Domains in cartilage proteoglycans: do they define function?

PETER J. NEAME
Shriners Hospital for Crippled Children, 12502 N. Pine Drive, Tampa, FL 33612, U.S.A.

Proteoglycans (PGs) have, until recently, been poorly characterized at the amino acid sequence level. Since 1985, when the first proteoglycan core protein structure was determined [1], structural information on these complex molecules has become relatively plentiful. Proteoglycans are a good family of protein construction through the assembly of domains. It is tempting to assume that the individual domains define discrete functions: this will be discussed here.

The major proteoglycans of cartilage are the large, aggregating chondroitin sulphate proteoglycans and the small, dermatan/chondroitin sulphate PGs (PG-S1 and PG-S2). The protein cores of these PGs have recently been sequenced. The sequence of the large PG from rat chondrosarcoma was determined by Doege et al. [2] and the sequences of small PGs were determined by Krusius & Ruoslahti (PG-S2 from human skin) [3], Day et al. (PG-S2 from bovine bone) [4], Fisher et al. (PG-S1 from human bone) [5] and Neame et al. (PG-S1 from bovine cartilage) [6]. All of these molecules have discrete domains which are amenable to theoretical analysis of their structure and function. In some cases, it has been possible to analyse their function experimentally and examine their structure at a gross level by rotary shadowing.

Cartilage aggregating proteoglycan

This proteoglycan forms large aggregates with hyaluronic acid (HA) and in this form makes up much of the bulk of cartilage due to the water associated with its glycosaminoglycan (GAG) chains. The first discrete domain found in this PG was the HA-binding region (HABR) [7]. Recently, other domains have been identified [8]. This discussion will be restricted to the two domains at the N-terminal, referred to as G1 (HABR) and G2. These structures have a striking resemblance to link protein (LP), which is a 40–50 kDa glycoprotein required for the formation of large stable aggregates of the PG with HA. Primary structures are available for rat, chick and pig LP [9–11]. G1 is similar to the whole of link protein, while G2 is similar to the C-terminal-half of either LP or G1. This region contains two loops, the proteoglycan tandem repeats or PTRs.

The N-terminal of LP and G1 is a loop which falls into the family of the immunoglobulin folds [12] and is rather less conserved between LP and G1. The immunoglobulin fold contains seven antiparallel β-sheets. Perkins et al. [13] have analysed the similarity between IgG V-regions and LP and G1 and found that there is evidence for β-sheets in this region. It is therefore reasonable to assume that G1 and LP will interact in a similar way to the light and heavy chains of an IgG V-region. Experimental evidence for this dimer exists [14]. The regions which would correspond to the β-sheets are particularly conserved in four out of the seven potential sheets. The conserved regions correspond to internal β-sheets or structures involved in the close approach of two V-regions.

Based on this model, the aggregate of LP and G1 would have a group of six loops, corresponding to the hypervariable loops in IgG, at one end of the pseudo V-region and an oligosaccharide at the other. The PTRs would originate at the same end of the aggregate at the oligosaccharide.

Circular dichroism (c.d.) spectra of LP (Fig. 1) did not, however, show the characteristic positive ellipticity of β-sheet-containing proteins which can be seen at 240 nm. In

![Fig. 1. C.d. spectrum of LP](image-url)

LP (approximately 100 μg/ml) in 0.4 M guanidine hydrochloride, 50 mM-Tris/HCl, pH 7.5, was analysed with and without HA (0.5 mg/ml). The HA was added to the LP in 4 M guanidine and the mixture dialysed to a final concentration of 0.4 M guanidine.

Abbreviations used: PG, proteoglycan; HA, hyaluronic acid; HABR, HA-binding region; LP, link protein; PTR, proteoglycan tandem repeat; GAG, glycosaminoglycan; vWF, von Willebrand factor.
fact, the c.d. spectreata looked more like α-helix with a negative ellipticity from 240 to 213 nm, at which point the ellipticity started to move back up before the signal was lost at 210 nm. While the spectrum has not been deconvoluted, if the assumption that the first loop is mostly β-sheet holds true, then it can be inferred that the PTRs are mostly helical and therefore may be dominating the spectrum. There was no detectable effect of adding HA. The pattern of homology of PTRs with the lymphocyte homing receptor, the Hermes antigen [15], tends to reinforce the possibility of an α-helix. Strongly conserved residues are found with three less strongly conserved residues between them, a pattern very characteristic of amphipathic helices.

There are two geometric requirements for binding the G1-LP dimer to HA, (1) Each monomer must have a binding site one decasaccharide long [16], (2) The two monomer binding sites must be juxtaposed axially with each other to accommodate two adjacent decasaccharides of a linear molecule (a total length of 95 Å). However, as a segment of HA longer than the sum of the two HA-binding sites (G1 and LP) is protected from digestion with chondroitinase ABC [17], the same may be as much as 90 Å apart to give an aggregate binding site of 100–200 Å.

The globular structure produced by the postulated interaction of the pseudo V-regions of G1 and LP would, by analogy with IgG, be approximately 40 Å × 40 Å × 48 Å and could not account for the binding site not been described. If HA-binding site must therefore derive from the PTRs, in contrast to what we have suggested previously [18]. Small angle neutron scattering for G1 [19] indicated that it has an elliptical shape and therefore at least one of the PTRs would extend away from the pseudo V-regions. There is experimental evidence that the former is the case. It is postulated that proteolysis of LP bound to HA, for the involvement of PTRs in the interaction with HA [14]. A model, taking into account the locations of the N-linked oligosaccharides, the size of the HA-binding sites and the likely nature of the interaction between LP and G1, is shown in Fig. 2. This is a very simple model and ignores the probable interaction between the PTRs and the N-terminal domain which would be expected in a globular protein.

Recent evidence indicates that G2 (two PTR loops) does not bind to HA, G1 or LP [20]. This agrees with what is observed by electron microscopy [8]. If G2 is capable of folding in the same way as the PTRs in G1 or LP, then the implication is that the PTRs require the rest of G1 (or LP) to fold up correctly. Alternatively, the postulated placement of N-linked oligosaccharides in G2 [2] which is different from that in G1 and LP, may disrupt the structure to an extent which is incompatible with binding to HA.

There is no explanation for the tendency of LP to form very high-mass aggregates when not bound to HA or PG. Synthetic peptides, designed to mimic the N-terminal loop of a PTR, have not assisted our analysis of structure and function, beyond indicating that these domains are rather intractable. While the structure of the N-terminal domain which would be expected in a globular protein.

Small cartilage PGs

These PGs are part of a family of ubiquitous molecules which have, so far, been found in two forms, PG-S1 and PG-S2. Both PGs have GAG attached near the N-terminal IgG-V-region-like domains. N-linked oligosaccharides attached to these are as a spacer separating the PTR domains from each other and enabling these to interact with HA through two similar, but non-contiguous, stretches of the polyelectronic. This simplistic model does not, however, take into account the directionality of the HA. If LP and G1 are bound to HA in the same way, then they should be oriented in the same direction.

Leucine-rich repeats are found in variety of other molecules. The exact function of the repeats is unknown. However, most of these molecules also have analogues for the other structures found in the small PGs, specifically the disulfide-bonded loops. It seems reasonable to suppose that all these molecules fold up in a similar way. It has been suggested that these repeats form a β-sheet arrangement with the hydrophobic residues in internal positions and the external residues defining the specificity of its interactions [22]. In a comparison of PG-S1 and PG-S2, there are clear differences. With our current knowledge, these differences are not very informative: the repeats may form a hydrophobic core in the PG and may not be very accessible to the solvent.

There are two small, disulfide-bond-containing loops in the small PGs, which are also found in other molecules. Apart from the position of the cysteines in the N-terminal loop and, by implication, the disulfide bonds, there are no strikingly conserved features, though these structures are very conserved between the PGs. They may provide much of the non-GAG-related functionality of the molecule.

It is good evidence that PG-S2 limits the size of collagen fibrils [23]. However, the similarity of the PGs to platelet GP-1b suggests another function. It has been found that
the von Willebrand factor (vWF)-mediated platelet adhesion to endothelium is not entirely due to collagen. Given that GP-1b binds to vWF, it is possible to propose an additional mode of platelet adhesion (Fig. 3). In this model, vWF is bound to collagen and/or small PG while GP-1b binds to the other half of the vWF dimer via a similar interaction.

It should, in the next few years, be possible to test the validity of these models and define some of the functions of the PGs at the submolecular level. Eventually, tertiary structure will be defined and PGs can be included among the ranks of well-characterized molecules.

Supported by N.I.H. grant AR 35322 and a grant from The Shriners Hospital for Crippled Children.


Received 17 October 1989

Assembly of cartilage proteoglycan with hyaluronate and structure of the central filament in proteoglycan aggregate

MATTHIAS MÖRGEILIN, MATS PAULSSON and JÜRGEN ENGEL

Abteilung Biophysikalische Chemie, Biozentrum, CH-4056 Basel, Switzerland

Proteoglycans are a family of large complex macromolecules abundant in the extracellular matrix of all connective tissues, but also present at cell surfaces. Each proteoglycan consists of a protein core, which is substituted with the characteristic glycosaminoglycan chains and often also with shorter oligosaccharide structures [for reviews, see [1, 2]]. The major proteoglycan in hyaline cartilage is a high-M₉ aggregating species (M, 1 x 10⁹–4 x 10¹⁰) with a large extended core protein to which typically about 100 chondroitin sulphate chains, 30 keratan sulphate chains and about 60 short oligosaccharides are attached. The best understood interaction of this proteoglycan is that with hyaluronate, which is an unbranched polysaccharide present in cartilage. Aggregates are formed by the non-covalent interaction between a hyaluronate-binding region of the protein core and a decasaccharide segment of the hyaluronate chain [3, 4]. The interaction is stabilized by the additional binding of link protein which has affinity for both the binding region of the proteoglycan and the hyaluronate to which it is attached, such that a ternary complex is formed between these three components. Many proteoglycans may bind to each hyaluronate chain and form aggregates with a length of as much as 4000 nm and a diameter of 500–600 nm [for reviews, see [1, 2]].

Pioneering electron microscopic work by the Klein fold technical [5] revealed binding of proteoglycan monomers to the hyaluronate strands [6] and distribution of side-chain substituents along the protein core [7]. Protein domains within the core of cartilage proteoglycan were first resolved by rotary shadowing of proteoglycans sprayed on to mica [8]. Three globular domains (G1, G2 and G3) and two extended regions (E1 and E2) can be distinguished [9, 10; Fig. 1]. G1 is located at the N-terminus and is separated from