AD_____

Award Number: DAMD17-96-1-6084

TITLE: Metastatic Regulation of Differential Splicing of CD44

PRINCIPAL INVESTIGATOR: Susan M. Berget, Ph.D.

CONTRACTING ORGANIZATION: Baylor College of Medicine
Houston, Texas 77030

REPORT DATE: August 2001

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

NOTICE

USING GOVERNMENT DRAWINGS, SPECIFICATIONS, OR OTHER DATA INCLUDED IN THIS DOCUMENT FOR ANY PURPOSE OTHER THAN GOVERNMENT PROCUREMENT DOES NOT IN ANY WAY OBLIGATE THE U.S. GOVERNMENT. THE FACT THAT THE GOVERNMENT FORMULATED OR SUPPLIED THE DRAWINGS, SPECIFICATIONS, OR OTHER DATA DOES NOT LICENSE THE HOLDER OR ANY OTHER PERSON OR CORPORATION; OR CONVEY ANY RIGHTS OR PERMISSION TO MANUFACTURE, USE, OR SELL ANY PATENTED INVENTION THAT MAY RELATE TO THEM.

LIMITED RIGHTS LEGEND

Award Number: DAMD17-96-1-6084
Organization: Baylor College of Medicine

Those portions of the technical data contained in this report marked as limited rights data shall not, without the written permission of the above contractor, be (a) released or disclosed outside the government, (b) used by the Government for manufacture or, in the case of computer software documentation, for preparing the same or similar computer software, or (c) used by a party other than the Government, except that the Government may release or disclose technical data to persons outside the Government, or permit the use of technical data by such persons, if (i) such release, disclosure, or use is necessary for emergency repair or overhaul or (ii) is a release or disclosure of technical data (other than detailed manufacturing or process data) to, or use of such data by, a foreign government that is in the interest of the Government and is required for evaluational or informational purposes, provided in either case that such release, disclosure or use is made subject to a prohibition that the person to whom the data is released or disclosed may not further use, release or disclose such data, and the contractor or subcontractor or subcontractor asserting the restriction is notified of such release, disclosure or use. This legend, together with the indications of the portions of this data which are subject to such limitations, shall be included on any reproduction hereof which includes any part of the portions subject to such limitations.

THIS TECHNICAL REPORT HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION.

Norma Olson MM
07/23/01

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2001	Final (15 Jul 96 - 14 Jul 01)

4. TITLE AND SUBTITLE
Metastatic Regulation of Differential Splicing of CD44

5. FUNDING NUMBERS
DAMD17-96-1-6084

6. AUTHOR(S)
Susan M. Berget, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Baylor College of Medicine
Houston, Texas 77030

E-mail: sberget@bcm.tmc.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Report contains color

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution authorized to U.S. Government agencies only (proprietary information, Aug 01). Other requests for this document shall be referred to U.S. Army Medical Research and Materiel Command, 504 Scott Street, Fort Detrick, Maryland 21702-5012.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Alternative mRNA splicing is a major step in gene expression used to produce different forms of a protein for specific cellular purposes. The process is normally carefully regulated to limit production of different mRNA isoforms to only appropriate tissues. Cancer, however, alters splicing regulation and causes the appearance of forms of mRNAs and their protein products not normally present in a cell. The cell surface receptor CD44, is a protein that undergoes extensive alternative processing and whose processing alters in both breast cancer tumors and their metastases. We have been investigating this splicing of several alternative exons within the CD44 gene to understand how the splicing of these exons is regulated and how that splicing alters during tumorigenesis and metastasis. We have observed exon sequences and RNA-binding proteins that are required for recognition and splicing of these exons. The three proteins are the SR splicing factor human Tra 2 the y-box protein YB1, and the ATP-dependent DEAD-box RNA helicase p72. These proteins may be significant for exon recognition because raising the *in vivo* concentration of and one protein enhances CD44 alternative splicing. More importantly the levels of Tra2 and p72 increase in mammary tumors and their metastases making them candidate factors responsible for altering CD44 splicing during tumorigenesis.

14. SUBJECT TERMS
breast cancer

15. NUMBER OF PAGES
46

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

(5) INTRODUCTION

Alternative splicing generates considerable diversity in human gene products. Alterations in alternative splicing are important for development and in maintaining tissue-specific gene expression. Cancer, including breast cancer, has been reported to induce alterations in pre-mRNA splicing to alter the distribution of spliced products in tumors as compared to normal cells. The work funded by this grant has focused on the alterations of alternative splicing that occur during breast cancer through the study of one gene that undergoes extensive alternative processing and whose processing dramatically changes in breast cancer and metastases thereof.

The human CD44 gene undergoes extensive alternative splicing (reviewed in 1). At least 10 alternative exons reside in a block located in the middle of the gene after the fifth constitutive exon. All of these exons have lengths divisible by 3; therefore, inclusion of either a single exon or sets of exons does not alter the reading frame of the resultant protein. Protein coding sequences within the alternative exons add extracellular domains to the protein and alter its interaction with the extracellular matrix. Alternative splicing of CD44 responds to developmental and extracellular stimuli to increase exon inclusion; in addition CD44 alternative splicing increases in many tumors and their metastases. The inclusion of certain CD44 alternative exons has been reported to be of prognostic value in human tumor evaluation. Perhaps most interesting, the inclusion of a single alternative exon was shown to be sufficient to render metastatic potential in a rodent model. These observations made the study of CD44 alternative splicing an ideal one for investigating the interplay between splicing and carcinogenesis.

Inclusion of the fourth, fifth, and sixth alternative exons (referred to as variable exons v4, v5, and v6) has been best correlated with tumor-specific changes in CD44 splicing. We have been studying the mechanism and factors involved in the recognition of exon v4 and v5 with an emphasis on v4 (2-4). We noticed that many of the CD44 variable exons contained either GA or CA rich sequences within them, sequences suggesting the existence of exon enhancer elements within the alternative exons. Our studies have concentrated on identifying the sequence elements dictating CD44 alternative splicing and the factors that bind these sequences. We then used a model system of mammary carcinogenesis to determine if and how these factors might be altered during tumorigenesis.

Our results indicate that there are CA-rich exon enhancers sequences within at least two of the variable exons whose inclusion increases during human breast cancer. These sequences are required for exon recognition. We have identified three factors that bind these sequences – the y-box binding protein YB-1 (3), the SR splicing factor Tra2, and the ATP-dependent RNA helicase p72 (4). The concentrations of at least the first two of these increase during mouse mammary tumorigenesis as does the splicing of the variable exons they bind (2). In a final set of experiments we began asking if transcription factors could play a role in the recognition of CD44 alternative exons, especially those transcription factors that are important for development of breast cancer. In these experiments we have discovered that the alternative splicing of CD44 was altered by steroid hormones and receptors when the transcription of the gene was driven by a steroid-responsive promoter. These results suggest an interesting interplay between

transcription and splicing that would alter splicing under conditions where hormones and their receptors are altering transcription.

BODY

ACE elements. Unlike many alternative exons, CD44 variable exon v4 lacks a recognizable purine-rich enhancer element, Instead the exon is C/A-rich containing C/A-rich blocks that we have termed ACE elements. Despite their common occurrence, few laboratories have delved into the mode of recognition of this class of enhancer. ACE elements have been isolated in iterative selection approaches to identifying common exon enhancers and are a major class of enhancer present in both alternative and constitutively-recognized exons (5). We have shown that mutation of the CD44 exon v4 ACE element significantly depresses exon recognition *in vivo* (3) Thus, this exon represents a natural exon regulated by ACE elements and presents a model system for studying the factors and mechanisms that regulate recognition of this major class of exon element.

YB-1 and CD44 Splicing. Fortuitously we began our study of CD44 splicing just as a local colleague, Tom Cooper, isolated an ACE element using an *in vivo* selection strategy for identification of exon enhancers (5). We immediately thought to ask if the protein that bound the Cooper selected ACE (the y-box binding protein YB-1) would bind to and affect splicing of exon v4. Our experiments indicated that YB-1 enhanced CD44 exons v4 and v5 inclusion using a reporter mini-gene that included exon v4 and v5, along with the natural introns sequences surrounding and between them. Co-expression of YB-1 along with the mini-gene increased v4 inclusion from 25% to over 80%. In

addition, YB-1 bound to CD44 exon v4 sequences and binding was ACE-dependent. These results implicated YB-1 in exon recognition and CD44 alternative splicing. These results were recently published.

Ours was the first report of a non-SR protein binding to an exon enhancer sequence and activating splicing. Other known non-SR proteins that bind exon elements have been demonstrated to be inhibitory for recognition, not stimulatory. This was also the first suggestion that YB-1 could be involved in RNA processing. Studied initially as a transcription factor, the YB-1 protein binds to both single stranded RNA and DNA (reviewed in 6), the preferred promoter sequence bound by YB-1 is CCAAT. It has also been implicated in RNA stability via binding to C/A-rich sequences in the 3'-UTR of affected mRNAs. Our experiments indicated that YB-1 did not alter the stability of RNAs containing CD44 exon v4, not did it affect transcription of our reporter genes although preliminary experiments indicated that it may increase transcription of the endogenous CD44 gene. Our finding of the involvement of YB-1 in CD44 alternative splicing suggests that YB-1 may be an important splicing factor.

Tra2 and CD44 splicing. One other prominent C/A-rich exon enhancer has been well characterized – that occurring in the Drosophila doublesex female-specific alternative exon (reviewed in 7). The C/A-rich enhancer in this exon closely resembles the CD44 exon v4 ACE element that binds YB-1. In collaboration with Bill Mattox at M.D. Anderson in Houston, we asked if human Tra2 (8) might be involved in exon v4 recognition. Indeed co-expression of Tra2α or β with our CD44 mini-gene containing exons v4 and v5 increased inclusion of both exons from 25 to 60%. Although not as powerful an activator as YB-1, these observations suggest that Tra2 proteins might

participate in exon v4 recognition. Co-transfection with both proteins gave high inclusion levels suggesting that they do not compete for exon v4 binding. In fact, Tra2 activated YB-1 binding. When we incubated baculovirus-purified recombinant Tra2α or β and *e. coli*-purified recombinant YB-1 with an RNA containing v4 sequences, the UV cross-linking of YB-1 to RNA was greatly enhanced. The cross-linking of Tra2, in contrast, decreased, suggesting that interaction between the two proteins altered Tra2 association with RNA. Gel shift experiments indicated that both proteins could bind to the RNA simultaneously. Tra2 would also be expected to bind to exon v5 to a characterized enhancer in this exon of sequence GAAGAA, a sequence isolated as a preferred Tra2 binding site by iterative selection experiments (9). Therefore, we propose that both YB-1 and Tra-2 bind to the alternative CD44 exon(s) that we are studying; furthermore, that these two proteins may interact during this process. We are in the process of writing a manuscript detailing the Tra2 results.

We also looked at the effects of other SR proteins on CD44 v4 and v5 splicing. Most tested SR proteins were negative for stimulation in a transfection assay, including ASF/SF2, 9G8, SRp40, and SRp55. Both SRp20 and SRp75 demonstrated modest positive effects on inclusion. SRp20 increased inclusion of both exons from 25 to 40%. SRp75 increased splicing of both exons weakly but had a pronounced effect on inclusion of only one of the two exons, although we do not yet know which one. It should be noted that all of the CD44 variable exons have either GAA or CAA rich enhancers suggesting coordinate inclusion of these exons. In fact, certain experiments have indicated mandatory co-inclusion (10). It is therefore possible that YB-1 and Tra2 are major effectors for CD44 alternative splicing.

Exon v5 has been reported by other laboratories to contain both an enhancer and a silencer element that regulate inclusion using reporter RNAs containing v5 as the only CD44 exon. HnRNP A1 has been reported to be a negative regulator for this exon (11) and, indeed when we co-transfect hnRNP A1 we see increased production of an RNA containing only v4 and a decrease in RNA containing both exons. As described in a later section we have also observed negative inclusion effects in experiments with SC35.

Helicase P72 and CD44 Splicing. In addition to binding to ACE elements in CD44 exon v4, YB-1 was recently observed to bind to another RNA containing C/A-rich sequences, the transcriptional co-activator RNA SRA, and to act as a transcriptional co-activator as a result of this interaction (12). The O'Malley laboratory at Baylor works with SRA (13) and had also observed YB-1 binding to SRA along with an RNA helicase, p72. P72 is a member of the DEAD box family of RNA helicases (14) and is extremely similar to p68 (15-17), a helicase observed as a constituent of an *in vitro* reconstituted spliceosome (18). We decided to test if p72 would have an effect on CD44 alternative splicing. Co-transfection of p72 with our reporter construct demonstrated an enhancement of inclusion of both exons from 25 to 75% (Figure 1). The effect was specific for p72; p68 did not activate CD44 splicing (Figure 1).

P72 is a prototypic DEAD-box helicase with regions that bind RNA and ATP (Figure 2). Its central helicase motifs are over 90% similar to those in p68 and both proteins are considerably more similar to each other than to other helicases (14). In contrast to p68, p72 has N-terminal and C-terminal domains that are more frequently found in hnRNP proteins that play roles in splicing. These domains include multiple RGG repeats thought to be able to bind RNA, and C-terminal SG-rich and poly-proline

regions important for protein-protein interactions. Mutation of the ATPase domain of the helicase depressed the ability of the protein to stimulate CD44 exon v4 splicing (Figure 2). Deletion of the C-terminal region of the protein including the RNA-binding domain and the SG-rich and proline-rich regions also depressed activation. Interestingly, mutation of one of the conserved helicase motifs, a SAT sequence thought to be important for a conformational change necessary for RNA unwinding (19,20), had minimal impact.

The effect of p72 on splicing was exon-specific. We have not observed effects on other types of alternative processing including recognition of alternative 5' splice sites, recognition of an alternative 3'-terminal exon (from the CT/CGRP gene), or recognition of weak internal exons lacking known ACE elements. Increasing p72 concentration *in vivo* could not by-pass the inhibitory effect of mutation of the exon v4 ACE element on exon v4 splicing. Given these observations, we propose that p72 is an exon-specific splicing enhancement factor recruited to ACE-containing exons.

Hormone-dependent splicing. During the last year we have begun a very exciting set of experiments to dissect the interplay between splicing and transcription at the molecular level. We began this endeavor by examining the effect of promoters and co-activators on splicing phenotypes. We were obviously influenced in these experiments by tantalizing experiments in other laboratories indicating that altering promoters altered splicing phenotypes (21,22). We had a more immediate stimulus, however, in our observed binding of the splicing/transcription factors YB-1 and helicase p72 to both CD44 exon sequences and to the steroid hormone transcriptional co-activator, SRA. Although the latter binding could very well be involved in the splicing of the SRA RNA

rather than in its functioning, the serendipity of the observations in our lab about CD44 and the O'Malley lab about SRA caused us to initiate a set of experiments looking at promoter-hormone effects on CD44 splicing.

Because most experiments published in this area have looked at only a minimal number of promoter-gene combinations, we wanted to look at a number of combinations to see if any stimulation of transcription could effect splicing changes or if there was a specificity to the interaction. As shown in Table I, amazing receptor-promoter-target specificity was observed, suggesting an informational rather than generic link between transcription and splicing. The most common co-effect upon addition of hormone and receptor was an increase in transcription and an increase in exon skipping using the CD44 reporter gene. Both effects required both the receptor and the hormone and non-cognate sets of promoter-receptor-hormone were ineffective. Receptors and hormone were unable to affect splicing when transcription was driven from a promoter that does not respond to either regardless of whether the promoter was extremely active (CMV) or weak (HSV). This observation suggests that we are not looking at an indirect effect of receptor or hormone on splicing factors.

We do not think that the observed effect was caused by an increase in RNA precursor levels per se. Some receptor-promoter combinations activated transcription but were ineffective for altering splicing phenotypes. In addition, hormone activated transcription from the receptor responsive promoters was always considerably less than from the CMV promoter and yet there was more exon inclusion with the more active promoter. Perhaps the most persuasive observation for a specific interaction between hormone regulated transcription and exon recognition came from the experiments

comparing the ability of the alpha and beta forms of the estrogen receptor to affect splicing phenotypes. The two forms of the receptor are very similar. Both activated the $(ERE)_2TATA$ promoter for transcription but only the alpha form was capable of altering splicing phenotypes. Although the transcription field has not yet deciphered the cellular transcriptional roles for two forms of this receptor, their differential impact on splicing suggests specificity in the interaction between transcription and splicing.

We also used two additional splicing reporters as part of this study – one containing the alternative exon from the calcitonin/CGRP gene that we use to study alternative polyadenylation (23) and a non-natural construct containing two exons derived from adenovirus in which the first exon has been partially duplicated to create a first exon with two 5' splice sites. Splicing of the CT/CGRP exons responds to increased levels of SRp20, ASF/SF2, U1 snRNPs, and PTB to produce more RNA including the CT 3'-terminal exon; 5' splice site usage in the adenovirus mini-gene responds to altering concentrations of SR proteins with different members of the family activating different 5' splice sites. Splicing/polyadenylation of the CT/CGRP mini-gene did not alter when driven by a steroid-responsive promoter suggesting both specificity in the observed response of CD44 and that we were most likely not looking at interactions between transcription and polyadenylation in our assays. The adenovirus constructs did respond, although to different receptor – hormone combinations that did the CD44 reporter. Like the CD44 reporter construct, there was a splicing shift when transcription was driven by the $(PRE)_2TATA$ promoter in the presence of GR and dexamethasone. The alteration produced more RNA resulting from usage of the promoter-proximal 5' splice site (using standard splicing nomenclature this is called the distal site because it is further away from

the 3' splice site of the affected intron). This same phenotype was observed for the two other combinations that increased transcription, including the combination of $(PRE)_2TATA$ – androgen receptor – R1881 which had no effect on CD44 splicing. Again these results suggest specific interactions rather than generic uncoupled effects.

The results with the adenovirus reporter are hard to put in a cellular context because of the artificiality of the construct. It is possible that the effects being measured reflect generic effects of transcription rate on splice site choice. The effects with CD44, however, were less related to transcription rate than to the presence of correct promoter-receptor-hormone complexes. There are also indications that CD44 gene expression is altered in response to steroid hormones *in vivo* (49-51). Therefore, the alterations in CD44 splicing we are observing may be relevant to the normal biological control of CD44.

It should be noted that both p72 and its related helicase p68 have been documented to be transcriptional co-activators for steroid-responsive promoters and to bind to steroid receptors and the RNA co-activator SRA (12,24). This effect was cell type specific and did not occur in HeLa cells but did occur in MCF-7 cells, a breast cancer cell lines that endogenously expresses the estrogen receptor alpha. We also observed a cell type specificity for estrogen-mediated alterations in splicing but with the opposite specificity observed for p72-mediated transcriptional activation; i.e. effects in HeLa but not MCF-7 cells. Although we don't yet know enough to unambiguously understand what this difference means it does suggest that the two phenotypes may be connected, possible mutually incompatible. It is interesting that both processes occurred only with the alpha form of the estrogen receptor, not the beta. We did ask if the RNA co-activator

that binds p72, SRA, would affect splicing phenotypes and saw no apparent effect in HeLa cells. We used the major form of SRA. Because this RNAs is extensively alternatively spliced we may not have tested the correct form of the RNA. The O'Malley laboratory maintains an interest in this RNA and its binding factors and will continue this portion of the study. If we find future evidence for a role for SRA in alternative splicing, we will initiate studies to investigate it. The absence of an effect, especially a negative effect, however, has temporarily sidelined our interest in SRA.

KEY RESEARCH ACCOMPLISHMENTS

- ✓ Establishment that CD44 alternative splicing changes during mammary tumorigenesis; while all tumors demonstrated high levels of alternative splicing, neoplasias varied.
- ✓ Both tumors and their metastases demonstrated high levels of CD44 alternative splicing indicating no significant alterations upon metastasis.
- ✓ The concentration of the major SR splicing factors increased in tumor cells.
- ✓ Establishment of an *in vivo* transfection system that carries out alternative processing of CD44 variable exons and responds to exogenous factors.
- ✓ Determination that human Tra2α influences CD44 alternative splicing
- ✓ Determination that SRp20 and SRp75 also influence CD44 alternative splicing
- ✓ Determination that purified recombinant Tra2α binds to CD44 exon sequences *in vitro.*
- ✓ Determination that Tra2α is induced during mouse mammary tumorigenesis.

- ✓ Determination that recombinant Tra 2α activates the binding of the y-box protein YB1 to CD44 exon sequences
- ✓ Determination that human YB1 influences CD44 alternative splicing *in vivo*
- ✓ Determination that AC-rich sequences within exon v4 are important for *in vivo* splicing of exon 4, *in vitro* assembly of exon 4 and YB1 binding.
- ✓ Determination that YB1 is induced during mouse mammary tumorigenesis.
- ✓ Determination that the ATP-dependent DEAD-box helicase p72 influences CD44 alternative splicing in vivo.
- ✓ Determination that the ATP and RNA binding domains of p72 are required for activation of CD44 splicing by p72.

REPORTABLE OUTCOMES

Manuscripts:

Stickeler, E., F. Kitrell, D. Medina, and S. M. Berget. 1999. Stage-specific changes in SR spplicing factors and alternative splicing in mammary carcinogenesis. Oncogene 18:3574-3582.

Stickeler, E. S. D. Fraser, A. Honig, A.L. Chen, S. M. Berget, and T.A. Cooper. 2001. The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. EMBO J. 20:3821-3830.

Honig, A., D. Auboeuf, B. W. O'Malley, and S. M. Berget. 2001. Regulation of alternative splicing by the ATP-dependent DEAD box RNA helicase p72. Submitted.

Auboeuf, D., A. Honig, S. M. Berget, and B. W. O'Malley. 2001. Steroid hormones control transcription and alternative splicing in a receptor selective and promoter specific manner. To be submitted.

Stickeler, E.S., A. Honig, A., and S. M. Berget 2001. Regulation of CD44 alternative splicing by the SR proteins Tra2 and SC35. To be submitted.

Presentations (major meeting platform presentations only)

Cold Spring Harbor Meeting on Pre-mRNA Splicing, August, 1997. "Changes in CD44 alternative splicing during tumorigenesis of breast cancer are accompanied by alterations of SR protein expression patterns." Elmar Stickeler.

Cold Spring Harbor Meeting on Pre-mRNA Splicing, August, 1999. "CD44 Alternative Splicing", Elmar Stickeler.

M.D. Anderson Symposium on Cancer Mechanisms, October 2000, "Cancer and RNA Splicing", Susan M. Berget.

Cold Spring Harbor Meeting on Pre-mRNA Splicing, August, 2001. Regulation of alternative splicing by the ATP-dependent DEAD box RNA helicase p72. Arnd Honig.

Abstracts

Stickeler, E., D. Medina, and S. M. Berget. 1997. Changes in CD44 alternative splicing during tumorigenesis of breast cancer are accompanied by alterations of S/R protein expression patterns. Cold Spring Harbor Pre-mRNA Processing Meeting, Cold Spring Harbor New York, August 1997.

Stickeler, E., D. Medina, T. Cooper, W. Mattox, and S. M. Berget. 1999. CD44 alternative splicing. Cold Spring Harbor Pre-mRNA Processing Meeting, Cold Spring Harbor New York, August 1999.

Stickeler, E., A. Honig, A. Chen, D. Medina, T. Cooper, W.W. Mattox, S. Fraser, and S. M. Berget. 2000. CD44 alternative splicing. Era of Hope DOD Breast Cancer Research Program Meeting, June 2000.

Honig A., D. Auboeuf, B. O'Malley, and S.M. Berget. Regulation of alternative splicing by the ATP-dependent DEAD box RNA Helicase p72. Cold Spring Harbor Pre-mRNA Processing Meeting, Cold Spring Harbor New York, August 2001.

Funding Applied For:

NIH Grant, Susan M. Berget, P.I. Processing of pre-mRNA. Submitted 11/1/01.

Employment Opportunities Received Based on Training Supported by this award.

Elmar Stickeler, Assistant Professor, Obstetrics and Gynecology, University of Frieberg, Freiberg, GERMANY.

Arnd Honig, Obstetircs and Gynecology Resident, GERMANY.

Andy Chen, medical student, University of Texas, Galveston, TX.

Personnel Funded by this grant

Susan M. Berget, Ph.D. Principal Investigator

Elmar Stickeler, M.D., post-doctoral fellow.

Arnd Honig, M.D., post-doctoral fellow

Andy Chen, B.S. technician; Rebecca Sierra, B.S. technician

CONCLUSIONS

Our results indicate that alternative splicing of CD44 is a complicated process involving multiple factors that participate in both the slicing and transcription processes. We have identified three factors that increase CD44 alternative splicing irrespective of the promoter; two of these are RNA binding proteins and the third is an RNA helicase. The two RNA binding proteins, the y-box binding protein YB-1 and the SR protein human Tra2, are induced in mammary tumorigenesis, indicating that they may play an important role in the up-regulation of CD44 alternative splicing seen during this process. This is the first report of a non-SR protein acting as an enhancer of splicing by binding to exon regulatory elements. Our discovery that an RNA helicase enhances exon recognition is the first demonstration that this important class of protein is involved in alternative pre-mRNA processing, presumably reflecting the large number of RNA and protein rearrangements that must occur during assembly of the active spliceosome. Finally we have shown that steroid hormones and their receptors can regulate CD44 alternative splicing when transcription is driven by a steroid-responsive promoter. This is an exciting discovery linking transcriptional control to regulation of alternative splicing. Use of the system should permit deciphering the molecular mechanism whereby the two processes interact and are regulated both in normal cells and in disease.

REFERNCES

1. Günthert, U. 1993. CD44: a multitude of isoforms with diverse functions. Curr. Top. Microbiol. Immunol. 184:47-63.
2. Stickeler, E., F. Kittrell, D. Medina, and S.M. Berget. 1999. Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. Oncogene 18:3574-3582.
3. Stickeler, E., S.D. Fraser, A. Honig, A.L. Chen, S.M. Berget and T.A. Cooper. 2001. The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. EMBO J. 20:3821-3830.
4. Honig, A., D. Auboeuf, B.W. O'Malley, and S.M. Berget. 2001. Regulation of alternative splicing by the ATP-dependent DEAD box helicase p72. Submitted.
5. Coulter, L.R., M.A. Landree, and T.A. Cooper. 1997. Identification of a new class of exonic splicing enhancers by in vivo selection. Mol. Cell. Biol. 17:2143-2150.
6. Sommerville, J. and A.P, Wolffe. 1996. Masking of mRNA by Y-box proteins. FASEB J. 10:435-443.1
7. Tacke, R., and J.L Manley. 1999. Functions of SR and Tra2 proteins in pre-mRNA splicing egulation. Proc. Soc. Exp. Med. 220:59-63.
8. Dauwalder, B., F. Amaya-Manzanares, and W. Mattox. 1996. A human homologue of the Drosphila sex determination factor transformer-2 has conserved splicing regulatory functions. Proc. Natl. Acad. Sci. USA 93:9004-9009.
9. Tacke, R., M. Tohyama, S. Ogawa, and J.L. Manley. 1998. Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing.
10. Bell M.V., A.E. Cowper, M.P. Lefranc, J.I. Bell, and G.R. Screaton. 1998. Influence of intron length on alternative splicing of CD44. Mol Cell Biol. 18:5930-5941.

11. Matter, N., M. Marx, S. Weg-Remers, H. Ponta, P. Herrlich, and H. Konig. 2000. Heterogeneous ribonucleoprotein A1 is part of an exon-specific splice-silencing complex controlled by oncogenic signaling pathways. J. Biol. Chem. 275:35353-35360.

12. Watanabe, M., J. Yanagisawa, H. Kitagawa, K. Takeyama, S. Ogawa, Y. Arao, M. Suzuwa, Y. Kobayashi, T. Yano, H. Yoshikawa, Y. Mashuhiro, and S. Kato. 2001. A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor a coactivator through the N-terminal domain (AF-1) with an RNA coactivator, SRA. EMBO J. 20:1341-1352.

13. Lanz, R.B., N.J. McKenna, S.A. Onate, U. Albrecht, J. Wong, S.Y. Tsai, M.J. Tsai, and B.W. O'Malley. 1999. A steroid receptor coactivator SRA functions as an RNA and is present in an SRC-1 complex. Cell 97:17-27.

14. Lamm, G.M. S. M. Nicol, F. V. Fuller-Pace, and A. I. Lammond. 1996. p72: a human nuclear DEAD box protein highly related to p68. Nucleic Acids Res. 24:3739-3747.

15. Hirling, H. M. Scheffner, T. Restele, and H. Stahl. 1989. RNA helicase activity associated with the human p68 protein. Nature 339:562-564.

16. Iggo, R. D., and D. P. Lane. 1989. Nuclear protein p68 is an RNA-dependent ATPase. EMBO J. 8:1827-1831.

17. Iggo, R. D., D. J. Jamieson, S. A. MacNeill, J. Southgate, J. McPheat, and D. P. Lane. 1991. p68 helicase: identification of a nucleolar forma and cloning of related genes containing a conserved intron in yeasts. Mol. Cell. Biol. 11:1326-1333.

18. Neubauer, G., A. King, J. Rappsilber, C. Calvio, M. Watson, P. Ajuh, J. Sleeman, A. Lamond, and M. Mann. 1998. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. Nature Genetics 20:46-50.

19. Ohno, M. and Y. Shimura. 1996. A human RNA helicase-like protein, HRH1, facilitates nuclear export of spliced mRNA by releasing the RNA from the spliceosome. Genes Devel. 10:997-1007.

20. Schwer, B., and T. Meszaros. 2000. RNA helicase dynamics in pre-mRNA splicing. EMBO J. 23:6582-6591.

21. Cramer, P., C.G. Pesce, F.E. Baralle, and A.R. Kornblihtt. 1997. Functional association between promoter structure and transcript alternative splicing. Proc. Natl. Acad. Sci. USA 94:11456-11460.

22. Cramer, P., J.F. Caceres, D. Cazalla, S. Kadener, A.F. Munro, F.E. Baralle, and A.R. Kornblihtt. 1999. RNA polymerase II promoters modulate SF2/ASF and 9G8 affects on an exonic splicing enhancer. Mol. Cell 4:251-258.

23. Lou, H., R. F. Gagel, and S. M. Berget. 1996. An intron enhancer recognized by splicing factors activates polyadenylation. Genes Devel. 10:208-219.

24. Endoh, H., K. Maruyama, Y. Masuhiro, Y. Kobayashi, M. Goto, H. Tai, J. Yanagisawa, D. Metzger, S. Hashimoto, and S. Kato. 1999. Mol. Cell. Biol. 19:5363-5372.

APPENDIX

FIGURE 1. Increasing the *in vivo* concentration of p72 RNA helicase increases inclusion of CD44 variable exons v4 and v5. (A) CD44 alternative splicing: Line 1, The exon structure of the human CD44 gene is depicted. Constitutive exons are in black and alternative exons (termed variable exons) are in gray. Line 2, The employed CD44 mini-gene in which CD44 variable exons v4 and v5 and their surrounding intron sequences have been inserted into a β-globin mini-gene driven from the CMV promoter is depicted. Globin exons are white, CD44 exons are gray. Line 3, The sequence of CD44 exon v4. Exon sequences are capitalized; intron sequences are lower case. The CA-rich "ACE" exon enhancer is underlined. (B) Transfection of p72 or p68. The mini-gene shown in (A) was co-transfected into HeLa cells with 0, 1, 2, or 4 μg of an expression plasmid coding for full length human p72 (lanes 1-4) or human p68 (lanes 5-8). RNA splicing patterns were characterized by RT/PCR amplification of total cell RNA using primers specific for the exons flanking the CD44 variable exons. Bands corresponding to inclusion of no CD44 exons (the skip product), one CD44 variable exon, or both CD44 variable exons are indicated. Identity of each band was confirmed by sequencing of PCR products. Numbers below the gels indicate the percentage of product RNA containing both exons v4 and v5 as determined by scanning of gels in the Phosphoimager.

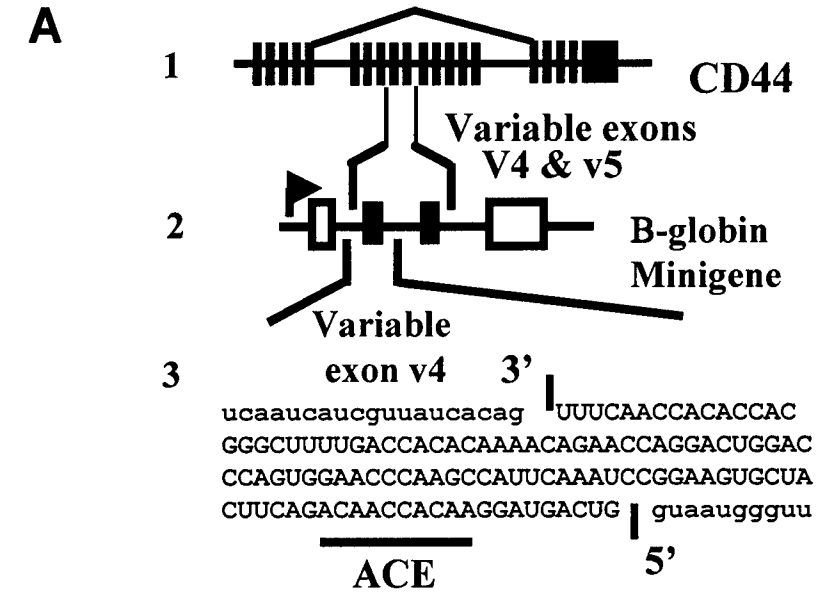
FIGURE 2. Maximum stimulation of CD44 alternative splicing required multiple domains within p72. Four mutants of p72 were prepared and assayed by co-transfection with the CD44 mini-gene. (A) Diagram of the constructed mutations. Domains of p72

are indicated by boxes. The p142 mutation was a lysine to arginine substitution at amino acid 142 within the ATP-binding domain; p277 was a serine to leucine alteration at amino acid 277 in the SAT domain thought to affect protein conformation; p300 was a translational stop mutation introduced at amino acid 300 to produce a protein lacking the RNA binding domain; and p437 was a truncation mutant at amino acid 437 lacking the SG-rich C-terminal tail of the protein. All mutants contained an N-terminal flag tag equivalent to that present in the wild type protein. (B) Western blot of total cell protein following transfection with the indicated p72 genes. Detection was with the anti-flag antibody. (C and D) Quantification of the ability of each mutant protein to affect the inclusion of CD44 variable exons v4 and v5. The percentage of RNA resulting from the inclusion of variable exons v4 and v5 was calculated from Phosphoimager tracing of a gel of RT/PCR amplification of RNA from co-transfections of the CD44 mini-gene and various forms of p72 as in Figure 2. Standard deviations from multiple experiments are indicated.

Table 1. Effect of Steroid Hormones and Receptors on CD44 Splicing

Promoter	Receptor	Hormone	Transcription phenotype	Splicing Phenotype
Hormone regulatable promoters (when cognate receptors present)				
MMTV	progesterone receptor	progesterone	increased	increased skipping
MMTV	none	progesterone	no effect	no effect
MMTV	progesterone receptor	none	no effect	no effect
MMTV	glucocorticoid receptor	dexamethasone	increased	increased skipping
MMTV	androgen receptor	R1881	increased	no effect
MMTV	estrogen receptor α	estradiol (E2)	no effect	no effect
MMTV	estrogen receptor β	estradiol (E2)	no effect	no effect
$(ERE)_2TATA$	estrogen receptor α	estradiol (E2)	Increased	increased skipping
$(ERE)_2TATA$	estrogen receptor β	estradiol (E2)	Increased	no effect
$(PRE)_2TATA$	glucocorticoid receptor	dexamethasone	Increased	increased skipping
$(PRE)_2TATA$	androgen receptor	R1881	Increased	no effect
Promoters not regulatable by hormones				
HSV	glucocorticoid receptor	dexamethasone	no effect	increased skipping
HSV	estrogen receptor α	estradiol (E2)	no effect	no effect
CMV	estrogen receptor α	estradiol (E2)	no effect	no effect
CMV	glucocorticoid receptor	dexamethasone	no effect	no effect

The CD44 reporter gene depicted in Figure 1 was placed downstream of the various promoters described in the table in a eukaryotic expression vector. The reporter mini-genes were co-transfected with the indicated receptors and the transfected cells were treated with the indicated hormone for 24 hours beginning 24 hours post transfection. RNA was harvested at 48 hours and assayed for splicing phenotype as in Figure 1.


A
1
CD44
Variable exons
V4 & v5
2
B-globin
Minigene
Variable
3
exon v4
3'
ucaaucaucguuaucacag
UUUCAACCACACCAC
GGGCUUUUGACCACACAAAACAGAACCAGGACUGGAC
CCAGUGGAACCCAAGCCAUUCAAAUCCGGAAGUGCUA
CUUCAGACAACCACAAGGAUGACUG
guaaugggguu
ACE
5'

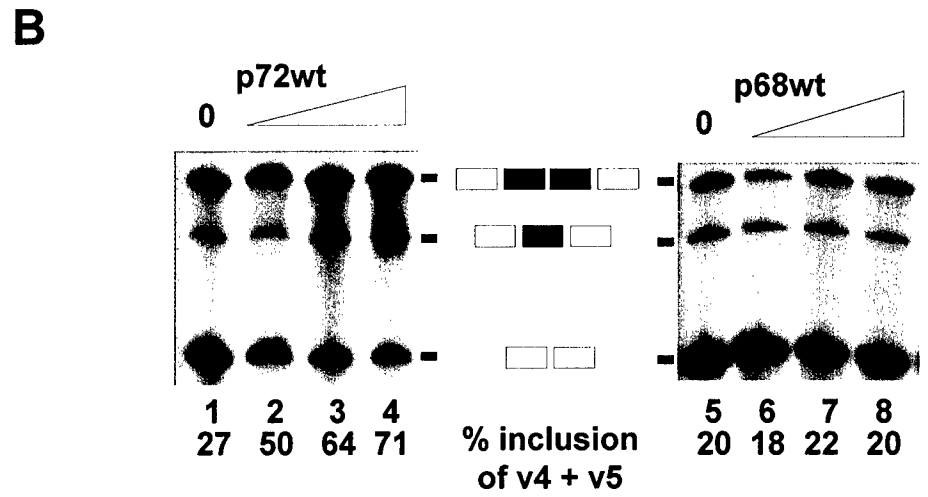

B
p72wt
0
p68wt
0
1 2 3 4
27 50 64 71
% inclusion
of v4 + v5
5 6 7 8
20 18 22 20

Figure 1

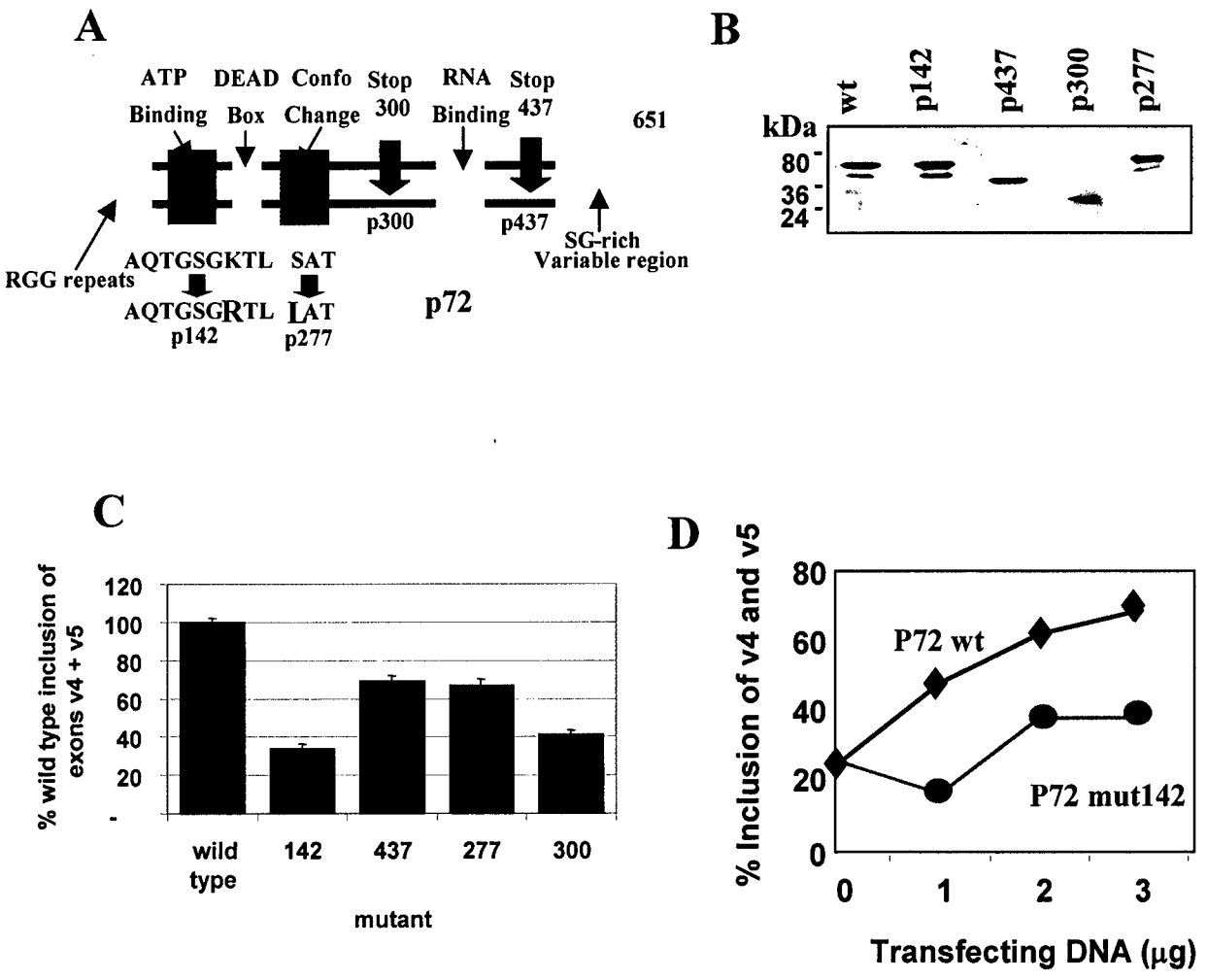

A
ATP Binding
DEAD Box
Confo Change
Stop 300
RNA Binding
Stop 437
651
p300
p437
SG-rich Variable region
RGG repeats
AQTGSGKTL
SAT
AQTGSGRTL
LAT
p142
p277
p72
B
kDa
wt
p142
p437
p300
p277
80
36
24
C
% wild type inclusion of exons v4 + v5
120
100
80
60
40
20
0
wild type
142
437
277
300
mutant
D
% Inclusion of v4 and v5
80
60
40
20
0
P72 wt
P72 mut142
0
1
2
3
Transfecting DNA (µg)

Figure 2

The EMBO Journal Vol. 20 No. 14 pp. 3821–3830, 2001

The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the *CD44* alternative exon v4

Elmar Stickeler[1,2], Sherri D.Fraser[3], Arnd Honig[2], Andy L.Chen[2], Susan M.Berget[2] and Thomas A.Cooper[3,4]

[1]Department of Obstetrics and Gynecology, University of Freiburg, Hugstetterstrasse 55, D-79106 Freiburg, Germany, [2]Department of Biochemistry and Molecular Biology and [3]Departments of Pathology and Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA

[4]Corresponding author
e-mail: tcooper@bcm.tmc.edu

Exon enhancers are accessory pre-mRNA splicing signals that stimulate exon splicing. One class of proteins, the serine-arginine-rich (SR) proteins, have been demonstrated to bind enhancers and activate splicing. Here we report that A/C-rich exon enhancers (ACE elements) are recognized by the human YB-1 protein, a non-SR protein. Sequence-specific binding of YB-1 was observed both to an ACE derived from an *in vivo* iterative selection protocol and to ACE elements in an alternative exon (v4) from the human *CD44* gene. The ACE element that was the predominant YB-1 binding site in *CD44* exon v4 was required for maximal *in vivo* splicing and *in vitro* spliceosome assembly. Expression of wild-type YB-1 increased inclusion of exon v4, whereas a truncated form of YB-1 did not. Stimulation of exon v4 inclusion by wild-type YB-1 required the ACE necessary for YB-1 binding *in vitro*, suggesting that YB-1 stimulated exon inclusion *in vivo* by binding to an exonic ACE element. These observations identify a protein in addition to SR proteins that participates in the recognition of exon enhancers.

Keywords: alternative splicing/*CD44*/exonic splicing enhancer/pre-mRNA/YB-1

Introduction

Pre-mRNA splicing involves the cumulative recognition of multiple short elements by a large number of RNA and protein factors. It has become clear that sequences in addition to splice sites play an important role in discerning exons from introns during the earliest steps of precursor RNA recognition (reviewed in Black, 1995; Reed, 1996; Cooper and Mattox, 1997). Accessory elements located within the recognized exon or flanking introns can either stimulate or depress recognition. Despite the ramifications for coding capacity, a large number of exons, including both alternatively spliced exons and constitutively recognized exons, contain short enhancer sequences that activate both splicing and spliceosome formation.

The splicing factors known to recognize exon enhancers are the arginine-serine-rich (SR) proteins (Tacke and Manley, 1999). This large family of proteins is characterized by the presence of one or more RNA recognition motifs (RRMs) and a *trans*-activation domain characterized by SR dipeptides. When bound to exon enhancers, SR proteins are thought to interact with constitutive splicing factors bound to 3′ and 5′ splice sites, including U2AF and U1 snRNPs, thereby activating the earliest steps in exon recognition. The major members of this family of proteins bind with high affinity to the major class of characterized exon enhancer—the purine-rich enhancers characterized by internal repeats of sequence GAR.

Sequences other than purine-rich repeats have also been characterized as exon enhancers (Dominski and Kole, 1994; Staknis and Reed, 1994; van Oers *et al.*, 1994; Wang *et al.*, 1995; Coulter *et al.*, 1997; Cooper, 1999; Gersappe and Pintel, 1999; Schaal and Maniatis, 1999a,b). Some of these also bind SR proteins. Interestingly, no other class of factor has been implicated as binding exon enhancer sequences other than SR proteins.

In this report we focus on a newly discovered class of exon enhancers—the A/C-rich enhancers, or ACE elements—that have been identified both in natural genes and in experiments designed to select exon enhancer elements (van Oers *et al.*, 1994; Wang *et al.*, 1995; Coulter *et al.*, 1997; Schaal and Maniatis, 1999b). We show that the major nuclear protein binding the ACE element is the single-stranded nucleic acid binding protein YB-1. The element did not bind classic SR proteins nor was the binding of YB-1 stimulated by SR proteins, suggesting that YB-1-mediated exon recognition could be an alternative mechanism of exon enhancer activity.

ACE sequences were observed in a natural alternative exon from the human *CD44* gene. Pre-mRNA from the human *CD44* gene undergoes extensive alternative splicing within a block of at least 10 exons (Günthert, 1993; Mackay *et al.*, 1994). Increased inclusion of some of these exons has been correlated to cancer and metastasis (Günthert *et al.*, 1991; Fox *et al.*, 1994). The alternative exon v4 contains three ACE elements that were required for efficient splicing *in vivo* and spliceosome assembly *in vitro*. YB-1 bound the exon v4 and stimulated exon inclusion *in vivo* in an ACE-dependent fashion. Our results demonstrate that YB-1 is required for ACE-dependent exon v4 inclusion and provide the first identification of a sequence-specific RNA binding protein affecting *CD44* alternative splicing.

Results

A 50 kDa nuclear protein binds to ACE elements

We have previously identified a class of ACE sequences via an *in vivo* iterative selection strategy (Coulter *et al.*, 1997). In this strategy, a 13 nucleotide cassette of random nucleotides (6×10^7 potential sequences) was inserted

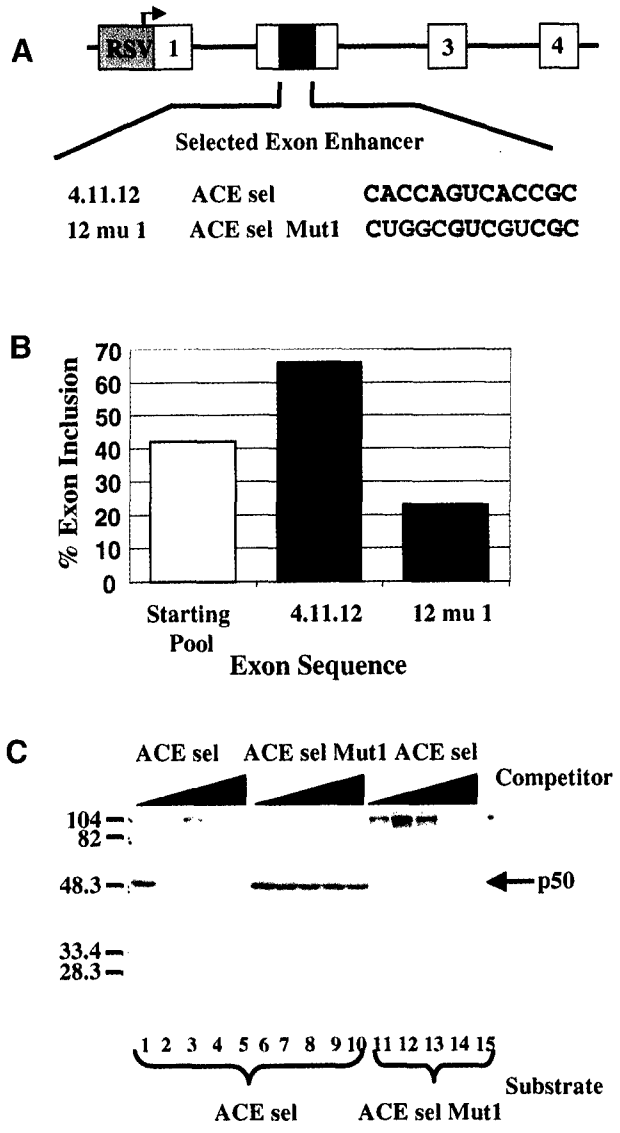


Fig. 1. A/C-rich exon enhancers bind a 50 kDa nuclear protein. (**A**) Sequence of one of the 13 nucleotide ACE exon enhancers identified by iterative *in vivo* selection from 13 random nucleotides using the diagrammed mini-gene (Coulter *et al.*, 1997). The 4.11.12 isolate, referred to here as ACE sel, is shown along with a mutant version. (**B**) *In vivo* exon inclusion activity of the sequences shown in (A). Activity was tested in a heterologous exon context (Coulter *et al.*, 1997). (**C**) *In vitro* UV cross-linking of a 51 nucleotide RNA including two copies of the ACE sel sequence from (A) (see Materials and methods). Radiolabeled RNA substrates as indicated beneath the gel were incubated for 10 min under standard *in vitro* splicing conditions using HeLa nuclear extract in the presence of increasing concentrations (0, 1, 5, 10 and 25 pmol) of competitor RNAs containing the sequences shown above the gel. Following cross-linking and RNase digestion, cross-linked proteins were resolved by 11% SDS–PAGE. A prominent 50 kDa band that was cross-linked to wild-type but not mutant RNA is indicated by an arrow.

into a weak alternative exon that was predominantly skipped in the absence of an enhancer (Figure 1A and B). Three rounds of transient transfection and RT–PCR were used to select for sequences that enhanced exon inclusion *in vivo*. Figure 1A shows the sequence of one of the A/C-rich sequences selected in this study. When placed into a natural weak exon from the cardiac troponin T gene, this sequence increased *in vivo* exon inclusion to 66% compared with 42% inclusion in the population of sequences used for selection (Figure 1B), demonstrating that this activity was independent of the mini-gene used for selection. Mutation of the A/C-rich portion of the sequence as diagrammed in Figure 1A eliminated enhancing activity (Coulter *et al.*, 1997). In this paper we used wild-type and mutant ACEs (herein termed ACE sel and ACE sel Mut1) containing a duplication of the sequence 4.11.12 and the mutant 12 mu1 shown in Figure 1.

To initially identify proteins in HeLa nuclear extracts that bind ACE elements, we performed UV cross-linking using wild-type and mutant ACE sel RNAs (Figure 1C). We identified a prominent 50 kDa protein (denoted p50) that strongly cross-linked to wild-type ACE sel RNA but not the ACE sel Mut1 mutant RNA (Figure 1C, compare lanes 1 and 11). Furthermore, binding of p50 to ACE sel was not efficiently competed by unlabeled ACE sel Mut1 RNA despite being strongly competed by wild-type RNA (Figure 1C, compare lanes 1–5 with 6–10). This result indicated that p50 was a candidate protein for an effector of ACE activity.

p50 is distinct from the SR proteins that recognize purine-rich enhancers

Our original iterative selection experiment isolated two classes of exon enhancers: ACE sequences and classic purine-rich enhancers (Coulter *et al.*, 1997). The latter are known to be recognized by the SR proteins (Tacke and Manley, 1999). To assess whether SR proteins are involved in ACE function, we first asked whether RNAs containing known purine-rich enhancers were effective competitors for the UV cross-linking of p50 to ACE sel RNA (Figure 2A). Competitors containing the well characterized purine-rich enhancer from the cardiac troponin T alternative exon 5 (Figure 2B; Ramchatesingh *et al.*, 1995) were ineffective competitors for p50 binding to ACE sel RNA (Figure 2A). Even a powerful mutant that causes higher enhancement than wild type in an exon inclusion assay (Ramchatesingh *et al.*, 1995) was an ineffective competitor. This experiment suggests that unlike classic SR proteins, p50 does not bind strongly to purine-rich enhancers.

Next, we asked whether semi-purified SR proteins would effectively UV cross-link to the ACE sel RNA or if they would enhance UV cross-linking of p50 to the ACE sel RNA (Figure 2C). For the latter we first determined that p50 was present in both HeLa nuclear extract and in cytoplasmic S100 extract that is deficient in nuclear SR proteins (Figure 2B, lanes 1 and 2). As shown in Figure 2C, the purified SR preparation showed only minimal UV cross-linking to the ACE sel sequence (Figure 2C, lanes 3–5) and had no effect on the ability of p50 in the S100 extract to cross-link to the ACE sel RNA (Figure 2C, lanes 6–8).

Identification of p50 as the human Y-box binding protein YB-1

The p50 protein was purified from HeLa nuclear extract using a UV cross-linking assay with the ACE sel RNA. As an initial fractionation attempt we took advantage of the affinity of p50 for poly(U)–agarose in 1 M NaCl. Proteins were eluted by repeated washing with buffer containing 2 M NaCl, and subsequently with guanidine

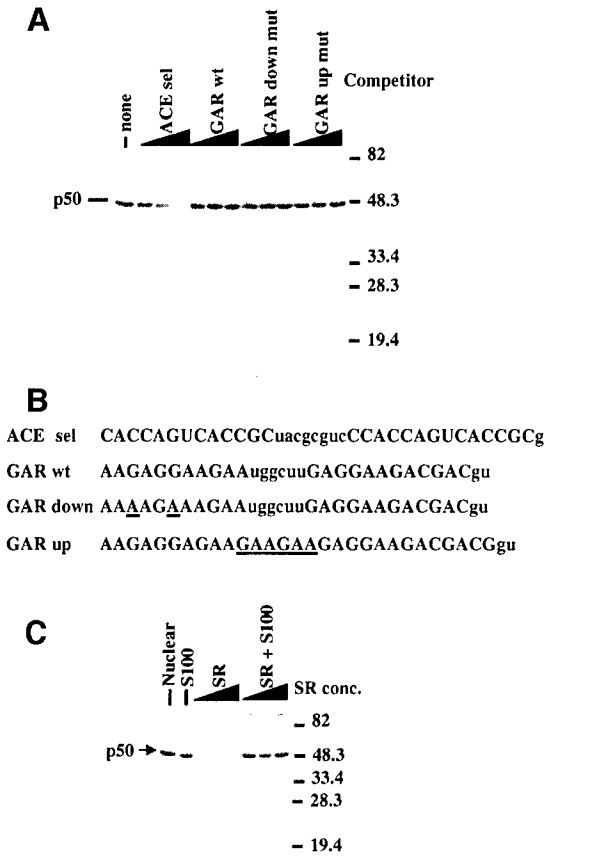


Fig. 2. p50 is distinct from the SR proteins that recognize purine-rich enhancers. (**A**) RNAs containing purine-rich enhancers do not compete UV cross-linking of p50 to ACE elements. The ACE sel RNA described in Figure 1 was subjected to UV cross-linking in the presence of competitor RNAs containing the sequences indicated above the gel. The positions of p50 and molecular weight markers are indicated. (**B**) Sequences of the competitor RNAs which are derived from exon 5 of the chicken cardiac troponin T gene [Ramchatesingh *et al.*, 1995, where they are referred to as D2WT (GAR wt), D2A2 (GAR down mutant), D2EY1 (GAR up mutant)]. Not shown are the first 17 nucleotides (ACE sel; see Materials and methods) or 15 nucleotides (GAR wt, GAR down mutant and GAR up mutant) derived from the transcription vector. (**C**) Preparations of HeLa SR proteins do not contain p50, and SR proteins do not stimulate the binding of p50. UV cross-linking was performed using *in vitro* splicing conditions (Materials and methods) in either HeLa nuclear (lane 1) or S100 (lanes 2, 6, 7 and 8) extract or with increasing amounts (50, 100 or 200 ng) of SR proteins (Zahler, 1999) in the absence (lanes 3, 4 and 5) or presence (lanes 6, 7 and 8) of S100 extract. The positions of p50 and marker proteins are indicated.

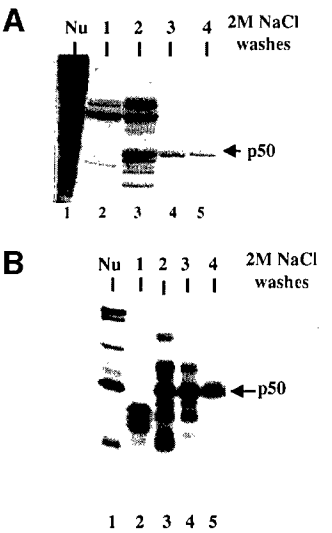


Fig. 3. Purification of p50. (**A**) HeLa nuclear extract was fractionated on poly(U)–agarose (Materials and methods). The starting material (lane 1) and four successive elution washes using 2 M NaCl are shown (lanes 2–5) in Coomassie Blue-stained 10% SDS–polyacrylamide gel. The position of a 50 kDa protein in fractions 2, 3 and 4 is denoted with an arrow. (**B**) UV cross-linking of the column fractions shown in (A). Column fractions were dialyzed against Roeder D (Dignam *et al.*, 1983) and used in a standard UV cross-linking assay with the ACE sel RNA. The cross-linked p50 in nuclear extract (lane 1) and fractions 2–4 is indicated.

hydrochloride. Fractions 2–4 of the 2 M NaCl wash contained a prominent 50 kDa protein (Figure 3A), and p50 UV cross-linking to the ACE sel RNA was seen in the same fractions (Figure 3B).

Preparative amounts of the 50 kDa protein were gel isolated. The extracted protein UV cross-linked to the ACE sel RNA in a sequence-specific manner (data not shown). The sequences of the tryptic peptides (Materials and methods) identified p50 as human YB-1 [also known as DNA binding protein B (DbpB); reviewed in Wolffe *et al.*, 1992].

To confirm the identification of p50 as YB-1, we prepared a rabbit polyclonal antibody to a C-terminal peptide of human YB-1, AENSSAPEAEQGGAE (Anaspec, CA). The antibody immunoprecipitated a 50 kDa protein that strongly cross-linked to the ACE sel RNA, confirming that p50 was YB-1 (Figure 4A). In addition, purified N-terminal His-tagged recombinant YB-1 UV cross-linked to the ACE sel RNA (Figure 4B). Competition with ACE sel and ACE sel Mut1 RNAs demonstrated that recombinant YB-1 bound with the same sequence specificity as endogenous p50 (Figure 4B, compare lanes 1–3 with 4–6).

YB-1 binds to ACE sequences in a natural alternative exon

To assess the potential role of YB-1 in enhancer recognition and alternative splicing, we investigated the function of both YB-1 and ACE elements in a natural alternative exon. The *CD44* variable exon 4 contains three A/C-rich sequences (denoted in red in Figure 5A) that resemble the ACE derived from iterative selection (Figure 5B). The strong similarity suggested that YB-1 might be involved in recognition of *CD44* exon v4.

A prominent protein of 50 kDa cross-linked to an RNA containing the first 49 nucleotides of *CD44* exon v4 containing ACE 1 and ACE 2 (Figure 6A, lane 2). The protein co-migrated with the band identified as YB-1, which UV cross-linked to the ACE sel RNA (compare lanes 2 and 4). An RNA containing sequences from intron 9 of *CD44* did not cross-link to p50 (lane 1). The 50 kDa band that cross-linked to the exon v4 substrate RNA was effectively immunoprecipitated with the YB-1-specific antibody (Figure 6B), indicating that it was indeed YB-1. Competition studies also indicated that the 50 kDa protein UV cross-linking to sequences in *CD44* exon v4 was YB-1 (Figure 6C).

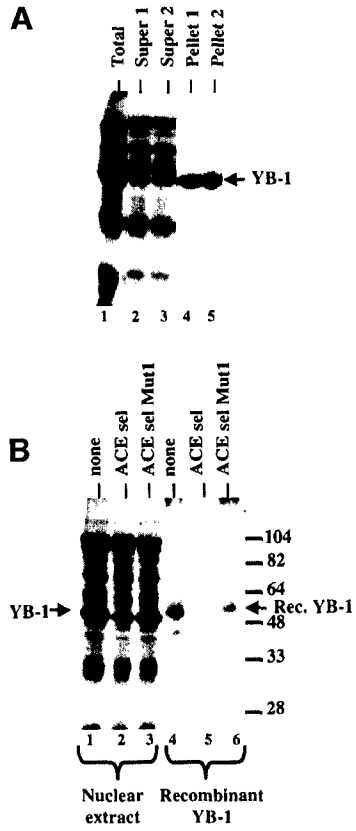
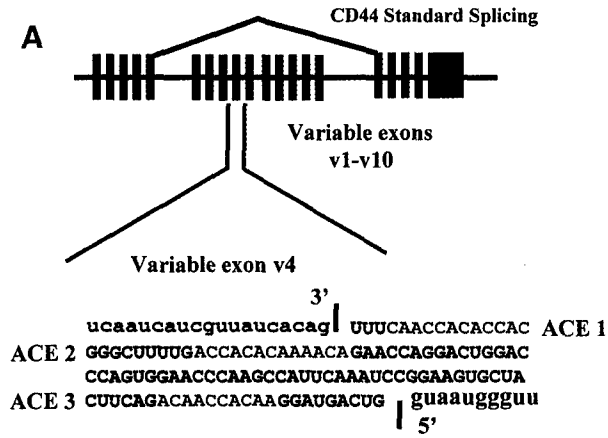


Fig. 4. Antibodies to YB-1 peptides immunoprecipitate p50. Polyclonal antibodies raised against a C-terminal peptide of human YB-1 were used to immunoprecipitate proteins cross-linked to the ACE sel RNA. (**A**) Immunoprecipitation of cross-linking proteins in HeLa nuclear extract. Duplicate immunoprecipitations are shown. Lane 1, total cross-linking; lanes 2 and 3, supernatants for immunoprecipitations 1 and 2; lanes 4 and 5, pellets from immunoprecipitations 1 and 2. (**B**) Competition of UV cross-linking of nuclear proteins (lanes 1–3) or gel-purified and renatured recombinant human YB-1 (lanes 4–6). Cross-linking reactions were performed without competitor (lanes 1 and 4) or with 10 pmol of the ACE sel competitor (lanes 2 and 5) or the ACE sel Mut 1 competitor (lanes 3 and 6). The positions of HeLa YB-1 and recombinant YB-1 are indicated; the latter is slightly larger as a result of the presence of a His tag.



B

AC-rich Repeats in CD44 Variable Exon v4

GACCACAACA	Exon Enhancer Selected *in vivo*
CAACCACACCA	CD44 exon v4 sequence 1 ACE 1
GACCACACAA	CD44 exon v4 sequence 2 ACE 2
CAACCACAA	CD44 exon v4 sequence 3 ACE 3
CAACCACA	CD44 Consensus

Fig. 5. Human *CD44* alternative splicing. (**A**) The exon–intron architecture of the human *CD44* gene is drawn at the top. The 10 alternative cassette exons are depicted in red. The exon studied in this report is the fourth variable exon v4. The sequence of this exon is shown at the bottom of the figure. Sequences within the exon are indicated with a gray background. The red sequence indicates A/C-rich elements ACE 1, ACE 2 and ACE 3 within exon v4 that are potential binding sites for YB-1. (**B**) A/C-rich repeats within *CD44* variable exons 4 and 5. The A/C-rich sequences from the two exons are aligned and a consensus repeat sequence (blue) is derived. At the top is the derived repeat consensus ACE from the exon selection experiments.

YB-1 activates CD44 exon v4 inclusion in vivo

To determine the effect of YB-1 on *CD44* splicing, we co-transfected expression plasmids coding for wild-type or a mutant version of YB-1 into HeLa cells along with a reporter construct containing *CD44* exons v4 and v5 placed within a human β-globin gene (Figure 7A). The mutant YB-1 contains a premature termination codon resulting in a truncated YB-1 protein (Figure 7B). Western blot analysis demonstrated that both full-length and truncated proteins were expressed at equivalent levels following transfection, and both were transported to the nucleus (data not shown). RT–PCR analysis of the RNA produced from the co-transfected *CD44* mini-gene demonstrated that only 28% of the expressed RNA included exons v4 and v5 in the absence of co-transfecting expression plasmid (Figure 7C, lane 1). Little RNA was produced that resulted from the inclusion of only one of the *CD44* exons. Inclusion of both exons v4 and v5 rose from 28 to >80% when increasing amounts of co-transfecting YB-1 plasmid were used (Figure 7C; quantified in D). The quantitative nature of the RT–PCR assay was demonstrated using co-transfection of two constructs for *CD44* cDNAs, one containing v4 + v5 and one lacking v4 + v5. The two constructs were co-transfected at varying ratios and the ratio of the RT–PCR products closely matched the ratio of the input DNA (data not shown). Co-transfection with the mutant form of YB-1 was unable to cause full activation of exon inclusion. This mutant protein contains the cold shock domain (CSD), which is thought to be the nucleic acid binding domain of the protein (Wolffe *et al.*, 1992) but is lacking other domains. Consistent with these reports, recombinant truncated YB-1 displays sequence-specific binding to ACE elements (data not shown). In its inability to perform wild-type function, the mutant YB-1 behaves like truncated SR proteins, which also fail to function when their SR domains have been deleted (Tacke and Manley, 1999).

YB-1 was unable to affect the inclusion of a heterologous weak exon inserted into the same mini-gene backbone (data not shown). The tested exon lacked notable A/C-rich sequences and would not be expected to respond to ACE binding proteins. YB-1 also had minimal effect on the ratio of mRNAs expressed from co-transfected *CD44* cDNAs containing and lacking v4 + v5 (data not shown), indicating that the effect in Figure 7C was not a result of mRNA stability. We cannot rule out the possibility of an indirect effect of YB-1 on *CD44* splicing

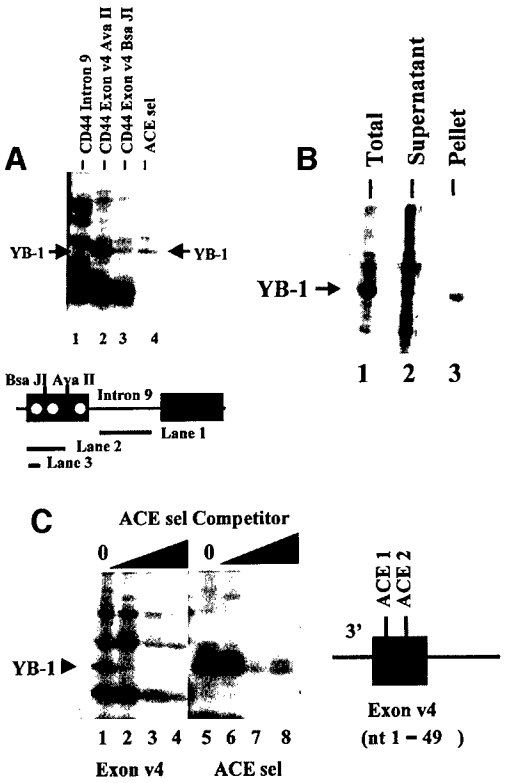


Fig. 6. *CD44* exon v4 can be UV cross-linked to YB-1. (**A**) The substrate RNAs diagrammed below the gel were UV cross-linked in a standard splicing assay using HeLa nuclear extract. A diagram of the region of the human *CD44* gene containing the fourth and fifth alternative exons is indicated below the gel. Intron 9 is the naturally occurring intron separating exons v4 and v5. ACE elements within exon v4 are indicated by white circles. The partial exon v4 RNAs shown in lanes 2 and 3 were generated by transcript termination at the indicated restriction sites. The ACE sel RNA used for lane 4 is longer than the RNA used for previous figures (see Materials and methods) resulting in the cross-linking of a protein slightly larger than YB-1 in addition to YB-1. The position of YB-1 is indicated. (**B**) Immunoprecipitation of cross-linked proteins with an antibody specific for YB-1. A substrate RNA containing the first half of exon 4 and including the ACE 1 and ACE 2 elements (diagrammed in C) was subjected to UV cross-linking in a standard HeLa *in vitro* splicing assay. Cross-linked proteins were immunoprecipitated using the anti-YB-1 antibody and displayed by SDS–PAGE. Lane 1, total proteins; lane 2, supernatant; lane 3, precipitated protein (3-fold more of the reaction was loaded in lanes 2 and 3 than 1; lanes 2 and 3 came from the same reaction). (**C**) Competition of YB-1 cross-linking to *CD44* exon v4 with an RNA containing ACE sel. Two substrate RNAs were employed: the exon 4 substrate described in (B) (lanes 1–4) and the ACE sel RNA itself (lanes 5–8). Competitor concentrations were 0, 0.3, 3.0 and 10 pmol. The position of cross-linked YB-1 is indicated.

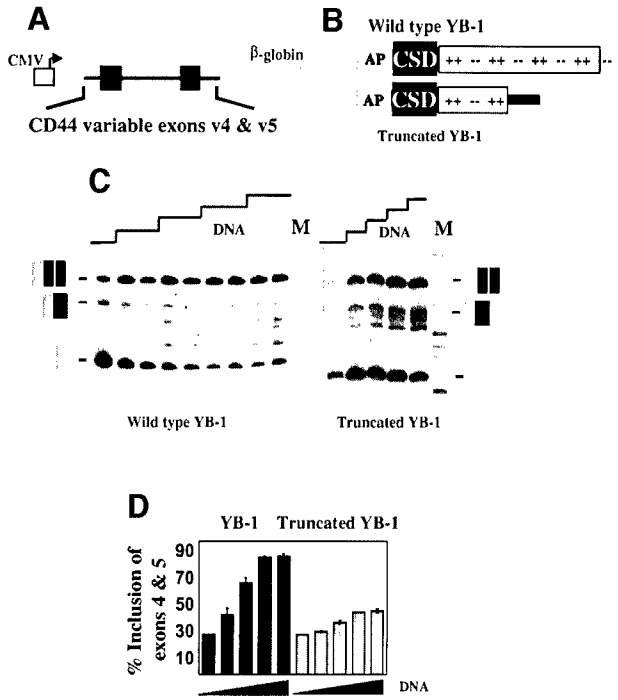


Fig. 7. Transfection of human YB-1 increased the inclusion of *CD44* exons v4 and v5. (**A**) The reporter mini-gene containing *CD44* variable exons v4 and v5. *CD44* sequences are in black, human β-globin sequences are in gray, and the CMV promoter is in white. The *CD44* insert is a contiguous genomic fragment mapping from 792 nucleotides upstream of exon v4 to 515 nucleotides downstream of exon v5. (**B**) Diagram of the wild-type and mutant YB-1 proteins made from co-expressing plasmids. The three indicated domains include an N-terminal alanine- and proline-rich region (AP), the single-stranded nucleic acid binding CSD and the C-terminal highly charged region. (**C**) RT–PCR analysis of RNAs produced upon transfection of HeLa cells with the reporter shown in (A) and increasing amounts (0, 1, 2, 3 or 4 μg) of either wild-type or mutant YB-1 expression plasmids. Duplicate lanes are shown for each concentration of the wild-type YB-1. PCR oligonucleotides were complementary to sequences in the flanking β-globin exons. Product RNAs resulting from inclusion of neither *CD44* exon, either exon v4 or v5, or both exons v4 and v5 are indicated. Note that exons v4 and v5 have such similar sizes that inclusion bands resulting from either of the exons will be in the same region of the gel. Increasing amounts of the YB-1 expression plasmid resulted in a linear increase in the amount of YB-1 mRNA (data not shown). (**D**) Quantification of the effect of YB-1 on inclusion of *CD44* exons v4 and v5. Results of four independent experiments are shown with 1 SD indicated.

caused by the transcriptional up-regulation of an unknown splicing protein. However, as YB-1 binds to *CD44* exon sequences, the most straightforward interpretation of the data in Figure 7 is that YB-1 participates directly in the splicing of *CD44* exons v4 and v5.

The CD44 exon v4 ACE 1 and ACE 3 elements are required for normal splicing and binding of YB-1

If YB-1 is important for exon v4 recognition, ACE sequences within exon v4 should be required for maximal activity. To test this hypothesis, each of the three ACE elements within exon v4 was mutated (Figure 8). These mutants were transfected into HeLa cells in the context of the mini-gene described in Figure 7. The ACE 3 mutation had the strongest phenotype (Figure 8B). It depressed the percentage of RNA that included both exons v4 and v5 from 25 to 6%. A major new PCR product in which only v4 or v5 was included appeared. This product was sequenced and found to be RNA that included only exon v5. Thus, the mutation had no effect on inclusion of v5 as the fraction of mRNAs that include v5 was comparable in wild type and mACE3, but it significantly depressed inclusion of exon v4.

The ACE 2 mutation had little effect, suggesting that it, alone, is not a major sequence regulating exon v4 inclusion. The ACE 1 sequence had an unusual effect. When mutated, exon v4 and v5 inclusion was increased such that 47% of the product RNA included v4 and v5 as compared with 25% for the wild type, suggesting that this sequence represents a modest exon silencer. A silencer with a

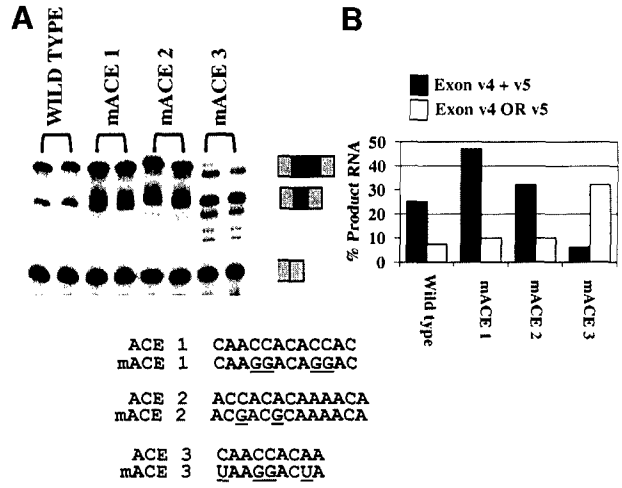


Fig. 8. Mutation of *CD44* exon v4 ACE 1 or 3 alters *in vivo* recognition of exon v4. (**A**) The diagrammed mutations were introduced into exon v4 of the *CD44* mini-gene described in Figure 7. Wild-type and ACE 1, 2 or 3 mini-genes were transfected in duplicate into HeLa cells and the splicing phenotype was determined by RT–PCR as described in Figure 7. Note that because exons v4 and v5 are almost the same size, products containing exons v4 or v5 alone are indistinguishable in size. Sequencing of the band produced with the ACE 3 mutant indicates that this RNA product contains exon v5 but not v4. Product RNAs including no, one or two *CD44* alternative exons are indicated. (**B**) Quantification of the results in (A).

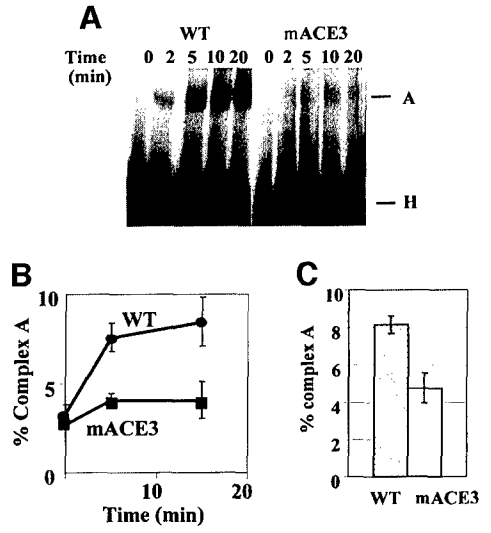


Fig. 9. Mutation of ACE 3 depressed spliceosome assembly on *CD44* exon v4. The wild-type or ACE3 mutant substrate RNAs including all of variable exon v4 and its flanking 3′ and 5′ splicing signals were incubated under *in vitro* splicing conditions for 0, 2, 5, 10 or 20 min. Mutations are identical to those described in Figure 8. (**A**) Formed complexes were analyzed by native gel electrophoresis (Carlo *et al.*, 2000). The non-specific H complex and initial ATP-dependent spliceosomal A complex are indicated. (**B**) Quantification of the reaction in (A). The gel was scanned in the phosphoimager and the percentage of RNA in complex A was plotted. (**C**) Quantification of the amount of complex A formed at 20 min from multiple experiments using mutant and wild-type substrate RNA.

different sequence has been detected experimentally in the beginning of the murine *CD44* exon v5 as well (König *et al.*, 1998).

Figure 9 displays the ability of the ACE 1 and 3 mutants to direct *in vitro* spliceosome assembly using a substrate RNA containing intact exon v4 with its surrounding splice sites. Such substrates have the ability to assemble the first ATP-dependent spliceosome complex, complex A. Wild-type RNA forms complex A (Figure 9A), as does the ACE 1 mutant (data not shown). The ACE 3 mutant is depressed for complex A formation, in agreement with the *in vivo* depression of production of spliced RNA containing exon 4 (Figure 9). This result suggests that at least a portion of the phenotype of an ACE 3 mutation *in vivo* is a result of a reduced ability to form an early spliceosome complex.

The ACE 1–3 mutations also depressed the UV cross-linking of YB-1 to exon v4 sequences (Figure 10). Some cross-linking was detected with the ACE 1 and ACE 2 mutants (Figure 10, top), consistent with the inability of these mutants to adversely affect either *in vivo* splicing or *in vitro* assembly. Reduced cross-linking was observed with the ACE 3 mutant, as detected by both the cross-linking of total proteins (Figure 10, top) and immunoprecipitation of cross-linked proteins with the YB-1-specific antibody (Figure 10, bottom). This absence is consistent with reduced inclusion of exon v4 *in vivo* and the reduction of *in vitro* assembly of exon 4 with the ACE 3 mutant.

The ACE 3 mutation was less responsive to increasing concentrations of YB-1 than wild type *in vivo* (Figure 11A, compare lanes 1–4 with 5–8). YB-1 increased the percentage of wild-type product RNA including both exons v4 and v5 by 40% (Figure 11B). In the mutant, however, little RNA was produced that contained both v4 and v5. In the presence of YB-1, the amount of mutant RNA including both exons increased by only 17% and the amount of RNA including exon v5 alone increased 12%. Thus, increasing the concentration of YB-1 was unable to reverse the effect of the ACE 3 mutant in exon v4 significantly. Given the ability of YB-1 to bind to both ACE 1 and ACE 2 elements, it is not surprising that the ACE 3 mutant retained some ability to respond to YB-1.

Discussion

Exon enhancers are abundant accessory elements that facilitate the recognition of exons during the earliest steps of pre-mRNA splicing. Although most of the experimental attention to the mechanism of enhancer recognition has concentrated on the purine-rich enhancers and the SR proteins that recognize them, a number of other exon sequences are potential enhancers (Dominski and Kole, 1994; Staknis and Reed, 1994; van Oers *et al.*, 1994; Wang *et al.*, 1995; Coulter *et al.*, 1997; Gersappe and Pintel, 1999; Schaal and Maniatis, 1999a,b). One common exon enhancer is the A/C-rich ACE element (van Oers *et al.*, 1994; Wang *et al.*, 1995; Coulter *et al.*, 1997; Schaal and Maniatis, 1999b). Here we identify the human Y-box binding protein YB-1 as a nuclear protein that binds to ACE elements and increases *in vivo* inclusion of exons containing ACE elements in a fashion that is dependent upon both the presence of exonic ACE elements and full-length YB-1 protein.

YB-1 was originally identified as a transcription factor that binds single-stranded DNA within the reverse CCAAT promoter element (Dorn *et al.*, 1987; Didier *et al.*, 1988), which is similar to the RNA ACE elements.

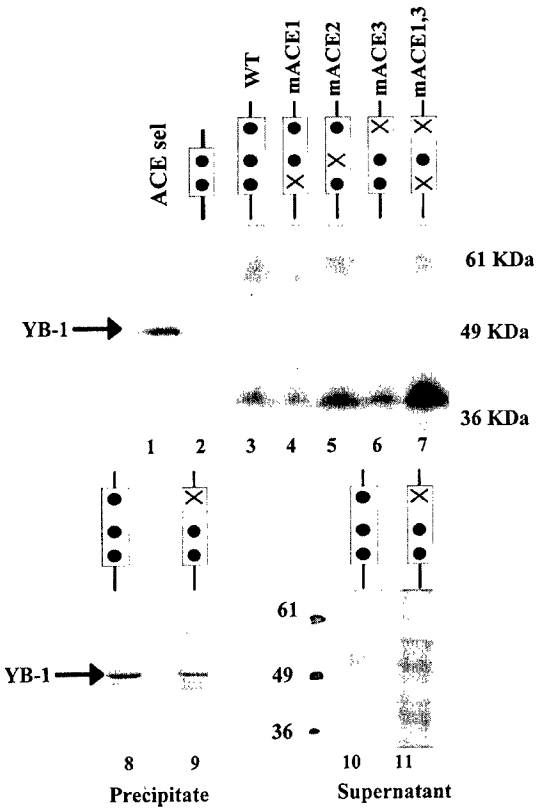


Fig. 10. Mutation of *CD44* ACE sequences depresses UV cross-linking of YB-1. Top: the binding of YB-1 to wild-type and mutant *CD44* exon 4 sequences was compared by UV cross-linking. The substrates are diagrammed above the gel. The position of YB-1 is indicated. Bottom: immunoprecipitation of YB-1 UV cross-linked to wild-type or mutant substrates. Cross-linked reactions were immunoprecipitated using the YB-1-specific antibody. Proteins from both the precipitate and the supernatant were displayed by SDS–PAGE.

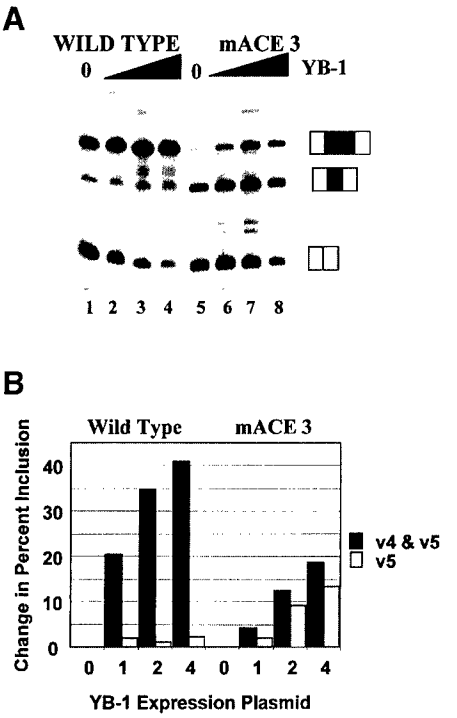


Fig. 11. Mutation of *CD44* ACE sequences depresses the ability to respond to YB-1 *in vivo*. (**A**) The mini-gene containing *CD44* variable exons v4 and v5 was co-transfected with an expression vector coding for YB-1 as in Figure 9. The mini-gene contained a wild-type exon v4 (lanes 1–4) or an exon v4 containing the ACE 3 mutation (lanes 5–8). Increasing amounts of YB-1 vector were used (0, 1, 2 or 4 μg). The positions of RNA species resulting from the inclusion of no, one or two *CD44* exons are indicated. (**B**) Quantification of the results in (A).

Recently, considerable attention has been focused on YB-1 as an RNA binding protein (Tafuri and Wolffe, 1990; Deschamps *et al.*, 1992; Murray *et al.*, 1992; Minich *et al.*, 1993; Evdokimova *et al.*, 1995). YB-1 is an abundant protein in both the nucleus and the cytoplasm of many cell types (Tafuri and Wolffe, 1990; Ranjan *et al.*, 1993). YB-1 binds to cytoplasmic mRNA and participates in translation regulation. In *Xenopus*, YB-1 is the major protein within storage mRNA particles in oocytes (Darnbrough and Ford, 1981; Kick *et al.*, 1987; Tafuri and Wolffe, 1990; Murray *et al.*, 1991, 1992; Marello *et al.*, 1992). In this case, multiple copies of the protein bind to RNA with a density of one protein molecule per 50 nucleotides (Darnbrough and Ford, 1981; Marello *et al.*, 1992). *In vitro* SELEX revealed that the *Xenopus* protein (FRGY2) binds best to the sequence AACUAC (Bouvet *et al.*, 1995), similar to the ACE elements described in this report. No RNA binding role for YB-1 within the nucleus had been reported prior to this study.

The structure of the YB-1 protein is ideal for a splicing regulatory protein (reviewed in Wolffe *et al.*, 1992). Like other RNA binding proteins, the sequence-specific binding domain of the protein contains β-sheets. In contrast to the four-stranded β-sheet associated with the SR and hnRNP proteins or the three-stranded β-sheet found in the KH proteins, YB-1 contains a five-stranded β-barrel known as the CSD (Wistow, 1990). YB-1 also contains accessory domains that are highly enriched in both basic amino acids and aromatic groups, referred to as existing in basic/aromatic (B/A) islands (Murray *et al.*, 1992). Separating the B/A islands are acidic regions predicted to form α-helices. It has been proposed that the accessory domains are responsible for sequence-independent RNA binding and that the CSD provides sequence specificity (Bouvet *et al.*, 1995). This arrangement of a specificity domain containing β-sheets and a basic domain providing generic RNA binding is highly reminiscent of the organization and functionality of the SR proteins. The mutant YB-1 used in this study is truncated shortly after the CSD, thereby eliminating many of the B/A islands. Its lowered ability to stimulate *CD44* splicing suggests a role for the accessory domain in splicing.

Also, like the SR proteins, YB-1 is highly phosphorylated. In *Xenopus*, phosphorylation is required for binding to RNA and specific dephosphorylation is associated with mRNA release from storage particles (Kick *et al.*, 1987; Cummings and Summerville, 1988). A similar phosphorylation–dephosphorylation cycle occurs with SR proteins (Tacke and Manley, 1999). Bacterially produced YB-1 was able to bind to ACE elements with the same sequence specificity as the HeLa protein. We were, however, unable to use bacterially produced protein to activate splicing activity *in vitro*, suggesting that correct phosphorylation might be required.

The ACE elements and the YB-1 preferred binding sequence are similar to the A/C-rich element required for exon recognition in the *Drosophila melanogaster* double-sex gene. The double-sex enhancer binds the sex-regulating proteins Tra and Tra 2 (Hoshijima *et al.*, 1991; Tian and Maniatis, 1992; Lynch and Maniatis, 1995). The latter is an SR protein, although it does not purify with the other 'classic' SR proteins. Human cells have no Tra orthologs, but do have two closely related Tra 2 proteins, Tra 2α and Tra 2β (Dauwalder *et al.*, 1996; Nayler *et al.*, 1998). It might be predicted that these proteins will also bind to and regulate ACE elements in human cells, although it should be noted that an attempt to define a binding site for human Tra 2 using iterative selection detected purine-rich binding sites, not ACE elements (Tacke *et al.*, 1998). Preliminary experiments indicate that human Tra 2 proteins do interact with the *CD44* exon v4 sequences and may also interact with YB-1 (E.Stickeler, A.Honig, W.Mattox and S.M.Berget, manuscript in preparation). In contrast, ASF/SF2 and SC35 had no effect on *CD44* exon v4 or v5 alternative splicing (data not shown). Thus, a slightly different member of the SR protein family, Tra 2, may be involved in ACE recognition.

We previously demonstrated that RNAs containing purine-rich exon splicing enhancers inhibited ACE-dependent splicing *in vitro*, suggesting a role for SR proteins in ACE-dependent splicing (Coulter *et al.*, 1997). Here we demonstrate that SR proteins do not bind strongly to ACEs. Neither do SR proteins affect binding of YB-1 to the selected ACE. Together, these results suggest that a role for SR proteins in ACE-dependent splicing occurs subsequent to binding of YB-1, consistent with reports describing roles for SR proteins following complex A assembly (Roscigno and Garcia-Blanco, 1995; Chew *et al.*, 1999). YB-1 also has properties reminiscent of another class of prominent RNA binding proteins that influence splicing: the hnRNP proteins (reviewed in Weighardt *et al.*, 1996; Krecic and Swanson, 1999). Like the hnRNP proteins, YB-1 could multimerize on RNA to form structures that look like beads under the electron microscope (Tafuri and Wolffe, 1992). The functionality of such association is unclear, but could provide interactions between exons bearing ACE elements. The *CD44* gene has multiple consecutive alternative exons containing putative ACE elements. The simultaneous recognition of multiple ACE exon enhancer domains by the self-association of bound YB-1 protein could explain the observation that blocks of *CD44* exons are simultaneously included upon stimulation or in certain cancers (Fox *et al.*, 1994; Mackay *et al.*, 1994; Bell *et al.*, 1998; Stickeler *et al.*, 1999). Interestingly, the increased inclusion of *CD44* exons in cancer correlates with increased expression levels of YB-1 (E.Stickeler and S.M.Berget, unpublished results; Bargou *et al.*, 1997).

The protein target by which the binding of YB-1 activates exon recognition is not known. SR proteins are thought to activate splicing by interaction with constitutive splicing factors, such as U170K and U2AF. The only RNA binding protein known to interact with YB-1 is hnRNP K (Funke *et al.*, 1996; Shnyreva *et al.*, 2000). Although this interaction is more relevant to the cytoplasmic activities of YB-1, it does suggest that YB-1 might interact with other members of the KH family of splicing factors. There are multiple splicing regulators that are KH proteins (Siebel *et al.*, 1995; Arning *et al.*, 1996; Min *et al.*, 1997). Perhaps the most intriguing candidate is splicing factor 1 (SF1), a highly conserved protein that plays a role in early spliceosome assembly by interactions with factors bound to both 5′ and 3′ splice sites (Berglund *et al.*, 1998). Such an interaction would place interactions between YB-1 and SF1 in a role analogous and parallel to the activation of early assembly by interactions between SR proteins and U2AF/U170K.

Materials and methods

Plasmids and transfections

In vitro transcription templates for ACE sel and ACE sel Mut1 RNAs were produced by cloning duplicate copies of the 4.11.12 or 12 mu1 selected ACE sequences described in Coulter *et al.*, (1997) downstream of the T7 promoter in Bluescript KS+. Cut with *Nhe*I, ACE sel was transcribed into a 59 nucleotide RNA of sequence GGGCGAAUUGGCUAGAGCACCAGUCACCGCUACGCGUCCACCAGUCACCGCCGCUAGC (used for Figures 1–4). The underlined regions are the selected sequences; both copies were mutated in the 12 mu1 variant (sequence in Figure 1). Cut with *Acc*I, a 69 nucleotide RNA was produced that added the sequence GAUACCGUCG to the 3′ end of the above RNAs (used for Figure 6). A cDNA clone isolated from a human HeLa cell line and coding for YB-1 was obtained from P.Newburger (Shen *et al.*, 1998). This clone contained a 2 bp deletion at amino acid 207 causing a frame shift and premature termination of translation 14 amino acids downstream. A wild-type version of this expression plasmid was prepared by replacing the mutated region with a cDNA prepared by PCR amplification of a commercial cDNA preparation isolated from human skeletal muscle. Both were subcloned into pET28a for bacterial production of N-terminal His-tagged YB-1 protein and pcDNA3.1/HisC for *in vivo* expression. The *CD44* mini-gene (Figure 7) was created by inserting a genomic fragment from the human *CD44* gene containing variable exons v4 and v5 into the first intron of a β-globin *in vivo* expression plasmid derived from Dup 33 (Dominski and Kole, 1992) obtained from R.Kole, University of North Carolina, NC. Several templates containing *CD44* exon v4 sequences were used for *in vitro* transcription. A short version containing nucleotides –20 upstream of the exon to +30 downstream of the exon was used to generate an RNA incapable of forming complex A (because of deletion of important sequences from the 3′ splice site) in Figures 6 and 7. A longer clone including 38 nucleotides of the intron upstream of exon v4 was used for *in vitro* assembly and some UV cross-linking studies (Figures 9 and 10). Transcripts were produced by truncating these plasmids within or after exon v4 sequences using *Bsa*JI (+11 nucleotides of exon v4), *Ava*II (+49 nucleotides of exon v4) or *Sfa*I (56 nucleotides after the exon v4 5′ splice site). The first RNA is truncated within ACE 1, the second contains ACE 1 and ACE 2, and the third contains all three exon v4 ACE elements. All constructs were sequenced for verification.

Transfections of HeLa cells using lipofectAMINE™ (Gibco-BRL) and isolation of total cell RNA 48 h post-transfection using TRIzol™ (Gibco-BRL) were performed following the manufacturer's instructions. Splicing patterns were determined by RT–PCR analysis using 5′-end-radiolabeled primers specific for β-globin sequences (5′-AGACACCATGCATGGTGCACC and 3′-CCTGATCAGCGAGCTCTAG). These primers amplified no RNA from untransfected HeLa cells. Amplification conditions were 40 s at 94°C, 40 s at 58°C and 1 min at 72°C for 25 cycles (Figures 7 and 8), or 30 s at 94°C, 30 s at 62°C and 1 min at 72°C for 20 cycles (Figure 11). Product DNA was denatured and displayed on 5–6% urea gels. RNA products were quantified in the phosphoimager. Identified amplification products resulting from the inclusion of one or two variable *CD44* exons were sequenced to verify identity.

In vitro assays

In vitro synthesis of RNA, splicing, spliceosome assembly and UV cross-linking assays have been described previously (Ramchatesingh *et al.*, 1995). To immunoprecipitate YB-1, a Pansorbin suspension (20 μl) (Calbiochem) was washed once with 1× gamma bind buffer (BB; 10 mM sodium phosphate, 150 mM NaCl, 10 mM EDTA). Five microliters of rabbit antiserum were bound to Pansorbin in 300 μl of BB by rocking overnight at 4°C. The antibody beads were washed three times with RIPA

buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% v/v NP-40, 0.5% deoxycholate, 0.1% SDS) and then once with buffer DG (20 mM HEPES–KOH pH 7.9 at 4°C, 80 mM monopotassium glutamate, 0.2 M EDTA) plus 0.05% Triton X-100. The cross-linking reactions were used to resuspend the Pansorbin pellet following RNase digestion. Five microliters of 3 mg/ml heparin and 0.15% Triton X-100 were added, and the mixture was agitated for 30 min at 4°C with vigorous shaking. Proteins were prepared from the supernatant by acetone precipitation. The Pansorbin pellet was washed three times with DG plus 0.05% Triton X-100, then resuspended in 40 µl of protein loading buffer. Semi-purified SR proteins were recovered in the 65–90% ammonium sulfate cuts of cleared HeLa nuclear extract (Zahler, 1999) and dialyzed against Roeder D buffer without glycerol (Dignam *et al.*, 1983).

Purification of YB-1

YB-1 was fractionated from HeLa nuclear extract using a protocol originally described to purify U2AF (Wang *et al.*, 1995). HeLa nuclear extract was dialyzed into HeLa equilibration buffer [20 mM HEPES pH 7.6, 3 mM $MgCl_2$, 0.2 mM EDTA, 1 mM dithiothreitol (DTT), 1.0 M KCl, 10% glycerol]. Cleared and dialyzed extract was passed over a poly(U)–Sepharose column equilibrated with buffer A1 (20 mM HEPES pH 7.9, 3 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 1.0 M KCl, 5% glycerol) at a ratio of ~1 ml of extract per 2 ml of bed volume. After extensive washing with buffer A1, YB-1 was eluted with four successive column volumes of buffer A1 with KCl to 2.0 M and buffer A1 with NaCl lowered to 0.1 M and supplemented with 2.0 M guanidine–HCl. The fourth 2.0 M KCl fraction and the first two guanidine–HCl fractions were pooled from multiple columns and used to gel purify the 50 kDa band for tryptic peptide mapping. Resulting peptides were sequenced by the Protein Chemistry Core Facility at Baylor College of Medicine. The four peptides sequenced each had 100% identity with human YB-1 (EDVFVHQTAIK, GAEAANVTGPGGVPVQG, PGTTGSGAGSGGP-GG and ENQGDETQCQQPPQR).

N-terminal His-tagged recombinant YB-1 was induced in *Escherichia coli* BL21 using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3–4 h. Protein was purified using Ni-NTA according to the manufacturer's instructions (Qiagen). Recombinant protein was gel-purified from SDS–PAGE. Protein isolated from gel bands was concentrated by precipitation with four volumes of acetone at –20°C for 30 min, and centrifugation at 13 000 *g* for 20 min. The pellet was resuspended in 60 µl of 6 M guanidine–HCl in Roeder D buffer and incubated at room temperature to completely denature the protein. Renaturation was accomplished by dilution into 3.0 ml of Roeder D buffer and incubation at room temperature for 1 h. The resulting solution was desalted and concentrated in an ultra free centrifugal filter device (Millipore).

Acknowledgements

Special thanks to members of the Cooper and Berget laboratories for helpful suggestions and critical comments. We thank Dr P.Newburger for YB-1 clones and antibodies, R.Kole for plasmid constructs, the Baylor College of Medicine genome center for genomic human clones, Dr Richard Cook and the Baylor College of Medicine Protein Core for advice and protein sequencing, the Cell Culture Center (Endotronics Inc.) for large volume HeLa cell growth, and Gopal Singh for technical assistance. This work was supported by grants from The American Cancer Society (NP-79230) and the NIH (RO1 HL45565) to T.A.C.; the US Army (DAMD17-96-1-6084) and the NIH (GM38526) to S.M.B., the German Research Foundation (Sti. 153/1-1) to E.S., and the Foundation for Gene and Cell Therapy and the Medical Research Council of Canada (The Jesse Davidson Post-doctoral Research Fellowship) to S.D.F.

References

Arning,S., Gruter,P., Bilbe,G. and Kramer,A. (1996) Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA*, **2**, 794–810.

Bargou,R.C. *et al.* (1997) Nuclear localization and increased levels of transcription factor YB-1 in primary human breast cancers are associated with intrinsic *MDR1* gene expression. *Nature Med.*, **3**, 447–478.

Bell,M.V., Cowper,A.E., Lefranc,M.P., Bell,J.I. and Screaton,G.R. (1998) Influence of intron length on alternative splicing of *CD44*. *Mol. Cell. Biol.*, **18**, 5930–5941.

Berglund,J.A., Abovich,N. and Rosbash,M. (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.*, **12**, 858–867.

Black,D. (1995) Finding splice sites within a wilderness of RNA. *RNA*, **1**, 763–771.

Bouvet,P., Matsumoto,K. and Wolffe,A.P. (1995) Sequence-specific RNA recognition by the Xenopus Y-box binding proteins. *J. Biol. Chem.*, **270**, 28297–28303.

Carlo,T., Sierra,R. and Berget,S.M. (2000) A 5′ splice site-proximal enhancer binds SF1 and activates exon bridging of a microexon. *Mol. Cell. Biol.*, **20**, 3988–3995.

Chew,S.L., Liu,H.X., Mayeda,A. and Krainer,A.R. (1999) Evidence for the function of an exonic splicing enhancer after the first catalytic step of pre-mRNA splicing. *Proc. Natl Acad. Sci. USA*, **96**, 10655–10660.

Cooper,T.A. (1999) Defining pre-mRNA *cis* elements that regulate cell-specific splicing. *Methods Mol. Biol.*, **118**, 391–403.

Cooper,T.A. and Mattox,W. (1997) The regulation of splice-site selection, and its role in human disease. *Am. J. Hum. Genet.*, **61**, 259–266.

Coulter,L.R., Landree,M.A. and Cooper,T.A. (1997) Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol. Cell. Biol.*, **17**, 2143–2150.

Cummings,A. and Sommerville,J. (1988) Protein kinase activity associated with stored messenger ribonucleoprotein particles of Xenopus oocytes. *J. Cell Biol.*, **107**, 45–56.

Darnbrough,C.H. and Ford,P.J. (1981) Identification in *Xenopus laevis* of a class of oocyte-specific proteins bound to messenger RNA. *Eur. J. Biochem.*, **113**, 415–424.

Dauwalder,B., Amaya-Manzanares,F. and Mattox,W. (1996) A human homolog of the Drosophila sex determination factor transformer-2 has conserved splicing regulatory functions. *Proc. Natl Acad. Sci. USA*, **93**, 9004–9009.

Deschamps,S., Viel,A., Garrigos,M., Denis,H. and LeMaire,M. (1992) mRNP4, a major RNA binding protein from *Xenopus* oocytes is identical to a transcription factor FRGY2. *J. Biol. Chem.*, **267**, 13799–13802.

Didier,D.K., Schiffenbauer,J., Woulfe,S.L., Zacheis,M. and Schwartz, B.D. (1988) Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box. *Proc. Natl Acad. Sci. USA*, **85**, 7322–7326.

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

Dominski,Z. and Kole,R. (1992) Cooperation of pre-RNA sequence elements in splice site selection. *Mol. Cell. Biol.*, **12**, 2108–2114.

Dominski,Z. and Kole,R. (1994) Identification of exon sequences involved in splice site selection. *J. Biol. Chem.*, **269**, 23590–23596.

Dorn,A., Durand,B., Marfling,C., Le meur,M., Benoist,C. and Mathis,D. (1987) The conserved MHC class II boxes X and Y are transcriptional control elements and specifically bind nuclear proteins. *Proc. Natl Acad. Sci. USA*. **84**, 6249–6253.

Evdokimova,V.M., Wei,C., Sitikov,A.S., Simonenko,P.N., Lazerev, O.A., Vasilenko,K.S., Usinov,VA., Hershey,J.W.B. and Ovchinnikov, L.P. (1995) The major protein of messenger ribonucleoprotein particles in somatic cells is a member of the Y-box binding transcription factor family. *J. Biol. Chem.*, **270**, 3186–3192.

Fox,S.B., Fawcett,J., Jackson,D.G., Collins,I., Gatter,K.C., Harris,A.L., Gearing,A. and Simmons,D.L. (1994) Normal human tissues, in addition to some tumors, express multiple different *CD44* isoforms. *Cancer Res.*, **54**, 4539–4546.

Funke,B., Zuleger,B., Benavente,R., Schuster,T., Goller,M., Stevenin,J. and Horak,I. (1996) The mouse poly(C) binding protein exists in multiple isoforms and interacts with several RNA binding proteins. *Nucleic Acids Res.*, **24**, 3821–3828.

Gersappe,A. and Pintel,D.J. (1999) CA- and purine-rich elements form a novel bipartite exon enhancer which governs inclusion of the minutevirus of mice NS2-specific exon in both singly and doubly spliced mRNAs. *Mol. Cell. Biol.*, **19**, 364–375.

Günthert,U. (1993) *CD44*: a multitude of isoforms with diverse functions. *Curr. Top. Microbiol. Immunol.*, **184**, 47–63.

Günthert,U. *et al.* (1991) A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. *Cell*, **65**, 13–24.

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

Kick,D., Barrett,P., Cummings,A. and Sommerville,J. (1987) Phosphorylation of a 60 kDa polypeptide from Xenopus oocytes

blocks messenger RNA translation. *Nucleic Acids Res.*, **15**, 4099–4109.

König,H., Ponta,H. and Herrlich,P. (1998) Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. *EMBO J.*, **17**, 2904–2913.

Krecic,A.M. and Swanson,M.S. (1999) hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.*, **11**, 363–371.

Lynch,K.W. and Maniatis,T. (1995) Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes Dev.*, **9**, 284–293.

Mackay,C.R., Terpe,H.-J., Stauder,R., Marston,W.L., Stark,H. and Gunthert,U. (1994) Expression and modulation of *CD44* variant isoforms in humans. *J. Cell Biol.*, **124**, 71–82.

Marello,K., La Rovere,J. and Sommerville,J. (1992) Binding of Xenopus oocyte masking proteins to mRNA sequences. *Nucleic Acids Res.*, **20**, 5593–5600.

Min,H., Turck,C.W., Nikolicand,J.M. and Black,D.L. (1997) A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes Dev.*, **11**, 1023–1036.

Minich,W.B., Maidebura,I.P. and Ovchinnikov,L.P. (1993) Purification of the major 50-kDa repressor protein from cytoplasmic mRNA of rabbit reticulocytes. *Eur. J. Biochem.*, **212**, 633–638.

Murray,M.T., Krohne,G. and Franke,W.W. (1991) Different forms of cytoplasmic mRNA binding proteins and particles in *Xenopus laevis* oocytes and embryos. *J. Cell Biol.*, **112**, 1–11.

Murray,M.T., Schiller,D.L. and Franke,W.W. (1992) Sequence analysis of cytoplasmic RNA binding proteins of Xenopus oocytes identifies a family of RNA binding proteins. *Proc. Natl Acad. Sci. USA*, **89**, 11–15.

Nayler,O., Cap,C. and Stamm,S. (1998) Human transformer-2-β gene (SFRS10): complete nucleotide sequence, chromosomal localization, and generation of a tissue-specific isoform. *Genomics*, **53**, 191–202.

Ramchatesingh,J., Zahler,A.M., Neugebauer,K.M., Roth,M.B. and Cooper,T.A. (1995) A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. *Mol. Cell. Biol.*, **15**, 4898–4907.

Ranjan,M., Tafuri,S.R. and Wolffe,A.P. (1993) Masking mRNA from translation in somatic cells. *Genes Dev.*, **7**, 1725–1736.

Reed,R. (1996) Initial splice site recognition and pairing during pre-mRNA splicing. *Curr. Opin. Genet. Dev.*, **6**, 215–220.

Roscigno,R.F. and Garcia-Blanco,M.A. (1995) SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA*, **1**, 692–706.

Schaal,T.D. and Maniatis,T. (1999a) Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol. Cell. Biol.*, **19**, 261–273.

Schaal,T.D. and Maniatis,T. (1999b) Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.*, **19**, 1705–1719.

Shen,Q., Wu,R., Leonard,J.L. and Newburger,P.E. (1998) Identification and molecular cloning of a human selenocysteine insertion sequence binding protein. A bifunctional role for DNA binding protein B. *J. Biol. Chem.*, **273**, 5443–5446.

Shnyreva,M., Schullery,D.S., Suzuki,H., Higaki,Y. and Bomsztyk,K. (2000) Interaction of two multifunctional proteins. *J. Biol. Chem.*, **275**, 15498–15503.

Siebel,C.W., Admon,A. and Rio,D.C. (1995) Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. *Genes Dev.*, **9**, 269–283.

Staknis,D. and Reed,R. (1994) 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. *Mol. Cell. Biol.*, **14**, 7670–7682.

Stickeler,E., Kittrell,F., Medina,D. and Berget,S.M. (1999) Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. *Oncogene*, **18**, 3574–3582.

Tacke,R. and Manley,J.L. (1999) Determinants of SR protein specificity. *Curr. Opin. Cell Biol.*, **11**, 358–362.

Tacke,R., Tohyama,M., Ogawa,S. and Manley,J.L. (1998) Human Tra 2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell*, **93**, 139–148.

Tafuri,S.T. and Wolffe,A.P. (1990) Xenopus Y-box binding transcription factors: molecular cloning, functional analysis and developmental regulation. *Proc. Natl Acad. Sci. USA*, **87**, 9028–9032.

Tafuri,S.T. and Wolffe,A.P. (1992) DNA binding, multimerization, and transcription stimulation by the Xenopus Y box proteins *in vitro*. *New Biol.*, **4**, 349–359

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

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

Wang,Z., Hoffmann,H.M. and Grabowski,P.J. (1995) Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. *RNA*, **1**, 21–35.

Weighardt,F., Biamonti,G. and Riva,S. (1996) The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism. *BioEssays*, **18**, 747–756.

Wistow,G. (1990) Cold shock and DNA binding. *Nature*, **344**, 823–824.

Wolffe,A.P., Tafuri,S., Ranjan,M. and Familari,M. (1992) The Y-box binding factors: a family of nucleic acid binding proteins conserved from *Escherichia coli* to man. *New Biol.*, **4**, 290–298.

Zahler,A.M. (1999) Purification of SR protein splicing factors. *Methods Mol. Biol.*, **118**, 419–432.

Received October 10, 2000; revised May 18, 2001; accepted May 29, 2001

Note added in proof

While this paper was in review, YB-1 was shown to be associated with RNA polymerase II via binding to TLS or EWS, and to affect alternative splicing of an adenovirus E1A reporter construct. [Chansky,H.A., Hu,M., Hickstein,D.D. and Yang,Y. (2001) Oncogenic TLS/ERG and EWS/Fli-1 fusion proteins inhibit RNA splicing mediated by YB-1 protein. *Cancer Res.*, **61**, 3586–3590.]

Oncogene (1999) 18, 3574–3582

http://www.stockton-press.co.uk/onc

Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis

Elmar Stickeler[1], Frances Kittrell[2], Daniel Medina[2] and Susan M Berget*[,1]

[1]*Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030, USA;* [2]*Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030, USA*

Using a mouse model of mammary gland development and tumorigenesis we examined changes in both alternative splicing and splicing factors in multiple stages of mammary cancer. The emphasis was on the SR family of splicing factors known to influence alternative splicing in a wide variety of genes, and on alternative splicing of the pre-mRNA encoding CD44, for which alternative splicing has been implicated as important in a number of human cancers, including breast cancer. We observed step-wise increases in expression of individual SR proteins and alternative splicing of CD44 mRNA during mammary gland tumorigenesis. Individual preneoplasias differed as to their expression patterns for SR proteins, often expressing only a sub-set of the family. In contrast, tumors demonstrated a complex pattern of SR expression. Little difference was observed between neoplasias and their metastases. Alternative splicing of CD44 also changed through the disease paradigm such that tumors produced RNA containing a mixture of variable exons, whereas preneoplasias exhibited a more restricted exon inclusion pattern. In contrast, other standard splicing factors changed little in either concentration or splicing pattern in the same cells. These data suggest alterations in relative concentrations of specific splicing factors during early preneoplasia that become more pronounced during tumor formation. Given the ability of SR proteins to affect alternative processing decisions, our results suggest that a number of pre-mRNAs may undergo changes in alternative splicing during the early and intermediate stages of mammary cancer.

Keywords: SR splicing factors; alternative splicing; CD44; mammary tumorigenesis; mouse model

Introduction

Alternate pre-mRNA processing contributes significantly to the developmental regulation of gene expression in humans. It is estimated that as many as 25% of human genes utilize alternative RNA processing to produce subtly or grossly altered gene products (Moore *et al.*, 1993; Rio, 1993; Norton, 1994; Berget, 1995; Kramer, 1996). During the last year, reports have emerged indicating that cancerogenesis induces changes in alternative processing (Lee and Feinberg, 1997; Silberstein *et al.*, 1997; Zhu *et al.*, 1997). Some of the genes targeted for these changes are receptor proteins suggesting that the induced alterations in splicing could have pronounced consequences for cellular behavior. Little, however is known about how RNA processing factors are altered during tumorigenesis so as to affect processing outcomes.

The arginine-serine-rich (SR) proteins (Figure 1a) constitute a family of essential splicing factors (Krainer *et al.*, 1990a; Ge *et al.*, 1991; Zahler *et al.*, 1992) that recognize both splice sites and exonic splicing enhancers, and influence alternative processing decisions when their relative concentrations are altered *in vivo* or *in vitro* (Ge and Manley, 1990; Krainer *et al.*, 1990b; Zahler *et al.*, 1993a; Cáceres *et al.*, 1994; Wang and Manley, 1995). SR proteins have been observed to influence splicing activity via their binding to both splice sites and special splicing accessory sequences known as enhancers (Zahler *et al.*, 1993b; Fu, 1995; Manley and Tacke, 1996; Valcarcel and Green, 1996). Recent studies have indicated substrate-specific binding and activity for individual SR proteins (Fu, 1993; Sun *et al.*, 1993; Zahler *et al.*, 1993a; Cáceres *et al.*, 1994; Wang and Manley, 1995; Chandler *et al.*, 1997; Liu *et al.*, 1998). Furthermore, individual SR proteins have distinct tissue distributions (Zahler *et al.*, 1992; Screaton *et al.*, 1995; Cáceres and Krainer, 1997). These observations have led to suggestions that alterations in the levels of SR proteins could be determinative for alternative splicing during development. This class of proteins, therefore, become attractive candidates for factors whose activity changes during tumorigenesis.

An ideal system in which to study the relationship between processing and the progression of cancer would be one in which both normal and abnormal development could be compared and studied in an isogenic and manipulatable background. Such a system is available in an established *in vivo* mouse model of mammary tumorigenesis (Kittrell *et al.*, 1992; Medina, 1996) that provides access to normal mammary tissues from mature virgin, pregnant and lactating females, as well as an extensively characterized set of preneoplasias and their corresponding adenocarcinomas and metastases. In this model system, preneoplastic outgrowth lines are serially transplanted and maintained for extended periods of time (up to 12 months) in the mammary fat pads of syngenic female BALB/c mice. Each outgrowth line is characterized by a specific rate for development of adenocarcinomas and subsequent metastases (Kittrell *et al.*, 1992; Medina, 1996). The outgrowths are clonal cell populations as determined by oncogene analysis (Cardiff, 1988; Jerry *et al.*, 1993).

Using this system, we characterized SR protein expression during tumorigenesis. At the same time we

*Correspondence: SM Berget
Received 16 July 1998; revised 14 January 1999; accepted 14 January 1999

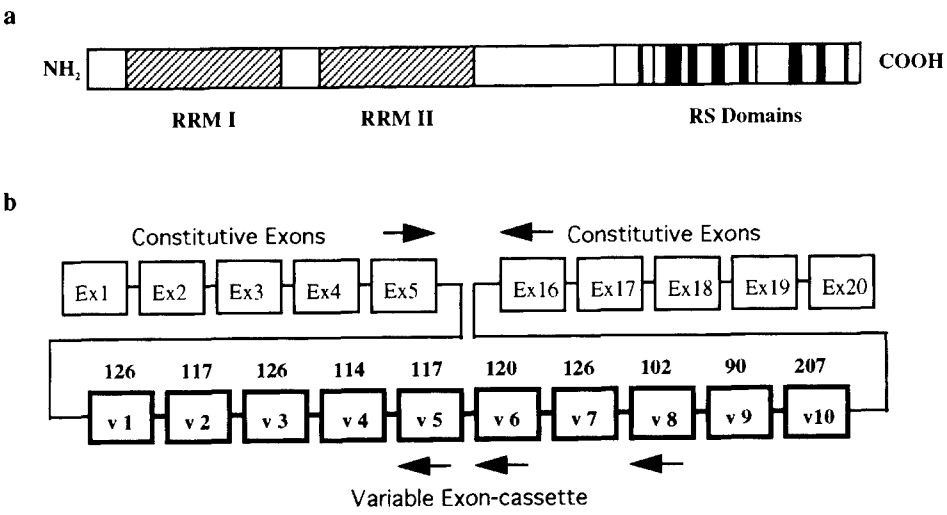


Figure 1 SR proteins and CD44 architecture. (**a**) Schematic of a generic SR protein with two N-terminal RRM (RNA recognition motif) motifs that bind RNA (hatched boxes) and extensive C-terminal RS (arginine-serine) domains (vertical lines). Thickness of vertical lines represents numbers of consecutive RS dipeptides. (**b**) CD44 exon structure. Constitutive exons 1–5 and 16–20 flank a central region encoding at least ten alternatively included (variable) exons (v1–v10). The size of each variable exon is indicated (drawing not to scale). Arrows indicate primers used for RT–PCR analysis

examined the alternative splicing of a pre-mRNA, that coding for CD44, whose splicing had been reported to be determinative for metastasis (Gunthert *et al.*, 1991). CD44, is a widely expressed cell adhesion molecule and trans-membrane glycoprotein that mediates a variety of cell-cell and cell matrix interactions (Fox *et al.*, 1994; Mackay *et al.*, 1994; Stamenkovic *et al.*, 1989). A number of CD44 isoforms are created via alternative pre-mRNA splicing. Within the CD44 gene is an internal cassette of ten alternatively spliced exons, all in the same translational reading frame, and coding for specific extracellular domains of the CD44 protein (Screaton *et al.*, 1992; Figure 1b). Combinations of these variable exons lead to a variety of CD44 isoforms (Screaton *et al.*, 1992; Haynes *et al.*, 1990; Gunthert, 1993). Certain isoforms, especially those including variable exons v5, 6, and 7, have been implicated in the metastasis of several malignancies (East and Hart, 1993; Cannistra *et al.*, 1995; Stickeler *et al.*, 1997; Wielenga *et al.*, 1993) and correlated with survival in human breast cancer (Joensuu *et al.*, 1993; Kaufmann *et al.*, 1995). Recently it has been shown that alternative splicing of CD44 responds to signal transduction. Treatment of T-lymphoma cells that normally express a form of CD44 lacking variable exons, with phorbol esters, concanavalin A, or c-*ras* induces inclusion of variable exons (König *et al.*, 1998). Thus, alternative splicing of CD44 may be involved in malignant transformation of tissues during tumorigenesis, and CD44 joins the collection of genes whose splicing is altered during cancer (Lee and Feinberg, 1997; Zhu *et al.*, 1997; Silberstein *et al.*, 1997).

Recently it has been shown that CD44 alternative splicing may involve inclusion of blocks of alternative exons rather than single isolated exons (Bell *et al.*, 1998). As the metastatic potential of CD44 has been correlated with only a single variable exon, it becomes of interest to see if the pattern of CD44 variable expression during tumorigenesis may have been underestimated by the heavy attention directed to this one exon.

Here we report that step-wise alterations in SR splicing factor levels and CD44 alternative splicing accompanied the transition from normal cells through preneoplasia and mammary adenocarcinoma, but not metastases. Individual preneoplasias exhibited differences in their SR expression pattern and limited CD44 alternative splicing. Alternative splicing of a control gene was not affected during preneoplasia or neoplasia, suggesting gene-specific alterations in RNA processing factors during early tumorigenesis, rather than a general increase in the generic processing machinery. Tumors showed high levels of SR proteins, including proteins never expressed in virgin mammary epithelia and high levels of inclusion of a variety of CD44 alternative exons. Our results suggest that alterations in alternative processing may be important features of early and intermediate stages of mammary cancer.

Results

For this study, a number of mammary preneoplasias and tumors were utilized, including tumors and metastases derived from two of the preneoplasias. The tumorigenic and metastatic properties of two of the preneoplastic outgrowth lines, TM-2L and TM-40, are summarized in Table 1. Both TM-2L and TM-40 produce type B mammary adenocarcinomas by 12 months, at an incidence of 21 or 51%, respectively. The main difference between the two lines is the significant metastatic growth, primarily in lungs and occasionally in liver, of TM-40 (Table 2).

Stepwise alterations in expression of SR proteins during development of mammary cancer

SR proteins are essential splicing factors that participate in early events of substrate recognition by the processing machinery (Fu, 1995; Zahler *et al.*, 1993a,b; Manley and Tacke, 1996; Valcarcel and

Table 1 Tumorigenic characteristics of TM preneoplastic outgrowth lines

Outgrowth line	*Transplant[a] generation*	*Tumors/ transplants (%)*	*TE_{50} (months)[b]*	*% Metastases[c]*
TM2L	21–28	21/98 (21)	9.5	0
TM40	4–13	80/157 (51)	11.0	73

[a]Number of times outgrowth lines were removed from fat pads and retransplanted into cleared fat pads. [b]TE_{50} refers to time for 50% of the transplants to form tumors. [c]Metastases were detected primarily in lung and secondarily in liver in mice 6–8 weeks after surgical removal of the primary tumors (11/15 animals). Metastases are rarely detected in animals bearing primary tumors

Table 2 Preneoplastic origin of adenocarcinomas and organ metastases

Tumor/metastases	*Outgrowth line*
T-5839	TM-2L
T-6788	TM-40
T-7780	TM-40
Liver metastases (T-7780)	TM-40
Lung metastases (T-7262)	TM-40

Green, 1996). In addition they bind to accessory splicing enhancer sequences to facilitate splicing of alternative exons (Ge and Manley, 1990; Cáceres *et al*., 1994; Wang and Manley, 1995). SR proteins are characterized by an extensive region containing arginine and serine (Figure 1a) which is extensively phosphorylated. Within this phosphorylated domain is the epitope for a monoclonal antibody, mAb104 (Zahler *et al*., 1993b; Valcarcel and Green, 1996). To examine alterations in splicing factors during development of mammary cancer, we examined the expression of SR proteins in normal and neoplastic mammary tissues by Western blots using mAb104. In HeLa cells, mAb104 recognizes approximately equal amounts of the major SR proteins (SRp20, ASF/SF2, SC35, 9G8, SRp30c, SRp40, SRp55, and SRp75 of molecular weights 20–75 kDa, Zahler *et al*., 1993b). Mammary epithelial cells from mature virgin glands expressed a sub-set of the major SR proteins, yielding dominant expression of SRp75 and SRp55 (Figure 2a, lane 1). Three of the preneoplasias, TM-2L, TM-4 and TM-10, yielded an SR expression pattern similar to that of mature, virgin mammary gland (Figure 2a, lanes 2, 4 and 5). The other tested preneoplasias, TM-3 and TM-40, in addition, demonstrated detectable levels of SRp40 and SR proteins of the 30 kDa class, suggesting induction of synthesis of these SR proteins during the preneoplastic process (Figure 2a, lanes 3 and 6). Tested adenocarcinomas and their metastases resembled the expression pattern of HeLa cells, characterized by expression of a broad spectrum of SR proteins detected by mAb104 (Figure 2a, lanes 7–9 and data not shown). We have tested a number of independently derived neoplasia for their SR expression profile (data not shown). All demonstrated SR expression patterns identical to those shown in Figure 2a, lanes 7–9, indicating that SR protein expression is induced during tumorigenesis regardless of the expression pattern during pre-neoplasia. Thus, the pattern of SR expression became more complex as neoplasia progressed until a complex array of SR proteins were expressed in tumor cells.

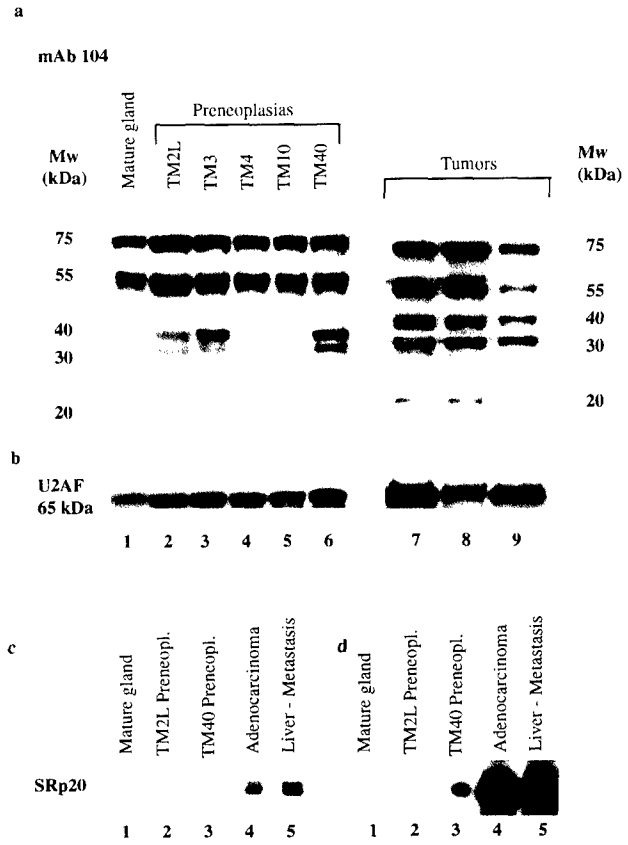


Figure 2 The expression of SR proteins is altered in mammary adenocarcinomas and in some preneoplasias. (**a**) Western blot analysis of SR protein expression in different breast tissues using an antibody specific for a phosphorylated SR epitope (mAb104). Lane 1, mature, virgin gland; lanes 2–6, preneoplasias TM-2L, TM-3, TM-4, TM-10 and TM-40, respectively; lanes 7–9 mammary adenocarcinomas T-5839, T-7780 and T-6129, respectively. Twice as much protein (100 μg) was analysed for normal and preneoplastic tissue (lanes 1–6) than for tumors (50 μg, lanes 7–9). (**b**) Western blot analysis of the same tissue analysed in (**a**) for the presence of the constitutive splicing factor U2AF using a polyclonal rabbit antibody raised against a recombinant fragment of the human 65 kDa subunit missing the SR domain (Zamore *et al*., 1992). Lanes as in (**a**). (**c**) Western blot analysis of selected tissues from (**a**) using an antibody specific for SRp20. The utilized antibody (Neugebauer and Roth, 1997) is an anti-peptide antibody specific for the region of SRp20 between the RNA binding domain and the SR domain. Lane 1, mature virgin gland; lane 2, preneoplasia TM2L; lane 3, preneoplasia TM40; lane 4, mammary adenocarcinoma T-5839; lane 5, liver metastasis. (**d**) Alterations in levels of mRNAs coding for splicing factor SRp20 during tumorigenesis. Levels of SRp20 revealed by RT/PCR using primers specific for SRp20 (see Materials and methods). Equal amounts (5 μg) of total cell RNA were used for each lane. Lanes as in (**c**). See Figure 3b for RT/PCR analysis of a control RNA in these RNA samples

The expression of several control RNA binding factors was monitored over the same disease paradigm. Individual preneoplasias showed identical levels of the 65 kDa subunit of U2AF (Figure 2b), hnRNP A1 (Figure 3b) and hnRNP A2 (Figure 4), as well as small nuclear ribonucleoproteins (snRNPs) (data not shown), indicating that there was not a general increase in the expression of factors involved in pre-mRNA splicing during preneoplasia. Tumors did show a slight increase (estimated to be twofold) for many factors as

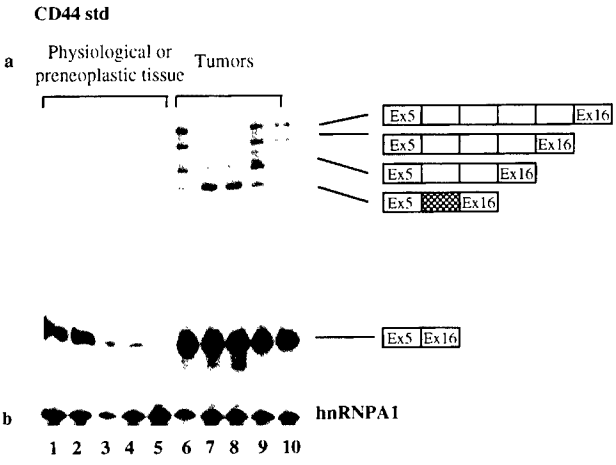


Figure 3 (**a**) Alternative splicing of CD44 changes during mammary development and neoplasia. RT–PCR analysis of CD44 mRNA isoforms in breast tissue using primers specific for constitutive exons 5 and constitutive exon 16. Amplification was for 25 cycles. Boxes symbolize included exons in the amplification bands. Exon structures of amplifications products are indicated. Hatching indicates mixtures of included variable exons. Lane 1, lactating gland; lane 2, pregnant gland; lane 3, mature virgin gland; lane 4, preneoplasia TM-2L; lane 5, preneoplasia TM-40; lanes 6–10, adenocarcinomas T-4031, T-4032, T-5839, T-6129, T-7780, respectively. All lanes used equal amounts (5 μg) of total cell RNA for amplification. (**b**) RT/PCR analysis of a control RNA coding for hnRNP A1 mRNA. Samples are identical to those in (**a**). The form of hnRNP AP shown is that coding for the version of the mRNA lacking exon 7b

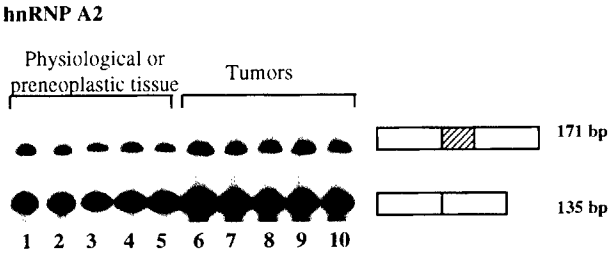


Figure 4 Alternative splicing of mRNA coding for hnRNP A2 does not change through mammary development or neoplasia. Low cycle RT–PCR of hnRNP A2 mRNA in mammary tissues using primers specific for two constitutive exons (white boxes) flanking a 36 nucleotide alternatively included exon (hatched box). Exon skipping produces a 135 nt band (bottom), inclusion in a 171 nt amplification product (top). Lane 1, lactating gland; lane 2, pregnant gland; lane 3, mature virgin gland; lane 4, preneoplasia TM-2L; lane 5, preneoplasia TM-40; lanes 6–10, adenocarcinomas T-4031, T-4032, T-5839, T-6129 and T-7780, respectively. Equal amounts of total cell RNA (5 μg) were analysed

compared to the preneoplasias, including the SR proteins SRp75 and SRp55 (Figure 2a). More interesting, however, was the pronounced induction of expression of the lower molecular weight SR proteins in the 20–40 kDa size range during neoplasia.

One of the SR proteins induced upon tumorigenesis was SRp20. This protein is only weakly detected by mAb104. By Western blotting, this protein was essentially undetectable in preneoplasias. Neoplasias weakly showed production of SRp20. To better examine induction of this factor we used an antibody specific for SRp20. This antibody was raised against an SRp20 peptide from the region of the protein located between the RNA-binding domain and the phosphorylated SR region (Neugebauer and Roth, 1997). As shown in Figure 2c, SRp20 levels detected by this antibody resembled those detected by the mAb104 antibody. No expression of SRp20 protein was detected in virgin or preneoplastic gland. In contrast, SRp20 was expressed in tumor tissue. To rule out the possibility of a false positive resulting from the presence of endothelial cells or infiltrating inflammatory cells in the excised adenocarcinomas, we analysed SR protein levels in cultured mammary epithelial cells established from the tumors in this study. In parallel with the tumours, each cell line demonstrated high levels of the entire family of SR proteins recognized by mAb104 (data not shown).

To provide more sensitive detection of SRp20, we monitored SRp20 mRNA levels using RT–PCR (Figure 2d). To control for non-linear estimations of mRNA levels using RT–PCR, SRp20 mRNA levels were also determined by RNase protection (data not shown). As observed by analysis with the Western blotting, tumorigenesis was accompanied by a major increase in the levels of SRp20 mRNAs, indicating that the increased levels of SRp20 in tumors seen with Western blotting reflected increased mRNA synthesis rather than increased phosphorylation of an existing protein population (Figure 2d).

RT–PCR revealed the presence of a low level of the tested SRp20 mRNAs in the TM-40 preneoplasia that had not been detected by Western blotting, suggesting that low level synthesis of SRp20 occurs in the TM-40 preneoplasia. Even using the power of PCR amplification, however, no SRp20 was detected in preneoplasia TM-2L or in mature virgin tissue. These results indicate that SRp20 is induced during development of mammary carcinogenesis. Furthermore, it indicates a progressive increase in expression as normal cells alter to preneoplasias and then to tumors.

The splicing of CD44 changes during progression of mammary cancer

Alteration in the relative levels of members of the SR proteins recognized by mAb104 is known to affect splicing phenotypes. Thus, the changes in overall amount and, more importantly, in relative amounts of the SR proteins displayed in Figure 2 should be accompanied by changes in alternative processing of target genes. To examine this possibility we looked at the splicing of CD44, a gene with multiple alternative exons. Some of these exons have sequences similar to the purine-rich splicing enhancers known to be the binding sites for the mAB104-reactive SR proteins (Black, 1995; Manley and Tacke, 1996; Reed, 1996; Rio, 1993; Bell *et al*., 1998; Liu *et al*., 1998). Splicing of CD44 was monitored by RT–PCR analysis of RNA from the same tissues used to examine SR proteins. CD44 mRNA is expressed as two basic forms, with or without one or more of the alternative internal exons. Expression of CD44 standard lacking all variable exons (CD44 std) was examined using primers complementary to constitutive exons 5 and 16, which border the central region of the CD44 gene encoding the ten alternative exons (Figure 1b). RT–PCR of CD44

standard mRNA should produce an amplification band of 221 nucleotides; inclusion of one or more variable exons should produce larger bands. Because all of the variable exons are approximately the same size (Figure 1b) inclusion of multiple variable exons produces an amplification pattern with a ladder of bands differing in length by the average variable exon length (approximately 115 nucleotides). This approach permits evaluation of the overall amount of CD44 mRNA present in examined tissues as well as an estimation of the extent of alternative splicing. Amplification revealed that mRNA encoding for CD44 std, lacking the variable exons, was expressed in all tissues examined (Figure 3). Tumor RNA from adenocarcinomas or their metastases demonstrated noticeable alternative processing to reveal a family of amplification bands resulting from the inclusion of multiple variable exons.

To distinguish CD44-specific effects on alternative splicing associated with neoplastic transformation from generic effects on splicing activity associated with an accelerated growth rate, we examined the amounts and spliced isoforms of an RNA that undergoes stochastic alternative splicing in many tissues. The hnRNP A2 gene is processed to produce two mRNA isoforms which alternatively include a 36 nucleotide variable exon. We observed relatively equal levels of total hnRNPA2 mRNA in all tissues examined and approximately equal relative levels of inclusion of the variable exon (Figure 4). This result suggests that the changes in CD44 splicing occurring in mammary neoplasia is due to a change in the sub-set of splicing factors required to recognize the CD44 alternative exons and not a change in the generic splicing machinery.

CD44 metastatic exons are expressed at different levels during normal mammary gland development and tumorigenesis

Low-cycle RT–PCR with primers specific for constitutive exons 5 and 16 produces complicated amplification patterns because of the potential complexity of the family of CD44 mRNA sequences. To better examine variable exons splicing, RNAs from tissue samples were analysed by RT–PCR using exon-specific primers for variable exons 5, 6, 7 or 8 and a consistent primer for constitutive exon 5 (Figure 5a–d), followed by cloning and subsequent sequencing of the PCR products. This approach allowed examination of the inclusion of all alternative exons in the interval v1–v8. Both low and high cycle amplifications were performed. We concentrated on analysis of species containing variable exons v5, 6 or 7, because these exons have been associated with metastasis in other studies (Stickeler *et al*., 1997; Heider *et al*., 1995; Sinn *et al*., 1995). Variable exon v8 was also of interest because it is one of three so-called 'epithelial ' exons, included in CD44 mRNA in a variety of epithelial cells.

Low cycle RT–PCR analysis of RNA from multiple tumors or their metastases using a primer specific for variable exon 5 (v5) indicated considerable production of RNAs containing v5 with or without other upstream variable exons (Figure 5a and data not shown). The inclusion patterns for all of the examined tumors and metastases were very similar, and indicated that most of the observed RNA included multiple variable exons from the v1–v5 region. In contrast, levels of inclusion of v5 were lower in normal dividing cells from pregnant or lactating mammary gland and in the two tested preneoplasias. Using 25–35 cycles of RT–PCR (Figure 5a and b), the major amplified species in these tissues included either v5 alone or v5 and one other upstream variable exon. Cloning and sequencing of the larger amplification band indicated that 4/5 sequenced clones contained v4 in addition to v5. This species was more obvious in pregnant and lactating RNA than in RNA from the preneoplasias (Figure 5a). We were unable to detect any RNA containing v5 in RNA isolated from mature virgin mammary gland (Figure 5a, lane 3; and 5b, lane 2). Because of the abundance of fat tissue in mature virgin gland, we isolated epithelial cells from the extracted virgin mammary tissue. High cycle amplification of this RNA yielded no demonstrable inclusion of exon v5 either (data not shown). Therefore, for v5 splicing, three patterns of inclusion were observed by RT–PCR and sequencing: high level of inclusion of multiple exons along with v5 in tumor cells, low level of inclusion of v5 or v5+4 in preneoplasias, and no inclusion of v5 in normal virgin mammary gland. This pattern suggests step-wise alterations in CD44 alternative splicing during neoplasia. In addition, our data suggest that preneoplasias adopt a splicing pattern similar to that induced in mammary epithelia during normal mammary development and that tumorigenesis is accompanied by induction of a pattern of frequent alternative splicing not found in normal development.

Analysis of alternative splicing of v6 or 7 gave results similar to that discussed above for v5 (Figure 5c and data not shown). Inclusion of both exons occurred at high frequency along with other upstream exons in tumor cell RNA. RNA from preneoplasias and pregnant gland demonstrated simpler RNAs containing fewer variable exons. Sequencing of PCR products indicated that for v6, these RNAs contained only v6. For v7, species containing v7 or v7+4 were observed. In contrast to v5, both v6 and 7 were included in RNA from mature virgin gland epithelial cells (Figure 5c, lane 5; and data not shown). Like v5, these observations suggest that splicing in preneoplasias is similar to splicing in normal proliferative mammary gland and that the conversion to neoplasia is accompanied by pronounced increased in mRNA including these variable exons. Therefore, for this system of mammary neoplasia, inclusion of v6 is not a marker for either tumorigenesis or metastasis. Instead, a better marker is increased inclusion of a number of CD44 variable exons including v6.

A recent report (Bell *et al*., 1998) indicated that inclusion of blocks of variable CD44 exons was a frequent phenotype in mice. Such multiple inclusion events are certainly consistent with the patterns we observed in the tumors. Preneoplasias and mature virgin mammary epithelial cells, however, demonstrated inclusion of individual exons without inclusion of an immediate neighbor. For example, v6 was included without v5 in mature gland and preneoplasias and v7 was included with v4 but not v6 and 5. If inclusion of multiple exons reflects a mechanism change with respect to inclusion of single exons, then

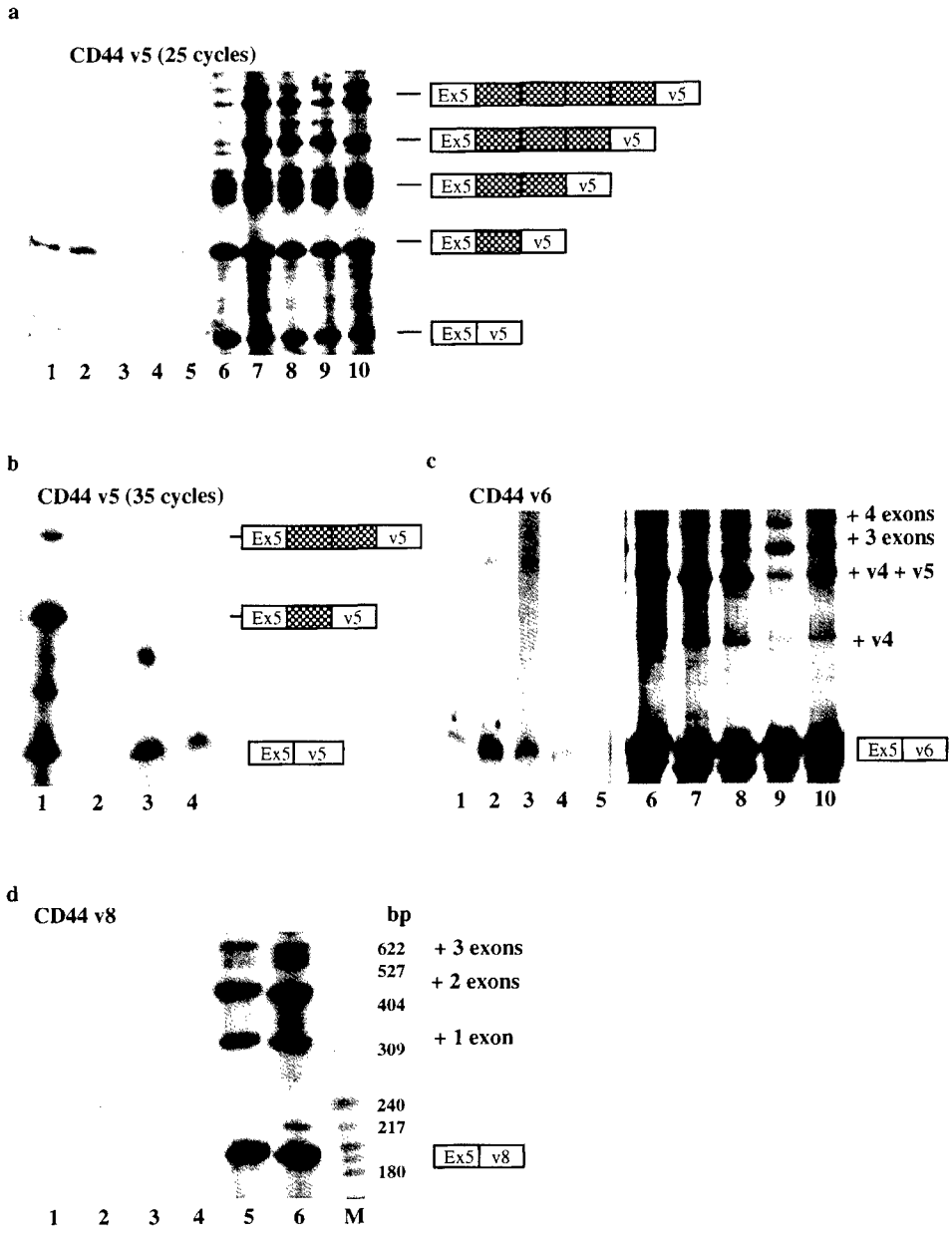


Figure 5 CD44 metastatic exons are expressed in normal development and tumorigenesis. RT–PCR analysis of CD44 mRNA variable isoforms using primers specific for constitutive exons 5 and variable exon v5 (**a** and **b**), variable exon v6 (**c**) or variable exon v8 (**d**). Amplification was for 25 (**a**, **c**, **d**) or 35 (**b**) cycles. Boxes symbolize included exons in the amplification bands. Exon structures of amplifications products are indicated. Hatching indicates mixtures of included variable exons. Majority products are indicated where they existed. (**a**, **c**) Lane 1, lactating gland; lane 2, pregnant gland; lane 3, mature virgin gland; lane 4, preneoplasia TM-2L; lane 5, preneoplasia TM-40; lanes 6–10, adenocarcinomas T-4031, T-4032, T-5839, T-6129 and T-7780, respectively. (**b**) Lane 1, pregnant gland; lane 2, mature virgin gland; lane 3, preneoplasia TM-2L; lane 4, preneoplasia TM-40. (**d**) Lane 1, lactating gland; lane 2, mature virgin gland; lane 3, preneoplasia TM-2L; lane 4, preneoplasia TM-40; lanes 5–6, adenocarcinomas T-4031 and T-4032. All lanes used equal amounts (5 μg) of total cell RNA for amplification

tumorigenesis is accompanied by this mechanism change.

Analysis of RNA for the presence of v8 yielded a surprising result. Like the other variable exons, v8 was included in a number of RNA species in mammary tumor cells (Figure 5d, lanes 5–6). Unlike v5–7, however, no v8 inclusion could be observed in preneoplasias or RNA from developing mammary, despite their epithelial origin (Figure 5d, lanes 3–4). The observation of induction of v8 splicing in tumor RNA indicates that tumorigenesis is accompanied by alterations in RNA processing of CD44 that are never seen in normal development.

Discussion

During the last several years, studies have suggested changes in pre-mRNA splicing in human malignancies (Lee and Feinberg, 1997; Zhu *et al*., 1997; Silberstein *et al*., 1997). Using an *in vivo* model of mammary development and tumorigenesis, we were able to study these questions with a focus on comparing splicing phenotypes and factors during normal mammary development to those occurring in development of mammary cancer. Neoplasia was accompanied by a dramatic increase in expression of some of the SR family of splicing factors recognized by mAb104,

resulting in alteration of the relative abundance of individual SR proteins and an increase in the complexity of expression of this important class of splicing factors. Given the number of genes whose splicing has been observed to be responsive to relative SR protein levels (Ge and Manley, 1990; Krainer *et al*., 1990b; Cáceres *et al*., 1994; Wang and Manley, 1995), this observation suggests that pronounced changes in alternative splicing of a number of pre-mRNAs should accompany mammary tumorigenesis.

The utilized model system also permitted examination of SR protein expression in preneoplasias with differing capacity to form tumors or for those tumors to metastasize. The preneoplasias characterized in this study differed in their SR protein expression pattern. Some of the neoplasias had mAb104-reactive SR expression patterns similar to that seen with virgin mammary tissue. The tested preneoplasias therefore appeared to represent different stages in the transition from normal to tumor tissue with respect to SR protein expression. The utilized preneoplasias arose independently and have different tumor and metastasis producing capabilities. SR expression levels did not correlate to tumor or metastatic incidence frequency. Thus, although SR expression changes may be a marker for preneoplasia, they are not predictive for tumor incidence or invasiveness.

Analysis of other constitutive splicing factors indicated that the changes in SR expression we observed were not the result of induction of all splicing during neoplasia. Levels of the constitutive factors U2AF65, two hnRNP proteins, and U snRNPs were only modestly altered during preneoplasia and neoplasia. Therefore, the alterations we observed occurred in a family of factors associated with alternative splicing and the recognition of exon enhancer sequences. This observation suggests that the splicing of multiple target genes regulated by SR proteins would be expected to be altered in neoplasia, without a general effect on exon inclusion levels.

We also used this model to investigate potential changes in alternative splicing in a gene whose splicing changes during cancer have been correlated to metastatic potential in human cancers and which has recently been shown to alter its splicing pattern in response to signal transduction (König *et al*., 1998). As with SR proteins, striking increases in alternative splicing of CD44 were observed in adenocarcinomas as compared to either preneoplasias or proliferative mammary epithelia from pregnant animals. All of the tumors examined displayed a similar pattern of exon inclusion, in which multiple alternative exons were included in mRNA. Included in this set of exons were exons not normally present in epithelia of pregnant mammary gland or in preneoplasias. The splicing pattern of the adenocarcinomas was similar to that of the lung and liver metastases of these tumors, indicating few changes accompanying metastasis. In contrast to CD44 alternative splicing, the splicing of a control RNA did not change during the transition from preneoplasia to neoplasia. This difference suggests that the pronounced changes in splicing of CD44 observed during development of neoplasia arise from the induction of specialized splicing factors, not an increase in the generic splicing machinery.

Although pronounced changes in CD44 splicing correlated with alterations in the expression level of SR proteins during the tumorigenesis paradigm used in this study, it is not yet clear if the two are related. Transfection of cells that normally skip CD44 variable exons 4 and 5 with cDNAs for SRp55 and ASF/SF2 had little effect on exon inclusion in a co-transfected mini-gene containing these two exons (data not shown). In contrast, transfection with cDNAs for both SC35 and tra2 did alter the percentage of exon inclusion (Stickeler and Berget, manuscript in preparation). Therefore, the splicing of CD44 may be regulated by a sub-set of the SR family of splicing factors.

When we examined CD44 variable exon splicing in epithelia from pregnant mammary gland, we observed low levels of inclusion of some but not all the variable exons. In particular, v6 and 7 were present in a subpopulation of CD44 mRNA from both pregnant tissue and preneoplasias, indicating that these so-called 'metastatic' exons are included in non-cancerous mammary epithelial cells. In contrast, one variable exons, v5, was not expressed in virgin mammary gland, indicating that alterations in CD44 alternative splicing can occur during preneoplasia and resemble those occurring during pregnancy. Other variable exons that are not included in pregnant tissue (such as the 'epithelial' exon v8) are also not expressed during preneoplasia, but are included in tumor RNA. Therefore, our analysis suggests that preneoplasias demonstrate a splicing pattern resembling that occurring during normal breast development. Known to participate in cell-cell interactions and cell-matrix interactions, CD44 isoforms could be involved in the ductal outgrowth occurring during early pregnancy.

Inclusion of three of the CD44 variable exons (v5, 6 and 7) have been postulated to be markers for metastasis potential in several human cancers, including breast cancer. The initial interest in these exons arose from an elegant experiment demonstrating that transfection of a CD44 cDNA including exon v6 into a tumor cell of no metastatic potential increased the rate of metastasis when these cells were placed back into syngenic animals. Such experiments expressed the variant CD44 mRNA at high levels. Our analysis would suggest that the major change in CD44 splicing occurs during development of adenocarcinomas, not their metastases and that it may be the over-expression of variant CD44 mRNAs during the neoplastic state which is important for the metastatic process, rather than the production of any mRNA including v6. This difference may suggest why experiments monitoring inclusion of CD44 variable exons have failed to demonstrate a consistent correlation to disease progression in several clinical studies (Friedrichs *et al*., 1995; Muller *et al*., 1997; Tran *et al*., 1997).

The recent report of a coupling between signal transduction and alternative splicing of CD44 concentrated on variable exon v5. Inclusion of this exon showed a pronounced response to induction by c-*ras*, TPA, or phorbol esters (König *et al*., 1998). Exon v5 contains internal sequences that resemble known purine-rich exon enhancers that bind SR proteins. These observations suggest that signal transduction cascades can lead to an induction of SR proteins needed for CD44 variable splicing. In light of these findings, our observation that inclusion of v5 is

changed upon preneoplasia supports a possible connection between SR protein levels and CD44 alternative splicing.

Besides CD44 a number of other genes show altered RNA processing during tumorigenesis. A number of cellular receptors and hormone genes undergo alternative processing and this processing appears to alter during cancer. Because the changes in these proteins caused by alternative processing can have a pronounced effect on cellular function, it is important to understand the role of alternative processing in cancer and the mechanisms involved. Determination of splicing factor constellations patterns in preneoplastic lesions of the mammary gland could be very helpful to identify patients in high-risk situations to develop invasive breast cancer and subsequent organ metastasis. Distinct expression patterns of splicing factors could serve as new markers for metastasis in breast cancer, and splicing factors may represent targets for intervention in a subgroup of patients abnormally expressing certain factors.

Materials and methods

Tissues

We used an *in vivo* mouse model of mammary development and tumorigenesis to study normal tissue, pregnant and lactating glands, preneoplastic lesions, eight primary adenocarcinomas and organ metastases to the liver and lung. The TM preneoplastic outgrowth lines developed after transplantation of established mouse mammary epithelial cell lines (MMEL) into cleared mammary fat pads of 3-week-old syngenic BALB/cMed mice (Kittrell *et al.*, 1992). The serially-transplanted outgrowths were removed either as preneoplasias at 8–12 weeks after transplantation or as tumors 5–7 months after transplantation. After removal, the preneoplastic outgrowths, primary adenocarcinomas, and metastases were frozen at −80°C for further analysis. The biological properties of the outgrowth lines are shown in Table 1.

Epithelial cell isolation from mature virgin mammary tissue

Isolated mammary fat pads were minced and incubated in DMEM medium with 2 mg/ml collagenase A (Boehringer Mannheim, Germany) and 100 U/ml hyaluronidase (Boehringer Mannheim, Germany) at 37°C for 3 h and slowly swirled. Afterwards the solution was centrifuged for 5 min at 1000 r.p.m. The supernatant, containing fat and single cells, was discarded and the pellet, containing mammary epithelial cells, was washed in PBS/5% FCS and stored at −80°C for further analysis.

RNA isolation, RT–PCR and sequencing

Tissues were minced on dry ice before they were treated with a tissue homogenizer (Polytron, Littau, Switzerland) in TRIzol solution (Gibco-BRL, Gaithersburg, MD, USA) with 1 ml per 100 mg of tissue for three times 10 s. After 5 min of incubation at 26°C, 0.2 ml of chloroform per 1 ml of TRIzol were added and the solution was shaken vigorously, followed by incubation at 26°C for 5 min. After centrifugation for 15 min at 4°C and 12 000 *g* the aqueous phase, containing the total RNA, was precipitated with 0.5 ml of isopropyl alcohol per 1 ml of TRIzol for 10 min at 26°C and centrifuged for 10 min at 12 000 *g* at 4°C. The pellets were rinsed with 75% ethanol (1 ml of ethanol per 1 ml of TRIzol reagent used for the initial homogenization), vortexed and centrifuged at 4°C for 5 min at 7500 *g* and finally dissolved in RNAse free water and stored at −80°C for further analysis.

Five micrograms of RNA were used for cDNA synthesis using M-MuLV reverse transcriptase (Perking Elmer, Branchburg, NJ, USA) and oligo-dT primers followed by PCR using primers specific for different SR proteins:

SRp20: 5′-GCCGTGTAAGAGTGG,
3′-AAGCTTCCTCCTTCTTGG;
hnRNP A1: 5′-GGTGGTCGTGGAGGTGGTT,
3′-CCAAAATCATTGTAGCTTCC, and
hnRNP A2: 5′-ACAGTCTGTAAGCTTTCCCC,
3′-CTGAAGCGACTGAGTCCGCG.

The standard form of CD44 mRNA (CD44 std) was detected using PCR primers for constitutive exons 5 and 16:

CD44 exon 5: 5′-ACCCCAGAAGGCTACATTTTGC
CD44 exon 16: 3′-CTCATAGGACCAGAAGTTGTGG

Inclusion of individual alternative exons was monitored using 3′ primers specific for individual alternative exons and a primer complementary to constitutive exon 5:

CD44v5: (MASV5) 3′-TTGTAGCATGTGGGGTCTCCTCTT,
CD44v6: (MASV6) 3′-CCTTCTGTCACATGGGAGTCTTCA,
CD44v8: (MASV8) 3′-CACTGAAAGTGGTC CTGTCCTGTT.

All amplification reactions used an ATP-labeled 5′ primer. The PCR conditions for all amplifications were as follows: 25–35 cycles (see Figure legends) of 94°C for 1 min, 59°C for 1 min and 72°C for 1.5 min. Gel electrophoresis of 35 μl of the 100 μl PCR reaction was performed in 6% denaturing acrylamide gels at 34–40 mA. Expected amplification products are as follows: 142 bp (SRp20), 155 bp (hnRNP A1 without alternative exon 7B), 134 and 170 bp (hnRNP A2 with or without the alternative exon), and 221 bp (CD44 standard lacking any alternative exon). The expected amplification bands for alternatively-spliced CD44 mRNA containing only one variable exon are 207 bp (v5), 209 bp (v6) and 190 bp (v8). pBR 322/*Hpa*II markers were used in all displayed gels for product size determination. Identified amplification products resulting from the inclusion of one or two variable CD44 exons were sequenced to verify identity. After cloning of CD44 PCR products into a pCR 2.1 vector (TA cloning kit, Invitrogen, CA, USA) sequencing of these clones was performed using the thermo sequenase method (Amersham, OH, USA) and CD44 exon specific primers mentioned above.

Protein isolation and Western blot

Total cellular protein was isolated from the interphase and phenol phase from the initial homogenate after precipitation of the DNA with 0.3 ml 100% of ethanol per 1 ml of TRIzol used for the initial homogenization. The samples were stored at 26°C for 5 min and afterwards centrifuged at 2000 *g* for 5 min at 4°C. The phenol-ethanol supernatant was then precipitated with 1.5 ml isopropyl alcohol per 1 ml TRIzol used for initial homogenization. After storage for 10 min at 26°C the protein precipitates were centrifuged at 12 000 *g* at 4°C for 10 min. Protein pellets were washed three times in 0.3 M guanidine hydrochloride in 95% ethanol for 20 min at 26°C and centrifuged at 7500 *g* for 5 min at 4°C. The protein pellets were vortexed after the final wash in 2 ml 100% ethanol, stored for 20 min at 26°C and finally centrifuged at 7500 *g* for 5 min at 4°C. The pellets were resuspended in 1% SDS solution and incubated at 50°C for complete dissolution. Insoluble material was removed by centrifugation at 10 000 *g* for 10 min at 4°C. Supernatants were stored for further analysis at −80°C. Gel electrophoresis of 20 μg of total protein was performed using a 10%

SDS-PAGE gel at 100 volts. Afterwards the gels were electroblotted on a PVDF transfer membrane (PolyScreen, NEN Life Science, Boston, MA, USA) at 100 volts for 2.5 h at 4°C. After blocking in 5% Blotto/PBST membranes were incubated with mouse IgM mAb104 (1:5 dilution), the SRp20 specific mouse IgG Ab (1:200 dilution) or polyclonal rabbit anti-U2AF Ab 1:6000 dilution). The Western blots were stained by chemiluminescence (NEN Life Science) using appropriate horseradish-peroxidase labeled anti IgM antibody for mAb104 (Pierce, Rockford, IL, USA) or anti IgG antibody for SRp20 Ab and U2AF Ab a dilution of 1:3000.

Acknowledgements

This work was supported by Grant Sti 153/1-1 from the German Research Foundation (to E Stickeler), Grant DAMD 17-96-1-0684 from the US Army (to SM Berget) and Grant NCI CA 47112 from the National Cancer Institute (to D Medina).

References

Bell MV, Cowper AE, Lefranc MP, Bell JI and Screaton GR. (1998). *Mol. Cell. Biol.*, **18,** 5930–5941.

Berget SM. (1995). *J. Biol. Chem.*, **270,** 2411–2414.

Black DL. (1995). *RNA*, **1,** 763–771.

Cáceres JF, Stamm S, Helfman DM and Krainer AR. (1994). *Science*, **265,** 1706–1709.

Cáceres JF and Krainer AR. (1997). In: *Mammalian pre-mRNA splicing factors*. Krainer AR (ed.). IRL Press: Oxford, pp. 174–212.

Cannistra S, Abu-Jawdeh G and Niloff J. (1995). *J. Clin. Oncol.*, **13,** 1912–1921.

Cardiff RD. (1988). *Anticancer Res.*, **8,** 925–933.

Chandler SD, Mayeda A, Yeakley JM, Krainer AR and Fu X-D. (1997). *Proc. Natl. Acad. Sci. USA*, **94,** 3596–3601.

East JA and Hart IR. (1993). *Eur. J. Cancer*, **29A,** 1921–1922.

Fox SB, Fawcett J, Jackson DG, Collins I, Gatter KC, Harris AL, Gearing A and Simmons D. (1994). *Cancer Res.*, **54,** 4539–4546.

Friedrichs K, Kugler G, Franke F, Terpe HJ, Arlt J, Regidor PA and Gunthert U. (1995). *Lancet*, **345,** 1237.

Fu X-D. (1993). *Nature*, **365,** 82–85.

Fu X-D. (1995). *RNA*, **1,** 663–680.

Ge H and Manley JL. (1990). *Cell*, **62,** 25–34.

Ge H, Zuo P and Manley JL. (1991). *Cell*, **66,** 373–382.

Gunthert U, Hoffmann M, Rudy W, Reber S, Zoller M, Haussmann I, Matzku S, Wenzel A, Ponta H and Herrlich P. (1991). *Cell*, **65,** 13–24.

Gunthert U. (1993). *Curr. Microbiol. and Immunol.*, **184,** 47–63.

Haynes BF, Hua-Xin L and Patton KL. (1990). *Cancer Cells*, **3,** 347–350.

Heider KH, Mulder JW, Ostermann E, Susanni S, Patzelt E, Pals ST and Adolf GR. (1995). *Eur. J. Cancer*, **31A,** 2385–2391.

Jerry DJ, Ozbun MA, Kittrell FS, Lane DP, Medina D and Butel JS. (1993). *Cancer Res.*, **53,** 3374–3381.

Joensuu H, Klemi P, Toikkanen S and Jalkanen S. (1993). *Am. J. Path.*, **143,** 867–874.

Kittrell FS, Oborn CJ and Medina D. (1992). *Cancer Res.*, **52,** 1924–1932.

Krainer AR, Conway GC and Kozak D. (1990a). *Genes Dev.*, **4,** 1158–1171.

Krainer AR, Conway GC and Kozak D. (1990b). *Cell*, **62,** 35–42.

Kramer A. (1996). *Annu. Rev. Biochem.*, **65,** 367–409.

Kaufmann M, Heider KH, Sinn HP, von Minckwitz G, Ponta H and Herrlich P. (1995). *Lancet*, **345,** 615–619.

König H, Ponta H and Herrlich P. (1998). *EMBO J.*, **17,** 2904–2913.

Lee MP and Feinberg AP. (1997). *Cancer Res.*, **57,** 3131–3134.

Liu H-X, Zhang M and Krainer AR. (1998). *Genes Dev.*, **12,** 1990–2012.

Mackay C, Terpe HJ, Stauder R, Marston WL, Stark H and Gunthert U. (1994). *J. Cell. Biol.*, **124,** 71–82.

Manley J and Tacke R. (1996). *Genes Dev.*, **10,** 1569–1579.

Medina D. (1996). In: *Mammary Tumor Cell Cycle, Differentiation and Metastases*. Dickson RB and Lippman ME. (eds). Academic Publishers: Norwell, pp. 37–69.

Moore MJ, Query CC and Sharp PA. (1993). In: *The RNA World*. Gesteland RF and Atkins JF. (eds). Cold Spring Harbor Laboratory Press: New York, pp. 303–357.

Muller W, Schneiders A, Heider KH, Meier S, Hommel G and Gabbert HE. (1997). *J. Pathol.*, **183,** 222–227.

Neugebauer KM and Roth M. (1997). *Genes Dev.*, **11,** 1148–1159.

Norton PA. (1994). *J. Cell. Sci.*, **107,** 1–7.

Reed R. (1996). *Curr. Opin. Genet. Dev.*, **6,** 215–220.

Rio DC. (1993). *Curr. Opin. Genet. Dev.*, **3,** 574–584.

Screaton GR, Bell MV, Jackson DG, Cornelis FB, Gerth U and Bell JI. (1992). *Proc. Natl. Acad. Sci. USA*, **89,** 12160–12164.

Screaton GR, Cáceres JF, Mayeda A, Bell MV, Plebanski M, Jackson DG, Bell JI and Krainer AR. (1995). *EMBO J.*, **14,** 4336–4349.

Silberstein GB, Van Horn K, Strickland P, Roberts Jr CT and Daniel CW. (1997). *Proc. Natl. Acad. Sci. USA*, **94,** 8132–8137.

Sinn HP, Heider KH, Skroch-Angel P, von Minckwitz G, Kaufmann M, Herrlich P and Ponta P. (1995). *Breast Cancer Res. Treat.* **36,** 307–313.

Stamenkovic I, Amiot M, Pesando JM and Seed B. (1989). *Cell*, **56,** 1057–1062.

Stickeler E, Runnebaum IB, Moebus VJ, Kieback DG and Kreienberg R. (1997). *Anticancer Res.*, **17,** 1871–1876.

Sun Q, Mayeda A, Hampson RK, Krainer AR and Rottman FM. (1993). *Genes Dev.*, **7,** 2598–2608.

Tran TA, Kallakury BV, Sheehan CE and Ross JS. (1997). *Hum. Pathol.*, **28,** 809–814.

Valcarcel J and Green MR. (1996). *Trends Biochem. Sci.*, **21,** 296–301.

Wang J and Manley JL. (1995). *RNA*, **1,** 335–346.

Wielenga VJ, Heider KH, Offerhaus GJ, Adolf GR, van den Berg FM, Ponta H, Herrlick P and Pals ST. (1993). *Cancer Res.*, **53,** 4754–4756.

Zamore PD, Patton JG and Green MR. (1992). *Nature*, **355,** 609–614.

Zahler AM, Lane WS, Stolk JA and Roth MB. (1992). *Genes Dev.*, **6,** 837–847.

Zahler AM, Neugebauer KM, Lane WS and Roth MB. (1993a). *Science*, **260,** 219–222.

Zahler AM, Neugebauer KM, Stolk JA and Roth MB. (1993b). *Mol. Cell. Biol.*, **13,** 4023–4028.

Zhu X, Daffada AA, Chan CM and Dowsett M. (1997). *Int. J. Cancer*, **72,** 574–580.

DEPARTMENT OF THE ARMY
US ARMY MEDICAL RESEARCH AND MATERIEL COMMAND
504 SCOTT STREET
FORT DETRICK, MARYLAND 21702-5012

REPLY TO
ATTENTION OF:

MCMR-RMI-S (70-1y) 8 Jan 2003

MEMORANDUM FOR Administrator, Defense Technical Information Center (DTIC-OCA), 8725 John J. Kingman Road, Fort Belvoir, VA 22060-6218

SUBJECT: Request Change in Distribution Statement

1. The U.S. Army Medical Research and Materiel Command has reexamined the need for the limitation assigned to the enclosed. Request the limited distribution statement for the enclosed be changed to "Approved for public release; distribution unlimited." These reports should be released to the National Technical Information Service.

2. Point of contact for this request is Ms. Judy Pawlus at DSN 343-7322 or by e-mail at judy.pawlus@det.amedd.army.mil.

FOR THE COMMANDER:

Encl

PHYLLIS M. RINEHART
Deputy Chief of Staff for
Information Management

ADB265840	ADB266633	ADB282069
ADB279138	ADB251763	ADB265386
ADB264578	ADB281601	ADB282057
ADB281679	ADB258874	ADB258251
ADB281645	ADB281773	ADB264541
ADB261128	ADB281660	ADB241630
ADB261339	ADB259064	ADB281924
ADB273096	ADB266141	ADB281663
ADB281681	ADB281664	ADB281659
ADB259637	ADB258830	
ADB256645	ADB266029	
ADB262441	ADB281668	
ADB281674	ADB259834	
ADB281771	ADB266075	
ADB281612	ADB281661	