MOLECULAR MECHANISMS REGULATING THE ALTERNATIVE SPLICING OF THE HYALURONAN BINDING PROTEIN, CD44

by

JULIE FAYE DIRKS

Bsc., University of British Columbia, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Genetics Programme)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

November 8, 1996

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Genetics

The University of British Columbia
Vancouver, Canada

Date Dec 6 1996

ABSTRACT

CD44 is an extensively glycosylated, widely expressed cell surface protein found on most cells. Multiple protein isoforms of CD44 are generated by the alternative splicing of one or more of 10 contiguous exons into a single site in the processed mRNA, resulting in the inclusion of additional amino acid sequence into the extracellular domain. The expression of certain CD44 isofoms such as CD44R1, which contains additional exons v8, v9 and v10, has been clinically correlated with an increase in tumour progression and metastatic potential. However, the molecular mechanisms which regulate the alternative splicing of CD44 exons, and thus modulate protein expression, are unknown.

A novel *in vivo* pre-mRNA splicing assay, termed the splice activated gene expression (SAGE) assay, was developed to investigate the pre-mRNA splicing of the CD44R1 variant region exons. This assay consists of a stably expressed *in vivo* splice construct composed of two or more exons of CD44R1 along with their accompanying intron(s) (v8-v9, v9-v10 or v8-v9-v10) placed in frame between the CD44 leader sequence, and a leaderless reporter construct of human liver/bone/kidney alkaline phosphatase (ALP). The expression and enzymatic activity of the ALP is thus dependent upon the prior correct splicing of the CD44 exons. Pre-mRNA splicing is assessed by either enzymatic assay, fluorescence activated cell sorting (FACS) or reverse transcription and polymerase chain reaction (RT PCR).

Using this assay system, combined with sequencing of the v8-v9-v10 genomic regions, alternative 5' and 3' splice sites were identified within the intron (intron 9) which separates exons v9 and v10. Splicing reactions which utilized these intronic sites, either together or in conjunction with the exonic v10-3' splice site, resulted in the production of aberrant splice

products which retained part of intron 9 including an in frame stop codon. Examination of several leukemic cell lines confirmed the presence and use of these intronic splice sites during native CD44 splicing. Similar intronic splice sites were not found within intron 8. Incubation of K562 cells expressing the SAGE constructs with the serine/threonine kinase inhibitor staurosporine, resulted in an increase in exonic splice site usage at the expense of the intronic splice sites. Thus the activity of several serine/threonine kinases, such as protein kinase C are implicated in the molecular control of CD44 alternative exon splicing.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS:

ACT	actin
ALP	alkaline phosphatase
ATP	adenosine triphosphate
BCIP/INT	5-bromo-4-chloro-3-indolyl phosphate/p-iodonitrotetrzolium
bps	branch point sequence
BS KS+	bluescript KS+
BS+L	bluescript KS+ with CD44 leader
CC/E	commitment complex/early
CCII	cool calf II
CKII	casein kinase II
CMV	cytomegalovirus
CT/CGRP	calcitonin/calcitonin gene-related peptide
dCTP	deoxy-cytidine triphosphate
DEAE dextran	diethylaminoethyl dextran
DMEM	Dulbecco's modified eagle medium
DNA	deoxyribonucleic acid
DNAse I	deoxyribonuclease I
dNTP	deoxy-nucleoside triphosphate
DSX	doublesex
DTT	dithiothreitol
EBNA-1	Epstein Barr nuclear antigen-1
EBV	Epstein Barr virus
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
ERM	ezrin/radixin/myosin
ESE	exon splicing enhancer
FITC	fluorescein isothiocyanate
GAG	glycosaminoglycan
GPI	glycophosphoinositol
HA	hyaluronan
HAT	hypoxanthine/aminopterin/thymidine
HBSS	Hank's balanced salt solution
HIV-1	human immunodeficiency virus-1
hnRNP	heteogeneous nuclear ribonucleoprotein
HSV	herpes simplex virus
I9	intron 9
ICAM-1	intracellular adhesion molecule-1
IL-1	interleukin-1
IPTG	isopropylthio-β-d-galactoside
ISE	intron splice enhancer
kb	kilobase
LBA	Luria broth agar

LFA-1	leukocyte function-associated antigen-1
Mab	monoclonal antibody
MIP-1β	macrophage inflammatory protein-1β
MMLV	Muloney murine leukemia virus
MOPS	3-[N-Morpholino]propane sulfonic acid
nt	nucleotide
OPI	oxaloacetate/pyruvate/insulin
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PdBU	phorbol di-butyrate
PHA	phytohemagglutinin
PKC	protein kinase C
PTB	polypyrimidine tract binding protein
RBD	RNA binding domain
RNA	ribonucleic acid
RNAse	ribonuclease
RRM	RNA recognition motif
RT PCR	reverse transcription polymerase chain reaction
S/T kinase	serine/threonin kinase
SAGE	splice activated gene expression
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SF2/ASF	splice factor-2/alternative splice factor
snRNA	small nuclear ribonucleic acid
snRNP	small nuclear ribonucleoprotein
SR protein	serine/arginine protein
SRPK1	SR protein kinase-1
SSC	salt saturated citrate
SV40	simian virus 40
SXL	sexlethal
TcR	T cell receptor
TK	thymidine kinase
TNF-α	tumour necrosis factor-α
TPA	12-o-tetradecanoyl phorbol-13-acetate
TRA	transformer
TRA-2	transformer-2
U2AF	U2 associated factor
v8,v9,v10	variant8/variant9/variant10
X-gal	5-bromo-4-chloro-3-indolyl-β-d-galactoside

ACKNOWLEDGMENTS

"What your mind can conceive,
what your heart will believe,
so shall you achieve."

Napoleon Hill

"Never give up"

Winston Churchill

The road from there to here has been long. Sometimes the way has been full of bumps and potholes, and sometimes it has been smooth. There has been rain, even thundershowers, but the sun has also shone. I would like to thank all of those who have journeyed with me, sharing in the hardships and in the joy.

My parents, Cec and Karen Memory, have always believed in me and encouraged me to stretch myself and acheive new heights.

My friends and fellow bench slaves have listened to me, encouraged me and challenged me to try new and unfamiliar paths. I would especially like to thank Anne Wallis, who began this adventure, and Kristian Baker, to whom I pass the baton.

The members of my supervisory committee, Drs. Connie Eaves, Fumio Takei and Jurgen Vielkind, as well as many other senior scientists at the Terry Fox Laboratory have listened to my questions and pointed out the way on the map.

The journey would not have been possible without the constant help and support of my supervisor, Dr. Graeme Dougherty. He has stimulated me, motivated me, and guided me. Thank you for teaching me to walk on the hairy edge of science.

This thesis is dedicated to my husband, Geoff Dirks. He has been my rock and my strength through it all. Without his continual encouragement I would have given up long ago. Now we begin a new journey together, and I am looking forward to the challenge.

Sincerely,

Julie F. Dirks

I INTRODUCTION

The molecular mechanisms facilitating the metastatic spread of neoplastic cells from a primary tumour mass to a secondary site have yet to be fully defined. In general this process is thought to result from the continued accumulation of alterations in the genes controlling multiple cellular processes such as locomotion and adhesion (1). To accomplish metastasis, the malignant cell must acquire several functional capabilities reminiscent of circulating hematopoietic cells (2). That is, a metastatic tumour cell must detach from the parental mass, pass through the extracellular matrix, and enter the peripheral circulation through either the draining lymphatic system or by migration through the endothelial cell lining of the walls of the venous or arterial vessels (3). Once a metastatic cell reaches the peripheral circulation it must withstand the hydrostatic pressure and shear forces associated with blood flow. Either singly or more often aggregated with other metastatic cells, it then must lodge in a capillary bed and extravasate into the peripheral tissues (4). Although multiple genetic events control each of these processes, alterations in the cell surface expression and/or functional activity of adhesion molecules has been suggested to play a major role in determining the acquisition of metastatic capacity as well as the final secondary site of the metastases (2).

Early work to define the cell surface adhesion proteins participating in tumour metastasis suggested a connection between the expression of the cell surface glycoprotein, CD44 and metastatic propensity (3). Initially, CD44 had been identified as a widely expressed cell surface molecule (5) with an observed size heterogeneity thought to be caused by extensive glycosylation (6). Subsequent cloning and sequencing of a number of different CD44 cDNAs revealed that the protein was expressed as a family of isoforms generated by the alternative pre-mRNA splicing of one or more of a set of ten contiguous exons found within the *CD44* gene (7-9). The unique function of each of these isoforms has yet to be identified.

However, the expression of individual CD44 isoforms has been causally linked to metastasis. Using monoclonal antibodies (mAbs) developed against a metastatic rat pancreatic carcinoma cell line, Gunthert *et al.* (4) identified a large molecular weight isoform of CD44 named pMeta-1, which conferred metastatic capability upon a non-metastatic variant of the same cell line. This isoform differed from the parental form by the addition of only 3 of the 10 possible additional exons and was also expressed by normal tissues, implying a link between inappropriate expression of pMeta-1 and metastasis (10). Since this initial finding, a number of other CD44 variants have been described which are associated with either primary or metastatic disease as well as other cellular processes (4, 5, 11, 12).

1. Structure And Function Of CD44

Although CD44 was initially described as a molecule expressed by hemopoietic cells, it is now known to be present on almost every cell of the body including: macrophages, granulocytes, lymphocytes, erythrocytes, epithelial, endothelial, neural and glial cells (13). In addition to its role in metastatic potential, a number of cellular functions have been demonstrated for CD44. Crosslinking of CD44 proteins by mAbs induced signaling events which enhanced CD2 or CD3 mediated T cell activation (14-17), stimulated the release of Interleukin-1 (IL-1) from monocytes (18), and regulated lymphocyte function-associated antigen-1/intracellular adhesion molecule-1 (LFA-1/ICAM-1) mediated adhesion (19, 20). Additionally, a role for CD44 has also been demonstrated in a variety of adhesion-related processes such as facilitation of the cell surface attachment of HIV-1 (21), modulation of the poliovirus receptor (22), interactions between early hemopoietic progenitors and the bone marrow stroma (23, 24), and binding of and movement though the extracellular matrix (25).

1.1 The structure of CD44

The structure of the CD44 molecule, solved from both a biochemical and genetic viewpoint., is that of a typical type I membrane protein (Figure 1; 26). The amino-terminal extracellular domain is composed of 248 amino acids (aa) which contain two repeated motifs with sequence similarity to cartilage link protein. The first third of the molecule contains six cysteine residues forming three possible disulfide bonds and resulting in a globular 'head' on a protein 'stalk' (27). The extracellular domain is followed by a typical membrane spanning domain of 21 aa and a cytoplasmic domain of 72 aa (26). Based on amino acid content, CD44 should resolve on SDS-PAGE with an apparent molecular mass of 37 kD. However, sizes ranging from 80 to 200 kD have been reported and are due in part to extensive N- and O-linked glycosylation (6). Furthermore, biochemical analysis has demonstrated that the higher molecular mass forms (180-200 kD) may be generated by the addition of large molecular weight glycosaminoglycans (GAGs) such as chondroitin sulphate, keratan sulphate, or heparan sulphate, to one of 4 serine-glycine dipeptide residues (27).

The cytoplasmic domain of CD44 may be phosphorylated on at least 2 serine residues (S325 and S327; 28, 29). Additionally, this domain has been demonstrated to interact with several components of the cytoskeleton such as ankyrin (30, 31), vimentin (32), and the ezrin/radixin/myosin (ERM) family of actin binding proteins (33). Immunoprecipitation studies have indicated that several kinases such as members of the serine/threonine kinase family protein kinase C (PKC; 34) and the tyrosine kinase $p56^{lck}$ (35) may also directly or indirectly interact with the CD44 cytoplasmic domain, and thus imply a link for CD44 into various signal transduction pathways.

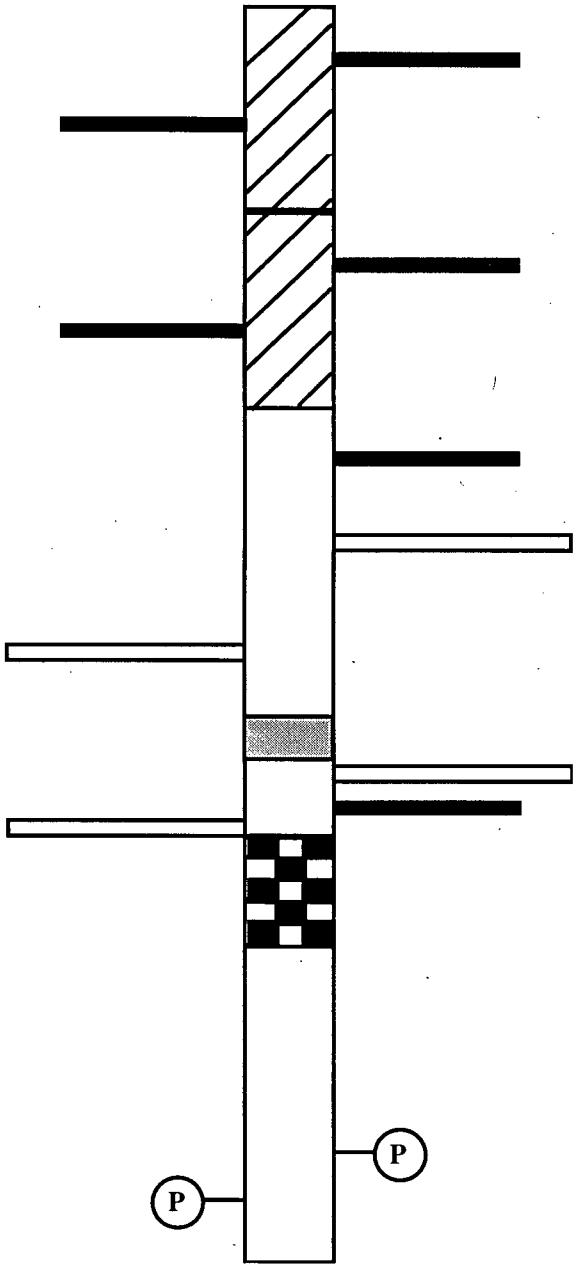


Figure 1: Cartoon representation of CD44 protein structure. Hatched boxes at the amino terminal indicate the regions of similarity to cartiledge link protein through which CD44 has been demonstrated to bind hyaluronan. Open boxes represent the structural protein core and checked box representing the transmembrane domain. Insertion of unique region variant exons occurs at a single site indicated by the grey box. N-linked glycosylation and sites for potential chondriotin sulphate sidechain modification are indicated by filled and open bars respectively. Numerous sites of O-linked glycosylation exist but are not shown for clarity. At least two serine residues within the cytoplasmic domain are phosphorylated, indicated by a circled P.

In addition to size heterogeneity generated by differential glycosylation of the CD44 core protein, further diversity is achieved by alternative splicing of CD44 pre-mRNA exons. The genomic region which encodes CD44 is found on chromosome 11p (36) and consists of 20 exons spread over 50 kb of DNA (Figure 2; 8, 9). The first 5 exons of CD44 encode the leader sequence and amino terminal end of the protein and also encode the putative common carbohydrate binding domain repeated within exons 2 and 5 (37). The 10 contiguous alternatively spliced exons follow next, encoding the CD44 unique or inserted region (7). These 10 variant exons are spliced into a single site within the pre-mRNA, resulting in additional amino acid sequence within the extracellular domain, proximal to the transmembrane domain (7). This region is followed by 3 common exons coding for the transmembrane domain. The last 2 exons are also alternatively spliced and encode either a full length (72 aa) or a truncated (3 aa) cytoplasmic domain (9). Some variant exons contain further alternative splice donor (exon 5) or splice acceptor (exon 8 or v3) sites which are utilized to generate further diversity (9). Recent work by Matsumura *et al.* (38) has also demonstrated the possible existence of a further exon found within the intron which separates the two variant region exons 8 and 9 (v3 and v4), termed exon 9a with the current exon 9 (v4) renamed exon 9b.

The predominant 80-90 kD CD44H isoform expressed by most cells does not contain any variant exons. The CD44v8-10 (CD44R1) isoform is found on cells of mesenchymal origin and contains additional exons 13, 14 and 15 (v8, v9 and v10), resulting in a protein with an apparent molecular mass of approximately 130 kD (39). Isoforms which contain a truncated cytoplasmic tail have been found at very low levels within certain cell types (27), and at this time no variant isoforms containing exon 6 (v1) have been described (40).

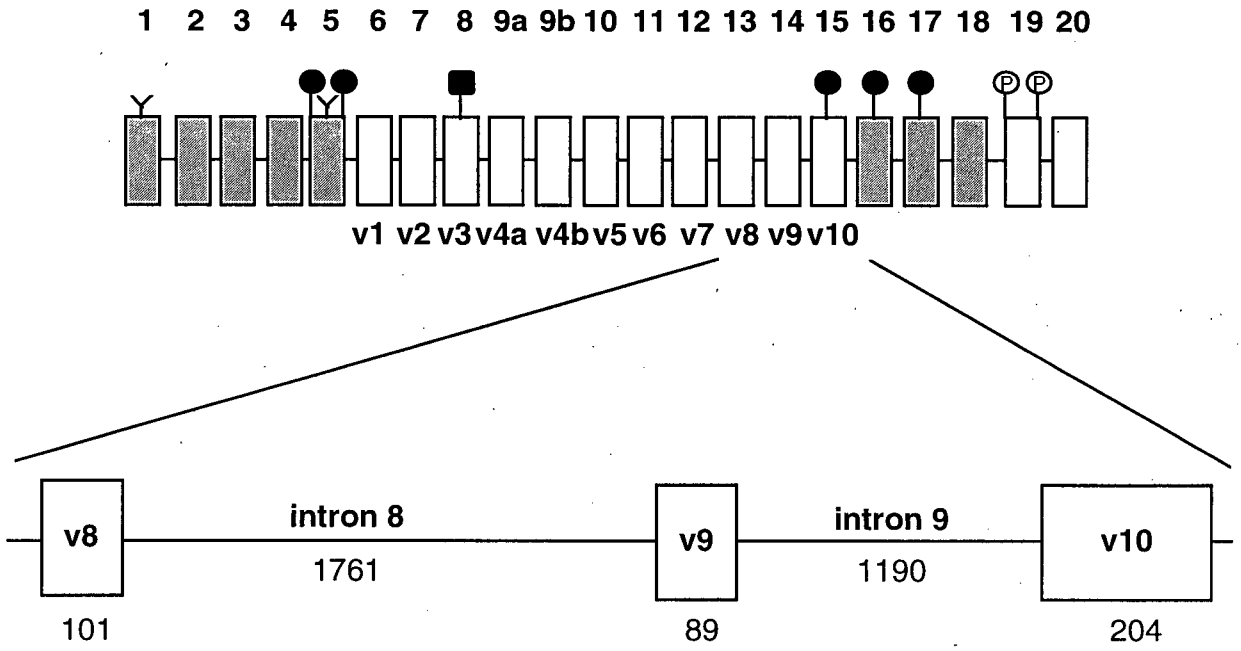


Figure 2: Constitutive (shaded boxes) and alternatively spliced (open boxes) exons of the *CD44* gene separated are by introns (lines) as described in the text. Genomic numbering is shown above and variant exon nomenclature is shown below. For clarity, the variant region nomenclature will be used throughout this work. Carbohydrate or hyaluronan recognition domains are indicated by a Y. Sites of possible chondroitin sulphate modification are indicated by closed circles, heparan sulphate is indicated by closed square, and sites of phosphorylation are indicated by a circled P. The expanded segment depicts the genomic region spanning exons v8, v9 and 10, with size in base pairs noted below (determined in Results section 1.1). The intron separating exons v8 and v9 is denoted as intron 8. Similarily, the intron separating exons v9 and v10 is denoted intron 9. Adapted from (41).

Clearly, multiple CD44 isoforms are generated by both post-transcriptional or post-translational events. In fact, the number of CD44 variants or isoforms putatively generated by alternative pre-mRNA splicing alone has been estimated to be approximately 10^3 (41). However, most terminally differentiated non-transformed cells express only a few isoforms, with increased numbers of variants expressed by activated, cycling or generative cells (42). Interestingly, the majority of transformed and neoplastic cells have also been demonstrated to express an increased number of heterogeneous CD44 isoforms (43, 44).

1.2 Ligand binding specificity of CD44

The protein sequence similarity between the common extracellular domain of CD44 and cartilage link protein initially suggested that CD44 may interact with components of the extracellular matrix (27, 45). Several lines of evidence have emerged demonstrating that CD44 functions as a cell surface receptor for the extracellular matrix component, hyaluronan. These include binding of cells transfected with plasmid DNA or retroviral vectors containing cDNAs encoding CD44 isoforms to hyaluronan coated plastic (46) or hyaluronan-conjugated fluorescein isothiocyanate (HA-FITC; 47), binding of CD44-Ig fusion proteins to hyaluronan, or the binding of a chimeric molecule composed of the CD44 extracellular domain fused to an alkaline phosphatase indicator to hyaluronan in an ELISA-type assay (48).

Hyaluronan is a negatively charged large molecular weight extracellular matrix polysaccharide composed of repeating units of N-acetyl-glucosamine and D-glucuronic acid, linked together by a β1-4 glycosidic bond to form a large unbranched polymer with an average M_r of several million (49). Hydrogen bonding along the backbone of the molecule holds it in a stiffened helical conformation, forming a random coil in solution and sequestering 1000 fold more water than polymer to form a highly hydrated sphere (50). Hyaluronan is found widely distributed throughout the extracellular matrix between cells and in the bone marrow, and is particularly rich in the lung, the humour of the eye, synovial fluids

and umbilical cord. The biosynthesis of hyaluronan occurs from the plasma membranes of fibroblasts (51) by the enzyme hyaluronate synthase, with catabolism taking place within the Kupffer and endothelial cells of the liver (52) and mediated in part by CD44 binding and turnover (53).

Initially, various CD44 isoforms were reported to bind to hyaluronan with different affinities (45), thus implying that the function of the unique inserted region was to modulate CD44 binding. However, this simplistic picture of CD44 function may not necessarily be correct. CD44 is now known to exist in 3 states with respect to hyaluronan binding: inactive, inducible and constitutively active (54-56). Each of these states is strictly regulated by mechanisms which are not yet completely described. Nevertheless, intensive work by various investigators has revealed several possible modes of regulation. Using different techniques, Lesley *et al.* (57) and Katoh *et al.* (58) have presented convincing evidence that glycosylation of CD44 isoforms negatively regulated hyaluronan binding. A further recent report by Takahashi *et al.* (59) indicated that modification of CD44H by some GAGs such as keratan sulphate also reduced its ability to bind hyaluronan. Others (48) have demonstrated that CD44 variants previously thought not to interact with hyaluronan, did in fact bind depending upon the cell line expressing the CD44 molecule. Furthermore, He *et al.* (60) have described 4 novel murine CD44 isoforms all of which bound to hyaluronan. These observations suggest that the ability of CD44 isoforms to bind to hyaluronan is actually more cell specific than isoform specific. Furthermore, other cell specific modifications of CD44, such as interactions with the cytoskeleton through ankyrin (30, 31) or actin/ERM (33, 61), and phosphorylation of cytoplasmic domain serine residues (29) have also been shown to affect hyaluronan binding capacity. Taken together, these data argue that the mere inclusion of additional amino acid residues within the unique region does not of itself regulate hyaluronan binding, but rather that the glycosylation or GAG addition to the overall CD44

extracellular domain in the context of other cytoplasmic activation events performs a regulatory role (62).

In addition to hyaluronan binding, the presence of GAG substitution on variant exon encoded motifs may also confer additional ligand binding specificity upon CD44 isoforms. For example, the ability to bind to distinct ligands has been noted for v3 (exon 8) containing isoforms substituted with heparan sulphate (63). These isoforms have been shown to adhere to basic-fibroblast growth factor (b-FGF) and heparin binding-epidermal growth factor (62). Several cytokines such as macrophage inflammatory protein-1β (MIP-1b 64) and osteopontin (Eta-1 65), and the heparin binding domain of the extracellular matrix component fibronectin have also been proven to bind to CD44, possibly through this domain (66).

In keeping with its ability to bind extracellular matrix molecules, CD44 has also been shown to adhere to chondroitin-4-sulphate (CS) attached to protein cores. This interaction occurs through a protein domain close to but separate from its hyaluronan binding domain (67, 68), (Droll *et al.*, manuscript in preparation). Serglycin (69, 70), and the Major Histocompatibility Complex class II invariant chain (71) are two recently reported CD44 ligands, both of which are modified by CS and may be interacting with CD44 through this domain.

Interestingly, Droll *et al.* (72) have recently reported that a CD44 variant containing additional exons v8, v9 and v10 (CD44R1) is also modified by CS on the portion encoded by exon v10. The presence of this modification was found to induce the homotypic aggregation of cells expressing CD44R1, or heterotypic aggregation with either CD44R2 (or CD44v10; containing additional exon v10) or CD44H expressing cell lines. Aggregation was not observed between cells expressing CD44H or control vector only. Thus certain CD44 variants can act as ligands for themselves or other isoforms. The specific interaction of this

CS modified form of CD44 with types I and IV collagen has also been demonstrated to be necessary for invasion of basement membranes by melanoma cells (73, 74). Taken together, these observations suggest that the expression and modulation of different CD44 isoforms by the cell may serve not only to influence binding to hyaluronan and other extracellular matrix components, or to sequester and present certain cytokines, but also to provide alternative and perhaps activating self-ligands. This unique ability of CD44 may be the selective pressure which induces many transformed cells to express multiple CD44 isoforms.

2. Involvement Of CD44 With Metastatic Disease

Since the initial finding that a CD44 variant was causally linked to metastatic potential (4), multiple reports of CD44 variants coincidentally linked to primary or metastatic lesions have appeared. The expression of CD44 variants has provided a useful prognostic indicator. Specifically, the presence of v6 containing isoforms have been linked to the metastatic propensity of pancreatic adenocarcinoma (75, 76), papillary thyroid carcinoma (77), breast carcinoma (78) and colorectal carcinoma (79, 80). In many cases the expression of v10 containing variants has also been observed in the same carcinomas (43, 75, 81), as well as others such as uterine cervical carcinoma (82) and renal carcinoma (83, 84). Other variant exon containing isoforms have also been correlated to certain neoplasms (82, 85, 86). In addition to solid tumours, CD44 isoform expression is also upregulated on various types of leukemia (87) and lymphoma (88) relative to normal controls. Furthermore, CD44H expression itself has been shown to be involved in the development of metastatic capacity. For example, an anti-sense expression system has been utilized to demonstrate a link between motility of CD44H positive glioma cells on hyaluronan coated plastic and *in vivo* metastatic behavior in murine brain (89).

The exact mechanisms used by CD44 isoforms to potentiate tumour metastasis have yet to be characterized. However, CD44 expression may have this effect through several possible mechanisms and at various stages of metastasis. For example, the *in vivo* treatment of metastatic tumour cell lines with either anti-CD44 variant antibody (90) or with CD44-Ig fusion proteins (91) has been demonstrated to block their ability to metastasize. This finding suggests that the adhesive functions of CD44 play a role in the binding of the metastasizing cell to the vascular endothelium. Inclusion of variant exons such as v3 and v10 may lead to the addition of GAGs to a portion of the CD44 core protein expressed by some cells and result in alternate binding specificities (72, 74, 92), or enhancement of motility through the extracellular matrix (93). Alternatively, crosslinking of CD44 proteins by mAbs or ligands has been shown to induce 'activation signals' such as those leading to IL-1 and tumour necrosis factor-α (TNF-α) production in monocytes (94), or co-stimulation of CD2 and CD3 mediated T cell activation (95-97). Similar signals may be transduced by the ligation of CD44 isoforms expressed by metastatic tumour cells. In contrast, Ayroldi *et al.* have recently reported that expression of CD44 on thymocytes provided a protective effect against CD3 or dexamethasone induced apoptosis (98). Thus CD44 may also modulate metastatic propensity by increasing the longevity of the cell.

Interestingly, ligation of CD44 has previously been demonstrated to indirectly down regulate the signaling effects of other adhesion proteins such as LFA-1/ICAM-1 (19), or CD3 (99). In this case, crosslinking of CD44 appears to function as a negative regulator of other stimulatory molecules. Since immunopreciptation studies have shown that the cytoplasmic domain of CD44 interacts in some fashion with intracellular proteins involved in signaling cascades, such as PKC (34) and $p56^{lck}$ (35), CD44 may conceivably perform these diverse regulatory roles by binding to and sequestering various proteins involved in signaling cascades. The presence of different unique regions may modulate the cell surface

interactions of CD44 isoforms with other proteins, and thus regulate the binding and release of the sequestered proteins or alternatively regulate cross-phosphorylation events.

Since the inclusion of variant exons is performed by the process of alternative pre-mRNA splicing, investigation of the molecular mechanisms which control the splicing of CD44 exons is relevant both to further the current understanding of metastatic disease, and to develop future therapies.

3. SPLICING OF PRE-MRNA

Shortly after the physical structure of genes in bacteria were established and characterized, researchers realized that the genes of higher eukaryotes were different. Eukaryotic genes contained numerous long regions of non-coding sequences called introns, which interrupted the coding sequence or exons (100). After initial transcription of the DNA into RNA, these intronic sequences were removed to form the continuous coding of the mRNA which is then translated into protein. The cellular process which removed introns was termed RNA splicing (reviewed in 101).

Perhaps even more surprising than the discovery of 'split genes', was the finding that the pre-spliced RNA transcript may encode more than one protein through its ability to be alternatively spliced. That is, certain exons present within the pre-spliced transcript may or may not be included in the final mRNA, thus resulting in proteins with various isoforms such as CD44 (102), or fibronectin (103), or different proteins altogether such as calcitonin and calcitonin gene related peptide (104) . Research into the control mechanisms which regulate alternative splicing of exons has revealed a complex cellular machinery composed of catalytic RNA molecules bound by numerous proteins, and accompanied by other associated proteins. This complex has been termed the spliceosome. Although many of the

mechanisms responsible for constitutive or non-alternative splicing are now understood, those mechanisms which regulate alternative splicing have remained unclear.

3.1 Constitutive pre-mRNA splicing

The current understanding of the basic splicing reactions which link exons together has been made possible by the development of the soluble *in vitro* splice assay by Padgett (105). This technique consists of a radiolabelled pre-mRNA incubated with a nuclear extract from HeLa cells, followed by separation of reaction components by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography. Used in conjunction with powerful yeast mutational screens as well as experiments involving the depletion and re-addition of various factors, this technique has allowed the characterization of the mechanisms of pre-mRNA splicing.

3.1.1 The splice reactions

The basic splice reaction between two exons (exons A and B in Figure 3) and the intervening intron takes place in two steps. In the first step the upstream exon is cleaved from the intron at the 5' splice site by the nucleophilic attack of a residue downstream within the intron. This residue, usually an adenine, forms a 2'-5' phosphodiester bond with the 5' end of the intron resulting in a free upstream exon (exon A) with a 3'-OH terminus and a tri-bonded lariat RNA intermediate (106, 107). In the second, kinetically distinct step, the free 3'-OH of exon A attacks the downstream exon (exon B) at the 3' splice site, resulting in cleavage of the lariat intron from exon B, and joining of exon A to B.

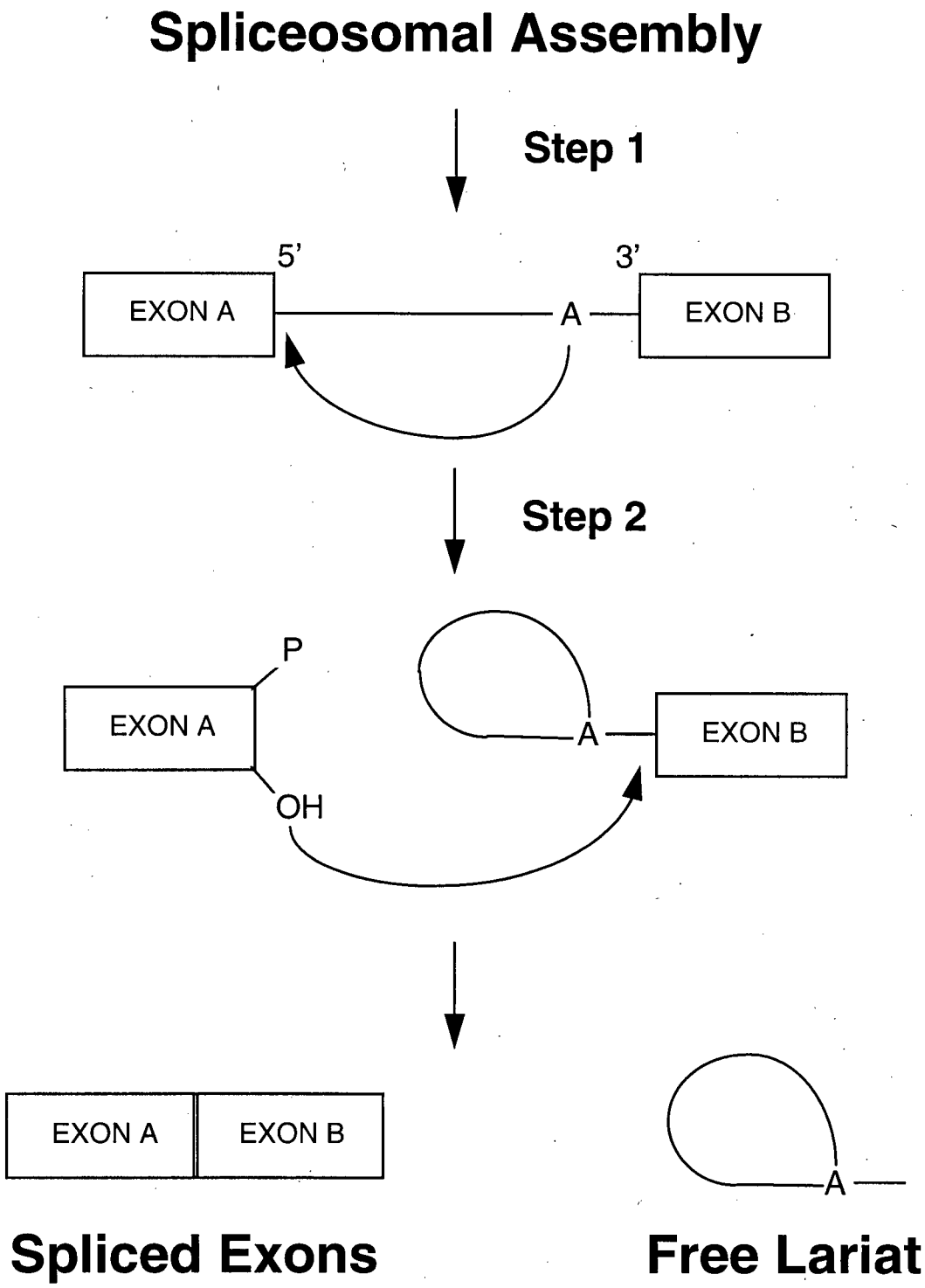


Figure 3: The basic chemistry of the splice reactions.

3.1.2 Sequence elements at the 5' and 3' splice sites

Early comparisons of the genomic regions which encode the genes of various species demonstrated that certain sequence motifs were highly conserved (101). These motifs were found at the positions which correspond to exon and intron boundaries in the pre-mRNA (108), and define what is now known as the 5' (splice donor) and 3' (splice acceptor) splice sites (Figure 4). These splice sites are highly conserved in yeast, but more loosely conserved in mammals. In general, when a mammalian splice site is mutated to mimic the yeast consensus, an overall increase in splice site usage or final mRNA product is observed (109, 110, 111, 112), thus, those sites which more closely match the yeast consensus are said to be stronger splice sites.

	MAMMAL	**YEAST**
5' Splice Site	AG|GURAGU	AG|GUAUAU
3' Splice Site	YAG|G	CAG|G
Branch Point	YNYURAY	UACUAAC

Figure 4: Splice site consensus sequences. Vertical lines indicate the site of cleavage, with invariant dinucleotides underlined and branch point adenine shown in bold. A comparison to the yeast consensus is shown for reference.

Two other sequence motifs, normally located within 50 nt of the 3' splice site, are also necessary for constitutive pre-mRNA splicing. These are the branch point region which contains the attacking nucleophilic adenine (109, 113) and a pyrimidine rich region which extends from the branch point to the 3' splice site, called the polypyrimidine tract (107, 114, 115).

3.1.3 Pre-mRNA splicing is catalyzed by the spliceosome

The splicing of pre-mRNA takes place within the nucleus in regions called nuclear speckles (101), and is performed by a set of 5 small nuclear RNAs (snRNAs) and >100 associated proteins. Together the snRNAs U1, U2, U4, U5, and U6 with their proteins are referred to as small nuclear ribonucleoproteins (114). These snRNPs have been demonstrated to interact with the splicing pre-mRNA in a cyclical direct and sequential manner by RNA:RNA interactions (114, 116, 117).

Initial recognition and binding of the pre-mRNA is performed by U1 snRNP, which interacts jointly with the 5' and 3' splice sites to form the commitment complex or complex E (CC/E; Figure 5; 111, 114, 116). Next, U2 binds to the branch point region in an ATP requiring reaction to form complex A (118). This interaction is critically dependent upon the stabilizing influence U1 as well as an associated protein factor U2AF (119). U2AF is composed of 2 subunits, $U2AF^{35}$ and $U2AF^{65}$ (120, 121). The $U2AF^{65}$ subunit binds to the polypyrimidine tract adjacent to the branch point region (120, 122), and is necessary for U2 interactions with the branch point region (119, 123, 124). $U2AF^{65}$ also interacts with U1 through the U1 snRNP associated protein U1-70K (125). Thus the binding of U1 and U2 snRNPs to the pre-mRNA is an interactive, rather than independent process (123). In addition to U2AF, several other proteins recently described in yeast have also been shown to interact with the pre-mRNA in a sequence independent manner and be necessary for U2 snRNP/branch point region binding (124, 126); mammalian homologues have yet to be described.

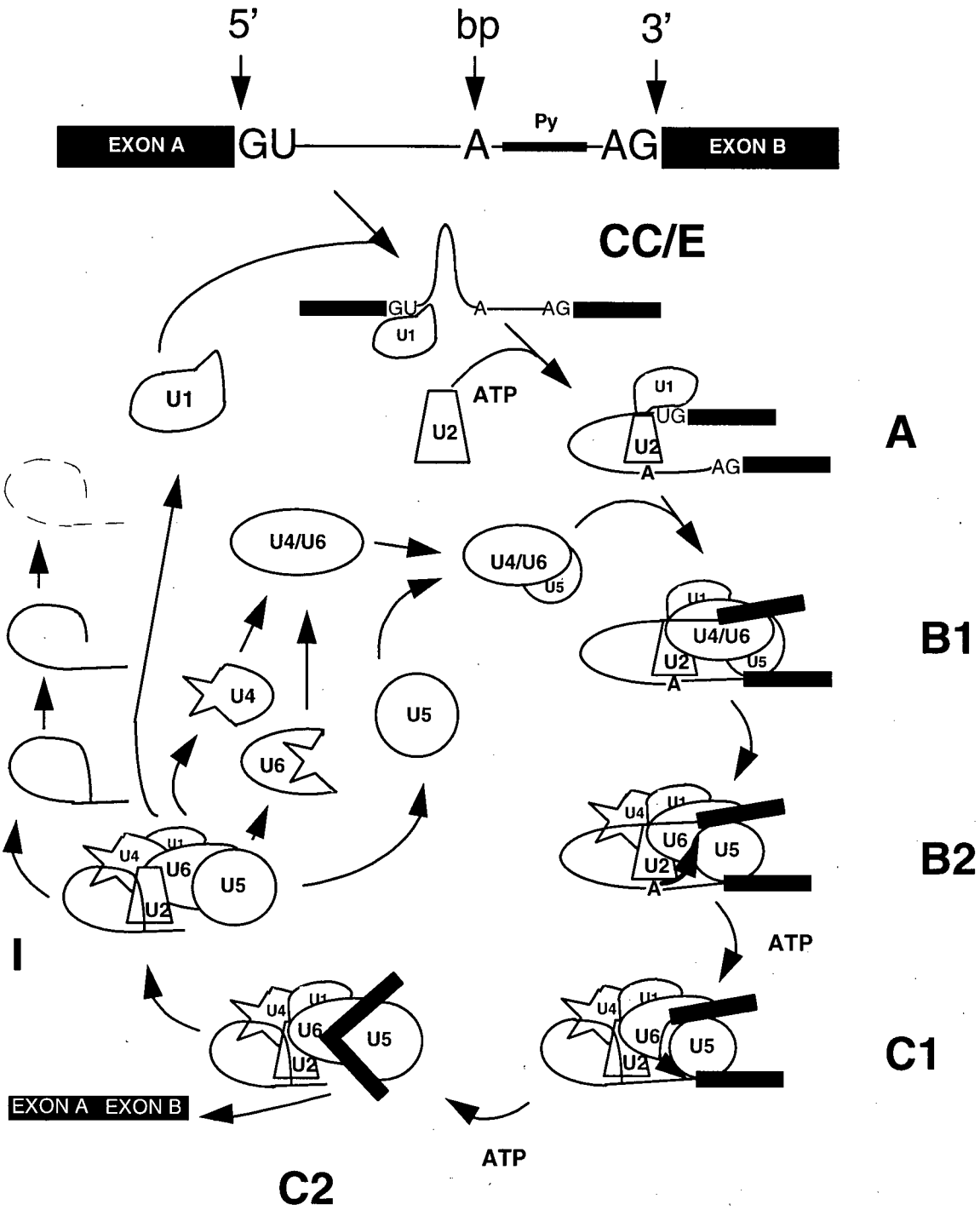
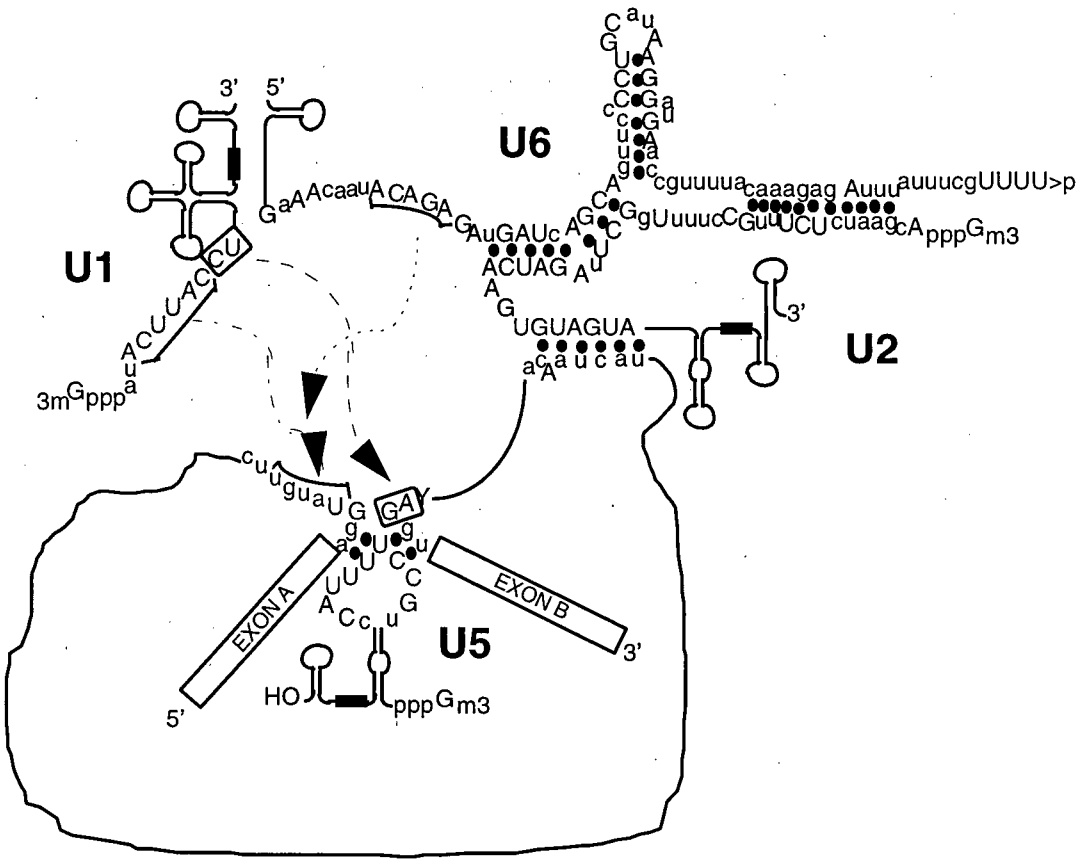


Figure 5: Cycling of the snRNPs during the splicing reaction. The snRNPs (U1, U2, U4/U6 and U5) interact with the pre-mRNA at the 5' splice site (5'), 3' splice site (3'), branch point sequence (bp) and polypyrimidine tract (Py). The various stages of the cycle (CC/E, A, B1, B2, C1, C2 and I) are defined by the binding of the snRNPs and the splicing reaction as described in the text. Adapted from (101).

Figure 6: Interactions of the various snRNPs during the splicing reaction. (A) Consensus sequences within the snRNAs interact with conserved nucleotides of the pre-mRNA. The pre-mRNA and snRNA sequences shown are those of *S. cerevisiae* with highly conserved nucleotides in uppercase. Internal snRNA sequences which do not interact with other snRNA or with the pre-mRNA are depicted as stylized stems and loops. Rearrangements which occur between U1, U6, and the pre-mRNA during the splice reaction are shown with arrows. (B) Extended mammalian consensus sequences of the 5' splice site, U1, and U6 consensus shown with nucleotide numbers indicated above. Adapted from (101).

A



B

5' SPLICE SITE

1 2 3 4 5 6

C/A AG | GU A/G AGU

U1 CONSENSUS

6 5 4 3 2 1

3' - GUC | CAUUCAUA - 5'

U6 CONSENSUS

6 5 4 3 2 1

3' - GA | GAGACAUAGCA - 5'

The B1 spliceosomal complex forms when a tri-snRNP particle composed of U4, U6 and U5 (U4.U6-U5) binds to the A complex. (127-129). U4 has been shown to bind and sequester the proposed catalytically active U6 snRNP (128), while a conserved loop of U5 coordinately recognizes the 5' and 3' splice sites by non-canonical base pairing interactions (127, 130), (Figure 6). Intramolecular rearrangements then occur to form the B2 complex. These include the dissociation of U4 from U6 (106, 131, 132) and the subsequent annealing of U6 with U2 and the 5' splice site (133, 134). Association of U2 and U6 snRNPs has been postulated to form the catalytic core of the spliceosome (135, 136). Interestingly, U6 snRNP also displays a secondary proofreading function for the site of cleavage at the 5' splice site in conjunction with that performed by U1 (117, 137, 138).

The first step of splicing (section 3.1.1) occurs in the B2 complex with the addition of ATP to form the C1 complex which now contains the lariat intron structure as well as the snRNPs. Further ATP addition accompanies the second step of splicing to form the C2 complex. This complex quickly dissociates into the joined exons, and lariat intron-snRNPs (complex I) followed by lariat intron release and degradation, and release and recycling of the snRNP components (108).

3.1.4 Factors involved in defining the exon

There is a conceptual problem which arises when considering the mechanisms of splice site recognition and pre-mRNA splicing. Namely, introns can be very long (average length of 1127 nt) with many that are known to be greater than 3 kb. Conversely internal exons are normally quite small (average length of 137 nt) and appear to be limited to a length of less than 300 nt (139, 140). How then are a series of very short exons recognized within a very long stretch of introns? Furthermore, as previously stated, exon boundaries are defined by

loosely conserved 5' and 3' splice sites within mammals. Sequences which closely resemble these sites will often occur within the long intronic region. How then are the 'true' exon splice sites selected from 'false' sites?

Several models have been proposed to answer these questions. The best model to fit with known experimental data is the exon definition model of Berget *et al.* (139). This model proposes that the factors involved in spliceosomal assembly 'search out' a pair of closely spaced splice sites in an exonic polarity (in other words 3' splice site upstream and 5' splice site downstream - see Figure 7), which is then defined by the binding of U1 and U2 snRNPs and their associated factors. Following recognition, exons are juxtaposed by interactions between their bound snRNPs and proteins, allowing splicing to occur. According to this model, the length of an internal exon is limited to 300 nt by the need for exon-spanning interactions of the snRNPs and proteins which define the exon. Single cryptic splice sites within a large intron are often ignored by the splicing machinery because they do not have an opposite polarity site in an appropriate location.

Note that terminal 5' and 3' exons are exempt from these known size limitations. Instead these exons are recognized by interactions between factors bound to the exon splice site and those at the m7G(5')ppp(5')N cap (141, 142) or the polyadenylation site (143-146).

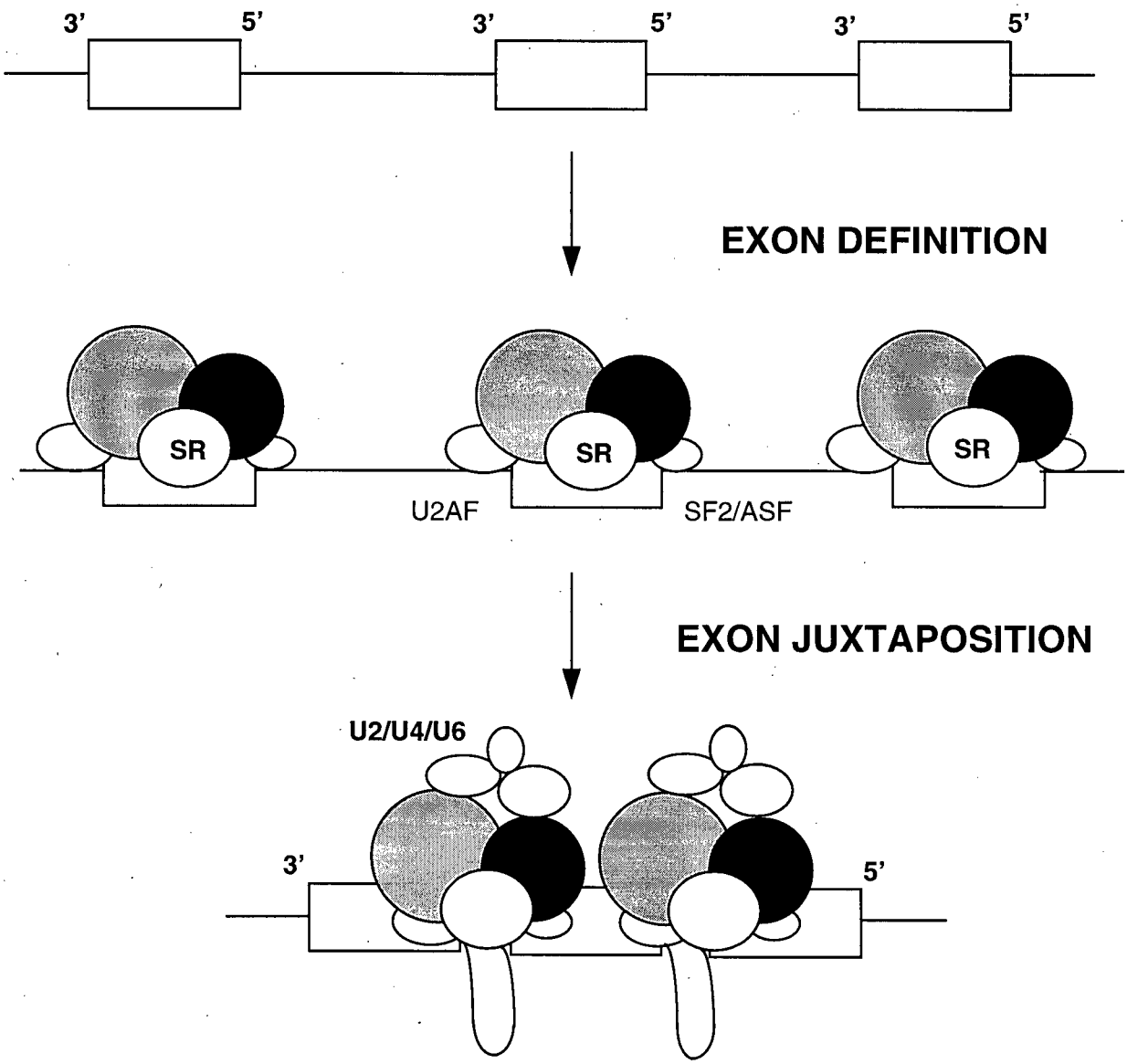


Figure 7: The exon definition model of Berget *et al.* Each exon (open boxes) is defined at its 3' and 5' splice sites by the binding of U2AF and SF2/ASF in concert with other SR proteins and the U1 and U2 snRNPs (open and shaded circles) as described in the text. After the exon is defined, opposing exons are juxtaposed by interactions between proteins and snRNPs bound to the 5' and 3' splice sites of the upstream and downstream exon.

3.2 Alternative splicing of pre-mRNA

Through the use of the in vitro splice assay system, the mechanisms of constitutive exon splicing have been characterized. However, the factors which control alternative splicing are less well-defined.

Mammalian splice sites often diverge from the known consensus sequences defining the boundaries of constitutively spliced exons (147). Thus, the mere sequences present at the 5' and 3' splice sites are not sufficient to define those exons that are alternatively used from those that are constitutively used. In fact, many exons which undergo alternative splicing are indistinguishable from constitutively spliced exons based on a match to consensus sequence alone (108, 147, 148). While no single 'alternative exon splice indicator' has been found, recent experimental evidence has accumulated which suggests that the alternative splicing of exons is determined by a balance of the strength of the 5' and 3' splice sites in concert with local sequence context, competing sites, and expression of non splice-essential trans-acting factors (148, 149).

3.2.1 Cis-acting sequences

While divergence from consensus may not be the defining hallmark of alternatively spliced exons, a general trend of suboptimal splicing signals often combined with unusually placed or competing splice sites flanking these exons has been observed (109, 110, 149, 150). This trend is particularly noted in alternative 5' splice site selection. Eperon *et. al.* have demonstrated that when two closely spaced (within 50 nt) 5' splice sites are in competition with each other, the downstream site is preferentially utilized unless the upstream site has a much greater affinity for U1 snRNP (151). In addition to U1, U6 has also been demonstrated

to interact with part of the 5' splice site and the following downstream nucleotides (117, 137), influencing 5' splice site choice by extending the recognized sequence context.

Similar to 5' splice site choice, 3' splice site choice also appears to be controlled by both sequence of the 3' splice site signals and their position or context. Since the consensus which defines the 3' cleavage site is very short (YAG|G, see Figure 4), control of 3' splice site usage is performed primarily by the branch point region and/or polypyrimidine tract. The site of exon cleavage appears to be determined by a scanning mechanism which starts at the branch point region and searches downstream for the first AG dinucleotide. Closely spaced AG dinucleotides at the 3' splice site are competitive, with a hierarchy of cleavage site preference being CAG≡UAG>AAG>GAG (112).

Observed steric constraints on the assembly of spliceosomal components at the 3' splice site initially led to the conclusion that the branch point region must be located between 18 and 40 nt of the 3' cleavage site (107, 109, 152). This spatial limitation appears to be true for constitutive but not alternatively spliced exons. Several examples of abnormally distant (>100 nt) 'long range' branch points have been described as a major controlling mechanism in the alternative splicing of rat and chicken tropomyosin I (153, 154), Adenovirus E1a (155) and rat fibronectin (156) pre-mRNAs. This type of branch point is normally accompanied by a compensatory polypyrimidine tract with an unusually high pyrimidine content (154).

Mutation of the branch point region has also been noted to modulate alternative 3' splice site usage by influencing the crucial RNA:RNA interaction of U2 snRNP with this sequence (152). In some cases, a branch point C or U residue may be used as a 'low affinity' branch point nucleophile to regulate 3' splice site usage (157) as demonstrated for human growth hormone receptor splicing (158). Lastly, alternative exon splicing may also be regulated by

the presence of multiple overlapping branch point regions which compete for U2 snRNP binding (154, 155).

The pyrimidine content of the polypyrimidine tract has a major effect upon 3' splice site usage (114). Using a mutational approach to examine 3' splice site efficiency, even the loss of a single uridine nucleotide was demonstrated to drastically alter 3' splice site choice (115, 159). When branch points are abnormally far from the 3' cleavage site, or have a low affinity to U2 snRNP, a strong tract with many consecutive pyrimidine residues can 'rescue' a 3' splice site (160) or exert a strong competitive effect as observed for rat α–tropomyosin (161, 162).

A final type of cis-acting sequence motif has recently been described in both positive and negative control of alternative splicing (147). These motifs are found outside of the described splice sites and are called splicing enhancers. Most commonly the enhancer has been found within the exon itself (exon splice enhancer or ESE) as described in CT/CGRP splicing (163, 164), rat β-tropomyosin (165), bovine growth hormone (166, 167), cardiac troponin T (168, 169), and CD45 (170) pre-mRNA splicing. These ESE motifs are generally composed of multiple copies of a 6-13 nt repeat of GAR where R=G or A (171-175). Other enhancers have also been located within introns (175, 176) as have negative repressors or 'poison sequences' (175, 177, 178). Interestingly these downstream ISEs have been often found to resemble 5' splice sites (164) and interact with U1 (123, 139, 179, 180). The negative repressor motif has been less well defined than that of the ESE, but appears to contain repeated motifs of CUCUCU (178).

3.2.2 Trans-acting factors

The isolation of trans-acting factors that influence alternative splice site choice has provided a major leap forward in the understanding of both alternative and constitutive splicing. These factors can be classified into 2 families: heterogeneous nuclear ribonucleoprotiens (hnRNPs) and SR proteins. Both of these families are characterized by a common RNA recognition motif (RRM) or RNA binding domain (RBD) of approximately 80 aa. The RRM has been characterized in a number of RNA binding proteins and is not particular to splice factors (181). The SR family of splice factors is further differentiated by the presence of a glycine-rich hinge region and C-terminal serine-arginine rich domain (182).

Heteronuclear RNPs (hnRNPs) are one of the major protein components of the nucleus. At least 20 hnRNPs have been isolated based on apparent molecular weight and isoelectric charge and are named hnRNP A through U (183). Although many have been demonstrated simply to coat RNA through non-specific binding, several have been isolated which preferentially interact with pre-mRNA (184, 185) including hnRNP A1 and its alternatively spliced transcript hnRNP A1B (186), hnRNPs A2, B, C1, C2, D, and I (187). Interestingly, all of these hnRNPs have been shown to interact preferentially with the polypyrimidine tract (184, 187). For example, using *in vitro* antibody inhibition experiments, the hnRNP C proteins have been proven to be required for splicing (183) through tight binding to the polypyrimidine tract (159). In addition, hnRNPs I and most notably A1 have been shown to also interact with pre-mRNA 5' splice sites, and to affect 5' splice site choice (188). Recently a high affinity hnRNP A1 binding site was identified which resembles both 5' and 3' splice sites (189).

There are currently eight characterized human SR proteins: p75, p55, p40, p35, p30a (SF2/ASF), p30b (SC35), p30c, and 9G8 (190) with homologues described in *Drosophila* as

well as other species (182, 191-193). These SR proteins were initially identified by an antibody (mAb 104) which recognized a common phosphoepitope (182), and have been demonstrated to be required for the commitment of different pre-mRNAs to splicing (194) by promoting the first recognition of pre-mRNAs together with U1 snRNP (195) in the early commitment complex. It is at this early stage of spliceosome formation that SR proteins are thought to play a significant role in determination of exon inclusion/exclusion or alternative splicing.

Two SR proteins, SC35 and SF2/ASF have been shown to be particularly involved in splice site selection, and thus in alternative splice regulation. SF2/ASF is known to interact specifically at the 5' splice site with U1 snRNP (116, 196-198). This interaction is dependent of the presence of its RS domain but not its RRM (198) and probably occurs by binding to the U1-70K component of U1 snRNP (125, 199, 200). Early research into SF2/ASF function by Ge *et al.* found that high levels of expression resulted in alternative splicing of SV40 early pre-mRNA *in vitro* (201). More accurately, SF2/ASF has been demonstrated to influence 5' splice site choice when highly expressed by the activation of the site more proximal to the 3' splice site (101). This function of SF2/ASF is antagonized by the expression of hnRNP A1 which directs selection of distal 5' splice sites when over expressed (202). Taken together, these data suggest that the regulation of 5' splice site choice depends upon the ratio of SF2/ASF to hnRNP A1.

In contrast, SC 35 has been demonstrated to interact with both U1 and U2 snRNPs at the 3' splice site (203) during the formation of spliceosomal complex A. Although overexpression of SC35 has been shown to replace some of the functions of U1 in 5' splice site determination (204, 205), the purified factor does not bind directly to U1 snRNP or the 5' splice site (206). SC35 binds to the U1-70K component through which it may mediate these effects and also binds to the $U2AF^{35}$ subunit of U2AF (199). Thus, SC35 has been

suggested to bridge the exon by interactions at both the branch point region and the 5' splice site.

3.3 The regulation of alternative splicing

As mentioned above, 5' splice site selection has been demonstrated to be influenced by the cellular ratio of SF2/ASF to hnRNP A1, while 3' splice site selection is affected by the binding of SC35 to U2AF^{35} and U1-70K. Interestingly SF2/ASF has also been shown to interact with U2AF^{35} in a bridging function (199) which may also be a regulatory factor. However, recent experiments with ESEs have demonstrated a further role for SR proteins in the control of alternative splicing.

3.3.1 Interactions between SR proteins and ESEs

Early clues that SR proteins may interact with sites distinct from the 5' and 3' splice site regions came from studies of sex determination in *Drosophila*, which is controlled by the alternative sex-specific splicing of 3 genes. An autoregulated female-specific splice in Sex-lethal (*SXL*) pre-mRNA produces a protein isoform which controls a female specific splice in the pre-mRNA of an SR protein, Transformer (*TRA*). Together with a non-sex specific splice factor Transformer-2 (TRA-2), the female isoform of TRA binds to exon enhancers in Double-sex (*DSX*) pre-mRNA, leading to female specific splicing of *DSX* and somatic sex-specific development (108, 161, 179, 207-209). The RS domain of TRA-2 has been shown to interact with itself, TRA, and the *Drosophila* homologue of SF2 together (210) on a 13 nt repeat motif found within the *DSX* exon 4 (207, 208). Since TRA-2 is a SR protein (193) it is conceivable that other human SR proteins may likewise regulate alternative splicing by binding to the newly discovered ESE (199).

Evidence for the sequence specific interaction of various mammalian SR proteins with ESEs has since accumulated. Each SR protein appears to bind to a distinct ESE (211, 212), and several examples of alternative splice site switching by addition of SR proteins have been documented. Based on this evidence, Staknis and Reed (195) have proposed a model for SR protein-mediated enhancement of splicing (Figure 8). They propose that splicing enhancers function by promoting the assembly of the spliceosomal complex on pre-mRNAs which have weak 5' or 3' splice sites.

Pre-mRNAs which have strong sites and are constitutively spliced do not require enhancers (Figure 8a). For pre-mRNAs with a weak 3' splice site, an enhancer complex consisting of SR proteins and U1 interacts through U1-70K (125, 173) at the ESE to promote the interaction of U2AF with the branch point sequence (Figure 8b). Downstream 5' splice site-like enhancer elements function in a similar fashion (Figure 8c). When 5' splice sites are weak, ESEs stabilize U2AF binding to the branch point region, a function normally performed by U1 at the 5' splice site (213). Thus various SR proteins can regulate ESEs by differential recognition or competition (171). Interestingly, the distance between the ESE and the 3' splice site also regulates alternative splicing (207, 214), indicating a steric component to the enhancer complex. Thus downstream splicing events which can relocate an ESE into the range of an upstream 3' splice site, may act as control points for upstream splicing events. This sort of alternative splicing control has been demonstrated in the splicing of rat β-Tropomyosin (154).

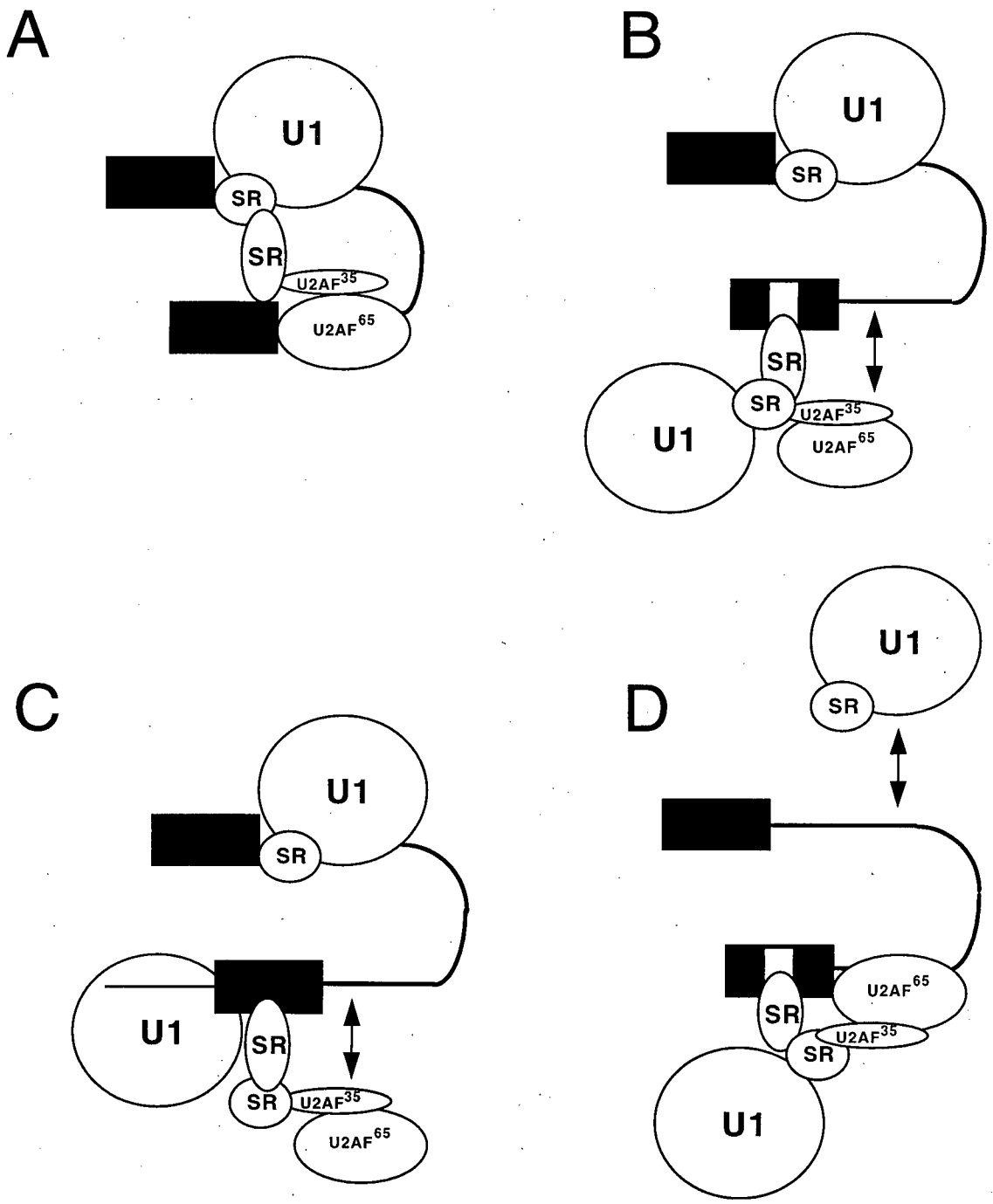


Figure 8: Model of ESE enhancement as proposed by Staknis and Reed. *Panel A*: strong 5' and 3' splice sites do not require any enhancement. *Panel B*: Weak 3' splice sites (diamond) are enhanced by SR proteins which bind to the ESE (open box within exon) and stabilize U2AF interactions with the branch point regions. *Panel C*: Downstream ISEs also stabilize U2AF binding to the branch point regions through interaction of U1 with a 5' splice site like motif. *Panel D*: ESEs bound with SR proteins stabilize U2AF/branch point region interactions when the 5' splice site is weak and cannot perform this function.

3.3.2 Regulation by other proteins

Proteins other than the SR proteins described above have been described which also induce splice site switching and thus control alternative splicing. These include hnRNP F in neural specific splicing (215) and Mer1 protein regulation of *MER2* splicing in *Saccharomyces. cerevisiae* (216). As previously mentioned, the sequence of polypyrimidine tracts has a potent affect on splicing, presumably mediated by the binding of the various hnRNP proteins (section 3.2.1). Furthermore $U2AF^{65}$ has been demonstrated to bind to polypyrimidine tracts close to the branch point region (122) during early spliceosome assembly and is necessary for U2 binding to this site (119, 123, 124). Thus competition for binding at the polypyrimidine tract also regulates alternative splicing through alternative branch point selection (160). Several proteins have been demonstrated to bind to distinct motifs within the polypyrimidine tract and perform this sort of regulation. These include hnRNP C (157, 217), PTB or hnRNP I and SXL in *Drosophila* (122).

3.3.3 Regulation by phosphorylation

Further fine-tuning of alternative splicing regulation may be accomplished by phosphorylation. Several steps during spliceosome assembly require ATP (218). Furthermore, serine/threonine phosphatases have been demonstrated to be necessary for splicing catalysis (219). Thus, it is likely that the phosphorylation/de-phosphorylation of regulatory proteins is a control mechanism for splicing regulation.

All SR proteins isolated thus far are phosphoproteins (182, 220) with serine residues of the C-terminal RS domain being the likely target (221). Although the exact role of the phosphorylation of the RS domain in the regulation of alternative splicing is not clear, this

domain has been demonstrated to interact with other proteins, particularly the U1-70K protein (199, 222, 223). Two kinases, the serine/threonine kinase SRPK1 (224, 225) and the serine/threonine/tyrosine kinase Clk/Sty (221) have been shown to phosphorylate SR proteins. Overexpression of these kinases caused a relocation of SR proteins in the nucleus (224). Interestingly both the SR protein SF2/ASF and the $U2AF^{65}$ subunit were phosphorylated in vitro by SRPK1 (225), suggesting that phosphorylation may alter interactions between these two proteins, or other proteins in vivo.

Phosphorylation of hnRNPs has been demonstrated to regulate their interactions with pre-mRNA. In particular, the phosphorylation of hnRNP A1 on a C-terminal serine residue negatively regulated its *in vitro* RNA strand annealing activity (226), a function thought to be necessary for the binding of U2 to the branch point region (227, 228). This phosphorylation could be performed *in vitro* by protein kinase A, and casein kinase II (227, 228) and interestingly also by protein kinase C ζ *in vivo* (229).

hnRNP C binding to the polypyrimidine tract is also negatively regulated by phosphorylation on serine and threonine residues (230) in what appears to be a cell cycle specific manner (217). Phosphorylation of this hnRNP occurs *in vitro* by casein kinase II (231, 232) and is modulated by nucleotides 78 to 95 of U6 (233) perhaps indicating a further catalytic function for U6, as well as a splice regulatory function.

hnRNP A2 is also phosphorylated *in vitro* by casein kinase II (231) although the functional consequences of this modification is not known.

Lastly, U1-70K is also phosphorylated on serine residues at its C-terminal by an snRNP kinase activity (234). Using a thio-phosphorylated U1-70K complementation assay, Tazi *et al.* have shown that de-phosphorylation of this protein is necessary for the catalysis of

splicing but not spliceosomal assembly (235). Phosphorylation of U1-70K may be accomplished by the serine/threonine phosphatases already described as necessary for activity of the catalytic core (219).

3.3.4 RNA secondary structure

The role of RNA secondary structure as a control mechanism for alternative splicing has been hotly debated among researchers. Using exon/intron model systems, some have demonstrated that a stem-loop structure can sequester a 5' splice site and cause alternative splicing (101) , (236), while others have shown that the region of complementarity necessary to form a stable splice-regulating secondary structure would be very large, and therefore these structures are unlikely to occur (224). Nevertheless, at least two examples of alternative splicing regulated by the presence of an RNA secondary structure have been demonstrated in natural systems. A stem-loop structure isolating exon 6B within the loop regulated a muscle-specific splicing event in chicken β-Tropomyosin pre-mRNA (237, 238). The opposite effect was found in splicing of the actin (ACT) pre-mRNA of *Kluyveromyces lactis*, where a mutation in an intron disturbed a natural regulatory secondary structure. In this case, a stem-loop structure was necessary to bring a 3' splice site closer to a long range branch point region (239), the loss of the stem-loop through mutation led to cryptic splice site activation. Thus, although model systems may or may not implicate secondary structure as a control factor, this regulatory mechanism of alternative splicing does occur in some cases.

4. CD44 And Alternative Splicing

Investigation of the regulation of alternative splicing of the CD44 pre-mRNA presents some unique challenges. Generally, model systems which study alternative splicing deal with the splicing of only one exon. However, within the CD44 pre-mRNA as many as 12 exons can be alternatively spliced. In accordance with this complexity of RNA organization, the regulation of CD44 alternative splicing is expected to be a complex balance of alternative 5' and 3' splice site choices dictated by splice site strengths and expression of tissue specific factors as well as the other modes of regulation.

5. THESIS OBJECTIVE

Multiple protein isoforms of CD44 are generated in a developmental, tissue-specific, and activation-dependent manner by alternative pre-mRNA splicing. The cell surface expression of several CD44 isoforms has been clinically and experimentally correlated with an increase in tumour progression, invasion and metastasis. Therefore, understanding the mechanisms controlling the alternative splicing of CD44 exons will not only add to the basic knowledge of alternative pre-mRNA splicing, but also result in a more complete understanding of the role of CD44 in carcinogenesis.

The objective of this thesis was to investigate the molecular mechanisms that regulate the pre-mRNA splicing of the alternative exons v8, v9 and v10 which encode the unique region of the CD44 isoform, CD44R1. This was accomplished using the following strategies:

1) complete sequencing of the introns separating v8-v9 and v9-v10,

2) construction and testing of a novel *in vivo* pre-mRNA splicing assay, the splice activated gene expression (SAGE) system,

3) detailed examination of the splice products arising from processing of SAGE construct pre-mRNA,

4) identification of exogenous modulators of v8-v9-v10 splicing using the SAGE assay system.

II MATERIALS AND METHODS

1. General Molecular Techniques

General molecular biology techniques utilized throughout this work, such as restriction enzyme digestion of DNA, agarose gel electrophoresis, DNA ligation etc. were performed according to standard protocols as described in (240) or (241). Large scale preparation of DNA was prepared according to the method of Seed *et al.* DNA fragments were isolated from agarose gel slices by the GeneClean II procedure (Bio 101; Vista, CA). All chemicals and plasticware used in this work were supplied by BDH, Inc. (Toronto, ON) and by Falcon (Becton Dickinson Labware, Oxnard CA) respectively, unless otherwise noted.

2. Vectors

The following vectors were utilized throughout this work and are shown in figure 9:

a) pBluescript KS+ (BS KS+; Stratagene Cloning Systems, La Jolla, CA) was used for cloning and sequencing of SAGE constructs. This vector contains the ampicillin resistance gene (β-lactamase, Amp^r phenotype), the ColE1 origin of replication, as well as a multiple cloning site which interrupts the β-galactosidase gene (*lacZ* and *lacI*) and allows for blue/white color selection of transformants on Luria Broth Agar (LBA) following induction with 50 μg/ml 5-bromo-4-chloro-3-indolyl-β-d-galactoside (X-gal) and 15 μg/ml isopropylthio-β-d-galactoside (IPTG; Life Technologies, Burlington, ON). The multiple cloning site also contains sequences complementary to the T7 and M13 reverse primer or T3 and M13 universal primer which were used to sequence the constructs. BS KS+ was propagated in E. coli DH5α grown in Luria Broth (LB), supplemented with 100 μg/ml ampicillin.

b) pCR-Script SK+ (Stratagene Cloning Systems) was used for cloning and sequencing of PCR products. This vector is essentially identical to the BS KS+ vector described above except the multiple cloning site is reversed with respect to the orientation of the *lacZ* gene and an additional *Srf*I restriction enzyme site (arrow) is inserted between the *Sma*I and *Not*I sites.

c) pCEP4 (Invitrogen Corporation; San Diego, CA) was used as an expression vector for the SAGE constructs. This vector contains the β-lactamase gene and the ColE1 origin of replication for maintenance in *E. coli* DH5α in LB supplemented with 100 μg/ml ampicillin, as well as the Epstein Barr Virus (EBV) Nuclear Antigen-1 gene (EBNA-1) and *ori*P sequences necessary for maintenance as an episomal plasmid in EBV transformed primate cells. In addition, the pCEP4 vector contains the hygromycin B phosphotransferase gene flanked by the Herpes Simplex Virus (HSV) thymidine kinase promoter (TK promoter) and polyadenylation signal sequence (TK poly (A)). This arrangement confers the hygromycin B resistance phenotype (hyg^r) and allows for stable expression of the episome in mammalian cells. The gene of interest is inserted into the multiple cloning site which is flanked by the Cytomegalovirus (CMV) immediate-early gene enhancer-promoter (CMV promoter) and the Simian Virus 40 (SV40) polyadenylation signal sequence (SV40 p(A)).

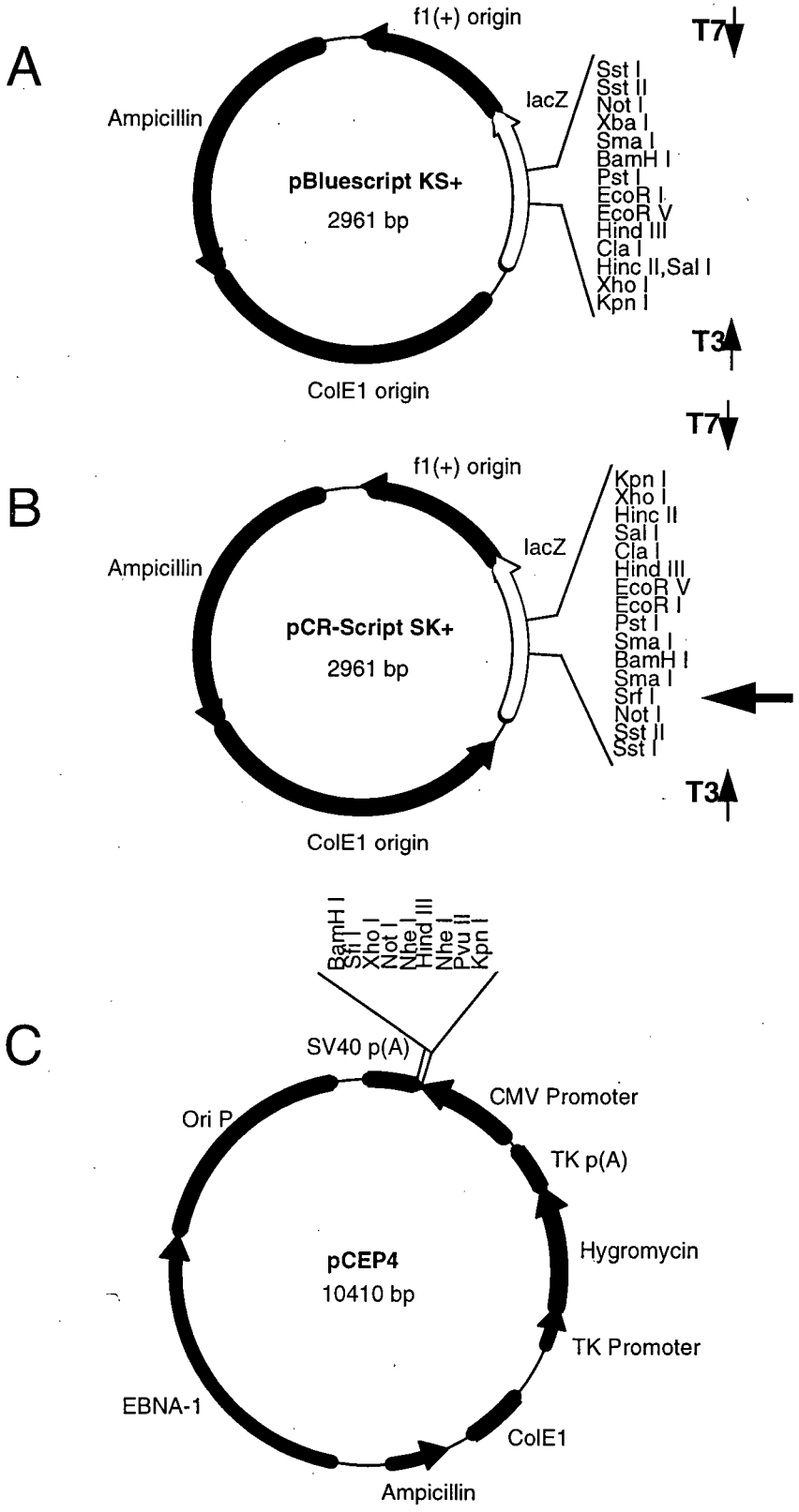


Figure 9: Vectors used in this work. See text for description

3. POLYMERASE CHAIN REACTION (PCR)

3.1 PCR from cDNA

5 µl of 10X PCR buffer (Life Technologies, Inc.), 1.5 µl of 50 mM $MgCl_2$, 0.5 µl of 20 mM dNTP's and 35.75 µl dH_2O were mixed together as a cocktail and transferred to a fresh 500 µl microfuge tube containing 5 µl of the first strand synthesis reaction. 1 µl of each primer (from a 1 mg/ml stock) was added and the reaction was incubated at 72 ^{o}C for 5 minutes (Hot Start PCR). 0.25 µl (1.25 U) of Taq Polymerase (Life Technologies, Inc.) was then added to the warm reaction mix. The tube was gently flicked to mix the contents followed by brief centrifugation. PCR amplification from a pre-warmed 72 ^{o}C Bioscycler Oven (Mandel Scientific Company Ltd.; Guelph, ON) was then initiated using a program of 25 cycles of 94 ^{o}C for 30 seconds, 55 ^{o}C for 1 minute, and 72 ^{o}C for 2.5 minutes. The reaction products were verified by agarose gel electrophoresis.

3.2 PCR from plasmid

PCR reaction conditions used were as stated above with the exception that the volume of template was decreased to 1 µl containing approximately 20 ng of DNA, and the volume of dH_20 in the reaction cocktail was adjusted accordingly. The reaction was assembled on ice, and was not pre-incubated at 72 ^{o}C prior to amplification.

3.3 PCR primers

NAME	PRIMER	RESTRICTION ENZYME SITE (underlined)
CD44R1:		
R1 extracellular 5'	5'-GTCTAGACGCTCCGGACACCATGGAC-3'	*Xba*I
R1 extracellular 3'	5'-GTCTAGACTATTCTGGAATTTGGGG-3'	
Alkaline Phosphatase:		
ALP 5'	5'-GGGTCGACTTAGTGCCAGAGAAAGAG-3'	*Sal*I
ALP 3'	5'-GGGTCGACCCTCAGAACAGGACGCTCAG-3'	
SAGE Construction:		
v8 5'	5'-CCATCGATGGCAGTCATAGTACAACGC-3'	*Cla*I
v9 3'	5'-CCGCTCGAGCGTCAGAGTAGAAGTTGTTGG-3'	*Xho*I
v9 5'	5'-CCATCGATGGGCAGAGTAATTCTCAGACG-3'	*Cla*I
v10 3'	5'-CCGCTCGAGGCGATTGACATTAGAGTTGG-3'	*Xho*I
RT PCR Analysis:		
v9 Sph 5'	5'-ACATGCATCGAGGCTTGGAAGAAG-3'	
v9 Sph 3'	5'-ACATGCATGCGATGTAGAGAAGC-3'	
Rc/AP 5'	5'-GCTTGGTCTCGCCAGTAC-3'	
Rc/AP 3'	5'-GTTACTGTTGGAGATTCC-3'	
Intron 9 5'	5'-GGTGCATCCATTAGCTGCCG-3'	

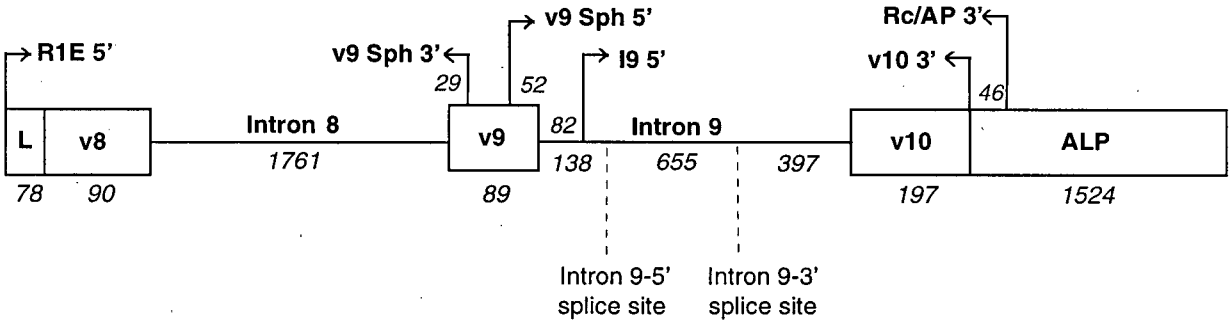


Figure 10: Size representation of v8-v9-v10 SAGE construct. A cartoon representation of the v8-v9-v10 SAGE construct with exons v8, v9, v10, the CD44 leader (L) and ALP represented as boxes, and introns 8 and 9 represented as lines. Arrows indicate primers R1E 5', v9 Sph 5', I9 5', v9 Sph 3', v10 3' and Rc/AP 3'used in the RT PCR analysis of transcripts. Italicized numbers represent the size of each region within the construct in base pairs. Dashed lines indicate the position of the intron 9-5' and intron 9-3' splice sites described in Results section 2.4.

4. PREPARATION OF SPLICE ACTIVATED GENE EXPRESSION (SAGE) VECTORS

4.1 Preparation of CD44 Leader in Bluescript KS+ (BS+L)

The CD44R1 extracellular region was produced from CD44R1 clone 2.3.pCDM8 (39) by PCR as previously described using the R1 extracellular primers (R1E 5' and 3'). The resulting 1215 bp fragment was digested with *Xba*I and ligated into *Xba*I cut Bluescript KS+ and an appropriately oriented clone was selected. This clone was digested with *Cla*I and *Xho*I to remove the mature protein-encoding portion of the cDNA and the vector containing the leader fragment was purified (BS+L vector).

4.2 Preparation of ALP.pCEP4

A leaderless human liver/bone/kidney Alkaline Phosphatase (ALP) fragment was removed via *Sal*I digestion from an R1.ALP.pUC19 construct previously described (48) and ligated into the *Sal*I site of BS KS+. An isolated clone in the forward orientation was selected and digested with *Xho*I and *BamH*I to produce an ALP fragment with asymmetrical ends. This fragment was then force-cloned into *Xho*I and *BamH*I cut pCEP4 .

4.3 Assembly of the SAGE constructs

4.3.1 v8-v9.ALP.pCEP4 (v8-v9 SAGE vector)

An exon cassette consisting of exon v8-intron 8-exon v9 (v8-v9, the presence of intron is denoted by dashes) was produced by PCR from genomic DNA isolated from peripheral blood leukocytes (a generous gift of C. Carpenito; 241) using the v8-5' and v9-3' primers and PCR protocol described in section 3. The resulting ~2 kb product was digested with *Cla*I and *Xho*I and ligated into *Cla*I and *Xho*I

digested BS+L vector (Section 4.1). Positive clones were selected and digested with *Xho*I into which a *Sal*I fragment of leaderless ALP from a chimeric CD44R1.ALP.pUC19 construct previously described (48) was inserted. This clone without the inserted ALP was also used to prepare templates for sequencing of the v8-v9 genomic region. A positive clone was selected, digested with *Kpn*I and *Sst*I, and the fragment containing the fused CD44 leader, v8-v9 exon cassette and ALP (SAGE cassette) was isolated and ligated into *Kpn*I and *Sst*I cut pUC18/13 (a generous gift of Dr. R. Kay). The SAGE cassette was removed from the v8-v9.ALP.pUC18/13 clone via flanking Sal I sites by digestion with *Sal*I, and ligated into *Xho*I digested pCEP4 .

4.3.2 v8-v9-v10.ALP.pCEP4 (v8-v9-v10 SAGE vector)

The v8-intron 8-v9-intron 9-v10 exon cassette (v8-v9-v10) was PCR amplified from genomic DNA using the PCR protocol previously described in section 3 with the following modifications: The PCR reaction mix was denatured by incubation at 95 ^{o}C for 7 minutes prior to the addition of the polymerase. The reaction temperature was then adjusted to 72 ^{o}C and 1 µl of Ampli-*Taq* Polymerase (Life Technologies, Inc.) was added. The exon cassette was then amplified by 35 cycles in a Cetus 96-Well PCR Machine (Perkin Elmer) using the following parameters: 94 ^{o}C for 1 minute, 55 ^{o}C for 30 seconds and 72 ^{o}C for 5 minutes. In addition, the total reaction volume was doubled from 50 µl to 100 µl with all components adjusted as required. A 3.3 kb product was obtained and digested with *Xho*I and *Cla*I and inserted into *Xho*I and *Cla*I cut BS+L (Section 4.1). A positive clone was selected and digested with *Xho*I and *Not*I. The fragment containing the CD44 leader fused to the v8-v9-v10 exon cassette was

isolated and ligated into *Xho*I and *Not*I digested ALP.pCEP4 previously prepared (Section 4.2).

4.3.3 v9-v10.ALP.pCEP4 (v9-v10 SAGE vector)

PCR of the v9-intron 9 v10 exon cassette (v9-v10) was carried out as described in section 3.2 using v9-5' and v10-3' primers and a template prepared from v8-v9-v10 SAGE vector. Assembly of the v9-v10 SAGE vector was performed exactly as described for the v8-v9-v10 SAGE vector (section 4.3.2). The v9-v10.BS+L clone used to generate this construct was also used to prepare templates for sequencing.

4.3.4 v89-10.ALP.pCEP4 (v89-10 SAGE vector)

A construct which consisted of exon v8 pre-spliced to exon v9, followed by intron 9 and exon v10 (v89-10, see figure 11) was prepared using a PCR ligation technique. Briefly, a v89 construct was amplified from CD44R1 clone 2.3.pCDM8 (39) using the v8-5' and v9-3' primers as previously described. Similarly a v9-intron 9-v10 PCR product was obtained by amplification of genomic DNA using the v9-5' and v10-3' primers. These two PCR products were purified and mixed together in a single cycle PCR reaction (section 3.1). This reaction was incubated for 7 minutes at 72 ^{o}C to allow for a single extension of the template. v8-5' and v10-3' primers were then added and the PCR amplification was continued as previously described (section 3.1). The PCR products were separated by agarose gel electrophoresis and a 1700 bp band was isolated and ligated into the BS+L and ALP.pCEP4 vectors as previously described. Correct extension of the v89-10 SAGE construct was confirmed by restriction enzyme digest and sequencing.

4.3.5 v89.ALP.pCEP4 (v89 SAGE vector) and v910.ALP.pCEP4 (v910 SAGEvector)

cDNA SAGE controls consisting of exons v89 or v910 or v8910 (without the intron) were constructed exactly as previously described for v8-v9-v10 SAGE vector using the v8-5' and v9-3' or v9-5' and v10-3' primers on a template from CD44R1 clone 2.3.pCDM8.

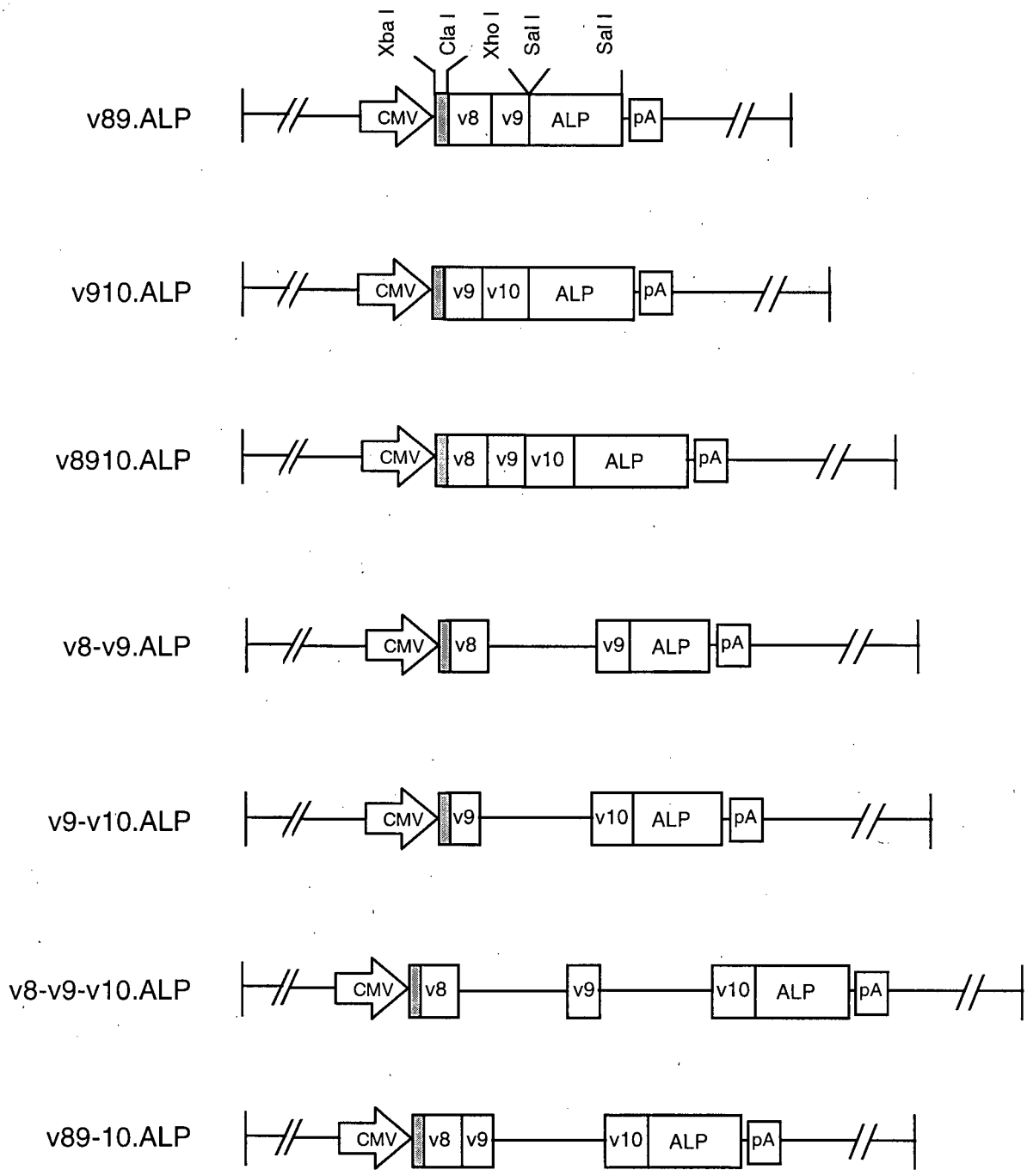


Figure 11: Representation of SAGE vector constructs. SAGE vectors were constructed as described in the text and inserted into the expression vector pCEP4 under the transcriptional control of the CMV promoter (CMV) and SV40 polyadenylation sequence (pA). Exons v8, v9, v10 and ALP are shown as open boxes, the CD44 leader is shown as a shaded box, and the intron(s) is represented as a line.

5. PREPARATION OF TOTAL RNA

5×10^6 cells were pelleted by centrifugation and lysed in 1 ml of TRIzol reagent (Life Technologies, Inc.) as per the manufacturer's instructions. After 5 minutes incubation at room temperature, 0.2 ml of chloroform was added. The tubes were shaken for 15 seconds and phases separated by centrifugation for 15 minutes at 4 ^{o}C at the top speed of a microfuge (Hermle A 231 M; Mandel Scientific Company Ltd.). The upper phase was recovered and the RNA was precipitated by the addition of 0.6 ml of isopropanol followed by 10 minutes incubation at room temperature, and centrifugation at 4 ^{o}C for 10 minutes. After drying, the RNA was dissolved in 10 to 20 µl dH_2O.

6. REMOVAL OF CONTAMINATING DNA

The total volume of dissolved RNA in distilled H_2O was adjusted to 20 µl to which 2 µl of 10X DNAse I buffer, 1 U of DNAse I (Life Technologies, Inc.) and 2 U of calf intestinal RNAse inhibitor (RNAsin; Life Technologies, Inc.) was added. RNA was then incubated at 37 ^{o}C for 1 hour to degrade any remaining DNA. The reaction was stopped with 2 µl of 0.2 M EDTA and the RNA was cleaned by phenol:chloroform:iso-amyl alcohol (25:24:1) extraction, followed by a second chloform:iso-amyl alcohol (24:1) extraction and ethanol precipitation.

7. FIRST STRAND SYNTHESIS OF cDNA

Total RNA in dH_2O was quantified by absorbance at 260 nm (OD_{260}=40 µg/ml; LKB Ultrospec III; Pharmacia Biotech; Uppsala, Sweden). 5 µg of total RNA was adjusted to a total volume of 12.5 µl in dH_2O, denatured at 65 ^{o}C for 10 minutes and placed on ice. A cocktail of 10 µl 5X first strand synthesis buffer (Life Technologies,

Inc.), 5 µl 0.1 mM dithiothreitol (DTT), 1.0 µl of 20 mM dNTP's (Phamacia Biotech), 1 µl of 0.5 mg/ml oligo d(T) (Pharmacia Biotech), 5 U of RNAsin, 18.75 µl dH2O and 200 U of MMLV Reverse Transcriptase (Life Technologies, Inc.) was added and first strand synthesis was carried out for 1 hour at 37 ^{o}C.

8. NORTHERN BLOTS

A cocktail of 6 µl formamide dye mix (de-ionized formamide plus 10% xylene cyanol and 10% bromophenol blue), 2 µl de-ionized formaldehyde and 0.6 µl 10X MOPS (0.36 M 3-[N-morpholino] propane sulphonic acid, 0.01 M EDTA, 0.18 M glacial acetic acid, pH 7.0) was added to 10 µg of total RNA dissolved in 5 µl dH_2O. The mixture was heated to 65 ^{o}C for 10 minutes before loading into a formaldehyde/agarose gel (5% v/v formaldehyde, 1% w/v agarose, 1X MOPS buffer). The RNA was electrophoresed through the gel in 1X MOPS buffer for approximately 2 hours at 70V or until the bromophenol blue dye front had migrated 7 cm. The gel was then washed briefly in dH_2O and soaked for 20 minutes in 10X SSC (from 20X SSC stock: 3 M NaCl, 0.3 M tri-sodium citrate pH 7.0). RNA was transferred to a supported nylon membrane (Zetaprobe GT; Bio-Rad Laboratories; Hercules, CA) by capillary transfer using 10X SSC. RNA was fixed to the membrane by UV crosslinking (UV Stratalinker 1800; Stratagene Cloning Systems), and the membrane was washed for 5 minute in 1% SDS. The membrane was incubated for at least 2 hours in a hybridization solution consisting of 4 ml de-ionized formamide, 2 ml of 2 M NaH_2PO_4 in 4 mM EDTA pH 7.2 and 80 µl 100 mg/ml bovine serum albumin (BSA; Sigma Chemical Company) at 42 ^{o}C in a Micro-4 Hybaid oven (InterScience Inc.; Markham, ON). Radiolabeling of probes was performed using a Random Primer Oligolabeling Kit (Life Technologies, Inc.) as per manufacturer's instructions. The probe was denatured by the addition of 2 µl 10N NaOH and was added to the

hybridization solution. Hybridization was carried out overnight at 42 ^{o}C. The membrane was then washed at 55 ^{o}C twice with 2X SSPE/0.3% SDS (20X SSPE stock: 3M NaCl, 0.2M NaH_2PO_4, 0.02M EDTA pH 7.4), twice with 1X SSPE/0.5%SDS and once with 0.3X SSPE/1% SDS and was placed on film (Kodak X-OMAT AR) and developed (Fuji RG II X-Ray Film Processor).

8.1 Probes

The various probes used in this work were constructed as follows:

a) unique region probe: clone 2.3.pCDM8 (CD44R1) was digested with *Xmn*I and *Pst*I and the 1200 bp band was recovered. This DNA fragment was further digested with *Hinf*I and the 340 bp band was purified. The ends of this fragment were filled in using the standard T4 polymerase technique, and the fragment was ligated into *Hinc*II digested pUC 18/13. A suitable clone was selected and digested with *BamH*I to recover the 340 bp fragment. The probe DNA was purified from a 1% agarose gel using the GeneClean II technique.

b) ALP probe: The full length fragment of ALP generated during construction of the SAGE vectors (Materials and Methods section 4.2) was used as a probe.

c) Exon 8 probe: An exon 8 specific probe was generated by PCR from clone 2.3.pCDM8 template as described in Materials and Methods section 3.2 using the v8 5' and v9 Sph 3' primers. The resulting PCR product was digested with *Dde*I and the 110 bp product was purified by agarose gel electrophoresis and GeneCleaning.

d) Exon 910 and exon 10 specific probes: A specific probe spanning exons 9 and 10 (v910 probe) was generated by PCR from clone 2.3.pCDM8 as described above using v9 Sph 5' and v10 3' primers. The resulting PCR products were separated on a 1% agarose gel, and the 270 bp band was purified by standard protocols. To make a v10 specific probe, the v910 probe was digested with *Fok*I, and the 140 bp band was purified as described above.

e) Actin probe: A 1.8 kb fragment containing the cDNA encoding actin was isolated by digestion of plasmid pA1-actin with *Pst*I.

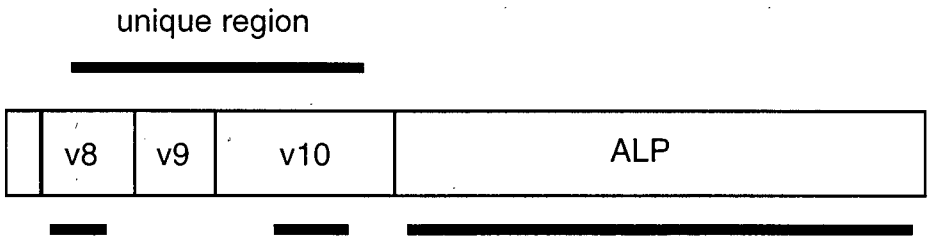


Figure 12: Probes used in this work. Probes were constructed as described in the text, but are shown with positions marked relative to v8910.ALP SAGE construct for reference.

9. SOUTHERN BLOTS

1 µl of PCR product diluted in 9 µl of dH_2O was run out on a 1 % w/v agarose gel in a standard 1 X Tris acetate buffer with 0.2 µg/ml ethidium bromide until the bromophenol blue dye front had migrated at least 7 cm. The gel was then photographed for reference, washed briefly in dH_2O, and soaked in 0.4 M NaOH for 20 minutes to denature the DNA, followed by 20 minutes in 2 X SSC buffer. The DNA was then transferred overnight to Zetaprobe GT membrane by standard capillary transfer with 10 X SSC. The membrane was then probed using ^{32}P-labeled

probes using hybridization conditions as described for Northern blots. The labeled blots were washed twice in room temperature 2 X SSC/0.1% SDS and twice in 0.1 X SSC/ 0.1% SDS at 60-65 ^{o}C depending on the probe used. The membrane was placed on autoradiography film (Kodak X-OMAT AR) and developed on an automatic X-Ray film processor (Fuji RG II X-Ray Film Processor).

10. DENSITOMETRY

Densitometry of the Southern blots was determined by analysis on a Phosphor Imager (Molecular Dynamics) using the ImageQuaNT program. Calculation of percentage density was performed by dividing the density of each band by the total density of each lane.

11. CELL LINES

K562, a human erythroleukemia cell line, KG1 and KG1a, human acute myelogenous leukemic cell lines, were obtained from American *Type Culture* Collection (ATCC) and maintained in an humidified incubator at 37 ^{o}C and 5% CO_2. The growth medium consisted of Dulbecco's Modified Eagle Medium (DMEM; StemCell Technologies, Inc.; Vancouver, BC) supplemented with 10% Cool Calf II (CCII; Sigma Chemical Company).

12. ELECTROPORATION OF K562 CELLS

K562 cells maintained in log phase in DMEM + 10% CCII were washed once in serum-free warm (37 ^{o}C) DMEM, and resuspended at 1 X 10^{7} cells/ml in serum-free DMEM. 1 ml aliquots of cells were pipetted into sterile 5 ml polypropylene tubes, 20 µg of CsCl purified DNA was gently mixed into each aliquot, followed by 5 µg of DEAE-Dextran (242; Pharmacia Biotech). DNA was allowed to adhere to the K562 cells for 5 min prior to electroporation. After incubation, 800 µl of the cell solution was transferred to a sterile 0.4 cm electroporation cuvette (Bio-Rad Laboratories) and electroporated using a Bio-Rad Gene Pulser set at 960 mF and 0.25 kV. The cells were then plated into a 10 cm tissue culture dish. The cuvette was washed once with serum-free DMEM and the washing was added to the dish. 10 ml of warm DMEM + 10% CCII was then added and the cells were replaced in the incubator.

13. SELECTION OF SAGE-EXPRESSING K562 CELLS

Electroporated K562 cells were incubated for 2 days, washed once in phosphate buffered saline (PBS; StemCell Technologies, Inc.) and re-plated into a 10 cm dish in 10 ml of DMEM + 10% CCII containing 200 µg/ml hyrogromycin (Sigma Chemical Company). Polyclonal selection of cells was performed in the continual presence of hygromycin with media changes at 2 day intervals. For monoclonal populations, cells at the crisis stage (when greater than 90% death was observed) were washed once in PBS and suspended in 1 ml of DMEM + 10% CCII in a sterile 15 ml tube. Nine ml of methylcellulose plating medium (StemCell Technologies, Inc.) was added (0.9% methylcellulose in DMEM + 20% CCII + 200 µg/ml hygromycin) and well mixed. Cells were plated into 10 cm Petri dishes and incubated at 37 ^{o}C and 5% CO_2 for 5 to 10 days until visible colonies formed. The colonies were then plucked into 200 µl of

DMEM + 10% CCII + 200 μg/ml hygromycin, tested by cytospin analysis and enzymatic staining, and expanded.

14. CYTOSPIN ANALYSIS AND STAINING

Transfected K562 cells in log phase were suspended at 1 X 10^5 cells/ml. 100 μl of cells were transferred to cytospin chambers and spun onto glass slides using a cytospin machine (Shandon Elliot). The slides were allowed to dry, then cells were fixed for 30 seconds in 100% acetone. After drying, the enzymatic ALP stain 5-bromo-4-chloro-3-indolyl phosphate/p-iodonitrotetrazolium (BCIP/INT; Moss Inc.) was added drop wise to the slide. Alkaline phosphatase (ALP) development was allowed to proceed for 5-10 minutes at 37 ^{o}C. The cells were then counter-stained for 30 seconds in Gill's hematoxylin (Sigma Chemical Company) rinsed in water and allowed to dry. The presence of ALP positive cells was indicated by a deep orange/brown stain, in opposition to the blue hematoxylin-stained nuclei. Staining was observed by microscopy (Nikon TMS microscope). The slides were then stored unmounted, in the dark at room temperature.

15. INDUCTION OF SAGE EXPRESSION

Transfected K562 cells were treated with either nil (media only), 5 μg/ml 12-o-tetradecanoyl phorbol-13-acetate (TPA; Sigma Chemical Company) or 10 ng/ml staurosporine (Sigma Chemical Company) for 48-72 hours prior to testing for ALP surface expression by cytospin, FACS analysis, or RT PCR.

16. FLUORESCENCE ACTIVATED CELL SORTING (FACS) ANALYSIS

The cell population to be analyzed was suspended at approximately 1 X 10^7 cells/ml in DMEM + 10% CCII. 10 µl of cells were added to the bottom of a round-bottom flexible assay plate (Falcon #3911), followed by 100 µl of either nil (DMEM + 10% CCII) or primary antibody. The primary antibody was added as either undiluted tissue culture supernatants (mAb A3A11, a mouse anti-human alkaline phosphatase) or as 1/1000 dilutions in DMEM + 10% CCII of a MAb acites preparations (MAb 4A4, a mouse anti-human CD44 common region, or MAb 2G1, a mouse anti-human CD44 exon v10 as described in (72)). Cells were incubated on ice for 30 minutes, followed by 3 washes with ice-cold Hank's balanced salt solution (HBSS) + 2% CCII, after which 100 µl of fluorescein isothiocyanate (FITC)-labeled goat anti-mouse secondary antibody (Sigma Chemical Company) diluted 1/128 in HBSS + 2% CCII was added. A further 30 minute incubation of cells in the secondary antibody was followed by 3 washes in HBSS + 2% CCII. After the last wash, the cells were suspended in 500 µl HBSS + 2% CCII + 1 µg/ml propidium iodide (PI ;Sigma Chemical Company) and were analyzed on either a FACScan or FACSort (Becton Dickinson Immunocytometry Systems; San Jose, CA).

17. SEQUENCING

17.1 General Sequencing protocols

17.1.1 Preparation of templates

Plasmid templates for sequencing were prepared using a standard alkaline lysis technique (240). RNA contamination was removed by digestion with 50 µg/ml RNAse A (Sigma Chemical Company) and 100 U RNAse T1 (Boehringer Mannheim; Laval, PQ) for 30 minutes at 37 ^{o}C, followed by

phenol:chloroform:iso-amyl alcohol (25:24:1) extraction and ethanol precipitation. The purified DNA was resuspended in dH_2O.

17.1.2 Sequencing of the templates

Sequencing reactions were performed by the dideoxy chain termination method from double stranded Bluescript KS+ or pCR-Script SK+ templates using the Sequenase Version 2.0 sequencing kit (USB Biochemical Corporation; Cleveland, OH) modified for ^{32}P-dCTP as per the manufacturer's instructions.

17.1.3 Polyacrylamide Gel Electrophoresis and Autoradiography

The completed sequencing reactions were separated on an 8% polyacrylamide wedge gel (Gel-Mix 8; Life Technologies, Inc.) using a Sequi-Gen Nucleic Acid Sequencing Cell (Bio-Rad Laboratories). Electrophoresis was carried out at 50-55W constant power such that the gel temperature was maintained at 50 ^{o}C. Buffers used were those recommended by the manufacturer. After electrophoresis, the gel was dried for 2 hours on a Bio-Rad Model 583 gel dryer and visualized by autoradiography on Kodak X-OMAT AR film. The sequence was assembled using the computer programs DNA Strider and the GCG Wisconsin Package.

17.2 Exonuclease III deletion of the v8-v9.BS+L template

v8-v9.BS+L vector DNA was prepared by standard CsCl large scale preparation and was digested with *Kpn*I and *Xho*I. A set of progressively deleted templates was prepared by Exonuclease III digestion (Life Technologies, Inc.) using standard buffers and conditions (240). Aliquots were removed to the stop buffer at 45 second intervals. The ends of the deleted constructs were filled using Mung Bean Nuclease treatment (Life Technologies, Inc.) and re-ligated using standard protocols. Selected

clones which exhibited 200 bp progressive deletions as determined by agarose gel electrophoresis, were chosen, purified and sequenced.

17.3 Bal 31 deletion of the v9-v10.BS+L template

Despite several attempts, the v9-v10.BS+L vector proved to be refractory to Exonuclease III digestion. Thus a set of deleted constructs was prepared by *Bal*31 Nuclease digestion (New England Biolabs Ltd.; Mississauga, ON). Purified v9-v10.BS+L was digested to completion with *Kpn*I, and treated with *Bal*31 Nuclease as per standard protocol (241). Aliquots were removed to stop buffer at 5 minute intervals. The ends were repaired with the Klenow fragment of DNA Polymerase I (Life Technologies, Inc.) prior to secondary digestion with *Sst*I. Successively deleted fragments were purified by agarose gel electrophoresis and ligated into *Sma*I and *Sst*I digested BS KS+. Clones which displayed progressive deletions of approximately 200 bp were selected, purified and sequenced.

17.4 Sequencing of aberrantly spliced v9-10 PCR products

RT PCR of RNA from four separate clones of v8-v9-v10 SAGE K562 cells, and two separate clones of v9-v10 SAGE K562 cells was performed using PCR primers v9 Sph 5' and Rc/AP3'. The ends of the resulting PCR products were repaired by standard T4 Polymerase treatment (Life Technologies, Inc.) and the products were ligated into the PCR cloning vector pCR-Script SK+ via the *Srf*I site as described in the manufacturers' instruction manual. The resulting clones were analyzed by restriction enzyme digest and one clone each of the 292 bp, 450 bp, and 750 bp PCR products from each of the RT PCR's was selected, purified and sequenced.

18. - MONOCLONAL ANTIBODY PRODUCTION

To generate mAbs against liver/bone/kidney ALP for use in subsequent detection and sorting of SAGE populations, a single clone of K562 cells stably transfected with the cDNA-like v910.ALP.pCEP4 (v910 SAGE) vector which showed greater than 90% reactivity with the enzymatic ALP stain BCIP/INT, was used to immunize 3 C3H/HEJ mice. 5×10^6 irradiated cells (30 Gy from an Phillips RT 250 X-Ray Machine) were administered twice intraperitoneally, and once intravenously at 21 day intervals. Three days after the last injection, the mice were sacrificed and spleens were resected. Spleen cells were collected and fused to the hybridoma fusion partner, SP2/0 by polyethylene glycol (PEG) fusion as described (72). Single clones were selected in methylcellulose growth medium (StemCell Technologies, Inc.) containing oxaloacetate/pyruvate/insulin (OPI) and hypoxanthine/aminopterin/thymidine (HAT) supplementation (Sigma Chemical Company). The resulting single colonies were plucked into DMEM + 10% Fetal Clone I (Hyclone Laboratories Inc.; Logan, UT) with HAT supplementation and expanded. Supernatants were tested against v910 SAGE K562 cells and untransfected K562 controls by FACS. One positive clone, MAb A3, was demonstrated to recognize an epitope present on the alkaline phosphatase portion of the SAGE construct as it exhibited a high level of reactivity against the v910 SAGE K562 as well as K562 cells transfected with alkaline phosphatase in pCEP4 (ALK.pCEP4; 48) but not against untransfected controls. This hybridoma clone was subjected to further clonal selection in methylcellulose growth medium to isolate high level antibody producers. Clones were obtained and tested by FACS analysis against v910 SAGE K562 cells and one clone, MAb A3A11, was selected for use in the analysis of SAGE expression.

III RESULTS

1. CHARACTERIZATION OF THE GENOMIC DNA SPANNING THE CD44R1 EXON v8, v9 AND v10 REGION

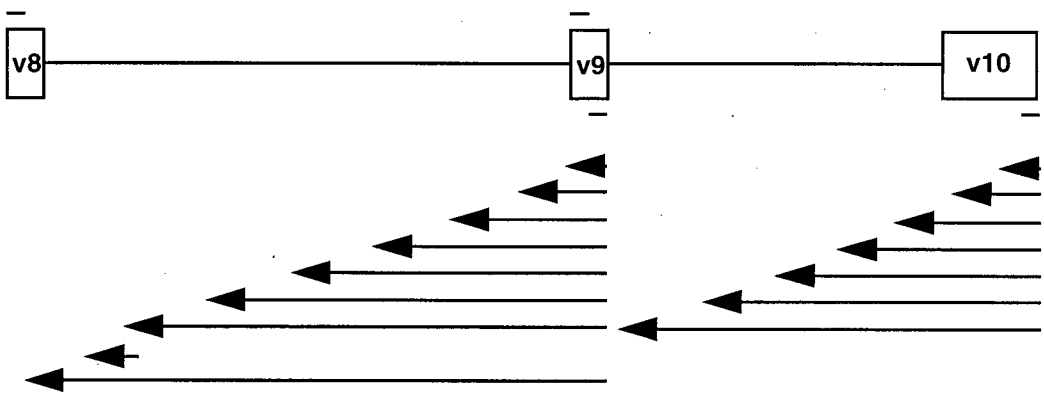
The nucleotide sequence of the cDNA encoding CD44R1 has been previously determined (39). Similarly, the intron/exon boundaries of the alternatively spliced exons (v8, v9, and v10) which encode the unique region of CD44R1 have been characterized (7-9). However, the full sequence of the introns separating v8, v9 and v10 was unknown. Thus to begin an analysis of the alternative splicing of the unique region exons of CD44R1, the genomic sequence of *CD44* spanning exons v8-v9-v10 was first determined.

1.1 Sequence of the v8-v9-v10 genomic region

Clones spanning the v8 to v9 and v9 to v10 region were generated from genomic DNA by PCR with exon specific primers v8 5' to v9 3' and v9 5' to v10 3' which had been derived from previously published cDNA sequence (Materials and Methods section 3.3). The resulting 2 kb and 1.4 kb PCR products were inserted into the cloning vector pBluescript KS+ and a progressive set of deletions was generated by Exonuclease III (v8-v9) or Bal 31 nuclease (v9-v10) digestion (Figure 13). Sequencing was performed by the dideoxy method of Sanger *et al.* The full sequence of the v8-v9-v10 region was assembled from these clones and is shown in Figure 13.

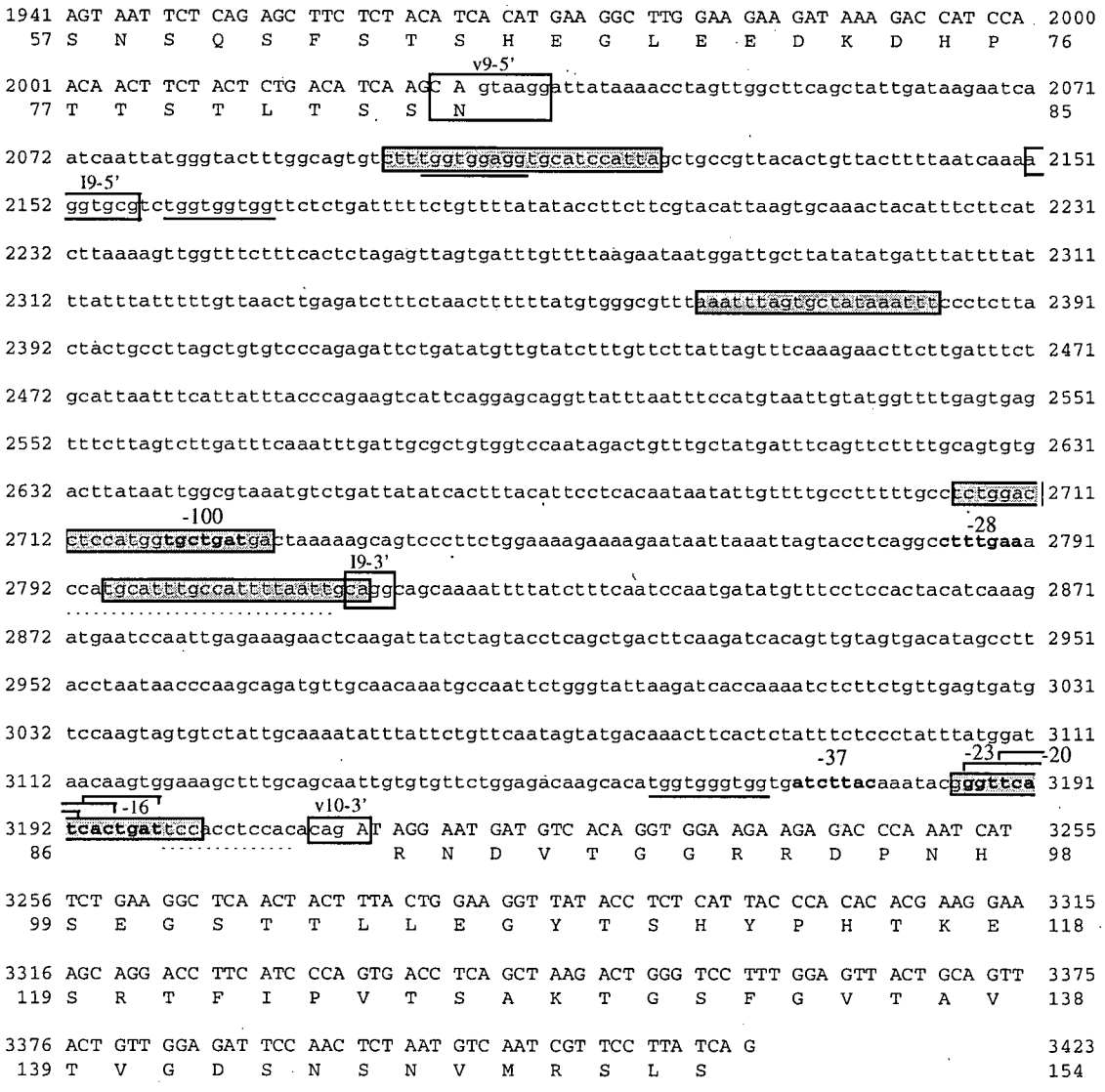
Figure 13: The genomic sequence of CD44 v8-v9-v10 region. The genomic region of CD44 spanning exons v8, v9, and v10 was cloned using a PCR-based approach and sequenced. (A) A cartoon representation of the CD44 v8-v9-v10 genomic region with the position of the PCR primers shown for reference (see Materials and Methods section 3.3 for primer sequence). Progressive sets of deletions (arrows) were constructed by Exonuclease III or *Bal*31 digestion. (B) Genomic sequence of the v8-v9-v10 region with various putative *cis*-acting sequences detailed. Nucleotides encoding exons are shown in uppercase, with those encoding introns in lowercase. Open boxes indicate 5' or 3' splice sites as marked and shaded boxes indicate a putative stem-loop structure. Putative branch point regions are denoted by bold text with distance of the branch point adenine to the 3' splice site noted above. Polypyrimidine tracts are indicated with a dashed underline. Single underline and double underline indicate a YGGTGGTGG sequence, or repeated sequence, respectively.

A



B

```
  1 CGCTCCGGACACC ATG GAC AAG TTT TGG TGG CAC GCA GCC TGG GGA CTC TGC CTC GTG CCG    61
  1               M   D   K   F   W   W   H   A   A   W   G   L   C   L   V   P    16

 62 CTG AGC CTG GCG CAG ATC GAT GGC AGT CAT AGT ACA ACG CTT CAG CCT ACT GCA AAT CCA  121
 17 L   S   L   A   Q   I   D   G   S   H   S   T   T   L   Q   P   T   A   N   P   36

122 AAC ACA GGT TTG GTG GAA GAT TTG GAC AGG ACA GGA CCT CTT TCA ATG ACA ACG C gtaaga 182
 37 N   T   G   L   V   E   D   L   D   R   T   G   P   L   S   M   T   T   Q       55

183 ataacgatgctgagccactttattgacttgtatcctgcttcatctcttcccgctattggccaagattgacagagctttgg 262

263 tggtggaatggtgctatgtggcttacttcagcccagagtgtgaaactgtcttcattgtctgtatttcttgtggaggtctc 342

343 tcaacctgattttcccttcatttctttgtgccttccagatttaatgactttgtgttttaagaaaaagtaggactctattc 422

423 aactccattttcattctctaagtagttatgtcagtgaaacacatgcccacttccaggatagtgatttctctatgttttaa 502

503 ctgaccatttcctttcaatgtccatctttcctccagccagtgcagaagaggggtaggattgcttgcatggttgtaaaaag 582

583 acatttttggggccagataatcctagagatcacatcagttctcagaggagaggcaaggcatgttattgacaatcaacaaa 662

663 ggtggcgggatccctggaacaacaacaacaacaaaacattgccagtatctctttctatcactaagaaatgttcaaattct 742

743 agtagccttaatttctactctatgtgcagggatgtcctagagatcagcctaaagtcgttcataatttctttgaggaatac 822

823 aaaagaaatctagaattaagaaggtccattttactgcttgccacactataatgggcgagtccttcatatgagcccttta 902

903 ggcaaaggaaaaaaatcagcaagcatgtattgatggtcagcaatgtaaatagcacagcgttaagcatcaaaggagatgca 982

983 aataagcctatgaatggtccttgccaaagaaggacttatatttccgttttatgtcactgttggcttttgttctgacattg 1062

1063 ggtttgggatttagtgactgtctttctcctggttgtaatatttattcggtggtgggcatggcttatgggtgtgctttatg 1142

1143 aagtttttgcacctgcaactcaggtcagtacctttcttaaaagaacacaaaagaaataaacttgaaacagtccaagtgat 1222

1223 tttagatatgggcaatgggtcaacctccagttgatatatgtcactgagggccctgtgtttgcccaggcaagcgtttccttt 1302

1303 ggtgcagtgatgaaaatcccaatgcacaggtgaaattaagaaataacttaataatggagtaaaatggtaactttaattta 1382

1383 atgacctttacaaggtagctaagtatttccttgtataaatttgatctgtgttacttgagatggctattctagataggctt 1462

1463 ctttttcttttgtggttacaactagactacaggctatgatagtgttcattttccctgccgtcacatttcttgacacacag 1542

1543 ttaggtgcttttggttactgataggagcattgggattcaaatattctttttttccatgtctctgtgttctctggtttctg 1622

1623 gagtgactgggagcaatgctgcaattgagcctggacaatccaggataggcaaagttttggctcagtcttcaccttgccac 1702

1703 tttaggtttgatgattatactgagaaataaagggtcttcagaaccaaatagggggaaaatagaattcttctgtatatttc 1782

                                                                    -100
1783 tatcaatcagaggagaattgtgtgcaacaaaccttggtcttcctatgcacctatccatctgactctactctagctagatt 1862

                                                  -27
1863 ttccttgctagcagataacaggatttgtaccgtagcttcaattaacactggattcactcctcattgaaacag AG CAG 1940
  56                                                                          Q    56
```


v9-5'
I9-5'
-100
-28
I9-3'
-37
-23
-20
-16
v10-3'

1.2 Analysis of cis-acting splice sequences

The 5' and 3' splice site sequences flanking the intron/exon boundaries of exons v8, v9, and v10 were located by comparison to previously described 5' (C/AAG | GTA/GAGT) and 3' (CAG | G) splice site consensus sequences (108). A comparison of these splice site sequences with their respective consensus (Figure 14) reveals that while the exonic portion of the splice sites varies, the majority of the intronic portion remains conserved.

A	**5' Splice Site**	**$^{C}_{A}$AG | GT$^{A}_{G}$AGT**
	v8-5'	CGC | GTAAGAATA
	v9-5'	GCA | GTAAGGATT
B	**3' Splice Site**	**CAG | G**
	v9-3'	CAG | A
	v10-3'	CAG | A
C	**Branch point**	**YNYURAC**
	I8 -27	AATTAAC
	I8 -100	ATCTGAC
	I9 -16	CACTGAT
	I9 -20	TCATCAC
	I9 -23	GGTTCAT
	I9 -37	ATCTTAC

Figure 14: Consensus splice site sequences of the v8-v9-v10 region. (A) 5' and (B) 3' splice sites flanking the intron/exon boundaries and (C) putative branch point regions CD44 exons v8, v9 and v10 are aligned, with published consensus shown in bold.

Branch point regions are normally located between 18 to 40 nt from the 3' splice site (number of nucleotides from the branch point adenine to the 3' splice site acceptor guanine) due to the steric constraints of the spliceosomal complex (107). Two putative branch point regions were found within intron 8. The first was found within the conserved distance at -27

and contains a pyrimidine to purine substitution at residue 1. The second, a 'long range' branch point due to its extended distance from the 3' splice site, was found at -100 and contains the same substitution. Intron 9 was found to contain multiple overlapping putative branch point regions located at -16, -20, -23 and -37 nt, each diverging from the consensus by a pyrimidine to purine substitution at residues 1 or 3 (Figure 14). In comparison with previously described polypyrimidine tracts (159, 160), those located between the 3' splice site and the branch point region of the CD44 variant region exons were found to be highly purine rich, varying from 60% pyrimidines (v8-v9) to 60-80% pyrimidines (v9-v10) depending on the branch point used (Figure 13B).

In addition, upon close inspection of the nucleotide sequence, several putative *cis*-acting sequence motifs were also identified and are indicated in Figure 13. These include:

1) a YGGTGGTGG motif found twice in intron 8 at 90 nt downstream from the 5' splice site, and 830 nt upstream from the 3' splice site. The same motif was found three times within intron 9 in close proximity to the splice sites at 79 nt and 139 nt downstream from the 5', and 49 nt upstream from the 3' splice site.
2) Using the method of Zuker on the GCG Wisconsin package, four putative strong stem-loop structures were found within intron 8 and four within intron 9. Note that the last intron 9 stem loop structure is formed close to the 3' splice site (-21 nt), overlapping with the multiple branch point regions.
3) An 11 base pair repeat (CCTAGAGATCA) is found twice within intron 8 at 388 nt and 600 nt downstream from the 5' splice site.
4) A repeated purine and pyrimidine motif (Y_4RY_7 or Y_7RY_4) occurs twice within the first 200 nt of intron 8. The same motif is found five times in intron 9, three times between 150 and 300 nt downstream from the 5' splice site, once in the middle of the intron, and lastly 120 nt upstream from the 3' splice site.

2. ASSEMBLY AND FUNCTIONAL ANALYSIS OF SPLICE ACTIVATED GENE EXPRESSION VECTORS

To determine and characterize the molecular mechanisms which regulate the splicing of the CD44R1 unique region exons, a set of *in vitro* splicing vectors were developed. These vectors, termed splice activated gene expression (SAGE) vectors, consist of either v8-intron 8-v9 (referred to as v8-v9, where the dash indicates the presence of the intron), v9-intron 9-v10 (v9-v10) or a full length construct of v8-intron 8-v9-intron 9-v10 (v8-v9-v10) fused in frame to a leaderless construct of human liver/bone/kidney alkaline phosphatase (ALP). ALP is normally expressed as a cell surface glycophosphoinositol (GPI)-linked tetramer due to the double hydrophobic sequences found at the amino and carboxy terminals of the protein (243, 244). In the SAGE vector system, the upstream signal sequence was removed and replaced by the CD44 leader placed 5' to the splicing exons. Thus expression and activity of ALP is dependent on the prior correct splicing of the CD44R1 unique region exons.

2.1 Construction of the Splice Activated Gene Expression Vectors

The SAGE vectors were constructed by PCR of the CD44 v8-v9, v9-v10 or v8-v9-v10 region as an 'exon cassette' from genomic DNA using PCR primers v8 5' to v9 3', v9 5' to v10 3' and v8 5' to v10 3' as described (Materials and Methods section 4). The exon cassettes were then fused to a leaderless construct of ALP in the expression vector pCEP4 (Figure 9C). Transcription of the constructs was under the control of the CMV promoter, with stable replication maintained by the presence of the hygromycin resistance gene and by the Epstein Barr Virus (EBV) sequences encoding EBNA-1 and ORI P which are necessary for episomal replication of the pCEP4 vector. In initial experiments, several hemopoietic (K562, KG1, U937, Raji, Jurkat and Daudi) and non-hemopoietic (Cos-7 and Eahy.926) cell lines were transiently transfected with the SAGE vectors, and histological assays for ALP expression with the enzymatic stain, BCIP/INT, were performed as described in Materials and Methods section 14. Interestingly, while each of these cell lines have been observed to expresses

various levels of CD44 isoforms , a similar pattern of ALP expression was noted. Namely, cells transfected with the v8-v9 SAGE vector demonstated positive staining for ALP, while little or no positive staining was detected on those transfected with the v8-v9-v10 and v9-v10 SAGE vector respectively. Control pCEP4 transfected cells stained negatively in all cell lines examined (data not shown).

K562, a CD44 negative human erythroleukemia cell line may be induced to express multiple CD44 isoforms after phorbol ester treatment (39). This cell line was chosen as a representative for further study of the splicing of the SAGE vector transcripts, and was transfected with the SAGE vectors by electroporation. Several monoclonal and polyclonal populations resulting from multiple transfections were selected for stable maintenance of the SAGE vectors by hygromycin resistance, and splicing of SAGE transcripts was analyzed by histological staining for ALP expression. Typical results of monoclonal (v8-v9 and pCEP4) and polyclonal (v9-v10 and v8-v9-v10) populations are shown in Figure 15. While greater than 90% of stably transfected K562 cells expressing the v8-v9 SAGE construct demonstrated positive orange/brown staining (Figure 15B), less than 10% of v8-v9-v10 SAGE and no v9-v10 SAGE construct expressing cells showed positive staining (Figure 15C and D). Multiple monoclonal and polyclonal populations of each transfected SAGE vector were examined with similar staining patterns and proportion of ALP positive cells observed in each population. Representative monoclonal (pCEP4 and v8-v9) and polyclonal (v9-v10 and v8-v9-v10) populations of transfected K562 cells were selected for further analysis.

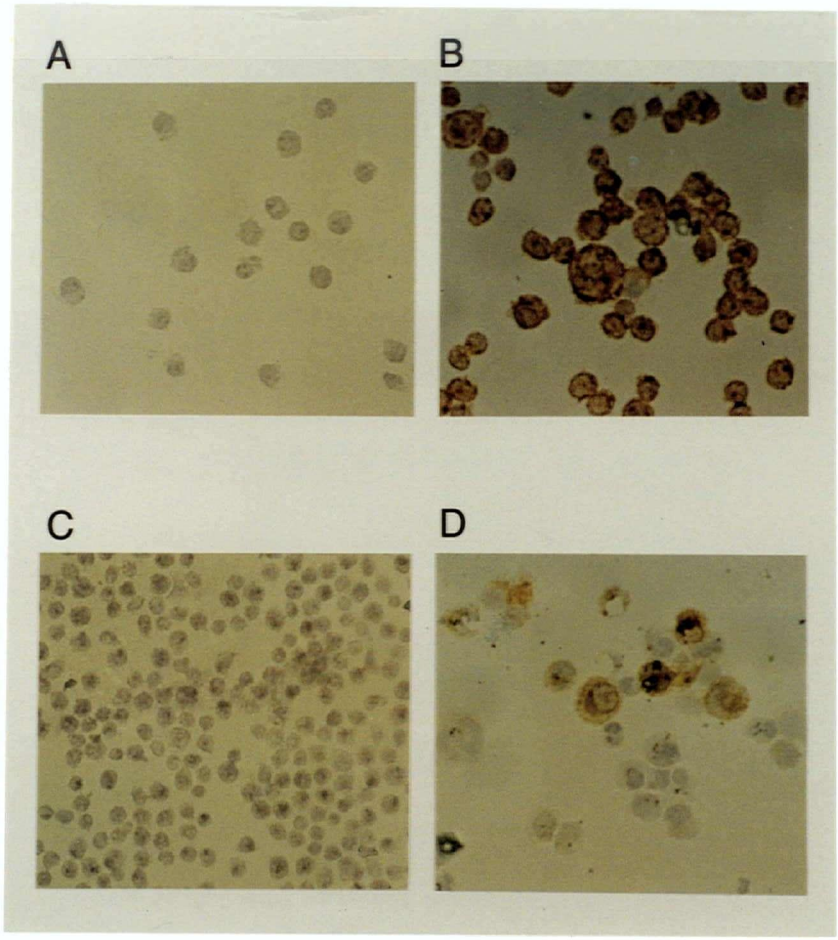


Figure 15: Enzymatic staining of SAGE vector transfected K562 cells. K562 cells transfected with (A) the pCEP4 vector alone, (B) v8-v9, (C) v9-v10, or (D) v8-v9-v10 SAGE vectors were cytospun onto glass slides, stained with enzymatic stain BCIP/INT and counterstained with hematoxylin as described in Materials and Methods section 14. ALP positive orange/brown BCIP/INT staining of the cytoplasm (B and D) makes these transfected cells appear larger than those whose nuclei only are stained by hematoxylin (A and C).

2.2 Assessment of SAGE splicing

The exact junction of the exon cassette and the leaderless ALP construct was confirmed by sequencing using the primers Rc/AP 5' and 3' as detailed in Materials and Methods, demonstrating that an unexpected stop codon or frameshift mutation had not been introduced into the construct. Since cell surface enzymatic staining detects the presence of active ALP and thus implies correct splicing of the transcripts from the SAGE exon cassette, a lack of ALP expression by the v9-v10 and v8-v9-v10 SAGE construct transfected K562 cells may indicate that incorrect splicing is occurring

2.3 RT PCR analysis of SAGE splicing

To investigate further the apparent lack of splicing in the v9-v10 and v8-v9-v10 SAGE transcripts, the more sensitive reverse transcriptase PCR (RT PCR) technique was employed. RNA was extracted from transfected K562 cells, subjected to oligo d(T)-primed first strand synthesis, and analyzed by PCR using primers nested in the common leader (R1-E5') and ALP (Rc/AP 3') regions of the SAGE constructs. The PCR products were then detected by blotting and autoradiography using a probe derived from the entire unique region of CD44R1 (Figure 11). The results are shown in Figure 16. An appropriately sized band of 295 bp corresponding to fully spliced v8-v9 was detected in the RT PCR of the RNA from the v8-v9 SAGE expressing K562 cells, while several PCR products resulting from multiple putative splice products were observed in the v9-v10 and v8-v9-v10 SAGE RT PCR (Figure 16B).

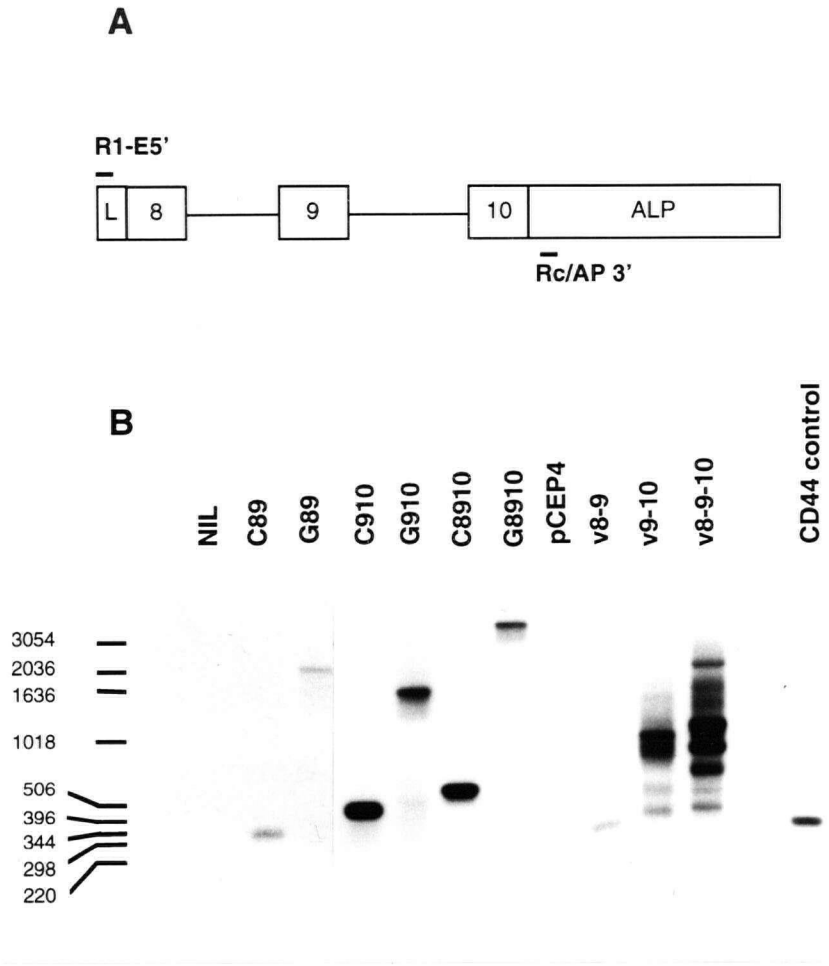


Figure 16: RT PCR of SAGE products. (A) Total RNA from K562 cells stably transfected with the SAGE vectors was collected and RT PCR was performed as described in Materials and Methods (Sections 5-7) using PCR primers R1-E5' and Rc/AP 3'. (B) The products, lanes pCEP4, v8-v9, v9-v10 and v8-v9-v10, were separated on a 2% gel and were transferred to Zetaprobe GT membrane and probed with the unique region probe (Figure 12). Controls for the PCR reaction in either a cDNA (C89, C910 and C8910), or genomic, intron containing form (G89, G910 or G8910) were amplified using the same primer set and included as size markers. NIL lane: no DNA added to PCR reaction. CD44 control: a *Hinf*I to *Pst*I fragment from CD44R1 clone 2.3.pCDM8 containing exons v8, v9 and v10.

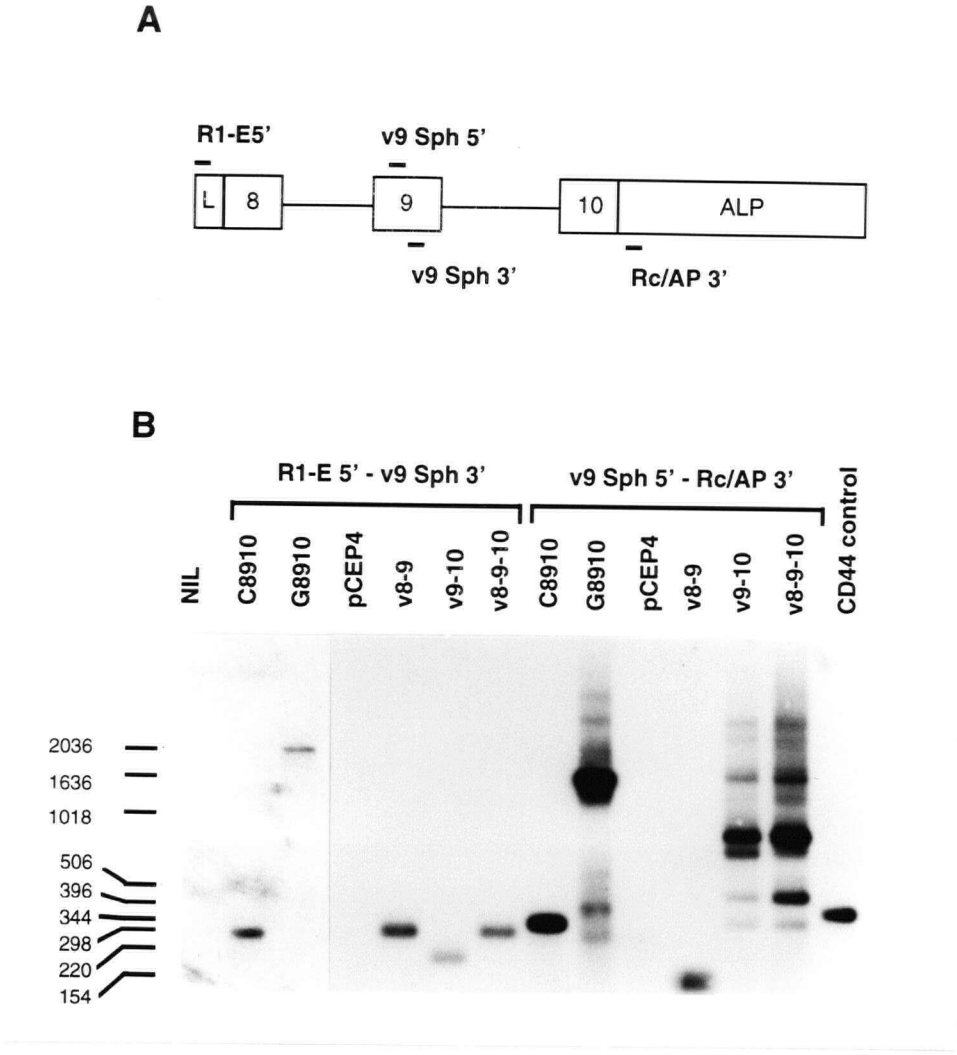


Figure 17: Exon specific RT PCR of SAGE products: (A) PCR amplification was performed on the first strand synthesis reactions described in Figure 16 using primers R1-E5' to v9 Sph 3', and v9 Sph 5' to Rc/AP 3' as described in Materials and Methods section 3. (B) The products were separated on a 2% agarose gel, transferred to Zetaprobe GT membrane, and probed with the unique region probe. C8910 and G8910 represent cDNA and genomic plasmid templates used as PCR and size marker controls. NIL and CD44 control lanes are described in Figure 16.

In order to precisely analyze the products formed by splicing of the SAGE transcripts, exon specific RT PCR spanning either exons v8 to v9, or v9 to v10 was performed on each RNA sample followed by blotting and autoradiography as before. Figure 17 demonstrates that the v8-v9 portion of the v8-v9-v10 SAGE construct was completely spliced (compare R1-E5' to v9 Sph 3' PCR of the cDNA control C8910 with v8-v9 and v8-v9-v10 lanes). However, RT PCR using the v9 Sph 5' to Rc/AP 3' primers spanning exons v9 to v10 resulted in the detection of at least 4 bands corresponding to completely spliced (292 bp) and completely unspliced (1487 bp) transcripts, and 2 other products of approximately 450 and 750 bp. Notably, products of similar sizes were found in the exon v9 to v10 specific RT PCR reactions of both the v9-v10 and v8-v9-v10 SAGE transcripts. These results indicate that production of pre-mRNA from the v9-v10 and v8-v9-v10 SAGE constructs is occurring, followed by similar splicing reactions. Thus the lack of cell surface ALP expression may be due to non-productive (i.e. non-protein encoding) pre-mRNA splicing.

2.4 Cloning and Characterization of SAGE products

To characterize the origin of the 292, 450 and 750 bp bands, the products from the exon v9 to v10 RT PCR of the v9-v10 and v8-v9-v10 SAGE transcripts discussed above were subcloned into the cloning vector, pCR (Materials and Methods section 2B and Figure 18). Several clones of each size were sequenced and results of representative clones are shown in Figure 19. An analysis and comparison of each of these products to the exon v9-v10 genomic sequence previously determined confirmed that the 292 bp product (19.8.pCR, Figure 19A and B) corresponded to completely spliced v9-v10 pre-mRNA. In other words this product arose from a pre-mRNA splice reaction utilizing the 5' splice site flanking exon v9 (v9-5' splice site) and the 3' splice site flanking exon v10 (v10-3' splice site) as previously described (7-9). Use of these splice sites in conjunction with other appropriate upstream and downstream sites leads to a functional mRNA encoding CD44R1 (39). In contrast, the 450

bp (4.3.pCR, Figure 19C and D) and 750 bp (22.8.pCR, Figure 19 E and F) products each corresponded to alternative splice products of the v9-v10 region. Comparison of the genomic sequence spanning the v9 and v10 exons with that of the 450 bp product revealed that this splice product was formed by the utilization of a downstream alternative 5' splice site found within intron 9 (referred to as the I9-5' splice site) while the use of the v10-3' splice site was maintained (Figure 19C). The 750 bp product also resulted from the use of the same alternative 5' splice site in conjunction with an alternative 3' splice site also located within intron 9 (I9-3' splice site). Protein translation of both the 450 bp and 750 bp products revealed the presence of a common stop codon (TGA) located 6 amino acids downstream of the v9-5' splice site (Figure 19B, D, and F). This stop codon presumably resulted in the premature termination of any protein product. Thus alternative splicing of the v9-v10 region of the SAGE construct transcripts is unlikely to produce a functional protein. These results provide an explanation for the paucity of ALP expression observed for the v9-v10 SAGE and v8-v9-v10 SAGE transfected K562 cells.

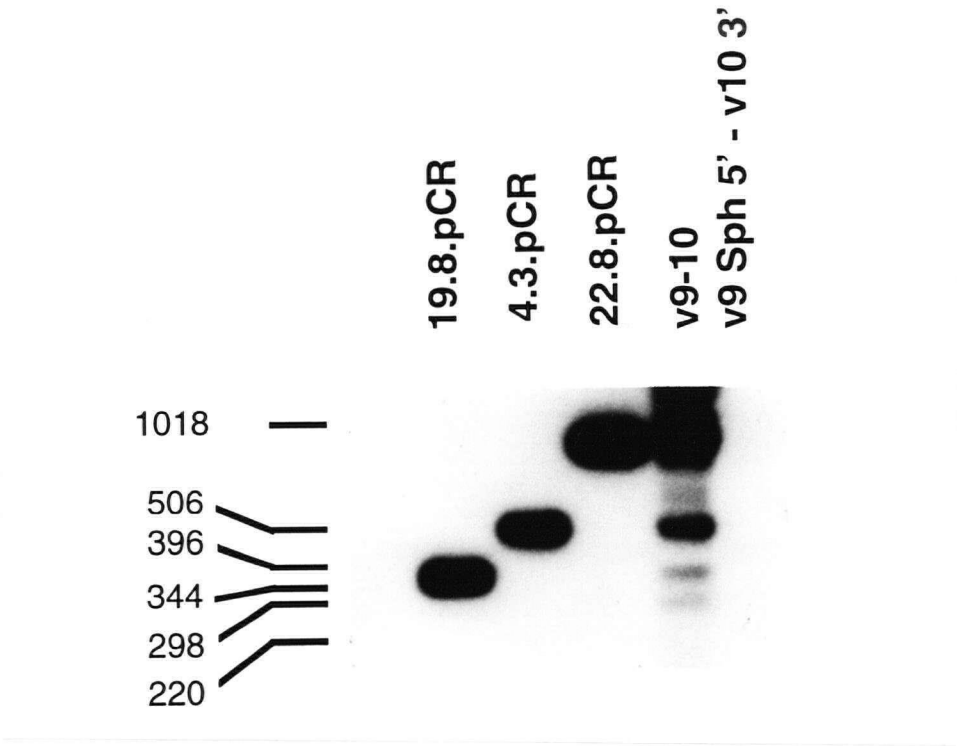
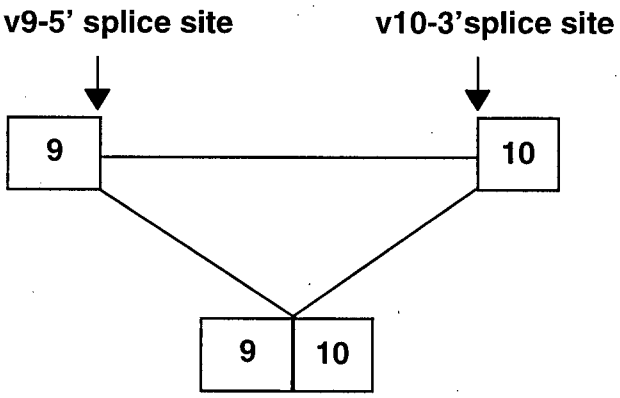


Figure 18: Aberrantly spliced PCR products were subcloned into the PCR cloning vector, pCR-Script SK+ as described in Materials and Methods. Representatives of each size were digested with Sma I and Sst I and were separated on a 2% agarose gel with v9-v10 PCR products (v9 Sph 5' to v10 3') for reference. The DNA was transferred to Zetaprobe GT membrane, probed with a v910 specific cDNA probe (Materials and Methods section 1.8D), washed and detected by autoradiography.

Figure 19: Sequence of aberrantly spliced PCR products. Splice products were cloned into the vector, pCR-Script SK+ and sequenced as described in Materials and Methods. Clone 19.8.pCR (A and B) is the result of a splice reaction utilizing the v9-5' and v10-3' splice sites, while clone 4.3.pCR (C and D) resulted from the use of an alternative 5' splice site located within intron 9 (I9-5' splice site), and the v10-3' splice site, as described in the text. Clone 22.8.pCR (E and F) resulted from use of the same I9-5' splice site and an alternative 3' splice site also found within intron 9 (I9-3' splice site). *Panels A, C and E:* cartoon representations of the splice reactions occurring between exons 9 and 10 leading to the observed products. *Panels B, D and F:* sequence of the described clones with protein translation below. Nucleotides encoding exons are shown in uppercase, and those encoding intron are in lowercase. Open boxes surround the partial sequence of the primers v9 Sph 5' and v10 3' used in the RT PCR reaction. Arrows note the position of the v9-5' and v10-3' splice sites flanking exons v9 and v10. The star indicates the site of fusion between the I9-5' and v10-3' splice sites in (D) or the I9-5' and I9-3' splice sites in (E).

A



B

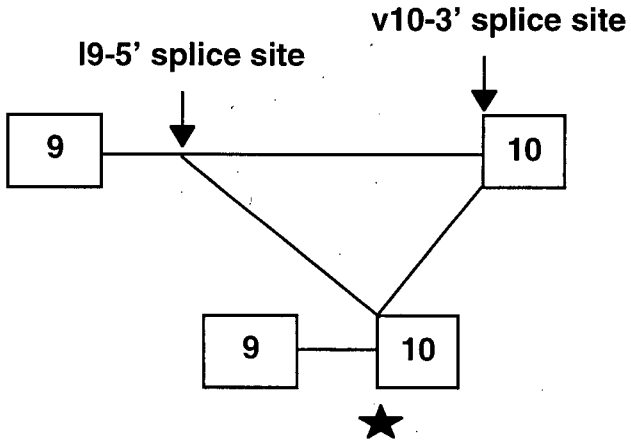
exon v9 exon v10

v9 Sph 5'

v10 3'

```
1/1                                                    31/11
GGC TTG GAA GAA GAT AAA GAC CAT CCA ACA ACT TCT ACT CTG ACA TCA AGC AAT AGG AAT
G   L   E   E   D   K   D   H   P   T   T   S   T   L   T   S   S   N   R   N
61/21                                                  91/31
GAT GTC ACA GGT GGA AGA AGA GAC CCA AAT CAT TCT GAA GGC TCA ACT ACT TTA CTG GAA
D   V   T   G   G   R   R   D   P   N   H   S   E   G   S   T   T   L   L   E
121/41                                                 151/51
GGT TAT ACC TCT CAT TAC CCA CAC ACG AAG GAA AGC AGG ACC TTC ATC CCA GTG ACC TCA
G   Y   T   S   H   Y   P   H   T   K   E   S   R   T   F   I   P   V   T   S
181/61                                                 211/71
GCT AAG ACT GGG TCC TTT GGA GTT ACT GCA GTT ACT GTT GGA GAT TCC AAC TCT AAT GT
A   K   T   G   S   F   G   V   T   A   V   T   V   G   D   S   N   S   N   V
241/81
AAT CG
N
```

C



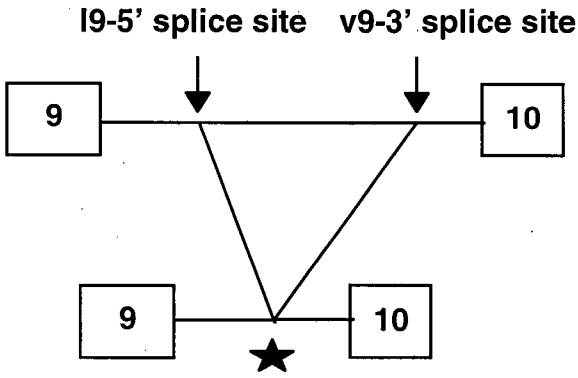
D

v9-5' splice site

v9 Sph 5'

```
1/1                                                    31/11
GGC TTG GAA GAA GAT AAA GAC CAT CCA ACA ACT TCT ACT CCG ACA TCA AGC Agt aag gat
G   L   E   E   D   K   D   H   P   T   T   S   T   P   T   S   S   S   K   D

61/21                                                  91/31
tat aaa acc tag ttg gct tca gct att gat aag aat caa tca att atg ggt act ttt gca
Y   K   T   *   L   A   S   A   I   D   K   N   Q   S   I   M   G   T   F   A

                                                       v10-3' splice site
121/41                                                 151/51
gtg gct ttg gtg gag gtg cat cca gga gct gcc gtt aca ctg tta ctt tta atc aaa agA
V   A   L   V   E   V   H   P   G   A   A   V   T   L   L   L   L   I   K   R

                                                       211/71
TAG GAA TGA TGT CAC AGG TGG AAG AAG AGA CCC AAA TCA TTC TGA AGG CTC AAC TAC TTT
*   E   *   C   H   R   W   K   K   R   P   K   S   F   *   R   L   N   Y   F

241/81                                                 271/91
ACT GGA AGG TTA TAC CTC TCA TTA CCC ACA CAC GAA GGA AAG CAG GAC CTT CCT CCC AGT
T   G   R   L   Y   L   S   L   P   T   H   E   G   K   Q   D   L   P   P   S

                                                                        v10 3'
301/101                                                331/111
GAC CTC AGC TAA GAC TGG ATC CTT TGG AGT TAC TGC AGT TAC TGT TGG AGA TTC CAA CTC
D   L   S   *   D   W   I   L   W   S   Y   C   S   Y   C   W   R   F   Q   L

361/121
TAA TGT CAA TCG
*   C   Q   S
```

E



F

v9-5' splice site

v9 Sph 5'

```
1/1                                                    31/11
GGC TTG GAA GAA GAT AAA GAC CAT CCA ACA ACT TCT ACT CTG ACA TCA AGC Agt aag gat
G   L   E   E   D   K   D   H   P   T   T   S   T   L   T   S   S   S   K   D

61/21                                                  91/31
tat aaa acc tag ttg gct tca gct att gat aag aat caa tca att atg ggt act ttt gca
Y   K   T   *   L   A   S   A   I   D   K   N   Q   S   I   M   G   T   F   A

121/41                                                 151/51
gtg tct ttg gtg gag gtg cat cca tta gct gcc gtt aca ctg tta ctt tta atc aaa agg
V   S   L   V   E   V   H   P   L   A   A   V   T   L   L   L   L   I   K   R

181/61                                                 211/71
cag caa aat ttt atc ttt caa tcc aat gat atg ttt cct cca cta cat caa aga tga atc
Q   Q   N   F   I   F   Q   S   N   D   M   F   P   P   L   H   Q   R   *   I

241/81                                                 271/91
caa ttg aga aag aac tca aga tta tct agt acc tca gct gac ttc ttg atc aca gtt gta
Q   L   R   K   N   S   R   L   S   S   T   S   A   D   F   L   I   T   V   V
```

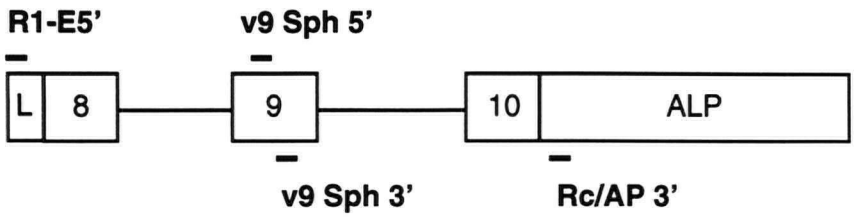
```
301/101                                                 331/111
gtg aca tag aat tac cta ata acc caa gca gat gtt gca aca aat gcc aat tct ggg tgt
V   T   *   N   Y   L   I   T   Q   A   D   V   A   T   N   A   N   S   G   C

361/121                                                 391/131
taa gat cac caa aat ctc ttc tgt tga gtg atg tcc aag tag tgt cta ttg caa ata ttt
*   D   H   Q   N   L   F   C   *   V   M   S   K   *   C   L   L   Q   I   F

421/141                                                 451/151
att ctg ttc aat agt gtg aca aac ttc act cta ttt ctc cct att tat gga taa taa gtg
I   L   F   N   S   V   T   N   F   T   L   F   L   P   I   Y   G   *   *   V

481/161                                                 511/171
gaa agc ttt tgc agc aat tgt gtg ttc tgg aga caa gca cat ggt ggg tgg tga tct tac
E   S   F   C   S   N   C   V   F   W   R   Q   A   H   G   G   W   *   S   Y

                                                        v10-3' splice site
541/181                                                 571/191
aaa tac ggg ttc atc act gat tcc acc tcc aca cag ATA GGA ATG ATG TCA CAG GTG GAA
K   Y   G   F   I   T   D   S   T   S   T   Q   I   G   M   M   S   Q   V   E

601/201                                                 631/211
GAA GAG ACC CAA ATC ATT CTG AAG GCT CAA CTA CTT TAC TGG AAG GTT ATA CCT CTC ATT
E   E   T   Q   I   I   L   K   A   Q   L   L   Y   W   K   V   I   P   L   I

661/221                                                 691/231
ACC CAC ACA CGA AGG AAA GCA GGA CCT TCA TCC CAG TGA CCT CAG CTA AGA CTG GGT CCT
T   H   T   R   R   K   A   G   P   S   S   Q   *   P   Q   L   R   L   G   P

                                                                        v10 3'
721/241                                                 751/251
TTG GAG TTA CTG CAG TTA CTG TTG GAG ATT ACT GCA GTT ACT GTT GGA GAT TCC AAC TCT
L   E   L   L   Q   L   L   L   E   I   T   A   V   T   V   G   D   S   N   S

781/261
AAT GTC AAT CG
N   V   N
```

3. PRE-SPLICING OF THE V8-V9 EXONS FACILITATES V8-V9-V10 SPLICING

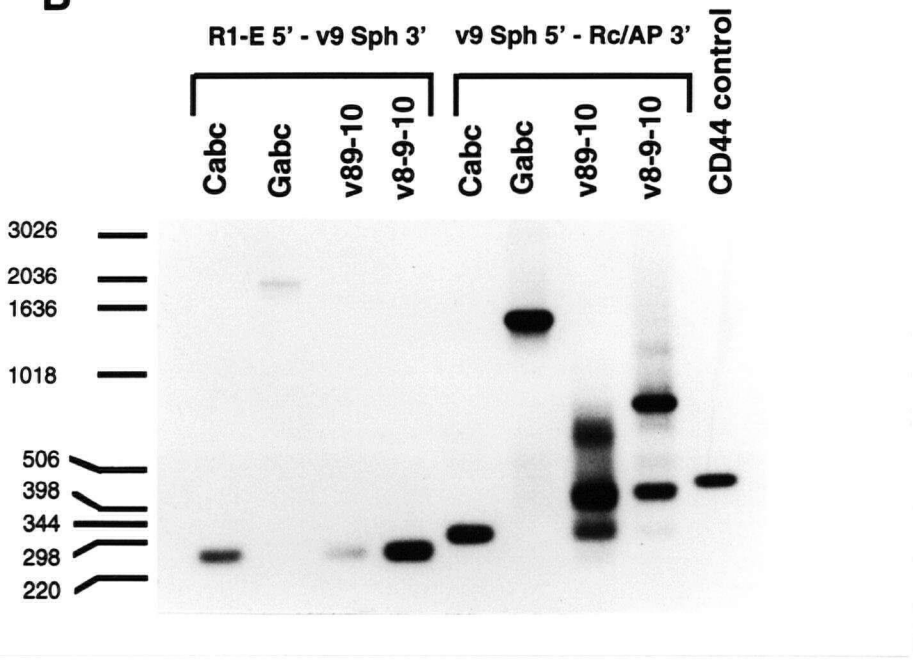
Previous work by Helfman *et al.* has demonstrated that ordered splicing occurs within the rat β–tropomyosin *in vitro* splicing system (154). Namely, prior splicing of alternative upstream exons 6 or 7 to a common downstream exon 8 is necessary for further splicing of common upstream exon 5 to downstream exons 6 or 7. The observations that the v8-v9 SAGE transcript is readily spliced in K562 cells while v9-v10 SAGE transcript splicing results in multiple non-protein encoding splice products, and that within the v8-v9-v10 SAGE transcript the v8-v9 portion is always completely spliced were suggestive that prior splicing of exon v8 to exon v9 may exert an influence on the fidelity of exon v9 to exon v10 splicing. This question was addressed by the construction of a new SAGE vector which consisted of exons v8 and v9 in a spliced conformation with v9-v10 unspliced (construct is denoted as v89-10; Figure 11). The v89-10 SAGE construct vector was introduced into K562 cells as previously described, and the cell surface expression of ALP was assessed. Histological assay of ALP expression showed that the percentage of v89-10 SAGE transfected K562 cells expressing active ALP approached that of the v8-v9-v10 SAGE K562 cells. Furthermore, analysis by RT PCR performed as previously described, coupled with densitometry of the autoradiograph confirmed that an increase in the percentage of v9-5' to v10-3' spliced 292 bp product had occurred, from 8.5% for the v8-v9-v10 SAGE construct transfected K562 cells, to 22.4% for v89-10 SAGE construct transfected K562 cells (Figure 20). These results indicate that prior splicing of the v8-v9 portion of the v8-v9-v10 exon cassette exerted a positive influence upon the v9-v10 portion and increased the proportion of protein-encoding mRNA.

Figure 20: Exon specific PCR of v89-10 SAGE transcripts. (A) Production of splice products in K562 cells stably transfected with v89-10 or v8-v9-v10 SAGE vectors was compared by RT PCR of total RNA using the R1-E 5' to v9 Sph 3' primers flanking exons v8 and v9, and v9 Sph 5' to Rc/AP 3' primers flanking exons v9 and v10. (B) The PCR products were separated on a 2% agarose gel and were probed with the unique region probe (Figure 12). (C) Densitometry measurements of the autoradiograph were made and the percentage of the spliced forms as compared to total density of each lane was calculated (in brackets).Cabc (C8910) and Gabc (G8910) represent cDNA and genomic plasmid templates used as PCR and size marker controls as described in Figure 16. Representative results of a typical experiment are shown.

A



B



C

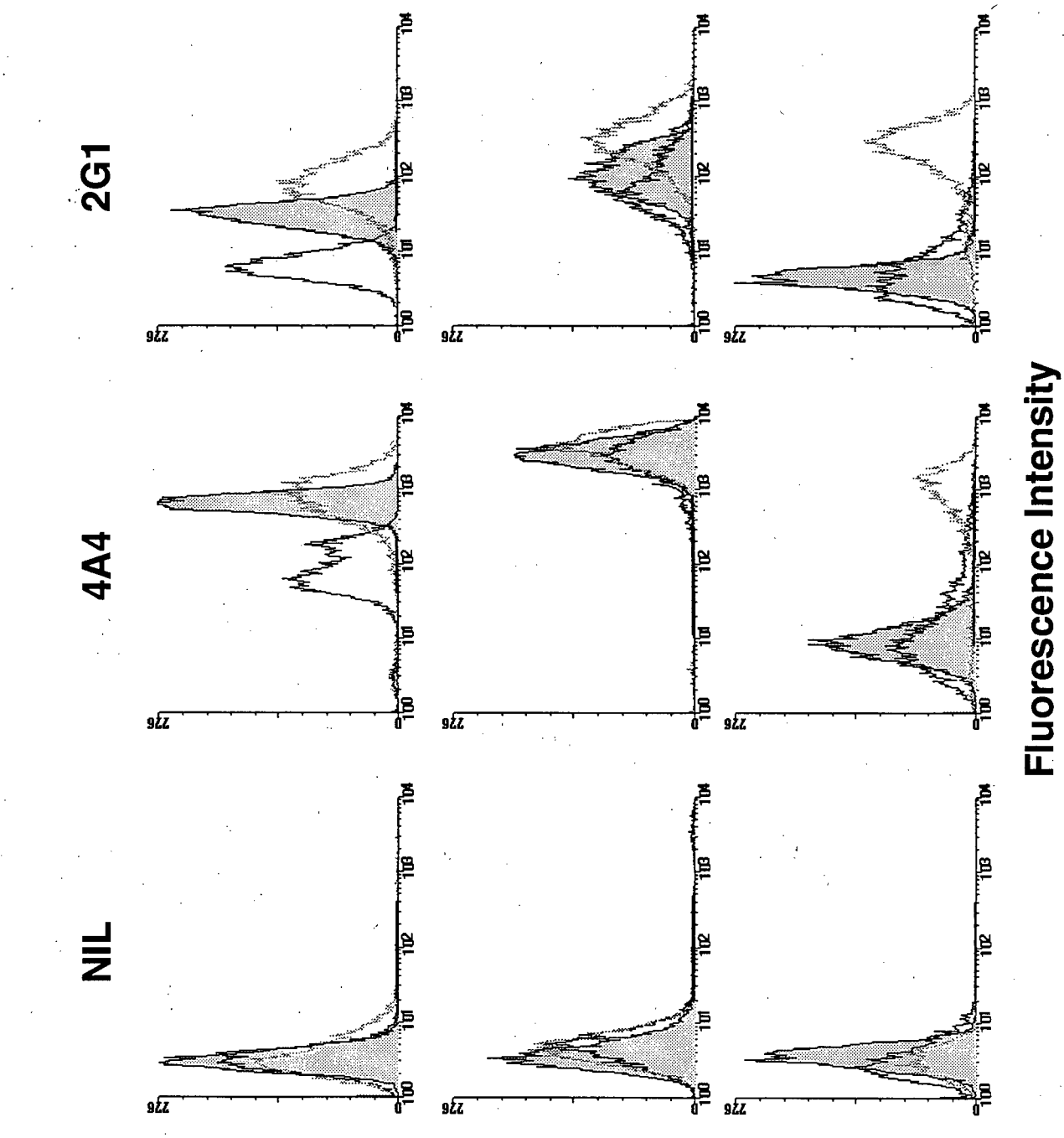
	v89-10 (%)	v8-v9-v10 (%)
750 bp	N/A	2110.93 (54.2)
450 bp	2762.08 (44.8)	1454.16 (37.3)
292 bp	1384.41 (22.4)	331.45 (8.5)

4. SPLICING OF THE CD44R1 UNIQUE REGION EXONS IS DEPENDENT UPON THE PHOSPHORYLATION STATUS OF THE CELL.

4.1 FACS analysis of CD44 isoform expression

Previous studies have shown that CD44 isoform expression is often dependent upon the activation status of a cell. For example, activation of resting T cells by anti-TcR antibodies plus anti-CD3 antibodies or by treatment with phorbol esters plus calcium ionophore has been demonstrated to cause an upregulation of CD44 expression as well as a transient induction of CD44 isoform expression (14-16). Furthermore, the treatment of K562 cells with phorbol ester has also been demonstrated to induce megakaryocyte differentiation and cell surface expression of CD44 isoforms (245, 246). To investigate this induction of CD44 isoform splicing, the myeloid cell lines K562, KG1, and its less differentiated subtype KG1a were treated with the phorbol ester TPA, a potent inducer of protein kinase C (PKC), and a serine/threonine kinase inhibitor, staurosporine. In initial experiments, the expression of CD44 isoforms was assayed by fluorescence activated cell sorting (FACS) using the monoclonal antibodies (mAb) 4A4, which recognizes a common CD44 epitope, and mAb 2G1 which recognizes an epitope encoded by exon v10. The results of a typical assay are shown in Figure 21. While TPA treatment induced a mild increase in overall CD44 as well as exon v10 expression in KG1 and KG1a cells, a 2.5 to 3 log shift in CD44 and v10 expression was noted in K562 cells. Interestingly, treatment with staurosporine had contrasting results in each of the three cell lines tested. CD44 and v10 epitope expression was decreased in KG1 cells by this treatment, but was unaffected in KG1a cells. In contrast, overall expression of CD44 and exon v10 was increased in a sub-population of K562 cells.

Figure 21: FACS of CD44 isoform expression in various cell lines. KG1, KG1a and K562 cells were incubated with nil (shaded histogram), 5 μg/ml TPA (grey overlay) or 10 ng/ml staurosporine (black overlay) for 72 hours. The cells were then harvested and treated for FACS using a mAb which recognizes either a common CD44 epitope (4A4), a v10 epitope (2G1) or media only control (NIL) as described in Materials and Methods.


NIL
4A4
2G1
KG1
KG1a
K562
Fluorescence Intensity

4.2 Identification of CD44 isoforms

FACS analysis facilitates a determination of the gross changes in overall CD44 and v10 epitope expression. However the actual isoforms which are expressed by each of the cell lines cannot be characterized by this method. Thus to analyze the different isoforms which are induced or repressed by treatment with TPA and staurosporine, RT PCR analysis was performed. RNA was collected from each treated and untreated cell line and subjected to first strand synthesis followed by PCR using primers which are nested in the leader sequence (R1-E5') and flank the extracellular domain (R1-E3'). The products were blotted and detected by autoradiography using the CD44R1 unique region probe. As shown in Figure 22, treatment of K562 cells with TPA induced at least 8 different isoforms in addition to CD44H (not detected by this probe). In contrast, treatment of the same cells with staurosporine induced only 3 isoforms corresponding to CD44R1 (v8-v9-v10), CD44R2 (alternative exon v10 only) and an intermediate sized isoform of 1073 bp as determined by size comparison to a standard curve. It is interesting to note that treatment with TPA or staurosporine appeared to induce differential expression of two CD44 isoforms of either 995 bp or 1073 bp in KG1 cells while the overall number of isoforms expressed was unaffected by treatment in KG1a. Taken together, these results indicate that either an increase or decrease in serine/threonine phosphorylation within these cells leads to an alteration the CD44 isoforms expressed on the cell surface.

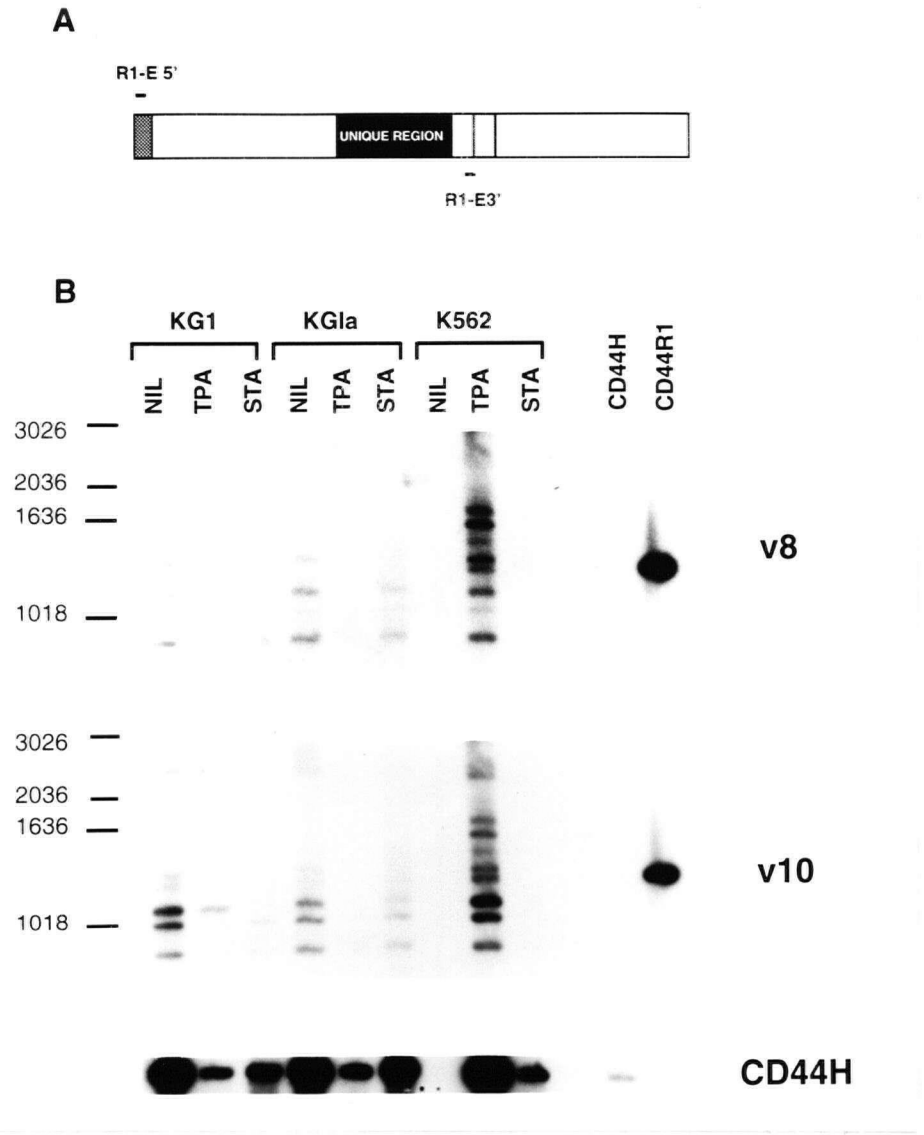


Figure 22: RT PCR of cell lines induced with TPA or Staurosporine. KG1, KG1a or K562 cells were treated for 72 hours with either nil, 5 mg/ml TPA, or 10 ng/ml staurosporine (STA). (A) Total RNA was collected and analyzed for changes in CD44 isoform encoding mRNA by RT PCR using primers R1-E5' to R1-E3' which amplify the extracellular region of CD44 . (B) The PCR products were separated on a 3% agarose gel and probed with a v8, v10, or CD44 common region probe.

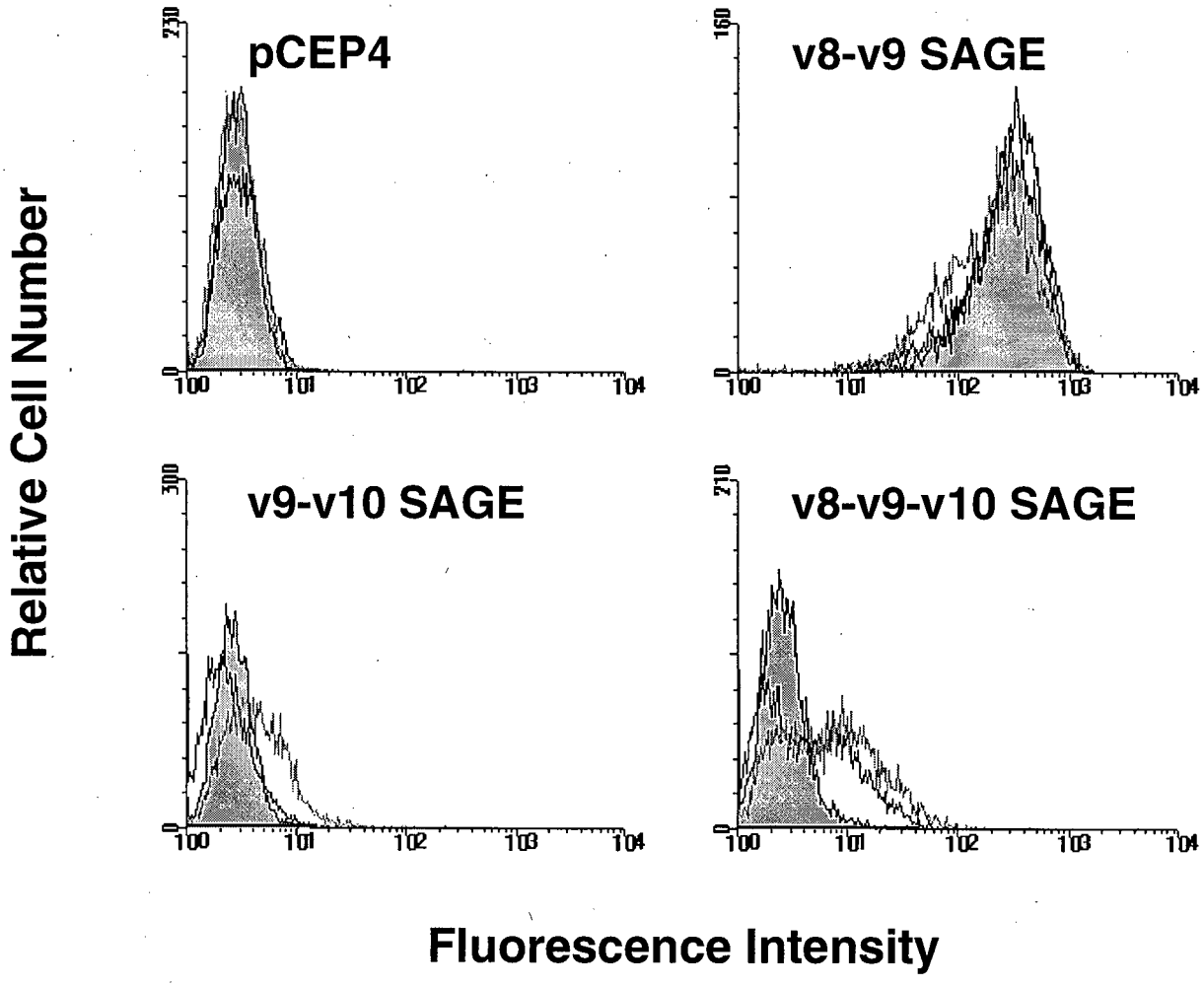
5. INDUCTION OF SAGE SPLICING.

5.1 Analysis of induced SAGE splicing

TPA and staurosporine treatment have been demonstrated to alter the expression pattern of CD44 isoforms on K562 cells. To determine if these inducers have the same affect on transcripts produced from the SAGE vector constructs, K562 cells expressing each of the SAGE vectors were treated with TPA or staurosporine. In initial experiments, splicing was determined by FACS analysis using MAb A3A11 as described in Methods and Materials sections 16 and 18. As demonstrated by Figure 23, splicing of the SAGE transcripts appeared to increase after incubation with either TPA or staurosporine. However, since TPA has been previously shown to stimulate transcription from the CMV immediate-early promoter in some cell lines (247), and this promoter is used to drive production of the SAGE construct RNA, the observed increase in ALP expression may be the result of upregulated transcription from the plasmid vector rather than an induction of splicing. To differentiate between these two possibilities, transfected K562 cells were incubated with nil, 5 μg/ml PdBU, or 10 ng/ml staurosporine and total RNA was analyzed by Northern blot. As shown in Figure 24, treatment of the v8-v9 SAGE expressing cells with PdBU increased the overall levels of spliced RNA (2.5 kb), while a decrease in the unspliced form (4.4 kb) is noted. After staurosporine treatment, only RNA corresponding to an appropriate size for the spliced form is observed. In a similar fashion, PdBU treatment increased the RNA levels of the unspliced (6 kb), aberrantly spliced (4 kb) and protein-encoding spliced (3 kb) v8-v9-v10 SAGE transcripts. Conversely, staurosporine treatment appears to have selectively reduced the amount of v9-5' to v10-3' unspliced and aberrantly spliced v8-v9-v10 SAGE transcripts leaving only the protein encoding spliced RNA. Figure 24 also demonstrates a lack of v9-v10 SAGE RNA in both the stimulated and unstimulated lanes, suggesting that the v9-v10 SAGE transcripts may be unstable. The result of this analysis indicates that PdBU treatment may increase the splicing of the SAGE construct pre-mRNA through a mechanism of

upregulated transcription leading to increased levels of pre-mRNA and thus more spliced product. Staurosporine treatment, however, appears to have shifted SAGE transcript splicing so that the level of unspliced and aberrantly spliced RNA is reduced, leaving the protein encoding splice product as the only form observed without an increase in the amount of pre-mRNA.

Figure 23: FACS analysis of TPA or staurosporine induced SAGE splicing. K562 cells stably transfected with the SAGE vectors were incubated with nil (filled histogram), 5 μg/ml TPA (grey overlay) or 10 ng/ml staurosporine (black overlay) for 72 hours. Induction of the ALP epitope recognized by MAb A3A11 was determined by FACS analysis as described in Materials and Methods section 16.


pCEP4
v8-v9 SAGE
v9-v10 SAGE
v8-v9-v10 SAGE
Relative Cell Number
Fluorescence Intensity

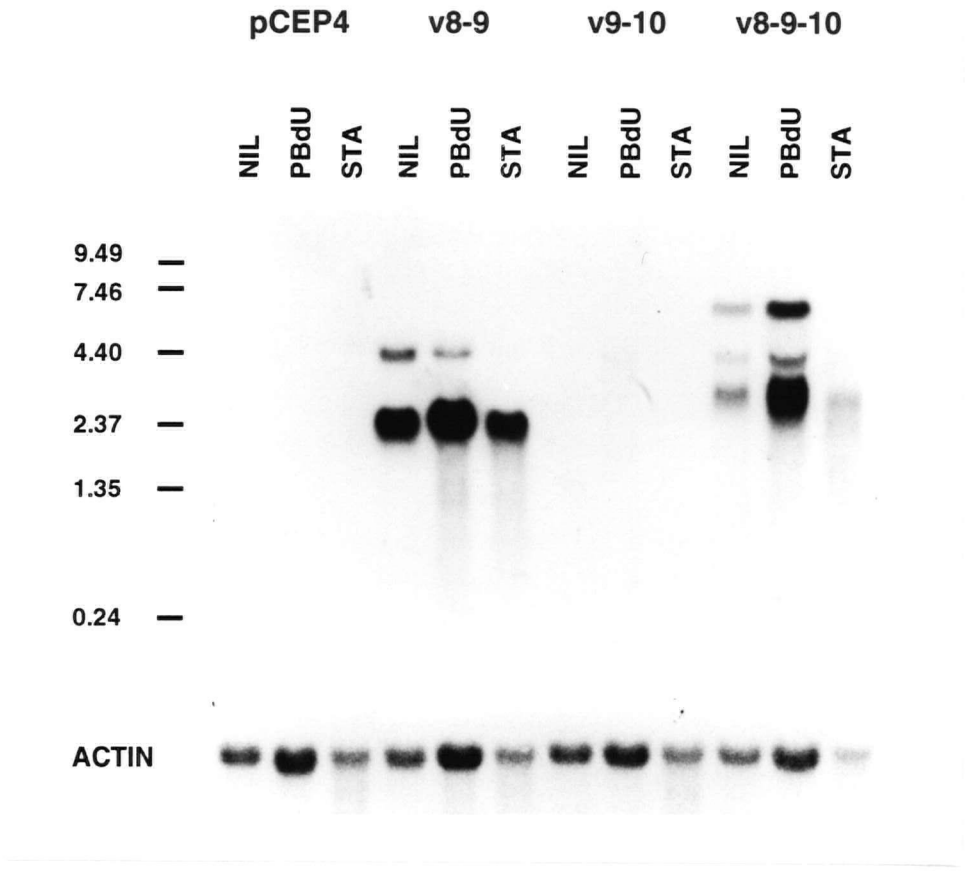


Figure 24: Northern analysis of SAGE RNA. K562 cells stably transfected with the v8-9, v9-10 and v8-9-10 SAGE vectors or pCEP4 control were incubated with nil, 10 μg/ml phorbol di-butyrate (PdBU) or 10 ng/ml staurosporine (STA) for 72 hours. RNA was collected and Northern blot analysis performed using the unique region probe as described in Materials and Methods. The blot was then stripped and re-probed with an actin probe.

5.2 RT PCR analysis of induced SAGE splicing

The induction of splicing observed in the previous experiment were confirmed by RT PCR of staurosporine treated SAGE vector transfected K562 cells using exon specific primers v9 Sph 5' and Rc/AP 3' (Section 3.3) as shown in Figure 25. Densitometric analysis of Figure 25 demonstrated that the percentage of the 292 bp band resulting from use of the v9-5' and v10-3' splice sites flanking exons v9 and v10 as previously described, increased twofold in the untreated v9-v10 SAGE transfected cells as compared to staurosporine treated controls, from 6% to 13% of total PCR products (Figure 25C). The same band in the v8-v9-v10 SAGE transcript RT PCR reaction increased from 7% (untreated) to 26.8% (staurosporine treated) of total PCR products, an increase of almost fourfold. Interestingly, similar treatment and comparison of the v9-3' to v10-5' spliced 292 bp product from the partially spliced v89-10 SAGE transcript versus the unspliced v8-v9-v10 SAGE transcript RT PCR (Figure 26) does not reveal any increase in the 292 bp product. These data suggest that the v89-10 SAGE transcript may be splicing at a maximal level thus splicing of v9 to v10 is not further induced by staurosporine.

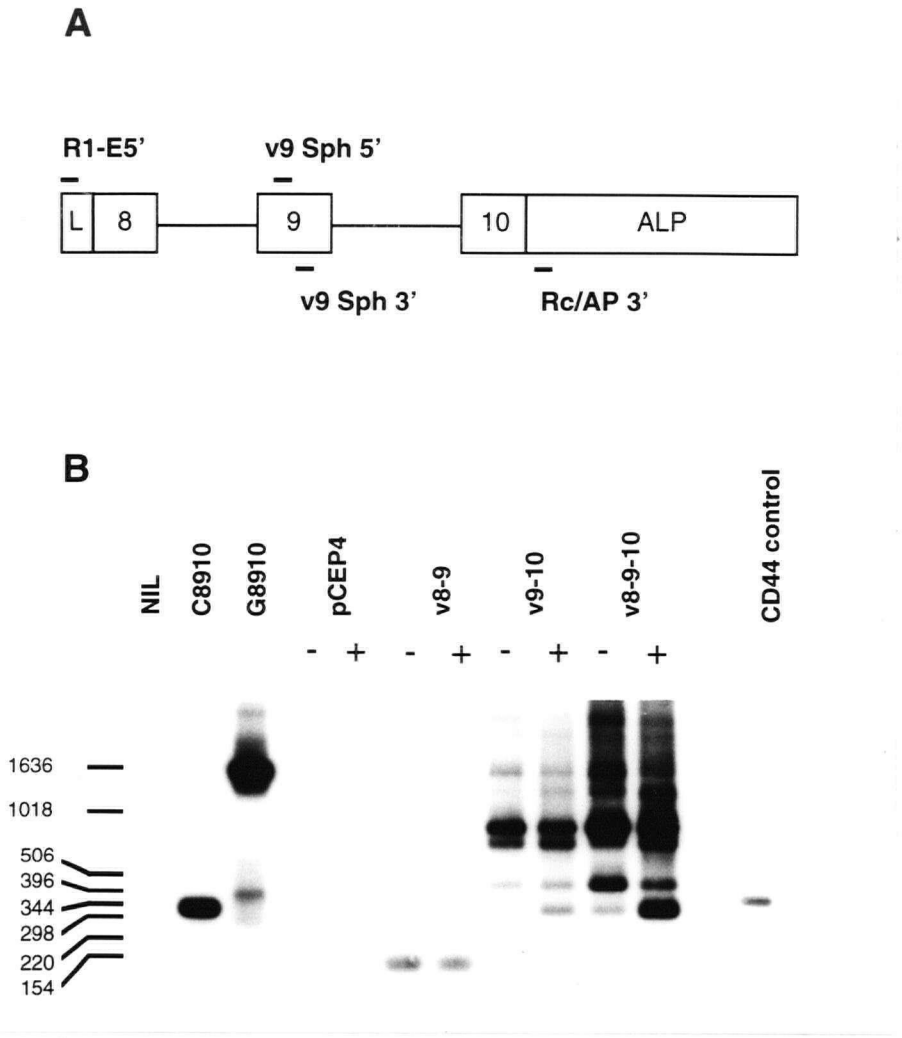


Figure 25: RT PCR of induced SAGE splicing. K562 cells stably expressing the SAGE constructs were incubated with nil (-) or 10 ng/ml staurosporine (+) for 72 hours. (A) RNA was collected and RT PCR performed using exon specific primers v9 Sph 5' to Rc/AP 3', (B) followed by blotting and detection with the CD44R1 unique region probe as described in Materials and Methods. Lanes are marked as described in Figure 17 (C) Densitometric analysis of the autoradiograph was performed and the density of each band (shown in brackets) as a percentage of the total density of each lane was calculated. Representative results of a typical experiment are shown.

C

	v9-10 (%)		v8-v9-10 (%)	
	NIL	STA	NIL	STA
1636 bp	698.5 (13.1)	2110.93 (11.5)	2227.5 (18.9)	1548.7 (11.3)
750 bp	2574.3 (46.4)	1454.16 (34.0)	6471.9 (54.8)	4603.7 (33.5)
600 bp	1139.5 (21.3)	331.45 (27.3)	N/A	2261.6 (16.4)
450bp	601.1 (11.3)	331.45 (14.0)	2282.1 (19.3)	1656.7 (12.0)
292bp	321.1 (6.0)	331.45 (13.0)	832.5 (7.0)	3685.3 (26.8)

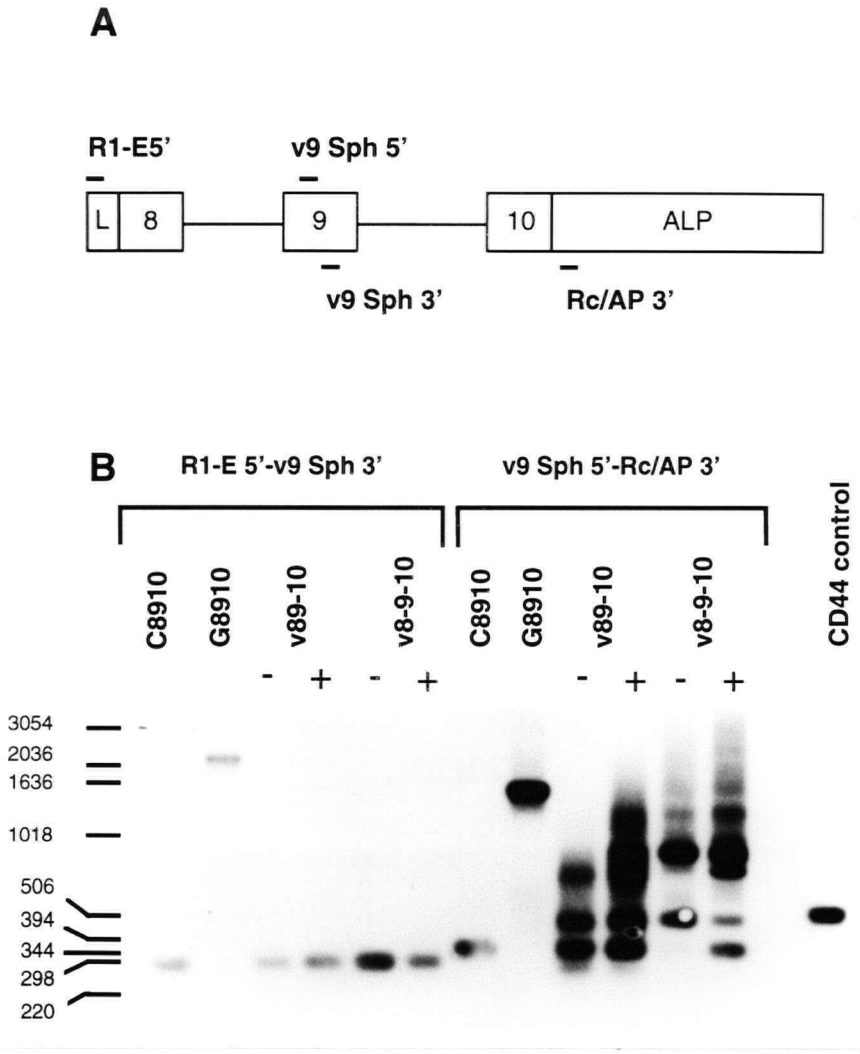


Figure 26: Comparison of staurosporine-induced v89-10 SAGE transcripts to v8-v9-v10 SAGE transcripts. K562 cells stably transfected with either v89-10 SAGE or v8-v9-v10 SAGE constructs were incubated with nil (-) or 10 ng/ml staurosporine (+) for 72 hrs. (A) RNA was collected and RT PCR was performed using exons specific PCR primers R1E 5' to v9 Sph 3' and v9 Sph 5' to Rc/AP 3' (B) followed by separation on a 2% agarose gel, blotting and autoradiography as previously described for Figure 17. Lanes are marked as described in Figure 17.

6. RETENTION OF INTRON 9 IN NON-TRANSFECTED CELL LINES

Aberrant splice products resulting from the use of alternative intronic 5' and 3' splice sites have been demonstrated to accumulate in K562 cells transfected with the SAGE constructs. However, this finding may be the result of overexpression of RNA in an artificial splicing system. To determine if the same aberrant splicing between v9 and v10 leading to retention of intron 9 sequences can be demonstrated to occur in various cell lines, RT PCR was undertaken. A new internal intron 9 5' primer nested upstream of the alternative I9-5' splice site was constructed and PCR from this primer to the v10 3' primer was performed followed by blotting and autoradiography using the v10 specific probe (Figure 27). KG1 cells demonstrated no detectable retention of intronic sequences while several bands corresponding to the 450 bp splice product (found at 250 bp using this primer set), and unspliced pre-mRNA (1400 bp) as well as two others at 300 bp and 900 bp were observed after TPA and staurosporine treatment. Interestingly, TPA treatment of K562 cells, originally showing no aberrant splice products or CD44 RNA as expected, demonstrated aberrantly spliced intron-retaining bands corresponding to the 450 bp (250 bp), 750 bp (also at 750 bp using this primer set) and unspliced pre-mRNA previously described as well as one other product at 300 bp. Staurosporine treatment, which had been previously been determined to increase the percentage of protein encoding splice product, abolished the intermediate sized intron retaining bands and only the 1400 bp band corresponding to unspliced pre-mRNA was observed. Taken together this data indicates that aberrant splicing and intron retention occurs in both the SAGE constructs and in non-transfected cell lines.

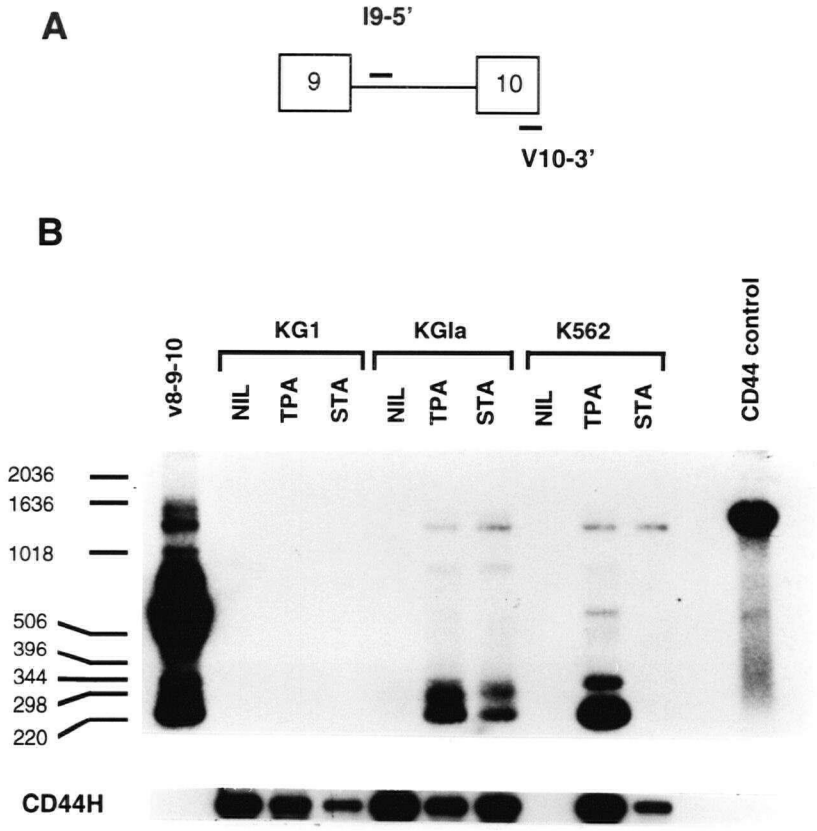


Figure 27: Intron retention in induced cell lines. KG1, KG1a and K562 cells were incubated with nil, 5 µg/ml TPA or 10 ng/ml staurosporine (STA) for 48 hours. (A) RNA was collected and RT PCR performed using I9-5' and v10-3' primers. (B) DNA was transferred to Zetaprobe GT membrane and probed with the exon v10 probe as described in Materials and Methods. A control PCR of CD44H using R1-E5' to R1-E3' primers was also performed and blotted with a common region CD44 probe.

IV DISCUSSION

1. THE SAGE ASSAY SYSTEM

The alternative splicing of pre-mRNA has emerged as a commonly used cellular strategy to generate diversity at the protein level. In general, most alternative splicing results in a mutually exclusive pattern of exon usage which is either developmentally regulated or tissue specific (248). For example, the rat β-tropomyosin pre-mRNA is alternatively spliced to express either exons 5-6-8 in fibroblasts, or 5-7-8 in skeletal muscle (154). In this case, the expression of exons 6 and 7 is mutually exclusive as either one or the other is expressed in the specific tissue. In contrast, CD44 isoform expression while often following a developmentally regulated pattern (249), is not mutually exclusive. Multiple protein isoforms of CD44 are generated by the alternative splicing of a set of ten contiguous exons into a single site in the mRNA (7). Using FACS and Western blot analysis, several of these isoforms have been demonstrated to be concurrently expressed upon the cell surface (39). Furthermore, multiple protein isoforms can be upregulated *de novo* by cellular activation through various signaling pathways such as triggering through the T cell receptor/CD3 complex, (250) or by incubation of cells with phorbol esters (251). Clearly the expression of CD44 isoforms, and therefore the alternative splicing of CD44 pre-mRNA, results from a complex regulatory system which is dependent upon signals which both activate and influence alternative splicing.

To investigate the molecular mechanisms which regulate the splicing of the CD44R1 unique region exons v8, v9, and v10, a novel *in vivo* splice assay system was developed. The basis of this system, termed splice activated gene expression (SAGE), is the juxtaposition of two or three exons and introns (the exon cassette) from the CD44 unique region with a leaderless

construct of human liver/bone/kidney alkaline phosphatase (ALP). ALP is normally expressed as a glycophosphinositol-(GPI-) linked tetramer on the cell surface (252, 253). The protein structure of ALP, as determined by x-ray crystallography, indicates that the tertiary structure of the molecule is triangular, with the amino and carboxy-termini pointing away from the active site (254, 255). Thus the addition of extra protein sequence to either terminus does not affect enzymatic function (255). In fact, previous experimentation has demonstrated that the entire extracellular domain of CD44R1 can be added to the amino terminal of ALP without any loss of enzymatic function (48). When a suitable leader sequence, in this case the CD44 leader, is placed upstream of the SAGE construct, correct splicing of the exon cassette is indicated by the cell surface expression of ALP. This may be tested by either enzymatic assay or by FACS using an anti-ALP antibody, mAb A3A11. In addition, sensitive detection of splice products can be accomplished using the technique of RT PCR with primers complimentary to the 5' end of the ALP gene and the heterologous CD44 leader sequence.

The *in vivo* method of splicing analysis described in this work has several advantages over the conventional *in vitro* splice assay. The basic *in vitro* splice assay consists of a radioactively labeled pre-mRNA molecule transcribed by RNA Polymerase II (142). A nuclear extract, traditionally made from HeLa cells is added and the splice reaction is incubated for one hour at 37 ^{o}C (105). The products are then detected by electrophoresis through a 6% polyacrylamide gel followed by autoradiography. This method has primarily been utilized to define the exact protein and RNA factors necessary to form the basic spliceosomal complex and for splicing to occur. However, the results obtained using the *in vitro* assay often do not reflect actual *in vivo* events with several cryptic splice reactions often detected which do not normally occur (139, 109). As an *in vivo* system, the SAGE assay only detects those splice reactions which actually occur within the cell of interest. Furthermore, use of the SAGE assay system allows for the investigation of upstream events such as cell

surface receptor crosslinking and other activation signals which lead to upregulation of splicing.

On a molecular level, the *in vivo* production of mRNA for the SAGE assay also incorporates two important mRNA modifications, namely a 5' methylated guanosine cap, and a poly (A) tail. Both of these sequences have recently been demonstrated to play a crucial role in the determination of terminal exons and have in fact been shown to interact with several splice factors and be necessary for appropriate splicing to occur (141, 142, 150). Inclusion of these pre-mRNA modifications insures that the spliceosomal machinery 'sees' the pre-mRNA transcribed from the SAGE construct as an authentic pre-mRNA (144) as opposed to the *in vitro* splice assay in which the terminal exons of the pre-mRNA are un-modified and thus are not processed correctly (139).

When introduced into K562 cells, the SAGE transcripts demonstrated a distinctive splicing pattern paralleling SAGE splice products detected in various transiently transfected hemopoietic and non-hemopoietic cell lines. Since each of these cell lines express variable levels of CD44 isoforms, these observations suggest that the splicing of the SAGE transcripts does not directly correlate with CD44 isoform expression.. When removed from the context of upstream regulatory sequences, splicing of the v8-v9 transcripts occured in a constitutive-like fashion in the v8-v9 and v8-v9-v10 SAGE transcripts, while splicing of the v8-v9-v10 transcript appeared to be both alternative and inducible (Figure 25). In contrast, cells transfected with the v9-v10 construct demonstrated no histologically discernable ALP expression (Figure 15C and Figure 23), correlating with no observed mRNA production as detected by Northern blot (Figure 24) and very low levels of protein-encoding spliced product detected by the presence of the 292 bp using RT PCR amplification (for example see Figure 25, uninduced (-) v9-v10 lane). The exact mechanisms leading to the instability of transcripts from the v9-v10 SAGE vector was not investigated in this study. However, recent

work by several groups (281-283) has suggested that a pre-mRNA/translational scanning mechanism may exist, preventing the splicing of pre-mRNAs containing pre-mature stop codons encoded within exons and signalling their degredation. It is conceivable that the in-frame stop codon formed when the intron encoded I9-5' splice site is utilized, located in close proximity to the translational start site, may be recognized by such a mechanism leading to an inhibition of v9-v10 SAGE transcript splicing and degradation of v9-v10 SAGE pre-mRNA. The ability of the v8-v9 portion of the v8-v9-v10 SAGE construct to undergo a splicing reaction may stabilize the remaining v9-v10 portion (108). Alternatively, the presence of sequences encoded by exon v8 may act as a 'spacer' between the start site and the stop codon and ameliorate the scanning mechanism, or may encode a mRNA stabilization motif and thus prevent the degredation of splice products from the v8-v9-v10 SAGE pre-mRNA.

2. CIS-ACTING SEQUENCES AFFECTING SAGE CONSTRUCT SPLICING

At least four previously characterized *cis*-acting sequences have been demonstrated to be crucial for the formation of the spliceosome and completion of the splice reaction. These are the 5' and 3' splice sites, the branch point region, and the polypyrimidine tract (101). Mutations within the conserved regions of these *cis*-acting sequences have been shown to downregulate splicing (115), activate cryptic splice sites (107), or abolish the first or second steps of splicing altogether (110, 111). Each of these conserved regions are known to base-pair with one or more of the snRNPs, the strength of each conserved site being determined by the number of residues involved in the pairing reaction (213).

2.1 5' splice site selection in the v9-v10 region

The sequencing of intron 9 separating exons v9 and v10 and comparison to aberrant splice products has demonstrated the presence of an alternative intronic 5' splice site, denoted as

the I9-5' splice site. Thus the regulation of exon v9 to exon v10 splicing involves an alternative 5' splice site choice (see Figure 19). The two splice sites, v9-5' and I9-5', are found exactly 50 nt apart from GT dinucleotide to GT dinucleotide (Figure 13B). The spacing limit for closely positioned 5' splice sites which are affected by steric hindrance due to overlap in the simultaneous binding of U1 snRNP has been previously determined to be 45 nt (151, 256). In other words, if two 5' splice sites are closer together than 45 nt, the binding of U1 snRNP to one site will hide the other so that it is not recognized by the splicing machinery. Since the v9-5' and intron 9-5' splice sites are greater than 45 nt apart, they are in direct competition for attachment of U1 snRNP and entry into the commitment complex. Eperon *et al.* has shown that when two alternative 5' splice sites are in competition with each other, the choice of splice sites is made dependent upon strength (match to consensus) with the stronger being utilized, or position if the sites are of the same strength with the more downstream of the 5' splice site being utilized (151).

Extensive experimentation by previous researchers using RNAse H and T1 mapping, psoralen or UV crosslinking, or mutation and co-conversion techniques has demonstrated that at least three snRNPs, U1, U5 and U6, interact with the conserved residues at the 5' splice site of the pre-mRNA (116, 117, 118, 137, 138). In particular, the interaction of U1 with the 5' splice site has been demonstrated to be a major determinant in the definition of the 5' splice site region (151). Recognition and binding by U1 occurs early in spliceosomal assembly prior to binding of the other snRNPs to define the commitment complex (118), and is assisted by several crucial protein factors such as SF2/ASF (126, 151, 206), SC35 (194, 203), U1-specific protein C and the phosphorylation dependent U1-70K protein (137). However, it is now generally recognized that the interaction of U1 with the pre-mRNA by itself is not specific enough to specify the site of nucleophilic attack of the branch point adenine (206, 117, 257). More recently, crucial interactions between SC35 and U1 have been demonstrated to be necessary for 5' splice site binding, with U1 proposed to function to

recruit U6 to the 5' splice site (117, 191). U6 has been demonstrated to pair with the 5' splice site at a consensus site which overlaps with that of U1 (117) and perform a dual proofreading function while the conserved first loop of U5 binds using non-canonical base pairing to both the 5' and 3' splice sites at the invariant pre-mRNA AG and GU dinucleotides (127). Mutation of individual residues within the 5' splice site motifs may alter 5' splice site choice (117) or lead to exon skipping if no alternative choices are available. Thus the strength of the 5' splice site is a sum of its ability to bind to each of these snRNPs

A comparison of the 5' splice sites of v9 and I9 with U1, U5 and U6 (Figure 28) demonstrates that the v9-5' splice site closely matches to the mammalian 5' splice site consensus. Thus, the v9-5' splice site should bind strongly to U1 (16 hydrogen bonds as shown in Figure 28B). In contrast, binding to the U6 consensus is limited, with only 12 possible hydrogen bonds observed. The match to consensus in the exon encoded portion which interacts with U5 snRNP is not conserved at all (consensus is AG - Figure 4). Examination of the I9-5' splice site demonstrates a limited ability to bind to U1 snRNP (13 hydrogen bonds) coupled with a more extensive ability to bind to U6 (14 hydrogen bonds) and a consensus match to U5 (Figure 28C). Using a model system composed of an H-*ras* exon placed in the context of human growth hormone, Hwang and Cohen (117) have recently shown that base pairing between positions 5 and 6 of the pre-mRNA, and 1 and 2 of U6 snRNP (see Figure 6) is necessary for efficient splicing. As demonstrated in Figure 28, the only 5' splice site capable of this pairing is the I9-5' splice site (GU to CA: nucleotides marked with an *). Taken together, this evidence indicates that the I9-5' splice site is a stronger, more competitive site than the v9-5' splice site and thus should be utilized more often.

An examination of exon v9-v10 splicing in either the v8-v9-v10 SAGE or the v9-v10 SAGE transcripts as shown in Figure 17 indicates that the I9-5' splice site is indeed used more

frequently in exon v9-v10 splicing. This conclusion is supported by densitometry of Figure 17, demonstrating that the I9-5' splice site is used 6-8 times more often than the v9-5' splice site (66.8% of splice products use the I9-5' splice site in v8-v9-v10, and 54.6% for v9-v10 compared to 8.1% and 8.9% for the v9-5' splice site). A similar 5' splice site weakness is observed in the v8-5' splice site, with 16 possible hydrogen bonds to U1, and 9 to U6. In contrast to the v9-5' splice site, the v8-5' splice site is constitutively used. This constitutive usage occurs presumably because no appropriately distanced downstream alternative splice sites exist within intron 8 (Figure 13B).

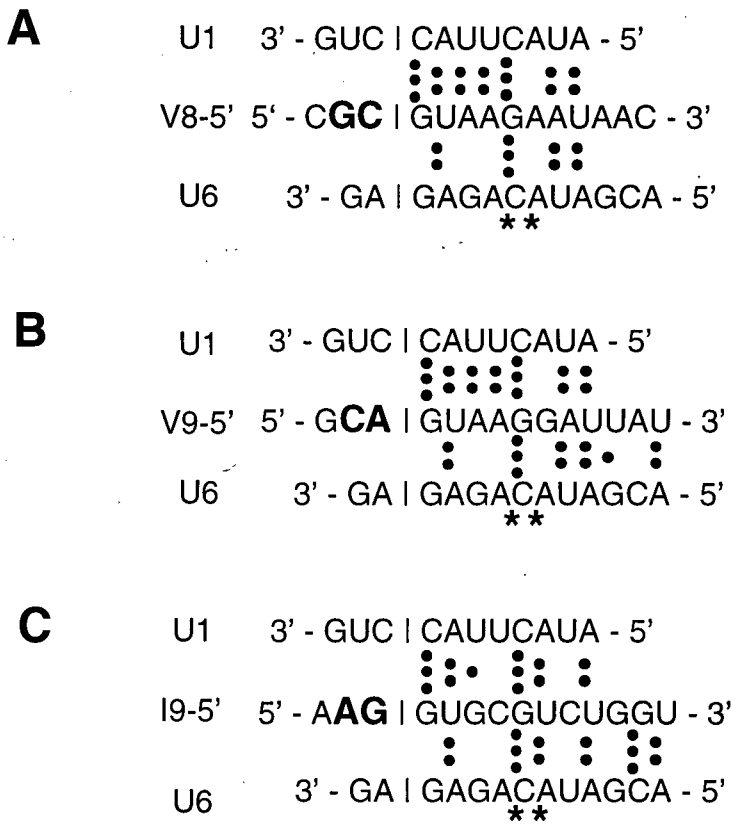


Figure 28: 5' splice site alignments of (A) v8-5', (B) v9-5' and (C) intron 9-5' (I9-5') with the mammalian consensus sequences for U1 and U6 snRNPs. The bonds between nucleotides are indicated as dots (G:C=3 bonds, A:U=2 bonds, and G:U=1 bond). Large bold nucleotides are those which interact with U5. Paired nucleotides of U6 necessary for high efficiency splicing are indicated with an *.

2.2 3' splice site selection in the v9-v10 region

In addition to alternative 5' splice site selection, 3' splice site selection also occurred within the v9-v10 SAGE transcript. Two 3' splice sites, one at the intron/exon boundary (the v10-3' splice site) and one within intron 9 (the I9-3' splice site), each with their own branch point region and polypyrimidine tracts, were utilized during exon v9 to v10 splice reactions. The resulting products observed in Figure 17 were formed by splicing from the v9-5' to v10-3' splice sites (292 bp product), the I9-5' to v10-3' splice sites (450 bp product) and the I9-5' to I9-3' splice sites (750 bp product). Further examination of Figure 17 demonstrates that, similar to the v9-5' splice site, usage of the v10-3' splice site is reduced (19.9% of splice products in the v9-10 SAGE transcripts and 27.8% in the v8-v9-v10 SAGE transcripts) while the I9-3' splice site is preferred (43.5% and 47.9% respectively).

In contrast, although at least one other putative 3' splice site exists within intron 8, only the 3' splice site found at the exon 8/intron 8 boundary was utilized in the exon v8 to v9 splice reactions. Since an appropriate downstream 5' splice site does not occur within the required 300 nt distance to the putative internal intron 8-3' splice site, this site is not used.

According to the exon definition model of Berget *et al.* (139, 258) exons are recognized at the 3' splice site by the initial binding of snRNP U2 (152) and several of its associated proteins such as $U2AF^{65}$ (124, 126), followed by scanning through the exon in a 5' to 3' manner until a 5' splice site is found. Thus suboptimal signals at the 3' splice site may function as a regulatory mechanism for exon recognition. In other words, when the 3' splice site itself, branch point region, or polypyrimidine tract differ significantly from established consensus sequences in these locations to form a weak 3' splice site, *trans*-acting factors bind with low affinity resulting in a reduction of spliceosomal assembly (147). A comparison of the exon flanking v9-3', v10-3' and intron encoded I9-3' splice regions (Figure 29) reveals that several of these 3' splice site regulatory mechanisms may interplay in the control of

splicing of exon v9 to exon v10. Firstly, all three 3' splice sites conform to the previously described consensus motif YAG|G (108). In all three cases, the variable Y is a cytosine which has been demonstrated to be optimal for 3' splice site strength (112). Putative branch point sequences (bps), normally located between 18-40 nt from the 3' splice site invariant AG dinucleotide (107, 162), were found at -27 and -38 for the v9 and I9-3' splice sites respectively. In addition, one long range putative bps for each was found at -100 and -93. Significantly, a set of multiple, putative, overlapping bps for the v10-3' splice site were located at -37, -23, -20 and -16. A sequence comparison of all the putative bps of exons v9, v10 and intron I9 (Figure 30) shows that each differs from consensus at one or two of the pyrimidine residues but the branch point adenine is conserved. However, no obvious sequence differences are observed between the constitutively used v9-3', and the alternatively used I9-3' and v10-3' splice sites. Therefore, it is unlikely that the 3' splice site selection is made based on the sequence of the bps alone.

Prior studies in the alternative splicing of chicken β-tropomyosin (153, 154, 162) and Adenovirus E1A transcripts (155) have characterized multiple bps usage as a major contributor to alternative 3' splice site selection. Thus the multiple bps in the v10-3' splice site may play a role in exon v10 alternative 3' splice site choice. The strongest match to the bps consensus is found for the -100 v9, -93 I9 and -16 v9 bps. Each of these branch point regions are found at suboptimal distance from their 3' splice sites (119), a factor previously noted to be involved in the control of alternative splicing of some pre-mRNAs which may also play a role in alternative splicing of the exon v9 to v10 region.

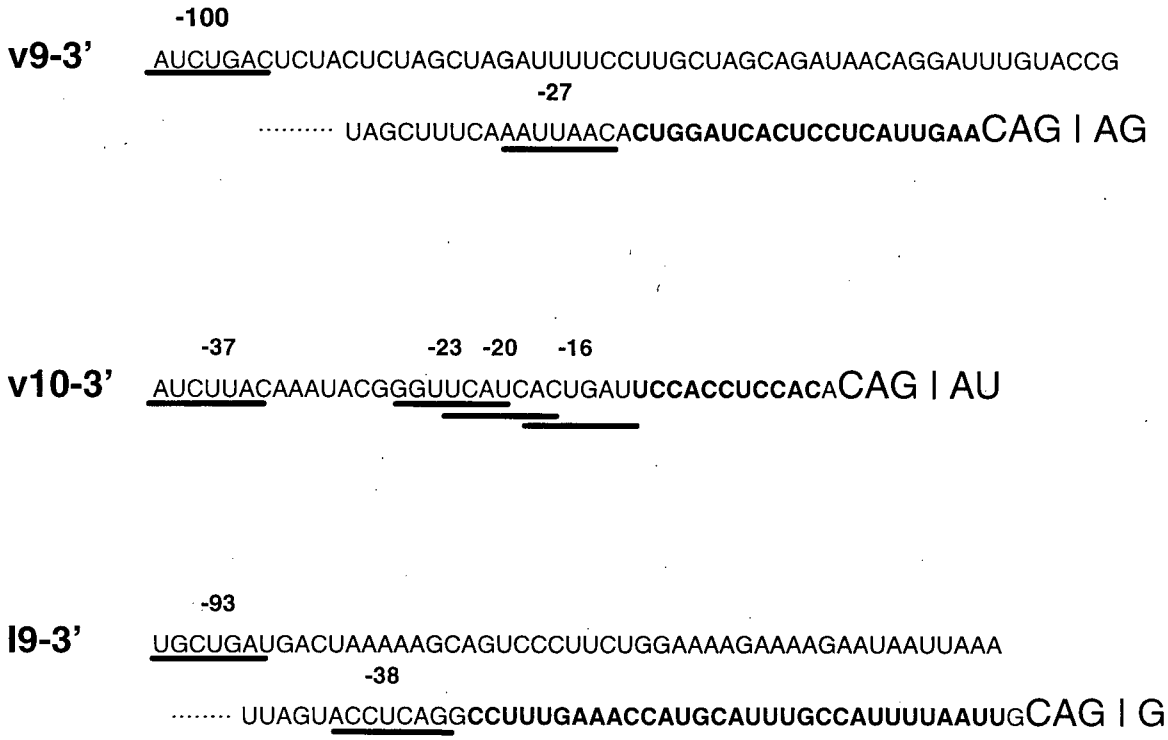


Figure 29: 3'splice site regions of v9-3', v10-3' and intron 9-3' (I9-3'). Branch point regions (consensus: YNYURAY) are underlined with distance from branch point adenine to cleavage site noted above, polypyrimidine tracts are shown in bold, and 3' splice site in large type (consensus: YAG I G)

3' - AUGAU-G - 5'	**U2 snRNP binding site**
YNYURAY	**Branch point consensus**
AUCUG**A**C *	-100 v9
AAUUA**A**C *	-27 v9
UGCUG**AU**	-93 I9
ACCUC**A**G * *	-38 I9
AUCUU**A**C *	-37 v10
GGUUC**AU** * *	-23 v10
UCAUC**A**C * *	-20 v10
CACUG**AU**	-16 v10

Figure 30: Comparison of branch point sequences of the v9-3' (v9), intron 9-3' (I9), and v10-3' (v10) splice sites. Branch point adenine shown in bold. Branch point consensus shown above with U2 snRNA sequence known to bind at the branch point is shown in 3' to 5' orientation. Residues deviating from consensus are marked with an *.

In all three of the 3' splice site regions examined, the polypyrimidine tracts were found to be extraordinarily purine rich (Figure 29), similar to that described for the female-specific 3' splice site of *doublesex* in *Drosophila* (179) The presence of intervening purines in a polypyrimidine tract has been noted to reduce the efficiency of splicing considerably (115, 122, 159, 160). In fact, neither the v9, v10, or I9 polypyrimidine tracts contain any stretch of pyrimidines greater than 5 (Y_5). This lack of pyrimidines is expected to severely limit the ability of several known polypyrimidine tract binding proteins to interact with the polypyrimidine tracts of all three of the 3' splice sites examined. In particular, the binding of $U2AF^{65}$ to the polypyrimidine tract at the bps has been convincingly demonstrated to be crucial for U2 recognition of the bps (119, 223, 259, 260). Thus the presence of a highly purinated, suboptimal polypyrimidine tract at the v10-3' splice site may lead to low affinity $U2AF^{65}$ binding. This factor, combined with multiple competitive bps, may significantly weaken the v10-3' splice site and lead to the preferred use of the I9-3' splice site observed in splicing of the SAGE transcripts.

It is interesting to note that all of the polypyrimidine tracts of the CD44 variant exons composing the unique region are highly purinated (Figure 31). In contrast, the polypyrimidine tracts of the common region constitutively spliced exons retain a high percentage of pyrimidine residues. This observation suggests that polypyrimidine tract recognition may be a common regulatory mechanism utilized in CD44 variant exon splicing.

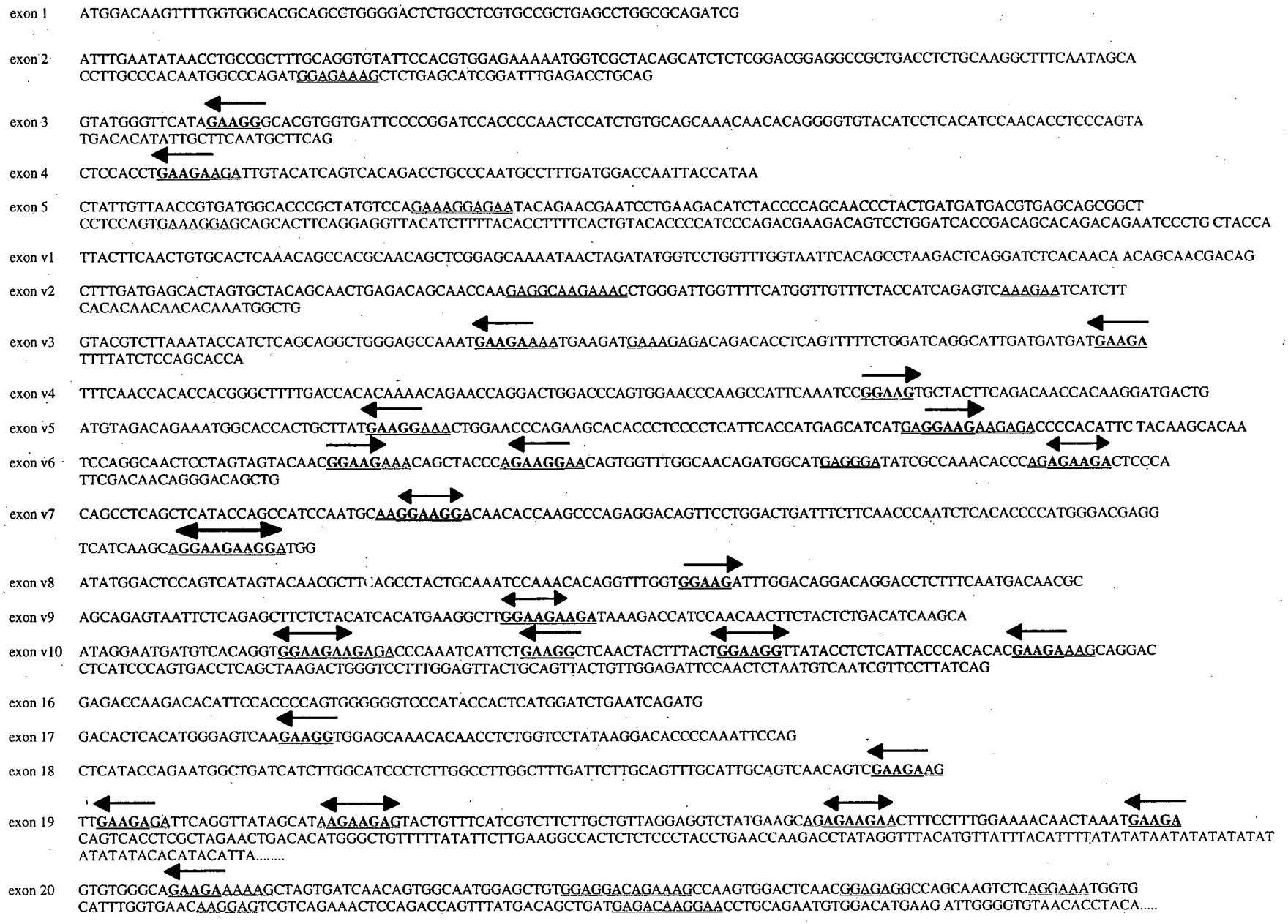
Figure 31: 5' and 3' splice sites of the CD44 pre-mRNA exons. Exons 1 through 5 and 16 through 18 are common to all CD44 isoforms and are constitutively spliced. Exons 6 through 15 (v1-v10) and 19 to 20 are alternatively spliced. Information regarding the intronic nucleotides surrounding exon 6 and downstream of exon 3 is currently unavailable. The alternatively spliced exons show a greater degree of purine substitution (purines shown in bold) within their associated polypyrimidine tracts than the constitutively spliced exons. Exon encoding nucleotides are shown in uppercase, and intron encoding nucleotides are shown in lowercase. The 5' and 3' splice site consensus are shown in bold above.

		3' SPLICE SITES	5' SPLICE SITES
		yyyyyyncagGG/U	A/CAGgua/gagu
	EXON 1	ugcagccaacuuccgaggcaGCCUCAUUGCCCAGCGGACC-----CGUGAGCCUGGCGCAGAUCGgugagugcccgccgcaggcu	
	EXON 2	auuucuauuucuucccauagAUUUGAAUAUAACCUGCCGC-----AGAAAGCUGUGAGCAUCGGAuuugagaccugcagguaaga	
	EXON 3	cuguuguuuuucucuuacagGUAUGGGUUCAUAGAAGGGC-----AUAUUGCUUCAAUGCUUCAG	
	EXON 4	cuuuuuuuccuaccucauagCUCCACCUGAAGAAGAUUGU-----GAUGGACCAAUUACCAUAAGuaugucucucuucuaaucu	
	EXON 5	guaccuuuucucucucccagCUAUUGUUAACCGUGAUGGC-----AGACAGAAUCCCUGCUACCAguaaggagaauaaaucacu	
V1	EXON 6	UUACUUCAACUGUGCACUCA-----CACAACAACACAAAUGGCUG	
V2	EXON 7	uucaacuauuauuacaacagCUUUGAUGAGCACUAGUGCU-----CACAACAACACAAAUGGCUGguaaugaguuauuauuaucu	
V3	EXON 8	aaugcaacucaccacaacagGUACGUCUUCAAAUACCAUC-----AGAUUUUAUCUCCAGCACCAguaagaauaaucaauuacag	
V4	EXON 9	uucaaucaucguuaucacagUUUCAACCACACCACGGGCU-----GACAACCACAAGGAUGACUGguaauggguucugcauauuu	
V5	EXON 10	uuuaaccaucaucacagcagAUGUAGACAGAAAUGGCACC-----CCCACAUUCUACAAGCACAAguaagcaagauggcggucgg	
V6	EXON 11	aacugauauucuucucacagUCCAGGCAACUCCUAGUAGU-----UUCGACAACAGGGACAGCUGguaauggaugguuuaacaag	
V7	EXON 12	cucaaacugcauggucacagCAGCCUCAGCUCAUACCAGC-----UCAACGAGGAAGAAGGAUGGguaauagccucugagauuuu	
V8	EXON 13	acuaauauuugauuccuucagAUAUGGACUCCAGUCAUAGU-----ACCUCUUUCAAUGACAACGCguaagaauaacgaugcgucag	
V9	EXON 14	uucauuccucauucaaacagAGCAGAGUAAUUCUCAGAGC-----UUCUACUCUGACAUCAAGCAguaaggauuauaaaaccuag	
V10	EXON 15	cugauuccaccuccacacagAUAGGAAUGAUGUCACAGGU-----UGUCAAUCGUUCCUUAUCAGguaauuuggcauuuauuauc	
	EXON 16	uuccugauugcucauuacagGAGACCAAGACACAUUCCAC-----UCAUGGAUCUGAAUCAGAUGgugaguucaaaacugcuuua	
	EXON 17	gauguuguuuuucccccuuagGACACUCACAUGGGAGUCAA-----AAGGACACCCCAAAUUCCAGgugaguuucaaacuuugagg	
	EXON 18	auuuaauuuacucauaccagAAUGGCUGAUCAUCUUGGCA-----GCAGUCAACAGUCGAAGAAGguaaggggcuguccuggggg	
	EXON 19	cuauuuucugggaaacuguaGUUGAAGAGAUUCAGGUUAU	
	EXON 20	auuuuuuaaauuauuagGUGUGGGCAGAAGAAAAAGC	

2.3 Exon enhancers and SAGE

Early experimentation with deletion or mutation of exon sequences revealed an active role for exon content in splice site selection (163, 167-170, 261). Intensive research in this area has demonstrated that inclusion of a 6-13 nt purine rich 'ESE' motif (Exon Splicing Enhancer) within exons results in exon inclusion. Mutation or deletion of this motif leads to exon skipping (173, 262). These ESEs have been demonstrated to function as binding sites for SR proteins (171, 195, 211) which interact with the ESE with different affinities to promote the binding of essential splicing factors such as U2AF^{35} to U2AF^{65} at the bps at the 3' splice site (173, 195, 207), and to stabilize the binding of U1 snRNP at the 5' splice site (see Figure 8B and C). One particular ESE motif 'RGAAGAAC' has been characterized as a high affinity SF2/ASF binding site (212), and the presence of the 'RGAAG' portion of this motif is known to result in inclusion of the bGH alternatively spliced exon D (166) and the fibronectin (FN) exon ED1 (174). Since SF2/ASF is known to interact with U1 snRNP through the associated U1-70K protein, (107, 125, 195) and also with U2AF^{35} (199), this 'RGAAG' motif within an exon may also influence both 5' and 3' splice site selection. Interestingly, this motif has also been shown to act as a U1 snRNP binding site (173) presumably through similar interactions. Purine rich regions which resemble the SF2/ASF ESE have been located within many alternatively spliced exons in well-characterized experimental systems such as mouse IgM exon M2 (173), human FN exon EDIIIA and EDI (174), bGH exon 5 (167), rat β-tropomyosin exon 8 (263), cardiac troponin T exon 5 (169), CD45 exon 4 (170), HIV type I *tat-rev* exons (196) and are also located within the CD44 unique region exons (Figure 32).

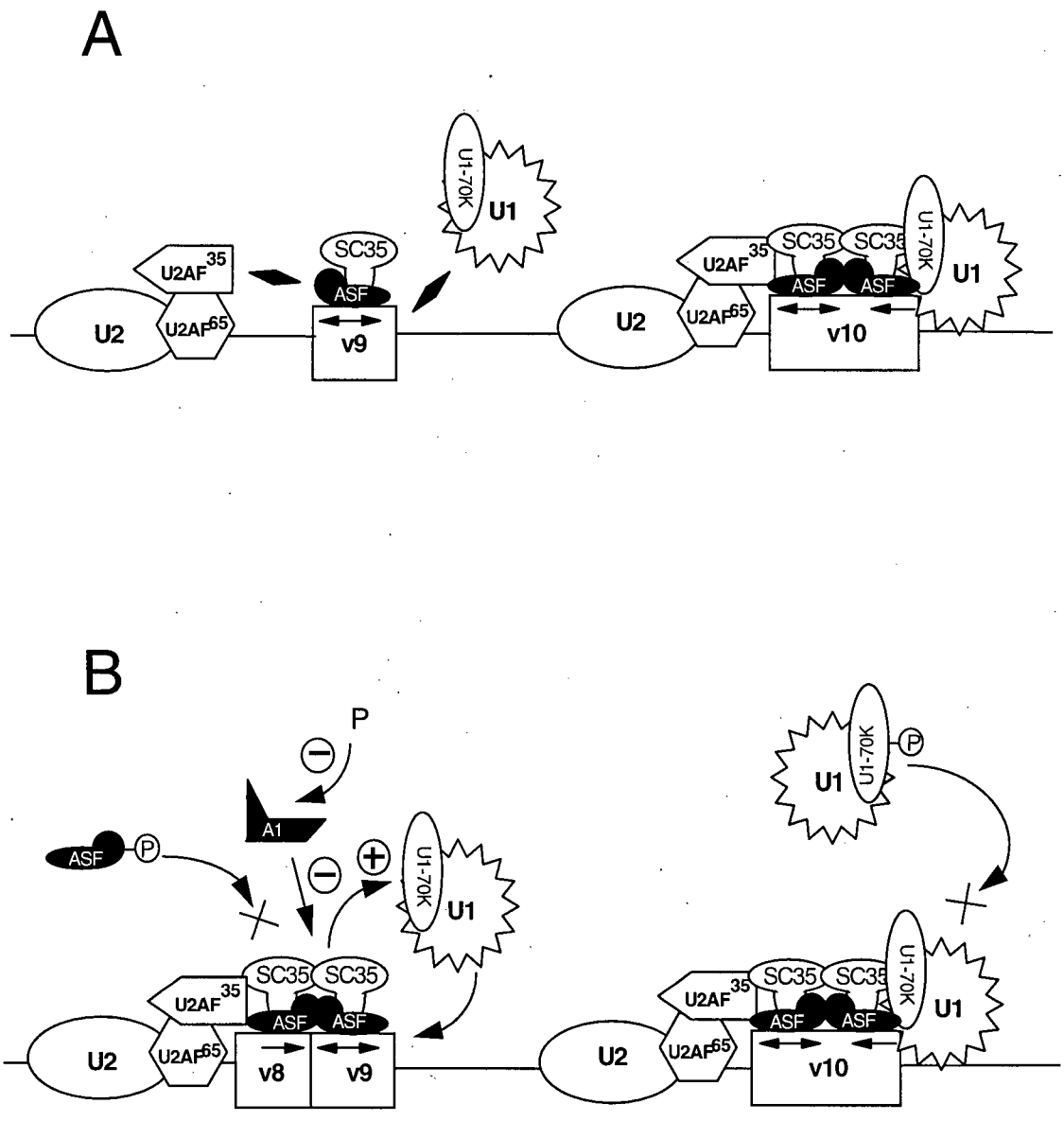
Figure 32: ESEs located within CD44 exons. Multiple putative ESEs are located within CD44 exons, particularly within the unique region. The ASF/SF2 ESE, 'RGAAG' is shown underlined and in bold with direction indicated by an arrow. Note how the direction of the repeat alternates throughout the unique region (exons 6 through 15). Other pure rich motifs are double underlined.


exon 1 ATGGACAAGTTTTGGTGGCACGCAGCCTGGGGACTCTGCCTCGTGCCGCTGAGCCTGGCGCAGATCG
exon 2 ATTTGAATATAACCTGCCGCTTTGCAGGTGTATTCCACGTGGAGAAAAATGGTCGCTACAGCATCTCTCGGACGGAGGCCGCTGACCTCTGCAAGGCTTTCAATAGCA
CCTTGCCCACAATGGCCCAGATGGAGAAAGCTCTGAGCATCGGATTTGAGACCTGCAG
exon 3 GTATGGGTTCATAGAAGGGCACGTGGTGATTCCCCGGATCCACCCCAACTCCATCTGTGCAGCAAACAACACAGGGGTGTACATCCTCACATCCAACACCTCCCAGTA
TGACACATATTGCTTCAATGCTTCAG
exon 4 CTCCACCTGAAGAAGATTGTACATCAGTCACAGACCTGCCCAATGCCTTTGATGGACCAATTACCATAA
exon 5 CTATTGTTAACCGTGATGGCACCCGCTATGTCCAGAAAGGAGAATACAGAACGAATCCTGAAGACATCTACCCCAGCAACCCTACTGATGATGACGTGAGCAGCGGCT
CCTCCAGTGAAAGGAGCAGCACTTCAGGAGGTTACATCTTTTACACCTTTTCACTGTACACCCCATCCCAGACGAAGACAGTCCTGGATCACCGACAGCACAGACAGAATCCCTG CTACCA
exon v1 TTACTTCAACTGTGCACTCAAACAGCCACGCAACAGCTCGGAGCAAAATAACTAGATATGGTCCTGGTTTGGTAATTCACAGCCTAAGACTCAGGATCTCACAACA ACAGCAACGACAG
exon v2 CTTTGATGAGCACTAGTGCTACAGCAACTGAGACAGCAACCAAGAGGCAAGAAACCTGGGATTGGTTTTCATGGTTGTTTCTACCATCAGAGTCAAAGAATCATCTT
CACACAACAACACAAATGGCTG
exon v3 GTACGTCTTAAATACCATCTCAGCAGGCTGGGAGCCAAATGAAGAAAATGAAGATGAAAGAGACAGACACCTCAGTTTTTCTGGATCAGGCATTGATGATGATGAAGA
TTTTATCTCCAGCACCA
exon v4 TTTCAACCACACCACGGGCTTTTGACCACACAAAACAGAACCAGGACTGGACCCAGTGGAACCCAAGCCATTCAAATCCGGAAGTGCTACTTCAGACAACCACAAGGATGACTG
exon v5 ATGTAGACAGAAATGGCACCACTGCTTATGAAGGAAACTGGAACCCAGAAGCACACCCTCCCCTCATTCACCATGAGCATCATGAGGAAGAAGAGACCCCACATTC TACAAGCACAA
exon v6 TCCAGGCAACTCCTAGTAGTACAACGGAAGAAACAGCTACCCAGAAGGAACAGTGGTTTGGCAACAGATGGCATGAGGGATATCGCCAAACACCCAGAGAAGACTCCCA
TTCGACAACAGGGACAGCTG
exon v7 CAGCCTCAGCTCATACCAGCCATCCAATGCAAGGAAGGACAACACCAAGCCCAGAGGACAGTTCCTGGACTGATTTCTTCAACCCAATCTCACACCCCATGGGACGAGG
TCATCAAGCAGGAAGAAGGATGG
exon v8 ATATGGACTCCAGTCATAGTACAACGCTTCAGCCTACTGCAAATCCAAACACAGGTTTGGTGGAAGATTTGGACAGGACAGGACCTCTTTCAATGACAACGC
exon v9 AGCAGAGTAATTCTCAGAGCTTCTCTACATCACATGAAGGCTTGGAAGAAGATAAAGACCATCCAACAACTTCTACTCTGACATCAAGCA
exon v10 ATAGGAATGATGTCACAGGTGGAAGAAGAGACCCAAATCATTCTGAAGGCTCAACTACTTTACTGGAAGGTTATACCTCTCATTACCCACACACGAAGAAAGCAGGAC
CTCATCCCAGTGACCTCAGCTAAGACTGGGTCCTTTGGAGTTACTGCAGTTACTGTTGGAGATTCCAACTCTAATGTCAATCGTTCCTTATCAG
exon 16 GAGACCAAGACACATTCCACCCCAGTGGGGGGTCCCATACCACTCATGGATCTGAATCAGATG
exon 17 GACACTCACATGGGAGTCAAGAAGGTGGAGCAAACACAACCTCTGGTCCTATAAGGACACCCCAAATTCCAG
exon 18 CTCATACCAGAATGGCTGATCATCTTGGCATCCCTCTTGGCCTTGGCTTTGATTCTTGCAGTTTGCATTGCAGTCAACAGTCGAAGAAG
exon 19 TTGAAGAGATTCAGGTTATAGCATAAGAAGAGTACTGTTTCATCGTCTTCTTGCTGTTAGGAGGTCTATGAAGCAGAGAAGAACTTTCCTTTGGAAAACAACTAAATGAAGA
CAGTCACCTCGCTAGAACTGACACATGGGCTGTTTTTATATTCTTGAAGGCCACTCTCTCCCTACCTGAACCAAGACCTATAGGTTTACATGTTATTTACATTTTATATATAATATATATATAT
ATATATACACATACATTA.....
exon 20 GTGTGGGCAGAAGAAAAAGCTAGTGATCAACAGTGGCAATGGAGCTGTGGAGGACAGAAAGCCAAGTGGACTCAACGGAGAGGCCAGCAAGTCTCAGGAAATGGTG
CATTTGGTGAACAAGGAGTCGTCAGAAACTCCAGACCAGTTTATGACAGCTGATGAGACAAGGAACCTGCAGAATGTGGACATGAAG ATTGGGGTGTAACACCTACA.....

CD44 exons v9 and v10 each contain this purine enhancer motif, with one repeat found in v8, two repeats found in v9 and four in v10. Most of the CD44 unique region exons with the exception of v1 and v2 contain multiples of the 'RGAAG' motif, while the constitutive exons contain only one or no motifs. Interestingly, these repeats occur throughout the unique region exons in an alternating back-to-back fashion (Figure 32), suggesting that a directionality of binding of SF2/ASF may be important for ESE function. Furthermore, each back-to-back 'RGAAG' motif occurs between 57 and 80 nt apart, often within a single exon or between two exons if spliced together, implying a minimal distance constraint on the motif.

Studies utilizing the SAGE construct assay suggest that clustering two of these SF2/ASF ESE motifs in a back-to-back fashion may result in enhanced splicing to the downstream exon. As demonstrated in Figure 20, the presence of the 'pre-spliced' v89 exon resulted in increased v9-5' splice site use from 8.5% of total splice products to 22.4%. Since SF2/ASF is known to influence 5' splice site selection through the U1-70K protein as mentioned above, it is conceivable that back-to-back bridging interactions on the v89 'exon' caused by binding of SF2/ASF to the 'RGAAG' ESE motifs formed when v8 and v9 are joined, stabilizes the interactions of U1 snRNP with the 5' splice site (Figure 33). Thus the exon flanking v9-5' splice site is strengthened and splicing using this site is increased. This proposed exon spanning function of SF2/ASF is similar to that previously hypothesized by Wu and Maniatis (199) to link the 5' and 3' splice sites during spliceosomal assembly. Furthermore, the clustering of ESE motifs by splicing implies a role for upstream or downstream splicing events in recognition and regulation of other exons, and may help to explain why certain CD44 exons such as v8 and v9 appear to be expressed in clusters (264).

Figure 33: Proposed mechanism of 'RGAAG' ESE clustering enhancement. SF2/ASF has been demonstrated to recognize and bind to 'RGAAG' motifs located within the exon by its RBD as well as binding to itself, SC35, U1-70K and U2AF^{35} through its RS domain. These interactions promote both 5' and 3' splice site recognition as described in the text. Likewise, SC35 also binds to itself, SF2/ASF, U1-70K and U2AF^{35} but recognizes an RNA motif distinct from that of SF2/ASF. (A) Exons which contain only one copy of the motif bind only one SF2/ASF molecule, and a moderate enhancement of 5' and 3' splice site recognition results, with 5' splice sites proximal to the 3' splice site favoured. (B) If the exon contains 2 or more motifs, multiple SF2/ASF binding occurs leading to high affinity recognition of the 5' and 3' splice sites and distal 5' splice site selection. Thus splicing of exons v8 to v9, which causes clustering of back-to-back 'RGAAG' motifs, leads to an increase in the exon flanking v9-5' splice site usage. v10, which has 4 repeats of the motif, does not require clustering. Overexpression of SF2/ASF in the absence of an ESE cluster induces SF2/ASF binding to motifs of less affinity and leads to proximal 5' splice site selection through similar interactions. Phosphorylation of SF2/ASF inhibits its ability to interact with other SR proteins through its RS domain, and inhibits enhancer complex formation. hnRNP A1, also negatively regulated by phosphorylation, competitively inhibits the interaction of SF2/ASF with the RNA. Adapted from (199).


A
U1-70K
U1
SC35
U2AF35
ASF
U2
U2AF65
v9
v10
B
P
A1
ASF
U1-70K
U1
SC35
U2AF35
ASF
U2
U2AF65
v8
v9
v10

2.4 Other putative cis-acting sequence motifs

Several other putative *cis*-acting motifs were located in the v8-v9-v10 genomic sequence which have a demonstrated regulatory role in other splicing systems and may also perform a regulatory function in CD44 unique region exon splicing. These include another type of position dependent 3' splice site enhancer called the dsxRE found within 300 nt downstream of the 3' splice site of *doublesex* (*DSX*) in *Drosophila* (179, 207). This 13 nt regulatory sequence, composed of overlapping motifs of $Y_5P_2Y_2P_2YP$, is repeated 6 times within the female-specific exon of *DSX* and binds the *Drosophila* SR proteins TRA and TRA-2 regulating female-specific splicing (179, 207, 208, 265). A similar motif of repeated alternating Y_7PY_4 or Y_4PY_7 pyrimidine/purine residues was found within the 300 nt limit clustered at the 5' and 3' end of intron 9, and also surrounding the internal I9-3' splice site (Figure 13B). Similarly, the motif was also found three times in intron 8 within 300 nt of the 5' splice site. By analogy to the dsxRE motif, this Y_7PY_4 or Y_4PY_7 motif may also function as putative binding sites for SR proteins and act as an enhancer for alternative 3' splice site selection.

A second sequence motif was noted to occur within the v8-v9-v10 CD44 genomic region in close association to either 5' or 3' splice site regions. This motif, YGGTGGT/AGG, was found twice within intron 8 and three times within I9. In intron 8 the motif was found 90 nt downstream of the 5' splice site, and interestingly was also found 50 nt upstream of a cryptic 3' splice site. This cryptic internal 3' splice site was not utilized in either the v8-v9 or v8-v9-v10 SAGE constructs. Similarly the same motif was located 70 nt downstream of the v9-5' splice site, 8 nt downstream of the I9-5' splice site and 50 nt upstream of the v10-3' splice site. Although no function for this motif was found in this present work, the close proximity of the motif to the 5' and 3' splice sites makes it a likely candidate for the binding of putative regulatory proteins.

When examined for RNA secondary structure using computer modeling strategies, several putative stem-loop structures were indicated to occur at positions close to the 5' and 3' splice sites within intron 9 (Figure 13B). The strongest stem-loop structures appeared to be of 20 nt as previously described (266), and are located between the v9-5' and I9-5' splice sites, at the -93 bps and the I9-3' splice site itself, as well as within the multiple overlapping bps of the v10-3' splice site. A putative regulatory stem-loop structure was also noted to occur between the -100 and -27 bps of the v9-3' splice site within intron 8 (Figure 13B). Thus RNA secondary structure is strongly implicated as a possible control mechanism for exon v9 to v10 splicing and may also function in exon v8 to v9 splicing

3. INDUCTION OF SAGE SPLICING BY *TRANS*-ACTING FACTORS

Recent work by Screaton *et al.* (190) using CD44 alternative exon v6 as a marker of T cell activation demonstrated changes in v6 expression induced by phytohemagglutinin (PHA) stimulation. When assayed by Northern blotting, 3 SR proteins p40, p55 and p30c displayed alterations in mRNA levels coincidental with the change in CD44 epitope expression. Thus CD44 splicing, like other alternatively spliced pre-mRNAs, may also be affected by the expression and regulation of SR proteins.

3.1 Serine/Threonine kinase inhibition induces SAGE splicing

The phosphorylation and de-phosphorylation of proteins has been proposed to be a control mechanism for protein function and has also been implicated in the control of alternative splicing. The SAGE vector constructs were used to determine if phosphorylation of proteins is also a regulatory mechanism in splicing of CD44 exons v8, v9 and v10. Incubation of SAGE vector transfected K562 cells with the potent serine/threonine (S/T) kinase inhibitor, staurosporine (267, 268) altered the pattern of splicing of exon v9 to v10, such that the exon

flanking v9-5' and v10-3' splice sites were recognized, and a productive ALP protein was produced (Figures 25 and 27). In addition, staurosporine treatment of various myeloid cell lines particularly noted in KG1 cells, altered v10 expression and led to the expression of other CD44 variant isoforms (Figure 21).

Recent evidence has demonstrated that the phosphorylation of spliceosomal associated proteins regulates pre-mRNA splicing. Several examples of this, such as hnRNP A1 and A2 (231), hnRNP C (217, 230-233), the nuclear matrix phosphoprotein P255 which associates with U5 (269), and an hnRNP A/B type analogue from brine shrimp (270), have recently been described. In addition, interest has centered upon the phosphorylation of the serine residues within the RS domains of SR proteins (271). Since SR proteins associate with other proteins through their RS domains (198), it is conceivable that the phosphorylation of these proteins on these domains may by involved in the regulation of splice site choice (271).

Phosphorylation of SF2/ASF is known to regulate its interaction with the U1-70K proteins and facilitate the binding of U1 snRNP to the 5' splice site (198, 223). This process is antagonized by hnRNP A1 which is also regulated by phosphorylation (226, 228, 229). As previously discussed, the v9 and v10 exons contain several putative binding sites for SF2/ASF (Discussion section 2.3). Taken together, these observations support the notion that the binding of SF2/ASF to exons v9 and v10, regulated by the phosphorylation status of the protein, performs a regulatory role in the alternative splicing of this region.

3.2 Putative kinase involved in the regulation of alternative splicing

Although the exact kinase which was modulated by staurosporine in this assay is not known, there are several likely candidates which have been previously demonstrated to associated with and to phosphorylate spliceosomal associated proteins. One particularly well described S/T kinase which may be involved in the regulation of splicing is the protein kinase C (PKC)

family. There are currently at least 12 described members of this family, divided into 3 sub-types based on their structure and known agonists: classical (cPKCs) α, $\beta_{I/II}$, and γ which are Ca^{2+} and phospholipid dependent, novel (nPKCs) δ, ε, η, θ, μ which are Ca^{2+}-independent, and atypical (aPKCs) ζ, ι, λ, which cannot be activated by phorbol esters (272). K562 cells are known to express four of these isotypes. α, β_{II}, ζ and at low levels, ε (245, 273, 274). hnRNP A1 has recently been described as an *in vivo* substrate of aPKC ζ, but not of the classical PKCs or of PKC ε, through direct association with the PKC ζ catalytic domain (229). Phosphorylation of hnRNP A1 by PKC in its C-terminal domain was associated with a conformational change in this domain, and a decrease in the affinity of A1 for single-stranded polynucleotides *in vitro.* Thus phosphorylation of hnRNP A1 regulated its ability to enhance the renaturation of single stranded nucleic acids (228). Importantly, A1 is known to bind stably to both U2 and U4, to regulate the binding of U2 to the bps (227), and to influence distal 5' splice site selection (188). Thus phosphorylation of A1 may result in negative regulation of bps selection.

hnRNP A1 as well as another molecule, hnRNP C is also phosphorylated in its C-terminal domain by another S/T kinase, casein kinase II (CK II; 228, 230, 232). Staurosporine inhibition of CKII may also be responsible for the alterations observed in exon v9 to v10 splicing. The hnRNP C proteins, C1 and C2 bind to uridine-rich polypyrimidine tracts (184, 232) and polyadenylation sites (217). Phosphorylation of hnRNP C proteins has been shown to negatively regulate their binding to RNA (230), and to be modulated by calmodulin (231), by the cell cycle (217), and interestingly is also modulated by a 3' region of U6 snRNP (233).

Two additional SR-specific kinases have recently been identified through the yeast two-hybrid system and chromatographic techniques. These are the dual specific (serine/threonine and tyrosine) LAMMER family kinase Clk/Sty (221) and the S/T kinase SRPK1 (224) Both

of these kinases have demonstrated an *in vitro* ability to phosphorylate SF2/ASF (221, 225, 275), and may also be inhibited by staurosporine in the SAGE assay.

In all examples cited above it is interesting to note that phosphorylation of spliceosomal-associated proteins negatively regulated their function (219, 224, 271). Most noticeably, phosphorylation of the U1-70K protein which is also phosphorylated by SRPK1 (224) as well as potentially by others (234), severely inhibited splicing (235) presumably mediated by an inability to form associations with other SR proteins through phosphorylation of its RS domain (223). Thus one potential mechanism for the action of staurosporine is through inhibition of S/T kinases, decreasing phosphorylation of these splice regulatory proteins particularly U1-70K and hnRNP A1, and thus promoting the use of the v9-5' and v10-3' splice sites.

4. FUNCTIONAL CONSEQUENCES OF INTRON RETENTION

Generally, the retention of introns within mRNAs which are subsequently translated results in a non-functional, truncated protein due to premature stop codons encoded by the intron. As such, intron retention has been causally linked to 6% of human genetic diseases (276). In contrast, in at least two described cases, the retention of intron sequences particularly in place of terminal exon sequences, resulted in the expression of functional isoforms with altered C-terminal domains. In both of these cases, phosphotyrosine phosphatase 1B (PTP-1B) and bovine growth hormone (bGH), these isoforms are suspected to function as regulatory molecules for their normally spliced counterparts (166, 277, 278). Information gained from the study of alternatively spliced SAGE constructs was used to assay for the retention of intron 9 in various myeloid cell lines (Figure 27). The results of this assay showed that retention of this intron was both inducible and cell line specific. For example, KG1 cells showed no evidence of intron retention, while multiple intron fragments were found in

stimulated KG1a cells and K562 cells. It is also interesting to note that different bands corresponding to the activation of other alternative intron encoded splice sites were present in KG1a cells, indicating that the use of internal intron splice sites varied from cell line to cell line. While no evidence is presented in this work to suggest that mRNA's which include intron 9 are translated, a stop codon common to all of these transcripts is found 18 nt downstream from the v9-5' splice site which would form a truncated, and possibly soluble protein if utilized. Recent work by Yoshida *et al.* (279), and Matsumura *et al.* (38) has demonstrated alternative intron splicing and retention of variant intron 4 in gastrointestinal tumors and bladder carcinoma cells. Thus the dis-regulation of pre-mRNA alternative splicing and subsequent retention of CD44 variant introns may be a common phenomena which occurs during carcinogenesis.

5. CONCLUSIONS

Pre-mRNA splicing is a complex process utilizing numerous regulatory mechanisms that involve both *cis*-acting sequences and *trans*-acting factors. In addition to constitutive splicing, the regulatory mechanisms controlling the splicing of CD44 pre-mRNA are further complicated by the requirements of alternative splicing of at least 10 exons. Clearly, multiple avenues exist where a loss of regulation could result in the inclusion of variant exons within the CD44 mRNA, and thus lead to the increase in cell surface expression of isoforms observed on neoplastic and metastatic cells.

In summary, a novel *in vivo* splice assay system was developed and used to investigate the molecular mechanisms which regulate the alternative splicing of CD44 exons v8, v9 and v10. Alternative, and possibly regulatory 5' and 3' splice sites were located within intron 9. Furthermore, modulation of alternative splicing of these exons by the addition of the serine/threonine kinase inhibitor staurosporine was demonstrated. Therefore, use of the SAGE assay system should facilitate a further dissection of the molecular mechanisms which regulate alternative splicing of these CD44 exons.

6. FUTURE DIRECTIONS

Although many of the factors involved in constitutive splicing have been described, those mechanisms which regulate alternative splicing have remained largely unknown. Using the SAGE assay system described in this work, the following studies could be undertaken:

1) The SAGE assay system provides a quick screening tool to investigate the involvement of known *trans*-acting proteins in the splicing of the CD44 v8, v9 and v10

exons. For example, SR proteins have been demonstrated to function as *trans*-acting factors which modulate splicing (194). Recent work by Screaton *et al.* has demonstrated a role for SR proteins in CD44 exon v6 splicing (190). However, the involvement of SR proteins in v8-v9-v10 splicing is unknown. Using the stably transfected SAGE K562 cell lines, vectors containing different SR proteins can be introduced and the induction of splicing assessed by FACS analysis with mAb A3A11.

2) The SAGE assay system may also be used as a tool to clone unknown positive and negative *trans*-acting regulators of splicing. Similar to the method described above, expression cloning libraries can be introduced into the v8-v9 SAGE K562 cells to screen for negative regulators, or the v8-v9-v10/v9-v10 SAGE K562 cells to screen for positive regulators. Separation of positive splicing from negative non-splicing cells can be performed by FACS using mAb A3A11.

3) Several putative *cis*-acting sequences were described which may serve as regulatory regions of v8-v9-v10 splicing. These regions can be mutated by site-directed mutagenesis, deleted, reversed with respect to the 5' and 3' splice sites, or moved to other locations within the SAGE construct and their effects assessed by histological assay and RT PCR. In particular, the SF2/ASF ESE sequences within the v89-10 SAGE construct could be modulated, and the effect on splicing observed.

4) Splicing of the v8-v9 portion of the v8-v9-v10 SAGE transcript occurred readily while the v9-v10 portion did not splice. The role of the various exon flanking splice sites in the regulation of v8-v9-v10 splicing is unclear. The effect of these *cis*-acting sequences could be investigated by exon or intron swapping. For example, new SAGE constructs consisting of v9-v8-v10 or v10-v9-v8 could be constructed and use of the various exon and intron specific splice sites determined.

5) RT PCR of total RNA from the induced and non-induced cell lines using the internal I9 5' primer revealed several bands which correlated to the use of the exon flanking and

intron encoded splice sites. These PCR products should be cloned and sequenced to confirm their putative identities. Furthermore, several bands were noted which did not correspond products arising from the use of the described splice sites, suggesting that other splice sites found in intron 9 may be active within the CD44 pre-mRNA in other cell types. These results should be confirmed by cloning and sequencing of the PCR products.

6) Both this work and that of others (280), (279) have described the retention of intron sequences within different CD44 mRNAs. Many questions regarding intron retention remain unclear and would be interesting points to study. For example, are the mRNAs which contain retained introns translated? Are the proteins they encode functional? Does the presence of the retained intron result in mRNA instability, and thus be considered a regulatory mechanism for translation? Does the retention of introns only occur within CD44 mRNA, or is it common to other alternatively spliced mRNAs? Is the retention of introns an outcome of oncogenic transformation, or does it also occur in normal cells? These and other questions in alternative splicing remain to be answered.

V BIBLIOGRAPHY

1. Nicolson, G.L., *Tumor cell instability, diversification, and progression to the metastatic phenotype: form oncogene to oncofetal expression.* Cancer Research, (1987) **47**: 1473-1487.

2. Herrlich, P., M. Zoller, S.T. Pals, *et al., CD44 splice variants: metastases meet lymphocytes. [Review].* Immunology Today, (1993) **14**: 395-9.

3. Picker, L.J. and E.C. Butcher, *Physiological and molecular mechanisms of lymphocyte homing. [Review].* Annual Review of Immunology, (1992) **10**: 561-91.

4. Gunthert, U., M. Hofmann, W. Rudy, *et al., A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells.* Cell, (1991) **65**: 13-24.

5. Haynes, B.F., M.J. Telen, L.P. Hale, *et al., CD44--a molecule involved in leukocyte adherence and T-cell activation [published erratum appears in Immunol Today 1990 Mar;11(3):80]. [Review].* Immunology Today, (1989) **10**: 423-8.

6. Jalkanen, S., M. Jalkanen, R. Bargatze, *et al., Biochemical properties of glycoproteins involved in lymphocyte recognition of high endothelial venules in man.* Journal of Immunology, (1988) **141**: 1615-1623.

7. Jackson, D.G., J. Buckley and J.I. Bell, *Multiple variants of the human lymphocyte homing receptor CD44 generated by insertions at a single site in the extracellular domain.* Journal of Biological Chemistry, (1992) **267**: 4732-9.

8. Tolg, C., M. Hofmann, P. Herrlich, *et al., Splicing choice from ten variant exons establishes CD44 variability.* Nucleic Acids Research, (1993) **21**: 1225-9.

9. Screaton, G.R., M.V. Bell, D.G. Jackson, *et al., Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons.* Proceedings of the National Academy of Sciences of the United States of America, (1992) **89**: 12160-4.

10. Arch, R., K. Wirth, M. Hofmann, *et al., Participation in normal immune responses of a metastasis-inducing splice variant of CD44 [see comments].* Science, (1992) **257**: 682-5.

11. Hofmann, M., W. Rudy, M. Zoller, *et al., CD44 splice variants confer metastatic behavior in rats: homologous sequences are expressed in human tumor cell lines.* Cancer Research, (1991) **51**: 5292-7.

12. Birch, M., S. Mitchell and I.R. Hart, *Isolation and characterization of human melanoma cell variants expressing high and low levels of CD44.* Cancer Research, (1991) **51**: 6660-7.

13. Haynes, B.F., H.X. Liao and K.L. Patton, *The transmembrane hyaluronate receptor (CD44): multiple functions, multiple forms. [Review].* Cancer Cells, (1991) **3**: 347-50.

14. Conrad, P., B.L. Rothman, K.A. Kelley, *et al., Mechanism of peripheral T cell activation by coengagement of CD44 and CD2.* Journal of Immunology, (1992) **149**: 1833-9.

15. Huet, S., H. Groux, B. Caillou, *et al., CD44 contributes to T cell activation.* Journal of Immunology, (1989) **143**: 798-801.

16. Shimizu, Y., G.A. Van Seventer, R. Siraganian, *et al., Dual role of the CD44 molecule in T cell adhesion and activation.* Journal of Immunology, (1989) **143**: 2457-63.

17. Denning, S.M., P.T. Le, K.H. Singer, *et al., Antibodies against the CD44 p80, lymphocyte homing receptor molecule augment human peripheral blood T cell activation.* Journal of Immunology, (1990) **144**: 7-15.

18. Webb, D.S., Y. Shimizu, G.A. Van Seventer, *et al., LFA-3, CD44, and CD45: physiologic triggers of human monocyte TNF and IL-1 release.* Science, (1990) **249**: 1295-7.

19. Koopman, G., Y. van Kooyk, M. de Graaff, *et al., Triggering of the CD44 antigen on T lymphocytes promotes T cell adhesion through the LFA-1 pathway.* Journal of Immunology, (1990) **145**: 3589-93.

20. Bruynzeel, I., G. Koopman, L.M. van der Raaij, *et al., CD44 antibody stimulates adhesion of peripheral blood T cells to keratinocytes through the leukocyte function-associated antigen-1/intercellular adhesion molecule-1 pathway.* Journal of Investigative Dermatology, (1993) **100**: 424-8.

21. Guo, M.M. and J.E. Hildreth, *HIV-induced loss of CD44 expression in monocytic cell lines.* Journal of Immunology, (1993) **151**: 2225-36.

22. Shepley, M.P. and V.R. Racaniello, *A monoclonal antibody that blocks poliovirus attachment recognizes the lymphocyte homing receptor CD44.* Journal of Virology, (1994) **68**: 1301-8.

23. Miyake, K., K.L. Medina, S. Hayashi, *et al., Monoclonal antibodies to Pgp-1/CD44 block lympho-hemopoiesis in long-term bone marrow cultures.* Journal of Experimental Medicine, (1990) **171**: 477-88.

24. Lewinsohn, D.M., A. Nagler, N. Ginzton, *et al.*, *Hematopoietic progenitor cell expression of the H-CAM (CD44) homing-associated adhesion molecule.* Blood, (1990) **75**: 589-95.

25. Thomas, L., H.R. Byers, J. Vink, *et al.*, *CD44H regulates tumor cell migration on hyaluronate-coated substrate.* Journal of Cell Biology, (1992) **118**: 971-7.

26. Idzerda, R.L., W.G. Carter, C. Nottenburg, *et al.*, *Isolation and DNA sequence of a cDNA clone encoding a lymphocyte adhesion receptor for high endothelium.* Proceedings of the National Academy of Science of the USA, (1989) **86**: 4659-4663.

27. Goldstein, L.A., D.F. Zhou, L.J. Picker, *et al.*, *A human lymphocyte homing receptor, the hermes antigen, is related to cartilage proteoglycan core and link proteins.* Cell, (1989) **56**: 1063-72.

28. Neame, S.J. and C.M. Isacke, *Phosphorylation of CD44 in vivo requires both Ser323 and Ser325, but does not regulate membrane localization or cytoskeletal interaction in epithelial cells.* EMBO Journal, (1992) **11**: 4733-8.

29. Pure, E., R.L. Camp, D. Peritt, *et al.*, *Defective phosphorylation and hyaluronate binding of CD44 with point mutations in the cytoplasmic domain.* Journal of Experimental Medicine, (1995) **181**: 55-62.

30. Lokeshwar, V.B., N. Fregien and L.Y. Bourguignon, *Ankyrin-binding domain of CD44(GP85) is required for the expression of hyaluronic acid-mediated adhesion function.* Journal of Cell Biology, (1994) **126**: 1099-109.

31. Bourguignon, L.Y., V.B. Lokeshwar, X. Chen, *et al.*, *Hyaluronic acid-induced lymphocyte signal transduction and HA receptor (GP85/CD44)-cytoskeleton interaction.* Journal of Immunology, (1993) **151**: 6634-44.

32. Carter, W.G. and E.A. Wayner, *Characterization of the class III collagen receptor, a phosphorylated transmembrane glycoprotein expressed in nucleated human cells.* Journal of Biological Chemistry, (1988) **263**: 4193-4120.

33. Tsukita, S., K. Oishi, N. Sato, *et al.*, *ERM family members as molecular linkers between the cell surface glycoprotein CD44 and actin-based cytoskeletons.* Journal of Cell Biology, (1994) **126**: 391-401.

34. Camp, R.L., T.A. Kraus and E. Pure, *Variations in the cytoskeletal interaction and posttranslational modification of the CD44 homing receptor in macrophages.* Journal of Cell Biology, (1991) **115**: 1283-92.

35. Taher, T.E., L. Smit, A.W. Griffioen, *et al.*, *Signaling through CD44 is mediated by tyrosine kinases. Association with $p56^{lck}$ in T lymphocytes.* Journal of Biological Chemistry, (1996) **271**: 2863-7.

36. Goodfellow, P.N., G. Banting, M.V. Wiles, *et al., The gene, MIC4, which controls expression of the antigen defined by monoclonal antibody F10.44.2, is on human chromosome 11.* European Journal of Immunology , (1982) **12**: 659-663.

37. Peach, R.J., D. Hollenbaugh, I. Stamenkovic, *et al., Identification of hyaluronic acid binding sites in the extracellular domain of CD44.* Journal of Cell Biology, (1993) **122**: 257-64.

38. Matsumura, Y., M. Sugiyama, S. Matsumura, *et al., Unusual retention of introns in CD44 gene transcripts in bladder cancer provides new diagnostic and clinical oncological opportunities.* Journal of Pathology, (1995) **177**: 11-20.

39. Dougherty, G.J., P.M. Landorp, D.L. Cooper, *et al., Molecular cloning of CD44R1 and CD44R2, two novel isoforms of the human CD44 lymphocyte "homing" receptor expressed by hemopoietic cells.* Journal of Experimental Medicine, (1991) **174**: 1-5.

40. Screaton, G.R., M.V. Bell, J.I. Bell, *et al., The identification of a new alternative exon with highly restricted tissue expression in transcripts encoding the mouse Pgp-1 (CD44) homing receptor. Comparison of all 10 variable exons between mouse, human, and rat.* Journal of Biological Chemistry, (1993) **268**: 12235-8.

41. Cooper, D.L. and G.J. Dougherty, *To metastasize or not? Selection of CD44 splice sites.* Nature Medicine, (1995) **1**: 635-7.

42. Mackay, C.R., H.J. Terpe, R. Stauder, *et al., Expression and modulation of CD44 variant isoforms in humans.* Journal of Cell Biology, (1994) **124**: 71-82.

43. Tanabe, K.K., L.M. Ellis and H. Saya, *Expression of CD44R1 adhesion molecule in colon carcinomas and metastases.* Lancet, (1993) **341**: 725-6.

44. Matsumura, Y. and D. Tarin, *Significance of CD44 gene products for cancer diagnosis and disease evaluation [see comments].* Lancet, (1992) **340**: 1053-8.

45. Stamenkovic, I., A. Aruffo, M. Amiot, *et al., The hematopoietic and epithelial forms of CD44 are distinct polypeptides with different adhesion potentials for hyaluronate-bearing cells.* EMBO Journal, (1991) **10**: 343-8.

46. Miyake, K. and P.W. Kincade, *A new cell adhesion mechanism involving hyaluronate and CD44.* Current Topics in Microbiology & Immunology 1990;166:87-90, (1990) .

47. Hyman, R., J. Lesley and R. Schulte, *Somatic cell mutants distinguish CD44 expression and hyaluronic acid binding.* Immunogenetics, (1991) **33**: 392-5.

48. Dougherty, G.J., D.L. Cooper, J.F. Memory, *et al., Ligand binding specificity of alternatively spliced CD44 isoforms. Recognition and binding of hyaluronan by CD44R1.* Journal of Biological Chemistry, (1994) **269**: 9074-8.

49. Brimacombe, J.S. and J.M. Webber, *Mucopolysaccharides.* (1964), Amsterdam: Elsevier Science Publishers.

50. Laurent, T.C. and J.R. Fraser, *Hyaluronan. [Review].* FASEB Journal, (1992) **6**: 2397-404.

51. Whiteside, T.L. and R.B. Buckingham. *Interactions between cells of the immune system and hyaluronate synthesis by human dermal fibroblasts.* in *The Biology of Hyaluronan.* 1989. Chichester: Wiley.

52. Roden, L., P. Campbell, J.R.E. Fraser, *et al. Enzyme pathways of hyaluronan catabolism.* in *The Biology of Hyaluronan.* 1989. Chichester: Wiley.

53. Culty, M., H.A. Nguyen and C.B. Underhill, *The hyaluronan receptor (CD44) participates in the uptake and degradation of hyaluronan.* Journal of Cell Biology, (1992) **116**: 1055-62.

54. Lesley, J. and R. Hyman, *CD44 can be activated to function as an hyaluronic acid receptor in normal murine T cells.* European Journal of Immunology, (1992) **22**: 2719-23.

55. Lesley, J., Q. He, K. Miyake, *et al., Requirements for hyaluronic acid binding by CD44: a role for the cytoplasmic domain and activation by antibody.* Journal of Experimental Medicine, (1992) **175**: 257-66.

56. Lesley, J., R. Hyman and P.W. Kincade, *CD44 and its interaction with extracellular matrix. [Review].* Advances in Immunology, (1993) **54**: 271-335.

57. Lesley, J., N. English, A. Perschl, *et al., Variant cell lines selected for alterations in the function of the hyaluronan receptor CD44 show differences in glycosylation.* Journal of Experimental Medicine, (1995) **182**: 431-7.

58. Katoh, S., J.B. McCarthy and P.W. Kincade, *Characterization of soluble CD44 in the circulation of mice. Levels are affected by immune activity and tumor growth.* Journal of Immunology, (1994) **153**: 3440-9.

59. Takahashi, K., I. Stamenkovic, M. Cutler, *et al., Keratan sulfate modification of CD44 modulates adhesion to hyaluronate.* Journal of Biological Chemistry, (1996) **271**: 9490-9496.

60. He, Q., J. Lesley, R. Hyman, *et al., Molecular isoforms of murine CD44 and evidence that the membrane proximal domain is not critical for hyaluronate recognition.* Journal of Cell Biology, (1992) **119**: 1711-9.

61. Murakami, S., Y. Shimabukuro, Y. Miki, *et al., Inducible binding of human lymphocytes to hyaluronate via CD44 does not require cytoskeleton association but does require new protein synthesis.* Journal of Immunology, (1994) **152**: 467-77.

62. Bennett, K.L., D.G. Jackson, J.C. Simon, *et al., CD44 isoforms containing exon v3 are responsible for the presentation of heparin-binding growth factor.* Journal of Cell Biology, (1995) **128**: 687-98.

63. Jackson, D.G., J.I. Bell, R. Dickinson, *et al., Proteoglycan forms of the lymphocyte homing receptor CD44 are alternatively spliced variants containing the v3 exon.* Journal of Cell Biology, (1995) **128**: 673-85.

64. Tanaka, Y., D.H. Adams, S. Hubscher, *et al., T-cell adhesion induced by proteoglycan-immobilized cytokine MIP-1 beta [see comments].* Nature, (1993) **361**: 79-82.

65. Weber, G.F., S. Ashkar, M.J. Glimcher, *et al., Receptor-ligand interaction between CD44 and osteopontin (Eta-1).* Science, (1996) **271**: 509-12.

66. Jalkanen, S. and M. Jalkanen, *Lymphocyte CD44 binds the COOH-terminal heparin-binding domain of fibronectin.* Journal of Cell Biology, (1992) **116**: 817-25.

67. Aruffo, A., I. Stamenkovic, M. Melnick, *et al., CD44 is the principal cell surface receptor for hyaluronate.* Cell, (1990) **61**: 1303-13.

68. Sy, M.S., Y.J. Guo and I. Stamenkovic, *Distinct effects of two CD44 isoforms on tumor growth in vivo.* Journal of Experimental Medicine, (1991) **174**: 859-66.

69. Toyama-Sorimachi, N. and M. Miyasaka, *A novel ligand for CD44 is sulfated proteoglycan.* International Immunology, (1994) **6**: 655-60.

70. Toyama-Sorimachi, N., H. Sorimachi, Y. Tobita, *et al., A novel ligand for CD44 is serglycin, a hematopoietic cell lineage-specific proteoglycan. Possible involvement in lymphoid cell adherence and activation.* Journal of Biological Chemistry, (1995) **270**: 7437-44.

71. Naujokas, M.F., M. Morin, M.S. Anderson, *et al., The chondroitin sulfate form of invariant chain can enhance stimulation of T cell responses through interaction with CD44.* Cell, (1993) **74**: 257-68.

72. Droll, A., S.T. Dougherty, R.K. Chiu, *et al.*, *Adhesive interactions between alternatively spliced CD44 isoforms.* Journal of Biological Chemistry, (1995) **270**: 11567-73.

73. Faassen, A.E., D.L. Mooradian, R.T. Tranquillo, *et al.*, *Cell surface CD44-related chondroitin sulfate proteoglycan is required for transforming growth factor-beta-stimulated mouse melanoma cell motility and invasive behavior on type I collagen.* Journal of Cell Science, (1993) **105** : 501-11.

74. Knutson, J.R., J. Iida, G.B. Fields, *et al.*, *CD44/Chondroitin Sulfate Proteoglycan and α2β1 integrin mediate human melanoma cell migration of type IV collagen and invasion of basement membranes.* Molecular Biology of the Cell, (1996) **7**: 383-396.

75. Rall, C.J. and A.K. Rustgi, *CD44 isoform expression in primary and metastatic pancreatic adenocarcinoma.* Cancer Research, (1995) **55**: 1831-5.

76. Gansauge, F., S. Gansauge, A. Zobywalski, *et al.*, *Differential expression of CD44 splice variants in human pancreatic adenocarcinoma and in normal pancreas.* Cancer Research, (1995) **55**: 5499-503.

77. Ermak, G., G. Gerasimov, K. Troshina, *et al.*, *Deregulated alternative splicing of CD44 messenger RNA transcripts in neoplastic and nonneoplastic lesions of the human thyroid.* Cancer Research, (1995) **55**: 4594-8.

78. Dall, P., K.H. Heider, H.P. Sinn, *et al.*, *Comparison of immunohistochemistry and RT-PCR for detection of CD44v-expression, a new prognostic factor in human breast cancer.* International Journal of Cancer, (1995) **60**: 471-7.

79. Mulder, J.W., P.M. Kruyt, M. Sewnath, *et al.*, *Colorectal cancer prognosis and expression of exon-v6-containing CD44 proteins [see comments].* Lancet, (1994) **344**: 1470-2.

80. Heider, K.H., J. Dammrich, P. Skroch-Angel, *et al.*, *Differential expression of CD44 splice variants in intestinal- and diffuse-type human gastric carcinomas and normal gastric mucosa.* Cancer Research, (1993) **53**: 4197-203.

81. Imazeki, F., O. Yokosuka, T. Yamaguchi, *et al.*, *Expression of variant CD44-messenger RNA in colorectal adenocarcinomas and adenomatous polyps in humans.* Gastroenterology, (1996) **110**: 362-368.

82. Dall, P., K.H. Heider, A. Hekele, *et al.*, *Surface protein expression and messenger RNA-splicing analysis of CD44 in uterine cervical cancer and normal cervical epithelium.* Cancer Research, (1994) **54**: 3337-41.

83. Kan, M., M. Aki, K. Akiyama, *et al.*, *High-level expression of the CD44 variant sharing exon v10 in renal cancer.* Japanese Journal of Cancer Research, (1995) **86**: 847-53.

84. Kanayama, H., M. Kan, M. Aki, *et al.*, *[Expression of CD44 variant form in human renal cell carcinoma]. [Review] [Japanese].* Nippon Rinsho - Japanese Journal of Clinical Medicine, (1995) **53**: 1710-5.

85. Fox, S.B., J. Fawcett, D.G. Jackson, *et al.*, *Normal human tissues, in addition to some tumors, express multiple different CD44 isoforms.* Cancer Research, (1994) **54**: 4539-46.

86. Manten-Horst, E., E.H. Danen, L. Smit, *et al.*, *Expression of CD44 splice variants in human cutaneous melanoma and melanoma cell lines is related to tumor progression and metastatic potential.* International Journal of Cancer, (1995) **64**: 182-8.

87. Ghaffari, S., G.J. Dougherty, P.M. Lansdorp, *et al.*, *Differentiation-associated changes in CD44 isoform expression during normal hematopoiesis and their alteration in chronic myeloid leukemia.* Blood, (1995) **86**: 2976-85.

88. Koopman, G., A.W. Griffioen, H. Ponta, *et al.*, *CD44 splice variants; expression on lymphocytes and in neoplasia. [Review].* Research in Immunology, (1993) **144**: 750-4; discussion 754-62.

89. Okada, H., J. Yoshida, M. Sokabe, *et al.*, *Suppression of CD44 expression decreases migration and invasion of human glioma cells.* International Journal of Cancer, (1996) **66**: 255-260.

90. Seiter, S., R. Arch, S. Reber, *et al.*, *Prevention of tumor metastasis formation by anti-variant CD44.* Journal of Experimental Medicine, (1993) **177**: 443-55.

91. Bartolazzi, A., D. Jackson, K. Bennett, *et al.*, *Regulation of growth and dissemination of a human lymphoma by CD44 splice variants.* Journal of Cell Science, (1995) **108** : 1723-33.

92. Faassen, A.E., J.A. Schrager, D.J. Klein, *et al.*, *A cell surface chondroitin sulfate proteoglycan, immunologically related to CD44, is involved in type I collagen-mediated melanoma cell motility and invasion.* Journal of Cell Biology, (1992) **116**: 521-31.

93. Henke, C.A., U. Roongta, D.J. Mickelson, *et al.*, *CD44-related chondroitin sulfate proteoglycan, a cell surface receptor implicated with tumor cell invasion, mediates endothelial cell migration on fibrinogen and invasion into a fibrin matrix.* Journal of Clinical Investigation, (1996) **97**: 2541-2552.

94. Levesque, M.C. and B.F. Haynes, *In vitro culture of human peripheral blood monocytes induces hyaluronan binding and up-regulates monocyte variant CD44 isoform expression.* Journal of Immunology, (1996) **156**: 1557-65.

95. Moll, J., A. Schmidt, H. van der Putten, *et al., Accelerated immune response in transgenic mice expressing rat CD44v4-v7 on T cells.* Journal of Immunology, (1996) **156**: 2085-2094.

96. Rothman, B.L., N. Kennure, K.A. Kelley, *et al., Elevation of intracellular cAMP in human T lymphocytes by an anti-CD44 mAb.* Journal of Immunology, (1993) **151**: 6036-42.

97. Galandrini, R., N. Albi, G. Tripodi, *et al., Antibodies to CD44 trigger effector functions of human T cell clones.* Journal of Immunology, (1993) **150**: 4225-35.

98. Ayroldi, E., L. Cannarile, G. Migliorati, *et al., CD44 (Pgp-1) inhibits CD3 and dexamethasone-induced apoptosis.* Blood, (1995) **86**: 2672-8.

99. Rothman, B.L., M.L. Blue, K.A. Kelley, *et al., Human T cell activation by OKT3 is inhibited by a monoclonal antibody to CD44.* Journal of Immunology, (1991) **147**: 2493-9.

100. Berget, S.M., C. Moore and P.A. Sharp, *Spliced segments at the 5' terminus of adenovirus 2 late mRNA.* Proceedings of the National Academy of Sciences of the United States of America, (1977) **74**: 3171-3175.

101. Sharp, P.A., *Split genes and RNA splicing.* Cell, (1994) **77**: 805-815.

102. Cooper, D.L., G. Dougherty, H.J. Harn, *et al., The complex CD44 transcriptional unit; alternative splicing of three internal exons generates the epithelial form of CD44.* Biochemical & Biophysical Research Communications, (1992) **182**: 569-78.

103. Schwarzbauer, J.E., R.S. Patel, F. D., *et al., Multiple sites of alternative splicing of the rat fibronectin gene transcript.* EMBO, (1987) **6**: 2573-2580.

104. Amara, S.G., V. Jonas, M.G. Rosenfeld, *et al., Alternative RNA processing in Calcitonin gene expression generates mRNAs encoding different polypeptide products.* Nature, (1982) **298**: 240-244.

105. Padgett, R.A., S.F. Hardy and P.A. Sharp, *Splicing of adenovirus RNA in a cell-free transcription system.* Proceedings of the National Academy of Sciences of the United States of America, (1983) **80**: 5230-5234.

106. Konarska, M.M., R.A. Grabowski, R.A. Padgett, *et al., Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors.* Nature, (1985) **313**: 552-557.

107. Reed, R. and T. Maniatis, *Intron sequences involved in lariat formation during pre-mRNA splicing.* Cell, (1985) **41**: 95-105.

108. Moore, M.J., C.C. Query and P.A. Sharp, *Splicing of precursors to mRNA by the spliceosome*. The RNA World, (1993), New York: Cold Spring Harbor Laboratory Press.

109. Zhuang, Y., A.M. Goldstein and A.M. Weiner, *UACUAAC is the preferred branch site for mammalian mRNA splicing*. Proceedings of the National Academy of Science of the USA, (1989) **86**: 2725-2756.

110. Zhuang, Y., L. H. and A.M. Weiner, *The natural 5' splice site of Simian Virus 40 Large T Antigen can be improved by increasing the base complementarity to U1 RNA*. Molecular and Cellular Biology, (1987) **7**: 3018-3020.

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

112. Smith, C.W.J., T.T. Chu and B. Nadal-Ginard, *Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns*. Molecular and Cellular Biology, (1993) **13**: 4939-4952.

113. Hornig, H., M. Aebi and C. Weissmann, *Effect of mutations at the lariat branch acceptor site on β-globin pre-mRNA splicing in vitro*. Nature, (1986) **324**: 589-591.

114. Patterson, b. and C. Guthrie, *A U-rich tract enhances usage of an alternative 3' splice site in yeast*. Cell, (1991) **64**: 181-187.

115. Roscigno, R.F., M. Weiner and M.A. Garcia-Blanco, *A mutational analysis of the polypyrimidine tract of introns*. Journal of Biological Chemistry, (1993) **268**: 11222-11229.

116. Mount, S.M., I. Pettersson, M. Hinterberger, *et al.*, *The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro*. Cell, (1983) **33**: 509-518.

117. Hwang, D.Y. and J.B. Cohen, *U1 snRNA promotes the selection of nearby 5' splice sites by U6 snRNA in mammalian cells*. Genes and Development, (1996) **10**: 338-350.

118. Black, D.L., B. Chabot and J.A. Steitz, *U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing*. Cell, (1985) **42**: 737-750.

119. Ruskin, B., P.D. Zamore and M.R. Green, *A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly*. Cell, (1988) **52**: 207-219.

120. Zamore, P.D. and M.R. Green, *Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor*. Proceedings of the National Academy of Science, USA, (1989) **86**: 9243-9247.

121. Zhang, M., P.D. Zamore, M. Carmro-Fonseca, *et al.*, *Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit.* Proceedings of the National Academy of Science, USA, (1992) **89**: 8769-8773.

122. Singh, R., J. Valcarcel and M.R. Green, *Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins.* Science, (1995) : 1173-1176.

123. Hoffman, B. and P.J. Grabowski, *U1snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon.* Genes and Development, (1992) **6**: 2554-2568.

124. Gozani, O., R. Feld and R. Reed, *Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A.* Genes and Development, (1996) **10**: 233-243.

125. Zahler, A.M. and M.B. Roth, *Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites.* Proceedings of the National Academy of Science, USA, (1995) **92**: 2642-2646.

126. Wells, S.E., M. Neville, M. Haynes, *et al.*, *CUS1, a suppressor of old-sensitive U2 snRNA mutations, is a novel yeast splicing factor homologous to human SAP 145.* Genes and Development, (1996) **10**: 220-232.

127. Newman, A.J. and C. Norman, *U5 snRNA interacts with exon sequences at 5' and 3' splice sites.* Cell, (1992) **68**: 743-754.

128. Brigmann, P., B. Appel, J. Rinke, *et al.*, *Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing.* EMBO , (1984) **3**: 1357-1363.

129. Utans, U., S.-E. Behrens, R. Luhrmann, *et al.*, *A splicing factor that is inactivated during in vivo heat shock is functionally equivalent to the [U4/U6.U5] triple snRNP-specific proteins.* Genes and Development, (1992) **6**: 631-641.

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

131. Yean, S.-L. and R.-J. Lin, *U4 small nuclear RNA dissociates from a yeast spliceosome and does not participate in the subsequent splicing reaction.* Molecular and Cellular Biology, (1991) **11**: 5571-5577.

132. Shannon, K.W. and C. Guthrie, *Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA binding motifs.* Genes and Development, (1991) **5**: 773-785.

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

134. Sun, Q. and J. Manley, *A novel U2-U6 snRNA structure is necessary for mammalial mRNA splicing.* Genes and Development, (1995) **9**: 843-854.

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

136. McPheeters, D.S. and J. Abelson, *Mutational analysis of the yeast U2 snRNA suggests a structural similarity to the catalytic core of Group I introns.* Cell, (1992) **71**: 819-831.

137. Crispino, J.D. and P.A. Sharp, *A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing.* Genes and Development, (1995) **9**: 2314-2323.

138. Konforti, B.B. and M.M. Konarska, *U4/U5/U6 snRNP recognizes the 5' splice site in the absence of U2 snRNP.* Genes and Development, (1994) **8**: 1962-1973.

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

140. Hawkins, J.D., *A survey on intron and exon lengths.* Nucleic Acids Research, (1988) **16**: 9893-9908.

141. Ohno, M., H. Sakamoto and Y. Shimura, *Preferential excision of the 5' proximal intron from mRNA precursors with two introns as mediated by the cap structure.* Proceedings of the National Academy of Science of the USA, (1987) **84**: 5187-5191.

142. Konarska, M.M., R.A. Padgett and P.A. Sharp, *Recognition of cap structure in splicing in vitro of mRNA precursors.* Cell, (1984) .

143. Ache, M.P., P. Griffin, W. James, *et al.*, *Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site.* Genes and Development, (1995) **9**: 3008-3025.

144. Lou, H., G. R.F. and S.M. Berget, *An intron enhancer recognized by splicing factors activates polyadenylation.* Genes and Development, (1996) **10**: 208-219.

145. Wassarman, K.M. and J.A. Steitz, *Association with terminal exons in pre-mRNAs: a new role for the U1 snRNP?* Genes and Development, (1993) **7**: 647-659.

146. Lutz, C.S., K.G.K. Murthy, N. Schek, *et al.*, *Interaction between the U1 snRNP-A protein and the 160-kD subunit of Cleavage-Polyadenylation Specificity Factor increases polyadenylation efficiency in vitro.* Genes and Development, (1996) **10**: 325-337.

147. Norton, P.A., *Alternative pre-mRNA splicing: Factors involved in splice site selection.* Journal of Cell Science, (1994) **107**: 1-7.

148. Peterson, M.L., *Regulated immunoglobulin (Ig) RNA processing does not require specific cis-acting sequences: non-Ig RNA can be alternatively processed in B cells and plasma cells.* Molecular and Cellular Biology, (1994) **14**: 7891-7898.

149. Dominiski, Z. and R. Kole, *Cooperation of pre-mRNA sequence elements in splice site selection.* Molecular and Cellular Biology, (1992) **12**: 2108-2114.

150. Liu, X. and J.E. Mertz, *Polyadenylation site selection connot occur in vivo after excision of the 3'-terminal intron.* Nucleic Acids Research, (1993) **21**: 5256-5263.

151. Eperon, E.C., D.C. Ireland, R.A. Smith, *et al.*, *Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF.* EMBO, (1993) **12**: 3607-3617.

152. Parker, R., P.G. Sillciano and C. Guthrie, *Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA.* Cell, (1987) **49**: 229-239.

153. Goux-pelletan, M., D. Libri, Y. d'Aubenton-Carafa, *et al.*, *In vitro splicing of mutually exclusive exons from the chichen β-tropomyosin gene: Role of the branch point location and very long pyrimidine stretch.* EMBO, (1990) **9**: 241-249.

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

155. Gattoni, R., P. Schmidt and J. Stevenin, *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 Research, (1988) **16**: 2389-2409.

156. Norton, P.A. and R.O. Hynes, *In vitro splicing of Fibronectin pre-mRNAs.* Nucleic Acids Research, (1990) **18**: 4089-4097.

157. Query, C.C., S.A. Strobel and P.A. Sharp, *Three recognition events at the branch-site adenine.* EMBO, (1996) .

158. Hartmuth, K. and A. Barta, *Unusual branch point selection of processing of human growth hormone pre-mRNA.* Molecular and Cellular Biology, (1988) **8**: 2011-2020.

159. Sebillon, P., C. Beldjord, J.C. Kaplan, *et al.*, *A T to G mutation in the polypyrimidine tract of the second intron of the human β-globin gene reduces in vitro splicing efficiency: evidence for an increased hnRNP C interaction.* Nucleic Acids Research, (1995) **23**: 3419-3425.

160. Norton, P.A., *Polypyrimidine tract sequences direct selection of alternative branch sites and influence protein binding.* Nucleic Acids Research, (1994) **22**: 3854-3860.

161. Mullen, M.P., C.W.J. Smith, P. J.G., *et al.*, *α-Tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine tracts determines default exon choice.* Genes and Development, (1991) **5**: 642-655.

162. Smith, C.W.J. and B. Nadal-Ginard, *Mutually exclusive splicing of α-Tropomyosin exons enforced by an unusual lariat branch point location: Implications for constitutive splicing.* Cell, (1988) **56**: 749-758.

163. Emerson, R.B., F. Hedjran, R.M. Yeakley, *et al.*, *Alternative production of calcitonin and CGRP mRNA is regulated at the calcitonin-specific splice acceptor.* Nature, (1989) **341**: 76-80.

164. Lou, H., Y. Yang, G.J. Cote, *et al.*, *An intron enhancer containing a 5' splice site sequence in the human calcitonin/calcitonin gene-related peptide gene.* Molecular and Cellular Biology, (1995) **15**: 7135-7142.

165. Guo, W., G.J. Mulligan, S. Wormsley, *et al.*, *Alternative splicing of β-tropomyosin and cellular factors that block the use of a skeletal muscle exon in nonmuscle cells.* Genes and Development, (1991) **5**: 2096-2107.

166. Dirksen, W.P., R.K. Hampson, Q. Sun, *et al.*, *A purine-rich exon sequence enhances alternative splicing of bovine growth hormone pre-mRNA.* Journal of Biological Chemistry, (1994) **269**: 6431-6436.

167. Hampson, R.K., L. La Follette and F.M. Rottman, *Alternative processing of bovine growth hormone mRNA is influenced by downstream exon sequences.* Molecular and Cellular Biology, (1989) **9**: 1604-1610.

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

169. Xu, R., J. Teng and T.A. Cooper, *The Cardiac Troponin T alternative exon contains a novel purine-rich positive splicing element.* Molecular and Cellular Biology, (1993) **13**: 3660-3674.

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

171. Ramchatesingh, J., A.M. Zahler, K.M. Neugebauer, *et al., A subset of SR proteins activates splicing of the Cardiac Troponin T alternative exon by direct interactions with an exonic enhancer.* Molecular and Cellular Biology, (1995) **15**: 4989-4907.

172. Yeakley, J.M., F. Hedjran, J. Morfin, *et al., Control of Calcitonin/Calcitonin Gene-Related Peptide pre-mRNA processing by constitutive intron and exon elements.* Molecular and Cellular Biology, (1993) **13**: 5999-6011.

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

174. Lavigueur, A., H. La Branche, A.R. Kornblihtt, *et al., A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding.* Genes and Development, (1993) **7**: 2405-2417.

175. Balvay, L., D. Libri, M. Gallego, *et al., Intronic sequence with both negative and positive effects on the regulation of alternative transcripts of the chicken β-tropomyosin transcripts.* Nucleic Acids Research, (1992) **20**: 3987-3992.

176. Huh, G.S. and R.O. Hynes, *Elements regulating an alternatively spliced exon of the rat fibronectin gene.* Molecular and Cellular Biology, (1993) **13**: 5031-5314.

177. Del Gatto, F. and R. Breathnach, *Exon and intron sequences, respectively, repress and activate splicing of a fibroblast growth factor receptor 2 alternative exon.* Molecular and Cellular Biology, (1995) **15**: 4825-4834.

178. Chan, R. and D.L. Black, *Conserved intron elements repress splicing of a nueron-specific c-src exon in vitro.* Molecular and Cellular Biology, (1995) **15**: 6377-6385.

179. Hedley, M.L. and T. Amaniatic, *Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that bind specifically to tra-2 protein in vitro.* Cell, (1991) **65**: 579-586.

180. Siebel, C.W., R. Kanaar and D.C. Rio, *Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI.* Genes and Development, (1994) **8**: 1713-1725.

181. Birney, E., S. Kumar and A.R. Krainer, *Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors.* Nucleic Acids Research, (1993) **21**: 5803-5816.

182. Zahler, A.M., W.S. Lane, J.A. Stolk, *et al.*, *SR proteins: a conserved family of pre-mRNA splicing factors.* Genes and Development, (1992) **6**: 837-847.

183. Choi, Y.D., P.J. Grabowski, P.A. Sharp, *et al.*, *Heterogeneous nuclear ribonucleoproteins: role in RNA splicing.* Science, (1986) **231**: 1534-1539.

184. Swanson, M.S. and G. Dreyfuss, *RNA binding specificty of hnRNP proteins: a subset bind to the 3' end of introns.* EMBO, (1988) **7**: 3519-3529.

185. Portman, D.S. and G. Dreyfuss, *RNA annealing activities in HeLa nuclei.* EMBO, (1994) **13**: 213-221.

186. Buvoli, M., F. Cobianichi, M.G. Bestagno, *et al.*, *Alternative splicing in the human gene for the core protein A1 generates another hnRNP protein.* EMBO, (1990) **9**: 1229-1235.

187. Mayeda, A., S.H. Munroe, J.F. Caceres, *et al.*, *Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins.* EMBO, (1994) **13**: 5483-5495.

188. Mayeda, A. and A.R. Krainer, *Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2.* Cell, (1992) **68**: 365-375.

189. Burd, C.G. and G. Dreyfuss, *RNA binding specificitiy of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing.* EMBO, (1994) **13**: 1197-1204.

190. Screaton, G.R., J.F. Caceres, A. Mayeda, *et al.*, *Identification and characterization of three members of the human SR family of pre-mRNA splicing factors.* EMBO Journal, (1995) **14**: 4336-49.

191. Tacke, R., A. Boned and C. Goridis, *ASF alternative transcripts are highly conserved between mouse and man.* Nucleic Acids Research, (1992) **20**: 5482.

192. Peng, X. and S.M. Mount, *Genetic enhancement of RNA-processing defects by a dominant mutation of B52, the Drosophila gene for an SR protein splicing factor.* Molecular and cellular biology, (1995) **15**: 6273-6282.

193. Zuo, H.G., Z. P. and J.L. Manley, *Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators.* Cell, (1991) **66**: 373-382.

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

195. Staknis, S. and R. Reed, *SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a*

general splicing enhancer complex. Molecular and Cellular Biology, (1994) **14**: 7670-7682.

196. Staffa, A. and A. Cochrane, *Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of Human Immunodeficiency Virus Type I.* Molecular and Cellular Biology, (1995) : 4597-4605.

197. Zuo, P. and J.L. Manley, *The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5' splice sites.* Proceedings of the National Academy of Science, USA, (1994) **91**: 3363-3367.

198. Jamison, S.F., Z. Pasman, J. Wang, *et al., U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements.* Nucleic Acids Research, (1995) **23**: 3260-3267.

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

200. Heinrichs, V., M. Bach, G. Winkelmann, *et al., U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site.* Science, (1990) **247**: 69-72.

201. Ge, H., P. Zuo and J.L. Manley, *Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators.* Cell, (1991) **66**: 373-382.

202. Caceres, J.F., S. Stamm, D.M. Helfman, *et al., Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors.* Science, (1994) **265**: 1706-1709.

203. Fu, X.-D. and T. Maniatis, *The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site.* Proceedings of the National Acadamy of Science, USA, (1992) **89**: 1725-1729.

204. Crispino, D., B.J. Blencowe and P.A. Sharp, *Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP.* Science, (1994) **265**: 1866-1869.

205. Tarn, W.-Y. and J.A. Steitz, *SR proteins can compensate for the loss of U1 snRNP functions in vitro.* Genes and Development, (1994) **8**: 2704-2717.

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

207. Tian, M. and T. Maniatis, *A splicing enhancer exhibits both constitutive and regulated activities.* Genes and Development, (1994) **8**: 1703-1712.

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

209. Baier-Bietterlich, G., F. Uberall, B. Auer, *et al., Protein kinase C-θ isoenzyme selective stimulation of the transcription factor complex AP-1 in T lymphocytes.* Molecular and Cellular Biology, (1996) **16**: 1842-1850.

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

211. Sun, Q., A. Mayeda, R.K. Hampson, *et al., General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer.* Genes and Development, (1993) **7**: 2598-2608.

212. Tacke, R. and J.L. Manley, *The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities.* EMBO, (1995) **14**: 3540-3551.

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

214. Furdon, P.J. and R. Kole, *The length of the downstream exon and the substitution of specific sequences affect pre-mRNA splicing in vitro.* Molecular and Cellular Biology, (1988) **8**: 860-866.

215. Min, H., C. R.C. and S.L. Black, *The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event.* Genes and Development, (1995) **9**: 2659-2671.

216. Nandabalan, K. and G.S. Roeder, *Binding of a cell-type-specific RNA splicing factor to its target regulatory sequence.* Molecular and Cellular Biology, (1995) **15**: 1953-1960.

217. Pinol-Roma, S. and G. Dreyfuss, *Cell cycle-regulated phosphorylation of the pre-mRNA-binding (heterogeneous nuclear ribonucleoprotein) C proteins.* Molecular and cellular biology, (1993) **13**: 5762-5770.

218. Tazi, J., M.-C. Daugeron, G. Cathala, *et al., Adenosine phosphorothioates (ATPαS and ATPτS) differentially affect the two steps of mammalian pre-mRNA splicing.* Journal of Biological Chemistry, (1992) **267**: 4322-4326.

219. Mermoud, J.E., P. Cohen and A.I. Lamond, *Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing.* Nucleic Acids Research, (1992) **20**: 5263-5269.

220. Mayeda, A., A.M. Zahler, A.R. Krainer, *et al.*, *Two members of a conserved family of nuclear phosphoproteins are involved in pre-mRNA splicing.* Proceedings of the National Academy of Science, USA, (1992) **89**: 1301-1304.

221. Colwill, K., T. Pawson, B. Andrews, *et al.*, *The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution.* EMBO, (1996) **15**: 265-275.

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

223. Kohtz, J.D., S.F. Jamison, C.L. Will, *et al.*, *Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors.* Nature, (1994) **368**: 119-124.

224. Gui, J.F., H. Tronchere, S.D. Chandler, *et al.*, *Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors.* Proceedings of the National Acadamy of Science, (1994) **91**: 10824-10828.

225. Gui, J.F., W.S. Lane and X.-D. Fu, *A serine kinase regulates intracellular localization of splicing factors in the cell cycle.* Nature, (1994) **369**: 678-682.

226. Cobianchi, F., C. Calbio, M. Stoppini, *et al.*, *Phosphorylation of human hnRNP protein A1 abrogates in vitro strand annealing activity.* Nucleic Acids Research, (1993) **21**: 949-955.

227. Buvoli, M., F. Cobianchi and S. Riva, *Interation of hnRNP A1 with snRNP and pre-mRNAs: evidence for a possible role of A1 RNA annealing activity in the first steps of spliceosome assembly.* Nucleic Acids Research, (1992) **20**: 5017-5125.

228. Idriss, H., A. Kumar, J.R. Casas-Finet, *et al.*, *Regulation of in vitro nucleic acid strand annealing activity of heterogeneous nuclear ribonucleoprotein protein A1 by reversible phosphorylation.* Biochemistry, (1994) **33**: 11382-11390.

229. Municio, M.M., J. Lozano, P. Sanchez, *et al.*, *Identification of heteogeneous ribonucleoprotein A1 as a novel substrate for Protein Kinase Cζ.* Journal of Biological Chemistry, (1995) **26**: 15884-15891.

230. Mayrand, S.H., P. Dwen and T. Pederson, *Serine/threonine phosphorylation regulates binding of C hnRNP protein to pre-mRNA.* Proceedings of the National Academy of Science, USA, (1993) **90**: 7764-7768.

231. Bosser, R., M. Raura, J. Serratosa, *et al.*, *Phophorylation of rat liver heterogeneous nuclear ribonucleoproteins A2 and C can be modulated by Calmodulin.* Molecular and Cellular Biology, (1995) **15**: 661-670.

232. Swanson, M.S., T.Y. Nakagawa, K. LeVan, *et al.*, *Primary structure of human nuclear ribonucleaoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA , and pre-rRNA-binding proteins.* Molecular and Cellular Biology, (1987) **7**: 1731-1739.

233. Mayrand, S.H., P.A. Fung and T. Pederson, *A discrete 3' region of U6 small nuclear RNA modulates the phosphorylation cycle of the C1 heterogeneous nuclear ribonucleoprotein particle protein.* Molecular and Cellular Biology, (1996) **16**: 1241-1246.

234. Woppmann, A., T. Patschinsky, P. Gringmann, *et al.*, *Characterisation of human and murine snSNP proteins by two-dimensional gel electrophoresis and phosphopeptide analysis of U1-specific 70K protein variants.* Nucleic Acids Research, (1990) **18**: 4427-4438.

235. Tazi, J., U. Kornstadt, R. Rossi, *et al.*, *Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing.* Nature, (1993) **363**: 283-286.

236. MacMillan, A.M., C.C. Query, C.R. Allerson, *et al.*, *Dynamic association of proteins with the pre-mRNA branch region.* Genes and Development, (1994) **8**: 3008-3020.

237. D'Orval, B.C., Y. D'Aubenton Carafa, P. Sirand-Pugnet, *et al.*, *RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts.* Science, (1991) **252**: 1823-1828.

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

239. Deshler, J.O. and J.J. Rossi, *Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron.* Genes and Development, (1991) **5**: 1252-1263.

240. Ausubel, F.M., R. Brent, R.E. Kingston, *et al.*, *Current Protocols in Molecular Biology*. (1992), New York: Greene Publishing Associates andWiley-Interscience.

241. Davis, L.G., M.D. Dibner and J.F. Battey, *Basic Methods in Molecular Biology.* (1986), New York: Elsevier Science Publishing Co., Inc.

242. Gauss, G.H. and M.R. Lieber, *DEAE-dextran enhances electroporation of mammalian cells.* Nucleic Acids Research, (1992) **20**: 6739-6740.

243. Micanovic, R., L.D. Gerber, J. Gerger, *et al.*, *Selectivity of the cleavage/attachment site of phosphatidylinositol-glycan-anchored membrane proteins determined by site-specific mutagenesis at Asp-484 of placental alkaline phosphatase.* Proceedings of the National Academy of Science of the USA, (1990) **87**: 157-161.

244. Weiss, M.J., K. Ray, P.S. Henthorn, *et al.*, *Structure of the human Liver/Bone/Kidney Alkaline Phosphatase gene.* Journal of Biological Chemistry, (1988) **263**: 12002-12010.

245. Murray, N., G.P. Baumgardern, D.J. Burns, *et al.*, *Protein Kinase C isotypes in human erythroleukemia (K562) cell proliferation and differentiation.* Journal of Biological Chemistry, (1993) **268**: 15847-15853.

246. Cheng, T., Y. Wang and W. Dai, *Transcription factor egr-1 is involved in phorbol 12-myristate 13-acetate-induced megakaryocytic differentiation of K562 cells.* Journal of Biological Chemistry., (1994) **269**: 30848-30853.

247. Stein, J., H.D. Volk, C. Liebenthal, *et al.*, *Tumour necrosis factor alpha stimulates the activity of the human cytomegalovirus major immediate early enhancer/promoter in immature monocytic cells.* Journal of General Virology, (1993) **74**: 2333-2338.

248. Balavay, L., D. Libri and M.Y. Fiszman, *Pre-mRNA secondary structure and the regulation of splicing.* BioEssays, (1993) **15**: 165-169.

249. Wheatley, S.C., C.M. Isacke and P.H. Crossley, *Restricted expression of the hyaluronan receptor, CD44, during postimplantation mouse embryogenesis suggests key roles in tissue formation and patterning.* Development, (1993) **119**: 295-306.

250. Geppert, T.D. and P.E. Lipsky, *Association of various T cell-surface molecules with the cytoskeleton. Effect of cross-linking and activation.* Journal of Immunology, (1991) **146**: 3298-305.

251. Haegel, H. and R. Ceredig, *Transcripts encoding mouse CD44 (Pgp-1, Ly-24) antigen: strain variation and induction by mitogen.* European Journal of Immunology, (1991) **21**: 1549-53.

252. Kihn, L., A. Dinwoodie and R.A. Stinson, *High-molecular-weight Alkaline Phosphatase in serum has properties similar to the enzyme in plasma membranes of the liver.* American Journal of Clinical Pathology, (1991) **96**: 470-478.

253. Nakamura, T., K. Makamura and R.A. Stinson, *Release of Alkaline Phosphatase form human osteosarcoma cells by Phosphphatidylinositol Phospholipase C: Effect of Tunicamycin.* Archives of Biochemistry and Biophysics, (1988) **265**: 190-196.

254. Kim, E.E. and H.W. Wyckoff, *Structure of Alkaline Phosphatases.* Clinica Chimica Acta, (1989) **186**: 175-188.

255. Manoil, C., J.J. Mekalanos and J. Beckwith, *Alkaline Phophatase fusions: Sensors of subcellular location.* Journal of Bacteriology, (1990) **172**: 515-518.

256. Cunningham, S.A., A.J. Alse, B.V. Potter, *et al.*, *Influences of separation and adjacent sequences on the use of alternative 5' splice sites.* Journal Of Molecular Biology, (1991) **217**: 265-281.

257. Mattox, W., L. Ryner and B.B. Baker, *Autoregulation and multifunctionality among trans-acting factors that regulate alternative pre-mRNA processing.* Journal of Biological Chemistry, (1992) **267**: 19023-19026.

258. Berget, S.M., *Exon recognition in vertebrate splicing.* Journal of Biological Chemistry, (1995) **270**: 2411-2414.

259. Valcarcel, J., R. Singh, P.D. Zamore, *et al.*, *The protein sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA.* Nature, (1993) **362**: 171-175.

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

261. Cote, G.J., D.T. Stolow, S.M. Berget, *et al.*, *Identifictaion of exon sequences and an exon binding protein involved in alternative RNA splicing of calcitonin/CGRP.* Nucleic Acids Research, (1992) **20**: 2361-2366.

262. Tian, H. and R. Kole, *Selection of novel exon recognition elements from a pool of random sequences.* Molecular and Cellular Biology, (1995) **15**: 6291-6298.

263. Tsukahara, T., C. Casciato and D.M. Helfman, *Alternative splicing of β-tropomyosin pre-mRNA: multiple cis-elements can contribute to the use of the 5'- and 3'-splice sites of the nonmuscle/smooth muscle exon 6.* Nucleic Acids Research, (1994) **232**: 2318-2325.

264. Ermak, G., G. Gerasimov, K. Troshina, *et al.*, *Deregulated alternative splicing of CD44 messenger RNA transcripts in neoplastic and nonneoplastic lesions of the human thyroid.* Cancer Research, (1995) **55**: 4594-4598.

265. Dominski, Z. and R. Kole, *Identification and characterization by antisense oligonucleotides of exon and intron sequences required for splicing.* Molecular and Cellular Biology, (1994) **14**: 7445-7454.

266. Eperon, L.C., I.R. Graham, A.D. Griffiths, *et al.*, *Effects of RNA secondary stucture on alternative splicing of pre-mRNA: Is folding limited to a region behind the transcribing RNA polymerase?* Cell, (1988) **54**: 393-401.

267. Gordge, P.C. and W.J. Ryves, *Inhibitors of Protein Kinase C.* Cellular Signalling, (1994) **6**.

268. Tamaoki, T., *Use and specificity of staurosporine, UCN-01, and calphostin C as protein kinase inhibitors.* Methods in Enzymology, (1991) **201**: 340-347.

269. Chabot, B., S. Bisotto and M. Vincent, *The nuclear matrix phosphoprotein p255 associates with splicing complexes as part of the [U4/U6.U5] tri-snRNP particle.* Nucleic Acids Research, (1995) **23**: 3206-3213.

270. Pype, S., H. Slegers, L. Moens, *et al.*, *Tyrosine phosphorylation of a Mr 38,000 A/B-type hnRNP protein selectively modulates its RNA binding.* Journal of Biological Chemistry, (1994) **50**: 31457-31465.

271. Woppmann, A., C.L. Will, U. Kornstadt, *et al.*, *Identification of an snRNP-associated kinase activity that phosphorylates arginine/serine rich domains typical of splicing factors.* Nucleic Acids Research, (1993) **21**: 2815-2822.

272. Buchner, K., *Protein kinase C in the transduction of signals toward and within the cell nucleus.* European Journal of Biochemistry, (1995) **228**: 211-221.

273. Zehng, B., T.C. Chambers, R.L. Raynor, *et al.*, *Human leukemia K562 cell mutant (K562/OA200) selected for resistance to okadaic acid (protein phosphatase inhibitor) lacks protein kinase C-epsilon, exhibits multidrug resistance phenotype, and expressed drug pump P-gylcoprotein.* Journal of Biological Chemistry, (1994) **269**: 12332-12338.

274. Walker, S.D., N.R. Murray, D.J. Burns, *et al.*, *Protein Kinase C chimeras: catalytic domains of* α *and* β*II Protein Kinase C contain determinants for isotype-specific function.* Proc Natl Acad Sci USA, (1995) **92**: 9156-9160.

275. Steitz, J.A., *Splicing takes a holliday.* Science, (1992) **257**: 888-889.

276. Nakai, K. and H. Sakamoto, *Construction of a novel database containing aberrant splicing mutations of mammalian genes.* Gene, (1994) **14**: 171-177.

277. Shifrin, V.I. and B.G. Neel, *Growth factor-inducible alternative splicing of nontransmembrane Phosphotyrosine Phosphatase PTP-1*β *pre-mRNA.* Journal of Biological Chemistry, (1993) **268**: 25376-25384.

278. Dirksen, W.P., Q. Sun and R. F.M., *Multiple splicing signals control alternative intron retention of bovine growth hormone pre-mRNA.* Journal of Biological Chemistry, (1995) **270**: 5346-5352.

279. Yoshida, K., J. Bolodeoku, S. T., *et al.*, *Abnormal retention of intron 9 in CD44 gene transcripts in human gastrointestinal tumors.* Cancer Research, (1995) **55**: 4237-4277.

280. Matsumura, Y., M. Sugiyama, S. Matsumura, *et al.*, *Unusual retention of introns in CD44 gene transcripts in bladder cancer provides new diagnostic and clinical oncological opportunities.* Journal of Pathology, (1995) **177**: 11-20.

281. Dietz, H.C., R.J. Kendzior, Jr. *Maintenance of an open reading frame as an additional level of scrutiny during splice site selection.* Nature Genetics, (1994) **8**: 183-188.

282. Aoufouchi, S. J. Yelamos, and C. Milstein. *Nonsense mutations inhibit RNA splicing in a cell-free system: Recognition of mutant codon is independent of protein synthesis.* Cell (1996) **85**: 415-422.

283. Naeger, L.K., R.V. Shoborg, Q. Zhao *et al., Nonsense mutations inhibit splicing of MVM RNA in cis when they interrupt the reading frame of either exon of the final spliced product.* Genes and Development, (1992) **6**: 1107-1119.