Hyaluronan-binding proteins: tying up the giant

Anthony J. Day‡ & Glenn D. Prestwich§

From the ‡MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK and §Department of Medicinal Chemistry, The University of Utah, Salt Lake City, Utah 84112-5820.

‡ To whom correspondence should be addressed: MRC Immunochemistry Unit, Department of Biochemistry, South Parks Road, University of Oxford, Oxford OX1 3QU, UK. Tel.: 44 1865 275349; Fax: 44 1865 275729; Email: ajday@bioch.ox.ac.uk.

Introduction

The ubiquitous glycosaminoglycan (GAG) hyaluronan has diverse biological roles in vertebrates. These include acting as a vital structural component of connective tissues, the formation of loose hydrated matrices that allow cells to divide and migrate (e.g., during development), immune cell adhesion and activation, and a role in intracellular signaling; further details can be found in the other articles of this MiniReview series (1-3). This wide range of activities may seem surprising for an unbranched polysaccharide comprised entirely of a repeating disaccharide, D-glucuronic acid (β1→3) N-acetyl-D-glucosamine (β1→4), which, unlike other GAGs, is neither attached to a protein core nor O- or N-sulfated. Such diversity results in fact from the large number of hyaluronan-binding proteins (often termed hyaladherins) that exhibit significant differences in their tissue expression, cellular localization, specificity, affinity and regulation. Therefore, characterization of the molecular basis of hyaluronan recognition by proteins, and how this is modulated *in vivo*, is an important key to understanding the biology of this GAG. In this article, we review the structural organization of vertebrate hyaladherins and how this may contribute to their different biological activities.

The hyaladherins - hyaluronan-binding proteins

Many hyaladherins contain a common structural domain of approximately 100 amino acids in length, termed a Link module, that is involved in ligand binding (4). However, a growing number of hyaladherins lack this domain and are unrelated to each other at the primary sequence
The Link module superfamily. The Link module (also referred to as a proteoglycan tandem repeat – see (5)) was first identified in the link protein isolated from cartilage. The link protein is comprised of an immunoglobulin domain and two contiguous Link modules, and this molecular arrangement is also found in the G1-domains of aggrecan, versican, neurocan and brevican (Fig. 1). These proteoglycans form huge, link protein-stabilized, complexes with hyaluronan that provide the load bearing function in articular cartilage, give elasticity to blood vessels and contribute to the structural integrity of many tissues, such as skin and brain (see (6,7)). A brain-specific link protein (BRAL1) has been characterized recently (8), which may be part of a larger link protein gene family (9).

The ubiquitous hyaluronan receptor CD44 has diverse functions including the attachment, organization and turnover of extracellular matrix at the cell surface, and mediates the migration of lymphocytes during inflammation (10). CD44 can exist in numerous isoforms, due to alternative splicing of 10 variant exons in different combinations. Each isoform contains a single Link module close to the N-terminus of the protein. The hyaluronan-binding properties of CD44 are determined by the isoform and the cell-type on which it is expressed (11).

Apart from CD44, the only other member of the Link module superfamily that has been clearly identified as a hyaluronan receptor is LYVE-1. This recently discovered molecule is restricted in its expression to lymph vessel endothelium and appears to be involved in hyaluronan degradation (12).

The protein product of tumor necrosis factor-stimulated gene-6 (TSG-6), that contains a single Link module, is secreted in response to inflammatory stimuli, e.g., in the articular joints of arthritis patients (13,14). TSG-6 has been implicated in the regulation of leukocyte migration, and the pattern of its expression suggests that it is likely to be involved in extracellular matrix remodelling. Although, the role of TSG-6 is, at present, poorly understood, its hyaluronan-binding properties (and their structural basis) are probably the best characterized of any hyaladherin.

Figure 1 shows several new members of the Link module superfamily (i.e., stabilin-1, CAB61358 and KIA0527), all of which contain a single Link module. Stabilin-1 (accession code AJ275213) is a transmembrane protein that was initially identified as a partial sequence (KIA0246) cloned from a myeloid cell line (15). Stabilin-1 and CAB61358 may be identical to
the proteins designated WF-HABP and BM-HABP (16), for which only the amino acid sequence for the Link modules have been reported. These both show about 50% identity with the Link module of TSG-6. However, it is unclear whether they bind hyaluronan as no functional data are yet available. The same is true for KIA0527 that was cloned from brain tissue (17).

**Non-Link module hyaladherins.** Inter-α-inhibitor (IαI), a serine protease inhibitor plentiful in serum, was one of the first proteins found to associate with hyaluronan (18). As shown in Figure 1 IαI is an unusual proteoglycan with a chondroitin-4-sulfate chain linked to bikunin (containing two kunitz inhibitor domains) and two heavy chains (HC1 and HC2) attached to the GAG by ester bonds via their C-terminal aspartic acid residues (19). IαI is essential during ovulation, acting to stabilise the hyaluronan-rich cumulus extracellular matrix with which it forms a covalent complex (20,21). The covalent association links the C-terminal ends of HC-1 and HC-2 to the C6-hydroxyl of an internal N-acetylglucosamine of hyaluronan via an ester bond, while the bikunin chain is released (20). It has also been reported that IαI can bind non-covalently to hyaluronan (22), as is the case for all other hyaluronan-protein interactions, but the precise regions of the molecule involved have not been determined.

CD38, a type II membrane glycoprotein, has been reported to be a hyaladherin (23). This protein is an enzyme with NADase activity, and this property has been studied much more extensively than its hyaluronan-binding function for which no biological role has yet been ascribed (24).

PHBP was purified from human plasma using hyaluronan-affinity chromatography (25). It is a serine protease that may be involved in wound healing where high levels of hyaluronan accumulate (26). It is comprised of EGF modules, a kringle domain and a serine protease domain. However, there is no information on which of these are involved in ligand binding.

Recently two related hyaladherins, IMP-150 and SPACRCAN, have been isolated from human retina. IMP-150, which may be identical to SPACR (27), is expressed by cone and rod photoreceptor cells and is present in the interphotoreceptor matrix (28). SPACRCAN (a chondroitin sulfate proteoglycan) is likely to be a receptor on photoreceptors and pinealocytes (29). SPACR and SPACRCAN both contain two SEA modules, a structural domain often associated with O-glycosylation (30). The positions of the hyaluronan-binding sites on these proteins are not yet established.

The **Receptor for Hyaluronan-Mediated Motility** (RHAMM) mediates cell migration and proliferation in normal and tumor cells (reviewed in (2)). A number of RHAMM isoforms are
present in the cytoplasm and nucleus and are also transiently expressed on the surface of activated leukocytes and subconfluent fibroblasts.

The presence of hyaladherins inside cells is not surprising given the increasing evidence for intracellular hyaluronan (1). In fact, there are three hyaluronan-binding proteins, in addition to RHAMM, which have been found in intracellular locations to date. The first is a vertebrate homologue of yeast and *Drosophila* CDC37 (31), which may be involved in cell cycle and kinase function as it associates with RAF and pp60^v-scr (32). The second protein, P-32 (also known as HABP-1), was originally co-purified with pre-mRNA splicing factor SF2 (33) and subsequently shown to be a hyaluronan-binding protein (34). P-32 is likely to have a role in cellular signal transduction (35), and may be involved in nucleus-mitochondrion interactions (36). As with RHAMM, P-32 can also be detected on the surface of some cells, e.g., transformed fibroblasts. The third protein, IHABP4, was identified using the same monoclonal antibody that permitted isolation of vertebrate CDC37 (37) and has been detected in the cytoplasm of IHABP4-transfected cells. Its role in hyaluronan trafficking or intracellular signalling remains to be determined.

**The structures of hyaluronan-binding domains**

The Link module structure (5), which has defined the ‘consensus fold’ for the entire Link module superfamily, consists of two α-helices and two triple-stranded anti-parallel β-sheets (Fig. 2). The single Link module from TSG-6 is sufficient for high-affinity interaction with hyaluronan (38). In contrast, CD44 has a hyaluronan-binding domain of ~160 amino acids comprised of a Link module with N- and C-terminal extensions that are essential for folding and functional activity (39,40). An even larger domain is utilized by the link protein, aggrecan and aggrecan-related proteoglycans, as these all have a pair of contiguous Link modules in their hyaluronan-binding regions. In link protein and aggrecan both Link modules are involved in binding (6,41). Therefore, hyaluronan-binding domains from Link module-containing proteins can be divided into 3 subgroups (Type A, B and C) on the basis of size (Fig. 3). The size of the binding domain appears to correlate broadly with the length of hyaluronan recognized. For instance, hexasaccharides (HA<sub>6</sub>) and decasaccharides (HA<sub>10</sub>) are the minimum sizes of hyaluronan required for high affinity interaction with Type A and Type C domains, respectively (38). Hyaluronan binding to CD44 can be displaced by either HA<sub>6</sub> or HA<sub>10</sub>, depending on the cell background on which it is expressed (42). A recent study indicates that HA<sub>10</sub> is the smallest oligosaccharide that
binds optimally to cell-surface CD44 in a monovalent fashion, while an increase in binding avidity is seen with oligomers (≥2HA₂₀) that can interact with two CD44 molecules simultaneously (43).

The amino acids of CD44 involved in the interaction with hyaluronan have been determined by site-directed mutagenesis (39,44). Functionally important residues are present both on the Link module (Figure 3) and on the flanking sequences indicating that a large coherent binding site is likely to be formed (i.e., the extensions may constitute a sub-domain in intimate contact with the Link module (4)). Recently, amino acids in TSG-6 that are perturbed on interaction with hyaluronan have been identified (38), and mutagenesis has allowed residues directly involved in binding to be distinguished from those affected due to ligand-induced conformation changes (45). These studies show that in Link module-containing proteins, binding residues are brought together from different parts of the sequence and that a folded structure is necessary to generate a ligand-binding surface. It should be noted that linear sequences (so called BX₃B motifs (46) - see below) do not appear to be involved in Link module-hyaluronan interactions (38).

The hyaluronan-binding sites in TSG-6 and CD44 have similar locations on the Link module, suggesting that the position of the ligand-interaction surface may be conserved across the superfamily (45). However, there appear to be some major differences in the details of the residues (and sequence positions) involved in mediating binding in the two proteins. In fact, comparisons of the Link module sequences of CD44 and TSG-6 with all the other members of the superfamily indicate that the molecular details of hyaluronan binding (i.e., the interaction networks) are likely to be distinct in each protein. This is perhaps not surprising given the different sub-types of hyaluronan-binding domains (described above) as well as many other differences in specificity and regulation. For example TSG-6 binds to chondroitin-4-sulphate with high affinity, but this GAG is not recognized by aggrecan or link protein (47).

The structure of a P-32 homotrimer, which is likely to be its oligomeric state in physiological solutions, has been solved recently by X-ray crystallography (36). Each monomer consists of seven consecutive β-strands that form a highly twisted antiparallel β-sheet flanked by one N-terminal and two C-terminal α-helices. The position of the hyaluronan-binding site has not yet been established.

The hyaluronan-binding domain of RHAMM has been localized to a 62 amino acid segment (the P1-domain) close to the C-terminus of the protein. This domain (residues 518-580 in the murine RHAMMv4 isoform) has been expressed and shown to be functionally active, and its
solution structure has been determined (48). As shown in Fig. 2, it consists of a helix-loop-helix motif. The P1-domain includes a region of 34 amino acids that has been implicated previously in hyaluronan binding by truncation mutagenesis (49). Two clusters of basic amino acids (531-KQIKHVVKL-541 and 553-KLRSQLVKRK-562) have been identified within this region, and these have been reported to be involved in mediating the interaction with hyaluronan (46). This led to the suggestion (46) that linear sequences termed BX,B motifs (where B is either lysine or arginine and X can be any amino acid apart from acidic residues) are likely to be a minimal requirement for hyaluronan binding in RHAMM and other hyaladherins. This study involved the expression of short basic peptides (and related mutants) fused onto the C-terminus of residues 1-238 of RHAMM followed by determination of hyaluronan-binding activity using a transblot assay. However, this experimental approach should be viewed with caution, since the pendant BX,B peptides were not displayed in the normal context of the hyaluronan-binding domain and likely lacked important contributions from secondary and tertiary structure. Although Arg46 of CD44 (which is part of a BX,B motif) was identified as important for hyaluronan binding in these experiments, subsequent work has indicated that it is not involved in the interaction (44,50). The recent determination of the structure of the P1 domain will provide the basis for a program of site-directed mutagenesis to define the position of the hyaluronan-binding surface in RHAMM.

BX,B-like sequences are present in IαI (22), SPACR (27), SPACRCAN (29), CD38 (23), CDC37 (31), P-32 (34) and IHABP4 (37). However, there are no data to indicate that they mediate hyaluronan binding in these proteins. This spacing of basic amino acids is extremely common in protein sequences, with over 16,000 matches in the NRL-3D database alone (i.e., proteins of known tertiary structure). If X is any amino acid apart from acidic residues, there are still over 10,000 matches. Therefore, the presence of a BX,B motif should not be interpreted as an indicator that a protein will bind hyaluronan nor should it be assumed that this is necessarily the site of hyaluronan-binding activity in a known hyaladherin.

Towards understanding the molecular basis of hyaluronan-protein interactions

Basic amino acids have long been presumed to be major determinants in hyaluronan binding by aggrecan and link protein, making ionic bonds with the carboxylic acid group of glucuronic acid (see references in (38)). This is also likely to be the case for CD44 where two arginines (Arg41 and Arg78), along with two tyrosines (Tyr42 and Tyr78) play a critical role in the interaction with hyaluronan (39,44). In addition to these four essential residues (dark blue on
Fig. 3), other, less critical, amino acids (including three lysines, four arginines, two asparagines and a tyrosine) have also been implicated (shown in light blue). The number of residues of CD44 engaged in binding indicates that there is a large network of interactions (distinct for this protein) that maintains the association with hyaluronan. The energetics are clearly finely balanced, since the loss of a single hydrogen bond or ionic interaction can be enough to diminish binding significantly (44).

The thermodynamics of hyaluronan-protein binding have been determined for the TSG-6 Link module, where the major energetic contribution driving the interaction is the enthalpy, which is large and exothermic (38). The small change in heat capacity seen on binding is consistent with an intermolecular interface comprising a significant polar or charged component rather than the burial of a large hydrophobic surface area. The interaction of the TSG-6 Link module with HA₈ is highly salt-strength dependent. However, ionic interactions may only contribute about 25% of the free energy of binding at physiological sodium ion concentrations (51). Therefore, while it is clear that ionic associations are important, non-ionic interactions (e.g., hydrogen bonding and van der Waals) also contribute significantly to the binding energy. This is consistent with data from site-directed mutagenesis showing that in addition to a lysine, 3 tyrosines and a phenylalanine have a critical role in binding (45).

Recently the crystal structure has been determined for a hyaluronan lyase (from *Streptococcus pneumoniae*) in complex with two hyaluronan disaccharides in the substrate binding cleft (52). There is an extensive network of hydrogen bonds and electrostatic interactions stabilizing the binding. In addition sugar rings are seen to stack against aromatic sidechains, which is a common feature in protein-carbohydrate complexes.

At present there are few data on the conformations adopted by hyaluronan after binding to proteins. In solution, hyaluronan is likely to be highly dynamic in nature, adopting a large number of low energy states (53,54). It is possible that different hyaladherins could capture (and stabilize) distinct transient conformations of hyaluronan, i.e., the structure of hyaluronan will be different depending on which protein it associates with. A consequence of this could be that particular hyaladherins are able to modulate the binding of other proteins by altering the conformation of hyaluronan. However, structural information on protein-hyaluronan complexes is necessary for definitive answers.
Regulation of hyaluronan-protein interactions

A unique feature of the interaction of TSG-6 with hyaluronan is that it is highly pH-dependent, which may serve to regulate the function of this protein in certain tissues (55). There is maximal binding at pH 6 and a dramatic reduction in binding as the pH is increased, but with no alteration in the Link module fold. A likely explanation is that raising the pH changes the charge-state on a histidine side chain, leading to the loss of a critical interaction (51).

CD44 is present on many cell types (e.g., lymphocytes) in a non-functional form that requires activation to acquire ligand-binding activity, unlike most hyaladherins that interact constitutively with hyaluronan. Understanding how the hyaluronan-binding function of CD44 is regulated has been a major focus of recent research. This complex issue involves the cell background on which CD44 is expressed, post-translational modifications (including N-glycosylation, GAG attachment and tyrosine sulfation), the splice isoform involved, membrane composition, phosphorylation of the intra-cytoplasmic domain, cytoskeletal attachment, molecular clustering and receptor density on the cell surface (10,43,56-58). A detailed discussion of each of these is clearly outside the scope of this review, and here we will focus just on N-glycosylation.

The hyaluronan-binding domain of human CD44 contains six potential sites for N-linked carbohydrate attachment, and expression of this region in E. coli indicates that glycosylation is not obligatory at any of these sites for either correct folding or functional activity (40). This conclusion is consistent with results from one study (59) but at odds with another (60). Although glycosylation is unlikely to be required for hyaluronan binding, it is clearly one of the principle mechanisms modulating the interaction between CD44 and hyaluronan (11), and there is differential glycosylation of the receptor depending on cell type and cell activation state. Removal of N-glycans by the mutation of the first or fifth glycosylation sites, which are on the N- and C-terminal extensions, respectively, can switch CD44 in an ‘inducible’ cell background into a constitutively active form (59). However, the presence of N-linked carbohydrate at all positions on a ‘constitutive’ cell-line does not inhibit binding. This may be due to particular carbohydrate structures at sites 1 and 5 causing steric interference by blocking hyaluronan binding; alternatively, certain N-glycans may prevent receptor clustering or fix the protein in a non-binding conformation.

The activation of cells to become hyaluronan binding by changes in the CD44 glycoform present on their surface (i.e., by de novo synthesis) would be relatively slow. Rapid induction in vivo (e.g., on leukocytes) may occur by the removal of sialic acid (potentially by an endogenous
sialidase) as this has been shown to be a major up-regulator of hyaluronan-binding function in CD44 (59,61,62).

The molecular basis of CD44 regulation should become clearer once the tertiary structure of the hyaluronan-binding domain is known, as this will reveal the positions of the glycosylation sites relative to the ligand interaction surface.

Summary

In recent years significant advances have been made in the identification of new hyaladherins and in our knowledge of hyaluronan-protein interactions. Determination of the three-dimensional structures of different types of hyaluronan-binding domains and their ligand complexes are clearly essential if we are to understand the molecular mechanisms underlying the diverse biology of this important glycosaminoglycan.

Acknowledgements. We thank Michael Ziebell for unpublished data and Eva Turley for informative discussions.

Figure legends

Figure 1. The modular organization of the hyaladherins. The domain structures of KIA0527, Stabilin-1, CAB61358 (a partial sequence), SPACR and SPACRCAN were determined with the assistance of the SMART programme (63) with the additional identification of OSF-2 domains (64) by databank searching.

Figure 2. Comparison of the folds of the Link module from TSG-6 (5) and the RHAMM-P1 domain (45). Helices and β-sheets are shown in red and yellow, respectively. In the Link module, the side chains of residues involved in hyaluronan binding are depicted in blue (45).

Figure 3. The Link module superfamily can be divided into three sub-groups on the basis of the size of their hyaluronan-binding domains (4). Only the tertiary structure of Type A (i.e., a single Link module as found in TSG-6) has been determined to date. Residues in TSG-6 that mediate the interaction with hyaluronan (45) are colored in red. The Link modules illustrated for CD44 (Type B) and link protein (Type C) were modelled on the basis of the TSG-6 co-ordinates (4,45).
In the Type B domain, the structures of the N- and C-terminal extensions, represented by striped green boxes, are unknown. Amino acids of the CD44 Link module that are critical or important for hyaluronan binding are depicted in dark blue or light blue, respectively (44). In the Type C domain, the relative orientation of the two Link modules is yet to be determined, and only one possible configuration is shown. The residues of link protein predicted to be involved in the interaction with hyaluronan are colored as described in Mahoney et al., (45).
Footnotes

*Financial support to A. J. D. by the Medical Research Council and Arthritis Research Campaign and to G.D.P. by the University of Utah and US Department of Defense Grant DAMD 17-98-1-8254 is gratefully acknowledged.

1 The abbreviations used are: GAG, glycosaminoglycan; RHAMM, receptor for hyaluronan-mediated motility; TSG-6, tumor necrosis factor-stimulated gene-6.

2 Space limitations preclude the citation of numerous primary papers; these can be found in the review articles referenced.


References

**Link module superfamily**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Domains/Modules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggrecan</td>
<td>VWFA domain</td>
</tr>
<tr>
<td>Versican</td>
<td>C-type lectin module</td>
</tr>
<tr>
<td>Link Protein</td>
<td>CCP module</td>
</tr>
<tr>
<td>Brevican</td>
<td>CUB module</td>
</tr>
<tr>
<td>Neurocan</td>
<td>Link module</td>
</tr>
<tr>
<td>TSG-6</td>
<td>C4S</td>
</tr>
<tr>
<td>KIA0527</td>
<td>Inter-α-inhibitor</td>
</tr>
<tr>
<td>CAB61358</td>
<td>Bikunin</td>
</tr>
<tr>
<td>Stabilin-1</td>
<td>RHAMM</td>
</tr>
</tbody>
</table>

**Other HA binding proteins**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Domains/Modules</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC1</td>
<td>Link module</td>
</tr>
<tr>
<td>HC2</td>
<td>EGF module</td>
</tr>
<tr>
<td>RHAMM</td>
<td>No known module or domain</td>
</tr>
<tr>
<td>CDC37</td>
<td>Transmembrane domain</td>
</tr>
<tr>
<td>P-32</td>
<td>Cytoplasmic domain</td>
</tr>
<tr>
<td>SPACR</td>
<td>SEA module</td>
</tr>
<tr>
<td>IHABP4</td>
<td>OSF-2 domain</td>
</tr>
<tr>
<td>CD38</td>
<td>Keratin domain</td>
</tr>
<tr>
<td>SPACRCAN</td>
<td>Glycosaminoglycan</td>
</tr>
</tbody>
</table>

Legend:
- Link module
- CCP module
- C-type lectin module
- CUB module
- VWFA domain
- Kringle domain
- EGF module
- Ig module
- SEA module
- OSF-2 domain
- Kunitz domain
- BX7B motif
- ~100 amino acids
Type A: Single Link module (~ 90 residues), e.g., TSG-6

Type B: Link module with N- and C-terminal extensions (~ 160 residues), e.g., CD44

Type C: Link module pair (~ 200 residues), e.g., Link protein