Towards a Structure for a TSG-6-Hyaluronan Complex by Modeling and NMR Spectroscopy

INSIGHTS INTO OTHER MEMBERS OF THE LINK MODULE SUPERFAMILY

Charles D. Blundell¶¶, Andrew Almond¶¶, David J. Mahoney¶¶, Paul L. DeAngelis¶**,**, Iain D. Campbell*** and Anthony J. Day*** ¶¶ from the ¶¶Medical Research Council Immunochemistry Unit and the ¶¶Department of Biochemistry, University of Oxford, Oxford OX1 3QU, United Kingdom and the ¶¶Department of Biochemistry and Molecular Biology, Oklahoma Center for Medical Glycobiology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 63104

The Link module from human TSG-6, a hyaladherin with roles in ovulation and inflammation, has a hyaluronan (HA)-binding groove containing two adjacent tyrosines that are likely to form CH−π stacking interactions with sequential rings in the sugar. We have used this observation to construct a model of a protein-HA complex, which was then tested against existing experimental information and by acquisition of new NMR data sets of [13C,15N]HA (8-mer) complexed with unlabeled protein. A major finding of this analysis was that acetamido side chains of two GlcNAc rings fit into hydrophobic pockets on either side of the adjacent tyrosines, providing a selectivity mechanism of HA over other polysaccharides. Furthermore, two basic residues have a separation that matches that of glucuronic acids in the sugar, consistent with the formation of salt bridges; NMR experiments at a range of pH values identified protein groups that titrate due to their proximity to a free carboxylate in HA. Sequence alignment and construction of homology models for all human Link modules in their HA-bound states revealed that many of these features are conserved across the superfamily, thus allowing the prediction of functionally important residues. In the case of cartilage link protein, its two Link modules were docked together (using bound HA as a guide), identifying hydrophobic residues likely to form an intra-Link module interface as well as amino acids that could be involved in supporting intermolecular interactions between link proteins and chondroitin sulfate proteoglycans. Here, we propose a mechanism for ternary complex formation that generates higher order helical structures, as may exist in cartilage aggregates.

Hyaluronan (HA)1 is an unbranched high molecular mass glycosaminoglycan (up to 107 Da) composed entirely of a repeating disaccharide of GlcA and GlcNAc. It is found in the extracellular matrix of most vertebrate tissues and underlies diverse biological functions, including roles in matrix structure, development, and ovulation (1). In addition to these normal physiological processes, HA has also been implicated in various pathological conditions such as inflammatory bowel disease, atherosclerosis and in various cancers (2–4). Much of the diversity in the roles displayed by HA is thought to arise through its interactions with a variety of specific HA-binding proteins (also termed hyaladherins), which lead to the formation of protein-HA complexes with distinct architectures and functional activities (5–7). The majority of these hyaladherins bind to HA via an α/β-structural domain of ~100 amino acids termed a Link module (8–10). The superfamily of Link module-containing proteins (which currently has 14 members in humans) has been divided into three subgroups (Types A–C) according to the known (or predicted) size of their HA-binding domains (HABDs) (6, 7). Type A domains are formed from a single independently folding Link module and are typified by TSG-6 (the protein product of tumor necrosis factor-stimulated gene-6 (11, 12)); as described in detail below, the Link module from human TSG-6 has been extensively characterized (7, 9). Type B HABDs also consist of a single Link module, but the β-sheet structure is extended by N- and C-terminal flanking sequences that are necessary for correct folding and functional activity (10). The best characterized of these is the ~150-amino acid HABD from human CD44 (the major cell-surface receptor for HA), for which the tertiary structure has recently been determined (10); Lyve-1 (13), found on lymph vessel endothelium, is also likely to have this kind of HABD. Type C domains are composed of a contiguous pair of Link modules, and HABDs belonging to this subgroup have been described in four related link proteins (HAPLN1, HAPLN2, HAPLN3, and HAPLN4 (14)) and in the G1-domains of the chondroitin sulfate proteoglycans (CSPGs) aggrecan, versican, neurocan, and brevican (7); in some cases, an additional N-terminal Ig domain may be required to stabilize the Link module folds (15).

TSG-6 is a highly conserved ~35-kDa secreted protein composed mostly of a Link module and a CUB module (11, 12). It is not constitutively produced in healthy adult tissues, but its

---

1 The abbreviations used are: HA, hyaluronan; HABD, hyaluronan-binding domain; CSPG, chondroitin sulfate proteoglycan; HA.m, HA octasacharide; ITC, isothermal titration calorimetry; NOE, nuclear Overhauser effect; HSQC, heteronuclear single quantum correlation; TOCSY, total correlation spectroscopy; NOESY, nuclear Overhauser effect spectroscopy.

---

* This work was supported in part by Arthritis Research Campaign Grants D0569, 14871, and M0625. The costs of publishing 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 1734 solely to indicate this fact.

††† Supported by the Wellcome Trust and the Biotechnology and Biological Sciences Research Council.

** Supported by National Science Foundation Grant MCB-9876193.

*** Supported by the MRC Immunochemistry Unit, University of Oxford, South Parks Rd., Oxford OX1 3QU, UK. Tel.: 44-1865-275-349; Fax: 44-1865-275-729; E-mail: tony.day@bioch.ox.ac.uk.

Received for publication, December 21, 2004, and in revised form, February 7, 2005 Published, JBC Papers in Press, February 17, 2005, DOI 10.1074/jbc.M414343200
expression is induced in a wide variety of cell types in response to inflammatory mediators and growth factors, being associated, for example, with arthritis and blood vessel injury (16, 17). The precise roles of TSG-6 in inflammation are still unclear, although recent studies in mouse models of arthritis have revealed that its potent anti-inflammatory effects protect against cartilage matrix destruction (18–20). In addition, TSG-6 is also produced during ovulation (21, 22) and cervical ripening (23), processes that have many characteristics in common with inflammation. In this regard, TSG-6 has been shown to have a critical role in murine ovulation via its interactions with inter-α-inhibitor (24, 25) and pentraxin-3 (26), leading to stabilization of the HA-rich extracellular matrix that forms around the oocyte and that is required for fertilization in vivo (27).

Considerable progress has been made in our understanding of hyaladherins with the recent determination of high resolution structures for the Type A and B HABDs of TSG-6 (9) and CD44 (10), respectively. Comparison of the Link module structures for TSG-6 and CD44 reveals that their backbones overlay extremely well, with both the positions and relative orientations of the secondary elements being highly conserved. These comparisons now allow the structural role for each conserved residue to be understood in detail and provide a firm basis for generating accurate sequence alignments and homology models of the other superfamily members. This is particularly important in the case of the Type C HABDs, for which no tertiary structures are yet available.

The three-dimensional structure of the TSG-6 Link module (termed Link_TSG6) has been determined in both its free and HA-bound conformations (9) using identical NMR methodology; in the latter state, the structure determination was performed in the presence of an HA octasaccharide (HA8), which has been shown to be the shortest length of testicular hyaluronidase-derived oligomer that binds optimally. Analysis of the free and bound structures revealed that, upon interaction with HA, conformational changes (e.g. in a loop between strands β4 and β5) and subtle side chain rearrangements result in the opening of a shallow groove on the protein surface. This groove contains the five key HA-binding residues identified previously by site-directed mutagenesis (i.e. Lys11, Tyr12, Tyr59, Phe70, and Tyr78) (28) and Arg81, which is believed to make a salt bridge with the bound HA molecule on the basis of our NMR studies (9). In this regard, isothermal titration calorimetry (ITC) experiments performed over a range of NaCl concentrations indicate that Link_TSG6 makes between one and two salt bridges with HA8 (29), with the available mutagenesis data indicating that Lys11 and Arg81 are the best candidates for these interactions (9). Furthermore, the polarity and register of the HA within the binding groove were determined from the analysis of discrete changes in chemical shift perturbation caused by different lengths of HA oligomer, and a simple model showed that the groove had dimensions (~20 Å long, ~10 Å wide, and ~7 Å deep) suitable for accommodating an octasaccharide (9). However, this detailed information is insufficient to determine the exact position or conformation of an HA8 molecule within the binding groove, i.e. de novo determination of a Link_TSG6-HA8 structure. The identification of intermolecular nuclear Overhauser effect (NOE) restraints and residual dipolar couplings in the bound sugar would clearly help overcome this problem, but collection of these data requires at least partial assignment of the hydrogen nuclei in the bound HA; unambiguous assignment of a repetitive molecule of this kind is extremely difficult, having only been fully accomplished to date for a free HA tetrasaccharide (30). In addition, the high interaction affinity of this system (~5 × 10^6 M⁻¹) (9) makes the use of transferred NOE experiments to determine the conformation of the bound ligand problematic (31–33). Because our attempts to crystallize the Link_TSG6-HA8 complex have so far been unsuccessful, a modeling approach with a firm experimental basis (33) represents a viable alternative means of providing new insights into the conformation of the HA molecule within the Link module ligand-binding groove.

The HA-binding groove of Link_TSG6 is noticeably rich in aromatic residues (Tyr12, Tyr59, Phe70, Tyr78, and Trp88), making it likely that stacking interactions of saccharide rings against aromatic planes (π–π stacking) have an important role in the association with HA (9). In particular, stacking interactions between sugar rings and two tyrosines (Tyr59 and Tyr78) have been implicated because these amino acids both change conformation and acquire distinct H* chemical shifts upon binding to HA8 (6.26 and 6.16 ppm in Tyr78, and 6.35 and 5.91 ppm in Tyr78); observation of distinct H* shifts results from tight, asymmetric interactions against the aromatic plane of the tyrosine ring, since these rings typically have a fast “flipping” motion, leading to averaging of their chemical shift values. The ring hydroxyl group of Tyr78 also appears to make a hydrogen bond with the sugar because a slowly exchanging H* proton was observed in the HA-bound protein that was not seen in the absence of ligand (9). In the family of 20 structures determined for Link_TSG6 in its HA-bound state (Protein Data Bank code 1o7c), the conformations of Tyr59 and Tyr78 are extremely well defined (by 50 and 30 NOEs, respectively), as evidenced by the very low root mean square deviation seen over all their heavy atoms (0.19 Å). Of particular note is the relative orientation and separation of the aromatic planes of these two residues: they lie on the same axis, with a slight twist between them, and are ~6 Å apart. Given that adjacent sugar rings in HA have an average separation of ~5–5.5 Å (34), it is reasonable to suggest that Tyr59 and Tyr78 stack with sequential saccharides in the complex. This raises the possibility of modeling an HA molecule into the open groove of Link_TSG6 (i.e. in its HA-bound conformation) based on the optimization of these sequential stacking interactions.

Here, we present a model of the Link_TSG6-HA8 complex generated by building HA into the Link module binding groove using the bound conformation of the protein, the polarity of HA, and potential stacking interactions. This model was then tested against previously acquired experimental information (the expected register and salt bridges) and newly collected NMR data. This modeling process has revealed the principal features underlying the specificity of HA recognition by the TSG-6 Link module. These features were further examined in light of new homology models generated for the bound conformers of other members of the human Link module superfamily and used to construct an HA-bound model of the Type C HABD from cartilage link protein (HAPLN1). This model reveals the residues most likely to be involved in stabilizing the inter-Link module interface, which are highly conserved across the HAPLN/CSPG subfamily. This has important implications for the condensation of these proteins along an individual HA chain to form ternary complexes and higher order structures.

**MATERIALS AND METHODS**

**HA-bound Model of Link_TSG6**—All molecular modeling calculations were performed with CHARMM Version 28b2 (35) using CHARMM22 parameter files for both protein and HA (suitably modified for carbohydrates (36), with all saccharide rings modeled in the 4C1 conformation throughout). The coordinates of the 20 lowest energy structures determined for the bound conformation of Link_TSG6 were used (Protein Data Bank code 1o7c) (9) and were kept fixed throughout the modeling process. Because only a few structural restraints were gathered for the side chain of Arg81 (9), it is poorly defined and occludes
the HA-binding groove in some of these structures. To prevent this degeneracy from artificially biasing the model binding process, the Arg^{23} side chain was replaced with that of an alanine in all structures. Both (β1→3) and (β1→4)-linked disaccharides 5 of HA (i.e. GlcAβ1→3GlcNAc and GlcNAcβ1→4GlcA, respectively) were placed in the binding groove of each protein structure in the orientation established previously (i.e. the reducing terminus over Tyr^{79}) (9) on the planes of the carbohydrate rings held above the planes of the side chains of Tyr^{29} and Tyr^{78} at the distance expected for stacking interactions (3.5 Å). A grid search was performed over all possible orientations and glycosidic angle conformations (10° increments in ϕ and ψ, giving 1296 combinations) for each disaccharide. Models with low van der Waals energies (Lennard-Jones potential of 6.60 kcal mol{−1} Å{−1}) and low rmsd were selected and subjected to energy minimization (1000 steps) on the whole octasaccharide (10,000 steps). The final stage in the modeling process was to allow for the expected salt bridge (9, 29, 29); no electrostatic forces were included in this step. At this point, the nonreducing terminal GlcA ring of the tetrasaccharide formed from the stacked (β1→4)-linked disaccharide (i.e. GlcNAc-GlcA-GlcNAc-GlcA) was found to prefer orientations that sharply exited the binding groove, and it was therefore not possible to continue with these models (see “Results”). In contrast, the nonreducing terminal GlcA ring of the tetrasaccharide formed from the (β1→3)-linked disaccharide (GlcA-GlcNAc) remained within the groove, allowing this line of modeling to be pursued further. For this tetrasaccharide, the addition of the next ring (GlcNAc) to the nonreducing terminus was performed as described above. The resulting low energy models now completely filled the binding groove, and the sugar was extended to an octamer by the addition of a monosaccharide to the nonreducing terminus and a disaccharide to the reducing terminus with the solution ϕH and ψH glycans at ϕH 30°, ψH 60°. H-15N HSQC, HNCO, and HNCA spectra were acquired at 500 MHz with a constant time period of 28 ms in the 13C domain, acquisition times of 7.92 ms (t_{1H}), 21.60 ms (t_{2H}, t_{13C}), and 128.00 ms (t_{3H}, t_{15N}) and the 13C carrier frequency set at 42.5 ppm. The H-13C-HSQC spectra for 15N-labeled Link_TSG6 (with unlabeled HA) were recorded at 500 MHz at pH values from 3.5 to 7.5 in 0.25-unit increments (161.80 ms, t_{1H}, 128.00 ms, t_{3H}, 13C), and 6.00 ms (t_{2H}, t_{1H}; and the 13C carrier frequency set to 91.2 ppm). The chemical shift changes of affected resonances were plotted as a function of pH, and the curves were fitted by nonlinear least-squares analysis to a one-site model, giving a measurement of the apparent pK_{a} values of the titrating group. The 1H-13C-15N-1H Overhauser effect spectroscopy (NOEYES-HSQC, 1H-13C-NOEYES-HSQC, and two-dimensional NOESY spectra referred to under “Results” are those reported previously (9). The chemical shifts for free and HA-bound Link_TSG6 have been deposited in BioMagResBank with accession numbers 6296 and 5839, respectively.

**Link Module Homology Models—**The primary sequences of each of the known human Link modules (i.e. from CD44, Lyve-1, stabulin-1, stabulin-2, KIA0527, HAPLN1, HAPLN2, HAPLN3, HAPLN4, aggrecan, brevican, neurocan, and versican) were aligned with that of Link_TSG6, taking into account both sequence identity and structural knowledge of the Link module fold; the resulting alignment is identical to that in Ref. 9, except for the second Link module of the Type C CSPG domain, which was removed. Residual side chains not aligned over strands β and β were moved four positions toward the N terminus. This alignment was then used to generate homology models for each Link module based on the lowest energy structure of Link_TSG6 in its HA-bound conformation, where aligned regions and side chain atoms of identical or structurally related amino acids were modeled into the Link_TSG6 coordinates (Protein Data Bank code 1G7c) (9). The remaining side chain atoms that did not have a match were initially positioned using internal distance restraints. Residual side chains not included in the same fashion, assigning them standard backbone (β-sheet) and side chain geometries while maintaining an unbroken peptide backbone over the site of insertion. All atoms within insertions/ deletions (relative to the TSG-6 sequence) were then allowed to relax in each model, over eight designated regions (i.e. residues 6–9, 10–11, 27–30, 40–43, 53–55, 70–73, 77–83, and 84–86, numbered as for Link_TSG6), using energy minimization (1000 steps of steepest descents and conjugate gradients) while the coordinates in the remainder of the structure were kept fixed. The final stage in the modeling process was to allow all side chains throughout the structure to relax (10,000 steps of steepest descents), leading to energy-minimized homology models of the Link modules in their bound conformations. Models containing bound ligand were then constructed by inserting the coordinates of the HA_{8} molecule from the lowest energy HA_{8}-bound model of Link_TSG6 (see above) into the structure file of the final Link module homology models, where these were aligned over strands β, β, and β.

Models of the cartilaginous link protein Type C HABD were generated in MOLMOL by positioning the two component Link modules (HAPLN1-1 and HAPLN1-2), both containing HA_{8} within their ligand-binding grooves, such that the bound HA molecules could form a continuous chain (requiring deletion of an overhanging disaccharide from either the reducing or nonreducing terminus of each oligomer as appropriate) by simultaneously moving the distance between the C and N termini of the first and second Link modules, respectively) that could be bridged by the intermediate leucine residues (Ser-Asn-Phe). This linker was added in an extended conformation from the C-terminal residue of HAPLN1-1 (Thr^{90}) and allowed to relax into place using 10,000 steps of steepest descent minimization. All side chain atoms were then energy-minimized, and finally, the two bound HA hexamers were joined together and minimized. Molecular diagrams were generated using MOLMOL and RASMOLE.
RESULTS

Generation of HA₈-bound Models of Link_TSG6

The initial location of HA in the binding groove of the Link module from human TSG-6 (Link_TSG6) was performed by optimizing the stacking interactions of both possible HA disaccharides (i.e. the (β1→3)- and (β1→4)-linked sugars) against the side chains of Tyr59 and Tyr78 in the polarity previously established (i.e. with the nonreducing terminal ring over Tyr59) (9). This was done for each member of the family of the 20 lowest energy protein structures (Protein Data Bank code 1o7c), searching the full range of glycosidic linkage conformational space for both disaccharides, with the coordinates of the protein kept fixed. The resulting low energy models, in which the sugar side chains had been allowed to relax relative to the protein, were extended from the nonreducing terminus by the sequential addition of monosaccharides, which were also allowed to search conformational space. In the trisaccharide formed (GlcNAc-GlcA-GlcNAc) by the addition of GlcNAc to the (β1→3)-linked disaccharide, the new N-acetyl side chain was found to be accommodated in an open pocket at the bottom of the binding groove (pocket I), with favorable glycosidic bond angles. Significantly, this pocket is present only in the HA-bound conformation of the Link module (see below). When extended to a tetrascarbohydrate (GlcA-GlcNAc-GlcA-GlcNAc), the carboxyl group on the new GlcA ring adopted orientations that readily allowed a salt bridge to be formed with Lys11. This salt bridge is expected to be present on the basis of site-directed mutagenesis, ITC, and chemical shift data (9, 28, 29); the weak restraint used to position the carboxylate of GlcA within 4 Å of the end of the lysine side chain was easily satisfied in this family of models. In contrast, the trisaccharide (GlcA-GlcNAc-GlcA) formed from the (β1→4)-linked disaccharide was not well accommodated and, when extended to a 4-mer, tended to exit vertically the binding groove due to steric clashes with the protein. Therefore, this line of models could not be pursued further; the GlcNAc-GlcA-GlcNAc-GlcA tetrasaccharide was unable to find positions that would allow the formation of the salt bridge with Lys11. Thus, in this modeling approach, we have found that an HA molecule (with C₄₁ ring conformations) can be extended only within the Link_TSG6 binding groove when the required sequential CH-π stacking interactions with Tyr59 and Tyr78 are made by a (β1→3)-linked disaccharide. Importantly, this register of HA is the same as that determined by our previous shift mapping studies with HA oligosaccharides of various lengths, which predicted that a GlcNAc ring was in intimate contact with Tyr78 (9). The low energy Link_TSG6:HA₈ models, resulting from construction of the (β1→3)-linked disaccharide, were therefore extended to an octascarbohydrate, with the inclusion of a salt bridge with Arg₆¹ (predicted from shift mapping and ITC data) (9, 29), which, like Lys₈¹, was found to be in close proximity to a GlcUA carboxylate group.

Selection of Models to Form a Family

The model complex resulting from building HA₈ into the lowest energy protein structure is shown in Fig. 1A. As can be seen, the HA molecule is readily accommodated within the binding groove and lies in good contact with all key HA-binding residues (i.e. Lys₁¹, Tyr₁², Tyr₅⁹, Tyr₇₈, and Arg₈¹), except Phe₇₀, the side chain of which was not precisely defined in the family of protein structures (9). In addition, the bound HA₈ molecule lies within the region of the protein that displays
significant chemical shift perturbations, and changes in side chain conformation, upon interaction with the ligand (9). However, the modeling approach employed was not successful for all members of the family of 20 Link_TSG6 structures, and in five of these, the HA molecule penetrated through the protein. This is not surprising because all the protein atoms were kept rigid throughout the modeling process and therefore led to steric clashes with certain less well defined amino acid side chains (in particular, Gly69 and Phe70); this lack of definition is partially a result of structure calculations for Link_TSG6 in its bound state being carried out in the absence of HA8 (9). These models, which tended to be from the higher energy protein structures, were therefore discarded. In a further four models, other steric clashes caused the HA chain to kink sharply out of the groove (Supplemental Fig. 1); pocket I was left empty in one of these, whereas in the other three cases, the 6-OH group of the GlcNAc ring was found to partially occupy the pocket. In these models, HA had very strained glycosidic angle conformations, and although the HA molecules could still form a salt bridge with Lys11 and maintained the same register as that of the model built on the lowest energy protein structure, they were considered to be outliers and were therefore discarded. The remaining 11 models (including the one arising from the lowest energy Link_TSG6 structure) were all very similar and could therefore be considered to form a family (Fig. 1B).

Only Five Saccharide Rings Are Required to Fill the Binding Groove

It is clear from the family of models that the binding groove is only long enough to interact with a stretch of five saccharide rings (from GlcA ring 3 at Lys11 to GlcA ring 7 at Arg44); in fact, the remaining three rings of the bound octasaccharide (one at the reducing terminus and two at the nonreducing terminus) are farther than 5 Å from the protein surface. Besides interactions with the key binding residues, we predict from the family of models that the side chains of Cys47, Val57, Ile61, Cys68, and Trp88 are also likely to be in intimate contact with the HA molecule. In this regard, all these residues displayed large chemical shift perturbations (e.g. Cys47 Cα, 3.14 ppm; Ile61 Cα, 2.82 ppm; and Trp88 Cα, 0.43 ppm) and a change in conformation upon binding to HA (e.g. the plane of Trp88 pitched by ~15° and yawed by ~45°).

Hydrophobic Pockets Accommodate Acetamido Side Chains

Comparison of the free and bound structures determined for Link_TSG6 (9) reveals that the HA-binding groove contains two hydrophobic pockets (I and II) that are formed upon its interaction with HA8 (Fig. 2A). Pocket I results primarily from rearrangements within the β4–β5-loop (hinged on Pro60 and Gly68), which are caused by a change in conformation of the disulfide bridge Cys47-Cys48 (9). The surface of this pocket is formed by the backbone of Ala48 and Gly68, side chain atoms in Cys47, Cys48, Ile61, and Lys72, and the CβII atoms of Tyr59 (Fig. 2B). During the modeling procedure, the bulky and hydrophobic acetamido group of GlcNAc ring 4 was found to locate in this pocket in 15 of the 20 models (including the selected family of 11 structures). In fact, HA molecules could be extended only within the binding groove if an acetamido side chain was located in this position (see above). It should be noted that this pocket is therefore largely responsible for determining the register of HA within the binding groove. As shown in Fig. 2A, a second, more exposed pocket (II) is also present on the opposite side of the binding groove in a position that can accommodate the acetamido group on ring 6. This pocket also opens upon binding to HA due to changes in side chain orientations of Trp88, Tyr78, and Tyr59 and is principally formed by Trp88 and Val57. During the modeling, the side chain of GlcNAc ring 6 located itself in this pocket, being bound in this way in 15 of the 20 models (including all of the family of 11 selected models). Analysis of these models indicated that N-acetyl group penetrates into the pocket to different depths in each case, which is not surprising because no restraint was specifically included to localize it in the pocket.

Glycosidic Torsion Angles in Modeled HA

In Fig. 3, the distribution of glycosidic φH and ψH angles across the family of 11 structures (i.e. at each of the four linkages between the five bound saccharide rings; see above) is compared with the preferred conformations predicted for HA in solution (39–41). It is clear that, considering the limitations expected for modeling HA into rigid protein conformations, the HA8 molecule has bond angles at each linkage remarkably similar to those predicted for HA in solution. Of particular interest is the observation that the relative orientations of the planes of Tyr59 and Tyr78 prefer stacking interactions to be made with both rings 5 and 6 in their solution conformation; in Fig. 3A, only 2 of the 11 models lie outside the low energy contours. Accommodation of the GlcNAc ring 4 acetamido side chain in pocket I serves to introduce twists at the glycosidic φH
angle of both its (β1→3)-linkage (−90°) (Fig. 3C) and (β1→4)-linkage (−60°) (Fig. 3D), generating distributions of conformations skewed away from those predicted in solution. Also note that, with this arrangement of glycosidic conformations, both 6′-hydroxyl groups (on GlcNAc rings 4 and 6) are solvent-exposed, which, because of their high mobility relative to other groups within HA, might be expected on entropic grounds. Therefore, in these models, HA was found to be able to be accommodated in the groove in a conformation that is similar to an allowed free solution conformation. A consequence of this is that most of the HA intramolecular hydrogen bonds are maintained in the models, which would also be energetically advantageous.

Roles of Aromatic Residues in Binding to HA

Unlike Tyr59 and Tyr78, the orientation and position of Tyr12 in the binding groove leave it unable to form a CH–π stacking interaction with HA in the model; consistent with this, Tyr12 does not show distinct Hα chemical shifts upon binding to HA6. It does, however, acquire a slowly exchanging Hn proton upon interaction with HA6, and as shown in Fig. 4A, this group is clearly pointing toward the octasaccharide in a favorable orientation to make a hydrogen bond (predicted to be to O-2 of ring 4). The models also suggest that the slowly exchanging Hn proton of Tyr78, seen in the HA6-bound protein (9), makes a hydrogen bond with O-4 of ring 2. The slight twist introduced into the bound HA6 molecule by accommodating the acetamido group of ring 4 in pocket I appears to serve to direct the hydrophobic face of this saccharide ring toward Phe70 (Fig. 4A); the hydrophobic face of this ring was found to be directly opposite the side chain of Phe70 in 6 of the 11 selected models. It therefore seems plausible that Phe70 makes a stacking interaction against this ring, closing over the HA molecule in the groove and holding it in place.

Testing the Model

Chemical Shifts of Bound HA6—Samples containing 15N- or 13C,15N-labeled HA6 bound to unlabeled Link_TSG6 were prepared for use in isotope-edited experiments, in which only the HA6 component of the complex was visible. 1H-,15N-HSQC, 1H-,15N-TOCSY-HSQC, and 1H-,13C-HSQC spectra were recorded (Fig. 5), allowing the measurement of all 1H, 13C, and 15N chemical shifts within the bound HA6 molecule, which aids in the identification of intermolecular NOEs (see below). As expected for NMR spectra of such a repetitive oligosaccharide, the chemical shifts were highly degenerate, leading to extensive resonance overlap, and it is therefore currently not possible to assign the majority of these resonances to specific nuclei. Nevertheless, comparison of these spectra with published values for the free sugar (30, 42–45) allows some general conclusions to be drawn.
Seven resonances are visible in the $^1$H,$^{15}$N HSQC spectrum of HA$_5$ in complex with protein (Fig. 5A). Species $\alpha$ and $\beta$ arise from the $\alpha$- and $\beta$-anomers of the reducing terminal GlcNAc ring (9, 45, 46), and the corresponding $C^1$, $C^2$, $C^{Me}$, $H^2$, and $H^3$ chemical shifts of ring 8 were measured (using the spectra shown in Fig. 5B in combination with HNCA- and HCCH-TOCSY experiments) and found to have near identity to the free sugar (30, 44); however, these cannot be accounted for by proximity to protein glutamate/aspartate residues, reveal the location of the unneutralized carboxylate group on ring 5.

The amide groups from the other three GlcNAc rings in Link_TSG6-HA$_5$ are partially resolved due to the presence of the protein. Thus, clear intermolecular NOEs (Fig. 5C) allow potential intermolecular NOEs (between Link_TSG6 and HA$_5$) to be examined in order to determine whether an appropriate chemical shift value is present in the protein, which causes chemical shift perturbations away from the solution values, generating resonances $\epsilon$–$g$ (Fig. 5A). The perturbation to resonances $e$–$g$ is quite large in the proton dimension (0.4–1.1 ppm upfield), consistent with the binding of one or more N-acetyl side chains in unique environments and/or their close proximity to aromatic residues, both of which the model predicts for rings 4 and 6. However, because there are only four amide groups within the HA$_5$ molecule (expected to give rise to five NH resonances), the presence of seven species in the $^1$H,$^{15}$N HSQC spectrum indicates additional complexity in the mechanism of HA binding (e.g. with regard to dynamics or multiple conformations), which is as yet not fully understood.

It is relatively straightforward to assign regions of the $^1$H,$^{13}$C-HSQC spectrum (Fig. 5C) to groups within a disaccharide (e.g. GlcNAc $C^2$ and $H^3$) by comparison with published values for the free sugar (30, 44); however, these cannot be assigned as yet to an individual ring (e.g. ring 2 or 4). Many of the observed resonances are very similar or identical to those of free HA$_5$, which is not surprising given that only five rings are predicted from the modeling to be in direct contact with the protein surface, i.e. the three non-interacting rings would be expected to have unperturbed chemical shift values. Although the methyl region in the $^1$H,$^{13}$C-HSQC spectrum has considerably more than four peaks (corroborating the observations from the $^1$H,$^{15}$N-HSQC spectrum described above), it is clear that some of these show distinct chemical shift perturbations (Fig. 5C). This indicates that certain methyl groups are bound to the protein in unique environments, such as, for instance, the hydrophobic pockets predicted from the molecular modeling.

**Intermolecular NOEs—Total (100%) assignment of $^1$H nuclei within the bound protein has been achieved, and high resolution structures of the protein have been determined (9), thereby allowing each NOE in the NOESY spectra to be accounted for with high confidence. Furthermore, observation of all $^1$H chemical shifts in the bound HA$_5$ molecule (i.e. in the HSQC spectra shown in Fig. 5) allows potential intermolecular NOEs (between Link_TSG6 and HA$_5$) to be examined in order to determine whether an appropriate chemical shift value is present in the sugar. In this regard, 13 NOEs from protons with chemical shifts that match those observed for the bound HA$_5$ ring protons (over the range 3.2–4.7 ppm) to amino acids within the protein ligand-binding site could not be assigned to groups within Link_TSG6 (Supplemental Table 1). Therefore, these represent extremely good candidates for intermolecular NOEs, especially because there were no remaining NOEs over this chemical shift range unaccounted for within the rest of the protein. Thus, clear intermolecular NOEs are evident, for example, to Ile$^{61}$ H$_{\alpha}$ (from 4.50, 3.74, and 3.55 ppm in HA$_5$) and Tyr$^{78}$ H$_{\alpha}$ (3.84 and 3.74 ppm), indicating the intimate contact of these residues with the HA ring hydrogen atoms, as...
seen in the model. In addition, two further NOEs from 1.98 ppm to Ala49 HN and Val46 HN (both found near the bottom of pocket I) are of particular interest because they could originate only from GlcNAc methyl groups (i.e. with a modest chemical shift perturbation of 0.06 ppm from its free solution value) (42). If these are from the same methyl group in the bound HA8, which is likely because Ala49 HN and Val46 HN are close together in the structure, they would position a methyl group at the bottom of pocket I (Supplemental Fig. 2), in direct confirmation of this aspect of the model.

Location of GlcA Carboxylate Groups—Previous work has indicated that the complex is likely to have two salt bridges between carboxylates of GlcA residues and the basic amino acids Lys11 and Arg81 (9, 28, 29). These were easily accommodated in the model, and it should be noted that only weak restraints were used to place the Lys11/Arg81 side chains in positions where they were able to make appropriate ionic interactions. Furthermore, in the case of Lys11, the mean distance between the carboxylate oxygen and N/H9256 atoms was 3.2 Å across the family of 11 models, showing that the “restraint” of 4 Å had no influence. Thus, the protein is predicted to bind to alternate carboxylate groups on HA8 (i.e. rings 3 and 7), leaving the charge on ring 5 pointing toward the /H9252–/H9252-loop (Fig. 4B). It should be noted that, although the side chain of Lys72 is close to this carboxylate, it is almost certainly not making a salt bridge because there are no chemical shift perturbations to the end of the Lys72 side chain upon binding HA8 and the H nuclei have the typical “free solution” value of 2.99 ppm in both free and bound states. Furthermore, mutation of Lys72 to Ala does not cause a reduction in affinity (28).

Basis for Homology Modeling of Link Module Superfamily Members

Recently, the structures of the Link modules of TSG-6 and CD44 have been determined at high resolution (9, 10), revealing that the backbone topology is highly conserved (as would be expected with a 32% sequence identity). Furthermore, comparison of the functions of side chain groups in these structures now permits the role of each conserved residue to be understood in detail (Supplemental Table 2). For example, the hy-
drophobic residues involved in forming the Link module core are well understood and highly conserved across the family, and the conservation of other particular small or charged residues arises from features such as helix caps (e.g., Thr15/Glu18 and Thr32/Gln35, numbered as in Link_TSG6), turns (e.g., Ala31), and the close approach of secondary elements (e.g., Gly50 allows the parallel hydrogen bonding of strands β3 and β6). This knowledge has been combined to produce a precise multiple sequence alignment for the superfamily as a whole (Fig. 7), with sequence identities to Link_TSG6 of between 30 and 40%, with the exceptions of KIA0527 (22%), stabilin-1 (47%) and stabilin-2 (43%). Therefore, this sequence alignment can be used for the generation of reliable homology models of the other Link module family members in their open (i.e., HA-}

**Fig. 6.** Chemical shift changes with pH of selected nuclei from HA₆-bound Link_TSG6, with data fit to a one-site model (smooth curves).

**Fig. 7.** Multiple sequence alignment used in the construction of Link module homology models. Residues implicated by site-directed mutagenesis or NMR studies to be involved in HA binding in TSG-6 and CD44 are shown in boldface red, whereas those determined to play no role in the interaction are shown in lowercase (28, 49, 50). Amino acids likely to be involved in HA binding in other members of the Link module superfamily (i.e., based on detailed analysis of the homology models) are depicted in red. Residues that are required for the formation of binding pockets I and II (indicated by asterisks) are shown in green and blue, respectively, and the critical hinge residues are denoted (#). Amino acids predicted to form Link module-Link module intramolecular and intermolecular (i.e., HAPLN/CSPG) interfaces in the Type C domains are highlighted in magenta boxes and cyan dashed boxes, respectively. Three residues that have been predicted by others (54) to contribute to the inter-Link module interface in the Type C domains in versican are underlined. The locations of the secondary structure elements of Link_TSG6 are indicated at the top.
FIG. 8. HA-binding sites in selected Link module homology models. The putative HA-binding residues indicated in Fig. 7 are colored gray in a space-filling representation of the protein. The stabilin-2 Link module is predicted to be functionally active and has an HAα, constructed in its binding groove (in the same conformation as that determined for Link_TSG6). KIA0527 has very few of the features expected for a functional Link module and, for example, has two negatively charged residues (−) in its “binding” groove. Aggrecan1 and brevican1 are both predicted to contribute to the binding capabilities of their Type C HABDs, and HAβ, has been included in their binding groove to highlight both the similarities and differences expected in the details of the protein-HA interaction (e.g., aggrecan1 has two additional basic amino acids (+)).

bound (state based on the HAα-bound Link_TSG6 coordinates (Protein Data Bank code 107c) (9). Analysis of the models for the second Link module from the CSPGs (i.e., aggrecan-2, brevican-2, versican-2, and neurocan-2) and the related aggrecan-4, which constitute a subgroup of the superfamily (see Fig. 7), indicated that these overlaid on the Link_TSG6 structure with a significantly lower root mean square deviation compared with the models generated from the alignment of Blundell et al. (9), where the insertion between strands β3 and β6 was in a different position. It should be noted that all other deletions and insertions (such as the long loop in HAPLN4-1 between strands β4 and β5 and the insertion between strands β5 and β6 in the second Link module of the CSPGs) could be easily accommodated within the Link module tertiary structure. Therefore, the 25 Link module models constructed here in their ligand-bound conformations allow the HA binding capabilities of these domains to be re-evaluated and also provide a firm basis for future programs of site-directed mutagenesis.

**Prediction of HA-binding Residues in Other Link Modules**

Inspection of the binding groove in each homology model for the presence of aromatic and basic amino acids (i.e., the residue types shown to be preferred in the interactions of HA with TSG-6 and CD44) (28, 49, 50) allows a prediction of the most probable residues that contribute to ligand binding (aliphatic residues capable of hydrogen bond interactions, such as Gln and Asn, have also been included where appropriate) (Fig. 7). This analysis therefore provides insight into whether certain Link modules are likely to interact with HA. As noted recently (7), the Link module of stabilin-2 is likely to be an HABD because it has a significant number of putative functional residues (Fig. 8), four of which are at equivalent sequence positions to key amino acids in Link_TSG6. Conversely, analysis of the homology model for KIA0527 indicated that it is unlikely to bind HA via its Link module given its low number of suitable basic and aromatic amino acids (Fig. 8). In fact, its binding groove contains two negatively charged residues that would repel HA, in marked contrast to the other Link module binding surfaces that are devoid of acidic residues (except where a candidate salt-bridging partner is also present, e.g., Asp45–Arg10 in neurocan-2). The functional status of the stabilin-1 Link module is rather more difficult to determine on the basis of sequence and modeling because its groove does contain several potential HA-binding residues, albeit a lower number than other bona fide hyaladherins. However, recent studies on stabilin-1 (and constructs containing the Link module) have clearly indicated that this protein does not interact with HA (51).

All of the other Link module models have structures and sequences consistent with a capability to interact with HA. From the multiple sequence alignment shown in Fig. 7 (where the putative HA-binding residues are colored red), it is clear that most functional Link modules appear to share more features in common with TSG-6 than CD44. For example, Tyr59 and Tyr78 of Link_TSG6, although not present in CD44 (or Lyve-1), are highly conserved across the superfamly as a whole. In the case of Tyr59, an equivalent tyrosine is found in 16 out of 18 of the HAPLN/CSPG Link modules, whereas the other two have a phenylalanine or histidine ring at this site; Tyr78 is absolutely conserved in the first Link modules of HAPLN/CSPGs, whereas in the second, it is replaced by Phe, Leu, or Val, i.e., aromatic or large and planar faced hydrophobic residues that could also stack against a GlcNAc ring. Furthermore, the critical residues Arg78 and Tyr79 of the CD44 HABD (50) are almost unique to this receptor, with no other examples of a basic amino acid at this position and only two Link modules having a tyrosine at a comparable sequence location (i.e., Lyve-1 and versican-2). It should also be noted that CD44 has been hypothesized to have two modes of ligand binding (10), where mode 1 is likely to be equivalent to the HA-binding site described above for TSG-6. Therefore, CD44 and the related HA receptor Lyve-1 (which both have Type B HABDs) (10) appear to be outliers of the superfamily, and therefore, some caution is required when interpreting models for these proteins based on TSG-6.

As shown in Fig. 7, the residues contributing to the formation of pockets I and II (colored blue and green, respectively) are very highly conserved across the superfamily, indicating that accommodation of GlcNAc side chains in this way is likely to be a general feature of HA binding to Link modules, including those of CD44 and Lyve-1 (except that the latter does not have the hinging proline residue of pocket I). KIA0527, however, does not have most of the residues required to form pocket I and lacks the disulfide bridge that permits opening of the groove, providing further evidence that it is unlikely to be functionally active with regard to HA binding. Importantly, because the combination of pocket I and the central aromatic residue Tyr59 (which is highly conserved as described above) determines the binding register in Link_TSG6, it is to be expected that other Link modules bind HA in the same polarity and (approximate) register. Consistent with this, basic residues at positions 11 and 81 are found in many Link modules, i.e., in the same location relative to the pocket and central aromatic residue. Nevertheless, although the register and orientation of HA within the binding groove are probably conserved in most cases, the details of the molecular interaction are expected to be subtly different in each case; for example, aggrecan1 has two additional basic amino acids in its binding site (compared with brevican1) at positions equivalent to Lys72 and Arg73 in Link_TSG6 (Fig. 8). In this regard, although a basic residue at position 11 is very highly conserved across the superfamly (and functionally implicated in both TSG-6 and CD44) (28, 49, 50), recent mutagenesis data for neurocan indicate that, in this case, an arginine at this location does not
contribute to HA binding in either of its Link modules. This suggests that there is considerable diversity in the way specific Link modules interact with HA, perhaps stabilizing different conformations of the sugar as suggested previously, and comprehensive programs of mutagenesis are now required to determine which of the amino acids identified as potentially functional by our analysis are actually involved in binding.

**Molecular Model of Cartilage Link Protein**

A model of the Type C HABD (i.e. two tandem Link modules) from cartilage link protein (HAPLN1) in its HA-bound form was made by combining the homology models for the individual Link modules using an HA molecule bound in each groove as a guide (Fig. 9A). This model indicates that, at most, nine sugar rings could make contact with the protein, which is consistent with data on human recombinant HAPLN1, where the minimal length of HA oligomer that can compete for polymeric HA binding is a decasaccharide. As shown in Fig. 9, there are six putative HA-binding residues in the first Link module, but only four in the second, suggesting that the former may play a more significant role in the interaction.

An important observation from this modeling exercise was that four aromatic and two large hydrophobic residues were found at the predicted Link module-Link module interface (Fig. 9B, cyan), whereas a conserved glycine residue at the start of helix 2 in the second domain was found to greatly aid the close approach of the two modules (where a larger side chain would result in a steric clash). Because these amino acids are seen to be well conserved across the HAPLN/CSPG Link modules (Fig. 7, magenta boxes) and are not highly conserved in the Type A and B HABDs (i.e. those that have a single Link module), these sequence positions are predicted to form the interdomain interface in the other Type C HABDs.

Both component Link modules in the Type C domain are distinctly wedge-shaped, and therefore, their tandem association to form a continuous HA-binding site naturally results in the introduction of a gentle curve into the bound HA chain (Fig. 9, A and B). Interestingly, aromatic/hydrophobic residues also cluster on the free “ends” of HAPLN1 (Fig. 9B, cyan), making two additional faces ideal for interaction with other proteins; these residues are also well conserved within the HAPLN/CSPG Link modules. In fact, during construction of the model, it was found that the opposite order of Link modules is also a reasonable organization for the structure of HAPLN1 (i.e. with the two modules in Fig. 9A and B swapped around) because these hydrophobic patches come together and make an interface (although in this case, it is smaller, and some charged residues are slightly buried). This potential to form Link module-Link module interfaces on either side of an individual Link module may underlie the ability of the HAPLN proteins to aggregate with CSPGs on a single HA chain, i.e. one side of each module would form intramolecular contacts, whereas the other would form intermolecular ones. A clear consequence of this hypothesis is that they could fit together in a repeated array, and this tessellation would propagate the gentle curve introduced into the HA at each binding site, giving rise to higher order superhelical structures. This model is consistent with the finding that the Link modules of versican mediate its interaction with HAPLN1; it could also potentially explain how versican alone could bind HA in a cooperative manner as has been suggested recently. For aggregcan, the Ig module is necessary for binding to HAPLN1 in the absence of HA, but this does not preclude that intermolecular Link module-Link module interactions take place in the ternary complex. In this regard, we predict that Ig modules could be accommodated in the inner face of the helix, where their “homotypic” interaction should further stabilize the higher order complex. However, at this point, there are no experimental data permitting the exact position of the Ig domains to be determined.
DISCUSSION

A model of the Link_TSG6-HA$_8$ complex has been generated that satisfies all the available structural data. Although it is likely that this model is a good representation of the solution structure, being consistent with newly acquired data, it does not yet constitute a de novo structure because it relies upon the assumption that the protein side chain orientations in the binding groove are accurate. This is likely to be the case for pocket I, Tyr$_{12}$, Tyr$_{59}$, and Tyr$_{78}$ (because these have each been defined by many NOE restraints); however, certain other residues (in particular, Phe$_{79}$) are still poorly defined. Nevertheless, this model is instructive and compares favorably with other protein-carbohydrate structures.

As mentioned above, stacking interactions between sugar rings and aromatic side chains are ubiquitous in protein-carbohydrate complexes (56–58), and three such contacts can readily be envisaged within the Link_TSG6-HA$_8$ complex (i.e., with Tyr$_{59}$, Phe$_{79}$, and Tyr$_{78}$). For instance, CH-π stacking interactions have been found to be partly responsible for the precise positioning of HA (for catalysis) in hyaluronate lyases from *Streptococcus pneumoniae* (59–62) and *Streptococcus agalactiae* (63, 64), and in the case of Link_TSG6, it appears that the stacking interactions with Tyr$_{59}$ and Tyr$_{78}$ precisely position the bound HA in a similar fashion. In addition, the model also provides insight into the molecular basis of the selectivity of Link_TSG6 for HA over other carbohydrate ligands. Specifically, two hydrophobic pockets select for the alternating GlcNAc side chains of HA, whereas two basic residues at the correct separation (∼20 Å) provide specificity for the GlcA groups. The binding of a GlcNAc group into a hydrophobic pocket has also been observed in the crystal structures of bee venom hyaluronidase (65) and tachylectin-2 (which binds to single GlcNAc moieties) (66), suggesting that this is a common method of selecting for and interacting tightly with saccharides containing acetamido side chains. In this regard, it has been shown previously that extending the length of the GlcNAc side chains in HA, through chemical modification, inhibits its binding to aggrecan (67), presumably because they are too long to fit into the pockets and still retain the stacking interactions. The narrowness of the groove, particularly at the site of the disulfide bridge (Cys$_{72}$-Cys$_{86}$), would also be expected to prevent the more bulky sulfated glycosaminoglycans (e.g., chondroitin and heparan sulfates) from binding at this surface; it should be noted that, although chondroitin 4-sulfate and heparin do bind to Link_TSG6 (68, 69), these interactions are mediated via a totally separate binding surface.

An additional interesting observation can be made concerning the relative arrangement of Lys$_{11}$ and Tyr$_{12}$. It has been observed that the methylene groups of lysine and arginine side chains often align against aromatic rings by van der Waals contacts, with the positive charge at the periphery of the ring (70); this interaction can serve to orient the basic side chain for specific salt bridging and increase the strength of the interaction when formed, probably by prior exclusion of water (47, 48). In this regard, the H$_2^\delta$ proton of Lys$_{11}$ is significantly upfield-shifted (to 0.42 ppm), implying that the side chain methylene groups are indeed aligned against Tyr$_{12}$ (as apparent in our HA$_8$-bound models). This interaction may orient the Lys$_{11}$ side chain into the correct position for receiving the GlcA carboxylate group upon binding and may account for the abundance of Lys-Tyr (and Arg-Tyr) pairs seen in the HA-binding sites of Link modules (Fig. 7). However, the observation of a Lys-Tyr (or Arg-Tyr) pair in an HA-binding groove does not necessarily indicate a salt bridge because recent site-directed mutagenesis studies of the basic residue at position 11 in the neurocan Link modules have demonstrated that these amino acids are not involved in HA binding (52).

The HA$_8$-bound model clearly shows that the binding groove is only long enough to accommodate a pentasaccharide (GlcA to GlcA). At first sight, this is in apparent contradiction to the results from ITC, which showed not only that a heptasaccharide (GlcA to GlcA) is the minimal size of oligomer that binds with maximal affinity, but also that its affinity is ∼40 times greater than this pentamer (9). Recently, however, we have shown that end effects are experienced even in the middle of oligomers as long as HA$_8$ (42), and thus, it is not surprising that lower affinities are seen with shorter oligomers (e.g., HA$_3$), where the more dynamic nature of their free ends might be expected to incur a greater entropic penalty upon binding. Therefore, data from ITC and plate-assay based competition experiments are likely to overestimate the minimal length of HA oligomer that interacts with a protein by a few sugar residues. We also note that, because differences in dynamic motion can apparently be propagated to a distance of three to four saccharide rings (42), it is likely that binding of a protein to HA will disrupt the local chain dynamics of the unbound sugar on either side of the binding site, which may assist in the polymerization of hyaladherins along an HA chain.

The recent determination of the structures for the Link modules of both TSG-6 and CD44 has provided detailed insight into the domain architecture, allowing the generation of improved homology models for the other superfamily members. Examination of these models in light of the knowledge gained from the HA$_8$-bound model of Link_TSG6 reveals that the mechanisms employed for specifically selecting HA as ligand are likely to be maintained across the superfamilies (i.e., hydrophobic pockets and precise stacking interactions with aromatic rings and basic residues at the correct separation). In addition, the models allow more accurate predictions to be made for the key HA-binding residues in each module and thereby provide a basis for new experiments. Using the modeled HA as a restraint, we have also been able to predict the likely protein-protein interface sites in the Type C HABDs, revealing that large hydrophobic patches on both sides of the individual Link modules probably allow for cooperative condensation along a common HA chain. These models supersede a previous attempt (54) to construct the Type C HABD of versican for several reasons. First, the models of the individual versican Link modules used by Matsumoto et al. (54) were based on the coordinates for unbound Link_TSG6 determined by Kohda et al. (8). These have been shown, by the more recent high resolution structures for Link_TSG6 (9) and the CD44 HABD (10), to have poor definition of side chain residues on the surface of the protein and an incorrect length and orientation of helix α2. These differences in surface features have a significant impact on the positioning of the two Link modules relative to each other, especially because helix α2 forms a large part of the interface site in both sets of models. Furthermore, our increased understanding of the roles of conserved amino acids in the Link module structure (Supplemental Table 2) has allowed a more precise alignment of sequences, leading to the generation of considerably more accurate homology models. It is therefore no surprise that the inaccurate sequence alignment used by Matsumoto et al. (54) led to the erroneous conclusion that a patch of three residues (Met-Gly-Lys; underlined in Fig. 7) is crucial to the formation of the inter-Link module contacts, although none of these residues are conserved across the HAPLN/CSPG family, and it necessitates the burial of negative charges at the interface in other Type C domains. Finally,
although we have been able to use both the linker peptide and bound HA in our models to determine the most likely interface site, neither of these approaches were available to Matsumoto et al.

The future direction of this work is clearly to derive more structural restraints for the Link_TSG6-HA6 complex to provide specific information for accurately positioning the octasaccharide within the binding groove (i.e. intermolecular NOEs, glycosidic bond coupling constants, and residual dipolar couplings). However, collection of such restraints largely relies upon successful assignment of the bound ligand, which, as noted in the Introduction, is far from trivial. To this end, we have recently developed a method for the synthesis of isotopically enriched HA oligosaccharides (9, 42). In this study, we have used these unique compounds to begin the first assignment of an HA molecule bound to protein, and these data clearly indicate that new NMR experiments are needed if we are to achieve a full assignment. Furthermore, new methods are required to determine the precise location of intermolecular hydrogen bonds and bridging water molecules, which are common mechanisms for specifically selecting saccharide epimers and for tight binding to unadorned sugar rings (33, 56, 58).

Modeling of Protein-Hyaluronan Complexes

REFERENCES

1. Tammi, M. I., Day, A. J., and Turley, E. A. (2002) J. Biol. Chem. 277, 4581–4584
2. Toole, B. P., Wight, T. N., and Tammi, M. I. (2002) J. Biol. Chem. 277, 4593–4596
3. Wight, T. N., and Merrilees, M. J. (2004)
4. Toole, B. P. (2004)
5. Blundell, C. D., Seyfried, N. T., and Day, A. J. (2004) in
6. Carrette, O., Nemade, R. V., Day, A. J., Brickner, A., and Larsen, W. J. (2001)
7. Blundell, C. D., Seyfried, N. T., and Day, A. J. (2001)
8. Kohda, D., Morton, C. J., Parkar, A. A., Hatanaka, H., Inagaki, F. M., Campbell, I. D., and Day, A. J. (2000) EMBO J. 19, 1228–1240
9. Mindrescu, C., Dias, A. A., Olszewski, R. J., Klein, M. J., Reis, L. F., and Dias, A. A., Olszewski, R. J., Klein, M. J., Reis, L. F., and Parkar, A. A., Hatanaka, H., Inagaki, F. M., Camp-
10. Teriete, P., Pickford, A. R., Campbell, I. D., and Day, A. J. (2003) J. Biol. Chem. 278, 49261–49270
11. Milner, C. M., and Day, A. J. (2003) J. Cell Sci. 116, 1863–1873
12. Bajgoraj, J., Greenfield, B., Munro, S. B., Day, A. J., and Almond, A. (1998) J. Biol. Chem. 273, 338–343
13. Prevo, R., Banerji, S., Ni, J., and Jackson, D. G. (2004) J. Biol. Chem. 279, 52580–52592
14. Rauch, U., Hirakawa, S., Ohashi, T., Kappfer, J., and Roos, G. (2004) Matrix Biol. 23, 629–639
15. Seyfried, N. T., Blundell, C. D., Day, A. J., and Almond, A. (2004) Glycobiology 14, 999–1009
16. Scinska, W., Adams, B., and Lerner, L. (1993) Carbohydr. Res. 242, 29–51
17. Cowman, M., Hittner, D. M., and Fedor-Davis, J. (1996) Macromolecules 29, 2904–2902
18. Holmbeck, S. M. A., Pettila, P. A., and Lerner, L. E. (1994) Biochemistry 33, 12446–12455
19. Cowman, M. K., Cozart, D., Nakashima, K., and Balazs, E. A. (1984) Arch. Biochem. Biophys. 230, 203–212
20. Thompson, S. E., and Smithrud, D. B. (2002) J. Biol. Chem. 277, 597–604
21. Almond, A., Brass, A., and Sheehan, J. K. (1997) Glycobiology 7, 973–980
22. Almond, A., Brass, A., and Sheehan, J. K. (2000) J. Phys. Chem. B 104, 5634–5640
23. Blundell, C. D., DeAngelis, P. L., Day, A. J., and Almond, A. (2004) Glycobiology 14, 999–1009
24. Scinska, W., Adams, B., and Lerner, L. (1993) Carbohydr. Res. 242, 29–51
25. Quiocio, F. A. (1995) Biochem. Soc. Trans. 21, 442–448
26. Wei, W., and Drickamer, K. (2000) Annu. Rev. Biochem. 69, 451–473
27. Poronkay, K., and Ochsenbauer, J. (2002) J. Biol. Chem. 277, 32828–32835
28. Poronkay, K., and Ochsenbauer, J. (2002) J. Biol. Chem. 277, 32828–32835
29. Blundell, C. D., Kahmann, J. D., Perczel, A., Mahoney, D. J., Cordell, M. B., Twarok, P., Campbell, I. D., and Day, A. J. (2002) in Hyaluronan (Kennedy, J. F., Phillips, G. O., Williams, P. A., and Hascell, V. C., eds) Vol 1, pp. 161–172, Wodhead Publishing Ltd., Abington Hall, Cambridge, United Kingdom
30. Toffanin, R., Kvam, B. J., Flashani, A., Atzori, M., Biviano, F., and Paoletti, S. (1993) Carbohydr. Res. 245, 113–128
31. Kogelberg, H., Solis, D., and Jimenez-Barbero, J. (2003) Curr. Opin. Struct. Biol. 13, 646–653
32. Johnson, M. A., and Pinto, B. M. (2004) Biochemistry 43, 5653–5663
33. Bush, C. A., Martin-Pastor, M., and Imberty, A. (1999) Annu. Rev. Biophys. Biomol. Struct. 28, 209–293
34. Atkins, E. D., Phelps, C. P., and Sheehan, J. K. (1972) Biochem. J. 128, 1255–1263
35. Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) J. Comp. Chem. 4, 187–217
36. Almond, A., Petersen, B. O., and Duus, J. O. (2004) Biochemistry 43, 5653–5663
37. DeAngelis, P. L., Jing, W., Drake, R. K., and Achyuthan, A. M. (1998) J. Biol. Chem. 273, 8454–8455
38. Kahmann, J. D., O’Brien, R., Werner, J. M., Heinegard, D., Labudie, J. E., Campbell, I. D., and Day, A. J. (2000) Structure 8, 763–774
39. Almond, A., Sheehan, J. K., and Brass, A. (1997) Glycobiology 7, 597–604
40. Almond, A., Brass, A., and Sheehan, J. K. (1997) Glycobiology 8, 973–980
41. Almond, A., Brass, A., and Sheehan, J. K. (2000) J. Phys. Chem. B 104, 5634–5640
42. Blundell, C. D., DeAngelis, P. L., Day, A. J., and Almond, A. (2004) Glycobiology 14, 999–1009
43. Scinska, W., Adams, B., and Lerner, L. (1993) Carbohydr. Res. 242, 29–51
44. Cowman, M., Hittner, D. M., and Fedor-Davis, J. (1996) Macromolecules 29, 2904–2902
45. Holmbeck, S. M. A., Pettila, P. A., and Lerner, L. E. (1994) Biochemistry 33, 12446–12455
46. Cowman, M. K., Cozart, D., Nakashima, K., and Balazs, E. A. (1984) Arch. Biochem. Biophys. 230, 203–212
47. Thompson, S. E., and Smithrud, D. B. (2002) J. Am. Chem. Soc. 124, 442–449
48. Black, J. C., and Schirmer, T. (2000)
49. Thompson, S. E., and Smithrud, D. B. (2002) J. Am. Chem. Soc. 124, 442–449
50. Black, J. C., and Schirmer, T. (2000)
51. Thompson, S. E., and Smithrud, D. B. (2002) J. Am. Chem. Soc. 124, 442–449
52. Black, J. C., and Schirmer, T. (2000)
53. Thompson, S. E., and Smithrud, D. B. (2002) J. Am. Chem. Soc. 124, 442–449
54. Black, J. C., and Schirmer, T. (2000)
Towards a Structure for a TSG-6·Hyaluronan Complex by Modeling and NMR Spectroscopy: INSIGHTS INTO OTHER MEMBERS OF THE LINK MODULE SUPERFAMILY
Charles D. Blundell, Andrew Almond, David J. Mahoney, Paul L. DeAngelis, Iain D. Campbell and Anthony J. Day

J. Biol. Chem. 2005, 280:18189-18201.
doi: 10.1074/jbc.M414343200 originally published online February 17, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M414343200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2005/03/03/M414343200.DC1

This article cites 68 references, 29 of which can be accessed free at
http://www.jbc.org/content/280/18/18189.full.html#ref-list-1