CD44 cell adhesion molecules

S Goodison, V Urquidi, D Tarin

Abstract
The CD44 proteins form a ubiquitously expressed family of cell surface adhesion molecules involved in cell-cell and cell-matrix interactions. The multiple protein isoforms are encoded by a single gene by alternative splicing and are further modified by a range of post-translational modifications. CD44 proteins are single chain molecules comprising an N-terminal extracellular domain, a membrane proximal region, a transmembrane domain, and a cytoplasmic tail. The CD44 gene has only been detected in higher organisms and the amino acid sequence of most of the molecule is highly conserved between mammalian species. The principal ligand of CD44 is hyaluronic acid, an integral component of the extracellular matrix. Other CD44 ligands include osteopontin, serglycin, collagens, fibronectin, and laminin. The major physiological role of CD44 is to maintain organ and tissue structure via cell-cell and cell-matrix adhesion, but certain variant isoforms can also mediate lymphocyte activation and homing, and the presentation of chemical factors and hormones. Increased interest has been directed at the characterisation of this molecule since it was observed that expression of multiple CD44 isoforms is greatly upregulated in neoplasia. CD44, particularly its variants, may be useful as a diagnostic or prognostic marker of malignancy and, in at least some human cancers, it may be a potential target for cancer therapy. This review describes the structure of the CD44 gene and discusses some of its roles in physiological and pathological processes.

Keywords: CD44 structure; hyaluronic acid; alternative splicing; neoplasia

Figure 1 Schematic diagram of the structure of the CD44 gene. The standard exons (s1–10) encode the ubiquitously expressed standard protein isoform, CD44s. Combinations of the variant exons (v1–10) can be alternatively spliced between s5 and s6 to encode variant protein isoforms, CD44v. Unlike the mouse gene, exon 6 (v1) of the human gene contains a stop codon and is not normally included in human CD44 mRNA.

The families of proteins known collectively as cell adhesion molecules are cell surface glycoproteins that have large extracellular domains, a membrane spanning region, and an intracellular, cytoplasmic functional domain. These molecules are termed adhesion molecules because of their relatively strong bonding to specific ligands. However, the interaction is more complex than mere cell attachment, and may be more involved with the cell “sensing” the extracellular environment and sending out information to adjacent cells. Although all cell adhesion proteins are involved in cell–cell or cell–matrix interactions, they have also been shown to be involved in cell motility and migration, differentiation, cell signalling, and gene transcription. The expression patterns of cell adhesion molecules are being increasingly implicated in disease processes and are candidates for use in diagnostic pathology. For example, the organisation and function of epithelial cells is maintained by interaction with the underlying substrate and one of the characteristics of malignancy is the escape from the constraints imposed by the basement membrane.

The CD44 glycoproteins are well characterised members of the hyaluronate receptor family of cell adhesion molecules. This group is defined functionally, rather than structurally, and binds to ligands of the extracellular matrix (ECM). The major ligand is hyaluronate, which is an abundant extracellular polysaccharide found in mammalian ECM, but CD44 appears to have many varied functions dependant on the extracellular structure of the protein, which can be produced in a myriad of isoforms. The wide range of functional proteins is produced from a single gene by both alternative splicing and post-translational modification.
molecule interacts with the cytoskeleton through the binding of ankyrin and the ERM family (ezrin, radixin, moesin) to the cytoplasmic domain.

Figure 2 CD44 protein structure. The standard isoform binds its principal ligand, hyaluronic acid at the N-terminal, distal extracellular domain. The inclusion of combinations of the variant exons (v1–10) within the extracellular domain can alter the binding affinity for hyaluronic acid and confer interaction with alternative ligands. The molecule interacts with the cytoskeleton through the binding of ankyrin and the ERM family (ezrin, radixin, moesin) to the cytoplasmic domain.

exons and by the use of cryptic splice sites. In the human gene, exon 6 (v1) contains a stop codon at the 17th amino acid and so, unlike in mice, this exon is not normally included in processed transcripts.

PROTEIN

The most abundant standard isoform of human CD44 protein (CD44s) contains 363 amino acids and has a theoretical molecular mass of 37 kDa. The protein consists of three regions, a 72 amino acid (aa) C-terminal cytoplasmic domain, a 21 aa transmembrane domain, and a 270 aa extracellular domain (fig 2). The cytoplasmic region is encoded by part of exon 18 and by exons 19 and 20. This highly conserved cytoplasmic tail can exist as a short or a more prevalent long form by the inclusion of the C-terminal exon. The hydrophobic transmembrane domain is encoded by exon 18 and is 100% conserved between mammalian species. The extracellular domain can be subdivided further into conserved and non-conserved regions. The N-terminal ectodomain, encoded by exons 1 to 5 is highly conserved (≥85%) between mammalian species and is thought to fold into a globular tertiary structure by the formation of disulphide bonds between three pairs of cysteine residues. Also present in this section is a 100 aa region of homology with other hyaluronic acid (HA) binding proteins. This is termed the “link module” or CLP domain because of its resemblance to the HA binding domain of cartilage link protein. The variable region is the point at which up to 381 aa encoded by the 10 alternatively spliced variant exons are inserted at a site between exons 5 and 16 of the RNA transcript, corresponding to amino acid position 223. The membrane proximal region of the extracellular domain, encoded by exons 16 and 17 is less conserved (35%) between mammalian species and includes several carbohydrate modification sites.

POST-TRANSLATIONAL MODIFICATIONS

The apparent molecular mass of the CD44 protein, as estimated by gel electrophoresis, is ~80 kDa and the largest possible protein, containing peptides from all variant exons (referred to as “epican”) can be over 200 kDa. This is far greater than the expected value as calculated from aa residues and is the result of the extensive post-translational modification of CD44 isoforms. There are at least five conserved consensus N-glycosylation sites in the N-terminal ectodomain and two chondroitin sulphate attachment sites on the exon 5 product. Several potential O-linked glycosylation sites reside in the membrane proximal extracellular region and there are also consensus attachment sites for heparan sulphate, keratin sulphate, and sialic acid residues on the standard extracellular region. The alternatively spliced, variable exon products also have extensive potential modification sites, including serine/threonine rich regions for O-glycosylation, glycosaminoglycan attachment consensus SGXG motifs, and tyrosine sulphation.

The degree of glycosylation can affect the ligand binding characteristics of the protein and therefore alter its function. Thus, the regulation of the amount and the type of post-translational modification can add further diversity to the range of potential functions of CD44 isoforms.

CD44 ligand binding

HYALURONIC ACID

The principal ligand of CD44 is HA, a ubiquitous component of the ECM. HA is a linear, polymeric glycosaminoglycan and there are at least three sites for its binding on the CD44 molecule: one in the “link” domain encoded by exon 2, and the other two overlap in the region encoded by exon 5. The HA binding sites consist of basic aa clusters, with specific arginine residues being essentially required, as shown by mutation studies. However, although all CD44 isoforms contain the HA recognition site, not all cells expressing CD44 bind the HA ligand constitutively. Cells can express CD44 in an active, an inducible, or an inactive state with respect to HA binding. The differences in the HA binding state of CD44 are cell specific and have been shown to be related to post-translational modification patterns. Inhibition of N-glycosylation enhances HA binding, and abrogation of glycosylation by mutation at specific sites can convert the inducible form to the constitutively active form. In certain cell types, inactive CD44 surface protein can be rapidly induced to bind HA by interaction with specific antibodies. The induced binding state can also be achieved more slowly by incubation of the cells with stimulating agents such as phorbol esters or interleukins. Detailed mutational analysis of basic aa clusters in the cytoplasmic domain has defined specific arginine and lysine residues...
through which protein kinase C (PKC) stimulating reagents differentially regulate the binding of CD44 to hyaluronic acid.\(^1\)

**OSTEOPONTIN**

There are reports of interactions between CD44 and another member of the ECM, osteopontin. Co-localised immunoreaction of osteopontin and CD44v in tumour cells concomitant with cluster formation in the lymphatic vessels has been observed, suggesting that interaction between osteopontin and CD44 may parallel lymphogenous metastasis.\(^2\) Weber et al have reported that osteopontin binds to CD44 in a specific, dose dependent manner and that the binding is sensitive to competition by both hyaluronate and anti-CD44 antibodies.\(^2\)

Osteopontin is a cytokine secreted by several cell types and it induces cellular chemotaxis. It has been suggested that the coupling of osteopontin secretion and the expression of CD44 could cause migration and metastasis of tumour cells to specific sites.\(^2\)

**OTHER EXTRACELLULAR LIGANDS**

The attachment of the polysaccharide chondroitin to sites on the CD44 protein (exon 5) can lead to a number of indirect extracellular matrix binding interactions. Collagen,\(^7\) fibronectin, and laminin can bind to CD44 in this way.\(^2\)

Serglycin, a family of proteoglycans named for its Ser-Gly dipeptide repeats, has also been reported to bind to CD44 indirectly.\(^2\)

Chondroitin sulphate modified serylcin is secreted from haemopoietic cells and its interaction with CD44 leads to the aggregation, adherence, and activation of lymphoid cells.\(^2\) CD44 isoforms containing the alternatively spliced exon v3 can be modified with heparan sulphate at a Ser-Gly-X-Gly motif and are then able to bind basic fibroblast growth factor (b-FGF) and heparin binding epidermal growth factor (HB-EGF).\(^2\)

The manipulation of glycosylation patterns by the alternative splicing of CD44 is an important mechanism for the achievement of multifunctional products from a single gene.\(^10\)

**Interaction of CD44 with cellular proteins**

**MULTIMERISATION**

In common with other cell surface molecules, there is evidence that ligand binding may be enhanced by the multimerisation of CD44. The effect of antibodies that induce CD44–HA binding requires multivalency, suggesting that CD44 clustering is required for such induced binding.\(^3\)

Sleeman et al demonstrated that reduction sensitive dimerisation of CD44v can occur, and that glycosylation inhibition that abrogates CD44v–HA binding also prevents oligomerisation.\(^3\)

Furthermore, although CD44 that lacks the cytoplasmic domain does not bind HA efficiently,\(^4\) synthetic CD44 molecules that can form disulphide bonded dimers retain HA binding, even when the cytoplasmic domain of the CD44 dimer is absent.\(^5\)

**HER2**

It is now apparent that CD44 interacts with the HER2 (p185, c-ErbB2) cell surface protein, a tyrosine kinase and a member of the EGF receptor family. Overexpression of HER2 by tumour cells is associated with a poor prognosis in many studies of breast and ovarian cancer. The HER2 receptor can bind the ligand heregulin and incubation of SKBr3 breast cancer cells with heregulin increases the adhesion of these cells to plastic and the invasiveness of the tumour cells into Matrigel membranes.\(^6\)

Interestingly, this heregulin treatment also increased expression of the CD44 adhesion molecule and invasion of Matrigel membranes was partially blocked by anti-CD44 antibodies, indicating a dual role for these adhesion molecules in the invasion process. Bourguignon et al examined the interaction between CD44s and the her2 proto-oncogene product in an ovarian carcinoma cell line. Anti-CD44 mediated immunoprecipitation indicated that both CD44s and HER2 are found on the cell surface and that these two molecules are physically linked to each other via interchain disulphide bonds. It was also shown that HA binding stimulates HER2 tyrosine kinase activity, leading to an increase in ovarian carcinoma cell growth.\(^7\)

Furthermore, they found that CD44s expression and CD44s mediated cell adhesion to HA were both significantly reduced after transfection of the cell line with the adenovirus 5 E1A gene, which is known to repress the synthesis of HER2. Although studies using in vitro analysis of fusion proteins have shown that HA binding is largely dependent on the structure of the CD44 molecule alone,\(^8\) data suggest that the expression and function of the CD44 and HER2 cell surface molecules are coupled to some extent.

Immunohistochemical analysis of melanoma cell lines has shown that CD44 co-localises and co-precipitates with the proteoglycan surface molecule NG2.\(^9\) NG2 is also associated with \(\alpha 4\beta 1\) integrin and, as with HER2, this association with other cell surface molecules suggests that CD44 may be only one part of large, multifunctional complexes, in which a balance of complementary or competitive ligand binding exists.

**Interaction with intracellular proteins**

The status of the cytoplasmic domain of CD44 can influence the binding of extracellular HA (see above), presumably by membrane localisation, involving clustering or dimerisation. However, the main role for the cytoplasmic domain is to transduce signals from extracellular stimuli.\(^10\) The cytoplasmic domain has been shown to bind to a number of intracellular proteins directly. An ankyrin binding domain has been identified\(^10\) and binding appears to be enhanced by palmitoylation of CD44 amino acid residues.\(^10\) In haemopoietic cells, it has been shown that protein 4.1 can bind directly to CD44 and this interaction prevents subsequent ankyrin binding.\(^10\) A correlation between the ability to bind ankyrin and HA has been proposed\(^10\) and, although it is not demonstrable in all cell types,\(^4\) it is clear that dynamic intracellular and extracellular interactions can be facilitated by CD44 proteins. As an example
of this, it has been reported that a CD44v3,8–10 isoform, present in a metastatic breast cell line, can interact with ankyrin via the cytoplasmic domain and is found closely associated with an active matrix metalloproteinase in a complex within pseudopodia structures during cell migration.44

CD44 interactions with the ERM family (ezrin, radixin, and moesin) of proteins have also been described. CD44 has been shown to co-precipitate and co-localise with ERM proteins,45 and a region of basic amino acids in the cytoplasmic tail has been identified as essential for the binding of moesin46 and ezrin.47 Ankyrin and ERM proteins connect elements of the plasma membrane with the actin filament network of the cell and thus a direct link between CD44 and intracellular scaffold structures can be envisioned, facilitating cell motility and migration, as well as membrane localisation of CD44.48

CD44 expression

Most mammalian cells predominantly express the CD44s isoform, but some epithelial cells also express a larger isoform (CD44E), which has the products of exons v8–10 included (fig 1), often in excess of CD44s.49 The expression of CD44 isoforms containing combinations of the other variant exons is far more restricted in normal tissues. However, the expression of variant isoforms is detectable in haemopoietic cells,49 particularly in peripheral blood mononuclear cells50 and in reactive lymph node cells.51 The expression of v6 containing isoforms has been observed in the ducts of breast52 and pancreas,53 54 and v4 isoforms may be expressed in normal urothelium.55 A few cell types do not express any CD44 isoforms, including hepatocytes, pancreatic acinar cells, and cells of the tubules of the kidney and pancreas.56 57

CD44 function

The multiple functions of the CD44 family of proteins are centered around the binding of HA and, to a lesser extent, other extracellular proteins. Functional diversity is achieved by post-translational modification and dynamic interaction with other cellular factors, as discussed above. The balance of interactions between cell surface molecules and between those molecules and the ECM will determine the positional address and the migrational status of specific cell types.

CELL ADHESION

The CD44s protein is involved primarily in the maintenance of three dimensional organ/tissue structure. Epithelia undergoing proliferation and cells under repair appear to upregulate both CD44 and HA production enabling the production of, and then attachment to, a structural scaffold during expansion.55 56 HA accumulates in angiogenesis,57 wound healing,58 and embryonic cell migration59 60 and because migration of cells across an HA substrate has been described,53 HA appears to serve as a ligand and support for CD44 mediated cell movement.

CD44 molecules can also mediate the aggregation of cells. This can be achieved via multivalent HA binding by CD44 on adjacent cells or via inter-CD44 binding via attached glycosylation moieties.52 HA dependant binding can cause aggregation of macrophages, lymphocytes, and fibroblasts.53 54

The binding of HA can also induce gene expression. A number of inflammatory genes that can be induced in macrophages by HA oligomers have been identified. These include several members of the chemokine gene family, and the inducible form of nitric oxide synthase. HA fragments as small as hexamers are capable of inducing these effects and monoclonal antibodies to the HA receptor CD44 inhibits HA induced gene expression.55

ROLE IN LYMPHOCYTE FUNCTION

Several lymphocyte functions appear to be dependant upon CD44 expression. Increased surface levels of CD44 proteins are characteristic of T cell activation after encounter with its cognate antigen.66 Cell surface CD44 on lymphocytes can mediate the adhesion of lymphocytes to vascular endothelial cells via binding of HA, and this interaction is used for activated T cell extravasation into sites of inflammation in mice67 and in humans.68 This targeting of lymphocytes to effector sites by CD44–HA binding is enhanced by the induction of HA synthesis in vascular endothelium by the proinflammatory cytokines, tumour necrosis factor α (TNF-α) and interleukin 1β (IL-1β).69 Accordingly, the presence of CD44 splice variants appears to be obligatory for the migration and function of Langerhans cells and dendritic cells from peripheral organs to lymph nodes for antigen presentation.70

CD44 in neoplasia

The expression of multiple CD44 isoforms and the resulting HA binding profile can influence tumour growth and development. Initial studies showed that tumour tissues contained a number of unusual CD44 transcripts relative to those present in corresponding normal tissues.71 72 In addition, tumour cell lines with raised concentrations of CD44 proteins were shown to be capable of forming more aggressive tumours in animal experiments.73 74 Changes in the pattern of expression of a cell adhesion molecule in a tissue are likely to disrupt normal epithelial–mesenchymal interactions and thus contribute to structural and functional disorganisation, a characteristic of cancer.

The requirement for CD44–HA binding in tumour development has been shown by the suppression of tumour formation in the presence of soluble human CD44s fusion proteins.75 76 Administration of a mutant CD44 fusion protein, which was incapable of binding HA, did not abrogate tumour development.77 Local injection of hyaluronan oligomers can also inhibit tumour formation by melanoma
CD44 cell adhesion molecules

ROLE IN METASTASIS

Much interest and investigative effort has centred on the possible involvement of CD44 isoforms in tumour metastasis. This is partly because of the many parallels between lymphocyte migration and the process of metastasis. In contrast to the ubiquitous expression of CD44s in most cell types, activated lymphocytes express multiple CD44 variant isoforms. Gunthert et al first investigated CD44 function in metastasis directly by transfecting plasmids expressing CD44s or CD44v isoforms into non-metastatic, rat pancreatic carcinoma cells. Expression of the CD44s gene had no effect in this tumour but CD44v expression conveyed metastatic potential. The ability of the CD44v expressing cells to metastasise was not dependent on HA binding because a CD44v construct was unable to bind and internalise hyaluronan and developed few or no metastases in invaded tissue. However, the question of a role for CD44 in metastasis remains open because there is evidence that CD44 alone is not responsible for metastatic capability. A murine lymphoma cell line with both alleles of the CD44 gene removed was shown to retain its invasiveness and metastatic potential in vivo, despite losing all hyaluronan binding ability. CD44v does appear to be overexpressed in some metastatic lineages, but not always so. It is also possible that a pattern of expression that is advantageous for tumour cell migration may be tissue specific. Zahalka et al showed that the presence of an anti-CD44 antibody or hyaluronidase prevented lymph node infiltration but had no effect on spleen invasion by murine lymphoma cells.

DIAGNOSTIC APPLICATIONS

The use of CD44 analysis as a diagnostic and/or prognostic marker was first suggested when it was shown that CD44 expression was greatly altered in human tumour tissues compared with the tissue of origin. Since that time, there have been very many studies of CD44 expression in human tissue samples and, although there are trends in patterns of expression, a truly tumour specific isoform, or pattern of expression, has yet to be defined. We and others have performed many studies that have analysed CD44 protein concentrations or mRNA expression in solid tissue extracts but, in the absence of tumour specific isoforms, such analysis is not yet reliable enough to challenge pathological diagnosis on morphological parameters. The most promising area for the use of CD44 in diagnosis is in the analysis of body fluids. We have achieved high levels of accuracy in the detection of bladder cancer by the analysis of non-invasively obtained urine samples and, in this area, molecular diagnosis can be as reliable as conventional cytology.

Recently, studies have considered the analysis of soluble CD44 in body fluids for diagnosis. The soluble CD44 isoforms found in the serum lack a cytoplasmic tail and do not appear to be associated with other proteins or with the ligand hyaluronan. There does appear to be a correlation between soluble CD44 concentrations and tumour burden in patients with lymphomas and these values return to within normal ranges if the patient responds to treatment. The mechanism by which tumours produce more soluble CD44 is unknown but it could result from the mis-processing of transcripts, including intron retention or exon skipping. Splicing errors or inclusion of stop codons would lead to the translation of truncated proteins that may lack the transmembrane region and would therefore be secreted.

ABERRANT CD44 TRANSCRIPT PROCESSING

Analysis of tumour cell CD44 mRNA consistently reveals an extensive array of alternatively spliced transcripts that are translated into soluble CD44 isoforms. The possible involvement of CD44 isoforms in metastasis can induce apoptosis. Expression of trans-
spliced transcripts, including many that are larger than would be predicted from the known structure of the gene. Further analysis has revealed that the level of misprocessing of CD44 transcripts is greatly increased in tumour cells. Examples of this are intron retention and the use of cryptic splice sites. Compromised CD44 intron removal was first observed in exfoliated urothelia of bladder cancer patients, and was subsequently found to occur in breast carcinoma tissues and in colon carcinoma, and can involve introns of the variant and standard regions of the gene. Further evidence of tumour cell CD44 misprocessing has been reported, revealing that tumour cells produce CD44 transcripts that have been aberrantly processed using ’cryptic’ splice sites. As in the case of intron retention, these events can lead to the translation of truncated, non-functional peptides. The understanding of the processing of CD44 pre-mRNA is an area of great interest and the mechanisms of CD44 splicing are beginning to be elucidated.

PLASTICITY OF EXPRESSION

Tumour cell expression of CD44 isoforms does not remain static during changing growth conditions in vitro or in vivo. The transfection of the human CD44s gene into mouse 3T3 cells revealed remarkable plasticity of expression during tumour progression and metastasis in a nude mouse model. Expression of CD44s induced tumorigenesis, HA binding, and micrometastatic competence to multiple organs. However, large primary tumours and overt metastases lost CD44s expression and HA binding, and this was shown to be the result of hypermethylation of the transfected human CD44s gene. These studies indicate that CD44s expression and the binding of CD44s to HA is important for tumorigenesis and metastatic competence, but that downregulated expression is beneficial to advanced tumour growth. A large body of evidence supports the view that altered CD44 expression correlates with tumour growth and metastasis. However, there appears to be considerable cell-type and tumour-type specificity for the functions of this family of molecules. Whether this molecule can become the focus of clinical intervention remains to be seen, but it is a candidate marker that will probably be included in multiplex diagnosis.

Summary

The CD44 family of transmembrane proteins is perhaps the best example yet studied of the way an organism can achieve multiple and diverse functions from a single gene. Detailed functional studies are revealing increasingly complex roles for the CD44 family of proteins in physiological and pathological processes. The multifunctional aspect of CD44 cell adhesion molecules is achieved by the balance of a number of variable parameters that the cell uses in the expression of this gene. These include the amount of total CD44, alternative splicing choices, the balance of many types of post-translational modifications, and the active/inactive HA binding state of the cell surface protein. Furthermore, the homo/hetero multimerisation of the protein, the interaction with other adhesion molecules (which can compete for ligand binding), and the amount of soluble CD44 protein in the extracellular space also need to be considered. Further characterisation of the function of CD44v proteins and how they interact with other cellular components and with the extracellular environment is necessary before a comprehensive understanding of the physiological roles of the CD44 cell adhesion molecules can be achieved.

CD44 cell adhesion molecules


47 Coxon FP, Tarnuzzer D, et al. The hyaluronate receptor is expressed in human melanoma cell variants expressing high and low levels of hyaluronate receptor (Eta-1). J Cell Physiol 1995;162:663–70.


CD44 cell adhesion molecules.

S Goodison, V Urquidi and D Tarin

*Mol Path* 1999 52: 189-196
doi: 10.1136/mp.52.4.189

Updated information and services can be found at:
http://mp.bmj.com/content/52/4/189

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/