Available online at www.sciencedirect.com
SCIENCE DIRECT
Biochemical and Biophysical Research Communications 343 (2006) 167–170
BBRC
www.elsevier.com/locate/ybbrc

Identification of novel splice variants of the human CD44 gene

Elena Vela [a,b], Xavier Roca [a,b,1], Marcos Isamat [a,*]

[a] *Fundación Echevarne, Barcelona, Spain*
[b] *Servei d'Anatomia Patològica, Hospital Universitari Germans Trias i Pujol, Badalona, Barcelona, Spain*

Received 20 February 2006
Available online 2 March 2006

Abstract

The CD44 gene contains 10 variable exons (v1–v10) that can be alternatively spliced to generate hundreds of different CD44 protein isoforms, several of which have been implicated in the metastatic spread of tumour cells. Here, we describe a cryptic splice site, in intron 6 of the human CD44 gene, used during mRNA processing. This cryptic splice site is used in conjunction with variable exon 3, or independently from it in the form of a *pseudo*-exon of 49 bp, which generates a stop codon by frame shift in the contiguous variable exon downstream. This *pseudo*-exon has been found inserted immediately 3′ to any other variable exon from v4 to v10, in the final CD44 mRNA. The implication of this cryptic splice site in haltering CD44 protein translation is questioned in the context of Nonsense Mediated Decay and the overall regulation of CD44 expression.

Keywords: CD44; Alternative splicing; Cryptic splice site; NMD

Alternative pre-mRNA splicing is one of the processes by which multiple mRNA isoforms can be generated from a single gene. It is a widespread mechanism for regulating gene expression in higher eukaryotes. Variability in splicing patterns is a major source of protein diversity from the genome [1]. Alternative splicing regulation is mediated by the interaction of positive and negative regulatory elements present in the RNA, which are recognized by different trans-factors, some of which are members of the hnRNP and SR protein families [2].

Several genes encode transcripts that are alternatively spliced to produce different mRNAs. The CD44 gene is one of the best studied models of loci that undergo complex alternative splicing giving rise to a diversity of proteins. CD44 is a transmembrane protein involved in cell–cell adhesion and signal transduction. All CD44 proteins are encoded by a single, highly conserved gene. CD44 transcripts are subject to alternative splicing of its 10 variable exons [3]. Complex alternatively spliced CD44 messages have been described as the result of a cellular response to particular stimuli both in normal and pathological conditions [4,5]. One such condition is the breast duct carcinoma model wherein CD44 follows two major routes of alternative splicing [6]. The first route, observed in transformed breast tissue, affects the whole of the central region of the CD44 gene composed of 10 variable exons (v-exons, v1–v10) and inserts v-exons in a 5′-to-3′ directional manner, so that the probability of inclusion of a certain v-exon is higher if the other v-exons downstream from it have also been included in the mRNA. The second splicing route concerns v3 inclusion specifically and is operational both in normal and transformed breast tissue. Here, we present the identification of a novel 3′ splice site (3′ss) in intron 6, 49 nucleotides upstream from the classical 3′ss of exon v3. We also show the use of the (−49)3′ss during CD44 processing in human biopsy material and established cell lines.

Materials and methods

Clinical samples and cell lines. Cryopreserved (−80 °C) human tissue from invasive ductal carcinoma and fibroadenoma of the breast were randomly selected from the frozen tissue bank of the Department of

* Corresponding author.
E-mail address: misamat@echevarne.com (M. Isamat).
[1] Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.

0006-291X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2006.02.128

Pathology at the Hospital Universitari Germans Trias i Pujol, Badalona, Spain. MCF-7, HT-29, HCT-116, SW48, and LoVo cell lines were obtained form ATCC.

CD44 v3 minigene constructs. A fragment of genomic CD44 sequence (GenBank Accession No. NT009237) from the SKBr3 human breast carcinoma cell line, containing exon v3 and corresponding introns 6 and 7, was amplified by PCR with primers containing XbaI restriction sites, I6F3X (5′-CTTTCAGCTTTCTAGAAGAGAATATTCAGTT-3′) and I7R1X (5′-ATATTTTCCCCAACTTCTAGAAACATTCTA-3′), from the nucleotide 1012 in intron 6 to nucleotide 160 in intron 7. PCR products were cloned into a pUC18 vector with SureClone Ligation Kit (Pharmacia). The genomic inserts were enzymatically restricted with XbaI and inserted in the multiple cloning site of the Exon Trap vector (Mobitec), resulting in the pET v3construct. Plasmid DNA was recovered using Plasmid Mini Kit (Qiagen), cycle sequenced, and analyzed by ABI PRISM 310 Genetic Analyzer (Applied Biosystems).

CD44 RT-PCR. RT-PCR was performed from 3 μg of total RNA extracted from cryopreserved breast biopsy samples with the Ultraspec RNA solution (Biotecx Laboratories). The first-strand reaction was performed with the first-strand synthesis kit (Pharmacia). Total CD44 isoforms were amplified by PCR using forward primer in CD44 exon 5 (5′-CCTGAAGAGATCTACCCCAGCAACCCTACTG-3′) and reverse primer in CD44 exon 19 (5′-TGGTGCGGCCGTTACACCCCAATCTT CATGTCC-3′). PCR cycling conditions were 1 cycle at 94 °C for 4 min, 50 °C for 4 min, and 72 °C for 2.5 min and 35 cycles at 94 °C for 40 s, 55 °C for 70 s, and 72 °C for 2.5 min using Taq polymerase Eurobiotaq (Eurobio). CD44-specific PCR products were separated by electrophoresis in a 1% agarose gel, visualized by ethidium bromide staining, denatured in 0.5 N NaOH, 1.5 mol/L NaCl, and transferred overnight in the same solution to a nylon membrane by standard methods. Filters were prehybridized in 7% SDS, 0.5 mol/L phosphate buffer, and 1% bovine serum albumin, and hybridized overnight in the same solution at 42 °C with 5 pmol of digoxigenin-labeled oligonucleotide using the DIGoligonucleotide labeling and luminescent detection kit (Roche) according to the manufacturer's protocol.

Analysis of NMD. MCF-7 cells were treated with cycloheximide (CHX) at 100 μg/ml. After 2 h, cells were harvested and semiquantitative RT-PCR was performed on total RNA to determine the abundance of each isoform.

Results and discussion

Novel CD44 isoforms expressed in breast cancer cells

In our previous studies of CD44 v3 expression in a breast cancer model, we detected the presence of novel CD44 isoforms. Each amplicon, of the RT-PCR product, was cloned and sequenced. These isoforms derived from the use of a novel 3′ splice site (3′ss) in intron 6, 49 nucleotides upstream from the classical 3′ss of exon v3. No mutation was detected in the exon-intron boundaries and intron 6. We have classified this splice site as cryptic due to of its lesser usage relative to v3 3′ss, although normally the term "cryptic" refers to sites that are not used unless the authentic splice site is altered or mutated. In turn, the 3′ss upstream of v3 overlaps with a putative 5′ splice site (5′ss) sequence at the 5′ end of this exon, named (−49)5′ss.

Theoretical calculations of the relative forces of all these splice sites, using position-weight matrices [7,8], show that both (−49)5′ss and 3′ss have higher scores than those of the splice sites around the v3 exon, see Fig. 1. These results suggest that (−49)3′ss has sufficient theoretical strength to perform as a bona fide 3′ss, and likewise, the 5′ end of v3

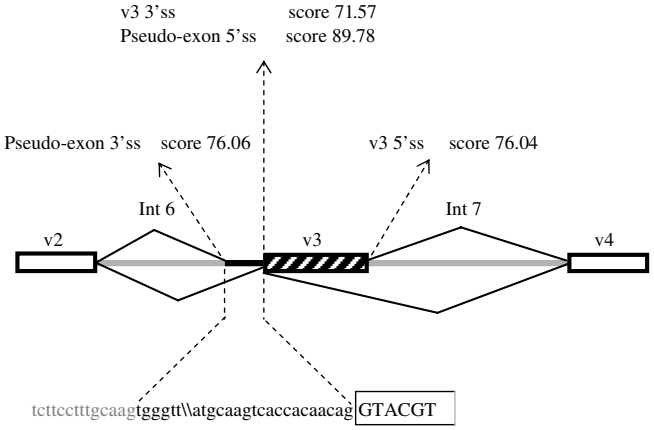

Fig. 1. Schematic representation of the v3 exon. Theoretical values of splicing sites strength as donor (5′ss) or acceptor (3′ss) according to position-weight matrices [7,8] are shown with arrows. Intron-junction sequences, including intron 6 (lower case), 49 bp *pseudo*-exon (lowercase, black), and v3 exon (uppercase) are indicated.

can act both as a donor and as an acceptor splice site. Accordingly, the sequence between (−49)3′ss and (−49)5′ss has been termed *pseudo*-exon (49 bp), whose inclusion in the CD44 mRNA depends on the use of the (−49)3′ss.

In order to test these values functionally, a construct was generated that included intron 6, exon v3, and intron 7 in the context of constitutive exons 1 and 2 of the human insulin gene in an Exon Trap vector (Mobitec) (Fig. 2A). The results of the transient transfection into the breast cell line MCF-7 and RT-PCR of the spliced products show the use of the cryptic (−49)3′ss and the capacity of the V3-3′ss to behave as both donor and acceptor during splicing (Fig. 2B).

Pseudo-exon expression in human cell lines and tumor samples

In order to analyze the presence of these novel variants in MCF-7, we performed a quantitative RT-PCR

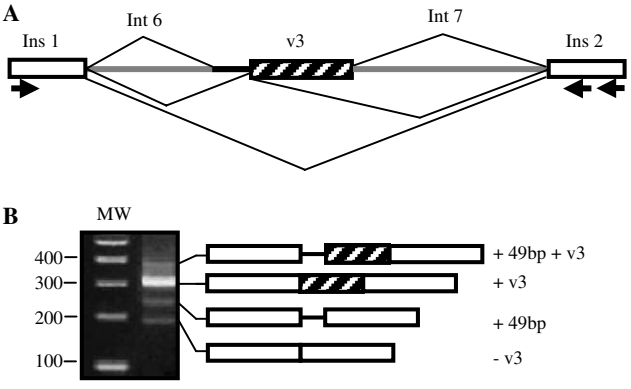

Fig. 2. Transfection of *pseudo*-exon containing v3 in MCF-7 cells. (A) v3 minigene exon trap construct illustrating positions of commercial exon trap (Mobitec) RT and PCR primers on Insulin 1 and Insulin 2 exons. CD44 intron 6 containing the 49 bp *pseudo*-exon, v3 and intron 7 and possible splicing routes amongst them are indicated. (B) RT-PCR from v3-exon trap construct showing splicing products in transfected MCF-7 cells.

with primers that amplified total CD44 isoforms followed by hybridization with oligonucleotide probes located inside the *pseudo*-exon and in constitutive exon 16. Comparison of each resulting isoform with cloned controls showed that −49 inclusion is a rare event (data not shown).

This result prompted confirmation of such splice site usage in the endogenous CD44 pre-mRNA from biopsy samples and several human cell lines using a strategy of RT-PCR that enables the amplification of all alternatively spliced CD44 mRNAs expressed from the endogenous CD44 gene [6]. This was performed on RNA extracted from breast carcinoma biopsy material (invasive ductal carcinoma and fibroadenoma) to visualize the whole range of CD44 messages produced. Following this, hybridization with probes specific for constant exon 16, variable exons v3, v4, v6, and v8, and the intronic sequence contained within the 49 bp *pseudo*-exon reveals in vivo inclusion of the *pseudo*-exon in isoforms with and without v3 (v3 − v10, v4 − v10, v3 + v6 − v10, v3 + v8 − v10, v8 − v10, and v3) in breast carcinoma samples and in MCF-7 (Fig. 3). This inclusion has been also detected in colon carcinoma cell lines (HT-29, HCT-116, SW48, and LoVo) in isoforms with and without v3 (v3 − v10, v4 − v10, v6 − v10, and v3 + v8 − v10).

Pseudo-exon containing isoforms are substrate of NMD

Use of the (−49)3′ss inevitably shifts the reading frame of the exon downstream and generates a premature termination codon (PTC). PTCs are common in mRNAs, as they arise not only from mutant genes that contain nonsense and frameshift mutations, but also from wild-type genes, as a result of splicing and transcription errors. It is well documented that eukaryotic cells detect and degrade nonsense transcripts via a pathway known as nonsense-mediated mRNA decay (NMD). This surveillance mechanism is thought to protect the organism from the deleterious dominant negative or gain-of-function effects of truncated proteins that could result if nonsense transcripts were stable [9]. To investigate the impact of the NMD on the quantities of the splice products carrying the PTC, we analyzed their expression after treatment with the translation inhibitor, cycloheximide, known to suppress NMD.

RNA was isolated from MCF-7 cell line after treatment with 100 μg/ml cycloheximide for 2 h (+CHX) or without CHX treatment (−CHX). Relative abundance of *pseudo*-exon and total CD44 isoforms after NMD inhibition was measured by RT-PCR followed by hybridization with internal *pseudo*-exon probe and exon 16 probe (Fig. 4A). Normalization to untreated cells revealed an increased abundance of CD44 mRNAs containing the −49 *pseudo*-exon (compatible with NMD targets) over other CD44 isoforms or the control gene, β-actin (Fig. 4B). This indicated that *pseudo*-exon containing transcripts were downregulated by NMD.

In the present study, we identify novel splice variants of the human CD44 gene. We describe a cryptic splice

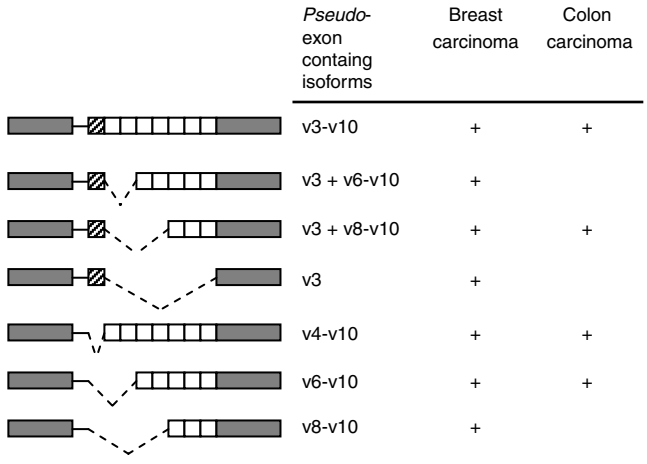

Fig. 3. Schematic representation of CD44 *pseudo*-exon containing isoforms expressed in breast carcinoma samples, MCF-7 cell line and colon carcinoma cell lines (HT-29, HCT-116, SW48, and LoVo). Gray boxes: CD44 constant regions; dashed box: exon v3; white boxes: variable exons v4–v10; solid line: 49 bp *pseudo*-exon; dashed lines: splicing routes.

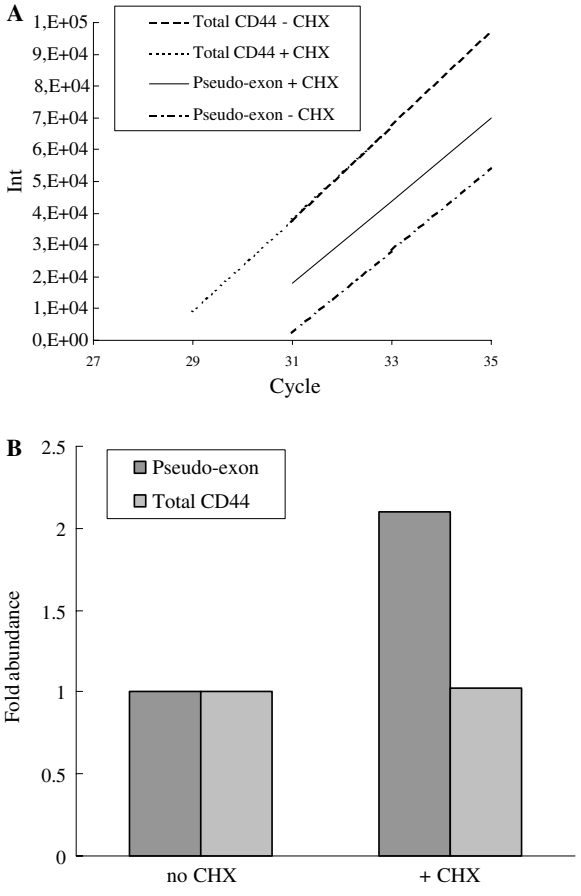

Fig. 4. Fold abundance of total CD44 transcripts and *pseudo*-exon containing transcripts after treatment of the MCF-7 cell line with cycloheximide. (A) Semiquantitative RT-PCR followed by hybridization with internal *pseudo*-exon probe and exon 16 probe. (B) Hybridization results normalized to untreated cells.

site, in intron 6 of CD44, used during mRNA processing. This cryptic splice site is used in conjunction with variable exon 3, or independently from it in the form of a *pseudo*-exon of 49 bp, which generates a stop codon by frame shift in the contiguous variable exon downstream. Our results suggest that CD44 *pseudo*-exon containing isoforms are acted upon by NMD. It has been described that the magnitude of the downregulation caused by a PTC is determined by the splicing efficiency of introns downstream a PTC, not those upstream [10]. These authors propose that splicing efficiency affects deposition of exon-junction complex (EJC) downstream of the stop codon, and preliminarily establish three categories of NMD targeted genes according to the overall deleterious physiological effect of aberrant transcripts: (1) genes that undergo programmed rearrangements, (2) dominant negative truncated proteins, and (3) genes that give rise to PTC bearing transcripts in the absence of mutation, as would be the case for alternatively spliced CD44. Possibly, efficient splicing recruits the EJC more often than inefficient splicing does, thereby eliciting NMD in a greater proportion of transcripts. In the case of CD44, the PTC created by the insertion of the 49 bp *pseudo*-exon is followed by several variable exons which contain weak splice sites with low splicing efficiency. Splicing of CD44 pre-mRNA at the (−49)3′ss could be yet another uncontrolled event in a transformed cell whose splicing mechanisms may already be altered. In most cancers, the altered expression of CD44 is not the result of CD44 mutations. Instead, the cellular processes involved during carcinogenesis and metastatic spread may affect CD44 expression patterns in cancer cells at high rates of cell division [11]. CD44 intron retention has been previously described in colon carcinomas, bladder tumors and breast carcinomas [12–15]. We suggest that the CD44 *pseudo*-exon inclusion leading to PTC-containing transcripts may be targeting these transcripts for NMD and prevent an overtly excess of different CD44 isoforms in the cell membrane. This hypothesis may help explain the many inconclusive reports of prognostic value of certain CD44 isoforms observed at the mRNA level that have not been confirmed by immunohistochemistry or other protein studies [16]. Furthermore, our observation may illustrate a fail-safe mechanism of CD44 expression in the context of deregulated splicing during carcinogenesis as implied by Xu and Lee [17].

Acknowledgments

The authors are grateful to Dr. A. Ariza, Dr. J. L. Mate, and Dr. E. Castella for kindly providing biopsy material from the frozen tissue bank of the Department of Pathology at the Hospital Universitari Germans Trias i Pujol, Badalona, Spain. We also thank Dr. A.M. Muñoz-Marmol and J.M. Hilari for their helpful suggestions and review of the manuscript.

References

[1] S. Stamm, S. Ben-Ari, I. Rafalska, Y. Tang, Z. Zhang, D. Toiber, T.A. Thanaraj, H. Soreq, Function of alternative splicing, Gene 344 (2005) 1–20.
[2] D.W.J. Smith, J. Valcarcel, Alternative pre-mRNA splicing: the logic of combinatorial control, Trends Biochem. Sci. 25 (2000) 381–388.
[3] J. Bajorath, Molecular organization, structural features, and ligand binding characteristics of CD44, a highly variable cell surface glycoprotein with multiple functions, Proteins 39 (2000) 103–111.
[4] Z. Kalnina, P. Zayakin, K. Silina, A. Line, Alterations of pre-mRNA in cancer, Genes Chromosomes Cancer 42 (2005) 342–357.
[5] R. Marhaba, M. Zöller, CD44 in cancer progression: adhesion, migration and growth regulation, J. Mol. Histol. 35 (2004) 211–231.
[6] X. Roca, J.L. Mate, A. Ariza, A.M. Muñoz-Mármol, C. von Uexkull-Guldeband, I. Pellicer, J.J. Navas-Palacios, M. Isamat, CD44 isoform expression follows two alternative splicing pathways in breast tissue, Am. J. Pathol. 153 (1998) 183–190.
[7] M.B. Shapiro, P. Senapathy, RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression, Nucleic Acids Res. 15 (1987) 7155–7174.
[8] P. Senapathy, M.B. Shapiro, N.L. Harris, Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project, Methods Enzymol. 183 (1990) 252–278.
[9] P.A. Frischmeyer, H.C. Dietz, Nonsense-mediated mRNA decay in health and disease, Hum. Mol. Genet. 8 (1999) 1893–1900.
[10] J.P. Gudikote, S. Imam, R.F. Garcia, M.F. Wilkinson, RNA splicing promotes translation and RNA surveillance, Nat. Struct. Mol. Biol. 12 (2005) 801–809.
[11] H. Ponta, L. Sherman, P.A. Herrlich, CD44: from adhesion molecules to signalling regulators, Nat. Rev. Mol. Cell Biol. 4 (2003) 33–45.
[12] S. Goodison, K. Yoshida, M. Churchman, D. Tarin, Multiple intron retention occurs in tumor cell CD44 mRNA processing, Am. J. Pathol. 153 (1998) 1221–1228.
[13] K. Yoshida, T. Sugino, J. Bolodeoku, B.F. Warren, S. Goodison, A. Woodman, T. Toge, E. Tahara, D. Tarin, Detection of exfoliated carcinoma cells in colonic luminal washings by identification of deranged patterns of expression of the CD44 gene, J. Clin. Pathol. 49 (1996) 300–305.
[14] Y. Matsumura, M. Sugiyama, S. Matsumura, A.J. Hayle, P. Robinson, J.C. Smith, D. Tarin, Unusual retention of introns in CD44 gene transcripts in bladder cancer provides new diagnostic and clinical oncological opportunities, J. Pathol. 177 (1995) 11–20.
[15] J. Bolodeoku, K. Yoshida, T. Sugino, S. Goodison, D. Tarin, Accumulation of immature intron-containing CD44 gene transcripts in breast cancer tissues, Mol. Diagn. 1 (1996) 175–181.
[16] A. Herrera-Gayol, S. Jothy, Adhesion proteins in the biology of breast cancer: contribution of CD44, Exp. Mol. Pathol. 66 (1999) 149–156.
[17] Q. Xu, C. Lee, Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences, Nucleic Acid Res. 31 (2003) 5635–5643.