Multiple Variants of the Human Lymphocyte Homing Receptor CD44 Generated by Insertions at a Single Site in the Extracellular Domain*

David G. Jackson‡, Jonathan Buckley§, and John I. Bell
From the Molecular Immunology Group, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, United Kingdom

The human CD44 cell-surface glycoprotein participates in a wide variety of cell-cell interactions including lymphocyte homing and tumor metastasis. The CD44 antigen is known to display extensive size heterogeneity when compared between different tissue sources although the structural basis for this variation is not yet clear. Recently, two further isotypes in addition to the basic hemopoietic form of the CD44 antigen have been cloned and sequenced and these have been found to contain all or part of a 200-400-base pair insert within the extracellular domain, suggesting that the characteristic heterogeneity in the molecule may be generated by a mechanism of alternative splicing.

We have obtained further evidence for alternative splicing, and we report here the cloning and sequencing of six different CD44 sequence variants from a variety of cell lines using a combination of expression cloning and the polymerase chain reaction. Comparison of these variants indicates that each is probably assembled by the insertion of five different exon units in tandem into a discrete site within the membrane proximal region of the extracellular domain. One of these variants contains an exon that expresses a tandem repeat of the consensus sequence SG amino acid homology with a recently described rat CD44 variant that mediates tumor metastasis. Another variant contains a new exon that encodes a tandem repeat of the consensus sequence SG for covalent modification with chondroitin sulfate and is expressed predominantly on mammary tumors.

We suggest that a mechanism of alternative exon splicing generates much of the observed structural heterogeneity of CD44 and that the particular set of CD44 variants expressed in a single cell may represent a precise postnatal code directing the final destination of migrating cells and metastatic tumors.

* This work was generously supported by grants from the Wellcome Trust The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1754 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M83324-M83328.

‡ To whom correspondence should be addressed.
§ Present address University of Southern California, Los Angeles, CA.

restricted to lymphocytes but is found in a wide variety of tissues including the central nervous system, lung, epidermis, liver, and pancreas (3) in addition to cell types as diverse as ovarian carcinomas (4) and erythrocytes (5).

Recently, monoclonal antibodies to CD44 have been shown also to recognize both the major cell-surface receptor for hyaluronic acid (6, 7) and an abundant surface glycoprotein initially characterized as an extracellular matrix receptor (ECMR III) for collagen and fibronectin (8, 9). Other CD44 monoclonal antibodies have been found to stimulate T-cell activation in vitro (10) or to block the development of lymphoid and myeloid precursors when added to in vitro bone marrow cultures (11). Data also link CD44 with a role in tumor metastasis (12), and increased expression of CD44 transcripts is found in many human tumors (13).

It is clear from the many immunochemical studies that CD44 is involved in a diverse range of biological functions. These same studies have also revealed a considerable diversity in the structure of CD44 molecules expressed on different cell types (3, 4, 14, 15). Three broad size categories of CD44 molecules have been identified in previous immunoprecipitation analyses of surface radioiodinated cells: 1) a predominant 80-90 kDa category that is expressed by most cells of hemopoietic origin (1, 16), 2) an intermediate size category 110-160 kDa that includes forms expressed primarily on epithelial cells (4, 14, 15, 17, 18), and 3) a category that includes very large isoforms (>200 kDa) covalently modified by the addition of chondroitin sulfate (16, 18). However, each of these size categories represents more than one species, and the true number of different CD44 variants may be very large.

Recently, cDNAs encoding the CD44 glycoprotein have been cloned from lymphoid tissues of the human (13, 19), baboon (20), and mouse (21). In each case a single cDNA species was obtained that encoded an integral membrane protein with an NH2-terminal extracellular domain homologous to the cartilage-link and proteoglycan core proteins found in connective tissue. The protein product of the human CD44 cDNA was shown to migrate with an apparent molecular mass of 80-90 kDa, similar to that of the major species found on lymphoid tissues and has been termed the hemopoietic form (13, 19). Two variants of the basic CD44 molecule from the human and one CD44 variant from the rat have been cloned and sequenced recently. The first of these, termed the epithelial form because of its predominant expression in human epithelial tissues and tumors encodes a protein that is identical to the hemopoietic form except for the presence of an additional 155 amino acids within the extracellular domain (17, 18). The second human variant, reported during the preparation of this article encodes a protein that is identical to the epithelial form but which contains only the last 69 amino acids of the epithelial insert (22). The rat variant which was cloned from a rat metastatic pancreatic adenocarcinoma
cDNA library encodes a protein with an additional 162 amino acids located at a similar site in the extracellular domain (12). Interestingly, this variant is expressed only in metastatic pancreatic and mammary tumors and confers metastatic growth characteristics when transfected into non-metastatic rat tumor cells (12).

These findings suggested that additional insertions might be found within the extracellular domain of other CD44 molecules, perhaps consisting of new sequence units. This was a particularly attractive hypothesis since it provided a potential explanation for the characteristic structural and functional diversity of the CD44 molecule.

We report here that additional sequence insertions indeed occur within the extracellular domain of CD44 and that these are restricted to a discrete site in the membrane proximal region. We describe the characterization of six different CD44 variants that are generated most likely by alternative splicing of five or more different exons, often inserted in tandem. Furthermore, three of the variants are new and have not been described previously. One confers metastatic homology with the rat "metastatic" CD44 variant (12) and another, containing a site for O-glycosylation with chondroitin sulfate (23) appears to be expressed predominantly on human breast tumors. The potential diversity generated by the postulated mechanism of exon "shuffling" is discussed in relation to the role of CD44 in lymphocyte homing and tumor metastasis.

**MATERIALS AND METHODS**

**Cell Lines**—The following human hemopoietic and tumor cell lines were used as sources of RNA for cDNA synthesis and amplification of CD44 variants by the polymerase chain reaction (see below): JY (B-lymphoblastoid), IM-9 and U-266 (plasmacytoma), Molt-4 (T-cell leukemia), HPBALL (T-cell leukemia), LGL (large granular lymphocyte), PBL (peripheral blood lymphocyte), K562 (erythroleukemia), U-937 (monoblastic leukemia), P05 and SW2 (small cell lung carcinoma), MDA-MB-231 (mammary adenocarcinoma), HT-29 (colon adenocarcinoma), Mel-D8-1 (melanoma), and EJ (bladder carcinoma). In addition, samples of human thymus, thyroid (predominantly follicular adenocarcinoma), melanoma, MDA-231 (mammary adenocarcinoma), HT-29 (colon carcinoma), anMolt-4 (T-cell leukemia), HPBALL and from large granular lymphocytes similarly constructed in the expression vector pCDM7 (31) were kindly provided by D. Olive, personal communication.

**RNA and DNA Isolation and Purification**—RNA was purified by extraction into guanidinium thiocyanate followed by either caesium chloride density centrifugation (26) or phenol extraction/ethanol precipitation at low pH (28). Genomic DNA was purified according to standard methods (29).

**cDNA Libraries and Expression Cloning of CD44 cDNAs**—Three different cDNA libraries were used for expression cloning. The first, constructed from phoxygenaglutinin-stimulated peripheral blood lymphocytes in the plasmid expression vector pCDSM, has been described previously (26). Two additional cDNA libraries from T-cell leukemia HPBALL and from large granular lymphocytes similarly constructed in the expression vector pCDM7 (31) were kindly donated by Dr. David Simmons, ICRF Cell Adhesion Laboratory, Oxford, United Kingdom.

In the expression cloning of CD44 cDNAs, the three cDNA libraries were mixed and transfected into 50% confluent COS1 monolayers in 10-cm dishes (approximately 50 μg of plasmid DNA/10^6 cells) in the presence of DEAE-dextran (0.1 mg/ml) and chloroquine (10 μg/ml) for three subsequent rounds of screening with the 8B25 monoclonal antibody. Each round consisted of antibody panning (1/1000 diluted ascites) of the transfected cells followed by rescue of the transfected plasmids with the recombinant into COS1 cells via spheroplasting fusion. Each of these manipulations was carried out exactly as detailed originally by Seed and Aruffo (30).

After the final round of screening, one first-of the rescued plasmid preparation was transferred into Escherichia coli MC1061-P3, and 10 colonies were selected for plasmid isolation by the alkaline lysis method. Each of the cloned plasmids was then retransfected into COS1 cells which were screened for the expression of CD44 by indirect immunofluorescence microscopy (using the 8B25 monoclonal) as described previously (30). The procedure resulted in the isolation of two different CD44 variant cDNAs. The first corresponded to the previously described hemopoietic variant (13, 19) while the second contained an additional 200-bp insert within the region encoding the extracellular domain of the molecule. These variants are designated CD44A and CD44B in the text.

**Amplification from RNA of CD44 Variant Sequences Using the Polymerase Chain Reaction**—10 μg of total cellular RNA was used as template for the synthesis of cDNA in reactions that comprised 100 mM Tris-HCl buffer, pH 8.3, together with 50 mM KCl, 10 mM MgCl₂, 10 mM dithiothreitol, 35 units/ml RNase inhibitor (Promega Biotech, Madison, WI), 100 μg/ml oligo(dT) (18–22), and 5 mM deoxynucleoside triphosphates. First strand syntheses were carried out at 42 °C in the presence of avian myeloblastosis virus reverse transcriptase (Life Sciences, 1500 units/ml) and the products purified by three rounds of phenol extraction and ethanol precipitation. The cDNA products (one-twentieth of the total in each case) were then used as template for PCR in a reaction mix that comprised Tris-HCl (10 mM, pH 8.3), KCl (50 mM, pH 8.3), 1.5 mM MgCl₂, 200 μM of each dNTP, 0.5 μg/ml template DNA, 5 U/ml Taq polymerase (Amersham, Oxford, United Kingdom), 0.5 μg/ml of each primer (A3D8, 33-3B3), and 1 U/ml Taq polymerase (Life Sciences, Oxford, United Kingdom).

**Cloning and Sequencing of Amplified CD44 Products**—PCR products were end repaired by the addition of 1 unit of DNA polymerase (Klenow fragment) direct to the reaction mixture which was reincubated at 37 °C for 30 min before size selection of individual products on 1.2% agarose gels containing ethidium bromide. DNA bands were recovered on DEAE paper prior to phosphorylation (10 μg/ml poly nucleotide kinase in 50 mM Tris-HCl buffer, pH 7.6, 10 mM MgCl₂, 5 mM dithiothreitol, and 100 μM ATP, 37 °C, 30 min) and ligation into the Smal site of M13mp8. After transformation into E. coli JM101, individual M13 clones were screened for CD44 variant inserts by direct amplification of toeheld plaques using the PCR primers AMP1 and AMP3 as described above. Selected clones were sequenced by the dyeoxy chain termination method (33) using single-stranded phage as template in Sequenase-catalyzed reactions. Where necessary sequence compressions were resolved by substituting 7-deaza dGTP for dGTP in the polymerase reaction (34).

**Generation Blunting to Modified CD44 Products**—After electrophoresis on agarose gels, PCR products were transferred to Nitrocellulose membranes (Hybond-C extra, Amersham, United Kingdom) for hybridization with oligonucleotide probes using standard methods. Blots were hybridized with radiolabeled oligonucleotide probes (35) end-labeled in the presence of polynucleotide kinase, 10 units, 37 °C, 30 min.
min) for 1 h at 30 °C and washed (6 x SSC) for 5-15 min at 5 °C below the calculated Tm for the oligonucleotide/DNA duplex. The oligonucleotide probes were P1 (5′-TGTACATCGTACGACCCCTGCA-3′) which is complementary to a sequence beginning 284 bp upstream of the insert site and P2 (5′-GAGACCAAGACACATTCTC-3′) which is complementary to a sequence immediately 3′ of the insert site.

In addition, two probes PEX4 (5′-ATATACCTCAGCACGGCTGG-3′) and PEX5 (5′-CGAACTACTGGAGCAGGTAC-3′) complementary to exons 4 and 5 in the CD44 variants E and F (see Figs. 4 and 5) were end labeled for use in the experiment shown in Fig. 7.

Cell-surface Radiodination/Immunoprecipitation—Suspensions of transfected COS1 cells or of cell lines expressing CD44 (approximately 10^9 in 200 μl of phosphate-buffered saline, pH 7.5) were surface labeled with 500 μCi of Na-125I in the presence of 100 μl of immobilized lactoperoxidase/glucose oxidase reagent (Enzymobeads, Pierce Chemical Co.) prior to extraction into 20 mM Tris-HCl, pH 7.5, containing NaCl (150 mM), EDTA (2 mM), Triton X-100 (0.5%, w/v), iodoacetamide (2 mM), phenylmethanesulphonyl fluoride (1 mM), and leupeptin (50 μg/ml) as described previously (30). For immunoprecipitation, CD44 monoclonal antibodies were added at a dilution of 1/100-1/200 of ascites fluid to samples of precleared cell lysates (see Ref. 30 for details) and incubated for 1 h before precipitation of the immune complexes with Protein A-Sepharose CL-4B. Immuno-

RESULTS

Expression Cloning of a Novel CD44 Isotype CD44B That Contains a 200-bp Insert—cDNAs encoding the human CD44 antigen were initially cloned from a mixture of cDNA libraries in the plasmid vectors pCDM7 and pCDM8 expressed transiently in COS1 cells. The monoclonal antibody 8B25 (25) see “Materials and Methods”) was used for panning. This antibody immunoprecipitated the same range of CD44 iso-

Two full-length CD44 clones CD44A and CD44B were obtained. CD44A corresponded to the original CD44 sequence referred to as the hemopoietic form (13, 19). CD44B contained approximately 200 bp of additional sequence inserted within the region of the cDNA that encodes the membrane proximal portion of the extracellular domain. This sequence is identical to that reported very recently by Dougherty et al. (CD44R1, Ref. 22) during the preparation of this article and corresponds to the last 207 bases of the 400 bp insert recently reported in an epithelial form of CD44 (see Figs. 4 and 5 below and Refs. 17, 18).

The protein products encoded by the CD44B and CD44A cDNAs were compared after transient expression in COS1 cells and immunoprecipitation with the monoclonal antibody 8B25 (Fig. 1). The CD44B protein migrated with an apparent molecular mass (130 kDa) that was approximately 40 kDa larger than the CD44A protein (90 kDa), a difference that was much greater than would be expected from the size (0.73 kDa) of the insert alone. The CD44B protein was also immunoprecipitated from the Burkitts lymphoma B-cell line JY and the plasmacytoma cell line IM-9 (not shown) but was not expressed in the T-cell leukemia HPBALL (see also the PCR products in Fig. 2b, below). CD44B appears to be expressed only at low levels (at least 10-fold less than CD44A) on the surface of most cells. No products were precipitated from the CD44 negative T-cell line Jurkat (not shown).

Treatment of immunoprecipitated CD44B and CD44A (hemopoietic form) with glycopeptidase-F to remove N-linked sugar chains yielded products of approximately 100 and 70 kDa, respectively (Fig. 1), confirming the likelihood that further covalent modifications such as O-glycosylation or chondroitin sulfate addition were present in the insert.

PCR Amplification of cDNAs Encoding Multiple CD44 Iso-

A

B

FIG. 1. Expression of the CD44B protein from cloned cdNA. In panel A, the CD44 variants immunoprecipitated by the monoclonal antibody 8B25 from lysates of surface radiodinated cells were separated on a 10% gel and visualized by autoradiography. The immunoprecipitates were prepared from the following cells; lanes 1 and 2, COS1 cells transfected with full-length cDNAs encoding CD44A (the hemopoietic form) and CD44B, respectively; lane 3, the T-cell leukemia HPBALL; and lane 4, the B-lymphoblastoid cell line JY. In panel B, the electrophoretic mobility of CD44B immunoprecipitated from transfected COS1 cells is compared both before (lane 1) and after (lane 2) deglycosylation with glycopeptidase-F. The migration positions of the hemopoietic variant before (H−) and after (H+) similar deglycosylation in a parallel experiment are also shown at the right of the figure.
Human CD44 Sequence Variants

Fig. 2. PCR amplification of transcripts encoding multiple insert variants of CD44 from a panel of cell lines and tissues. CD44 products containing different inserted sequences were amplified from cDNA templates prepared from the panel of different cell lines shown by means of PCR using either of two different primer pairs (see "Materials and Methods" that spanned the postulated variable insert site) in parallel panels. In panel A the PCR primers used were AMP1 and AMP2, and in panel B the primers were AMP1 and AMP3. The products were detected by autoradiography after electrophoresis on a 1.2% agarose gel followed by transfer to nitrocellulose membrane and hybridization with the $^{32}$P-labeled oligonucleotide probe P1 complementary to a common sequence upstream of the insert site (see "Materials and Methods"). The migration positions of products representing the CD44B and CD44D (epithelial variant) inserts were determined using previously characterized controls that were run in parallel lanes.

The predicted product sizes in this case were approximately 420 bp for the CD44B variant and 620 bp for the epithelial variant. Since our earlier results with the AMP2 3' primer showed that the sequences corresponding to the insert were expressed among the cell lines tested (Fig. 2b). CD44B was the predominant product amplified from the hematopoietic cDNA template and was also present in metastatic breast tumor tissue, the colon carcinoma HT-29, and the bladder tumor EJ. Some of the hematopoietic cell lines (e.g. peripheral blood lymphocytes, large granular lymphocytes, and the erythroblast cell line K562) yielded only the CD44B product whereas others, particularly the metastatic mammary tumor cell line MDA-231, appeared to contain a mixture of all of the different sized products (see Fig. 2b).

Our interpretation of these results was that transcripts encoding multiple different CD44 variants were present in many cell types and that these variants were assembled by tandem insertions of different sequence "cassettes."

Parallel amplification of the region spanning exon 1 and the 3' breakpoint of the insert site using the AMP4 and AMP2 primers yielded only the predicted 235-bp product from the three representative cDNA templates tested (Fig. 3) suggesting that no further sequence units were inserted 3' to exon 1 in any of the CD44 variants described here.

The Multiple CD44 Transcripts Are Generated by Tandem Insertion of Individual Sequence Cassettes into a Discrete Insert Site-In order to define the boundaries of the putative insert site and to characterize the inserted sequences, the PCR products generated with the AMP1 and AMP3 primers were cloned and sequenced in M13. Five different variant inserts, termed B-F, were subsequently defined (Figs. 4 and 5). Three of the variants (C, E, and F) were new and have not been described previously. Variant B corresponded to the CD44B isotype isolated earlier in this study by expression cloning (and recently by Dougherty et al. (22)), and variant D was apparently identical to the epithelial form of CD44 isolated by Stamenkovic et al. (17) except for a single GC inversion (position -5 of the sequences in Fig. 5) that converts the arginine in the published sequence to an alanine residue. Interestingly, the sequence of the epithelial isoform recently from human keratinocytes (18) also contains an alanine residue at this position.
Comparison of the inserts from each of the five cloned variants allowed the identification of at least five individual sequence cassettes (Fig. 4). The simplest of the insert variants (CD44B) contained only cassette 1 while the most complex (variants E and F) each contained four cassettes arranged in tandem (Fig. 4). Furthermore, the insert in variant C was identical to the insert within the epithelial form (variant D) except for the absence of the first 35 amino acids from NH$_2$ terminus (Fig. 5 and references). The most plausible explanation for these findings is that the sequence inserts or cassettes described here represent individual exons that are alternatively spliced to generate different CD44 isotypes. This explanation is further supported by the occurrence of sequences conforming closely to the conserved exon splice-donor consensus (C$^\text{357}$A$^\text{399}$) at the 3' ends of many of the putative exons and by the fact that each of the inserts maintains the CD44 open reading frame (see Fig. 5). In the case of both exons 3 and 5 (variants C and F, respectively) the 3' splice site occurs within a codon and changes the identity of the encoded amino acid. This provides an explanation for the substitution of aspartic acid for asparagine at residue +45 of the variant F insert and the substitution of lysine for glutamine at residue +1 of the variant C insert (Fig. 5).

Each of the five putative exons contained a high proportion of serine and threonine residues (28–43%) and with the exception of exon 5 (variant F) all contained the tripeptide motifs that mark potential O-glycosylation sites (Fig. 5). By contrast only exon 1 contained the consensus N X (S/T) sequence for N-glycosylation. Interestingly exon 4 also contained a copy of the SGXG motif (23) that represents the potential site for O-glycosylation with chondroitin sulfate. The minimal consensus motif SG is also present at the COOH terminus of exon 1 (Fig. 5) that was originally identified in the epithelial form of CD44 (17). The effects of inserting these highly glycosylated sequences on the biological functions of CD44 are not clear. However, recent evidence indicates that in the case of the epithelial form of CD44 (CD44D) the presence of the 135-amino-acid insert (defined here as exons 1–3) decreases the affinity of the molecule for hyaluronic acid (7). It is also significant that the insert site defined in the present study is located within the same region of the extracellular domain that appears to be modified by chondroitin sulfate addition and by conventional O-glycosylation (13, 19).

![Fig. 3. PCR amplification of the 3' end of the insert site.](image)

The region spanning the 3' end of the putative exon 1 sequence and the sequence just downstream of the insert site was amplified from cDNAs using the primers AMP2 and AMP4 (see "Materials and Methods") to check for additional insertions 3' of exon 1. The products were visualized by hybridization with the $^{32}$P-labeled oligonucleotide probe P2 (see "Materials and Methods") after electrophoresis and transfer to nitrocellulose as described under "Materials and Methods." The cDNA templates were as follows: lane 1, IM-9; lane 2, U-937; and lane 3, SW2. The positions of molecular weight markers are shown on the left of the figure.

![Fig. 4. Six different CD44 variants generated by insertion of different combinations of five different sequence cassettes into a single site within the extracellular domain.](image)

The figure shows a diagrammatic representation of the basic hemopoietic CD44 variant (CD44A) together with the different combinations of the five sequence cassettes that were found in the CD44 variants (CD44B–F) identified by PCR, using the AMP1 and AMP3 primers. The individual cassettes are represented by the shaded bars and were identified by sequencing of cloned PCR products (see "Materials and Methods" and text for details). The full sequences for each variant are shown in Fig. 5. The sequences flanking the insert site are shown in the basic CD44 variant for the sake of clarity.
We have examined the remainder of the CD44 sequence for further insertions both 5' and 3' of the "membrane-proximal" site in a large panel of different tissues using PCR (AMP5, AMP6, AMP7, and AMP8 primers, see "Materials and Methods"). With the exception of the splice site within the cytoplasmic tail described by Goldstein et al. (35), no other variable insert sites were detected (not shown).

One of the Exons Represents the Human Homologue of an Exon Present within the Rat Metastatic CD44 Variant—None of the five variant exon sequences was found to be homologous with any sequence in the Genbank database. However, comparison with the cDNA encoding the recently described rat CD44 variant that has been shown to play a causal role in pancreatic and mammary tumor metastasis (12) (see also "Discussion") revealed a striking similarity with the sequence of exon 5 (Fig. 6). Furthermore, the region of homology which we identified was located within the 162-amino-acid insert of the rat CD44 that appears to be required for the metastatic growth promoting properties of the molecule (12). Altogether 29 out of 44 amino acids in exon 5 were identical to sequence at the COOH terminus of the rat CD44 (see Fig. 6 and Ref. 12). This result suggests that CD44 variants containing exon 5 may also play a role in human tumor metastasis.

We have not yet identified sequences homologous to the NH2-terminal portion of the rat metastatic variant insert although it is highly likely that additional exons containing these sequences will be found in other human CD44 variants. One possibility is that such exons are only rarely spliced into transcripts containing exon 1 (e.g. the major rat variant contains no additional insertions besides the metastatic exons (12)) and thus would not have been amplified efficiently by the PCR primers used in our strategy.

The usage of exons 4 and 5 in transcripts from different tissues was investigated by probing nitrocellulose blots of CD44 PCR products (AMP1 and AMP3 primers) with oligonucleotides specific for the individual exons. The results (Fig. 7) show that in the case of the exon 4 probe PEX4 (see "Materials and Methods") three bands of 560, 660, and 760 bp were detected specifically in metastatic mammary tumor cell transcripts (Fig. 7a). Interestingly, the two smaller products were present only in the metastatic mammary tumor line MDA-231 (Fig. 7a). These results were confirmed using a second probe, complementary to a different part of the exon 4 sequence (data not shown). In contrast, the exon 5 probe PEX5 (see "Materials and Methods") hybridized with transcripts present in each of the tumor cell samples (with the exception of the melanoma, see Fig. 7b) in addition to the B-lymphoblastoid cell line JY and the monocye cell line U-937 (Fig. 7b). In this case up to five transcript sizes were detected, ranging in size from 660 to 1200 bp, and different combinations of these were present in each cell line.

Curiously, the transcript levels appeared to be much higher in the metastatic mammary tumor line MDA-231 than in any of the other tumors. This overall pattern of exon 5 usage was similar to that reported recently for the full rat metastatic variant insert (12). Interestingly, this rat CD44 insert was also reported to be present in a variety of different sized transcripts, and these were expressed as different sized protein products.

**DISCUSSION**

The results presented in this study provide strong support for the hypothesis that the structural and functional heterogeneity which characterizes the CD44 molecule (3, 4, 14, 15) is generated by a mechanism of alternative splicing. The splicing appears to be confined to a single insert site in the

---

**Fig. 5. The full sequences of the five CD44 insert cassettes.** The figure shows the nucleotide and derived amino acid sequences of the variant inserts depicted in Fig. 4. The bars underneath each sequence mark the putative exon boundaries and are shaded as in Fig. 4. Sequences upstream of the insert site which were identical in each variant are omitted except for the last 18 bp encoding the motif DRIPAT. The complete sequence of exon 1 is shown in the case of variant B only. Sites of potential N-glycosylation (boxes), O-glycosylation ( — ), and chondroitin sulfate addition (****) are marked. The possible exon splice donor consensus sequences denoted by the dinucleotide AG at the 3' ends of exons are underlined.
There is the potential for 26 different variants, limited only by the particular combination of different CD44 variants expressed on individual cells. The potential for such diversity is clearly present in the cell lines tested but remains to be confirmed at the single cell level by PCR using the RNA from individual cells as template.

The splicing of exons into the extracellular site described in the present study may subserve a number of different functions. One interesting possibility is that the spliced exons could modulate the ligand binding properties of the different variants. Indeed, it has already been demonstrated that the presence of the 135-amino-acid insert within the membrane proximal region of the epithelial variant radically reduces the capacity for hyaluronic acid binding that resides in the NH2-terminal domain of the molecule (7). Furthermore, in the rat a 162-amino-acid insertion at a similar site of the molecule generates a variant of the CD44 molecule that confers metastatic growth potential when overexpressed in non-metastatic tumor cells (12). Significantly, the metastatic growth potential of this rat CD44 variant can be reversibly blocked by a monoclonal antibody that binds to an epitope within the spliced insert (40). The introduction of new exons into the membrane proximal insert site of the human CD44 molecule may also have significant effects on the trafficking of lymphocytes between the bloodstream and lymph nodes. For example, it is interesting that the insert site is located within the region of the CD44 molecule that contains the epitope recognized by the monoclonal antibody Hermes 3 (19). This antibody blocks the binding of lymphocytes expressing the gp90Hermes (CD44) antigen to the high endothelium of mucosal lymph nodes in vitro (2), an interaction that appears to involve recognition of CD44 by a 58-66 kDa glycoprotein termed the mucosal vascular addressin (3).

Each of these results suggests that the membrane proximal region of the CD44 molecule plays a critical role in determining its ligand binding properties. The introduction of new exons into this region could modulate adhesiveness either by extending the various combinations of exons allow for between 12 and 24 potential protein isoforms, respectively. Splicing of the CD44 is more similar to that found in CD45, where eight isoforms are generated from a single gene by alternative use of only three variable exons within a single extracellular splice site. However, the number of alternative exons and hence the extent of the potential diversity in CD44 are clearly much greater.

Some general patterns for the tissue specificity of the CD44 variants were evident from our studies. For example the so-called hemopoietic variant of the CD44 (CD44A) was found to be expressed on virtually all the cell lines surveyed including most of the tumor cells. By contrast expression of the CD44B variant although present in most of the hemopoietic tissues was absent from many of the tumor cell lines. The other new CD44 variants that contained exons 4 and 5 showed differing degrees of tissue specificity. Our initial data on exon 4 suggests a relatively restricted usage of this element on mammary tumor tissue. Exon 5 which is homologous to part of the rat metastatic variant insert (12) was particularly abundant in transcripts from mammary tumors but unlike exon 4 was present also in other tumor types in addition to macrophages (U-937) and some B-cells (JY). This particular pattern of expression is quite similar to that reported for the rat metastatic CD44 variant, and it is tempting to speculate that CD44 molecules containing exon 5 may play an analogous role in human tumor metastasis.

The splicing of exons into the extracellular site described in the present study may subserve a number of different functions. One interesting possibility is that the spliced exons could modulate the ligand binding properties of the different variants. Indeed, it has already been demonstrated that the presence of the 135-amino-acid insert within the membrane proximal region of the epithelial variant radically reduces the capacity for hyaluronic acid binding that resides in the NH2-terminal domain of the molecule (7). Furthermore, in the rat a 162-amino-acid insertion at a similar site of the molecule generates a variant of the CD44 molecule that confers metastatic growth potential when overexpressed in non-metastatic tumor cells (12). Significantly, the metastatic growth potential of this rat CD44 variant can be reversibly blocked by a monoclonal antibody that binds to an epitope within the spliced insert (40). The introduction of new exons into the membrane proximal insert site of the human CD44 molecule may also have significant effects on the trafficking of lymphocytes between the bloodstream and lymph nodes. For example, it is interesting that the insert site is located within the region of the CD44 molecule that contains the epitope recognized by the monoclonal antibody Hermes 3 (19). This antibody blocks the binding of lymphocytes expressing the gp90Hermes (CD44) antigen to the high endothelium of mucosal lymph nodes in vitro (2), an interaction that appears to involve recognition of CD44 by a 58-66 kDa glycoprotein termed the mucosal vascular addressin (3).

Each of these results suggests that the membrane proximal region of the CD44 molecule plays a critical role in determining its ligand binding properties. The introduction of new exons into this region could modulate adhesiveness either by...
inducing a conformational change in the hyaluronic acid-binding site or by generating entirely new binding sites for ligands that have yet to be identified. The particular choice of exon usage in an individual cell may be triggered by signals in the surrounding environment during lymphocyte maturation, and this could provide a structural basis for the specificity of lymphocyte homing. Moreover, the particular set of CD44 variants expressed on any one cell may represent a precise “postal code” directing the final destination of migrating cells as well as dictating the metastatic potential of human tumors. The potential for diversity revealed within this single site of the extracellular domain of the CD44 provides the structural basis for the specificity that may underly these important processes.

Acknowledgments—We acknowledge Dr. David Simmons for kindly providing the HPBALL and H-T-29 cDNA libraries and the melanoma RNA, Dr. Kenneth Smith for the RNA samples from bladder and mammary tumor cells, Dr. Peter Beverly for the PO5 Small Cell Carcinoma cell line, and Dr. Rolf Stahel for the SW2 Small Cell Carcinoma cell line. We also thank Randall Hoyer and especially Jon Edwards for valuable assistance with the computer graphics.

REFERENCES
1. Jalkanen, S., T., Bargatze, R. F., Herron, L. R., and Butcher, E. C. (1986) Eur. J. Immunol. 16, 1195-1202
2. Jalkanen, S., Bargatze, R. F., DelosTojos, J., and Butcher, E. C. (1987) J. Cell Biol. 105, 983-998
3. Picker, L. J., Nakache, M., and Butcher, E. C. (1989) J. Cell Biol. 106, 927-937
4. Bai, S. T., Hugger, F., Keizer, G. D., Thepen, T., Horst, E., and Fogel, C. C. (1989) J. Immunol. 143, 851-857
5. Telen, M. J., Eisenbarth, G. S., and Haynes, B. F. (1983) J. Clin. Invest. 71, 1878-1886
6. Underhill, C. B., Green, S. J., Comoglio, P. M., and Tarone, C. (1987) J. Biol. Chem. 262, 13142-13146
7. Aruffo, A., Stamenkovic, I., Melnick, M., Underhill, C. B., and Seed, B. (1990) Cell 101, 1303-1313
8. Carter, W. G., and Wayner, E. A. (1988) J. Biol. Chem. 263, 4193-4201
9. Gallatin, W. M., Wayner, E. A., Hoffman, P. A., St. John, T., Butcher, E. C., and Carter, W. G. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 4654-4658
10. Huet, S., Groux, H., Cailou, B., Valentin, H., Prieur, A., and Bernard, A. (1989) J. Immunol. 143, 798-801
11. Miyake, K., Medina, K. L., Hayashi, S., Ono, S., Hamaoka, T., and Kincade, P. W. (1990) J. Exp. Med. 171, 477-488
12. Gunther, U., Hofmann, M., Rudy, W., Reber, S., Zoller, M., Haussmann, I., Matzku, S., Weinzel, A., Ponta, H., and Herrlich, P. (1991) Cell 65, 13-24
13. Stamenkovic, I., Amiot, M., Pesando, J. M., and Seed, B. (1989) Cell 56, 1057-1062
14. Omori, M., Trowbridge, I. S., Letarte, M., Kagnoff, M. F., and Isaak, C. M. (1988) Immunogenetics 27, 488-494
15. Kansas, G., Wood, G. S., and Dailey, M. O. (1989) J. Immunol. 142, 3050-3057
16. Jalkanen, S., Jalkanen, M., Bargatze, R., Tammi, M., and Butcher, E. B. (1988) J. Immunol. 141, 1615-1623
17. Stamenkovic, I., Aruffo, A., Amiot, M., and Seed, B. (1991) EMBO J. 10, 343-348
18. Brown, T. A., Bouchard, T., St. John, T., Wayner, E., and Carter, W. G. (1991) J. Cell Biol. 113, 207-221
19. Goldstein, L. A., Zhou, D. F., Pick, L. J., Minty, C. N., Bargatze, R. F., Ding, J. F., and Butcher, E. C. (1989) Cell 56, 1063-1072
20. Izard, R. L., Carter, W. G., Nottenburg, C., Wayner, E. A., Gallatin, W. M., and St. John, T. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 4859-4863
21. Wolfe, E. J., Gaus, W. C., Pelfrey, C. M., Holland, S. M., Steinberg, A. D., and August, J. T. (1990) J. Biol. Chem. 265, 341-347
22. Dougherty, G. J., Lansdorp, P. M., Cooper, D. L., and Humphris, R. K. (1991) J. Exp. Med. 174, 1-5
23. Bourdon, M. A., Oldberg, A., Pierschbacher, M., and Ruoslahti, E. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 1321-1325
24. McMichael, A. J., and Gootch, F. M. (1987) in Leucocytes Typing III White Cell Differentiation Antigens (McMichael, A. J., Beverley, P. C. L., Cobbold, S., Crompton, M. J., Gilks, W., Gotch, F. M., Hogg, N., Horton, M., Ling, N., MacLennan, I. C. M., Mason, D. Y., Milstein, C., Spiegelhalter, D., and Waldmann, H.) pp. 51-62, Oxford University Press, Oxford
25. (1989) Leucocyte Typing IV White Cell Differentiation Antigens (Knapp, W., Dorken, B., Gilks, W. R., Rieber, E. P., Schmidt, R. E., Stein, H., and von dem Borne, A. E. G. K.) Oxford University Press, Oxford
26. Gribbin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294-5299
27. Dechau, R., Kirkley, J., and Fabre, J. W. (1980) Eur. J. Immunol. 10, 745-749
28. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156-159
29. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning A Laboratory Manual, Cold Spring Harbor Press, New York
30. Jackson, D. G., and Bell, J. I. (1990) J. Immunol. 144, 2911-2915
31. Seed, B., and Aruffo, A. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 3365-3369
32. Hirt, B. (1967) J. Mol. Biol. 26, 365-369
33. Sanger, F., McKinley, F., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467
34. Mizusawa, S., Nishimura, S., and Seela, S. (1991) Nucleic Acids. Res. 14, 1319-1324
35. Goldstein, L. A., and Butcher, E. C. (1990) Immunogenetics 32, 283-297
36. Pagani, F., Zagato, L., Vegani, C., Casari, G., Siddoli, A., and Baralle, F. E. (1991) J. Cell Biol. 113, 1223-1229
37. Santoni, M. J., Barthsels, D., Vopper, G., Bong, A., Goridis, C., and Wille, W. (1989) EMBO J. 8, 385-392
38. Ralph, S. J., Thomas, M. L., Morton, C. C., and Trowbridge, I. S. (1987) EMBO J. 6, 1251-1257
39. Barclay, A. N., Jackson, D. I., Willis, A. C., and Williams, A. F. (1987) EMBO J. 6, 1259-1264
40. Reber, S., Mattzku, S., Gunthert, U., Ponta, H., Herrlich, P., and Zoller, M. (1990) Int. J. Cancer 46, 919-927