium phosphate

transfections of HeLa and 293 cells were coupled with a one minute glycerol shock (Ausubel et al, 1987). F9 transfections were performed as described by Gorman et al (1985). Transient transfections used 18 μg expression plasmid DNA per 10 cm dish (23 μg for F9 cells) plus 2 μg pSV2-neo (Southern and Berg, 1982) included as an internal control. RNAs were isolated 48-52 hr post-transfection. Transfections for generating G418-resistant cell populations were done in 6 cm dishes using 9 μg of minigene plasmid plus 1 μg pSV2-neo. Selection in G418 (Gibco) was carried out for 10-16 days; colonies (100-1000) were pooled and RNA isolated after 1-2 additional days. G418 concentrations used were: for HeLa, 293, NIH3T3 and Rat-1, 400 μg/ml ; for F9, 300 μg/ml. Cytoplasmic RNA was isolated as described (Ausubel et al 1987); total RNA was isolated either by centrifuging through a 5.7 M CsCl cushion (Chirgwin et al, 1979) or acid phenol/guanidinium (Chomczynski and Sacchi, 1987); RNAs from transient transfections were DNased before analysis (Ausubel et al, 1987).

In some experiments, transfections were carried out in which RNA expression levels were ten-fold below optimal. For COS, RNA was harvested 12 hr (rather than 48 hr) post-transfection; for 293 and F9, only 2 μg of expression plasmid was used (adding carrier pGEM3 to keep total DNA at 20-25 μg).

Nuclease protection analyses

S1 analysis was carried out as described by Sambrook et al (1989) with the following modifications. RNA samples (5-35 μg) were made up to 35 μg with yeast RNA; samples and probe (5'-end-labelled; 2-5x10^5 cpm per sample) were dried down, dissolved in S1 hybridization buffer, heated to 80°C for 10 minutes, placed in a 65°C water bath and cooled to 43-48°C overnight. S1 digestion was in 275 μl using 400-800 units S1 nuclease (Boehringer-

Mannheim) at 37°C for 30 minutes; products were ethanol-precipitated and analyzed on 4-5% acrylamide gels containing 8M urea and were visualized using autoradiography. Appropriate bands were quantitated using a Molecular Dynamics Phosphorimager. RNase protection analyses were carried out as described by Ausubel et al (1987), except for the RNase digestion buffer composition (50 mM Tris-HCl, pH 7.5; 500 mM NaCl); RNase-resistant fragments were analyzed on 5% denaturing gels and quantitated as above.

ACKNOWLEDGEMENTS

We are very grateful to Urban Lendahl, Ty White, Alan Buckler, Phillip Sharp, Earl Ruley, Lorraine Gudas and Benny Geiger for gifts of plasmids and for cell lines used during the course of this work. We also wish to thank Pamela Norton, Mike DiPersio and Charles Query for many helpful comments and suggestions on the manuscript from which this chapter was derived. Additional thanks go to Pamela Norton and Phillip Sharp for experimental advice and suggestions.

Chapter Three

Cell-Type-Specific Regulation Of Alternative Splicing By A Novel Repeated Hexanucleotide Element

ABSTRACT

The alternatively spliced exon EIIIB is regulated in a cell-type-specific manner in the rat fibronectin gene. Splicing of EIIIB into fibronectin mRNA is dependent upon sequences in the intron immediately downstream of EIIIB. We show that a short, highly repeated TGCATG motif in this intron is important for cell-type-specific recognition of EIIIB as an exon. This motif enhances usage of the EIIIB 5' splice site; furthermore, this repeated TGCATG sequence can activate an alternatively spliced exon in the unrelated rat preprotachykinin pre-mRNA. Interestingly, this sequence can also be found within *cis*-acting elements previously identified in other alternatively spliced genes. This short repeated TGCATG motif is therefore a cell-type-specific regulatory *cis*-element that, in addition to controlling fibronectin alternative splicing, may participate in the regulation of other alternative RNA processing events.

INTRODUCTION

Alternative pre-mRNA splicing represents a fundamental mode of post-transcriptional metazoan gene regulation. This phenomenon generates distinct mRNAs from a single RNA transcript, often generating multiple proteins from a single gene in the process, and can be regulated in a developmental or cell-type-specific fashion (for reviews see Smith et al., 1989a; McKeown, 1992).

Splicing is a multistep process involving numerous proteins and the small nuclear ribonucleoprotein particles (snRNPs) U1, U2, U4, U5 and U6; much has been learned about this reaction from biochemical studies utilizing simple model pre-mRNAs. However, considerably less is known about the faithful recognition or regulation of splice sites *in vivo*. Naturally occurring splice sites often vary from known consensus sequences; in addition, long pre-mRNA transcripts can harbor numerous splice-site-like sequences, only a fraction of which are authentic signals. Therefore additional mechanisms must exist that specify introns and exons; these mechanisms must also exhibit flexibility in order for regulated splicing to occur (for reviews see Green, 1991; Moore et al., 1993).

In addition to resemblance to a consensus, splice site usage can be affected by many diverse parameters. For example, internal exon identification in vertebrates may be aided by coordinate recognition of the exon's splice sites, a process termed exon definition (Robberson et al., 1990; Talerico and Berget, 1990; Hoffman and Grabowski, 1992). This process may be compromised if the exon is of an inappropriate length (Robberson et al., 1990; Black, 1991; Dominski and Kole, 1991). Splice site usage can also be affected by adjacent intron sequences (Black, 1991; Gallego et al., 1992; Helfman et al, 1990), exon sequences (Somasekhar and Mertz, 1985; Reed and Maniatis, 1986; Mardon et

al., 1987; Cooper and Ordahl, 1989; Streuli and Saito, 1989) or RNA secondary structure (Solnick, 1985; Eperon et al., 1988; Libri et al., 1991; D'Orval et al., 1991). Many parameters can therefore affect splicing in ways that are complex and not readily predictable.

In *Drosophila melanogaster*, sophisticated genetic analyses have identified genes that can encode negative or positive splicing regulators (for reviews see Baker, 1989; McKeown, 1992). However, the identification of vertebrate splicing regulators has been more difficult. Since multiple pre-mRNA sequences affect splicing, it can often be difficult to distinguish between constitutive and regulatory elements by mutational analysis. Furthermore, the genetic methods that have identified *Drosophila* regulators cannot yet be easily applied to the identification of regulatory splicing factors in vertebrates. Among the proteins implicated in the biochemistry of splicing, a number can affect splice site selection *in vitro* (Ge and Manley, 1990; Krainer et al., 1990; Mayeda et al., 1993; Fu et al., 1992; Zahler et al., 1993), suggesting that differential concentrations of general splicing factors might regulate some alternative splicing events *in vivo* (Maniatis, 1991).

In order to address some of these questions, we have been studying the regulation of alternative splicing in the rat fibronectin (FN) gene. The EIIIB exon in this gene (also called ED2, EDB, or EDIIIB) is differentially spliced into FN mRNA to varying degrees in different cell types and developmental stages; for example, EIIIB is absent from adult liver FN mRNA but is present in nearly all early embryonic FN messages (for review see Hynes, 1990). Previous transfection studies have established that EIIIB is an inefficiently recognized exon and that its cell-type-dependent inclusion into mRNA requires (i) a balance between competing splice sites in the EIIIB region and (ii) sequences located in the intron downstream of EIIIB (Huh and Hynes, 1993).

The present study extends our previous work on EIIIB-activating intron elements. Surprisingly, we have found that EIIIB inclusion can be activated by repeated copies of a hexanucleotide sequence (TGCATG) that normally occurs several times in the EIIIB-activating intron element. These TGCATG repeats, which are capable of activating EIIIB 5' splice site usage, constitute a major requirement for cell-type-specific EIIIB recognition and represent a novel example of mammalian splicing regulation. TGCATG sequences can also direct regulated splicing in the rat preprotachykinin gene transcript and furthermore can be found within elements identified in other alternatively spliced genes. Since the regulatory element we have identified is a simple hexanucleotide sequence, these findings may have implications that bear on the regulation of alternative splicing in general.

RESULTS

A repeated sequence coincides with elements required for EIIIB usage

To study the regulation of EIIIB splicing, we used a three-exon minigene that contained the EIIIB region of the rat fibronectin (FN) gene. Previous studies showed that this minigene, called 7iBi89, reproduced cell-type-specific EIIIB regulation when expressed by a number of cell lines. In addition, these studies established that EIIIB inclusion requires *cis*-acting sequences that lie at least 445 bases away from EIIIB, spanning more than 500 bases in the 3' region of the intron immediately downstream of EIIIB (called IVS2 in 7iBi89; Figure 3-1). Deletions in this region caused near-complete EIIIB skipping in cell lines that normally splice EIIIB into mRNA; deletion of a particular section of this region (termed the intronic control region or ICR) attenuated the cell-type-dependency of exon inclusion, indicating that cell-type-specific elements existed within IVS2 (Chapter Two).

Inspection of the 7iBi89 sequence revealed nine GCATG repeats within the 1071-base IVS2 (Figure 3-1). Eight of these have the sequence TGCATG; five have the sequence TGCATGA. This clustering of short sequences was striking, since one would expect to find a given pentanucleotide only once every 1024 nucleotides by random chance. In addition, the distribution of these repeats correlates well with the EIIIB-activating region and the ICR. A comparison between rat and human FN genes revealed significant homology in the EIIIB-activating region of IVS2 relative to other regions in this intron (Figure 3-1, lower). In particular, seven of nine GCATG repeats are conserved. In previous studies, IVS2 deletions that reduced EIIIB inclusion removed at least one GCATG repeat; removal of more repeats had greater effects. Given (i) the remarkably high incidence of this short repeat in IVS2 and (ii) the striking

correlation between these repeats and conserved EIIIB-activating *cis*-sequences, we hypothesized that GCATG or TGCATG repeats promote EIIIB inclusion.

TGCATG repeats activate EIIIB exon recognition

To test whether TGCATG sequences could activate EIIIB inclusion, we synthesized a DNA oligonucleotide that contained two TGCATG hexamers (oligonucleotide T in Figure 3-2A). Because the repeats in IVS2 were often followed by adenosine or cytidine nucleotides, oligonucleotide T contained one each of the sequences TGCATGA and TGCATGC separated by pyrimidine-rich spacers. Up to three copies of oligonucleotide T were inserted in tandem into 7iBi89ΔGA, a version of 7iBi89 from which 495 bases of IVS2 were removed; deletion ΔGA, which abolishes EIIIB inclusion (Chapter Two; Figure 3-2B), removed four endogenous GCATG repeats. As controls, we tested oligonucleotides that contained either mutant hexamers (TGACTG; oligonucleotide M) or scrambled hexamers (AGTCGT; oligonucleotide S). Inserts were designated by name and copy number (e.g., insert T3 has three copies of oligonucleotide T).

Minigenes were transfected stably into F9 teratocarcinoma cells, which normally exhibit high levels of EIIIB inclusion. S1 nuclease analysis of transfectant RNA was carried out using a minigene-specific cDNA probe that contains the third 7iBi89 exon and overlaps EIIIB. The 7iBi89 minigene exhibited significant EIIIB inclusion, as assessed by the ratio of $EIIIB^+$ to $EIIIB^-$ mRNA (Figure 3-3, lane 1). EIIIB inclusion was reduced dramatically by the ΔGA deletion (lane 2). However, the insertion of one to three copies of oligonucleotide T into 7iBi89ΔGA (two to six TGCATG repeats) restored inclusion in an additive manner (lanes 3 to 5). With six TGCATG repeats, EIIIB inclusion was indistinguishable from that exhibited by the undeleted minigene

(compare lanes 5 and 1). In contrast, insertion of repeated mutant or scrambled hexamers (lanes 6 through 11) did not restore EIIIB inclusion. Therefore, in F9 cells, repeated TGCATG sequences can substitute for certain EIIIB-activating sequences in IVS2.

We then tested whether TGCATG repeats could also substitute for other repeat-containing sequences in IVS2. The IVS2 deletion ΔAR2a removes sequences not deleted by ΔGA (Figure 3-2B). ΔAR2a removes three TGCATG repeats and also reduced EIIIB inclusion (Figure 3-3, lane 12). Insertion of four TGCATG repeats in ΔAR2a (insert T2; lane 13) enhanced EIIIB inclusion to levels slightly higher than those of 7iBi89, an effect not observed with mutant hexamers (insert M2; lane 14). We conclude that repeated TGCATG sequences can substitute for either of two distinct EIIIB-activating, GCATG-containing regions in IVS2.

We next examined whether TGCATG repeats also had effects in EIIIB-skipping cell lines by transfecting these constructs into HeLa cells (which normally include EIIIB at low levels) and into COS cells (which completely skip EIIIB in 7iBi89). In HeLa cells, TGCATG repeats in 7iBi89ΔGA detectably stimulated EIIIB inclusion (Figure 3-3, lanes 15-28). Six synthetic TGCATG repeats (T3) had weaker effects in HeLa than in F9 cells, although HeLa EIIIB inclusion was greater in 7iBi89ΔGA+T3 than in 7iBi89 (compare lanes 15 and 19). Otherwise, the effects of test and control repeats in HeLa cells paralleled those in F9 cells. In contrast, complete EIIIB-skipping in COS was unaffected by deletions or repeat insertions in IVS2 (Figure 3-3, lanes 29-36). Thus, TGCATG repeats activated the splicing of EIIIB only in cell types that normally exhibited EIIIB inclusion.

Activity of TGCATG repeats is affected by flanking bases

We next tested whether TGCATG activity was affected by bases immediately flanking the repeat. The original test oligonucleotide (T) contained one each of the sequences TGCATGA and TGCATGC; since five repeats in the EIIIB-III-8a intron were TGCATGA sequences, we tested an oligonucleotide that contained two TGCATGA heptamers (TA; Figure 3-2A). Another oligonucleotide was tested that contained two TGCATGC repeats (TC). We also tested flanking sequence effects by substituting the relatively pyrimidine-rich spacers of oligonucleotides T and M with purine-rich sequences (oligonucleotides TP and MP). Each oligonucleotide was inserted in three copies into 7iBi89ΔGA (inserts TA3, TC3, TP3, and MP3) and tested by stable transfection in F9 and HeLa cells and by transient transfection in COS cells.

Each of the inserts TA3, TC3 and TP3 significantly enhanced EIIIB inclusion in F9 cells (Figure 3-4). Surprisingly, neither TA3 nor TC3 worked as well as T3 in this respect (Figure 3-4, lanes 3, 6, 7). Therefore six alternating TGCATGA and TGCATGC repeats stimulated EIIIB inclusion to a greater extent than did six repeats of either sequence alone. Interestingly, insert TP3 (with purine spacers) enhanced EIIIB inclusion much more effectively than did T3 (compare lanes 3 and 8). Mutant repeats with purine spacers (MP3) had no such effect (lane 9).

In HeLa cells, inserts TA3 and TC3 had little effect on EIIIB splicing (Figure 3-4, lanes 15, 16). As in F9 cells, insert TP3 stimulated HeLa EIIIB inclusion more effectively than did insert T3 (lane 17). For any given minigene, HeLa inclusion was always lower than F9 inclusion. None of these inserts affected the complete EIIIB-skipping pattern exhibited by transiently transfected COS cells (lanes 21-24).

Figure 3-1.

The EIIIB region in the rat fibronectin gene

Organization of the EIIIB region, the 7iBi89 minigene and *cis*-elements required for EIIIB inclusion (labelled wide rectangles, exons; lines and narrow bars, introns). Part of the rat fibronectin (FN) gene is shown at top; dashed lines above and beneath outline the EIIIB alternative splice patterns, along with the cell types in which each of these patterns predominate. Scale bar, 200 nucleotides (nt). Beneath this is the minigene 7iBi89; the introns preceding and following EIIIB are called IVS1 and IVS2 respectively. Minigene expression is directed by a modified human β-actin promoter and human growth hormone poly(A) signal. The narrow shaded bar within IVS2 shows the location of sequences required for EIIIB inclusion; the 3' endpoint of this region is undefined (Huh and Hynes, 1993). The black bar within this shaded bar represents a segment (the ICR) that is particularly important for EIIIB inclusion (see text). GCATG repeats are shown as vertical lines beneath 7iBi89. The nucleotide sequence below 7iBi89 is of IVS2 (1071 nt), flanked by ten bases each of exons EIIIB and III-8a. The rat sequence is shown (uppercase letters, bases conserved between rat and human FN genes; lowercase letters, bases absent or not conserved in human; dashes, gaps generated in comparing the two sequences). Nucleotide +1 is the first base in IVS2. Sequences important for EIIIB inclusion are enclosed within dashed lines; the ICR is shaded. Heptamers that contain GCATG repeats are enclosed in boxes. Letters in bold type indicate stretches of rat-human homology (13 or more identical bases out of 15). Certain regions (bases +1 to +50 and +500 to +1071) contain several relatively uninterrupted stretches of rat-human sequence identity.

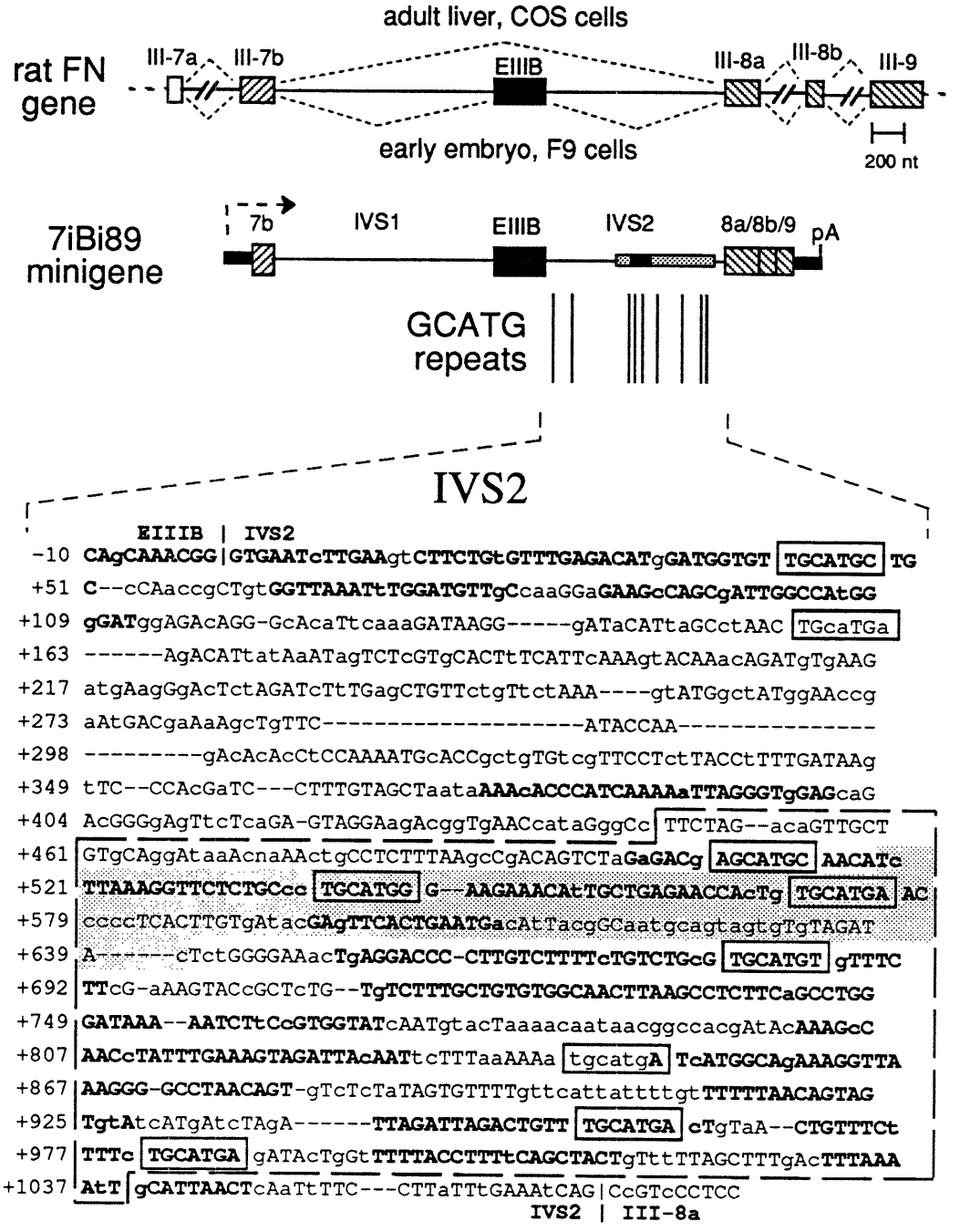

adult liver, COS cells
rat FN gene
III-7a
III-7b
EIIIB
III-8a
III-8b
III-9
early embryo, F9 cells
200 nt
7iBi89 minigene
7b
IVS1
EIIIB
IVS2
8a/8b/9
pA
GCATG repeats
IVS2
EIIIB | IVS2
-10 CAgCAAACGG|GTGAATcTTGAAgtCTTCTGtGTTTGAGACATgGATGGTGT TGCATGC TG
+51 C--cCAaccgCTgtGGTTAAATtTGGATGTTgCcaaGGaGAAGcCAGCgATTGGCCAtGG
+109 gGATggAGAcAGG-GcAcaTtcaaaGATAAGG-----gATaCATtaGCctAAC TGcaTGa
+163 ------AgACATtatAaATagTCTcGTgCACTtTCATTcAAAgtACAAacAGATgTgAAG
+217 atgAagGgAcTctAGATcTtTGagCTGTTctgTtctAAA----gtATGgctATggAAccg
+273 aAtGACgaAaAgcTgTTC--------------------ATACCAA---------------
+298 ---------gAcAcAcCtCCAAAATGcACCgctgTGtcgTTCCTctTACCtTTTGATAAg
+349 tTC--CCAcGaTC---CTTTGTAGCTaataAAAcACCCATCAAAAaTTAGGGTgGAGcaG
+404 AcGGGgAgTtcTcaGA-GTAGGAagAcggTgAACcataGggCc TTCTAG--acaGTTGCT
+461 GTgCAggAtaaAcnaAActgCCTCTTTAAgcCgACAGTCTaGaGACg AGCATGC AACATc
+521 TTAAAGGTTCTCTGCcc TGCATGG G--AAGAAACAtTGCTGAGAACCAcTg TGCATGA AC
+579 ccccTCACTTGTgAtacGAgTTCACTGAATGacAtTacgGCaatgcagtagtgTgTAGAT
+639 A------cTctGGGGAAacTgAGGACCC-CTTGTCTTTTcTGTCTGcG TGCATGT gTTTC
+692 TTcG-aAAGTACcGCTcTG--TgTCTTTGCTGTGTGGCAACTTAAGCCTCTTCaGCCTGG
+749 GATAAA--AATCTtCcGTGGTATcAATgtacTaaaacaataacggccacgAtAcAAAGcC
+807 AACcTATTTGAAAGTAGATTAcAATtcTTTaaAAAa tgcatgA TcATGGCAgAAAGGTTA
+867 AAGGG-GCCTAACAGT-gTcTcTaTAGTGTTTTgttcattattttgtTTTTTAACAGTAG
+925 TgtAtcATgAtcTAgA------TTAGATTAGACTGTT TGCATGA cTgTaA--CTGTTTCt
+977 TTTc TGCATGA gATAcTgGtTTTTACCTTTTcAGCTACTgTttTTAGCTTTgAcTTTAAA
+1037 AtT gCATTAACTcAaTtTTC---CTTaTTtGAAAtCAG|CcGTcCCTCC
IVS2 | III-8a

Figure 3-2.

TGCATG repeats can substitute for the EIIIB-activating region

A. At left are the names of repeat-containing oligonucleotides used in this study. Monomeric forms of the inserts are shown at right; differences between each sequence and that of oligonucleotide T (top line) are shaded.

B. The 7iBi89 minigene and its deletion derivatives with single or multiple oligonucleotide inserts. Inserts were placed at the ΔGA or ΔAR2a deletion junctions. GCATG repeats are shown as vertical lines in the 7iBi89, ΔGA and ΔAR2a minigene diagrams. A summary of EIIIB inclusion in each transfected cell line (from Figures 3-3 and 3-4; cell line indicated at top) is at the right (–, $EIIIB^+/EIIIB^-$ ratio < 0.10; +, 0.10 < ratio < 0.25; ++, 0.25 < ratio < 0.40; +++, 0.40 < ratio < 0.65; ++++, 0.65 < ratio < 0.80; +++++, ratio > 0.80; n.t., not tested).

A

Oligo	Sequence
T	GATCC TGCATGA CTACTGCT TGCATGC TCGTCATA
M	GATCC TGACTGA CTACTGCT TGACTGC TCGTCATA
S	GATCC AGTCGTA CTACTGCT AGTCGTC TCGTCATA
TA	GATCC TGCATGA CTACTGCT TGCATGA TCGTCATA
TC	GATCC TGCATGC CTACTGCT TGCATGC TCGTCATA
TP	GATCA TGCATGA GGGAAAGG TGCATGC AAAGGGAA
MP	GATCA TGACTGA GGGAAAGG TGACTGC AAAGGGAA

B

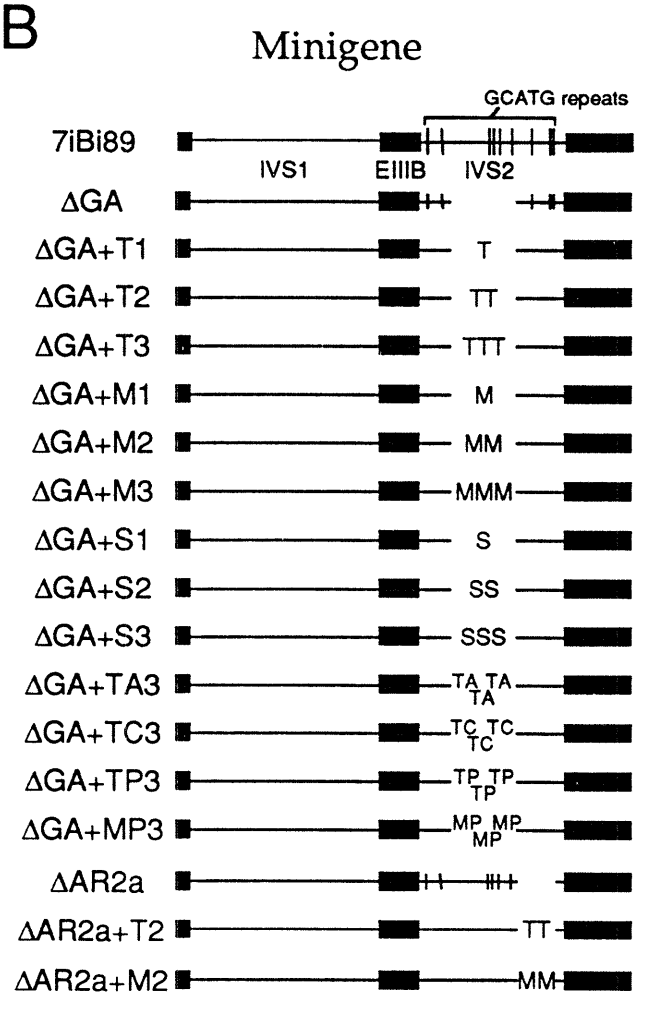


Figure 3-3.

S1 analyses of transfected 7iBi89 minigenes

The EIIIB$^+$ 7iBi89 mRNA is shown at top left (conventions as in Figure 3-1), along with the cDNA-derived 5'-end-labelled S1 probe (called B-89); the electrophoretic positions of undigested probe ("S1 probe") and of fragments generated from protection by EIIIB$^+$ or EIIIB$^-$ mRNAs are indicated at the left of each panel. The ratio of these fragments (EIIIB$^+$/EIIIB$^-$) indicates the level of EIIIB inclusion. Cell lines are indicated at the top of each panel. Cells were transfected with each of the indicated constructs (7iBi89ΔGA and 7iBi89ΔAR2a were abbreviated to ΔGA and ΔAR2a, respectively; Ctrl; untransfected cells). These panels show the effects of inserts derived from oligonucleotides T, M and S (see Figure 3-2A for sequences). For analysis, cytoplasmic RNA (from stably transfected F9 and HeLa cells) or total RNA (from transiently transfected COS cells) was used. Lane M, end-labeled molecular weight markers (pBR322/MspI fragments).

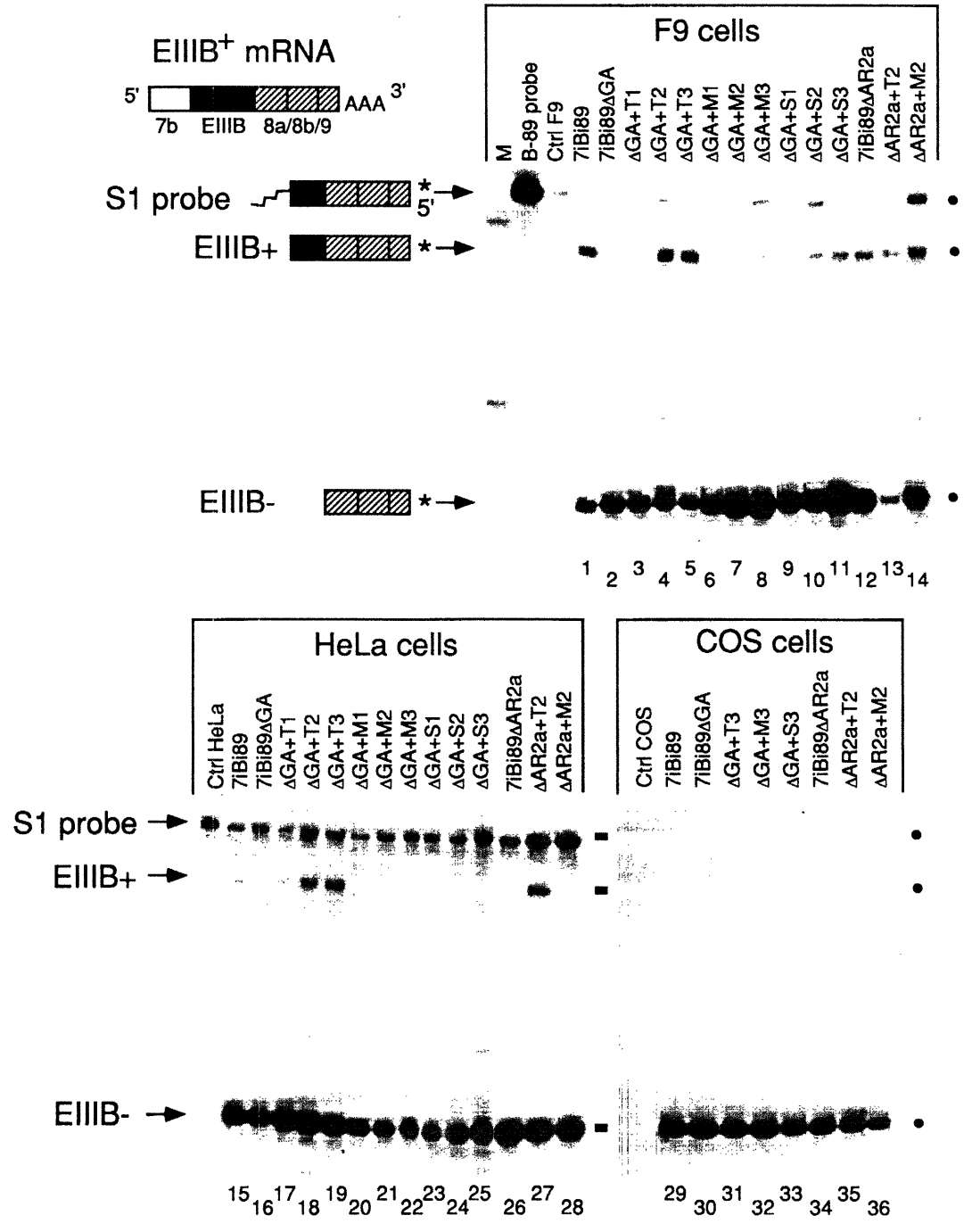

EIIIB+ mRNA
5'
AAA 3'
7b
EIIIB
8a/8b/9
F9 cells
M
B-89 probe
Ctrl F9
7iBi89
7iBi89ΔGA
ΔGA+T1
ΔGA+T2
ΔGA+T3
ΔGA+M1
ΔGA+M2
ΔGA+M3
ΔGA+S1
ΔGA+S2
ΔGA+S3
7iBi89ΔAR2a
ΔAR2a+T2
ΔAR2a+M2
S1 probe
5'
EIIIB+
EIIIB-
1 2 3 4 5 6 7 8 9 10 11 12 13 14
HeLa cells
Ctrl HeLa
7iBi89
7iBi89ΔGA
ΔGA+T1
ΔGA+T2
ΔGA+T3
ΔGA+M1
ΔGA+M2
ΔGA+M3
ΔGA+S1
ΔGA+S2
ΔGA+S3
7iBi89ΔAR2a
ΔAR2a+T2
ΔAR2a+M2
COS cells
Ctrl COS
7iBi89
7iBi89ΔGA
ΔGA+T3
ΔGA+M3
ΔGA+S3
7iBi89ΔAR2a
ΔAR2a+T2
ΔAR2a+M2
S1 probe
EIIIB+
EIIIB-
15 16 17 18 19 20 21 22 23 24 25 26 27 28
29 30 31 32 33 34 35 36

Figure 3-4.

S1 analyses of 7iBi89 minigenes containing variant TGCATG inserts

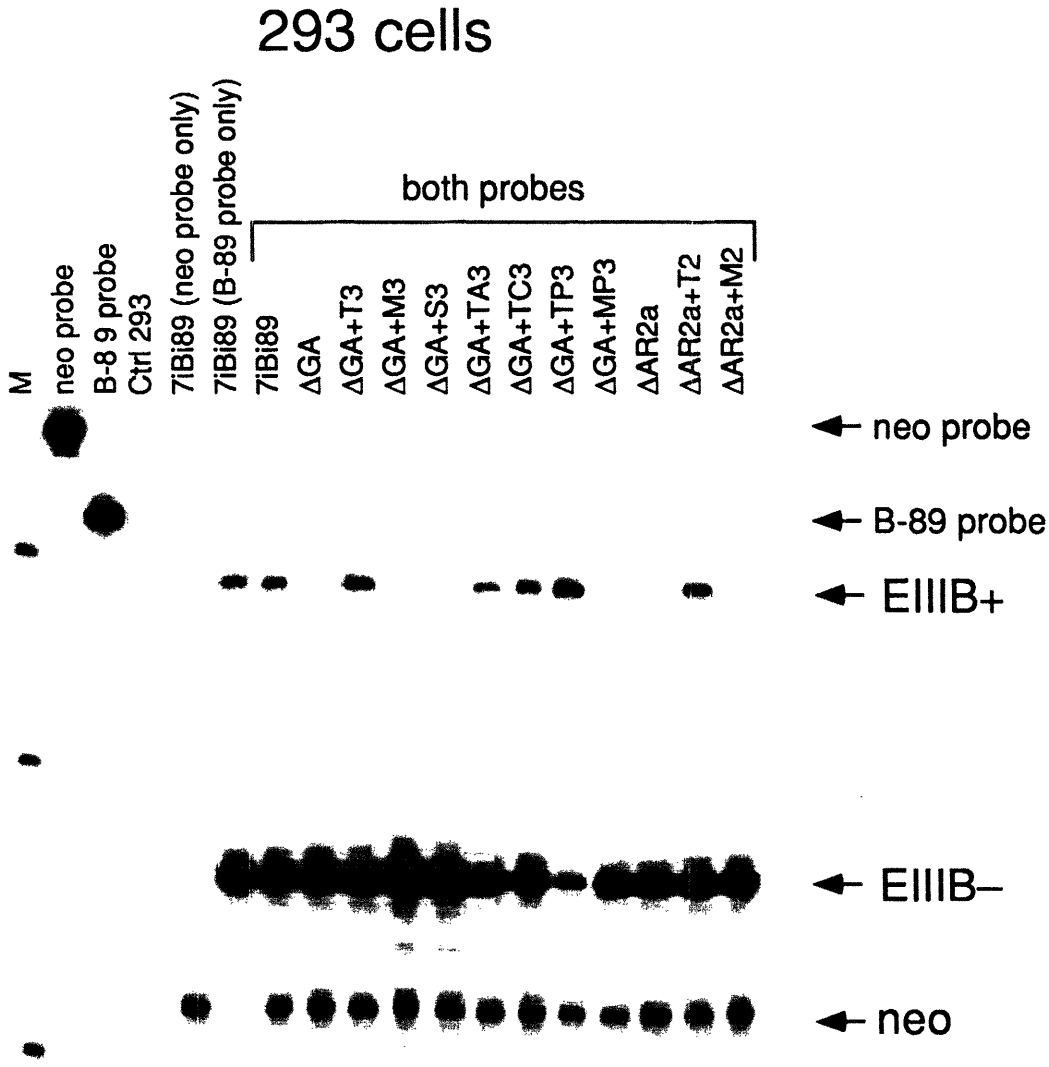
S1 analyses of transfections using minigenes containing inserts TA3, TC3, TP3 and MP3 (see Figure 3-2A for sequences). Procedures and figure conventions were as in Figure 3-3.

probe
EIIIB+
EIIIB-
F9 cells
Ctrl F9
7IB/89
7IB/89ΔGA
ΔGA+T3
ΔGA+M3
ΔGA+S3
ΔGA+TA3
ΔGA+TC3
ΔGA+TP3
ΔGA+MP3
HeLa cells
Ctrl HeLa
7IB/89
7IB/89ΔGA
ΔGA+T3
ΔGA+M3
ΔGA+S3
ΔGA+TA3
ΔGA+TC3
ΔGA+TP3
ΔGA+MP3
COS cells
Ctrl COS
7IB/89
7IB/89ΔGA
ΔGA+TA3
ΔGA+TC3
ΔGA+TP3
ΔGA+MP3
1 2 3 4 5 6 7 8 9
10 11 12 13 14 15 16 17 18
19 20 21 22 23 24

Figure 3-5.

Transient expression of TGCATG-containing 7iBi89 minigenes in 293 cells: comparison of expression levels

The indicated constructs (Figure 3-2B) were transiently transfected into 293 cells; transfections were carried out using 18 μg 7iBi89/pBAGH derivatives plus 2 μg pSV2neo (Chapter One). S1 probes were B-89 (Figure 3-3) and/or a segment of pSV2neo sequence (See Materials and Methods).


293 cells
M
neo probe
B-8 9 probe
Ctrl 293
7iBi89 (neo probe only)
7iBi89 (B-89 probe only)
both probes
7iBi89
ΔGA
ΔGA+T3
ΔGA+M3
ΔGA+S3
ΔGA+TA3
ΔGA+TC3
ΔGA+TP3
ΔGA+MP3
ΔAR2a
ΔAR2a+T2
ΔAR2a+M2
neo probe
B-89 probe
EIIIB+
EIIIB–
neo

It was possible that the differences in inclusion that we observed were due to changes in the level of expression associated with these inserts. To address this, we analyzed expression in transiently transfected 293 cells (which also normally include EIIIB), using a cotransfected pSV2neo plasmid as an internal control. We observed no gross differences in overall expression among a representative set of 7iBi89 minigenes (Figure 3-5). Thus it is unlikely that these effects on EIIIB are a consequence of differential expression levels.

We conclude that, although sequence context is not absolutely critical for hexamer function, repeat-dependent EIIIB inclusion is influenced to some extent by sequences flanking the TGCATG motif.

TGCATG repeats can activate cell-type-specific exon usage

The above results suggested that the TGCATG sequence represents a cell-type-specific element. However, it was possible that these repeats enhanced EIIIB inclusion nonspecifically, while other elements provided cell-type-specificity. In past studies, the role of IVS2 sequences in cell-type-specific EIIIB regulation was explored by using a chimeric FN-related minigene called M2. The M2 construct contains 7iBi89 sequences that spanned IVS2 (black rectangles in Figure 3-6). The upstream portion of M2 (white and striped rectangles) is derived from a constitutively spliced region of the rat FN gene (see Figure 3-6 legend). Regulated inclusion of the hybrid III-9/EIIIB central exon of M2 was detectable in all cell lines and also correlated with EIIIB regulation; exon inclusion was low in COS cells, intermediate in HeLa cells and high in F9 cells. Certain IVS2 deletions in M2 had previously been found to attentuate cell-type-specific regulation by differentially reducing inclusion in HeLa and F9 relative to COS cells. This effect was only observed when at least three GCATG repeats were removed, particularly those within the empirically

determined ICR (intronic control region; Figure 3-1). These data established that cell-type-specific elements within IVS2 contained GCATG sequences (Chapter Two). In order to address directly whether TGCATG repeats acted in a cell-type-specific fashion, synthetic TGCATG-containing oligonucleotides were tested for their ability to substitute for ICR function in the M2 minigene.

The deletion of four GCATG repeats from M2 (using the IVS2 deletion ΔGA, which also removes the ICR; minigene M2ΔGA in Figure 3-6) significantly reduced inclusion in F9 and HeLa cells while leaving COS inclusion relatively unaffected, reducing the differential between F9 and COS inclusion from 11-fold to 3-fold (Figure 3-6). The insertion of two synthetic TGCATG hexamers (M2ΔGA+T1) modestly increased the F9/COS differential (from 3-fold to 5-fold). Six repeats in the form of insert T3 (M2ΔGA+T3) significantly enhanced M2 exon inclusion in all three cell lines, even in COS cells. Because COS inclusion was also enhanced, insert T3 (in contrast to T1) did not significantly increase the F9/COS differential. Insert T3 also enhanced EIIIB inclusion HeLa to levels similar to F9. Therefore a moderate number of TGCATG repeats can partially substitute for ICR sequences. Neither mutant nor scrambled repeats enhanced M2 exon usage (M2ΔGA+M1, +M3, +S1, and +S3); in fact, inclusion was nonspecifically reduced by these control inserts.

The effects of variant inserts upon M2ΔGA splicing were found to differ from each other, both quantitatively and qualitatively. Insert TC3 (with six TGCATGC repeats) had effects similar to those of insert T3. In contrast, insert TA3 (with six TGCATGA repeats), in addition to enhancing the F9/COS differential (from 3- to 5-fold), exhibited a pattern of cell-type-specificity most closely resembling that of intact M2, with inclusion levels in HeLa being intermediate between those in F9 and COS. Insert TP3 (with purine spacers) enhanced the F9/COS differential to the greatest extent (from 3-fold to 10-fold).

Insert MP3 did not exhibit this effect. We conclude that TGCATG repeats are indeed capable of cell-type-specific action and that cell-type-specificity is influenced by hexamer context.

TGCATG repeats promote EIIIB 5' splice site usage

In general, IVS2 deletions previously shown to reduce exon inclusion removed, at most, four GCATG repeats (Chapter Two). When we deleted seven IVS2 repeats from the M2 minigene, an unusual splicing phenotype resulted. This phenotype was detected by using an RNase protection probe that spanned the central exon of M2 (Figure 3-7). Differently sized probe products were obtained depending upon which central exon splice sites (3' or 5') were used. Analysis of M2 expression in COS or HeLa cells yielded primarily a product corresponding to usage of both splice sites of the M2 central exon (Figure 3-8, lanes 1 and 10). In contrast, when a large deletion removing seven repeats was tested (M2AL2R2a; Figure 3-7), this fragment was not detected; instead, an IVS2-containing RNA accumulated (Figure 3-8, lanes 5 and 14). This RNA, which lacks IVS1, did not accumulate if three- or four-repeat sections of the AL2R2a region were deleted (M2AR2a, lanes 4 and 13; M2AL2, lanes 3 and 12), nor did it accumulate if a large deletion that only removed four repeats was tested (M2GA, lanes 2 and 11). A series of control RNase protection assays using probes that spanned the other splice junctions in the M2 minigene confirmed that the splice sites of both upstream and downstream flanking exons were utilized in all of these constructs in COS cells, although usage of the downstream exon 3' splice site was only partial for M2AL2R2a (Figure 3-9). Taken together, these results suggested that when seven repeats were deleted, IVS1 splicing and exon-skipping still occurred, but

Figure 3-6.

TGCATG repeats regulate cell-type-specific exon recognition

A. The M2 and M2ΔGA minigenes. The portion of M2 derived from 7iBi89 is indicated by black rectangles (exons) and black bars (introns). The upstream portion of M2 (striped and unshaded rectangles and bars, exons and introns respectively) is derived from a constitutively spliced region of the rat FN gene and contains exon III-8b, the III-8b/III-9 intron and part of exon III-9. GCATG repeats are indicated as vertical lines within IVS2. Oligonucleotides were inserted at the ΔGA deletion junction.

B. Histograms indicating the amount of central exon inclusion exhibited by each minigene ($exon^+/exon^-$ ratios); error bars represent standard deviations from transient transfections that were carried out at least twice for COS and HeLa cells and at least three times for F9 cells. Unshaded bars, COS cells; shaded bars, HeLa cells; black bars, F9 cells. The F9/COS differential for each construct (calculated as a ratio between F9 and COS values) is shown below each triplet histogram set.

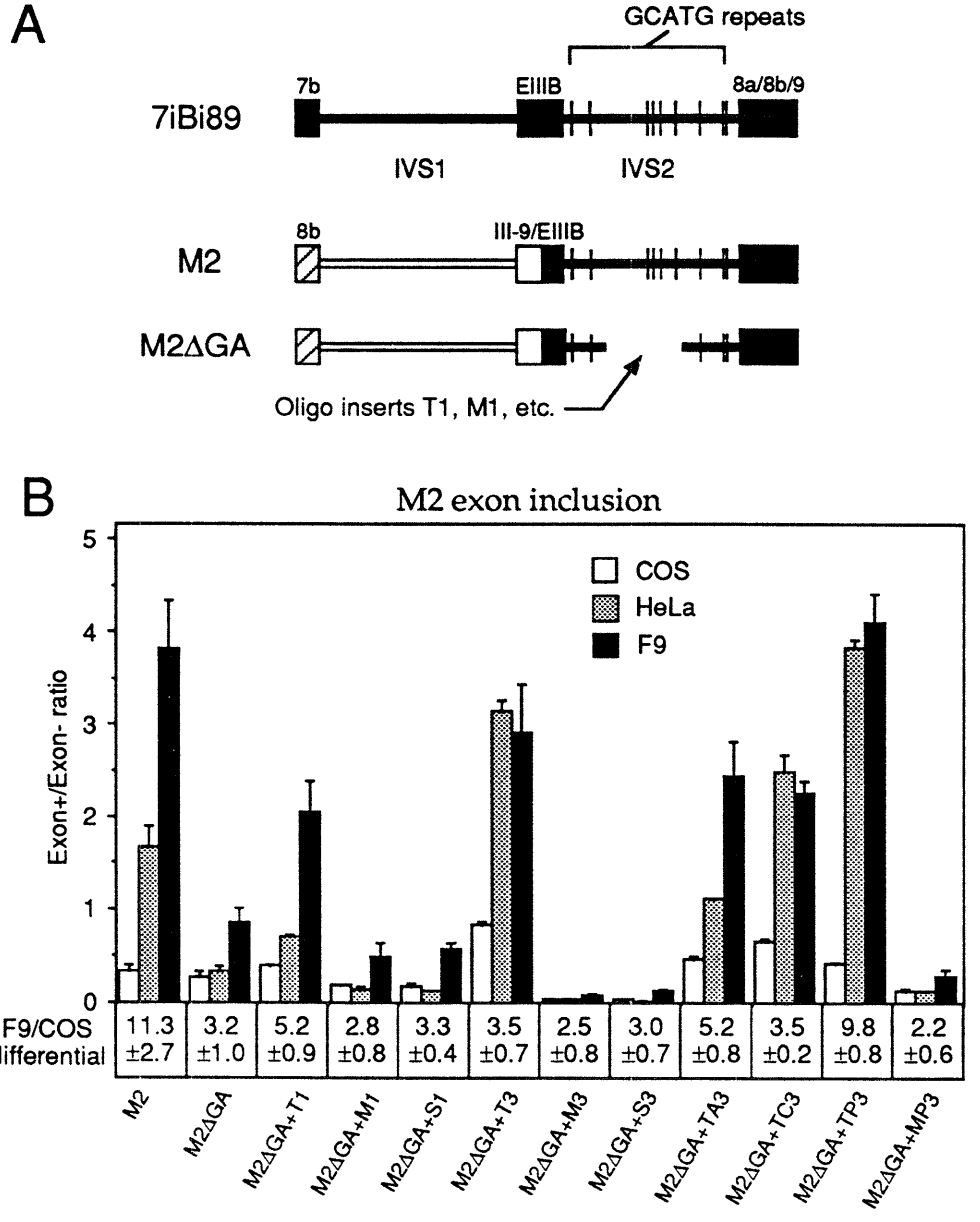

A
GCATG repeats
7b
EIIIB
8a/8b/9
7iBi89
IVS1
IVS2
8b
III-9/EIIIB
M2
M2ΔGA
Oligo inserts T1, M1, etc.
B
M2 exon inclusion
COS
HeLa
F9
Exon+/Exon- ratio
0
1
2
3
4
5
F9/COS differential
11.3 ±2.7
3.2 ±1.0
5.2 ±0.9
2.8 ±0.8
3.3 ±0.4
3.5 ±0.7
2.5 ±0.8
3.0 ±0.7
5.2 ±0.8
3.5 ±0.2
9.8 ±0.8
2.2 ±0.6
M2
M2ΔGA
M2ΔGA+T1
M2ΔGA+M1
M2ΔGA+S1
M2ΔGA+T3
M2ΔGA+M3
M2ΔGA+S3
M2ΔGA+TA3
M2ΔGA+TC3
M2ΔGA+TP3
M2ΔGA+MP3

Figure 3-7.

Splicing patterns in M2 minigenes with IVS2 deletions

Summary of splice site usage in the M2 minigene and its deletion derivatives. Conventions are as in Figure 3-6; dashed lines above and below each construct represent the splicing that occurs in that minigene (construct M2GA is the same as M2ΔGA in Figure 3-6A). The number of GCATG repeats that were removed from each construct is indicated beside each minigene. At right is indicated, for each minigene, whether usage of the 5' splice site of the M2 central exon (the EIIIB 5' splice site) can be detected. The RNase protection probe that spans the M2 central exon is shown below M2AL2R2a. Transfection of the M2AL2R2a construct results in the accumulation of a partially spliced IVS2-containing RNA, which is diagrammed at bottom. Synthetic inserts were inserted at a BglII site in M2AL2R2a (indicated by an asterisk).

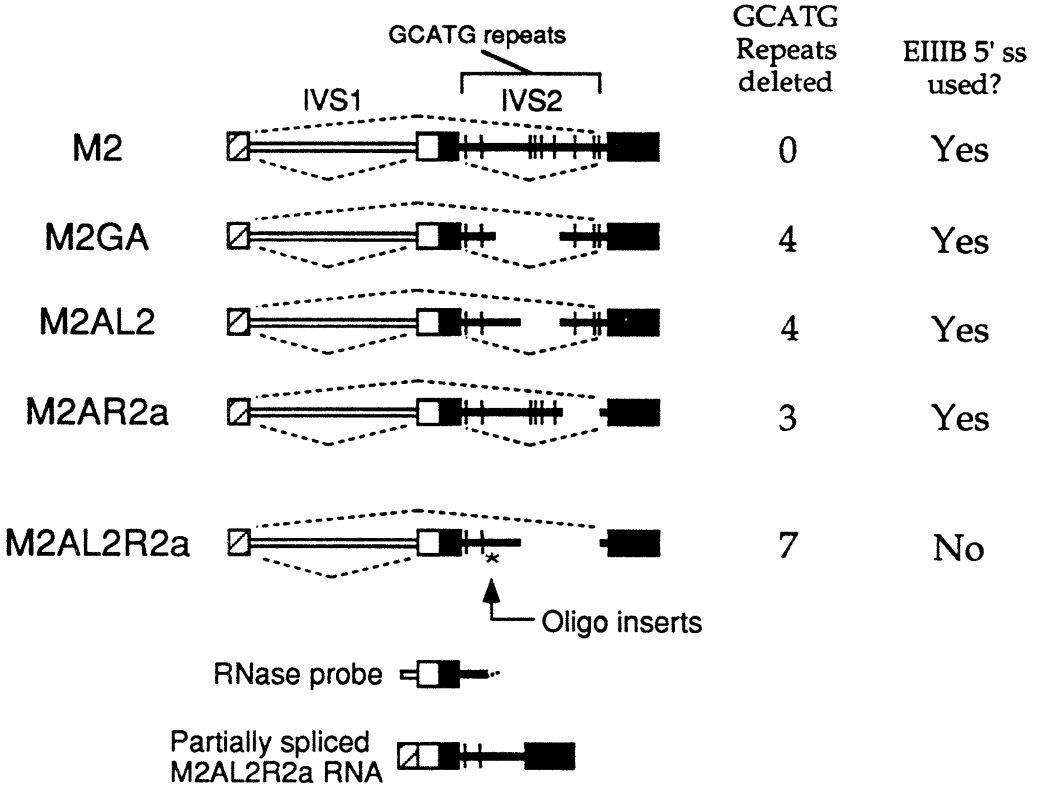

GCATG repeats
IVS1
IVS2
GCATG Repeats deleted
EIIIB 5' ss used?
M2
0
Yes
M2GA
4
Yes
M2AL2
4
Yes
M2AR2a
3
Yes
M2AL2R2a
7
No
Oligo inserts
RNase probe
Partially spliced M2AL2R2a RNA

Figure 3-8.

TGCATG repeats promote EIIIB 5' splice site usage

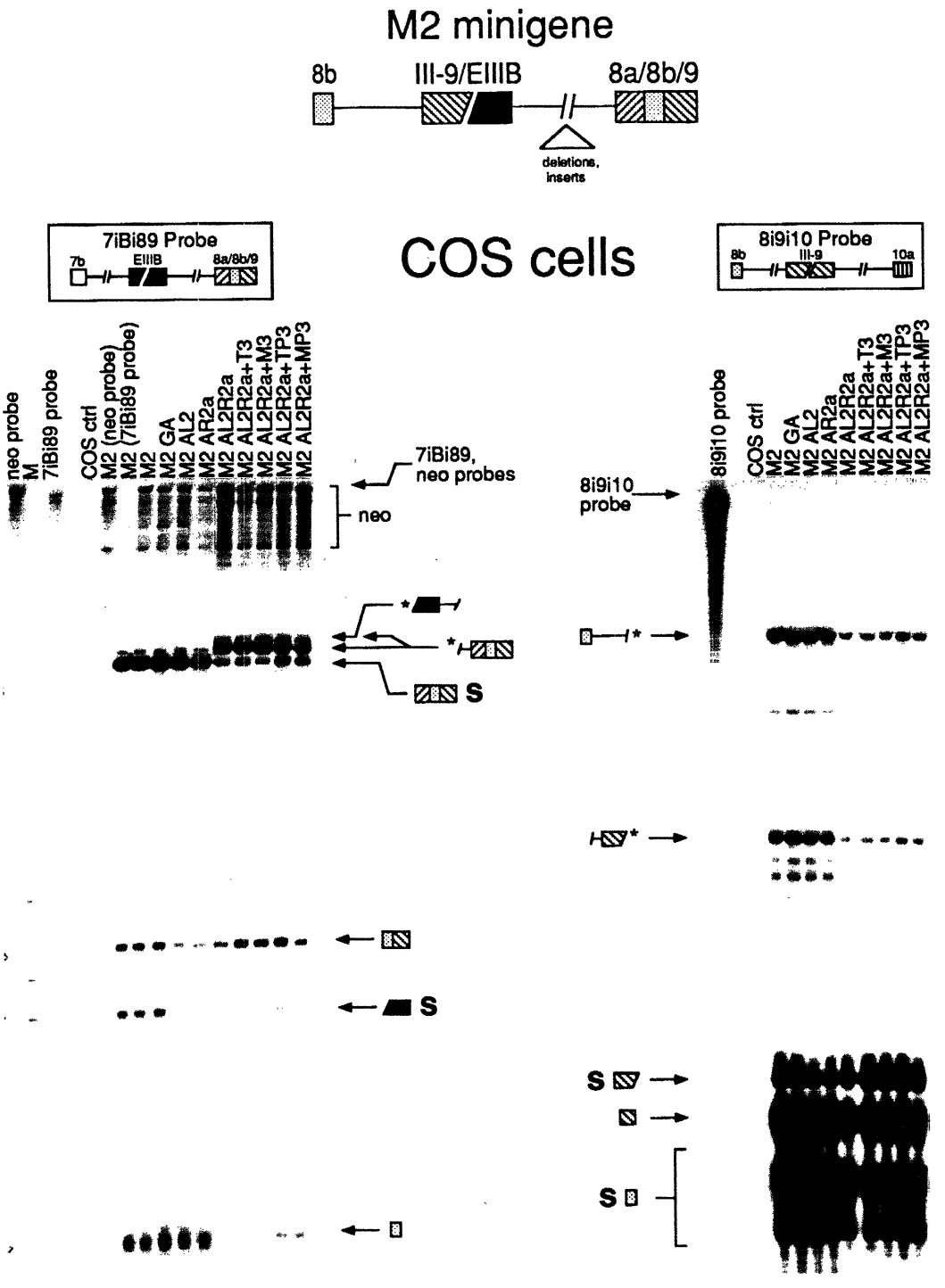
RNase protection assays of M2 exon usage. Total RNAs from transiently transfected COS or HeLa cells were analysed; the constructs transfected are indicated above each lane. Expected RNase protection products are shown at left; bands corresponding to partially spliced (IVS2-containing RNA; small arrowhead) and completely spliced central exon (large arrowhead) are also indicated at right. A similar assessment of M2 splicing could not be carried out in F9 cells, since endogenous FN interfered with this analysis.

M2 probe
Unspliced
IVS1 spliced
IVS2 spliced
IVS1+IVS2 spliced
M
probe
COS
COS M2
" M2 GA
" M2 AL2
" M2 AR2a
" M2 AL2R2a
" M2 AL2R2a+T3
" M2 AL2R2a+M3
" M2 AL2R2a+TP3
" M2 AL2R2a+MP3
HeLa
HeLa M2
" M2 GA
" M2 AL2
" M2 AR2a
" M2 AL2R2a
" M2 AL2R2a+T3
" M2 AL2R2a+M3
" M2 AL2R2a+TP3
" M2 AL2R2a+MP3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Figure 3-9.

Splicing of M2 minigene derivatives in COS cells

The M2 minigene is shown at top (shading of the exons is different from that shown in Figures 3-6 and 3-7). Two RNase protection analyses are also shown. The probes used in the analysis at left were either neo-specific (as an internal control for expression levels; see Materials and Methods) or a riboprobe corresponding to a version of 7iBi89 containing internally deleted introns. This probe was used to assess splicing of the downstream portion of the M2 minigenes. Protection by cotransfected pSV2neo yielded the bands near the top of the gel; protection by properly spliced M2 RNA yields bands corresponding to individual exons or exon fragments (bands marked by an "S"). Protection by unspliced M2 RNA yields the bands marked by asterisks. Due to the internally repeated nature of the M2 minigene, a number of other bands appear in the analysis; these correspond to protection by upstream portions of the spliced M2 RNA. Except where noted, a mixture of neo and 7iBi89 probes were used. Lane M, pBR322/MspI markers. At right is an analysis using a probe corresponding to an internally deleted 8i9i10 minigene template, which was used in order to assess splicing of the upstream portion of the M2 minigene; similar labeling conventions apply. Total RNA was analyzed from COS cells transiently transfected with each of the M2 constructs listed in Figure 3-7 plus M2AL2R2a constructs containing T3, M3, TP3 or MP3 inserts. Expression of all constructs yielded profiles that correspond to the splicing patterns diagrammed in Figure 3-7. The TGCATG-dependent EIIIB 5' splice site activation is not clearly visible here but is more clearly documented in Figure 3-8.


M2 minigene
8b
III-9/EIIIB
8a/8b/9
deletions, inserts
7iBi89 Probe
7b
EIIIB
8a/8b/9
COS cells
8i9i10 Probe
8b
III-9
10a
neo probe
M
7iBi89 probe
COS ctrl
M2 (neo probe)
M2 (7iBi89 probe)
M2
M2 GA
M2 AL2
M2 AR2a
M2 AL2R2a
M2 AL2R2a+T3
M2 AL2R2a+M3
M2 AL2R2a+TP3
M2 AL2R2a+MP3
7iBi89, neo probes
neo
8i9i10 probe
S

that the EIIIB-derived 5' splice site of the M2 central exon was not used, resulting in production of a partially spliced RNA (summarized in Figure 3-7).

As these results suggested that GCATG repeats were required for EIIIB 5' splice site usage, we analyzed the expression of M2AL2R2a derivatives that contained synthetic repeat inserts T3, M3, TP3 or MP3. Both minigenes M2AL2R2a+T3 and M2AL2R2a+TP3 expressed detectable amounts of completely spliced M2 exon, although partially spliced RNA was still present (Figure 3-8, lanes 6, 8, 15, and 17). Mutant repeats did not exhibit this effect (M2AL2R2a+M3 or M2AL2R2a+MP3; lanes 7, 9, 16, and 18). Splice site usage of the flanking exons in M2AL2R2a remained relatively constant regardless of insert identity (Figure 3-9). Therefore TGCATG repeats can partially rescue EIIIB 5' splice site usage. Rescue by insert TP3 (as assessed by ratios of partially to completely spliced RNA) was slightly greater than that by insert T3, particularly in HeLa cells, thus correlating with the effects of these inserts in 7iBi89ΔGA (Figure 3-2). Since only partial rescue was effected by six TGCATG repeats, it is possible that additional sequences in IVS2 are required for proper 5' splice site usage; alternatively, partial effects may have been a consequence of inappropriate synthetic repeat positioning within this intron. Nonetheless, these results suggest that one function of TGCATG sequences is to promote more efficient usage of the EIIIB 5' splice site.

TGCATG repeats can activate an unrelated alternative exon

We next tested whether these repeats could activate exons unrelated to EIIIB by testing TGCATG inserts in the rat preprotachykinin (PPTK) gene transcript. The PPTK gene contains two alternative cassette-type exons, E4 and E6. We focused on exon E4, since E4 is only spliced into PPTK mRNA about 25% of the time in a wide range of cell types (Krause et al., 1987; Nasim et al.,

1990). Therefore, any positive effects of TGCATG repeats on E4 selection should be detectable. If so, it should also be possible to assess whether TGCATG-dependent exon activation is cell-type-dependent. We transferred a number of oligonucleotide inserts into a minigene expressing exons E2 through E7 of the rat PPTK gene transcription unit; inserts were positioned in the intron between E4 and E5 (Figure 3-10A). RNA from transiently transfected COS, HeLa, NIH3T3 and F9 cells was analyzed for E4 splicing by RNase protection, using a probe derived from a PPTK cDNA containing exons E1 through E5.

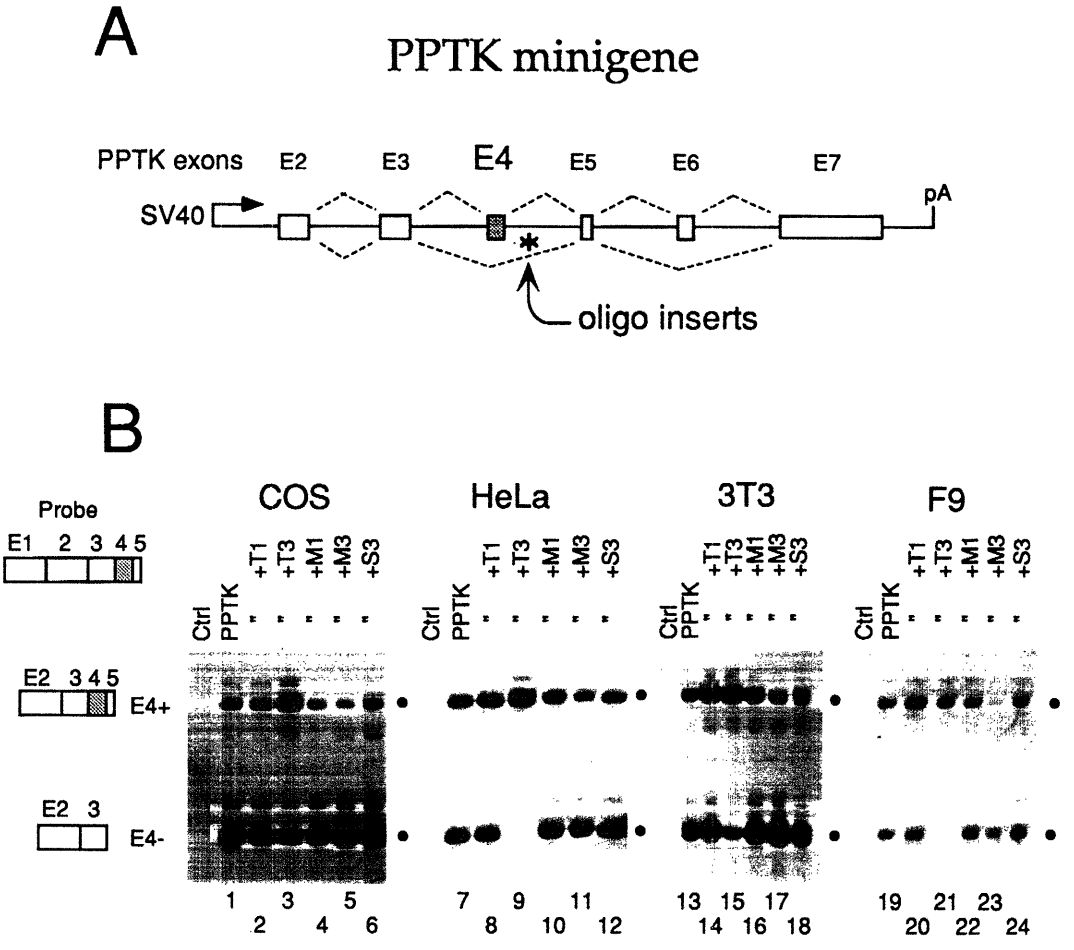
As expected, E4 inclusion in the unmodified PPTK minigene was relatively low in these four cell types (Figure 3-10B, lanes 1, 7, 13, and 19). Although some quantitative differences were apparent, E4 inclusion did not differ drastically between cell lines. Insert T3 was found to have a significant positive effect upon E4 inclusion (PPTK+T3; lanes 3, 9, 15, and 21) relative to the control inserts M3 (lanes 5, 11, 17, and 23) or S3 (lanes 6, 12, 18, and 24). The effects of two-repeat inserts were also tested (T1 versus M1). Enhancement of E4 usage by insert T1 was slight but significant compared with insert M1. The degree of exon activation ranged from low (COS) to very high (HeLa, F9). The degree of TGCATG-dependent E4 activation among COS, NIH3T3 and F9 cells (COS < 3T3 < F9; Figure 3-10B) correlated with EIIIB regulation in 7iBi89 (Chapter Two). We note that, although EIIIB inclusion is normally low in HeLa cells, TGCATG-dependent E4 activation in HeLa cells was unexpectedly high. We conclude that TGCATG repeats are capable of activating alternatively spliced exons unrelated to EIIIB. Additionally, these results establish that TGCATG repeats provide at least part of the cell-type-specificity of EIIIB regulation in the FN gene.

Figure 3-10.

TGCATG repeats function in a heterologous gene

A. Diagram of the rat preprotachykinin (PPTK) minigene, along with the locations of alternatively spliced exons E4 (shaded) and E6. This minigene contains the SV40 early promoter and 3' poly(A) signals (Nasim et al, 1990). Other labelling conventions are as in Figure 3-1. Synthetic inserts were placed at a BglII site (asterisk) within the E4-E5 intron, located 136 nt downstream of E4 and 328 nt upstream of E5.

B. RNase protection analysis of the transfected PPTK minigene and its derivatives. A cDNA containing PPTK exons E1 through E5 was used as a probe. Analyses were carried out on total RNA from untransfected cells (Ctrl) or from transiently transfected COS, HeLa, 3T3 or F9 cells (cell lines are indicated above each panel; transfected minigenes are indicated above each lane). The protected fragments corresponding to $E4^+$ and $E4^-$ mRNAs are indicated at left. The full-length riboprobe ("Probe") is large and migrates much more slowly than the $E4^+/E4^-$ fragments and is not shown here.


A
PPTK minigene
PPTK exons
E2
E3
E4
E5
E6
E7
SV40
pA
oligo inserts
B
Probe
E1 2 3 4 5
E2 3 4 5
E4+
E2 3
E4-
COS
HeLa
3T3
F9
Ctrl
PPTK
+T1
+T3
+M1
+M3
+S3
1 3 5
2 4 6
7 9 11
8 10 12
13 15 17
14 16 18
19 21 23
20 22 24

DISCUSSION

We have identified a short hexanucleotide element (TGCATG) that activates the splicing of the EIIIB exon in the rat fibronectin gene. This element, as a repeated array, can also promote the regulated inclusion of a heterologous alternatively spliced exon. This novel short motif therefore regulates splice site selection when present in alternatively spliced pre-mRNAs.

TGCATG sequences in IVS2 activate splicing of EIIIB

The EIIIB minigene 7iBi89 exhibited high levels of EIIIB inclusion in F9 cells, low inclusion in HeLa cells and complete EIIIB-skipping in COS cells. Previous studies identified a region in the intron downstream of EIIIB (IVS2) that was important for the recognition of EIIIB for splicing (Huh and Hynes, 1993). Several TGCATG repeats are present in this relatively large (>500 nt) region (Figure 3-1). Deletions of separate repeat-containing regions in IVS2 abolished EIIIB inclusion in F9 cells. Synthetic TGCATG repeats that replaced these deletions restored F9 inclusion (Figures 3-2, 3-3). These repeats had reduced effects in HeLa cells (which normally express low EIIIB inclusion) and had no effect in COS cells (which skip EIIIB completely); repeat activity therefore appeared to vary according to cell type. In a related FN minigene, M2 (Figure 3-6), deletion of certain TGCATG-containing IVS2 sequences significantly reduced F9 and HeLa exon inclusion while having little effect on the low level of inclusion in COS cells. The effects of this deletion could be reversed by reinsertion of an appropriate number of synthetic TGCATG repeats, confirming that these repeated sequences are indeed capable of cell-type-dependent action.

TGCATG sequences also activated preprotachykinin (PPTK) exon E4 inclusion when placed downstream of E4 (Figure 3-10). Therefore exons EIIIB and E4 may both be intrinsically poorly recognized for similar reasons. *In vitro* studies of PPTK alternative splicing have indicated that the 5' splice site of E4 binds U1 snRNP poorly; as a consequence, recruitment of the splicing factor U2AF to the E4 3' splice site is also compromised (Kuo et al., 1991; Hoffman et al., 1992). In addition, mutations in the E4 or EIIIB 5' splice sites that improve donor complementarity to U1 snRNA can also significantly increase inclusion (Nasim et al., 1990; see also Chapter Two). Since TGCATG repeats can activate the EIIIB 5' splice site (Figures 3-7, 3-8), TGCATG sequences may act by facilitating U1-5' splice site interactions and the exon selection events that follow.

Sequence specificity of the element

Besides the TGCATGA consensus, no other obvious inter-repeat homologies are apparent in IVS2 (Figure 3-1). Since hexamer elements can act from within different sequence contexts (Figures 3-2 and 3-10), specific RNA secondary structures appear not to be critical for hexamer function. Therefore it is likely that the TGCATG motif comprises (or contains) the essential *cis*-active element.

Both the degree and cell-type-specificity of repeat-dependent exon inclusion can be affected by sequences flanking the synthetic hexamers. For example, TGCATG sequences were most active in a purine-rich context. Interestingly, three TGCATGA/TGCATGC repeat pairs stimulated EIIIB inclusion more than either six TGCATGA's or six TGCATGC's (Figures 3-2, 3-4). This finding could reflect either subtle context effects or a synergistic interaction between distinct repeat-binding factors with overlapping sequence

specificities. Context-dependent differences could also reflect differential hexamer accessibility or factor affinity. Cell-type-specificity was most closely mimicked by TGCATGA repeats (Figure 3-6), which could explain why five of nine IVS2 repeats have this sequence. Many of the naturally-occurring GCATG repeats in IVS2 are adjacent to evolutionarily conserved sequences (Figure 3-1), possibly reflecting a requirement for appropriate context.

Taken together, these data indicate that repeated TGCATG hexamers can account for the activity of most if not all of the *cis*-elements previously identified within IVS2. Although the presence of other EIIIB-regulating elements cannot be ruled out, we have shown that TGCATG sequences can regulate EIIIB if present in enough copies and in the proper sequence context. Therefore TGCATG repeats probably play a major essential role in the splicing of EIIIB into FN mRNA.

Is the TGCATG element a regulatory splicing signal?

Whereas a constitutive signal might be expected to exert equal effects in all cell types, a regulatory element should exert cell-type-dependent effects. The repeated TGCATG element meets this latter criterion in three separate minigenes. First, TGCATG repeats activated EIIIB usage in a minigene in which inclusion was compromised by an IVS2 deletion (7iBi89ΔGA). TGCATG-dependent EIIIB inclusion was highest in F9, lower in HeLa and undetectable in COS cells, correlating with the ability of each cell line to utilize EIIIB. Furthermore, in two other minigenes that exhibited detectable inclusion in all cell lines (M2 and PPTK), TGCATG repeats again had cell-type-dependent effects (Figures 3-6 and 3-10). Therefore these repeats appear to comprise regulatory rather than constitutive signals.

We note that, although HeLa cells normally exhibit low EIIIB inclusion, the response of HeLa cells to TGCATG repeats in many cases was disproportionately high. For example, PPTK E4 activation was unexpectedly high in HeLa cells, despite apparently EIIIB-like regulation of E4 in the other cell lines tested (Figure 3-10). Hence it is possible that other FN gene sequences contribute to EIIIB regulation, perhaps by affecting hexamer context or by acting independently.

A number of studies in other vertebrate genes have identified sequences that control splicing. Secondary structure and intron sequences near exon 7 in the rat β-tropomyosin gene (exon 6B in chicken) can repress usage of this exon (Helfman et al., 1990; Libri et al., 1991; D'Orval et al., 1991; Guo et al., 1991; Gallego et al., 1992). An element that inhibits splicing in Rous sarcoma virus can also function in a heterologous intron (Arrigo and Beemon, 1988; McNally et al., 1991). Elements that activate splicing in the *c-src* gene and others that inhibit splicing in the calcitonin/CGRP gene have also been described (Black, 1992; Emeson et al., 1989). For many of these cases, minimal essential elements that affect splicing have not yet been defined; however, parallels may exist between regulation of the *c-src* and calcitonin/CGRP systems and the regulation of EIIIB, as will be commented on below.

In comparison, the regulatory element we have identified is novel in a number of respects. Repeated TGCATG sequences are capable of functioning in a cell-type-dependent fashion in a heterologous context, a finding that has not been conclusively established for many of the other potential vertebrate elements reported to date. In addition, the remarkably diminutive character of this repeat may have widespread implications for regulated alternative splicing.

How might TGCATG elements work?

The removal of seven GCATG repeats from IVS2 resulted in inactivation of the EIIIB splice donor, a phenotype that was partially rescued by synthetic TGCATG repeats (Figure 3-7, 3-8). Therefore these elements may recruit factors that directly activate nearby 5' splice sites (Figure 3-11A). As discussed previously, activation of EIIIB (or PPTK exon E4) could occur by repeat-facilitated U1 snRNP binding. Alternatively, these repeats may render IVS2 more accessible for splicing, perhaps by recruiting factors that configure this intron appropriately. A number of hnRNP proteins exhibit sequence- or transcript-specificity (Swanson and Dreyfuss, 1988a, 1988b; Bennett et al., 1992b; Matunis et al., 1993); some of these may prefer to bind UGCAUG. Other possible UGCAUG-binding candidates include members of the SR protein family, which are splicing factors that can promote proximal 5' splice site selection in cell-free extracts (Fu et al., 1992; Zahler et al., 1992). Indeed, SR proteins have been implicated in *Drosophila dsx* regulation *in vitro* (Tian and Maniatis, 1993) and may also interact with the purine-rich exon element that has been found in some exons (Lavigueur et al., 1993; Sun et al., 1993b). Different SR proteins have also been shown to exhibit substrate-specificity *in vitro*, either for commitment to splicing (Fu, 1993) or for proximal 5' splice site selection (Zahler et al., 1993); these observations may reflect distinct sequence specificities among individual SR family members.

Additionally, these repeats may inhibit or delay the usage of the IVS2 splice acceptor (which is used regardless of splice choice), allowing more opportunity for EIIIB recognition (Figure 3-11A); this might explain why several repeats are located in the 3' portion of IVS2. Regulation of this type is not without precedent, since studies of SV40 and adenovirus alternative splicing have uncovered instances where 5' splice site selection can be

governed by alterations at a common 3' splice site (Fu and Manley, 1988; Ulfendahl et al., 1989).

In principle, it is also possible that these repeats represent transcriptional pause sites, in which case a delay in exon III-8a acceptor commitment might result from pausing of RNA polymerase II during transcription of the repeat-rich region within IVS2. Evidence for cotranscriptional splicing has been documented in insect systems (Beyer and Osheim, 1988; Lemaire and Thummel, 1990; Baurén and Wieslander, 1994); therefore such a mechanism could effectively result in delayed exon III-8a commitment. If commitment delay were the mechanism by which these repeats functioned, then it should be possible to assay these repeats upstream of an alternative 3' splice site and thereby discriminate between models invoking splice site attenuation or delayed transcription (see Chapter Four).

In any event, it is likely that differentially expressed cellular factors mediate these repeat-dependent effects, either by interacting with hexamer targets directly or by interacting with ubiquitous hexamer-binding proteins. The detailed characterization of such cellular factors should offer further insights regarding mechanisms of alternative splicing.

Do TGCATG sequences regulate other alternatively spliced genes?

Given our observations, short sequences could in principle govern regulation in other alternatively spliced genes. In fact, the TGCATG element occurs in at least two other instances of RNA regulation. Studies of the *c-src* gene have identified at least two intron elements, located downstream of the alternatively spliced N1 exon, that are important for neural-specific N1 exon inclusion (Black, 1992). It is interesting to note that a TGCATG sequence occurs in one of these elements and that a mutation that reduced N1 inclusion

also disrupted this hexamer (Figure 3-11B). Moreover, extensive deletions that removed both intron elements resulted in production of unspliced or partially spliced RNA, a result similar to that obtained upon removal of seven repeats from FN minigenes. These parallels with our data suggest that TGCATG-related sequences may participate in *c-src* splicing regulation, probably in conjunction with additional neural-specific elements.

In the calcitonin/CGRP gene, HeLa and thyroid C cells produce calcitonin by using exon 4 whereas neurons and F9 cells produce calcitonin gene-related peptide (CGRP) by using exons 5 and 6 instead (Leff et al., 1987; Emeson et al., 1989; Figure 3-11C). Regulation is evident at the calcitonin-specific exon 4 splice acceptor , which is used in HeLa cells but repressed in F9 cells (Emeson et al., 1989). We note that eight GCATG repeats are clustered within 700 bases surrounding this splice acceptor in the human gene (Figure 3-11C). Five of these are TGCATG. In studies using the rat gene, derepression of the exon 4 acceptor in F9 cells could be induced by nearby upstream mutations, a number of which removed one or more GCATG repeats (Emeson et al, 1989). Although a more recent and comprehensive analysis of the rat gene failed to identify any single element that was consistently critical for regulation (Yeakley et al., 1993), the influence of multiple dispersed elements could explain these recent findings, particularly if such elements were functionally redundant. CGRP-specific splicing in F9 cells might then occur by repeat-dependent downregulation of the exon 4 splice acceptor, similar to the way in which these repeats may attenuate the 3' splice site downstream of EIIIB in the same cells (Figure 3-11A). Although there is no conclusive evidence that these repeats regulate calcitonin/CGRP, it is nevertheless intriguing that multiple copies of a splicing signal for EIIIB should occur (at a much higher density than predicted by chance) in a gene characterized by a different mode of regulated

alternative splicing. It seems unlikely that TGCATG-dependent splicing regulation is unique to EIIIB; in fact, given these correlations among FN, *c-src* and calcitonin/CGRP, it is tempting to speculate that TGCATG-related motifs may participate in a range of embryo- or neural-specific RNA processing events.

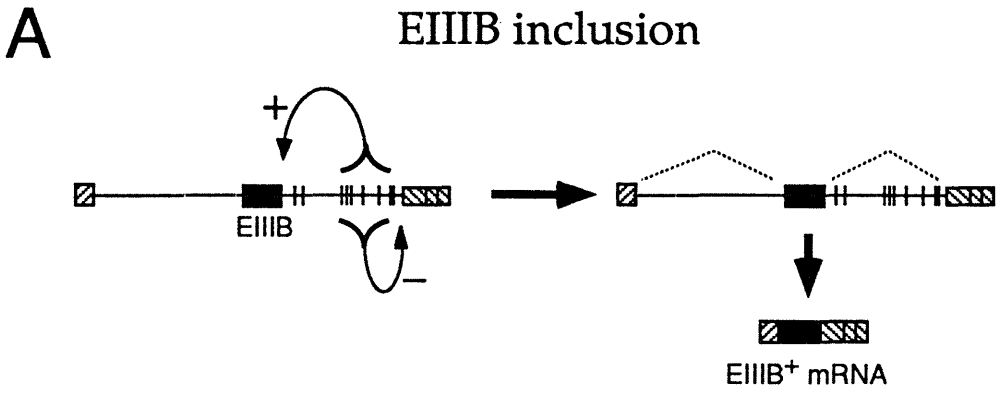
Figure 3-11.

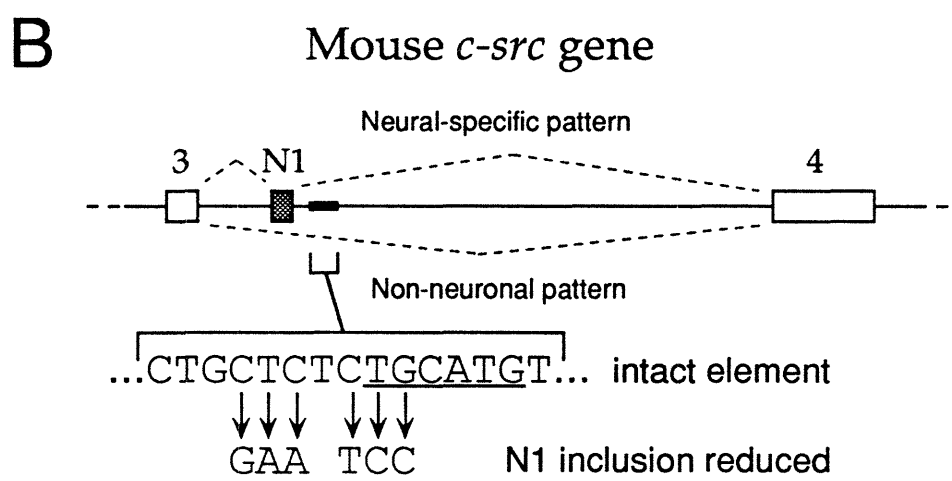
Repeats in other alternatively processed genes?

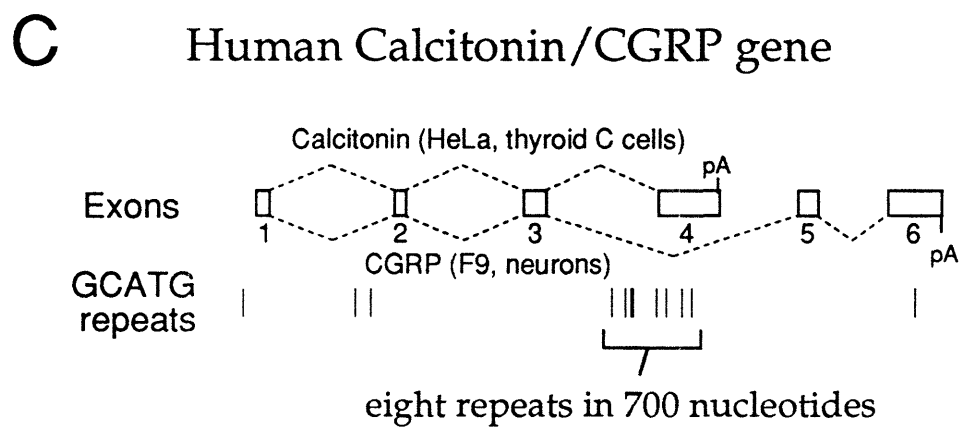
A. Repeat-dependent activation of EIIIB inclusion. The 7iBi89 minigene is shown (vertical lines, GCATG repeats). Arrows labeled with plus or minus signs denote either positive or negative effects on splice site usage that may be mediated by TGCATG repeats. These and other possible models are discussed in the text. In the absence of these repeats (or of repeat-dependent factors that activate EIIIB recognition), EIIIB-skipping is the predominant splicing event.

B. Organization of the murine c-*src* gene in the region of the alternatively spliced N1 exon (derived from Black, 1992). Conventions are as in Figure 3-1. Exon N1 is shaded. The black bar indicates the location of neural-specific N1-activating elements characterized in a study by Black (1992); the sequence of one of these elements is also shown (the TGCATG sequence is underlined), along with a mutation made in this element that reduced N1 inclusion splicing both in vivo and in vitro.

C. Organization of the calcitonin/α-CGRP gene; conventions are as in Figure 3-1. The locations of GCATG repeats (derived from the human sequence) are shown below this diagram. Note that the random expectation is one such repeat per 1024 nucleotides, yet several are clustered in the vicinity of exon 4.


A
EIIIB inclusion
+
EIIIB
−
EIIIB+ mRNA


B
Mouse c-src gene
Neural-specific pattern
3
N1
4
Non-neuronal pattern
...CTGCTCTCTGCATGT... intact element
GAA TCC
N1 inclusion reduced


C
Human Calcitonin/CGRP gene
Calcitonin (HeLa, thyroid C cells)
Exons
pA
1
2
3
4
5
6
pA
CGRP (F9, neurons)
GCATG
repeats
eight repeats in 700 nucleotides

MATERIALS AND METHODS

Oligonucleotides and Computer Analysis

Each double-stranded DNA oligonucleotide consisted of annealed sense and antisense oligonucleotides. Sense strand sequences are shown in Figure 3-2A; antisense oligonucleotides consisted of the sequence 5'-GATC-3' followed by the reverse complement to nucleotides 5-35 of each sense sequence.

Comparisons of rat FN (Schwarzbauer et al., 1987; R. Patel-King, J. E. Schwarzbauer, P. A. Norton, and R. O. Hynes, unpublished data; GenBank accession number L20801) and human FN genomic sequences (Paolella et al., 1988; GenBank accession number X07717) were carried out using version 7.0 of the UWGCG software package (Devereux et al., 1984); the comparison shown in Figure 3-1 was adapted from the output from a BESTFIT comparison (using default parameters). Analysis of repeats in the rat FN gene and in the human calcitonin/α-CGRP gene (Broad et al., 1989; GenBank accession number X15943) was carried out using the nucleotide interpretation program of the STADEN software package (Staden, 1990).

DNA Constructs and Probes

DNA manipulation procedures were as described previously (Ausubel et al., 1987; Sambrook et al., 1989). Restriction enzymes were obtained from New England Biolabs. Minigenes 7iBi89 and M2 in expression vectors pBAGH or pBAGH.Sv, along with their deletion derivatives and the S1 analysis probe, have been described previously (Huh and Hynes, 1993). The IVS2 deletions in this study (in 7iBi89 or in M2) removed the following nucleotides from the 1071 base IVS2 (numbering as in Figure 3-1): ΔGA, bases 235-729; ΔAR2a, bases 734-1039; ΔAL2, bases 417-729; ΔAL2R2a, bases 417-1039.

The following procedure was carried out with each of the DNA oligonucleotide pairs in order to create single or multiple inserts in 7iBi89ΔGA. After gel purification, 800 pmol of each of the sense and antisense oligonucleotides were annealed (in a total volume of 40 µl TE [10 mM Tris-HCl, 1 mM EDTA pH 8.0]) by cooling slowly from 90°C to 30°C. 20 µl aliquots of the annealed mixture were then phosphorylated in a total volume of 40 µl using 40 U 3'-phosphatase-free polynucleotide kinase (Boehringer-Mannheim). After phenol-chloroform (PC) extraction and ethanol precipitation, the mixture was ligated (in 20 µl overnight at 15°C) using 200U T4 DNA ligase. The ligated oligonucleotides were then PC-extracted, precipitated and digested with BamHI and BglII to eliminate all but head-to-tail ligation joins (for inserts TP and MP, BclI was used instead of BamHI). This ligation-digestion cycle was repeated two more times to ensure an adequate production of appropriately ligated multimers. The final digest was then blunt-ended with Klenow fragment and dNTPs; this mix was then ligated to AflII-cut, blunt-ended 7iBi89ΔGA/pBAGH. DNAs from transformed colonies were sequenced to ensure appropriate insert size, identity and orientation. 7iBi89ΔGA/pBAGH.Sv derivatives were made by transferring a XhoI-EcoRI fragment (containing the entire transcription unit) from 7iBi89ΔGA/pBAGH into pBAGH.Sv.

Inserts were excised from 7iBi89ΔGA/pBAGH via BglII excision; these were ligated to BglII-cut M2 GA/pBAGH.Sv or M2 AL2R2a/pBAGH.Sv to make the insert-containing M2ΔGA and M2 AL2R2a derivatives. The RNase protection probe used for assessing M2 central exon splice site usage (Figure 3-8) was derived from a HindIII-PstI fragment of the M2 minigene that contains the central exon (in pBluescript SK$^-$; Stratagene); KpnI-linearized plasmid was the template for T3 RNA polymerase (Stratagene). The probes used in Figure

3-9 (left) were T7 antisense transcripts of a 7iBi89 template (in pGEM3) that carried the deletion XE (Chapter Two) plus a BglII-DraI deletion in IVS2. The probe used for the right of this figure was transcribed (by T7) from an 8i9i10-derived pGEM3 template containing a 734 bp AflII-KpnI deletion in IVS1 and a 1054 bp AccI-EcoRI deletion in IVS2. All riboprobes were purified via preparative electrophoresis on a 4% or 5% denaturing polyacrylamide gel.

Preprotachykinin (PPTK) minigene constructs and probes were derived from pBPSVpA$^+$2-7 (a plasmid containing exons E2 through E7 of the rat PPTK gene inserted between SV40 promoter and poly(A) signals) and pRPC1, a β-PPTK cDNA clone in pBS (Stratagene) containing PPTK exons E1 through E7 (both kind gifts from P. J. Grabowski; Nasim et al., 1990). Oligo inserts were inserted into a BglII site in pBPSVpA$^+$2-7, in the intron between E4 and E5 (Figure 3-10A). The RNase probe used for assessing E4 inclusion was made by cutting pRPC1 with BglII and BamHI and religating the vector, generating a subclone containing exons E1 through E4 and part of E5. This was linearized with EcoRI and transcribed with T3 RNA polymerase.

In some experiments, levels of expression were were compared, using cotransfected pSV2neo as an internal control. An EcoRI-HindIII fragment of the plasmid pLJ (containing part of the neomycin resistance gene; Guan et al., 1990) was cloned into pGEM2 and was used for both S1 and RNase protection assays. For S1 analyses, this plasmid was cut with BglII, kinased and then recut with EcoRI. For RNase protection analyses, this plasmid was linearized with HindIII and transcribed *in vitro* using SP6 RNA polymerase (Promega). All probes were gel purified before use.

Cell culture and RNA analysis

Cell culture methods, transfection protocols and RNA analysis by S1 nuclease or RNase protection assays were as described previously (Chapter One). S1 analysis was carried out using 5' end-labelled probes, whereas RNase protection analyses were carried out with uniformly labelled riboprobes. Quantitation of bands was carried out using a Molecular Dynamics PhosphorImager.

ACKNOWLEDGEMENTS

We are grateful to Paula Grabowski for providing us with the rat preprotachykinin plasmids used in some of these studies. We also wish to thank Ben Blencowe, Melissa Moore, Charles Query, Phillip Sharp, Peggy Kolm, Mark Borowsky, Grant Wheeler and Pamela Norton for helpful discussions and/or criticisms of the manuscript from which this chapter was derived.

Chapter Four

Regulation of Alternative Calcitonin/CGRP pre-mRNA Processing By Repeated TGCATG Elements

ABSTRACT

The EIIIB exon in the rat fibronectin gene is regulated in a cell-type-specific fashion. Cell-type-specific EIIIB inclusion is dependent upon a hexanucleotide sequence that occurs several times in the intron downstream of EIIIB. An inspection of other alternatively spliced genes revealed that these hexamer elements may participate in other cell-type-specific splicing events, in particular those of *c-src* and of the calcitonin/CGRP gene. The presence of several GCATG repeats near the regulated splice acceptor of the calcitonin-specific exon in this gene suggested that TGCATG repeats could specifically attenuate an adjacent splice acceptor as part of their function. To test this hypothesis, TGCATG repeats were placed upstream of the calcitonin-specific exon splice acceptor in mutant calcitonin/CGRP constructs that exhibited elevated usage of this exon in CGRP-producing cell types. Inserted TGCATG hexamers (but not control hexamers) were found to decrease calcitonin splicing and promote CGRP splicing; furthermore, this effect was somewhat cell-type-dependent. These results indicate that TGCATG elements can attenuate 3' splice site recognition. These findings are consistent with the hypothesis that TGCATG-related sequences normally participate in the regulation of calcitonin/CGRP alternative RNA processing. Moreover, these experiments provide further support for the role of these hexamer elements as cell-type-specific determinants of alternative splicing.

INTRODUCTION

From an examination of regulated alternative splicing in *Drosophila*, it is clear that both positive and negative modes of splicing regulation are possible (For reviews, see Baker, 1989; McKeown, 1992; Mattox et al., 1992). It is also apparent that one splicing regulator is capable of affecting splice site choice in more than one gene. For example, the *Sex-lethal* gene product regulates the splicing of its own pre-mRNA as well as of that of *tra*. The *suppressor of white apricot* gene, in addition to affecting expression of the *white apricot* allele, appears to exhibit negative autoregulation (for review, see Mattox et al., 1992). The MER1 gene in yeast has been genetically inferred to have an additional function distinct from the activation of splicing of the MER2 intron (Engebrecht et al., 1991). Furthermore, a given regulator of splicing may exert either positive or negative effects upon splicing, depending on the situation and the target pre-mRNA. The product of the *tra-2* gene, in conjunction with that of the *tra* gene, acts to stimulate 3' splice site usage (and possibly also polyadenylation) of the *doublesex* gene (Ryner and Baker, 1991; Hedley and Maniatis, 1991); in addition, the *tra-2* gene product appears to repress the splicing of an intron in its own pre-mRNA in the male germ line (Mattox and Baker, 1991). Therefore it is not unreasonable to imagine that a regulator of splicing in vertebrates could control the processing of several genes simultaneously.

The previous chapter examined the role of a hexamer sequence (TGCATG) that cell-type-specifically regulates EIIIB inclusion in the fibronectin gene. This *cis*-element occurs eight times in the intron downstream of EIIIB. Deletions that removed sequences containing several of these repeats were found to reduce EIIIB inclusion. The replacement of these deleted

sequences with synthetic TGCATG-containing oligonucleotide inserts was sufficient to restore EIIIB inclusion. This repeated hexamer was capable of acting in a cell-type-dependent manner and in addition could function when placed in a heterologous context. Therefore this repeated TGCATG sequence represents a major control element of EIIIB alternative splicing that contributes to the cell-type-specificity of this exon's regulation. In addition, sequences that are important in neural-specific *c-src* and calcitonin/CGRP alternative splicing also contain TGCATG sequences. These observations raise the intriguing possibility that TGCATG elements participate in determining the cell-type-specificity of alternative splicing in these and possibly additional genes.

In particular, the calcitonin/CGRP gene exhibits a pattern of alternative splicing distinct from that exhibited by the fibronectin EIIIB exon. The pre-mRNA transcript that encodes either calcitonin or calcitonin gene-related peptide (CGRP) contains six exons. Production of calcitonin mRNA results from splicing of exons 1 through 4; polyadenylation occurs at the 3' terminus of exon 4. CGRP mRNA is produced when exon 3 is spliced to exons 5 and 6 instead, with 3' end formation at exon 6 (Figure 4-1). The choice between calcitonin and CGRP processing is cell-type-specific; *in vivo*, calcitonin is produced by thyroid C cells, whereas CGRP is produced by certain neurons (Amara et al., 1982). The splice acceptors of exons 4 and 5 thus appear to compete with each other for splicing to the common exon 3 donor. In principle, calcitonin/CGRP processing choice could also involve differential usage of polyadenylation sites.

How is the calcitonin/CGRP processing choice made?

Calcitonin/CGRP processing represents a complex system wherein both differential 3' splice site selection and polyadenylation occurs. Studies of such

systems have posed difficulties beyond those posed by systems involving differential splicing alone. A key determining event must dedicate the calcitonin/CGRP pre-mRNA to a particular processing pathway. Varying degrees of experimental support have been provided for either splicing or polyadenylation as this commiting event. For example, a calcitonin-specific splice in a human gene-derived substrate was detected *in vitro* only if sequences downstream of exon 4 were not present (Bovenberg et al., 1988), suggesting that polyadenylation at exon 4 induces calcitonin-specific splicing. However, since a substrate containing exons 3 through 5 exhibited predominantly CGRP-specific processing regardless of splicing extract source, it could not be assumed that key processing events were rate-limiting *in vitro.*

On the other hand, transfection studies have supported splice site commitment as the key event. If both calcitonin and CGRP polyadenylation sites were placed in a construct without splice sites, both sites were used efficiently in both calcitonin- and CGRP-preferring cell lines (Leff et al., 1987); in addition, replacement of the exon 4 poly(A) site with that of exon 6 did not affect cell-type-specificity (Emeson et al., 1989). Thus specific polyadenylation sequences were neither sufficient nor critical determinants of processing choice. In addition, nuclear runon experiments have determined that transcriptional termination in the rat gene occurs 3' of exon 6, even in cells that predominantly produced calcitonin, arguing that differential transcriptional termination is not a determining event (Amara et al., 1984).

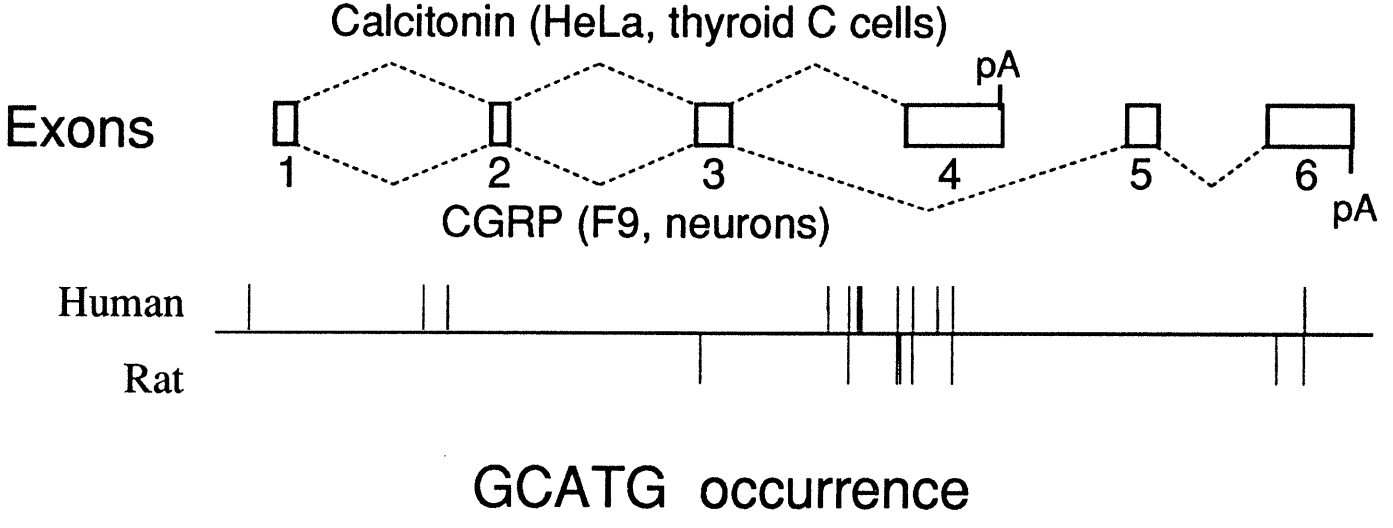
In contrast, mutations removing the exon 5 splice acceptor fail to promote exon 4 usage in CGRP-preferring cells, whereas mutations ablating the exon 4 splice acceptor allowed CGRP-specific processing even in calcitonin-preferring cell types (Leff et al., 1987; Emeson et al., 1989). These experiments indicate that the exon 4 splice acceptor is the site that is regulated. The

Figure 4-1.

GCATG repeats in the human and rat calcitonin/CGRP genes

The structure of the alternatively spliced rat and human calcitonin/CGRP genes. Exons are represented by rectangles; dashed lines above and below indicate possible splicing/polyadenylation patterns and the mRNAs produced, along with the cell lines or cell types in which these patterns predominate (pA, polyadenylation site; scale bar, 500 nucleotides). The incidence of GCATG pentanucleotides is indicated on the horizontal linear coordinate below this diagram. Vertical lines extending above and below this coordinate indicate occurrences of GCATG in the human and rat genes, respectively (derived from the known sequences of the human and rat genes; Broad et al., 1989; J. Yeakley and M. G. Rosenfeld, personal communication). The GCATG incidence between exons 1 and 3 in the rat gene are not known due to lack of sequence information in this area.

Calcitonin/CGRP gene structure



existence of *cis*-active inhibitory sequences upstream of the exon 4 acceptor was demonstrated by small nested deletions in this region (Emeson et al., 1989), supporting the role of upstream sequences in CGRP-producing cells (like F9 cells). On the other hand, exon 4 sequences have also been identified that are required for calcitonin processing (Cote et al., 1992; van Oers et al., 1994). Splicing of the calcitonin-specific acceptor *in vitro* has been shown to involve usage of an unconventional uridine acceptor branch, suggesting that calcitonin exon usage is intrinsically poor and that additional factors are required for its efficient recognition in calcitonin-producing cells (Adema et al., 1990). However, mutation of this site to an adenosine in the rat gene has been shown not to affect cell type-specific regulation if assayed in the presence of appropriate nonspecific mutations (Yeakley et al., 1993).

Thus two possible models for calcitonin/CGRP regulation are possible: one model posits negatively acting factors that inhibit the exon 4 splice acceptor in CGRP-producing cells, whereas the other model proposes the cell-type-specific activation of exon 4 in calcitonin-producing cells. A constitutively-expressed calcitonin/CGRP transgene has been shown to exhibit calcitonin-specific processing in most tissues with the exception of neurons (Crenshaw et al., 1987). If one accepts an unproven assumption made in this study (i. e., that only regulated cell types should contain cell-type-specific factors), this finding would appear to support a model involving exon 4 repression specificaaly in neurons.

Some other puzzling observations remain unexplained. For instance, mutations or deletions that eliminate the calcitonin exon 4 polyadenylation signal do not always result in CGRP-processing by default but can cause the accumulation of partially spliced RNA, even in cells that normally produce

CGRP (Leff et al., 1989); therefore cell-type-specific mechanisms may also exist to promote CGRP processing, although this is far from clear.

Do GCATG repeats regulate calcitonin/CGRP processing?

From the available evidence, it still remains a matter of debate whether calcitonin acceptor usage is negatively regulated in neurons or positively regulated elsewhere. Inspection of the human calcitonin gene sequence has revealed the presence of several TGCATG and GCATG repeats near the calcitonin-specific splice acceptor. Several GCATG repeats also occur in this region in the rat gene. In comparing the effects of nested deletions carried out by Emeson et al. (1989), we observed that deletions that activated exon 4 splicing had removed one or more GCATG repeats. Since TGCATG repeats have been demonstrated to function in a heterologous context and also occur in elements that affect *c-src* alternative splicing, it seemed possible that TGCATG-related sequences could control the calcitonin/CGRP processing decision. This chapter describes a direct test of this hypothesis in the rat calcitonin/CGRP gene. Synthetic hexamer-containing oligonucleotide inserts were tested for their effect on splicing in rat calcitonin/CGRP constructs that contained exon 4-activating mutations. From this test it was concluded that TGCATG repeats were indeed capable of reversing the effects of these mutations, most likely by negatively affecting the calcitonin-specific 3' splice site in a cell-type-dependent manner. These findings have additional implications for the mechanistic action of hexamer repeats in general.

RESULTS

A test of TGCATG function in the rat calcitonin/CGRP gene

Inspection of the human gene revealed eight GCATG repeats within a stretch of 700 nucleotides (nt) in the vicinity of the calcitonin-specific exon 4 splice acceptor (Figure 4-1). This represents an anomalously high occurrence of this sequence, since any given pentanucleotide would be expected to occur only once per 1024 nucleotides by random chance. Of these eight, five have the sequence TGCATG. In the rat gene, five GCATG repeats are present in the general vicinity of exon 4. Therefore several GCATG repeats were found in genes from both species; curiously, only four of these repeats are conserved between the human and rat genes: two conserved GCATG pentamers occur 32 and 252 nt upstream of the rat exon 4 splice acceptor and two occur 47 and 236 nt downstream . Although fewer repeats occur in the rat gene, one in partcular (CGCATG) is located only 17 nt upstream of exon 4, between the 3' splice junction and the branchpoint that has been mapped from *in vitro* studies of the human gene (Adema et al., 1988).

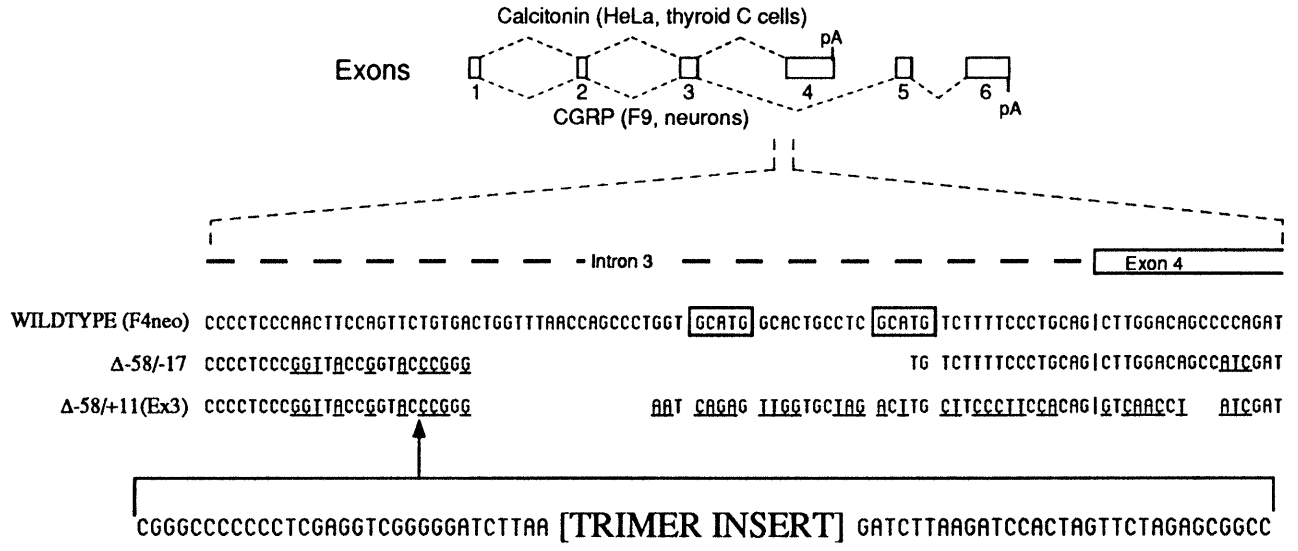
Studies using the rat gene have identified mutations and deletions that activate the use of exon 4 in F9 teratocarcinoma cells, a cell line that normally prefers CGRP- over calcitonin-processing (Leff et al., 1987; Emeson et al., 1989). The sequences of two of these mutations, -58/-17 and -58/+11(Ex3), are shown in Figure 4-2, along with the wildtype F4neo sequence. Mutation -58/-17 deletes or alters sequences between 17 and 58 nucleotides upstream of exon 4. In mutation -58/+11(Ex3), a 65 nt sequence spanning this 3' splice site was replaced by sequences normally spanning the splice acceptor from exon 3 in this gene. Both mutant constructs contained additional substitutions both upstream and downstream of the altered area; these base alterations, which

Figure 4-2.

Experimental test of hexamer repeats in mutant rat calcitonin/CGRP genes

The exon/intron structure of the rat calcitonin/CGRP line is indicated at top; the sequence of the calcitonin-specific acceptor region (exon 4) is shown below this diagram. The top sequence is that of the wildtype gene (F4neo). The next two lines are correspond to the same regions from the -58/-17 and -58/+11(Ex3) mutant forms of the F4neo construct (adapted from Emeson et al., 1989). Blank regions indicate deleted bases relative to the F4neo sequence; underlined residues indicate nucleotide changes relative to F4neo. GCATG repeats in this region are boxed. The general sequence of the inserts used in this study is shown in larger type below these exon 4 acceptor region sequences (derived from pBluescript polylinker); the arrow indicates the point in the calcitonin/CGRP gene at which these sequences were placed. Sequences of the monomeric forms of each insert are shown in the box at bottom. Shaded bases are differences from the top line (insert T). Trimerized versions of these inserts were inserted in tandem at the location indicated by "[TRIMER INSERT]".

Rat Calcitonin/CGRP gene



Monomer Inserts

T	GATCC TGCATG ACTACTGCT TGCATG CTCGTCATA
M	GATCC TGACTG ACTACTGCT TGACTG CTCGTCATA
S	GATCC AGTCGT ACTACTGCT AGTCGT CTCGTCATA
TP	GATCA TGCATG AGGGAAAGG TGCATG CAAAGGGAA
MP	GATCA TGACTG AGGGAAAGG TGACTG CAAAGGGAA

facilitated mutant construction (by creating KpnI and ClaI sites), had no effect upon calcitonin/CGRP splicing if introduced in the absence of other mutations (Emeson et al., 1989). Of the nested mutations and substitutions reported by Emeson et al. (1989), mutants -58/-17 and -58/+11(Ex3) significantly increased the percentage of calcitonin-specific processing in F9 cells (from 6% to 28% and 82% respectively).

To test whether TGCATG sequences could repress exon 4 usage, synthetic oligonucleotide inserts (Figure 4-2) were introduced into rat calcitonin/CGRP constructs containing the -58/-17 or -58/+11(Ex3) mutations. Each synthetic insert contained six hexamer repeats, separated by spacer sequences of varying composition; total insert length was 167 nt. Inserts T3, M3 and S3 contained TGCATG (test), TGACTG (mutant) and AGTCGT (scrambled) hexamers respectively; in this insert series, spacer sequences with a relatively high pyrimidine content were used to separate the hexamer units. Inserts TP3 and MP3 contained TGCATG and TGACTG hexamers respectively and in addition contained purine-rich spacers. Each insert contained three tandem repeats of a two-hexamer-containing oligonucleotide. In Chapter Three, it was found that TGCATG-containing inserts specifically enhanced EIIIB inclusion in a cell-type-dependent fashion when placed downstream of EIIIB, whereas none of the control inserts had this effect.

These insert-containing plasmids, along with their parental mutant constructs and the original F4neo plasmid, were transfected into a variety of cell lines that exhibited varying degrees of calcitonin/CGRP preference. Total RNA isolated from transfected cell lines was subjected to RNase protection analyses, using a probe that contained sequences spanning the splice acceptors of both exon 4 and exon 5. Protection by calcitonin/CGRP transcripts were predicted to yield RNase-resistant fragments of 588-793 nt (unspliced exon 4),

472 nt (spliced exon 4, specific for calcitonin), 300 nt (unspliced exon 5) and 183 nt (spliced exon 5, specific for CGRP). The 472 and 183 nt fragments therefore represented protection by calcitonin and CGRP mRNAs respectively whereas the other fragments were indicative of unspliced or partially spliced pre-mRNA. The relative intensities of the 472 and 183 nt bands (after normalizing for labeled uridine content) thus reflected calcitonin/CGRP splicing preference. Since oligonucleotide inserts were placed just upstream of the exon 4 acceptor, it was necessary to make probes specific for each F4neo construct in order to effectively distinguish between unspliced RNA and spliced calcitonin mRNA. Results for each cell line are described individually below and are quantitated in Figure 4-7.

TGCATG repeats restored calcitonin-specific processing in F9 cells

In stably transfected F9 teratocarcinoma cells (Figures 4-3, 4-7), the wildtype F4neo plasmid exhibited substantial amounts of CGRP expression but very little calcitonin splicing. Transfection of the -58/-17 and -58/+11(Ex3) mutant F4neo derivatives resulted in significantly elevated levels of calcitonin mRNA production. For all the -58/-17 derivatives, the band representing unspliced exon 5 migrated somewhat faster than expected; this was due to the occurrence of a spontaneous rearrangement in the intron separating exons 4 and 5 that was present in the parental -58/-17 construct. This rearrangement is not known to affect calcitonin/CGRP splice preference, either alone or in the context of other F4neo mutations (J. Yeakley, personal communication). The relative amount of calcitonin mRNA produced by the mutants -58/-17 and -58/+11(Ex3) compared well with the published values (26% and 86% respectively versus 28% and 82% as published; Emeson et al., 1989).

Insert T3 was found to reverse completely the effect of the -58/-17 mutation when placed upstream of the exon 4 acceptor. Expression of the construct -58/-17+T3 yielded a pattern matching that yielded by wildtype F4neo. The mutant and scrambled control inserts M3 and S3 did not exhibit this effect; therefore the observed repression of calcitonin exon 4 usage was mediated specifically by TGCATG sequences. Insert TP3 (which contains TGCATGs separated by purine-rich spacers) also affected the calcitonin/CGRP ratio, although not to the same extent as did insert T3; this effect was not observed using insert MP3. The finding that insert TP3 functioned less well than T3 was unexpected, since the reverse was found to be true in fibronectin-related minigenes (Chapter Three). Therefore the bases flanking the TGCATG sequence, although not critical, appeared to affect this element's action.

In the context of the -58/+11(Ex3) mutation, these inserts T3 and TP3 were found to exert effects that qualitatively paralleled those exerted in mutation -58/-17; although overall effects in the former construct were lower than in the latter construct. Therefore TGCATG elements repressed usage of a nearby downstream 3' splice site even if that 3' splice site sequence was derived from a constitutive exon. Together with the finding that TGCATG sequences are normally found near exon 4 in both the rat and human genes, these data provide evidence that TGCATG repeats can negatively regulate 3' splice site usage and furthermore may normally participate in repression of the calcitonin-specific splice acceptor in F9 cells.

In calcitonin-preferring Hela cells, TGCATG repeats have less activity

HeLa cells normally exhibit a preference for calcitonin processing (Emeson et al., 1989). The expression of these constructs in stably transfected HeLa cells (Figures 4-4, 4-7) differed from that exhibited by F9 cells in a

Figure 4-3.

The effects of hexamer elements on calcitonin/CGRP processing: F9 teratocarcinoma cells

RNase protection analyses of total RNA (10 μg) isolated from populations of F9 cells stably transfected with F4neo, its mutant derivatives -58/-17 and -58/+11(Ex3) or with mutants containing synthetic hexamer-containing inserts. Riboprobes used for this analysis contained intron sequences upstream of exon 4, exon 4, intron sequences upstream of exon 5 and exon 5 (probe structure shown at top right). RNase protection products were analyzed in a denaturing 4% polyacrylamide gel. Aliquots of representative probes were run at left. The remainder of the gel shows analyses of RNA from nontransfected F9 cells using the F4neo-derived probe (Ctrl) or RNA from F9 cells transfected with wildtype or mutant constructs containing the indicated synthetic inserts. Different probes were used for weach construct; none of the probes yield any significant signal when used to analyze RNA from nontransfected F9 cells. Positions of protected species are indicated at right. The proportions of spliced exon 4 and spliced exon 5 were taken to represent processing events specific for calcitonin and CGRP, respectively.

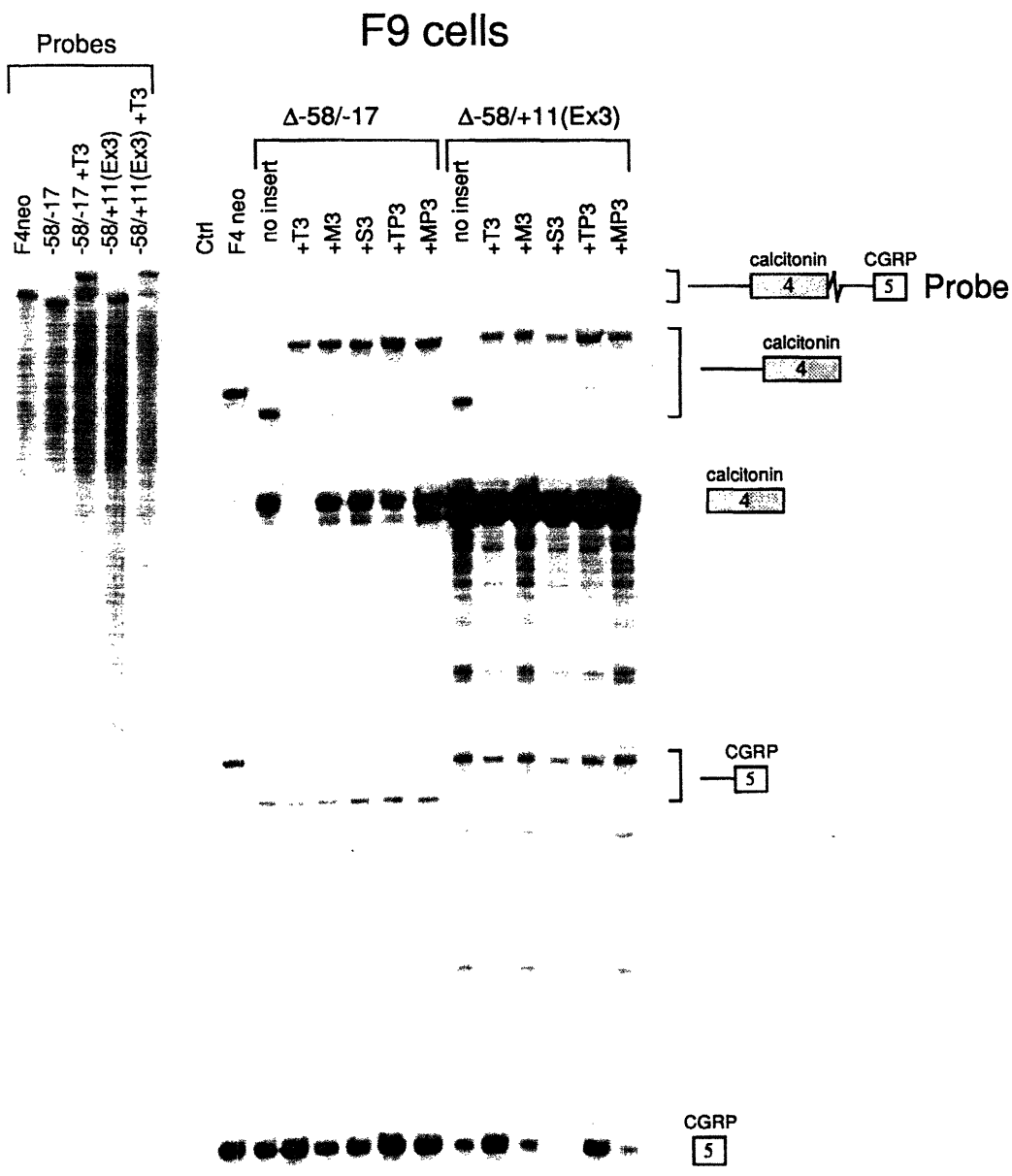

F9 cells
Probes
F4neo
-58/-17
-58/-17 +T3
-58/+11(Ex3)
-58/+11(Ex3) +T3
Δ-58/-17
Δ-58/+11(Ex3)
Ctrl
F4 neo
no insert
+T3
+M3
+S3
+TP3
+MP3
calcitonin
CGRP
4
5
Probe
calcitonin
4
calcitonin
4
CGRP
5
CGRP
5

Figure 4-4.

The effects of hexamer elements on calcitonin/CGRP processing: HeLa cells

RNase protection analyses of total RNA (10 μg) isolated from populations of HeLa cells stably transfected with each of the indicated constructs. Conventions are as in Figure 4-3.

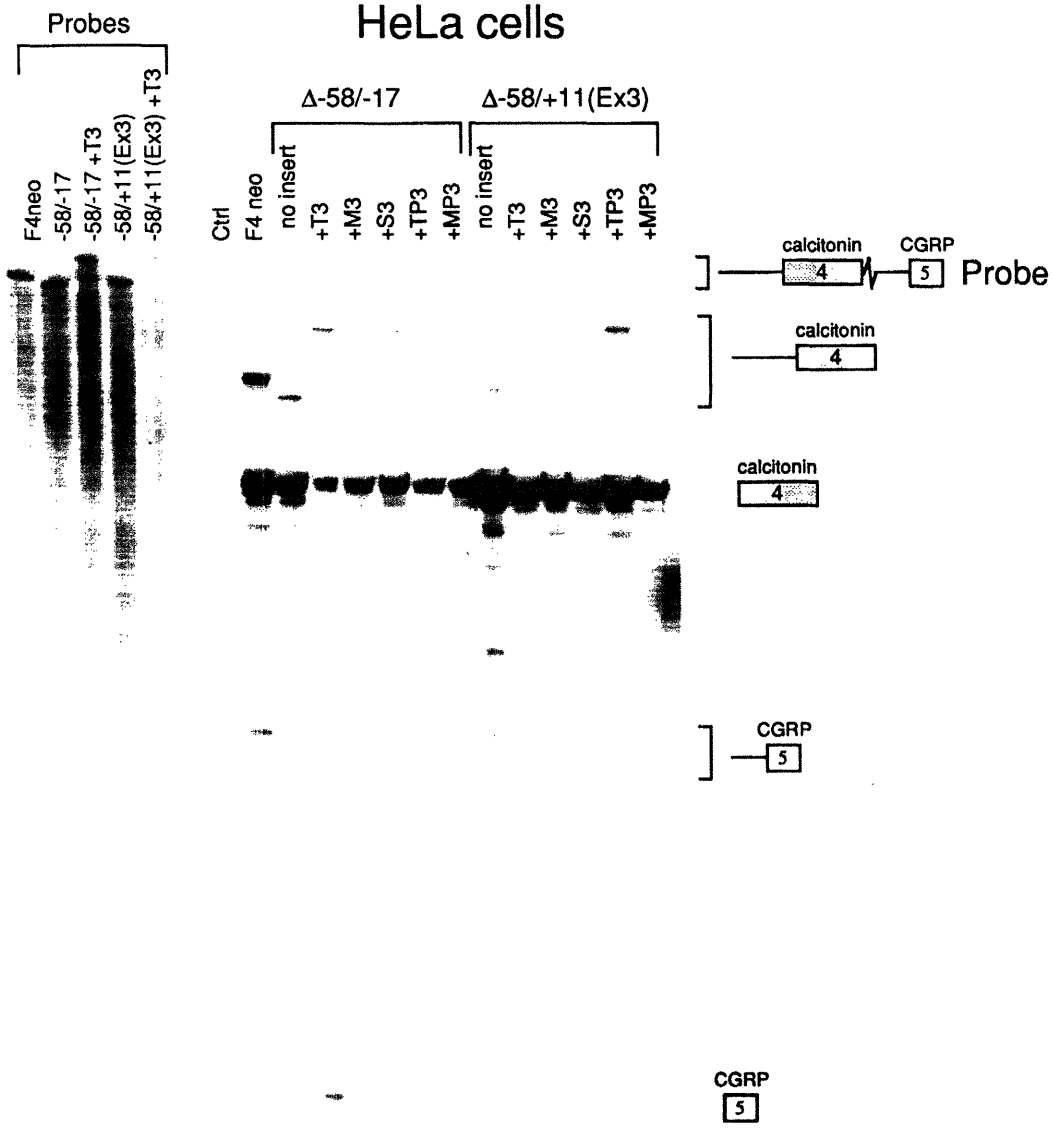

Probes
F4neo
-58/-17
-58/-17 +T3
-58/+11(Ex3)
-58/+11(Ex3) +T3
HeLa cells
Δ-58/-17
Δ-58/+11(Ex3)
Ctrl
F4 neo
no insert
+T3
+M3
+S3
+TP3
+MP3
no insert
+T3
+M3
+S3
+TP3
+MP3
calcitonin
CGRP
4
5
Probe
calcitonin
4
calcitonin
4
CGRP
5
CGRP
5

Figure 4-5.

The effects of hexamer elements on calcitonin/CGRP processing: NIH3T3 cells

RNase protection analyses of total RNA (15 μg) isolated from populations of NIH3T3 cells stably transfected with each of the indicated constructs. Conventions are as in Figure 4-3.

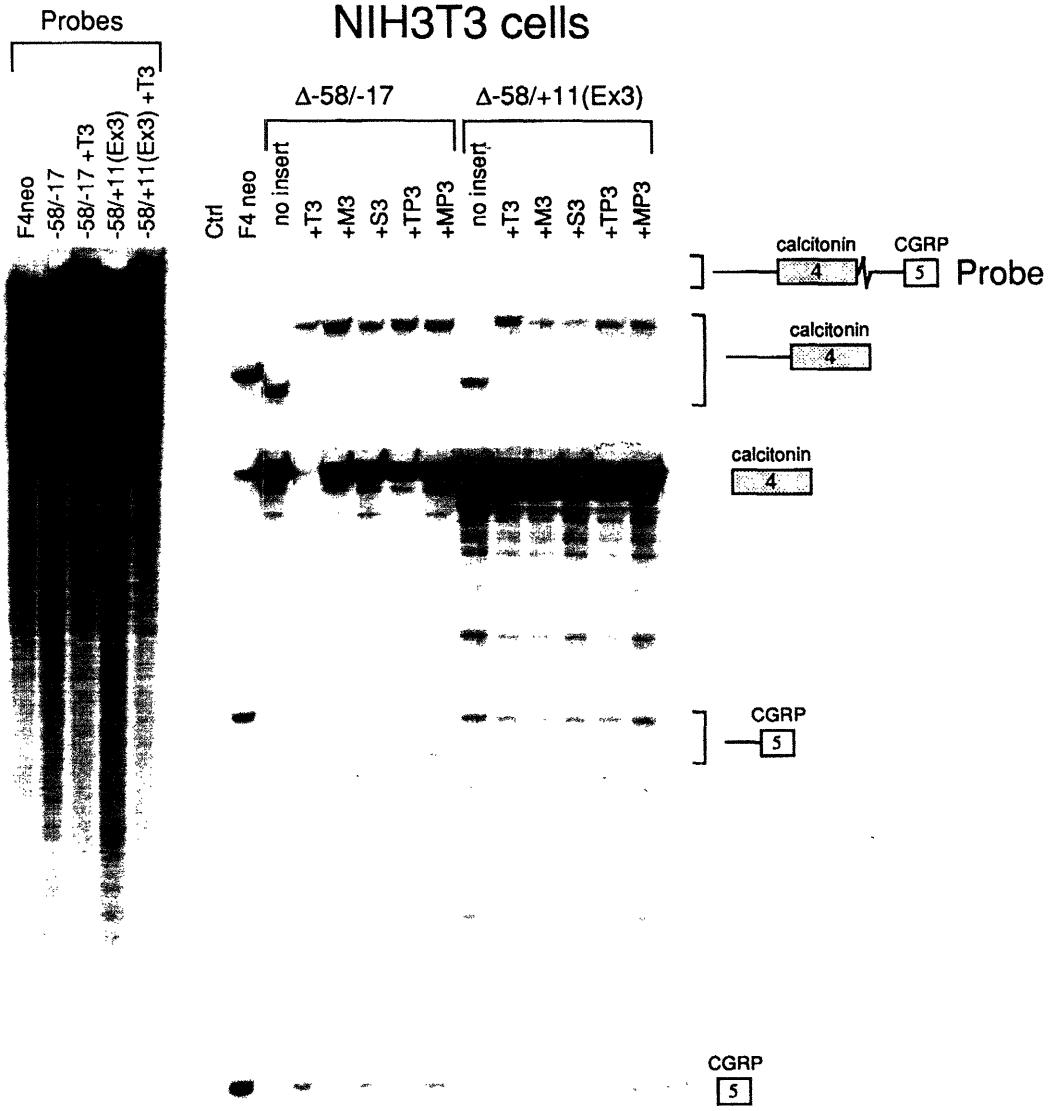

Probes
F4neo
-58/-17
-58/-17 +T3
-58/+11(Ex3)
-58/+11(Ex3) +T3
NIH3T3 cells
Δ-58/-17
Δ-58/+11(Ex3)
Ctrl
F4 neo
no insert
+T3
+M3
+S3
+TP3
+MP3
no insert
+T3
+M3
+S3
+TP3
+MP3
calcitonin
CGRP
4
5
Probe
calcitonin
4
calcitonin
4
CGRP
5
CGRP
5

Figure 4-6.

The effects of hexamer elements on calcitonin/CGRP processing: COS cells

RNase protection analyses of total RNA (5 μg) isolated from populations of COS cells transiently transfected with each of the indicated constructs. Conventions are as in Figure 4-3.

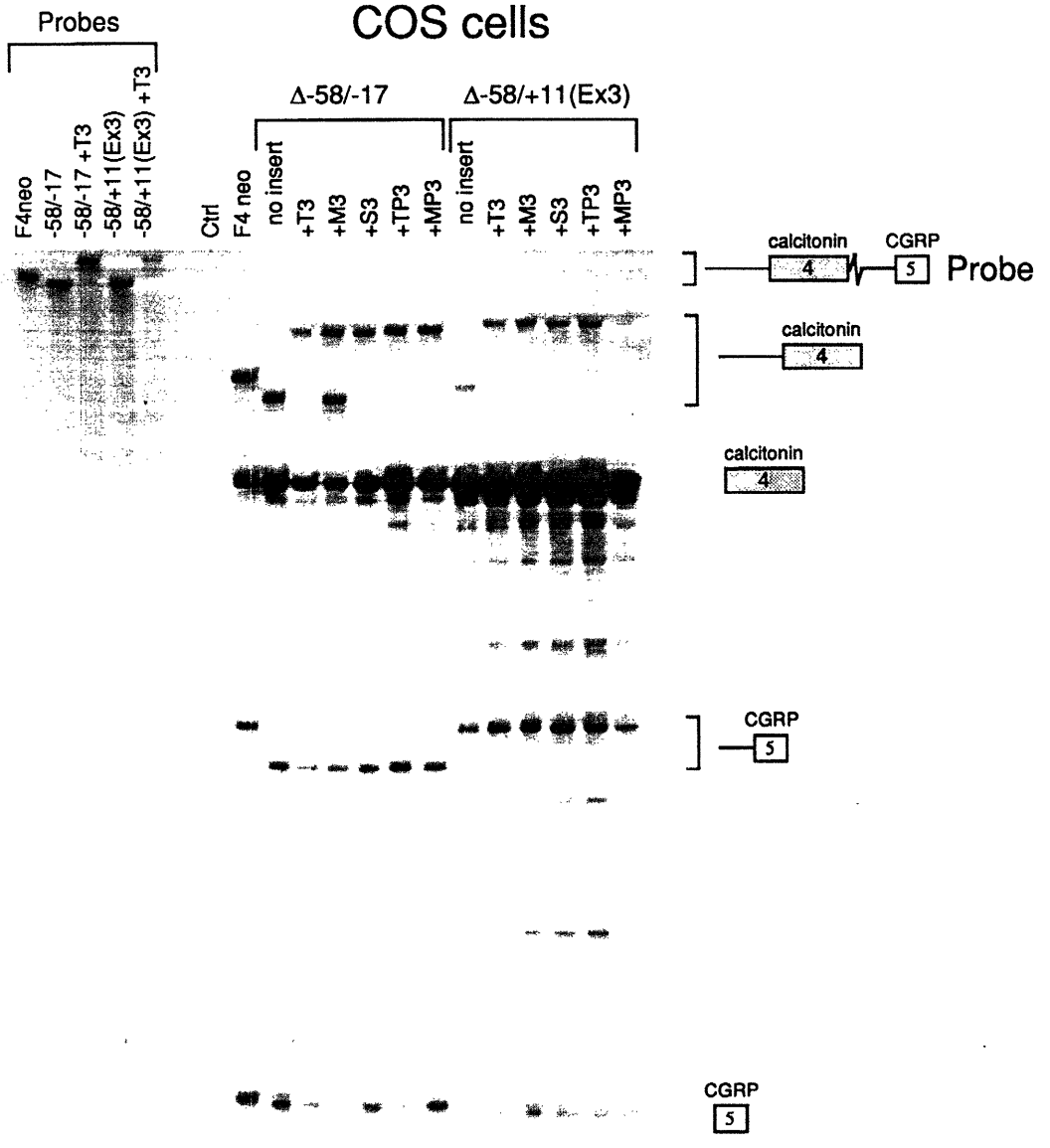
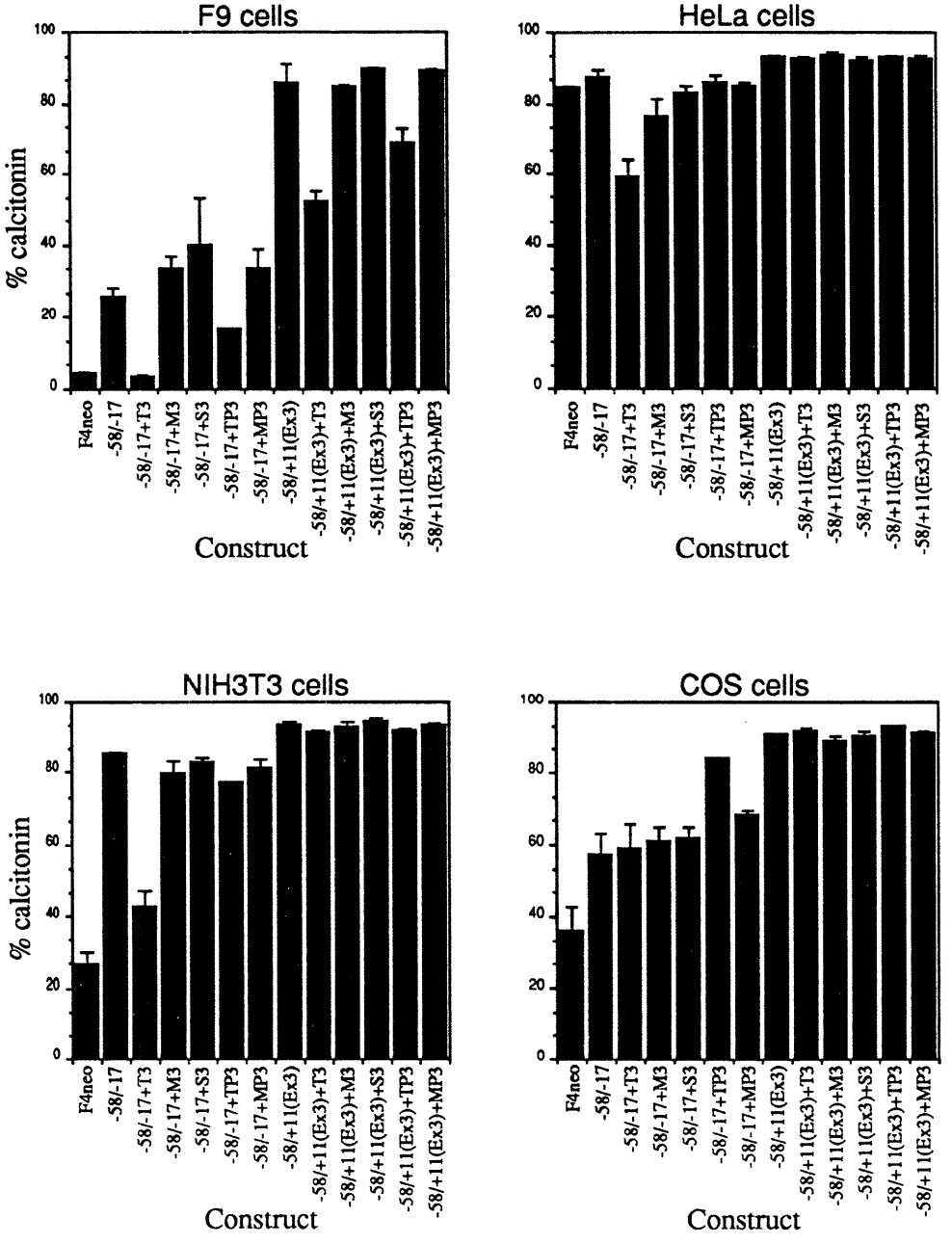

Probes
F4neo
-58/-17
-58/-17 +T3
-58/+11(Ex3)
-58/+11(Ex3) +T3
COS cells
Δ-58/-17
Δ-58/+11(Ex3)
Ctrl
F4 neo
no insert
+T3
+M3
+S3
+TP3
+MP3
no insert
+T3
+M3
+S3
+TP3
+MP3
calcitonin
CGRP
4
5
Probe
calcitonin
4
calcitonin
4
CGRP
5
CGRP
5

Figure 4-7.

Effects of synthetic inserts on calcitonin/CGRP processing

The percentage of calcitonin production was plotted as a function of each construct for each of the four cell lines used in this study. For F9, HeLa and NIH3T3 cells, measurements were obtained from two independently transfected polyclonal G418-resistant populations (>500 colonies each). For COS cells, measurements were obtained from two independent transient transfections. Error bars indicate standard deviations.


F9 cells
HeLa cells
NIH3T3 cells
COS cells
% calcitonin
Construct
F4neo
-58/-17
-58/-17+T3
-58/-17+M3
-58/-17+S3
-58/-17+TP3
-58/-17+MP3
-58/+11(Ex3)
-58/+11(Ex3)+T3
-58/+11(Ex3)+M3
-58/+11(Ex3)+S3
-58/+11(Ex3)+TP3
-58/+11(Ex3)+MP3

number of respects. Expression of the -58/-17 and -58/+11(Ex3) constructs differed little from that of F4neo. In the context of the -58/-17 mutation, insert T3 enhanced CGRP processing significantly, relative to inserts M3 and S3 and the wildtype F4neo construct. The TP3 insert, in contrast, appeared not to have any significant effect relative to controls. In the context of the -58/+11(Ex3) mutation, neither T3 nor TP3 had significant effects relative to controls. Therefore TGCATG inserts appeared to function less well in HeLa cells than in F9 cells. However, since calcitonin processing was always highly preferred in HeLa cells, it was possible that TGCATG elements functioned to similar extents in F9 and HeLa cells but that such effects went undetected in HeLa cells.

TGCATG repeats have intermediate effects in NIH3T3 cells

Expression of F4neo in murine NIH3T3 cells resulted in calcitonin processing at low but detectable levels (Figures 4-5, 4-7). Mutations -58/-17 and -58/+11(Ex3), as expected, significantly shifted splice preference in favor of calcitonin processing. Qualitatively, the effects of synthetic inserts in NIH3T3 cells paralleled those observed in F9 cells, but the magnitude of these effects appeared to be less. TGCATG inserts specifically reduced calcitonin processing in the context of -58/-17, with insert T3 having noticeably greater effects than TP3. In contrast to F9 or HeLa cells (where T3 completely reversed or even overcompensated the effect of the -58/-17 mutation) insert T3 did not reduce calcitonin processing to wildtype levels in NIH3T3 cells. Furthermore, inserts T3 and TP3 appeared to have little if any effect in the 58/+11(Ex3) construct when compared to controls. These results suggested that TGCATG repeats functioned to different extents in different cell lines.

Curiously, the expression in NIH3T3 cells of spliced calcitonin mRNA by either the F4neo or the -58/-17+T3 construct was quite low; in contrast, the

bands representing unspliced RNA were comparable in intensity to those seen in other lanes. For these constructs, a high ratio of unspliced to spliced RNA was reproducibly obtained in analyses of two independently derived, stably transfected cell populations, each of which represented pools of several hundred G418-resistant colonies. This effect of the T3 insert upon -58/-17 expression could be explained by a reduction in splicing efficiency of splicing of this region of the pre-mRNA, thus resulting in an accumulation of unspliced RNA. Such an effect was also apparent to some extent in the wildtype F4neo construct but was not observed with -58/-17 constructs that contained control inserts. These results are consistent with a TGCATG-dependent repression in calcitonin acceptor usage.

TGCATG sequences have little or no effect in COS cells

The transient expression of F4neo in COS cells yielded about 36% calcitonin splicing, a value that was higher than that obtained by others (11%; Delsert and Rosenfeld, 1992). The -58/-17 mutation modestly increased this value (Figures 4-9, 4-10). Whereas -58/-17 increased calcitonin usage from 27% to 85% in NIH3T3 cells, the shift in COS cells was noticeably lower (from 36% to 57%). Therefore, although the levels of calcitonin splicing were comparable between these cell lines, the -58/-17 mutation appeared to have lower effects in COS cells than in NIH3T3 cells. In addition, none of the inserts had any significant negative effect upon calcitonin splicing. In fact, insert TP3 significantly stimulated calcitonin splicing, a surprising finding which we cannot readily explain. In addition, none of these inserts appeared to have any significant effects on the Δ-58/+11(Ex3) mutation. These results suggested that TGCATG elements acted in a cell-type-dependent fashion and have no activity

in COS cells. This was be consistent with the effects of these TGCATG elements in other minigene contexts (Chapter Three).

For technical reasons, COS cell expression could only be carried out transiently; this posed a caveat regarding these interpretations, as measurements in the other cell lines were carried out on stably expressing cell populations. However, since the -58/-17 and -58/+11(Ex3) mutations were observed to have significant positive effects upon calcitonin processing in COS cells, the differential recognition of the exon 4 splice acceptor (versus that of exon 5) still appeared to be rate-limiting even when transient transfection conditions were employed. Therefore it seems unlikely that the TGCATG-independence of calcitonin/CGRP processing in COS cells is simply due to the transfection protocol. We therefore conclude, from the transfection of four cell lines, that TGCATG elements have a negative effect upon the selection of an exon closely downstream, probably by downregulating 3' splice site usage, and that the extent of TGCATG function depends to some degree upon cell type.

DISCUSSION

These experiments have examined the effect of the repeated TGCATG sequence element in the context of an alternative pre-mRNA processing pattern that is quite different from those that exist in fibronectin. The results obtained in these studies suggest that hexamer elements can negatively affect 3' splice site recognition. These results therefore provide insights regarding how TGCATG repeats may control EIIIB splicing. In addition, these findings have implications regarding the nature of the cell-type-specific determinants that normally regulate alternative processing in the calcitonin/CGRP gene.

TGCATG repeats exert negative effects on 3' splice site utilization

In the EIIIB region of the fibronectin gene, TGCATG repeats were found to occur mostly in the 3' portion of the intron downstream of the EIIIB exon (Chapter 3). In fact, several are closer to the 3' splice site of the exon downstream than to EIIIB. One possible model for repeat action posited a TGCATG-mediated delay in the the commitment of the exon III-8a 3' splice site, which might allow EIIIB to compete more efficiently for recognition. If so, then one would predict that these elements should inhibit a heterologous alternative 3' splice site if placed upstream. The results presented here fulfill this prediction and thus may explain the relative positioning of hexamer repeats in the fibronectin gene. TGCATG-containing sequences were found to affect a 3' splice site sequence derived from a constitutive exon (-58/+11[Ex3]) as well as the 3' splice site of exon 4 (-58/-17); therefore the effects of these elements are not limited strictly to alternative 3' splice sites. Although these experiments do not directly prove that TGCATG elements function in EIIIB

splicing via 3' splice site attenuation, they nevertheless strengthen the validity of such a model.

We had previously established that TGCATG repeats were capable of promoting the recognition of an upstream 5' splice site (Chapter Three). It seems unlikely that TGCATG elements perform a similar function in the calcitonin/CGRP gene, since the 5' splice site of exon 3 is utilized regardless of processing choice. It is formally possible that these elements are performing a distinct function from that performed in EIIIB splicing and that they mediate their effects by directly repressing polyadenylation; such a possibility seems relatively remote. Therefore the effects of hexamer elements in the calcitonin/CGRP gene most likely occur via 3' splice site attenuation.

Previous studies of hexamer function in EIIIB splicing could not rule out a role for these elements as transcriptional pause sites. However, the TGCATG-dependent calcitonin exon repression that we observe cannot be easily be explained on the basis of a mechanism involving transcriptional pausing. It is therefore likely that these hexamer elements function at the level of RNA.

Implications for regulation of calcitonin/CGRP RNA processing

As commented on above, the calcitonin/CGRP gene exhibits a mode of differential RNA processing that at first glance is quite distinct from that of EIIIB. Nevertheless, this chapter provides evidence that a short repeated sequence that regulates EIIIB can also control calcitonin/CGRP splicing. Therefore, in analogy to systems in *Drosophila*, a single *cis*-element can regulate multiple genes in vertebrates. There is some evidence for the involvement of at least one TGCATG-containing element in the splicing of the neural-specific N1 exon in *c-src* (see Chapter Three and Black, 1992). However, the case for

TGCATG involvement in calcitonin/CGRP alternative splicing is not as clear-cut.

Unlike the fibronectin gene, the GCATG repeats in the human and rat calcitonin/CGRP genes are not conserved in number or position. In addition, although certain deletions upstream of exon 4 that removed GCATG repeats had a positive effect upon calcitonin-specific processing (Emeson et al., 1989), a number of other mutations made in this area that removed GCATG repeats failed to consistently affect exon 4 usage (Yeakley et al., 1993). In addition, human calcitonin/CGRP minigenes in which a number of these GCATG repeats were removed appeared to exhibit proper regulated processing in HeLa and F9 cells (Cote et al., 1991). Therefore it is unclear whether repeated GCATG sequences contribute to calcitonin exon repression, despite their anomalously high occurrence. Although fewer repeats are present in the rat gene, it is interesting to note that one nonconserved rat repeat occurs between the putative branchpoint and 3' AG of exon 4, an occurrence that may compensate for the lack of repeats elsewhere.

Our findings suggest a potential role of TGCATG-related sequences in the regulation of calcitonin/CGRP alternative processing. An inverse correlation appears to exist between usage of exon 4 in F9 cells and the number of natural/artificial GCATG repeats upstream. However, experiments of the type shown here do not rigorously prove GCATG participation, nor do they formally rule out non-GCATG sequences in the regulation of exon 4. A mutagenesis study across the entire exon 4 splice acceptor region failed to find any single *cis*-active sequence critical for exon 4 repression or activation (Yeakley et al, 1993). While this information may apparently contradict the results obtained by Emeson et al. (1989), the more recent studies did not address whether functional redundancy was involved. Therefore multiple

distinct elements may regulate exon 4. Of the mutations that were tested in these studies, those that activated exon 4 usage had removed or altered sequences between 35 and 52 nt upstream of exon 4, plus at least one GCATG sequence (Yeakley et al., 1993; Emeson et al., 1989). Other short sequences in this region may therfore act in concert with GCATG elements. Regulation by multiple elements have been implicated in the regulation of alternatively spliced exons such as EIIIB (Chapter Three), the N1 exon in *c-src* (Black, 1992) and the female-specific exon in the *dsx* gene (Burtis and Baker, 1989; Hoshijima et al., 1991; Hedley and Maniatis, 1991; Ryner and Baker, 1991). The purine-rich exon recognition sequence, which has been identified in a number of constitutive and alternative exons, also appears to have greater activity if present in multiple copies (Watakabe, et al., 1993; Xu et al., 1993).

TGCATG repeats exhibit context-dependent cell-type-specific effects

In Chapter Three, the activity of the synthetic TGCATG insert was found to be greatest in F9 and HeLa cells, low or nonexistent in COS cells and intermediate in NIH3T3 cells in the rat proprotachykinin gene. With respect to exon 4 repression in the context of the Δ-58/-17 mutation a similar correlation among cell types was observed, although interpretation was limited by the different calcitonin preferences normally exhibited by each cell type. The results presented here are nevertheless consistent with cell-type-specific hexamer element action and therefore corroborate conclusions made in previous chapters.

Parenthetically, if TGCATG sequences were the only cell-type-specific determinant in calcitonin/CGRP processing then one would predict that COS cells should only produce calcitonin, based upon our previous studies. We note that the levels of calcitonin production in COS cells are quite detectable;

therefore it is probable that elements other than TGCATG contribute to calcitonin/CGRP cell-type-specificity.

It was surprising to note that TGCATG sequences worked less well in the context of purine-rich spacers than in the context of pyrimidine-rich spacers. Given that the opposite correlation was observed in the EIIIB system, this was surprising. It is possible that sequences in the vicinity of the insert site interact with the spacer sequences to antagonize or promote hexamer activity. In any event, the basis for this context effect, whether direct or indirect, remains a mystery.

TGCATG is a multifunctional element

It has already been pointed out that the TGCATG sequence occurs within elements that control RNA splicing in the murine *c-src* gene; in fact, experiments that mutated these *c-src* elements strongly implicate the importance of this sequence in the positive regulation of the N1 exon (Black, 1992). We have already established the importance of TGCATG sequences in EIIIB regulation in the rat fibronectin gene and have provided evidence that TGCATG repeats can promote the usage of an alternative 5' splice site (Chapter Three). The results presented in this chapter extend these previous findings by demonstrating an influence upon alternative 3' exon selection. These data also raise the possibiity that calcitonin/CGRP processing is normally regulated by TGCATG-related elements, although this remains to be conslusively established. Therefore TGCATG-related elements may function in three different alternatively spliced genes. This situation is not unlike regulated splicing systems in *Drosophila*, where a single regulator may control multiple genes. For example, the Sex-lethal protein controls both *tra* and *Sxl* splicing, whereas the Tra-2 protein controls both the splicing of *dsx* and its own *tra*-2

pre-mRNA (Sosnowski et al., 1989; Horabin and Schedl, 1993; Sakamoto et al., 1992; Tian and Maniatis, 1992; Mattox and Baker, 1992). It will therefore be interesting to characterize and identify the factors which interact with this hexamer sequence, since such factors may be involved in the control of multiple alternative splicing events.

MATERIALS AND METHODS

Computer and calcitonin/CGRP sequence analysis

The human calcitonin/α-CGRP sequence was obtained from GenBank (accession number X15943; Broad et al., 1989). Information regarding the locations of GCATG repeats in the corresponding gene in rat was kindly provided by J. Yeakley and M. G. Rosenfeld (UCSD, La Jolla; unpublished data). Analysis of repeats was carried out as described previously (Chapter Three).

DNA Constructs and Probes

Standard DNA manipulation techniques were as described in Chapters Two and Three. The plasmid F4neo contains the complete calcitonin/CGRP transcription unit (under the control of its own promoter) cloned into pSV2neo (Emeson et al., 1989). The deletion and substitution mutations -58/-17 and -58/+11(Ex3), also described by Emeson et al. (1989), were derived from a variant of F4neo, called Kpn/Cla; this derivative of F4neo contains KpnI and ClaI sites at the exon 4 splice acceptor. The sequences of the exon 4 splice acceptor region for each of these constructs is shown in Figure 4-2. F4neo, -58/-17, -58/+11(Ex3) and the F4neo RNase probe plasmids were kindly provided by R. Emeson (Vanderbilt, Nashville), J. Yeakley and M. G. Rosenfeld (UCSD, La Jolla).

7iBi89ΔGA/pBAGH construct derivatives that contained each of the inserts T3, M3, S3, TP3, and MP3 (Figure 4-2) were cleaved with AflII, blunted with Klenow fragment and cloned into BamHI-cut/blunted pBluescript SK^- in the T7 orientation. Each subclone was cut with SmaI and HincII, recircularized with T4 DNA ligase (thus deleting EcoRI and HindIII sites from the

polylinker), recut with NotI, blunted and then recut with KpnI; insert fragments were gel-purified and ligated with two vector fragments derived either from -58/-17 or -58/+11(Ex3) mutant constructs. One vector fragment was made by cutting the relevant F4neo mutant construct with KpnI, blunting, recutting with XbaI and isolating the smaller of the two bands generated; the second vector fragment was obtained by cutting with KpnI and XbaI and isolating the larger fragment. Insert orientation and junctions were verified by DNA sequencing of transformant minipreps.

The RNase protection probe used for analyzing expression of these minigenes contains two fragments from the calcitonin/CGRP gene cloned into pBluescript II (Stratagene). The first fragment contains the 626 nucleotides surrounding the exon 4 splice acceptor (154 bases upstream, 472 bases downstream); the second fragment contains the 300 nucleotides surrounding the exon 5 splice acceptor (117 bases upstream, 183 bases downstream). Since each of the above F4neo-derived constructs contained alterations immediately upstream of exon 4, analysis of each construct required the synthesis of RNase probes containing the appropriate sequence changes. To make each RNase probe, the polymerase chain reaction (PCR) was employed to generate the appropriate exon 4 region, using each of the expression plasmids as templates. The primers for PCR corresponded to the termini of the exon 4 splice acceptor probe fragment (upstream primer, 5' GCCGAAGCTTCCCTTCCCCCACACTTTTCTGG 3'; downstream primer, 5' GCCGGAATTCGTCCTTTTAGAAAAATAGTTT 3'; kindly provided by J. Yeakley). The upstream and downstream primers contained HindIII and EcoRI sites (underlined) respectively; PCR products were cleaved with HindIII and EcoRI and inserted into the RNase probe plasmid via these sites.

Transfections and RNA analysis

Stable transfections of F9, 3T3 andHeLa cells, transient transfections of COS cells and RNA isolation procedures were carried out as described in Chapter 2, except that co-transfection with pSV2neo was omitted and all RNAs (regardless of transfection protocol) were treated with DNase prior to analysis. RNase protection probe templates were linearized with HindIII and transcribed with T7 RNA polymerase; radiolabeled riboprobes were purified via preparative 4% denaturing gel electrophoresis before use. RNase protection assays were carried out exactly as described by Emeson et al. (1989); RNase-resistant fragments were separated on a 4% denaturing polyacrylamide gel. Samples in 90% formamide loading buffer were heated to 85°C for at least 10 minutes prior to loading the gel. Bands were quantitated using a Molecular Dynamics Phosphorimager. Proportions of calcitonin/CGRP processing were normalized for labeled uridine content (CGRP/calcitonin uridine ratio of 0.347).

ACKNOWLEDGEMENTS

We thank Jo Yeakley, Ron Emeson and Geof Rosenfeld for kindly providing the rat calcitonin/CGRP gene-related materials for the experiments described in this chapter and for many stimulating discussions over the phone.

Chapter Five

Biochemical Detection of UGCAUG-specific RNA-Binding Factors

ABSTRACT

Previous transfection studies have shown that a repeated hexamer sequence acts in a cell-type-specific manner to regulate alternative splicing of the EIIIB exon in the rat fibronectin gene. This repeated sequence is also capable of regulating alternative splicing when placed in the rat preprotachykinin and calcitonin/CGRP genes. Using a UV-crosslinking assay in F9-derived nuclear extracts, at least two factors (ca. 63 kDa and 52 kDa) can be shown to interact with UGCAUG-containing RNAs. Binding of these factors is specific and saturable. The binding of the 63 and 52K components is relatively slow and temperature-dependent. Dissociation is also relatively slow. These binding studies have also detected a factor (55 kDa) that specifically binds to the related sequence UGACUG. The sequence specificity of the 63K and 52K RNA-binding components suggests that they may be involved in the hexamer-dependent regulation of alternative splicing in the fibronectin gene.

INTRODUCTION

It is clear that in many cases of cell-type-specific alternative splicing, the primary RNA transcript is essentially identical in structure between cell types. Only a few cases exist where the processing decision is based strictly on *cis*-information alone (for example, see Gallego and Nadal-Ginard, 1990). Therefore cell-type-specific or developmentally regulated patterns of alternative splicing must be governed primarily by *trans*-acting influences unique for each cell type or developmental stage.

A detailed analysis of trans-acting influences upon alternative splicing has been most effectively carried out in *Drosophila melanogaster*, owing in part to the genetic identification of splicing regulators that determine sex (reviewed by Baker, 1989; Maniatis, 1991; Mattox et al., 1992; McKeown, 1992). These regulators act via specific interactions with sequences present on their target pre-mRNAs. The *Sex-lethal* gene product (*Sxl*) interacts with a uridine-rich stretch present both at the 3'-splice site of a male-specific exon in its own pre-mRNA and at the male-specific 3'-splice site of the *transformer* (*tra*) pre-mRNA, thus acting to prevent male-specific splice site usage in females (Sosnowski et al., 1989; Inoue et al., 1990; Valcarcel et al., 1993). The *tra* and *tra-2* (*transformer-2*) gene products interact with several 13-mer repeats in the female-specific exon of the *doublesex* (*dsx*) pre-mRNA and in doing so activate the usage of an upstream 3'-splice site in female flies (Hedley and Maniatis, 1991; Inoue et al., 1992; Ryner and Baker, 1991; Hoshijima et al., 1991; Tian and Maniatis, 1992). These regulators of splicing are therefore RNA-binding proteins that exhibit specificity for *cis*-elements present within their targets.

In contrast, the study of alternatively spliced vertebrate genes has proceeded primarily via mutational identification of *cis*-elements that affect

splice site selection. These elements have then been used as biochemical probes in an effort to identify sequence-specific RNA-binding factors in nuclear extracts. For example, an element that negatively affects exon 7 usage in the rat β-tropomyosin gene (Helfman et al., 1990; Guo et al., 1991; Guo and Helfman, 1993) appears to specifically interact with PTB (Garcia-Blanco et al., 1989; Patton et al., 1991; Mulligan et al., 1992). Intron sequences that negatively affect exon 4 usage in the calcitonin/CGRP gene (Emeson et al., 1989) have been shown to interact with two proteins in rat brain (41 kDa and 43 kDa; Roesser et al, 1993). Exon sequences that are required for efficient 3'-splice site usage *in vitro* in the human fibronectin ED1 exon and the bovine growth hormone terminal exon have been shown to interact with members of the SR protein family (Lavigueur et al., 1993; Sun et al., 1993b; Zahler et al., 1992). An exon sequence near the 5' splice site of the *Drosophila* P element third intron has also been identified as binding a multiprotein complex (Siebel et al., 1992). These approaches have permitted the identification of specific RNA-binding proteins that may regulate splicing.

A different approach to addressing regulation in mammals can be exemplified by studies of SV40 early pre-mRNA alternative splicing *in vitro*. An alternative splicing factor (ASF), isolated by its ability to promote proximal donor selection, was found to be identical to the splicing factor SF2, a member of the SR protein family (Ge and Manley, 1990, 1991; Krainer et al, 1991). SR family members are all capable of complementing a cytoplasmic S100 extract for splicing, suggesting that they share a role as general splicing factors (Zahler et al., 1992, 1993b). These findings suggest that splicing factors with overlapping functions may regulate some alternative splicing events. The action of ASF/SF2 on splice site selection can be antagonized *in vitro* by hnRNP A1 (Mayeda and Krainer, 1992), an RNA-binding protein that exhibits high

affinity for certain RNA sequences (Burd and Dreyfuss, 1994); in addition, different SR proteins have been shown to promote 5'-splice site selection in a substrate-dependent manner *in vitro* (Zahler et al., 1993a).

These and other studies have suggested that, given a sequence known to be involved in alternative splicing control, it may be possible to identify *trans*-acting factors that regulate alternative splicing in a cell-type-specific manner. Previous chapters have characterized the action of a repeated hexanucleotide sequence element that promotes the alternative inclusion of the EIIIB exon in the rat fibronectin gene. Repeated copies of this element are capable of promoting 5' splice site usage and in addition can function in a heterologous context in a cell-type-specific manner (Chapter Three). Furthermore, repeated hexamers are capable of negatively regulating alternative 3' splice site selection when placed in the rat calcitonin/CGRP gene and may in fact normally regulate the processing of this pre-mRNA. This chapter describes the identification of at least two RNA-binding factors in F9 teratocarcinoma cell nuclear extracts that exhibit specificity for the RNA sequence UGCAUG. These factors exhibit approximate molecular weights of 63 and 52 kilodaltons on an SDS-polyacrylamide gel. The identification of these factors may contribute to an understanding of the mechanism by which the UGCAUG hexanucleotide element controls alternative splicing.

RESULTS

In order to identify RNA-binding factors specific for UGCAUG sequences, nuclear extracts were prepared from a line of F9 cells capable of growing in suspension. F9att5.51 is a variant which is adhesion-defective and as a result grows in loose aggregates in suspension (Grover et al., 1987). Levels of fibronectin expression by F9att5.51 are somewhat higher than those expressed by normal F9 teratocarcinoma cells; however, EIIIB inclusion is identical to those expressed by normal F9 cells (Grover et al., 1987, and data not shown).

In previous studies (Chapters Two and Three), a number of oligonucleotide inserts were established to have a cell-type-specific effect upon alternative splicing. These DNA inserts, called T3 and TP3 (Figure 5-1), contained six repeats of the sequence TGCATG. Inserts T3 and TP3 differed in the content of the sequences separating each of the TGCATG repeats. Whereas T3 contained spacers that were slightly pyrimidine-rich, TP3 contained spacer sequences that consisted almost entirely of purines. Transcripts generated from the T3 and TP3 templates shared the repeated hexamer sequences (actually an alternating array of UGCAUGA and UGCAUGC heptamers), a GAUC at the 3' terminus of each 35-mer oligo unit and upstream polylinker sequences transcribed by T3 RNA polymerase.

As controls, a number of inserts that contained small changes relative to T3 or TP3 were used as control templates. M3 and S3 were identical to T3 except for the repeated hexamer sequences (Figure 5-1): whereas T3 (test) transcripts contained UGCAUG hexamers, transcripts M3 (mutant) and S3 (scrambled) contained UGACUG and AGTCGT hexamers respectively (underlined nucleotides indicate changes from the test hexamer). As a control

for TP3, template MP3 was used; MP3, like M3, contains UGACUG sequences but contains purine-rich spacers identical to those of TP3.

A UV-crosslinking assay was employed in order to detect factors that interacted specifically with the RNA sequence UGCAUG. In a typical assay, radiolabeled probes were incubated with nuclear extract at 30°C for 20 minutes and then irradiated with UV light for 15 minutes on ice. Samples were then digested with RNase T1 and subsequently analyzed by SDS polyacrylamide gel electrophoresis. Proteins or other components that were crosslinked to the RNA probe would carry label released by the RNase treatment. RNase T1 (rather than RNase A) was used for a number of reasons. Probes containing purine-rich spacers (like TP3 and MP3) might not be digested as efficiently by RNase A, which normally cleaves 3' of pyrimidines; RNase T1 cuts 3' of guanosines and therefore should cleave all spacer sequences at least once. In addition, the use of RNase A failed to yield any detectable signals specific for UGCAUG sequences, possibly because RNase A may have digested labelled nucleotides away from crosslinked protein (not shown). RNase T1 treatment of these RNA probes, on the other hand, would generate fewer oligonucleotides, most of which would contain at least one labeled nucleotide; thus T1 digestion would be less likely to release labelled nucleotides from crosslinked material.

UV-crosslinking assays were carried out using transcripts T3, M3, S3, TP3 and MP3 as probes. Figure 5-2 shows a typical crosslinking profile on a 9% SDS-polyacrylamide gel. A number of bands appeared to be specific for subsets of these probes. In particular, one band appeared in reactions containing probes T3 and TP3 that was absent from reactions that contained M3, MP3 or S3 probes. This band migrated with a molecular weight of approximately 63 kDa (marked as 63K[TP] in Figure 5-2) and was not

Figure 5-1.

UV-crosslinking RNA probes and competitors

Only the monomeric unit of each probe is shown; templates actually consisted of three copies of each unit (indicated here by "x 3" at right). Note that each monomeric unit contains two each of the heptamers UGCAUGA and UGCAUGC; thus each RNA transcript contained a total of six hexamer elements. Shaded bases indicate differences from the T3 RNA (top line).

Probe/Competitor RNA Sequences

T3 [GAUCC UGCAUG ACUACUGCU UGCAUG CUCGUCAUA] x 3

M3 [GAUCC UGACUG ACUACUGCU UGACUG CUCGUCAUA] x 3

S3 [GAUCC AGUCGU ACUACUGCU AGUCGU CUCGUCAUA] x 3

TP3 [GAUCA UGCAUG AGGGAAAGG UGCAUG CAAAGGGAA] x 3

MP3 [GAUCA UGACUG AGGGAAAGG UGACUG CAAAGGGAA] x 3

Figure 5-2.

UV crosslinking assays of RNA probes in F9att5.51 nuclear extracts

The indicated probes ("probe"; T, M, S, TP and MP represent probes T3, M3, S3, TP3, and MP3 respectively) were incubated with either of two nuclear extract preparations (F9-1, F9-2) for 20 minutes at 30°C and either left on ice (-UV) or irradiated with UV light on ice for 15 minutes (+UV). RNase T1 was then added for the indicated times at 30°C ("RNase T1 time"). The samples were analyzed on a 9% SDS-polyacrylamide gel along with prestained moecular weight markers (indicated by small type at right). The positions of the bands representing the T3- and TP3-specific 63 and 52 kDa adducts are indicated at right ("63[TP]" and "52[TP]" in large type). A band that migrates at 52 kDa is also present in the S3 lanes. The 55 kDa adduct specific for the M3 and MP3 probes is also indicated ("55[MP]").

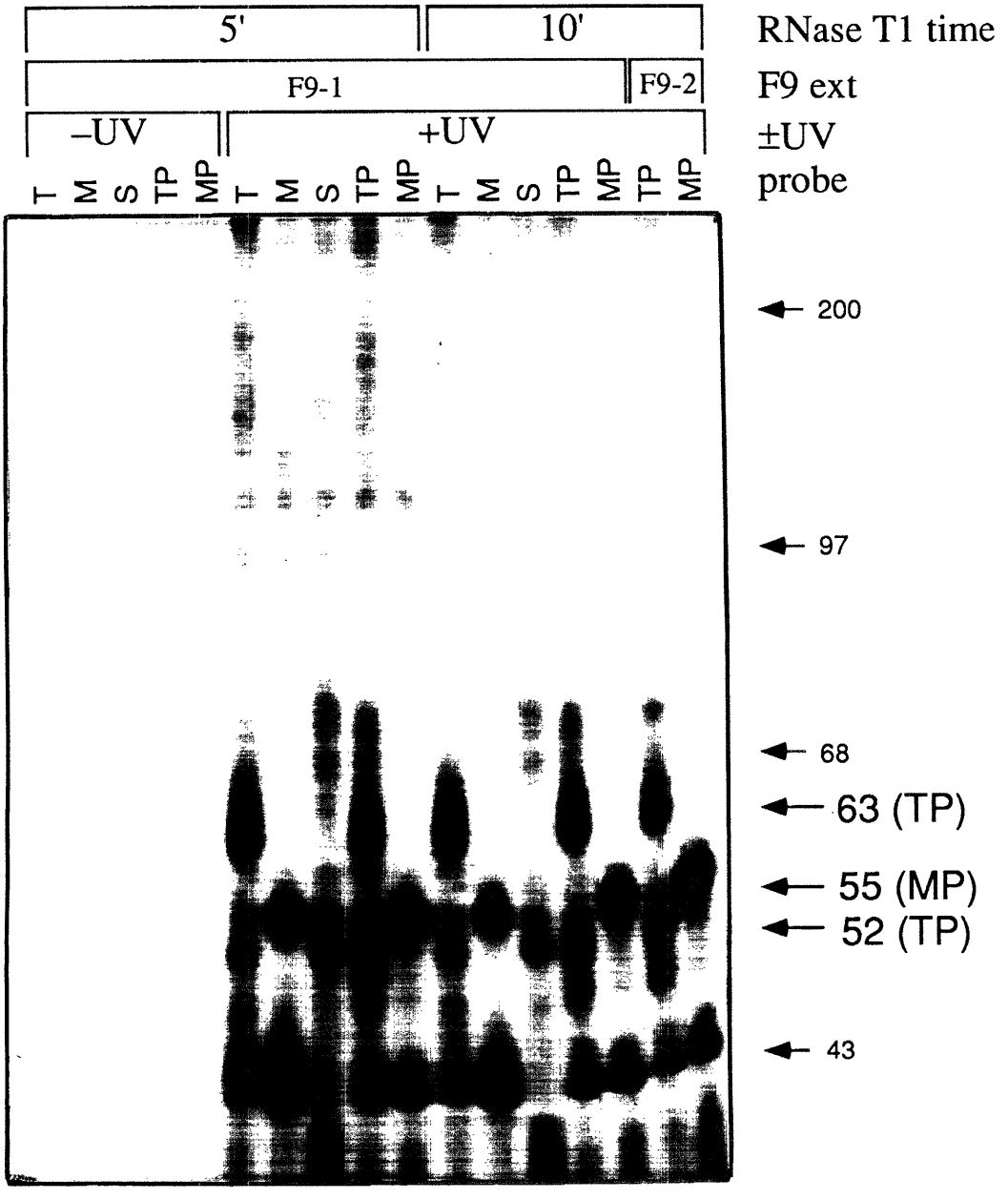

5'
10'
RNase T1 time
F9-1
F9-2
F9 ext
–UV
+UV
±UV
T M S TP MP T M S TP MP TP MP
probe
200
97
68
63 (TP)
55 (MP)
52 (TP)
43

Figure 5-3.

63 kDa, 52 kDa and 55kDa factors are sequence-specific

Either of probes TP3 ("TP") or MP3 ("MP") were incubated in F9 extract for 20 minutes at 30°C, irradiated for 15 minutes on ice, digested with RNase T1 for 10 minutes at 30°C and processed as in Figure 5-2. Probes were coincubated 0, 10 or 100 ng (0, 40 and 400-fold molar excess) of the indicated unlabeled RNAs ("cptr"). Figure conventions are as in Figure 5-2.

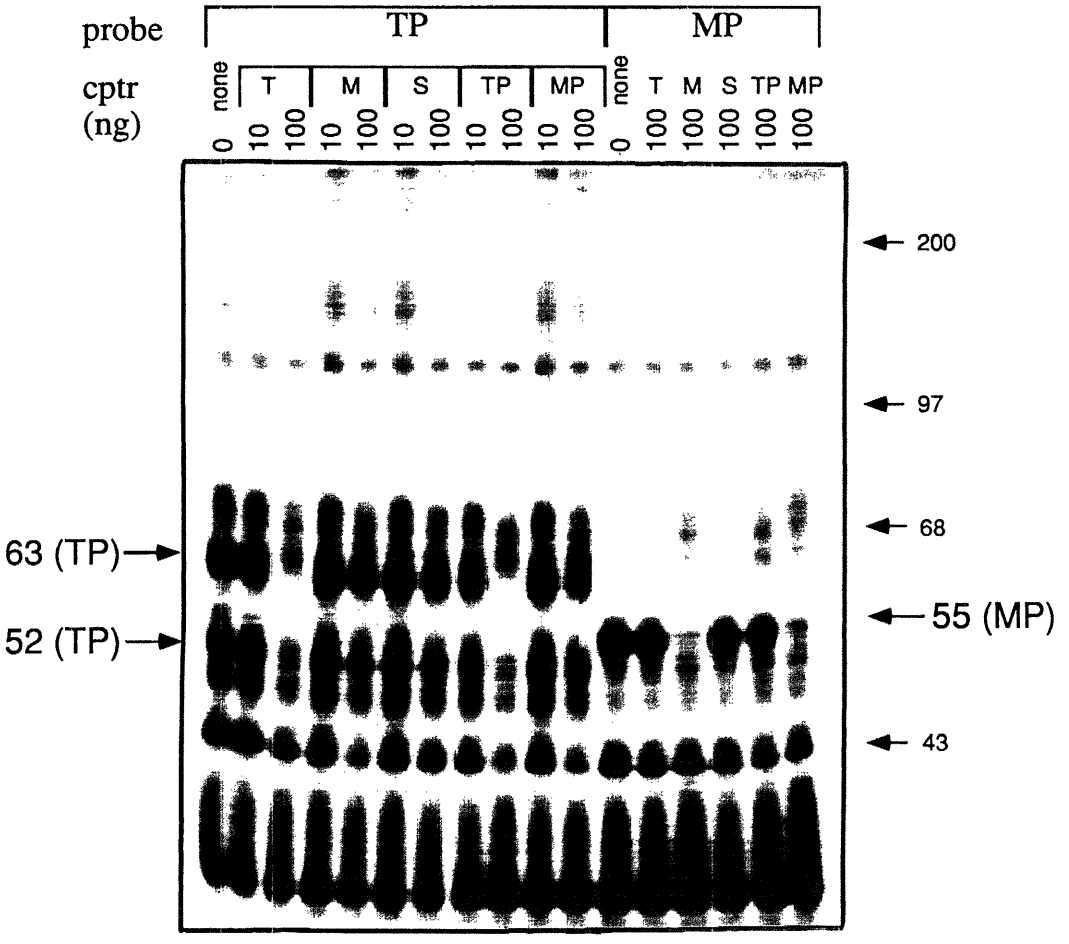
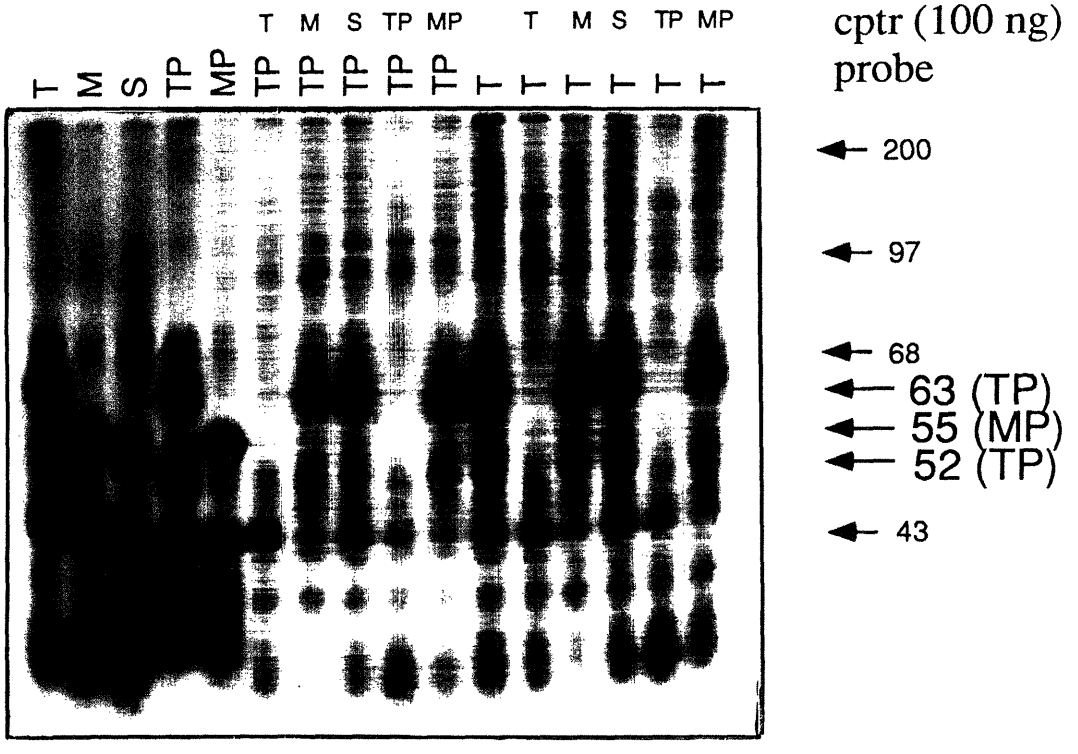

probe
TP
MP
cptr
(ng)
none
T
M
S
TP
MP
0
10
100
200
97
68
55 (MP)
43
63 (TP)
52 (TP)

Figure 5-4.

63 kDa and 52 kDa factors are specific for UGCAUG-containing RNAs

The indicated probes (line "probe") were assayed as in Figure 5-3, either alone or with 100 ng (400-fold molar excess) of the indicated competitors (line "cptr"). A 12% gel is shown. Other figure conventions are as in Figure 5-3.


T M S TP MP T M S TP MP T M S TP MP
cptr (100 ng)
probe
200
97
68
63 (TP)
55 (MP)
52 (TP)
43

significantly affected by the presence of purine-rich or pyrimidine-rich spacer sequences. In addition, a band migrating at the 52 kDa position (marked as 52K[TP] in Figure 5-2) appeared in the T3, TP3 and S3 lanes but not in reactions containing M3 or MP3. These bands migrated relatively heterogeneously and will be referred to collectively as the 63kDa and 52kDa factors, even though they may represent separate species. All bands were UV-dependent and did not reflect partially digested probe RNA, since omission of the UV irradiation step resulted in no signals anywhere on the gel. Therefore a 63 kDa component in F9 nuclear extracts appears to interact specifically with RNAs that contain the UGCAUG hexanucleotide sequence. This binding was ATP- and Mg^{2+}-independent; in addition, binding was reproducible between extract preparations (compare F9-1 and F9-2 lane sets, Figure 5-2) .

In addition, the M3 and MP3 probes were found to react specifically with a 55 kDa factor (marked as 55K[MP] in Figure 5-2) that apparently did not interact with probes T3, S3 or TP3. Since M3 and MP3 probes contained the control sequence UGACUG, it was somewhat surprising to discover components that bound this sequence with specificity. In any event, we conclude that a number of RNA-binding factors in nuclear extracts possessed affinities for UGCAUG- and UGACUG-containing RNAs.

The 63 and 52 kDa RNA-binding factors are sequence-specific

It was possible that UGCAUG-specific (63 kDa) and UGACUG-specific (55 kDa) bands reflected binding of the same factor but that differential RNase T1 digestion caused these bands to migrate differently. in addition, the T3- and TP3-specific 63 kDa bands may represent distinct factors that comigrated; this was also possible for the T3/TP3/S3-specific 52 kDa band. To address these possibilities, competition experiments were carried out using unlabeled

competitor RNAs. Unlabeled competitors were coincubated with labeled TP3 and MP3 probe RNAs and nuclear extract. Either 10 or 100 ng of competitor RNA (approximately 40- and 400-fold molar excess respectively) was coincubated with the probes as indicated in Figure 5-3. Because the structures of probes TP3 and MP3 allowed preferential incorporation of radiolabeled uridine at the hexamer repeats, these probes were used for several of the subsequent assays shown here.

As shown in Figure 5-3, the 63 kDa band specific for TP3 was modestly reduced when incubated with 10 ng of TP3 competitor and was abolished when 100 ng of TP3 competitor was used. The same was true when T3 was used as competitor. In contrast, the 63 kDa band was not affected by 10 ng of any of the control competitors MP3, M3 or S3. This band was only slightly reduced by 100 ng of these control competitor RNAs. As judged from several competition experiments, the 63 kDa component exhibited at least a ten-fold preference for UGCAUG sequences over control hexamer sequences. This number may in fact represent an underestimate of binding specificity, as the addition of competitors in amounts greater than 100 ng often resulted in an overall reduction in the crosslinking profile. Therefore the 63 kDa nuclear extract component specifically interacted with RNAs that contain the sequence UGCAUG and in addition is distinct from the M3/MP3-specific 55 kDa factor.

The 52 kDa band specific for TP3 was also reduced by coincubation with cold T3 and TP3 transcripts but not by any other competitor. This was slightly harder to visualize, as the background in the region of the gel is slightly higher than in the 60-70 kDa region. Importantly, the TP3-specific 52 kDa component is not significantly competed by the cold S3 competitor. Therefore the TP3-specific and S3-specific 52 kDa bands may represent distinct components that happened to comigrate by chance.

These data left open a formal possibility that the T3- and/or TP3-specific 63/52 kDa factors were relatively nonspecific RNA-binding proteins that were present in limiting amounts in the extract and were not seen in the M3 and MP3 reactions due to their displacement from the hexamer sites by the M3- and MP3-specific 55 kDa factor. The fact that S3 did not sequester these factors rendered this possibillty unlikely. Nevertheless, this was tested by competition of the MP3-specific 55 kDa band. If the above were true, then the addition of 100 ng of MP3 or M3 competitor to a reaction containing the MP3 probe should result in a reduction of the 55 kDa band plus a concomitant increase in 63 or 52 kDa intensity. As shown in Figure 5-3, the MP3-specific 55 kDa was specifically competed by coincubation with 100 ng of cold MP3 or M3 RNA but not by 100 ng of the other transcripts. More importantly, a 63 or 52 kDa band did not appear upon addition of 100 ng of either M3 or MP3 competitor (Figure 5-3). These data confirmed that the 63 kDa and 52 kDa factors bind UGCAUG sequences specifically.

Further support for the binding specificity of the 63 and 52 kDa factors is provided in Figure 5-4, which shows a 12% gel showing the binding profiles of all five probes. Competition profiles are shown using either T3 or TP3 as labeled probes. Competition of T3-specific bands by 100 ng of each cold competitor RNA yielded results that were essentially identical to those obtained using the TP3 probes, reinforcing further the specificity of the 63 and 52 kDa factors.

Parenthetically, the results presented in Figures 5-2 through 5-4 suggested that a single 55 kDa factor binds both the MP3 and M3 transcripts and is specific for the control sequence UGACUG. Although this factor was not of any immediate interest, it was useful to compare the data involving the

55 kDa factor with that obtained for the 63/52 kDa factors in subsequent experiments.

Factor dissociation is slow

Since other splicing regulators are capable of assembling into a stable complex (e.g., Tian and Maniatis, 1993), it was of interest to ask whether UGCAUG binding was a stable interaction. To address this possibility, TP3 probe was incubated with extract for various times; 100 ng of unlabeled TP3 or MP3 competitor was added after these times, followed by continued incubation for 10 minutes. The TP3-specific 63/52 kDa signals were compared with those obtained by an uninterrupted incubation (Figure 5-5, left half of gel). Binding was maximal after 15 minutes. The addition of TP3 competitor after 20 minutes, followed by a subsequent 10 minute incubation, resulted in a signal that was reduced but still significant if compared to control lanes in which no competitor (or MP3 competitor) was added. Therefore the TP3-factor interaction, once formed, was only partially resistant to addition of specific competitor, suggesting either that overall rates of dissociation were slow or that a fraction of the complexes formed were stable. Interestingly, the experiments utilizing MP3 as a probe yielded similar results (Figure 5-5, right half of gel).

Binding is slow and temperature-dependent

In order to examine the kinetics of factor association more closely, a time course was carried out. Binding was carried out using TP3 or MP3 as a probe. After incubation at 30°C for 0, 5, 10, 15 and 20 minutes, samples were put on ice and irradiated for 15 minutes and subsequently processed identically. As shown in Figure 5-6, the 63 and 52 kDa bands gradually appeared over the

course of several minutes, nearing completion after 10 minutes. The appearance of these bands was also dependent upon incubation at 30°C, since these signals did not appear if upon incubation at 0°C. 55kDa binding to MP3 probe was more rapid but still temperature-dependent (Figure 5-8). Therefore the interaction of these factors with probe RNAs takes place over a period of minutes. One possible reason for this is that binding of these factors is intrinsically slow; alternatively, association may normally be rapid but is inhibited by other extract components.

UGCAUG-specific binding is relatively stable

To explore the kinetics of binding and dissociation more closely, incubation of extract with the TP3 or MP3 probes was carried out for various times; 100 ng of unlabeled TP3 or MP3 competitor RNA was then added and the incubation continued until a total incubation time of 20 minutes was reached. These binding profiles were then compared with an uninterrupted time course.

Addition of TP3 competitor at time 0, as expected, yielded no significant signal (Figure 5-6). Conversely, TP3-specific signals were not reduced if TP3 competitor was added at the end of the incubation time. Addition of TP3 competitor after 5, 10 or 15 minutes resulted in signals that significantly weaker than those obtained by a straightforward incubation for these time points. Addition of the control competitor MP3 had little if any effect regardless of time of addition. Therefore the the association of these factors with the TP3 substrate is only partially stable. The 55 kDa factor that associates with the control MP3 substrate was found to behave in a similar fashion (Figure 5-7).

Figure 5-5.

Kinetics of factor binding: stable interactions

Either probes TP3 or MP3 ("TP", "MP") were incubated in extract for the indicated times ("preinc"), 100 ng of the indicated competitor RNA was then added if applicable ("cptr") and the reaction continued for the indicated times ("chase"). Sample processing and other figure conventions were as in Figure 5-3.

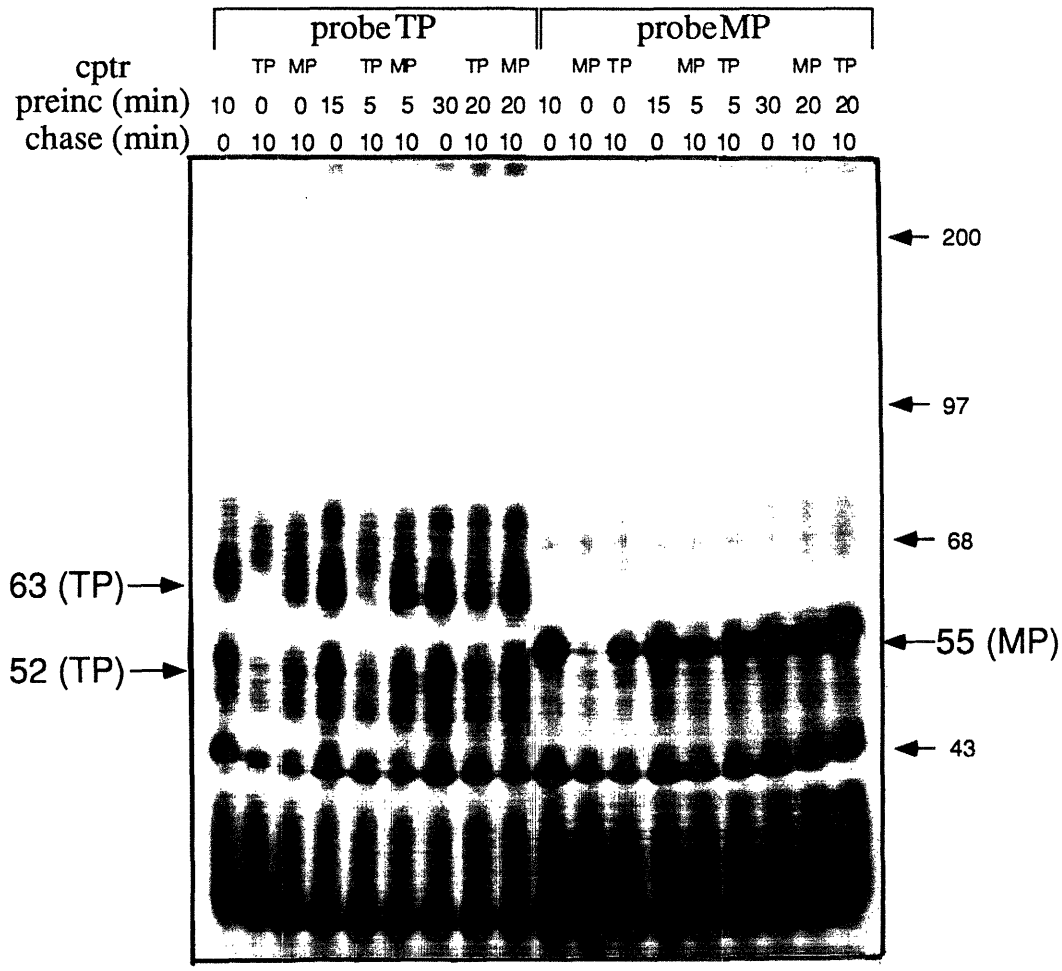

probe TP
probe MP
cptr
TP MP TP MP TP MP MP TP MP TP MP TP
preinc (min) 10 0 0 15 5 5 30 20 20 10 0 0 15 5 5 30 20 20
chase (min) 0 10 10 0 10 10 0 10 10 0 10 10 0 10 10 0 10 10
200
97
68
63 (TP)
52 (TP)
55 (MP)
43

Figure 5-6.

Kinetics of 63 kDa and 52 kDa (UGCAUG-specific) binding

Probe TP3 ("TP") was incubated in extract for the indicated times ("preinc"), 100 ng of the indicated competitor RNA was then added if applicable ("cptr") and the reaction continued for the indicated times ("chase"). Reactions were carried out at 30°C except where indicated ("temp (°C)"). Sample processing and other figure conventions were as in Figure 5-3.

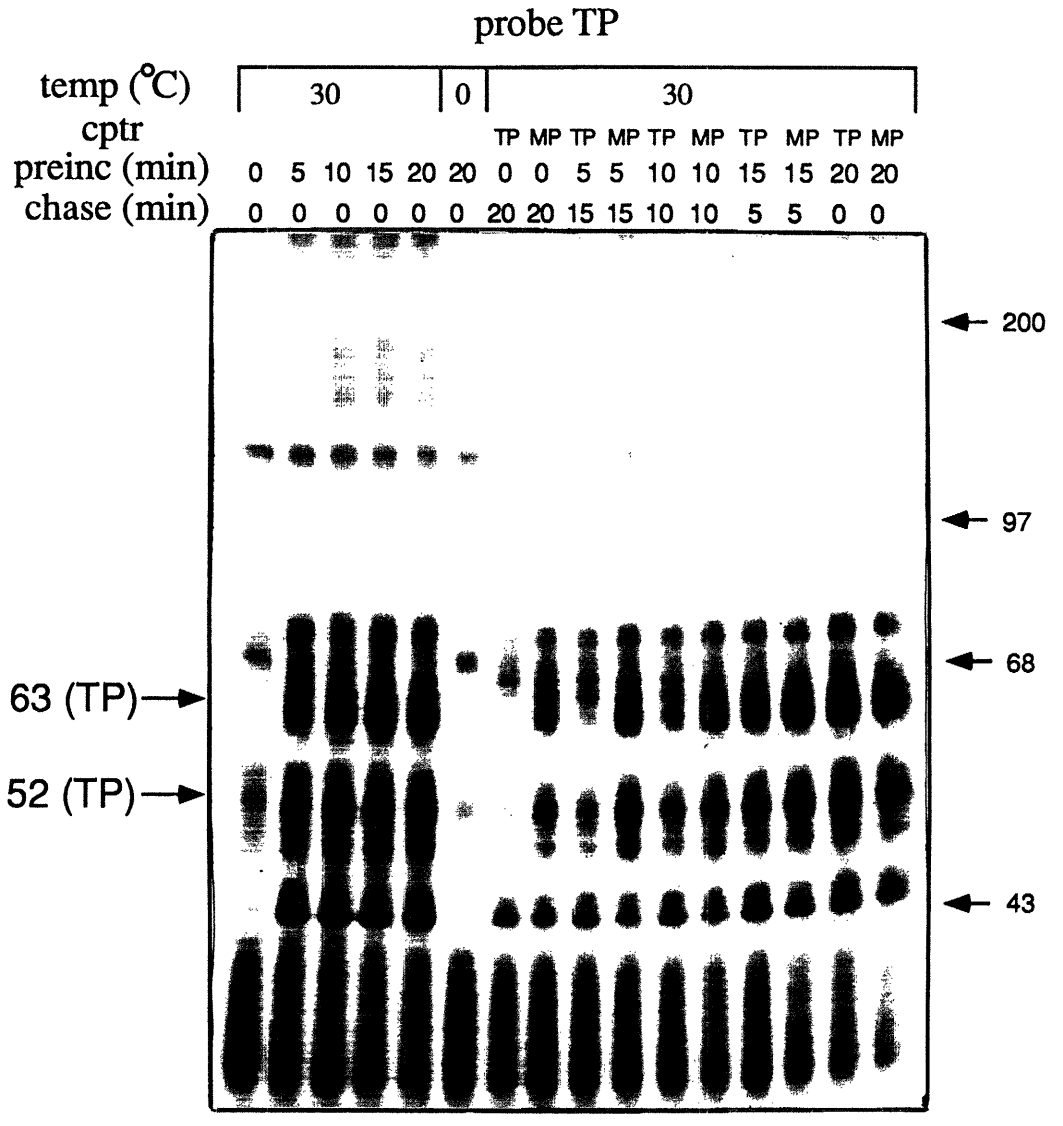

probe TP
temp (°C)
30
0
30
cptr
TP MP TP MP TP MP TP MP TP MP
preinc (min)
0 5 10 15 20 20 0 0 5 5 10 10 15 15 20 20
chase (min)
0 0 0 0 0 0 20 20 15 15 10 10 5 5 0 0
200
97
68
43
63 (TP)
52 (TP)

Figure 5-7.

Kinetics of 55 kDa (UGACUG-specific) binding

Probe MP3 ("MP") was incubated in extract for the indicated times ("preinc"), 100 ng of the indicated competitor RNA was then added if applicable ("cptr") and the reaction continued for the indicated times ("chase"). Reactions were carried out at 30°C except where indicated ("temp (°C)"). Sample processing and other figure conventions were as in Figure 5-3.

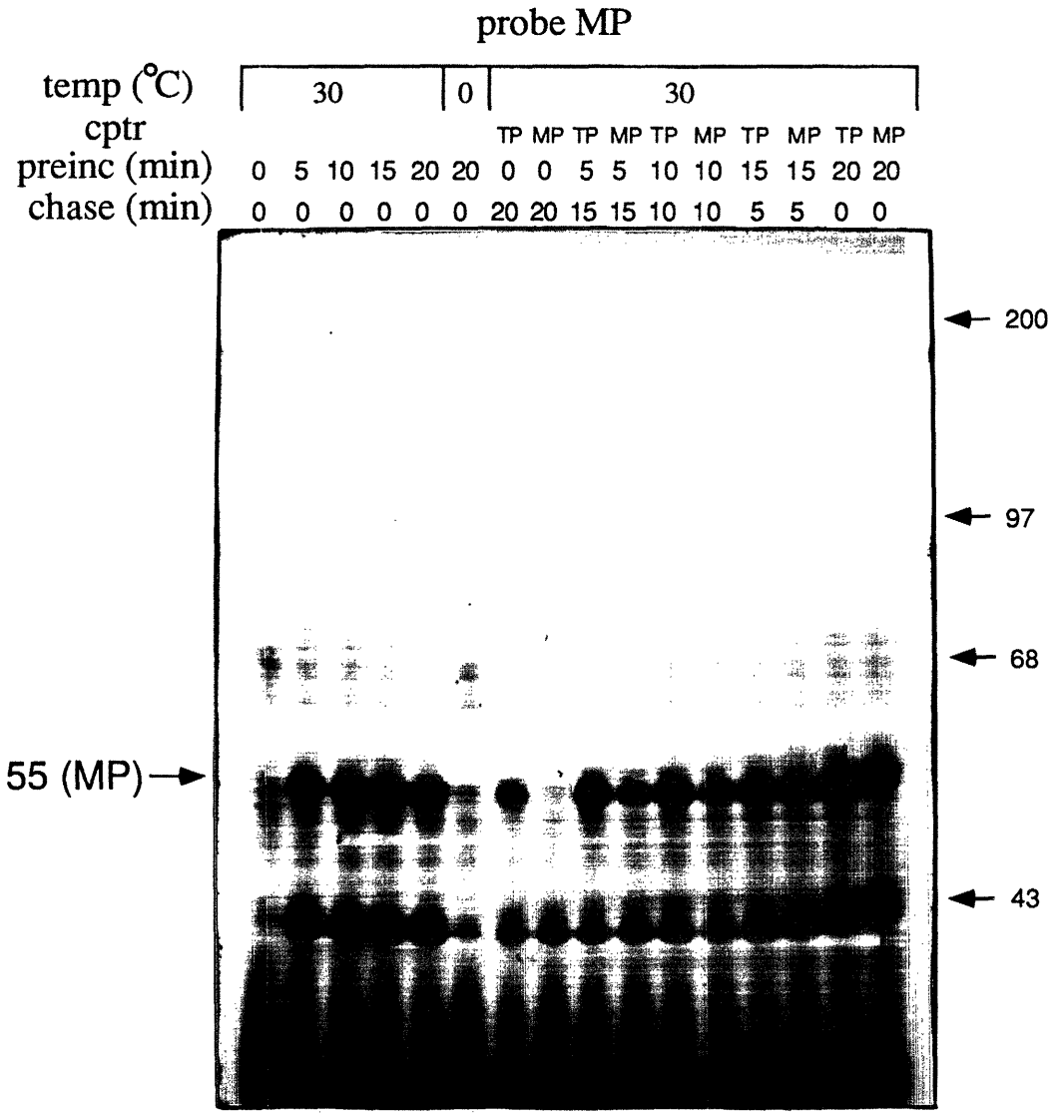
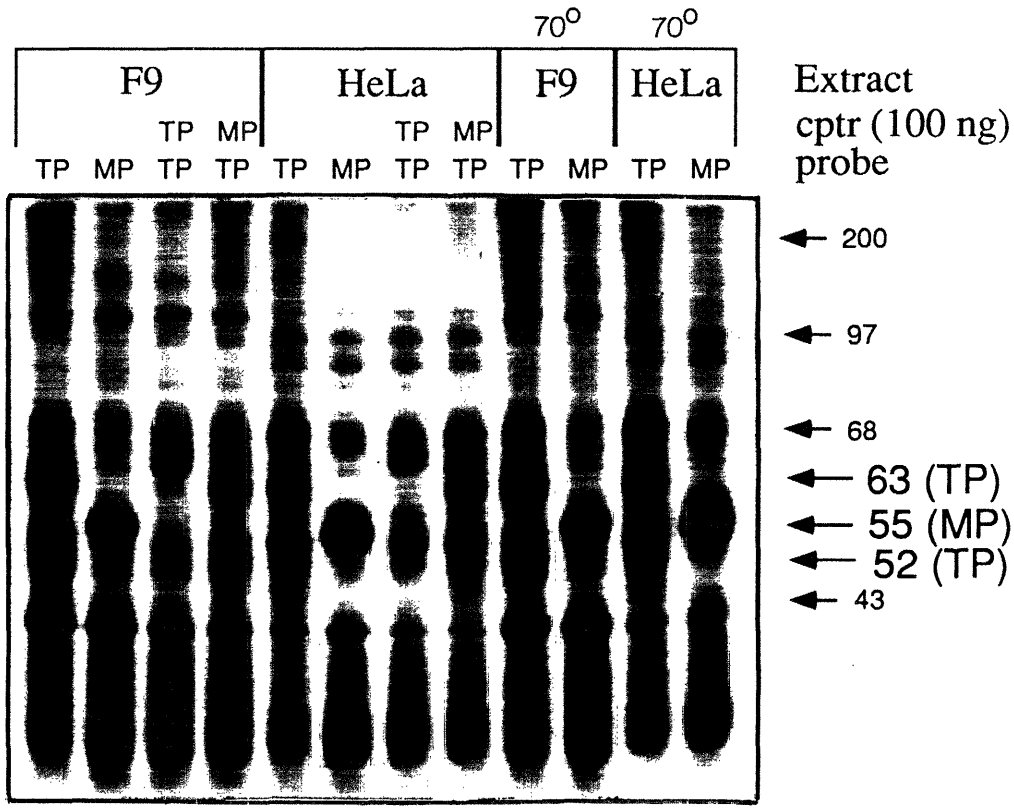

probe MP
temp (°C)
30
0
30
cptr
TP MP TP MP TP MP TP MP TP MP
preinc (min)
0 5 10 15 20 20 0 0 5 5 10 10 15 15 20 20
chase (min)
0 0 0 0 0 0 20 20 15 15 10 10 5 5 0 0
200
97
68
55 (MP)
43

Figure 5-8.

UGCAUG- and UGACUG-binding factors in HeLa and F9 extracts

The TP3 or MP3 probes ("probe") were assayed in extracts from either F9att5.51 cells ("F9") or HeLa cells ("HeLa"), either alone or in the presence of 100 ng TP3 or MP3 competitor ("cptr"). Samples were then processed as in Figure 5-3. In the last four lanes, samples were heated to 70°C for 15 minutes before the addition of RNase T1. Other figure conventions were as in Figure 5-3.


70°
70°
F9
HeLa
F9
HeLa
TP MP
TP MP
TP MP TP TP TP MP TP TP TP MP TP MP
Extract
cptr (100 ng)
probe
200
97
68
63 (TP)
55 (MP)
52 (TP)
43

The 63 kDa factor is present in HeLa cells

It was of interest to determine whether the UGCAUG-specific factors detected in F9 cells were also present in HeLa cell nuclear extracts. Therefore HeLa extracts were compared in parallel with F9 extracts. Both TP3 and MP3 probes were tested; in addition, TP3 probes were tested in the presence of 100 ng of unlabeled TP3 or MP3 competitor. It was clear that HeLa cell nuclear extracts contained factors that exhibited migration and competition characteristics similar similar to those exhibited by the TP3-specific 63 kDa and the MP3-specific 55 kDa factor (Figure 5-8). It was more difficult to assess whether the TP3-specific 52 kDa factor was present in HeLa extracts, due to the presence of background in this area of the gel. Other bands that were possibly more intense in HeLa reactions were visible, but it was not clear whether these were specific, as they comigrated with factors that appeared to bind both probes. Denaturing the extract after crosslinking (via heating at 70°C for 15 minutes) did not alter the binding profile upon RNase T1 digestion in either HeLa or F9 extracts, suggesting that the 63/55/52 kDa factors were probably protein and not RNA (Figure 5-8). In any event, HeLa cells appear to harbor at least some of the factors originally identified in F9att5.51 cells.

DISCUSSION

Previous chapters have analyzed and discussed the function and implications of hexamer element-dependent regulation of splicing; the data presented in this chapter extend these analyses by providing evidence for the presence of cellular hexamer-specific *trans*-acting factors that may govern how a given alternatively spliced pre-mRNA is treated in a mammalian cell.

RNA-binding factors specific for simple sequences

By virtue of previous studies, a number of convenient reagents were generated that permitted the biochemical identification of sequence-specific RNA-binding factors in a fairly rigorous fashion. These reagents are represented by the RNA probes used in this study. The use of these has demonstrated the presence of factors that exhibit specific and relatively stable interactions with RNA sequences that contain either UGCAUG (hexamer) or UGACUG (mutant) sequence motifs. This study was initiated with an eye toward the identification of UGCAUG-specific proteins that might be involved in the cell-type-specific alternative splicing of the rat fibronectin gene. In this respect, at least two candidate factors (63 and 52 kDa) were found to satisfy certain of the criteria imposed by experiments carried out *in vivo* in previous chapters.

However, these studies have also identified a factor that clearly exhibits specificity for a sequence not previously thought to have a biological function (i.e., the mutant hexamer-specific 55 kDa). This is a matter of concern, since it brings up questions regarding the relevance of biochemically identified factors that exhibit binding specificity for sequences of known biological interest. This unexpected result may imply that a range of discrete sets of binding

specificities exist in the cell, as has been proposed for some hnRNPs (Burd and Dreyfuss, 1994). In any event, both UGCAUG- and UGACUG-specific activities will be discussed, as both exhibit similar biochemical characteristics *in vitro*. However, for obvious reasons, more emphasis will be placed upon the possible identity of the UGCAUG-specific 63/52 kDa factors. The data presented in this study suggest but do not prove that these factors are proteins; a sequence-specific interaction by an RNA protected by tightly bound, heat-resistant protein could not be ruled out.

Possible known protein candidates

Numerous known splicing factors and RNA-binding proteins lie in the 50 to 70 kDa range; therefore the molecular weights of the UV-adducts identified here (63/52 kDa) provides little discriminating information regarding whether they correspond to any known proteins. The discussion of these will be restricted to proteins implicated in splicing *in vitro* and *in vivo*, although unfortunately the data presented here does not exclude proteins irrelevant to splicing or its regulation. Although sequence specificities have already been ascribed to a number of known proteins, it remains possible that additional specificities for these have not been discovered to date. Many of the proteins described below are constitutive splicing machinery components. Although the action of UGCAUG repeats appears to be cell-type-dependent, it is possible that constitutive splicing proteins adopt conformations or modifications that allow the effects of UGCAUG repeats to be observed only in certain cell types.

SnRNP proteins

Given that one function of the UGCAUG repeat *in vivo* may involve the recruitment of U1 snRNP for the activation of an upstream 5' splice site (Chapter 3), one particularly intriguing candidate for the 63 kDa factor is the U1-specific 70K protein (Query et al., 1989). In addition, a 70 kDa U5-associated intron-binding protein has been identified that binds to 3' splice sites (Gerke and Steitz, 1986; Tazi et al., 1986); given the negative effect of this cis-element on 3' splice site selection (Chapter 4), it is conceivable that UGCAUG repeats operate by sequestering U5-containing complexes in a nonproductive conformation. Another U5 protein (52 kDa) also falls within this range; in addition, the [U4/U6.U5] tri-snRNP complex contains a 60 kDa protein (Behrens and Lührmann, 1991). Other possible candidates include 60 and 66 kDa proteins specific for the 17S form of U2 snRNP also identified as splicing factor SF3a (Brosi et al., 1993) and as SAP60/SAP62 (Bennett and Reed, 1993).

The proteins associated with U2 snRNP, as well as U2 snRNA itself, are interesting with regard to the 55 kDa UGACUG-specific factor identified in this study. The M3 and MP3 probes used in this study contain a number of exact or close matches to the branchpoint consensus (YNYURAY). For example, M3 contains nine close matches (three each of the sequences TCCTGAC, GACTGAC and GCTTGAC), all of which are altered in the T3 transcript. MP3 contains three reasonable matches to the consensus (three TCATGAC sequences) that are altered in TP3. Since U2 snRNP is capable of binding the branch site in the absence of an adjacent pyrimidine tract (Nelson and Green, 1989), it is conceivable that U2 snRNP may bind these sequences in M3 or MP3 and thus bring U2-associated proteins in close proximity to the probe; however, this process would have to be ATP-independent. If any of these

proteins represent the activities described in this chapter, then it should be possible to demonstrate immunoprecipitation of these crosslinks using the appropriate antisera.

Non-snRNP proteins and splicing factors

A number of non-snRNP proteins have been implicated in splicing that fall within this molecular weight range, including SRp55 (Zahler et al., 1993), a proteolytic fragment of PSF (68 kDa; Patton et al., 1993), and U2AF (65 kDa; Zamore et al., 1992). Sequence preferences have been described for U2AF and PSF; however, the existence of alternative specificities cannot be ruled out, given the existence of multiple RRMs and other motifs in these proteins. In fact, alternative sequence affinities might actually be difficult to detect *a priori* unless addressed specifically, given the pyrimidine-rich preferences normally exhibited by these proteins.

hnRNP proteins

Numerous hnRNP proteins lie in this range, including PTB or hnRNP I (57-62 kDa; Piñol-Roma et al., 1988; Garcia-Blanco et al., 1989; Patton et al., 1991) as well as hnRNPs H (56 kDa), J (62 kDa), K, L, M (68 kDa), N (70 kDa), and P (72 kDa; Piñol-Roma et al., 1988). In particular, PTB (hnRNP I) has been implicated in the repression of β-tropomyosin exon 7 usage by binding to specific sequences upstream of this exon (Mulligan et al, 1992). However, it would again be necessary to invoke an additional sequence specificity in order to accomodate such a model for UGCAUG repeat action, since PTB is known to prefer only pyrimidine-rich tracts to date.

It is also possible that some of the other hnRNPs exhibit UGCAUG or UGACUG RNA-binding specificity. The binding of hnRNP proteins has been

shown to be somewhat substrate-dependent (Swanson and Dreyfuss, 1988a, 1988b; Bennett et al., 1992b; Matunis et al., 1993); in fact, a high-affinity binding site for hnRNP A1 has recently been identified by SELEX that has a hexanucleotide consensus (UAGGG[A/U]; Burd and Dreyfuss, 1994). It is likely that other hnRNP proteins have short, high affinity recognition sequences (reviewed by Dreyfuss, et al., 1993). This possibility could readily be tested by incubating the T3/M3/TP3/MP3 probes with immunopurified hnRNP complexes, which contain a collection of the known hnRNPs in the appropriate size range (Piñol-Roma et al., 1988).

RNA-binding kinetics

The UV-crosslinking of the UGCAUG-specific and UGACUG-specific factors reported here takes place over an unusually long time, and contrasts with what is known about many other RNA-binding proteins. In addition, binding is negligible at 0°C. A number of reasons may explain these findings. These RNA-factor interactions could be energy-dependent; however, since binding was carried out in the absence of ATP or magnesium, this seems unlikely. Alternatively, the stable detection of these factors may require the assembly of multiple components on a single site, only a few of which are detected by UV-crosslinking; this might also explain why dissociation rates were also slow. Additionally, it is possible that the binding of these factors is normally rapid but is rate limited in nuclear extract due to slow dissociation of these factors from endogenous nucleic acids in the extract; binding assays carried out in preincubated or micrococcal nuclease-treated extracts may help address these possibilities.

The fact that the UGCAUG-specific crosslinks observed were relatively resistant to subsequent challenge with competitors is encouraging with respect

to the potential purification of the 63 and 52 kDa factors. These results suggest that other assays such as gel mobility shift analyses could be used to further characterize these factors. It was not possible to obtain clear gel shift in nuclear extract using these probes, primarily because control substrates (such as M3 and MP3) also interacted with specific factors. Nevertheless, it may be possible to obtain gelshifts in partially purified fractions, as has been documented for the elements that repress exon 7 usage in β-tropomyosin (Guo et al., 1991; Mulligan et al., 1992) and for elements upstream of exon 4 in the calcitonin/CGRP gene (Roesser et al., 1993).

UGCAUG-binding factors in other cells

UGCAUG-binding factors were found to be present in HeLa nuclear extracts as well as in F9-derived extracts. This might be expected, since the UGCAUG element appears to function in the preprotachykinin (Chapter Three) and calcitonin/CGRP (Chapter Four) genes. The 63 kDa factor is present at levels comparable to that found in F9 extracts, which might suggest that this factor is present constitutively. This remains to be seen; characterization of these UGCAUG-binding factors in extracts from other sources (such as COS or liver) may help to clarify this issue. It is possible that cell-type-specificity is conferred by factors that do not recognize the element directly but interact via constitutive "adaptors" (e.g., *tra*; cf. Tian and Maniatis, 1992). It is also possible that binding does not correlate with function; for example, an alternatively spliced SF2/ASF form without an RS domain binds RNA but is nonfunctional for splicing (Zuo and Manley, 1993; Cacares and Krainer, 1993). Post-translational modifications may also affect regulatory function of this factor. Neural-specific *c-src* splicing has been observed to occur upon neural induction or phorbol ester treatment of*Xenopus*

neurectoderm by TPA or mesoderm independently of protein synthesis (Collett and Steele, 1993), suggesting that post-translational modifications and/or signal transduction pathways may be involved in the function of factors that regulate RNA splicing. Phosphorylation has been found to alter the biochemical properties of hnRNP A1 and C (Cobianchi et al., 1993; Mayrand et al., 1993), supporting the idea that phosphorylation could modulate the extent of factor activity. The 52 and 63 kDa bands were found to migrate somewhat heterogeneously, which could reflect either heterogeneous RNase T1 cutting or post-translational modifications to differing extents.

Alternatively, the ratios of distinct UGCAUG-binding factors may dictate cell-type-specific alternative splicing function; this might explain why levels of a factor may apparently remain relatively unchanged among cell types. Although the level of 52 kDa present in HeLa extracts appeared to be lower relative to 63 kDa, this was difficult to establish conclusively. For that matter, the UGCAUG-specific 52 kDa factor may represent a 63 kDa degradation product. In any event, the further characterization of the factors that bind this UGCAUG element should resolve their relevance and potential function.

MATERIALS AND METHODS

Transcription vectors and RNA synthesis

Cloning methods were as described (Ausubel et al., 1987). Transcription templates (for *in vitro* synthesis of hexamer-containing RNAs) were derived as described in Chapter Four. Briefly, inserts from the 7iBi89ΔGA/pBAGH expression vector derivatives that carried inserts T3, M3, S3, TP3 and MP3 (Chapter Three) were excised via AflII digestion and cloned into the BamHI site of pBluescript SK^- (Stratagene) via blunt-end ligation. Inserts were cloned in both orientations; thus transcription could be carried out using either T3 or T7 RNA polymerase (Promega, Stratagene). Transcriptions *in vitro* were carried out using BglII-linearized templates using conditions described by the manufacturers; T3 transcription reactions for probes contained 0.1 mCi α-[^{32}P]-UTP per 25 µl reaction (800 Ci/mmol; final specific activity approx. 1.3×10^9 dpm/µg). T7 transcription was employed for synthesis of unlabelled competitor RNAs; these reactions contained 0.5 mM unlabeled UTP. Both labelled and unlabelled transcripts were purified via 5% denaturing polyacrylamide gel electrophoresis; full-length labeled and unlabeled transcripts were identified and isolated by autoradiography and UV-shadowing respectively and were eluted in 0.4 ml 300 mM sodium acetate, 1 mM EDTA pH 8.0, 0.5% SDS (for labelled transcripts, 10 µg E. coli tRNA was included in the elution buffer as carrier). RNAs were precipitated twice from ethanol before use. Unlabelled RNAs were quantitated via ethidium bromide fluorescence, using total yeast RNA as standards.

Nuclear extract preparation and UV-crosslinking assays

HeLa cells (a kind gift from Phillip Sharp's laboratory, MIT) were grown in suspension in Joklik's MEM (BRL/Gibco) containing 5% horse serum (Sigma). An adhesion-defective variant of F9 cells capable of growing in suspension (F9att5.51; a kind gift of Eileen Adamson, La Jolla Cancer Research Foundation; Grover et al., 1987) were cultured in Joklik's MEM containing 9% bovine calf serum (Sigma) and 1% fetal bovine serum (Hazelton). Nuclear extracts were prepared essentially as described by Dignam et al. (1983), with the following modifications. After washing with PBS and buffer A, cell pellets were resuspended in Buffer A to a total cell volume that was 2.5 times the packed cell volume prior to homogenization. In addition, the final extract dialysis buffer contained 42 mM ammonium sulfate instead of 100 mM KCl. Protein concentrations were determined using a commercial kit (Pierce).

UV-crosslinking assays were carried out in a reaction volume of 25 µl and contained 20 mM HEPES-NaOH pH 7.6, 50 mM KCl, 1 mM DTT, 15-20 units RNasin, $3x10^5$ cpm labeled RNA (230 pg; 4 fmol), 50-100 ng E.coli tRNA and 40% (v/v) nuclear extract (45-55 µg total protein). The final concentrations of known components, including those in the dialysis buffer, were as follows: 28 mM HEPES-NaOH pH 7.6, 50 mM KCl, 16.8 mM $(NH_4)_2SO_4$, 0.04 mM EDTA, 9% glycerol, 1.2 mM DTT, 0.6-0.8 units/µl RNasin, 2-4 ng/µl E. coli tRNA, 9 pg/µl radiolabeled probe and about 2 mg/ml protein. Incubations (in 1.6 ml Eppendorf tubes) were for the times and temperatures indicated; tubes were then transferred to ice, the tube caps cut off, and the samples exposed for 15 minutes to shortwave (254 nm) UV radiation using a lamp (10.0-10.5 mW/cm^2 at the lamp surface) at a distance of 5 cm. Reactions were then moved to new tubes each containing 2.5 µl of 2 mg/ml RNase T1 (Boehringer-Mannheim) and incubated at 30°C, usually for 10 minutes. Samples were then

boiled for 5 minutes in SDS sample buffer containing 100 mM DTT (Ausubel et al., 1987) prior to analysis on a 9% or 12% discontinuous SDS polyacrylamide gel (3% stacking gel; total gel length 20 cm) along with prestained molecular weight markers (BRL/Gibco). Gels were run at 15-20 W until the bromophenol blue had entered the separating gel; gels were run at 35 W thereafter. After the bromophenol blue had run off the bottom, gels were fixed in 10% methanol, dried, and exposed to X-ray film or were analyzed using a Molecular Dynamics Phosphorimager.

ACKNOWLEGEMENTS

I would like to thank Eileen Adamson and Bob Ezzell for kindly providing the F9att5.51 cell line used in this study. I would also like to thank members of the Sharp laboratory for providing HeLa cells and for helpful discussions regarding biochemical assays.

Chapter Six

General Discussion

GENERAL DISCUSSION

The nuclear pre-mRNA splicing machinery can excise numerous intervening sequences from primary RNA transcripts that can be upwards of hundreds of kilobases in length. The RNA processing machinery of the cell must therefore be profoundly accurate, lest inappropriate mRNAs be generated at intolerably wasteful levels. Authentic splice sites must contain sufficient information in *cis* to allow their faithful identification, even in the context of cryptic splice site sequences that frequently occur in an RNA precursor. The information afforded by the splice site junctions alone is insufficient for this task. However, the identification of sequences that aid in the specification of splice sites has not been straightforward. It is possible that the difficulty in clarifying the mechanism of splice site recognition may lie in an inherent redundancy, coupled with structural and positional degeneracy of the information that is involved.

The pre-mRNA splicing machinery, in turn, must recognize this *cis*-information efficiently. The splicing mechanism is configured to recognize a correct splice site not only on the basis of its sequence but also by its relationship to its immediate milieu, i.e., to nearby sequences and their associated structures. This recognition mechanism is still largely unknown, perhaps because an understanding of the evolution of these processes has not yet been achieved. It is possible that alternative spliced systems may reflect some of the intermediate stages in the evolution of this mechanism. In this respect, many examples of alternative splicing could be viewed as paradigms that may lead to a better understanding of how splicing evolved in general. Given the propensity of evolution to provide several solutions to biological conundra such as this, it seems likely that each case of alternative splicing will

represent a collection of mechanisms that control splice site specification. Some of these may be common to many systems; others may be unique. The alternative splicing of the EIIIB exon in the fibronectin gene appears to represent such a collection. Some aspects of the EIIIB region appear to serve merely to establish a balanced regulatable state; other aspects serve to shift this balance in a cell-type-dependent manner. The studies described here, in contrast to those described by others, have permitted a formal discrimination between these two aspects of EIIIB regulation.

The role of the splice site sequences

The study of alternatively spliced genes has led to the identification of numerous sequences that govern splice site selection. Many of these findings have illustrated the importance of the splice sites themselves in the maintenance of alternative splicing. The findings obtained in initial experiments (Chapter Two) suggested a requirement for suboptimal 5' splice sites, not only of the regulated exon but also of that with which that exon competes. In the case of EIIIB splicing, the strengths of most if not all the participating splice sites are balanced relative to one other. In retrospect, this is not surprising, since cell-type-specific regulators could act by shifting a preexisting balance.

A priori, splice site balance does not necessarily require that the competing splice sites be suboptimal, although this is the case for EIIIB. Since EIIIB appears to be intrinsically poorly recognized, splice site balance via suboptimal sites may have been easier to achieve from an evolutionary perspective. This may be true of a number of alternatively spliced situations; suboptimal flanking splice sites may play a role in maintaining normal spliced product ratios in the murine *c-src* gene as well as in the *Drosophila Sex-lethal*

gene (Black, 1991; Horabin and Schedl, 1992; Chapter Two) and are of importance in allowing a balanced splicing of certain retroviral RNAs (Katz and Skalka, 1990; Fu et al., 1991). Other regulated exons appear not to be sensitive to splice site balance (Streuli and Saito, 1989; Tacke and Goridis, 1991); these exons may not be as sensitive to this parameter, which might be the case if flanking exon splice site recognition were not rate-limiting for either inclusion or skipping.

EIIIB is a poorly recognized exon partly as a consequence of suboptimal splice sites, since mutational improvement of either EIIIB splice site dramatically increased EIIIB inclusion. The EIIIB 3' splice site was of particular interest since its unusual structure and multiple branch usage *in vitro* is reminiscent of that exhibited by β-tropomyosin (Norton and Hynes, 1990; Helfman and Ricci, 1989; Goux-Pelletan et al., 1990). This unusual EIIIB 3'SS structure was not required for cell-type-specific regulation; on the other hand, this 3'SS is likely to contribute to general EIIIB inefficiency, since its replacement resulted in near-constitutive EIIIB inclusion. If these findings can be extrapolated to other systems, then the 3'SS elements found to repress β-tropomyosin exon 7 (Guo et al., 1991; Libri et al., 1991) may reflect a general requirement for poor exon 7 recognition rather than cell-type-specificity; the former is supported by the finding that exon 7 elements specifically interact with PTB, a protein that is ubiquitously expressed (Mulligan et al., 1992; Patton et al., 1991). This remains to be seen.

Maintenance vs. regulatory elements

A common approach to regulated vertebrate alternative splicing has consisted of analyzing the effects of mutations in regions suspected of harboring splicing elements. This has resulted in the identification of many

sequences that are clearly capable of influencing splicing in *cis*. The results obtained with EIIIB suggest that multiple elements govern the correct patterns of cell-type-specific EIIIB splicing. However, not all of these participate in cell-type-specific control. Therefore it may be premature to conclude that a given *cis*-affective element is cell-type-specific *per se* unless that element were also shown to possess differential activity among regulated cell types.

In studies characterizing elements in the rat calcitonin/CGRP and *c-src* genes, the effects of mutations were analyzed in cell lines that exhibited different splicing patterns. Mutations disrupting certain elements were found to cause splicing in one cell type to resemble that of the other, suggesting their importance as cell-type-specific control sequences. In studies of the calcitonin/CGRP gene, mutations were found that activated calcitonin-specific processing in F9 cells (which normally prefer CGRP; Emeson et al., 1989). Studies of the *c-src* gene have identified intron mutations that abolished N1 exon inclusion in the neuroblastoma LA-N-5 (Black, 1991, 1992). In neither case did these mutations affect HeLa cell patterns. However, since HeLa cells normally prefer calcitonin-processing and N1-skipping to begin with, these results could have been due to the disruption of nonspecific elements. This is a common caveat to mutational analyses and was specifically addressed in the analysis of EIIIB (Chapters Two and Three).

Initial deletion analyses suggested the presence of elements downstream of EIIIB that activated inclusion. For the purposes of identifying genuine regulatory elements, it was of paramount importance to ascertain that these elements acted cell-type-specifically. Consistent with this hypothesis, when tested in a background that allowed detection of inclusion in all cell lines (minigene M2), IVS2 deletions had effects that correlated with the ability of each cell line to include EIIIB. This type of study exemplifies an approach

whereby the cell-type-specific activity of *cis*-sequences can be rigorously assessed; this type of approach has not yet been fully adopted for the characterization of other vertebrate splicing elements to date. Potential similarities exist between regulation of EIIIB and that of calcitonin/CGRP and *c-src*; the analysis of EIIIB, in contrast to many other studies, goes a step further by illustrating cell-type-dependent *cis*-sequence action. These studies therefore extend the analysis of vertebrate splicing elements *in vivo* to a degree not attained previously and also represent a more sophisticated approach *vis a vis* the identification of regulatory control sequences in general.

Repeated hexamer elements positively control EIIIB splicing

The downstream intron element that was important for EIIIB inclusion was relatively large. It also appeared to be multipartite and partially redundant, since none of the deletions in this region completely abolished cell-type-specific exon inclusion (Chapter Two). Even so, it was astonishing to discover that a short repeated hexamer sequence was capable of accounting for most if not all of the activity exhibited by this element (Chapter Three). This finding has implications for the regulation of vertebrate alternative splicing in general.

Using assays previously used to characterize the IVS2 region required for EIIIB splicing, it was found that this hexamer element, like the larger IVS2 element, acted in a cell-type-dependent manner. In addition, a heterologous alternatively spliced gene (preprotachykinin) was also employed to assess functional autonomy. These results indicated that the hexamer element was capable of exerting its function in the context of other optional exons. Furthermore, these experiments establish the validity of the PPTK construct as

a potentially useful "reporter" gene that may in future facilitate the evaluation of *cis*-elements described in other alternatively spliced genes.

Previously characterized systems in *Drosophila* have found that the *cis*-targets for the regulation of *dsx* and *Sxl* alternative splicing are repeated sequence elements distributed in the vicinity of the site of regulation (Ryner and Baker, 1991; Hedley and Maniatis, 1991; Inoue et al., 1992; Sakamoto et al., 1992; Horabin and Schedl, 1993). The hexamer motif we describe is similar in that it normally functions as a highly repeated sequence array. The fact that *doublesex* regulation can be recapitulated *in vitro* in HeLa cell nuclear extracts (Tian and Maniatis, 1992) suggests that mechanisms for repeat-mediated regulation exist in mammals; however, it remained to be seen whether repeats were a common mode of mammalian splicing regulation. Our findings confirm that certain *Drosophila*-based paradigms of regulated alternative splicing can indeed be extrapolated to mammalian systems. The hexamer element is somewhat shorter than those that regulate *Drosophila* splicing. This may explain why regulatory elements have been harder to identify in mammals (if the TGCATG hexamer can be taken as representative of mammalian elements), since a standard internal homology search or a simple scanning mutational analysis is unlikely to identify 5- or 6-tuple redundancy in a long sequence.

The hexamer element can regulate calcitonin/CGRP alternative processing

The observation of several GCATG repeats in the region of the calcitonin-specific exon 4 in the human calcitonin/CGRP gene prompted a test to see whether synthetic hexamer repeats affected calcitonin exon 4 selection (Chapter Four). As predicted, hexamer repeats were found to downregulate exon 4 splice acceptor usage in a cell-line-dependent manner. These results

suggested but did not prove the natural role of hexamer-related repeats in calcitonin/CGRP regulation. In fact, as discussed in Chapter Four, the regulation of calcitonin/CGRP remains unresolved, primarily because small replacement mutations appeared to have no consistent effect upon alternative processing (Yeakley et al., 1993). An appropriate experiment that might address the role of GCATG repeats would involve mutation of all GCATGs in the vicinity of exon 4, followed by reintroduction of these repeats as a synthetic oligonucleotide. This remains to be carried out.

If calcitonin/CGRP regulation were indeed mediated by hexamer elements, then one might expect *in vitro* studies to detect hexamer-specific factors (Chapter Five). A system has been developed to assess the repression of calcitonin splicing: fractions from a rat brain extract have been shown to specifically repress exon 4 acceptor usage in a globin/calcitonin hybrid intron. This repression could be relieved by the specific addition of exon 4-containing competitors (Roesser et al., 1993). Curiously, a 41/43 kDa doublet was found to crosslink to the calcitonin acceptor region, a finding that contrasts with the hexamer-specific 63/52 kDa factors documented in our study (Chapter Five); however, since tissues and assays differed, these results may not be comparable. Therefore it remains to be seen whether calcitonin/CGRP and FN are regulated via similar *cis*-elements. Even if the regulation of calcitonin/CGRP processing proceeds in part via hexamer regulation, it seems likely that additional *cis*-acting elements exist that will be unique to each of these two genes.

The hexamer element may be multifunctional

Another novel feature of EIIIB regulation concerns the positioning of the hexamer repeats relative to the EIIIB exon. Many of these elements are closer

to the IVS2 splice acceptor than to the EIIIB donor. As a consequence, these findings may challenge a current belief regarding the regulation of alternative cassette exons. *Cis*-acting regulatory elements, as a general rule, have been thought to be local to the regulated exon (for examples, see Black, 1991; Tacke and Goridis, 1991; Cooper et al., 1988; Streuli and Saito, 1989). In contrast, the hexamer repeats in the EIIIB region are located at a considerable distance downstream from the exon they regulate and are located more closely to a constitutively used 3' splice site. Regulation at a common downstream splice site, at a distance from an alternative exon, has not been seriously considered as a possible regulatory mechanism for alternatively spliced exons and therefore represents a novel mode of regulation.

Hexamer function was evident even when repeats were placed well into a heterologous intron; hence hexamer function is somewhat position-independent. Nevertheless, it was possible that hexamer repeats exerted distinct effects upon 5' and 3' splice sites. The deletion of most of the IVS2 repeats resulted in an aberrant splicing phenotype characterized by an inability to use the EIIIB 5' splice site. This phenotype could be partially reversed by the reinsertion of synthetic hexamers, suggesting a positive effect of hexamer repeats upon 5' splice site usage (Chapter Three). In contrast, hexamer repeats negatively affected 3' splice site selection when placed in the rat calcitonin/CGRP gene. Taken together, these experiments suggest that hexamer elements mediate both positive and negative effects upon splice site selection, depending upon the situation. Alternatively, a given repeat may perform only one of these functions, depending on its relative position within the intron. We have not tested this specifically; however, it is interesting to note that one of the IVS2 repeats close to EIIIB lies within a very highly conserved stretch of sequence (Figure 3-1).

It is possible that hexamer elements mediate both 5' splice site activation and 3' splice site repression via a common mechanism. Although it is difficult to explain how 5' splice site activation could mediate a switch in 3' splice site usage in the calcitonin/CGRP system, a hexamer-mediated delay in 3' splice site commitment may be sufficient to explain both phenomena. In the case of the calcitonin/CGRP gene, this would lead to usage of CGRP-specific exons; in the FN gene, this might allow more time for EIIIB recognition before splice site pairing and catalysis.

Such mechanisms could be tested *in vitro*, pending the development of nuclear extracts that recapitulate *in vivo* regulation. This has already been done for *c-src* regulation, which may also involve the participation of this hexamer element (Black, 1992; Chapter Three). The addition of hexamer-containing competitor RNA might be predicted to specifically activate the 3' splice site of either calcitonin exon 4 or of FN exon III-8a. This type of effect has in fact been documented for usage of the calcitonin-specific acceptor in HeLa extracts: the inhibition of splicing of a globin/calcitonin hybrid intron (in the presence of rat brain fractions) could be relieved by the specific addition of RNA that contains calcitonin acceptor sequences (Roesser et al., 1993). However, this type of effect has not been seen in the case of *c-src* N1 splicing *in vitro*; instead, a reduction in N1-specific or overall splicing was observed (Black, 1992). It is possible that hexamer-dependent negative effects on 3' splice site usage are not rate-limiting in nuclear extracts under the conditions employed for *c-src* splicing. In any event, the exact mechanism(s) of hexamer element action remain to be identified.

Identification of hexamer-specific factors

Given the precedents established in the regulation of *doublesex* alternative splicing, one might predict that the hexamer elements might exert their action via the binding and subsequent action of sequence-specific factors. This prediction was tested by using hexamer-containing RNA probes in a UV-crosslinking assay in nuclear extracts derived from F9 cells (Chapter Five). Multiple factors could interact with a UGCAUG hexamer element in a sequence-specific manner. Assuming that these are proteins, the molecular weights of these factors (63 and 52 kDa) could correspond to a large number of previously identified RNA-binding proteins (discussed in Chapter Five). Many of these have already been shown to exhibit specificity for other sequences; however, this does not rule out additional specificities for these proteins, particularly since many of these contain multiple RRMs that may individually possess distinct sequence preferences (Kenan et al., 1991; Dreyfuss et al., 1993; Burd and Dreyfuss, 1994). A number of reagents for known proteins may therefore help identify the 52/63 kDa factors. In addition, initial estimates of nuclear extract abundance (approx. 0.01%-0.06% of total nuclear extract protein) suggests that purification of these factors is feasible. The purification of these factors may therefore facilitate the development of an *in vitro* system that recapitulates hexamer-dependent regulation.

In addition to the striking sequence specificity of these 52/63 kDa factors, the chance discovery of an unanticipated mutant hexamer-binding 55 kDa factor suggests that many cellular RNA-binding specificities may normally exist. A number of hnRNP proteins (notably hnRNPs A1 and C) have been shown to exhibit sequence- or substrate-specific binding to RNA (Swanson and Dreyfuss, 1988a, 1988b; Bennett et al., 1992b; Burd and Dreyfuss, 1994); therefore it is possible that the test (UGCAUG) and mutant (UGACUG)

specificities documented here represent the binding of hnRNP proteins. The role of hnRNPs in alternative splicing is not yet clear; their relative abundance and apparent ubiquity would seem to exclude a role in cell-type-specific regulation. However, posttranslational modification can affect hnRNP activities (Cobianchi et al., 1993; Mayrand et al., 1993); in addition, hnRNP I has been shown to be developmentally regulated (Patton et al., 1991). Therefore hexamer recognition by hnRNPs cannot be ruled out and may in fact supported by the relative abundance of these factors (0.01-0.06% of total nuclear extract protein, compared with 0.01-0.05% for hnRNP I/PTB; Garcia-Blanco et al., 1989) and the existence of hexamer-binding activities in HeLa as well as F9 extracts (Chapter Five).

Since hexamer-factor complexes appear to be relatively stable, it is possible that several proteins assemble on the RNA; techniques similar to those employed by Tian and Maniatis (1993) could be employed to identify and purify complex components. Alternatively, COS cell expression libraries could be employed to identify factors that activate EIIIB. Since COS cells normally skip EIIIB completely, such an approach would be ideal for the genetic identification of hexamer-dependent EIIIB-activating factors. Such an approach might not work if the basis for regulation required the expression of multiple components simulataneously or if posttranslational modification were critical; nevertheless, this type of assay might provide strong candidates for direct regulators of EIIIB.

Alternative splicing regulation by repeated short sequences: A model

Models for hexamer action were outlined in Chapters Two and Three and have here been supplemented to include information presented in subsequent chapters (Figure 6-1). In the absence of EIIIB-activating factors,

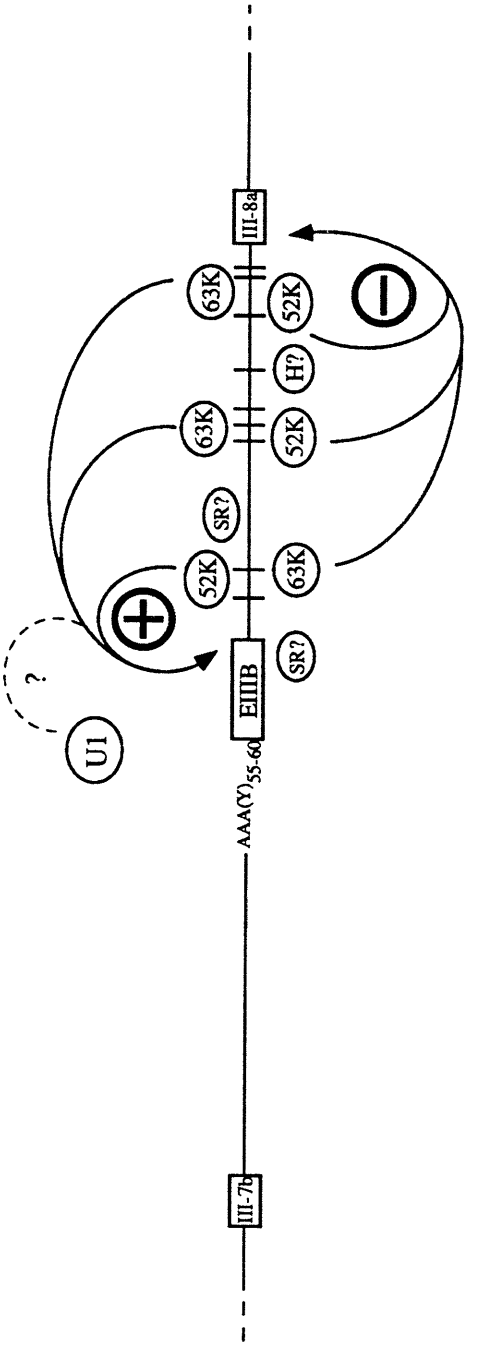
EIIIB is intrinsically poorly recognized and is therefore skipped by default in tissues such as liver and in COS cells. EIIIB activation models already discussed include: (i) direct recruitment by hexamer repeats of U1 snRNP to the EIIIB exon, (ii) the binding of substrate-specific splicing factors such as SR proteins and (iii) the binding of hnRNPs to the intron following EIIIB in order to facilitate the accessibility of EIIIB to the splicing machinery. Similar mechanisms may be at work in the neuron-specific recognition of the N1 exon in c-src, although such mechanisms must also account for the presence of only one hexamer in the intron following N1. Regulation of N1 involves multiple downstream elements (Black, 1992); therefore other sequence motifs may collaborate with the single hexamer in order to activate N1 for splicing.

In addition, we provide evidence that suggests a hexamer-mediated delay in 3' splice site commitment in the rat calcitonin/CGRP gene. These findings can easily be extended to consider EIIIB regulation, considering the distal position of the hexamer repeats. Activation of the EIIIB exon may then occur by delayed usage of the exon III-8a 3' splice site. A model invoking negative effects on splice site selection need not necessarily result in a reduction in mRNA levels, since splicing may not necessarily be rate-limiting *in vivo* for mRNA production (Pikielny and Rosbash, 1985). The normal regulation of the calcitonin/CGRP gene probably involves additional elements unrelated to those found in the EIIIB region, since EIIIB splicing and CGRP production do not always correlate. A model invoking 3' splice site attenuation is strikingly reminiscent of the models proposed for regulation of mutually exclusive splicing in the β-tropomyosin genes; in these, it has been proposed that the second exon of a mutually exclusive pair is repressed in a cell-type-specific manner, thus allowing the first exon to be recognized by default (Guo et al., 1991; D'Orval et al., 1991; Libri et al., 1991; Guo and

Figure 6-1.

Potential biochemical models for repeat-mediated EIIIB recognition

Diagram of the EIIIB region of the rat fibronectin gene, showing mechanisms by which hexameric elements may mediate their action. Exons and introns are represented by rectangles and horizontal lines respectively. GCATG and TGCATG repeats are indicated by vertical lines. The unusual EIIIB 3' splice site (with three adenosine branchpoints and extended pyrimidine tract [$AAA(Y)_{55-60}$]) is also shown. Curved arrows indicate positive (+) or negative (–) influences upon splice site recognition or commitment. Potential UGCAUG-binding factors are indicated as labelled ovals (63K, 52K). In addition, potential factors that may be recruited by these repeated elements (U1, SR proteins and hnRNP proteins as "U1", "SR" and "H" respectively) are indicated near sites with which they may interact. The locations of 63K and 52K binding are arbitrarily shown. Question marks and dashed lines indicate influences which have not explicitly been demonstrated.


III-8a
63K
52K
H?
63K
52K
SR?
52K
63K
SR?
EIIIB
III-7b
AAA(Y)55-60
U1
?

Helfman, 1993). In this light, it is conceivable that splicing in the EIIIB region might have been mutually exclusive at one time, with exons EIIIB and III-8a acting as partners.

While N1 inclusion and CGRP splicing are primarily neuronal, EIIIB inclusion occurs in early embryos. CGRP production and EIIIB inclusion are preferred events in F9 cells. Therefore correlations do exist in the regulation of these genes among various cell types. Based on these models, one might predict that EIIIB inclusion should be high in neurons. However, FN is not known to be synthesized by neurons *in vivo*; hence little information exists that would address this prediction. However, it is interesting to note that in one study of EIIIB splicing (Magnuson et al., 1991), a human neuroblastoma cell line (IMR-32) exhibited the highest EIIIB inclusion levels of all the cell lines tested (85-90%). These correlations suggest that the hexamer repeats documented here may participate not only in embryo-specific RNA processing events but also in neural-specific RNA regulation.

Short repeats and the evolution of RNA recognition

Given the complexity of hnRNA, a single TGCATG sequence would have little informational capacity to recruit *trans*-acting factors unless additional signals were present. Therefore it is not surprising that this motif is extensively repeated in IVS2. The role of repeated sequences in RNA splicing has been clearly documented in one other system, the regulation of *Drosophila doublesex* (*dsx*) pre-mRNA splicing by the *transformer* (*tra*) and *transformer-2* (*tra-2*) genes. In contrast, the repeats we document are shorter and can promote 5' splice site usage. Interestingly, each *dsx* 13-mer repeat can be depicted as two hexamers separated by an adenosine residue, the first hexamer

being a variable copy of the second (TCaaCA A TCAACA ; Burtis and Baker, 1989; Inoue et al., 1992).

Other examples of reiterated RNA elements include the AUUUA destabilizing sequence found in the 3' untranslated regions of GM-CSF and other short-lived mRNAs (Shaw and Kamen, 1986). The human immunodeficiency virus Rev protein may also interact with multiple sites within its RNA target element (Kjems et al., 1991). The regulation of *lin-14* translation in *C. elegans* appears to be mediated by small RNAs encoded by the *lin-4* gene, which exhibit complementarity to seven short repeated elements in the 3' UTR of the *lin-14* mRNA (Lee et al., 1993; Wightman et al., 1993). In *Drosophila*, the *Sex-lethal* (*Sxl*) gene product blocks male-specific splicing of the *tra* gene by binding to a single critical U-rich element located at the *tra* male-specific 3' splice site (Sosnowski et al., 1989; Inoue et al., 1990; Valcarcel et al., 1993); autoregulation by the Sxl protein of its own pre-mRNA, on the other hand, may involve interactions with multiple poly-U intron elements in the *Sxl* transcript (Sakamoto et al., 1992; Horabin and Schedl, 1993). The ERS-like purine rich exon element is somewhat degenerate and variable in length; however, it too appears to be repeated in some exons (Watakabe et al., 1993; Xu et al., 1993; Yeakley et al., 1993; Lavigueur et al., 1993).

With regard to the evolution of the splicing mechanism, the prevalence of repeated RNA motifs in various aspects of RNA metabolism suggests that the RNA sequence signals that aid splice site accuracy may often be short and redundant in function if not in sequence. The sequences at the 5' and 3' splice junctions and at the branch site, in addition to the purine-rich element mentioned above and the hexamer element described in this thesis, represent examples in splicing of short RNA sequence recognition; this recognition

clearly must occur in a collaborative fashion in order to specify accurate correct splicing in the appropriate cell type.

Taken together, these findings may address in part how constitutively spliced pre-mRNAs are correctly spliced. The recognition of primary RNA sequences may not occur as a linear contiguous sequence; instead, information may be distributed throughout a longer stretch of RNA as a collection of redundant, dispersed and/or degenerate short motifs. Only certain mosaics of motif combinations (and their associated factors) might permit splice site identification within a particular sequence region. Together with "exon definition", this method of discrimination could account for the accuracy of the splicing mechanism. Such a mechanism would not require functional sequence conservation, which may explain why the context-dependence of splice sites is still ill-defined at present. Nor would this mode of recognition require extensive sequence reiteration, as distinct sequences may possess overlapping functions with respect to the factors they bind. Although this seems like an inefficient way of specifying a splice site, such a combinatorial mode of position-independent recognition may have been the only way available to specify splice signals in the course of evolution. In this light, the regulation of EIIIB alternative splicing by highly repeated hexamer motifs represents an extreme case of distributed sequence recognition. Further work on the mechanisms of splice site selection will address whether this is an appropriate way of thinking about RNA.

In summary, our findings regarding EIIIB have allowed us to achieve a more refined understanding of splicing regulation for one vertebrate alternative exon. Perhaps more important, however, are the novel insights that these findings also offer with regard to splice site selection mechanisms in general. Extensive research has revealed a plethora of alternatively spliced

genes that are regulated in many different ways. The regulation of many of these could conceivably be a functional consequence of the collective action of multiple short sequences that are combinatorially unique to each system. Such a fascinating possibility would pose many interesting questions for future consideration.

REFERENCES

Adema, G. J., Bovenberg, R. A. L., and Baas, P. D. 1988. Unusual branchpoint selection involved in splicing of the alternatively processed calcitonin/CGRP-I pre-mRNA. Nucleic Acids Res, 16: 9513-9526.

Adema, G. J., Hulst, K. L., and Baas, P. D. 1990. Uridine branch acceptor is a cis-acting element involved in regulation of the alternative processing of calcitonin/CGRP-I pre-mRNA. Nucleic Acids Res. 18: 5365-5373.

Amara, S. G., Evans, R. M., and Rosenfeld, M. G. 1984. Calcitonin/calcitonin gene-related peptide transcription unit: tissue-specific expression involves selective use of alternative polyadenylation sites. Mol. Cell. Biol. 4: 2151-2160.

Amara, S. G., Jonas, V., Rosenfeld, M. G., Ong, E. S., and Evans, R. M. 1982. Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. Nature 298: 240-244.

Amrein, H., Hedley, M. L., and Maniatis, T. 1994. The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by Transformer-2. Cell 76: 735-746.

Arrigo, S., and Beemon, K. 1988. Regulation of Rous sarcoma virus RNA splicing and stability. Mol. Cell. Biol. 8: 4858-4867.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. 1987. Current protocols in molecular biology. (New York: John Wiley/Greene).

Baker, B. S. 1989. Sex in flies: the splice of life. Nature 340: 521-524.

Barker, G. F., and Beemon, K. 1991. Nonsense codons within the Rous sarcoma virus gag gene decrease the stability of unspliced viral RNA. Mol. Cell. Biol. 11: 2760-2768.

Barone, M. V., Henchcliffe, C., Baralle, F. E., and Paolella, G. 1989. Cell type specific trans-acting factors are involved in alternative splicing of human fibronectin pre-mRNA. EMBO J. 8: 1079-1085.

Baurén, G., and Wieslander, L. 1994. Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. Cell 76: 183-192.

Behrens, S. E., and Lührmann, R. 1991. Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. Genes Dev. 5: 1439-1452.

Bennett, M., and Reed, R. 1993. Correspondence between a mammalian spliceosome component and an essential yeast splicing factor. Science 262: 105-108.

Bennett, M., Michaud, S., Kingston, J., and Reed, R., 1992a. Protein complexes specifically associated with prespliceosome and spliceosome complexes. Genes Dev. 6: 1986-2000.

Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G., and Reed, R. 1992b. Transcript-dependent packaging of pre-mRNA in hnRNP complexes prior to spliceosome assembly in vitro. Mol. Cell. Biol. 12: 3165-3175.

Bennett, V. L., Pallante, K. M., and Adams, S. L. 1991. The splicing pattern of fibronectin mRNA changes during chondrogenesis resulting in an unusual form of the mRNA in cartilage. J. Biol. Chem. 266: 5918-5924.

Beyer, A. L., and Osheim, Y. N. 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. 2: 754-765.

Black, D. L. 1991. Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells? Genes and Dev. 5: 389-402.

Black, D. L. 1992. Activation of c-*src* neuron-specific splicing by an unusual RNA element in vivo and in vitro. Cell 69: 795-807.

Bovenberg, R. A. L., Adema, G. J., and Baas, P. D. 1988. Model for tissue-specific calcitonin/CGRP-I RNA processing from in vitro experiments. Nucleic Acids Res. 16: 7867-7883.

Broad, P. M., Symes, A. J., Thakker, R. V., and Craig, R. K. 1989. Structure and methylation of the human calcitonin/α-CGRP gene. Nucleic Acids Res. 17: 6999-7011.

Brosi, R., Gröning, K., Behrens, S. E., Lührmann, R., and Angela Krämer. 1993. Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. Science 262: 102-105.

Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A., and Housman, D. E. 1991. Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. Proc. Natl. Acad. Sci. 88: 4005-4009.

Burd, C. G., and Dreyfuss, G. 1994. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. EMBO J. 13: 1197-1204.

Burtis, K. C., and Baker, B. S. 1989. Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell 56: 997-1010.

Caceres, J. F. and Krainer, A. R. 1993. Functional analysis of the pre-mRNA splicing factor SF2/ASF structural domains. EMBO J. 12: 4715-4726.

Carmo-Fonseca, M., Pepperkok, R., Sproat, B. S., Ansorge, W., Swanson, M. S., and Lamond, A. I. 1991a. In vivo detection of snRNP-rich organelles in the nuclei of mammalian cells. EMBO J. 10: 1863-1873.

Carmo-Fonseca, M., Tollervey, D., Pepperkok, R., Barabino, S. M. L., Merdes, A., Brunner, C., Zamore, P. D., Green, M. R., Hurt, H., and Lamond, A. I. 1991b. Mammalian nuclei contain foci which are highly enriched in components of the pre-mRNA splicing machinery. EMBO J. 10: 195-206.

Carnemolla, B., Balza, E., Siri, A., Zardi, L., Nicotra, M. R., Bigotti, A., and Natali, P. G. 1989. A tumor-associated fibronectin isoform generated by alternative splicing of messenger RNA precursors. J. Cell. Biol. 108: 1139-1148.

Chain, A. C., Zollman, S., Tseng, J. C., and Laski, F. A. 1991. Identification of a *cis*-acting sequence required for germ line-specific splicing of the P element ORF2-ORF3 intron. Mol. Cell. Biol. 11: 1538-1546.

Chang, D. D., and Sharp, P. A. 1989. Regulation by HIV rev depends upon regulation of splice sites. Cell 59: 789-795.

Cheng, J. and Maquat, L. E. 1993. Nonsense codons can reduce the abundance of nuclear mRNA without affecting the abundance of pre-mRNA or the half-life of cytoplasmic mRNA. Mol. Cell. Biol. 13: 1892-1902.

Chirgwin, J. J., Przbyla, A. E., MacDonald, R., and Rutter, W. J. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18: 5294-5299.

Choi, Y. D., Grabowski, P. J., Sharp, P. A., and Dreyfuss, G. 1986. Heterogeneous ribonucleoproteins: role in RNA splicing. Science 231: 1534-1539.

Chomczynski, P., and Sacchi, N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162: 156-159.

Clackson, T., and Winter, G. 1989. 'Sticky feet'-directed mutagenesis and its application to swapping antibody domains. Nucleic Acids Res. 17: 10163-10170.

Cobianchi, F., Calvio, C., Stoppini, S., Buvoli, M., and Riva, S. 1993. Phosphorylation of human hnRNP protein A1 abrogates *in vitro* strand annealing activity. Nucleic Acids Res. 21: 949-955.

Collett, J. W. and Steele, R. E. 1993. Alternative splicing of a neural-specific src mRNA (src^+) is a rapid and protein-synthesis-independent response to neural induction in *Xenopus laevis*. Dev. Biol. 158: 487-495.

Cooper, T. A., and Ordahl, C. P. 1989. Nucleotide substitutions within the cardiac troponin T alternative exon disrupt pre-mRNA alternative splicing. Nucleic Acids Res. 17: 7905-7921.

Cooper, T. A., Cardone, M. H., and Ordahl, C. P. 1988. Cis requirements for alternative splicing of the cardiac troponin T pre-mRNA. Nucleic Acids Res. 16: 8443-8465.

Cote, G. J., Nguyen, I. N., Lips, C. J. M., Berget, S. M., and Gagel, R. F. 1991. Validation of an *in vitro* RNA processing system for CT/CGRP precursor mRNA. Nucleic Acids Res. 19: 3601-3606.

Cote, G. J., Stolow, D. T., Peleg, S., Berget, S. M., and Gagel, R. F. 1992. Identification of exon sequences and an exon-binding protein involved in alternative RNA splicing of calcitonin/CGRP. Nucleic Acids Res. 20: 2361-2366.

Crenshaw, E. B., Russo, A. F., Swanson, L. W., and Rosenfeld, M. G. 1987. Neuron-specific alternative RNA processing in transgenic mice expressing a metallothionein-calcitonin fusion gene. Cell 49: 389-398.

D'Orval, B. C., Carafa, Y. d'A., Sirand-Pugnet, P., Gallego, M., Brody, E., and Marie, J. 1991. RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts. Science 252: 1823-1828.

Delsert, C. D., and Rosenfeld, M. G. 1992. A tissue-specific small nuclear ribonucleoprotein and the regulated splicing of the calcitonin/calcitonin gene-related protein transcript. J. Biol. Chem. 267: 14573-14579.

Deshler, J. O., and Rossi, J. J. 1991. Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron. Genes Dev. 1252-1263.

DeSimone, D. W., Norton, P. A., and Hynes, R. O. 1992. Identification and characterization of alternatively spliced fibronectin mRNAs expressed in early *Xenopus* embryos. Dev. Biol. 149: 357-369.

Devereux, J., Haeberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12: 387-395.

Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. 11: 1475-1489.

Dominski, Z., and Kole, R. 1991. Selection of splice sites in pre-mRNAs with short internal exons. Mol. Cell. Biol. 11: 6075-6083.

Dominski, Z., and Kole, R. 1992. Cooperation of pre-mRNA sequence elements in splice site selection. Mol. Cell. Biol. 12: 2108-2114.

Dreyfuss, G., Matunis, M. J., Piñol-Roma, S., and Burd, C. G. 1993. hnRNP proteins and the biogenesis of mRNA. Ann. Rev. Biochem. 62: 289-321.

Emeson, R. B., Hedjran, F., Yeakley, J. M., Guise, J. W., and Rosenfeld, M. G. 1989. Alternative production of calcitonin and CGRP mRNA is regulated at the calcitonin-specific splice acceptor. Nature 341: 76-80.

Eng, F. J., and Warner, J. R. 1991. Structural basis for the regulation of splicing of a yeast messenger RNA. Cell 65: 797-804.

Engebrecht, J, Voelkel-Meiman, K., and Roeder, G. S. 1991. Meiosis-specific splicing in yeast. Cell 66: 1257-1268.

Eperon, L. C., Graham, I. R., Griffiths, A. D., and Eperon, I. C. 1988. Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? Cell 54: 393-401.

Eperon, L. P., Estibeiro, J. P., and Eperon, I. C. 1986. The role of nucleotide sequences in splice site selection in eukaryotic pre-messenger RNA. Nature 324: 280-282.

ffrench-Constant, C., and Hynes, R. O. 1989. Alternative splicing of fibronectin is temporally and spatially regulated in the chicken embryo. Development 106: 375-388.

ffrench-Constant, C., Van De Water, L., Dvorak, H., and Hynes, R. O. 1989. Reappearance of an embyonic pattern of fibronectin splicing during wound healing in the adult rat. J. Cell. Biol. 109: 903-914.

Fu, X. D. 1993. Specific commitment of different pre-mRNAs to splicing by single SR proteins. Nature 365: 82-85.

Fu, X. D., and Maniatis, T. 1992a. Isolation of a complementary DNA that encodes the mammalian splicing factor SC35. Science 256: 535-538.

Fu, X. D., and Maniatis, T. 1992b. The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. Proc. Natl. Acad. Sci. USA 89: 1725-1729.

Fu, X. D., Katz, R. A., Skalka, A. M., and Maniatis, T. M. 1991. The control of branchpoint and 3' exon sequences in the control of balanced splicing avian retrovirus RNA. Genes Dev. 5: 211-220.

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. USA 89: 11224-11228.

Fu, X. Y., Ge, H., and Manley, J. L. 1988. The role of the polypyrimidine stretch at the SV40 early pre-mRNA 3' splice site in alternative splicing. EMBO J. 7: 809-817.

Gallego, M. E., Balvay, L., and Brody, E. 1992. *Cis*-acting sequences involved in exon selection in the chicken β-tropomyosin gene. Mol. Cell. Biol. 12: 5415-5425.

Gallego, M. E., and Nadal-Ginard, B. 1990. Myosin light-chain 1/3 gene alternative splicing: cis-regulation is based upon a hierarchical compatibility between splice sites. Mol. Cell. Biol. 10: 2133-2144.

Galli, G., Guise, J. W., McDevitt, M. A., Tucker, P. W., and Nevins, J. R. 1987. Relative position and strengths of poly(A) sites as well as transcription termination are critical to membrane versus secreted μ-chain expression during B-cell development. Genes Dev. 1: 471-481.

Garcia-Blanco, M. A., Jamison, S. F., and Sharp, P. A. 1989. Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract in introns. Genes Dev. 3: 1874-1886.

Gattoni, R., Schmitt, P., and Stevenin, J. 1988. In vitro splicing of adenovirus E1A transcripts: characterization of novel reactions and of multiple branch points abnormally far from the 3' splice site. Nucleic Acids Res. 16: 2389-2409.

Ge, H., and Manley, J. L. 1990. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. Cell 62: 25-34.

Ge, H., Zuo, P., and Manley, J. L. 1991. Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. Cell 62: 25-34.

Gerke, V. and Steitz, J. A. 1986. A protein associated with small nuclear ribonucleoprotein particles recognizes the 3' splice site of premessenger RNA. Cell 47: 973-984.

Goguel, V. and Rosbash, M. 1993. Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. Cell 72: 893-901.

Goguel, V., Wang, Y., and Rosbash, M. 1993. Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing. Mol. Cell. Biol. 13: 6841-6848.

Gontarek, R. R., McNally, M. T., and Beemon, K. 1993. Mutation of an RSV intronic element abolishes both U11/U12 snRNP binding and negative regulation of splicing. Genes Dev. 7: 1926-1936.

Gorman, C. M., Rigby, P. W. J., and Lane, D. P. 1985. Negative regulation of viral enhancers in undifferentiated embryonic stem cells. Cell 42: 519-526.

Goux-Pelletan, M., Libri, D., Carafa, Y. d'A., Fiszman, M., Brody, E., and Marie, J. 1990. In vitro splicing of mutually exclusive exons from the chicken β-tropomyosin gene: role of the branch point location and very long pyrimidine stretch. EMBO J. 9: 241-249.

Grabowski, P. J., Nasim, F. H., Kuo, H. C., and Burch, R. 1991. Combinatorial splicing of exon pairs by two-site binding of U1 small nuclear ribonucleoprotein particle. Mol. Cell. Biol. 11: 5919-5928.

Graham, I. R., Hamshere, M., and Eperon, I. C. 1992. Alternative splicing of a human α-tropomyosin muscle-specific exon: identification of determining sequences. Mol. Cell. Biol. 12: 3872-3882.

Green, M. R. 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Ann. Rev. Cell Biol. 7: 559-599.

Grover, A., Rosenstraus, M. J., Sterman, B., Snook, M. E., and Adamson, E. D. 1987. An adhesive-defective variant of F9 embronal carcinoma cells fails to differentiate into visceral endoderm. Dev. Biol. 120: 1-11.

Guan, J. L., and Hynes, R. O. 1990. Lymphoid cells recognize an alternatively spliced segment of fibronectin via the integrin receptor α4β1. Cell 60: 53-61.

Guan, J. L., Trevithick, J. E., and Hynes, R. O. 1990. Retroviral expression of alternatively spliced forms of fibronectin. J. Cell. Biol. 110: 833-847.

Guo, W., and Helfman, D. M. 1993. Cis-elements involved in alternative splicing in the rat β-tropomyosin gene: the 3'-splice site of the skeletal muscle exon 7 is the major site of blockage in nonmuscle cells. Nucleic Acids Res. 21: 4762-4768.

Guo, W., Mulligan, G. J., Wormsley, S., and Helfman, D. M. 1991. Alternative splicing of β-tropomyosin pre-mRNA: cis-acting elements and cellular factors that block the use of a skeletal muscle exon in nonmuscle cells. Genes and Dev. 5: 2096-2107.

Guthrie, C. 1991. Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. Science 253: 157-163.

Hawkins, J. D. 1988. A survey on intron and exon lengths. Nucleic Acids Res. 16: 9893-9908.

Hedley, M. L., and Maniatis, T. 1991. Sex-specific splicing and polyadenylation of *dsx* pre-mRNA requires a sequence that binds specifically to Tra-2 protein in vitro. Cell 65: 579-586.

Helfman, D. M. and Ricci, W. M. 1989. Branch point selection in alternative splicing of tropomyosin pre-mRNAs. Nucleic Acids Res. 17: 5633-5650.

Helfman, D. M., Ricci, W. M., and Finn, L. A. 1988. Alternative splicing of tropomyosin pre-mRNAs in vitro and in vivo. Genes Dev. 2: 1627-1638.

Helfman, D. M., Roscigno, R. F., Mulligan, G. J., Finn, L. A., and Weber,K. S. 1990. Identification of two distinct intron elements involved in alternative splicing of β-tropomyosin pre-mRNA. Genes Dev. 4: 98-110.

Hoffman, B. E. and Grabowski, P. J. 1992. U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. Genes Dev. 6: 2554-2568.

Horabin, J. I., and Schedl, P. 1992. Regulated splicing of the *Drosophila Sex-lethal* male exon involves a blockage mechanism. Mol. Cell. Biol. 13: 1408-1414.

Horabin, J. I., and Schedl, P. 1993. *Sex-lethal* autoregulation requires multiple *cis*-acting elements upstream and downstream of the male exon and appears to depend largely on controlling use of the male exon 5' splice site. Mol. Cell. Biol. 13: 7734-7746.

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.

Humphries, M. J., Komoriya, A., Akiyama, S. K., Olden, K., and K. M. Yamada. 1987. Identification of two distinct regions of the type III connecting segment of human plasma fibronectin that promote cell type-specific adhesion. J. Biol. Chem. 262: 6886-6892.

Hynes, R. O. 1990. Fibronectins. (New York: Springer-Verlag).

Hynes, R. O. 1992. Integrins: versatility, modulation, and signaling in cell adhesion. Cell 69: 11-25.

Inoue, K., Hoshijima, K., Sakamoto, H., and Shimura, Y. 1990. Binding of the *Drosophila Sex lethal* gene product to the alternative splice site of *transformer* primary transcript. Nature 344: 461-463.

Inoue, K., Hoshijima, K., Higuchi, I., Sakamoto, H., and Shimura, Y. 1992. Binding of the *Drosophila* transformer and transformer-2 proteins to the regulatory elements of doublesex primary transcript for sex-specific RNA processing. Proc. Natl. Acad. Sci. USA 89: 8092-8096.

Kandels-Lewis, S., and Séraphin, B. 1993. Role of U6 snRNA in 5' splice site selection. Science 262: 2035-2039.

Katz, R. A., and Skalka, A. M. 1990. Control of retroviral RNA splicing through maintenance of suboptimal processing signals. Mol. Cell. Biol. 10: 696-704.

Kedes, D. H. and Steitz, J. A. 1988. Correct in vivo splicing of the mouse immunoglobulin kappa light chain pre-mRNA is dependent on 5' splice site position even in the absence of transcription. Genes Dev. 2: 1448-1459.

Kenan, D. J., Query, C. C., and Keene, J. D. 1991. RNA recognition: toward identifying determinants of specificity. Trends. Biochem. Sci. 16: 214-220.

Kim, Y. J., Zuo, P., Manley, J. L., and Baker, B. S. 1992. The *Drosophila* RNA-binding protein RBP-1 is localized to transcriptionally active sites of chromosomes and shows a functional similarity to human splicing factor ASF/SF2. Genes. Dev. 6: 2569-2579.

Kjems, J., Brown, M., Chang, D. D., and Sharp, P. A. 1991. Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element. Proc. Natl. Acad. Sci. USA 88: 683-687.

Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Lührmann, R., Garcia-Blanco, M. A., and Manley, J. L. 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature 368: 119-124.

Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K., and Baralle, F. E. 1985. Primary structure of human fibronectin: differential splicing may generate at least 10 polypeptides from a single gene. EMBO J. 4: 1755-1759.

Kornblihtt, A. R., Vibe-Pedersen, K., and Baralle, F. E. 1984. Human fibronectin: molecular cloning evidence for two mRNA species differing by an internal segment coding for a structural domain. EMBO J. 3: 221-226.

Krainer, A. R., Conway, G. C., and Kozak, D. 1990. The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. Cell 62: 35-42.

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.

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.

Krämer, A., and Utans, U. 1991. Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. EMBO J. 10: 1503-1509.

Krause, J. E., Chirgwin, J. M., Carter, M. S., Xu, Z. S., and Hershey, A. D. 1987. Three rat preprotachykinin mRNAs encode the neuropeptides substance P and neurokinin A. Proc. Natl. Acad. Sci. USA 84: 881-885.

Kunkel, T. A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. 82: 488-492.

Kuo, H. C., Nasim, F. H., and Grabowski, P. J. 1991. Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. Science 251: 1045-1050.

Laski, F. A. and Rubin, G. M. 1989. Analysis of the cis-acting requirements for germ-line-specific splicing of the P-element ORF2-ORF3 intron. Genes Dev. 3: 720-728.

Laski, F. A., Rio, D. C., and Rubin, G. M. 1986. Tissue specificity of Drosophila P-element transposition is regulated at the level of mRNA splicing. Cell 44: 7-19.

Lavigueur, A., LaBranche, H., Kornblihtt, A. R., and Chabot, B. 1993. A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. Genes Dev. 7: 2405-2417.

Lear, A. L., Eperon, L. P., Wheatley, I. M., and Eperon, I. C. 1990. Hierarchy for 5' splice site preference determined in vivo. J. Mol. Biol. 211: 103-115.

Leavitt, J., Gunning, P., Porreca, P., Ng, S. Y., Lin, C. S., and Kedes, L. 1984. Molecular cloning and characterization of mutant and wild-type human β-actin genes. Mol. Cell. Biol. 4: 1961-1969.

Lee, R. C., Feinbaum, R. L., and Ambros, V. 1993. The C. elegans heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell 75: 843-854.

Leff, S. E., Evans, R. M., and Rosenfeld, M. G. 1987. Splice commitment dictates alternative RNA processing in calcitonin/CGRP gene expression. Cell 48: 517-524.

Lemaire, M. F., and Thummel, C. S. 1990. Splicing precedes polyadenylation during *Drosophila* E74A transcription. Mol. Cell. Biol. 10: 6059-6063.

Lesser, C. F. and Guthrie, C. 1993. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. Science 262: 1982-1988.

Libri, D., Goux-Pelletan, M., Brody, E., and Fiszman, M. Y. 1990. Exon as well as intron sequences are cis-regulating elements for the mutually exclusive alternative splicing of the β tropomyosin gene. Mol. Cell. Biol. 10: 5036-5046.

Libri, D., Piseri, A., and Fiszman, M. Y. 1991. Tissue-specific splicing in vivo of the β-tropomyosin gene: dependence on an RNA secondary structure. Science 252: 1842-1845.

Magnuson, V. L., Young, M., Schattenberg, D. G., Mancini, M. A., Chen, D., Steffensen, B., and Klebe, R. J. 1991. The alternative splicing of fibronectin pre-mRNA is altered during aging and in response to growth factors. J. Biol. Chem. 266: 14654-14662.

Malim, M. H., Hauber, J., Le, S. Y., Maizel, J. V., and Cullen, B. R. 1989. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature 338: 254-257.

Maniatis, T. 1991. Mechanisms of alternative pre-mRNA splicing. Science 251: 33-34.

Mardon, H. J., Grant, R. P., Grant, K. E., and Harris, H. 1993. Fibronectin splice variants are differentially incorporated into the extraceellular matrix of tumorigenic and non-tumorigenic hybrids between normal fibroblasts and sarcoma cells. J. Cell. Sci. 104: 783-792.

Mardon, H. J., Sebastio, G., and Baralle, F. E. 1987. A role for exon sequences in alternative splicing of the human fibronectin gene. Nucleic Acids Res. 15: 7725-7733.

Mardon, H. J., and Sebastio, G. 1992. Regulation of alternative splicing in the IIICS region of human fibronectin pre-mRNA encoding cell binding sites CS1 and CS5. J. Cell. Sci. 103: 423-433.

Mattox, W. and Baker, B. S. 1991. Autoregulation of the splicing of transcripts from the *transformer-2* gene of *Drosophila* Genes Dev. 5: 786-796.

Mattox, W., Ryner, L., and Baker, B. S. 1992. Autoregulation and multifunctionality among *trans*-acting factors that regulate alternative pre-mRNA processing. J. Biol. Chem. 267: 19023-19026.

Matunis, E. L., Matunis, M. J., and Dreyfuss, G. 1993. Association of individual hnRNP proteins and snRNPs with nascent transcripts. J. Cell. Biol. 121: 219-228.

Mayeda, A. and Krainer, A. R. 1992. Regulation of alternative pre-mRNA splicing by hnRNP and splicing factor SF2. Cell 69: 365-375.

Mayeda, A., Helfman, D. M., and Krainer, A. R. 1993. Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. Mol. Cell. Biol. 13: 2993-3001.

Mayeda, A. and Ohshima, Y. 1988. Short donor sequences inserted within the intron of β-globin pre-mRNA serve for splicing in vitro. Mol. Cell. Biol. 8: 4484-4491.

Mayeda, A., Zahler, A. M., Krainer, A. R., and Roth, M. B. 1992. Two members of a conserved family of nuclear phosphoproteins are involved in pre-mRNA splicing. Proc. Natl. Acad. Sci. USA 89: 1301-1304.

Mayrand, S. H., Dwen, P., and Pederson. T. 1993. Serine/threonine phosphorylation regulates binding of C hnRNP proteins to pre-mRNA. Proc. Natl. Acad. Sci. USA 90: 7764-7768.

McKeown, M. 1992. Alternative mRNA splicing. Ann. Rev. Cell Biol. 8: 133-155.

McNally, M. T., Gontarek, R. R., and Beemon, K. 1991. Characterization of Rous sarcoma virus intronic sequences that negatively regulate splicing. Virology 185: 99-108.

Moore, M. J., Query, C. C., and Sharp, P. A. 1993. Splicing of precursors to mRNA by the spliceosome. In The RNA World, R. Gesteland and J. Atkins, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 303-357.

Mullen, M. P., Smith, C. W. J., Patton, J. G., and Nadal-Ginard, B. 1991. α-Tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine tracts determines default exon choice. Genes and Dev. 5: 642-655.

Mulligan, G. J., Guo, W., Wormsley, S., and Helfman, D. M. 1992. Polypyrimidine tract binding protein interacts with sequences involved in alternative splicing of β-tropomyosin pre-mRNA. J. Biol. Chem. 267: 25480-25487.

Nagoshi, R. N., and Baker, B. S. 1990. Regulation of sex-specific RNA splicing at the Drosophila doublesex gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. Genes Dev. 4: 89-97.

Nandabalan, K., Price, L., and Roeder, G. S. 1993. Mutations in U1 snRNA bypass the requirement for a cell type-specific RNA splicing factor. Cell 73: 407-415.

Nasim, F. H., Spears, P. A., Hoffman, H. M., Kuo, H. C., and Grabowski, P. J. 1990. A sequential splicing mechanism promotes selection of an optional exon by repositioning a downstream 5' splice site in preprotachykinin pre-mRNA. Genes Dev. 4: 1172-1184.

Neel, H., Weil, D., Giansante, C., and Dautry, F. 1993. In vivo cooperation between introns during pre-mRNA processing. Genes Dev. 7: 2194-2205.

Nelson, K. K., and Green, M. R. 1988. Splice site selection and ribonucleoprotein complex assembly during in vitro pre-mRNA splicing. Genes Dev. 2: 319-329.

Nelson, K. K., and Green, M. R. 1989. Mammalian U2 snRNP has a sequence-specific binding activity. Genes Dev. 2: 319-329.

Newman, A. and Norman, C. 1991. Mutations in yeast U5 snRNA alter the specificity of 5' splice site cleavage. Cell 65: 115-123.

Noble, J. C. S., Prives, C., and Manley, J. L. 1988. Alternative splicing of SV40 early pre-mRNA is determined by branch site selection. Genes and Dev. 2: 1460-1475.

Norton, P. A., and Hynes, R. O. 1987. Alternative splicing of chicken fibronectin in embryos and in normal and transformed cells. Mol. Cell. Biol. 7: 4297-4307.

Norton, P. A., and Hynes, R. O. 1990. In vitro splicing of fibronectin pre-mRNAs. Nucleic Acids Res. 18: 4089-4097.

Norton, P. A., and Hynes, R. O. 1993. Characterization of HeLa nuclear factors which interact with a conditionally processed rat fibronectin pre-mRNA. Biochem. Biophys. Res. Comm. 195: 215-221.

Pagani, F., Zagato, L., Vergani, C., Casari, G., Sidoli, A., and Baralle, F. E. 1991. Tissue-specific splicing pattern of fibronectin messenger RNA precursor during development and aging in rat. J. Cell Biol. 113: 1223-1229.

Paolella, G., Henchcliffe, C., Sebastio, G., and Baralle, F. E. 1988. Sequence analysis and in vivo expression show that alternative splicing of ED-B and ED-A regions of the human fibronectin gene are independent events. Nucleic Acids Res. 16: 3545-3557.

Patton, J. G., Mayer, S. A., Tempst, P., and Nadal-Ginard, B. 1991. Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing. Genes Dev. 5: 1237-1251.

Patton, J. G., Porro, E. B., Galceran, J., Tempst, P., and Nadal-Ginard, B. 1993. Cloning and characterization of PSF, a novel pre-mRNA splicing factor. Genes Dev. 7: 393-406.

Paul, J. I., Schwartzbauer, J. E., Tamkun, J. W., and Hynes, R. O. 1986. Cell-type-specific fibronectin subunits generated by alternative splicing. J. Biol. Chem. 261: 12258-12265.

Petersen, T. E., Thøgersen, H. C., Skorstengaard, K., Vibe-Pedersen, K., Sahl, P., Sottrup-Jensen, L., and Magnusson, S. 1983. Partial primary structure of bovine plasma fibronectin. three types of internal homology. Proc. Natl. Acad. Sci. USA 80: 137-141.

Peterson, M. L., and Perry, R. P. 1986. Regulated expression of μ_m and μ_s mRNA requires linkage of the poly(A) addition sites and is dependent on the length of the μ_m-μ_s intron. Proc. Natl. Acad. Sci. USA 83: 8883-8887.

Pikielny, C. W. and Rosbash, M. 1985. mRNA splicing efficiency and the contribution of nonconserved sequences. Cell 41: 119-126.

Piñol-Roma, S., Choi, Y. D., Matunis, M. J., and Dreyfuss, G. 1988. Immunopurification and heterogeneous nuclear ribonucleoproteins particles reveals and assortment of RNA-binding proteins. Genes Dev. 2: 215-227.

Pulak, R. and Anderson, P. 1993. mRNA surveillance by the *Caenorhabditis elegans smg* genes. Genes Dev. 7: 1885-1897.

Query, C. C., Bentley, R. C., and Keene, J. D. 1989. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. Cell 57: 89-101.

Reed, R. 1989. The organization of 3' splice site sequences in mammalian introns. Genes Dev. 3: 2113-2123.

Reed, R., and Maniatis, T. 1986. A role for exon sequences and splice-site proximity in splice-site selection. Cell 46: 681-690.

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

Roesser, J. R., Liittschwager, K., and Leff, S. E. 1993. Regulation of tissue-specific splicing of the calcitonin/calcitonin gene-related peptide gene by RNA-binding proteins. J. Biol. Chem. 268: 8366-8375.

Ruskin, B., Zamore, P. D., and Green, M. R. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell 53: 207-219.

Ryner, L. C., and Baker, B. S. 1991. Regulation of doublesex pre-mRNA processing occurs by 3'-splice site activation. Genes and Dev. 5: 2071-2085.

Saga, Y., Lee, J. S., Saraiya, C., and Boyse, E. A. 1990. Regulation of alternative splicing in the generation of isoforms of the mouse Ly-5 (CD45) glycoprotein. Proc. Natl. Acad. Sci. 87: 3728-3732.

Sakamoto, H., Inoue, K., Higuchi, I., Ono, Y., and Shimura, Y. 1992. Control of *Drosophila Sex-lethal* splicing by its own female-specific product. Nucleic Acids Res. 20: 5533-5540.

Sambrook, J., Fritsch, E. F., and Maniatis, T. 1989. Molecular cloning: a laboratory manual. 2nd ed. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Sawa, H. and Shimura, Y. 1992. Association of U6 snRNA with the 5' splice site region of pre-mRNA in the spliceosome. Genes Dev. 6: 244-254.

Schmitt, P., Gattoni, R., Keohavong, P., and Stévenin, J. 1987. Alternative splicing of E1A transcripts of adenovirus requires appropriate ionic conditions in vitro. Cell 50: 31-39.

Schwarzbauer, J. E., Patel, R. S., Fonda, D., and Hynes, R. O. 1987. Multiple sites of alternative splicing of the rat fibronectin gene transcript. EMBO J. 6: 2573-2580.

Schwarzbauer, J. E., Tamkun, J. W., Lemischka, I., and Hynes, R. O. 1983. Three different fibronectin mRNAs arise by alternative splicing within the coding region. Cell 35: 421-431.

Shaw, G., and Kamen, R. 1986. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell 46: 659-667.

Siebel, C. W. and Rio, D. C. 1990. Regulated splicing of the *Drosophila* P transposable element third intron in vitro: somatic repression. Science 248: 1200-1208.

Siebel, C. W., Fresco, L. D., and Rio, D. C. 1992. The mechanism of somatic inhibition of *Drosophila* P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. Genes Dev. 6: 1386-1401.

Sierakowska, H., Szer, W., Furdon, P. J., and Kole, R. 1986. Antibodies to hnRNP core proteins inhibit in vitro splicing of human β-globin pre-mRNA. Nucleic Acids Res. 14: 5241-5254.

Siliciano, P. G. and Guthrie, C. 1988. 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. Genes Dev. 2: 1258-1267.

Smith, C. W. J., and Nadal-Ginard, B. 1989. Mutually exclusive splicing of α-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing. Cell 56: 749-758.

Smith, C. W. J., Patton, J. G., and Nadal-Ginard, B. 1989a. Alternative splicing in the control of gene expression. Ann. Rev. Genet. 23: 537-577.

Smith, C. W. J., Porro, E. B., Patton, J. G., and Nadal-Ginard, B. 1989b. Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. Nature 342: 243-247.

Solnick, D. 1985. Alternative splicing caused by RNA secondary structure. Cell 43: 667-676.

Somasekhar, M. B., and Mertz, J. E. 1985. Exon mutations that affect the choice of splice sites used in processing the SV40 late transcripts. Nucleic Acids Res. 13: 5591-5609.

Sontheimer, E. J., and Steitz, J. A. 1993. The U5 and U6 small nuclear RNAs as the active site components of the spliceosome. Science 262: 1989-1996.

Sosnowski, B. A., Belote, J. M., and McKeown, M. 1989. Sex-specific alternative splicing of RNA from the *transformer* gene results from sequence-dependent splice site blockage. Cell 58: 449-459.

Spector, D. L. 1990. Higher order nuclear organization: three-dimensional distribution of small nuclear ribonucleoprotein particles. Proc. Natl. Acad. Sci. USA 87: 147-151.

Spector, D. L., Fu, X. D., and Maniatis, T. 1991. Associations between distinct pre-mRNA splicing components and the cell nucleus. EMBO J. 10: 3467-3481.

Staden, R. 1990. Searching for patterns in protein and nucleic acid sequences. Meth. Enzymol. 183: 193-211.

Steitz, J. A. 1992. Splicing takes a Holliday. Science 257: 888-889.

Sterner, D. A. and Berget, S. M. 1993. In vivo recognition of a vertebrate mini-exon as an exon-intron-exon unit. Mol. Cell. Biol. 13: 2677-2687.

Streuli, M. and Saito, H. 1989. Regulation of tissue-specific alternative splicing: exon-specific cis-elements govern the splicing of leukocyte common antigen pre-mRNA. EMBO J. 8: 787-796.

Strickland, S., Smith, K. K., and Marroti, K. R. 1980. Hormonal induction of differentiation in teratocarcinoma stem cells: generation of parietal endoderm by retinoic acid and dibutyryl cAMP. Cell 21: 347-355.

Sun, Q., Hampson, R., and Rottman, F. M. 1993a. In vitro analysis of bovine growth hormone pre-mRNA alternative splicing. J. Biol. Chem. 268: 15659-15666.

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.

Swanson, M. S., and Dreyfuss, G. 1988a. Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. Mol. Cell. Biol. 8: 2237-2241.

Swanson, M. S., and Dreyfuss, G. 1988b. RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. EMBO J. 7: 3519-3529.

Tacke, R., and Goridis, C. 1991. Alternative splicing in the neural cell adhesion molecule pre-mRNA: regulation of exon 18 skipping depends upon the 5' splice site. Genes and Dev. 5: 1416-1429.

Talerico, M., and Berget, S. M. 1990. Effect of 5' splice site mutations on splicing of the preceding intron. Mol. Cell. Biol. 10: 6299-6305.

Tamkun, J. W., Schwarzbauer, J. E., and Hynes, R. O. 1984. A single rat fibronectin gene generates three different mRNAs by alternative splicing of a complex exon. Proc. Natl. Acad. Sci. USA 81: 5140-5144.

Tanaka, K., Watakabe, A., and Shimura, Y. 1994. Polypurine sequences within a downstream exon function as a splicing enhancer. Mol. Cell. Biol. 14: 1347-1354.

Tazi, J., Alibert, C., Temsamani, J., Reveillaud, I., Cathala, G., Brunel, C., and Jeanteur, P. 1986. A protein that specifically recognizes the 3' splice site of mammalian pre-mRNA introns is associated with a small nuclear ribonucleoprotein. Cell 47: 755-766.

Tian, M., and Maniatis, T. 1992. Positive control of pre-mRNA splicing in vitro. Science 256: 237-240.

Tian, M., and Maniatis, T. 1993. A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. Cell 74: 105-114.

Tsai, A. Y. M., Streuli, M., and Saito, H. 1989. Integrity of the exon 6 sequence is essential for tissue-specific alternative splicing of the human leukocyte common antigen pre-mRNA. Mol. Cell Biol. 9: 4550-4555.

Tsurushita, N. and Korn, L. J. 1988. Nuclear factors in B lymphoma enhance splicing of mouse membrane-bound μ mRNA in *Xenopus* oocytes. Science 239: 494-497.

Ulfendahl, P. J., Kreivi, J. P., and Akusjärvi, G. 1989. Role of the branch site/3' splice site region in adenovirus-2 E1A pre-mRNA alternative splicing: evidence for 5' and 3' splice site cooperation. Nucleic Acids Res. 17: 925-937.

Urlaub, G., Mitchell, P. J., Ciudad, C. J., and Chasin, L. A. 1989. Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. Mol. Cell. Biol. 9: 2868-2880.

Valcárcel, 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.

van Oers, C. C. M., Adema, G. J., Zandberg, H., Moen, T. C., and Baas, P. D. 1994. Two different sequence elements within exon 4 are necessary for calcitonin-specific splicing of the human calcitonin/calcitonin gene-related peptide I pre-mRNA. Mol. Cell. Biol. 14: 951-960.

Vibe-Pedersen, K., Kornblihtt, A. R., and Baralle, F. E. 1984. Expression of a human α-globin/fibronectin gene hybrid generates two mRNAs by alternative splicing. EMBO J. 3: 2511-2516.

Vilardell, J. and Warner, J. R. 1994. Regulation of splicing at an intermediate step in the formation of the spliceosome. Genes Dev. 8: 211-220.

Watakabe, A., Tanaka, K., and Shimura, Y. 1993. The role of exon sequences in splice site selection. Genes Dev. 7: 407-418.

Wayner, E. A., Garcia-Pardo, A., Humphries, M. J., McDonald, J. A., and Carter, W. G. 1989. Identification and characterization of the T lymphocyte adhesion receptor for an alternative cell attachment domain (CS-1) in plasma fibronectin. J. Cell. Biol. 109: 1321-1330.

Wightman, B., Ha, I., and Ruvkun, G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates tremporal pattern formation in C. elegans. Cell 75: 855-862.

Wise, J. A. 1993. Guides to the heart of the spliceosome. Science 262: 1978-1979.

Wu, J., and Manley, J. L. 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes. Dev. 3: 1552-1561.

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

Wyatt, J. R., Sontheimer, E. J., and Steitz, J. A. 1992. Site-specific crosslinking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. Genes Dev. 6: 2542-2553.

Xing, Y. and Lawrence, J. B. 1993. Nuclear RNA tracks: structural basis for transcription and splicing? Trends Cell Biol. 3: 346-353.

Xu, R., Teng, J., and Cooper, T. A. 1993. The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. Mol. Cell. Biol. 13: 3660-3674.

Yeakley, J. M., Hedjran, F., Morfin, J. P., Merillat, N., Rosenfeld, M. G., and Emeson, R. B. 1993. Control of calcitonin/calcitonin gene-related peptide pre-mRNA processing by constitutive intron and exon elements. Mol. Cell. Biol. 13: 5999-6011.

Zachar, Z., Chou, T. B., and Bingham, P. M. 1987. Evidence that a regulatory gene autoregulates splicing of its transcript. EMBO J. 6: 4105-4111.

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 and Dev. 6: 837-847.

Zahler, A. M., Neugebauer, K. M., Lane, W. S., and Roth, M. B. 1993a. Distinct functions of SR proteins in alternative pre-mRNA splicing. Science 260: 219-222.

Zahler, A. M., Neugebauer, K. M., Stolk, J. A., and Roth, M. B. 1993b. Human SR proteins and isolation of a cDNA encoding SRp75. Mol. Cell. Biol. 13: 4023-4028.

Zamore, P. D. and Green, M. R. 1991. Biochemical characterization of the U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. EMBO J. 10: 207-214.

Zamore, P. D., Patton, J. G., and Green, M. R. 1992. Cloning and domain structure of the mammalian splicing factor U2AF. Nature 355: 609-614.

Zardi, L., B. Carnemolla, A. Siri, T. E. Peterson, G. Paolella, G. Sebastio, and F. E. Baralle. 1987. Transformed human cells produce a new fibronectin isoform by preferential alternative splicing of a previously unobserved exon. EMBO J. 6: 2337-2342.

Zhuang, Y. and Weiner, A. M. 1986. A compensatory base change in human U1 snRNA suppresses a 5' splice site mutation. Cell 46: 827-835.

Zhuang, Y. and Weiner, A. M. 1989. A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev. 3: 1545-1552.

Zuo, P. and Manley, J. L. 1993. Functional domains of the human splicing factor ASF/SF2. EMBO J. 12: 4727-473

3636-26