

36**Rapid #: -616020****IP: 128.175.82.31**

-616020



128.175.82.31

CALL #: **EJOURNALS**
LOCATION: **ZCU :: Butler :: lweb**
TYPE: Journal
USER JOURNAL TITLE: CELL RESEARCH
OCLC JOURNAL TITLE: Cell research [electronic resource].
ZCU CATALOG TITLE: Cell research [electronic resource].
ARTICLE TITLE: Structure and function of aggrecan
ARTICLE AUTHOR: Kiani, C
VOLUME: 12
ISSUE: 1
YEAR: 2002
PAGES: 19-32
ISSN: 1001-0602
OCLC #:
CROSS REFERENCE ID:
VERIFIED:

BORROWER: **DLM :: Main Library**
PATRON: **Polenova, Tatyana**
PATRON ID: tpolenov
PATRON ADDRESS: 245 Brown Lab, University of Delaware, Newark, DE 19716
PATRON PHONE: (302) 831-1968
PATRON FAX:
PATRON E-MAIL: tpolenov@udel.edu
PATRON DEPT: Chemistry & Biochemistry
PATRON STATUS: fac
PATRON NOTES:

System Date/Time: 9/19/2005 7:14:24 AM MST

REVIEW

Structure and function of aggrecan

CHRIS KIANI¹, LIWEN CHEN¹, YAO JIONG WU¹, ALBERT J YEE², BURTON B YANG^{1,*}

Sunnybrook and Women's College Health Sciences Centre and ¹Department of Laboratory Medicine and Pathobiology, ²Department of Surgery, Faculty of Medicine, University of Toronto, Canada

ABSTRACT

Aggrecan is the major proteoglycan in the articular cartilage. This molecule is important in the proper functioning of articular cartilage because it provides a hydrated gel structure (via its interaction with hyaluronan and link protein) that endows the cartilage with load-bearing properties. It is also crucial in chondroskeletal morphogenesis during development. Aggrecan is a multimodular molecule expressed by chondrocytes. Its core protein is composed of three globular domains (G1, G2, and G3) and a large extended region (CS) between G2 and G3 for glycosaminoglycan chain attachment. G1 comprises the amino terminus of the core protein. This domain has the same structural motif as link protein. Functionally, the G1 domain interacts with hyaluronan acid and link protein, forming stable ternary complexes in the extracellular matrix. G2 is homologous to the tandem repeats of G1 and of link protein and is involved in product processing. G3 makes up the carboxyl terminus of the core protein. It enhances glycosaminoglycan modification and product secretion. Aggrecan plays an important role in mediating chondrocyte-chondrocyte and chondrocyte-matrix interactions through its ability to bind hyaluronan.

Key words: *Proteoglycan, chondroitin sulfate, glycosaminoglycan, G1 domain, G3 domain.*

INTRODUCTION

Degenerative joint disease is a leading source of morbidity resulting in significant social and economic impact. One to 5% of the population under the age of 45 and 15-85% of older individuals suffer from some form of degenerative joint disease, mainly osteoarthritis. Osteoarthritis is characterized by the

slow progressive deterioration of articular cartilage[1, 2]. Current therapeutic regimens address mainly pain but not degeneration. A better understanding of the distinct micro-environment of articular cartilage and the complex interactions that exist between cell and surrounding extracellular matrix (ECM) is evolving so that strategies can be directed towards altering the natural history of degenerative joint disease.

The gliding surfaces of synovial joints are covered by articular cartilage. Articular cartilage is composed of hyaline cartilage, which provides a thin, smooth, stiff and wear-resistant layer that provides a low friction weight bearing joint surface that allows our joints to move smoothly and without pain. The destruction of articular cartilage that occurs with degenerative joint diseases that leads to joint dys-

*Correspondence to: Burton B. Yang, Research Building, Sunnybrook and Women's College Health Sciences Centre, 2075 Bayview Avenue, Toronto M4N 3M5 Canada. Tel: (416) 480-5874; Fax: (416) 480-5737; E-mail: Burton.Yang@swchsc.on.ca

The abbreviations used are: ECM, extracellular matrix; CS, chondroitin sulfate (chain attachment sequence); KS, keratan sulfate (chain attachment sequence); GAG; glycosaminoglycan; HA, hyaluronan; G₁, the globular domain in the N-terminus of aggrecan; G₂, second globular domain of aggrecan; G₃, the globular domain in the carboxyl terminus of aggrecan or selectin-like domain; IGD, inter-globular domain; PTR, proteoglycan tandem repeat; EGF, epidermal growth factor-like motif; CRD, carbohydrate recognition domain; CBP, complement binding protein.

function and pain.

Cartilage is composed of the cells named chondrocytes and the ECM produced by these cells. The biochemical properties of cartilage and the physical function of joints are critically dependent on the integrity of the matrix. The ECM molecules in cartilage include proteoglycans, hyaluronan (also called hyaluronic acid or HA), type II collagen, glycoproteins and various mixtures of elastic fibers. Proteoglycans are a family of glycoconjugates with a central core protein to which one or more glycosaminoglycan (GAG) side chains are covalently linked post-translationally[3]. In addition, most of the proteoglycans exist as aggregates[4] formed by the non-covalent association of proteoglycan with HA and link protein[5-7]. Among the cartilage proteoglycans, the most crucial to the proper functioning of articular cartilage is aggrecan, one of the large aggregating chondroitin sulfate proteoglycans. Aggrecan is a multimodular molecule and here, we will review the structures and functions of its modules.

Aggrecan and its environment

Cartilage contains up to 10% proteoglycan consisting of mainly the large aggregating chondroitin sulfate proteoglycan aggrecan. The aggrecan family includes other important members such as versican, also named PG-M, neurocan, brevican and the cell surface HA receptor CD44[8-11]. They are modular proteoglycans containing combinations of structural motifs, such as epidermal growth factor (EGF)-like domains, carbohydrate recognition domains (CRD), complement binding protein (CBP)-like domains, immunoglobulin folds and proteoglycan tandem repeats[12]. At high concentrations, proteoglycans create a large osmotic swelling pressure and draw water into the tissue (Fig 1). This occurs because all of the negatively charged anionic groups on the GAG chains of aggrecan carry with them mobile counter ions such as Na^+ . This creates a large difference in ion concentration between the cartilage and surrounding tissue and an imbalance amongst the freely diffusible anions and cations. Water is drawn into cartilage because of this osmotic imbalance and because aggrecan is too large and immobile to redistribute itself. The addition of water causes aggrecan-rich matrix network to swell and expand. This water-swollen matrix is critical to

the biomechanical properties of cartilage. Another feature of the composite collagen/aggrecan organization is also important. Not only is aggrecan greatly restricted in its ability to move within the matrix, but also the collagen/aggrecan network is stiff and resistant to deformation. Aggrecan also offers great resistance to any fluid flow and redistribution of water. Thus, cartilage is referred to as a visco-elastic tissue in that it behaves like the stiff elastic polymer resistant to sudden impact loading, yet shows some slow inelastic deformation with sustained loads[13].

Aggrecan structure and functions

In cartilage, aggrecan is found in huge multimolecular aggregates, comprised of numerous monomers non-covalently bound to HA. The small glycoprotein link protein, which is homologous to the N-terminus of aggrecan, helps to stabilize aggregate formation (Fig 2). The GAG side chains contribute to the formation of the large mass aggregate. Almost 90% of aggrecan mass is comprised of substituted GAG chains which are mostly chondroitin sulfate chains, but also include keratan sulfate chains with N- and O-linked oligosaccharides.

Aggrecan has three globular domains (G1, G2 and G3) and three extended domains (IGD, KS and CS) as shown in Fig 3. Following the signal peptide is the N-terminal G1 domain. An inter-globular domain (IGD) connects G1 and G2 domains. Situated between the G2 domain and G3 domain is a large sequence modified by KS CS side chains. Each aggrecan contains ~100 chondroitin sulfate chains, which are typically ~20 kDa each. There are fewer keratan sulfate chains (up to 60) and they are usually of smaller size (5-15 kDa). Aggrecan also contains a variable number of O- and N-linked oligosaccharides. The O-linked oligosaccharides have a linkage to protein similar to that of keratan sulfate. During biosynthesis, some O-linked oligosaccharides are extended and sulfated to form keratan sulfate chains.

All three of the globular domains of the aggrecan protein contain sequences that are highly conserved amongst aggrecan from different species, while the extended domains are less conserved. The mouse aggrecan gene spans at least 61 kb and contains 18 exons. Exon 1 encodes 5'-untranslated sequence and exon 2 contains a translation start codon. The coding sequence is 6545 bp for a 2132-amino-acid pro-

tein with calculated molecular weight of 259,131 Da, including an 18-amino acid signal peptide. There is a strong correlation between structural domains and exons. Notably, the chondroitin sulfate domain consisting of 1161 amino acids is encoded by a single exon of 3.6 kb.

Aggrecan G1 domain is in fact similar to link protein and other members of this proteoglycan family, in terms of its structural domains and subdomains. There are also stretches of sequences similar to the promoter region of both the type II collagen and link protein genes. These sequences may be important for cartilage gene expression[14]. The amino acid sequences of human aggrecan and rat aggrecan are about 75% identical. The human sequence contains

two regions of highly conserved repeats not found in rat aggrecan: 11 repeats of a hexameric sequence in the keratan sulfate attachment domain, E-E-P-(S/F)-P-S; and a 19-amino acid sequence reiterated 19 times, in the CS-1 portion of the serine-glycine-containing region. There are at least three forms of aggrecan transcripts, generated by alternative exon usage, with or without EGF-like motifs[15].

As a structural proteoglycan, aggrecan appears to be important in mediating chondrocyte-chondrocyte and chondrocyte-matrix interactions. Co-expression of a mini-aggrecan and link protein stabilizes cell-substratum interaction[16]. Addition of exogenous gene products into fibroblast cell lines and chondrocyte culture had the same effect as ex-

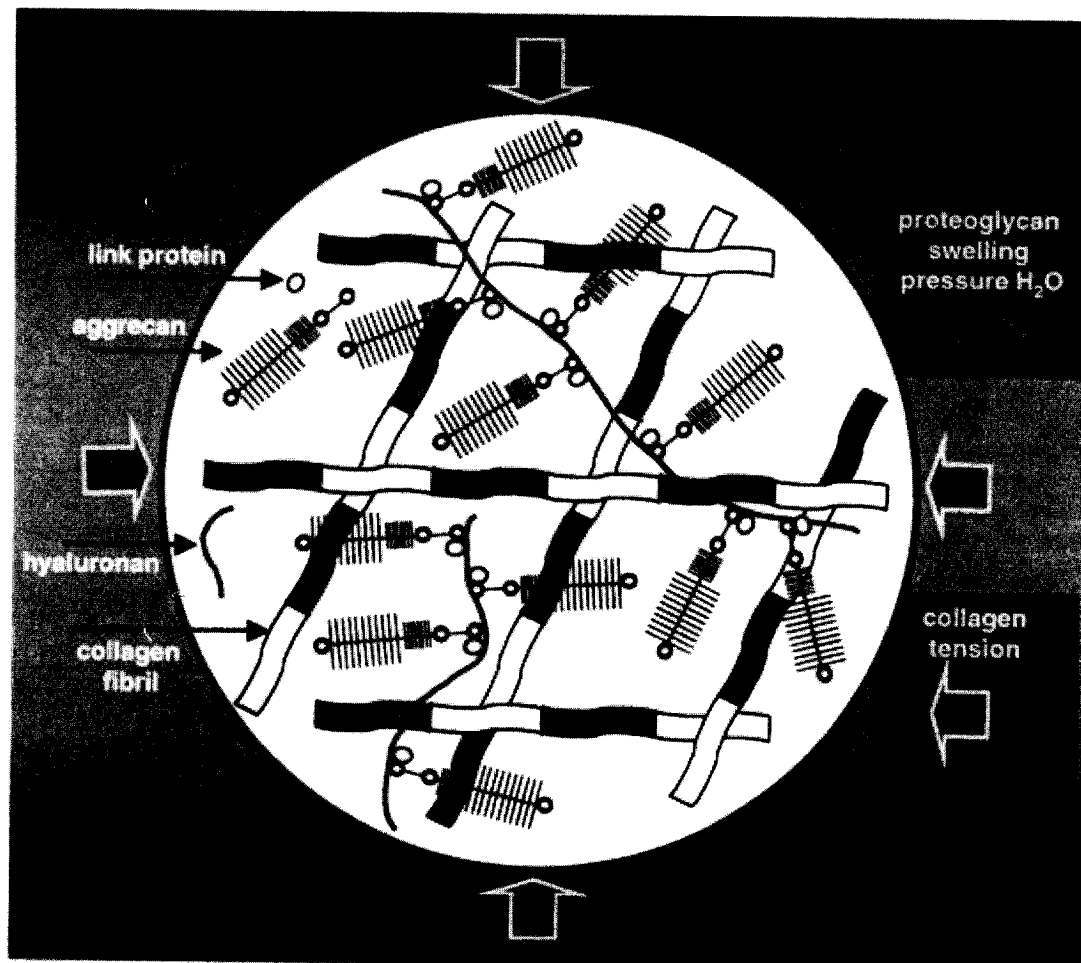


Fig 1. The combined properties of collagens and aggrecan in articular cartilage

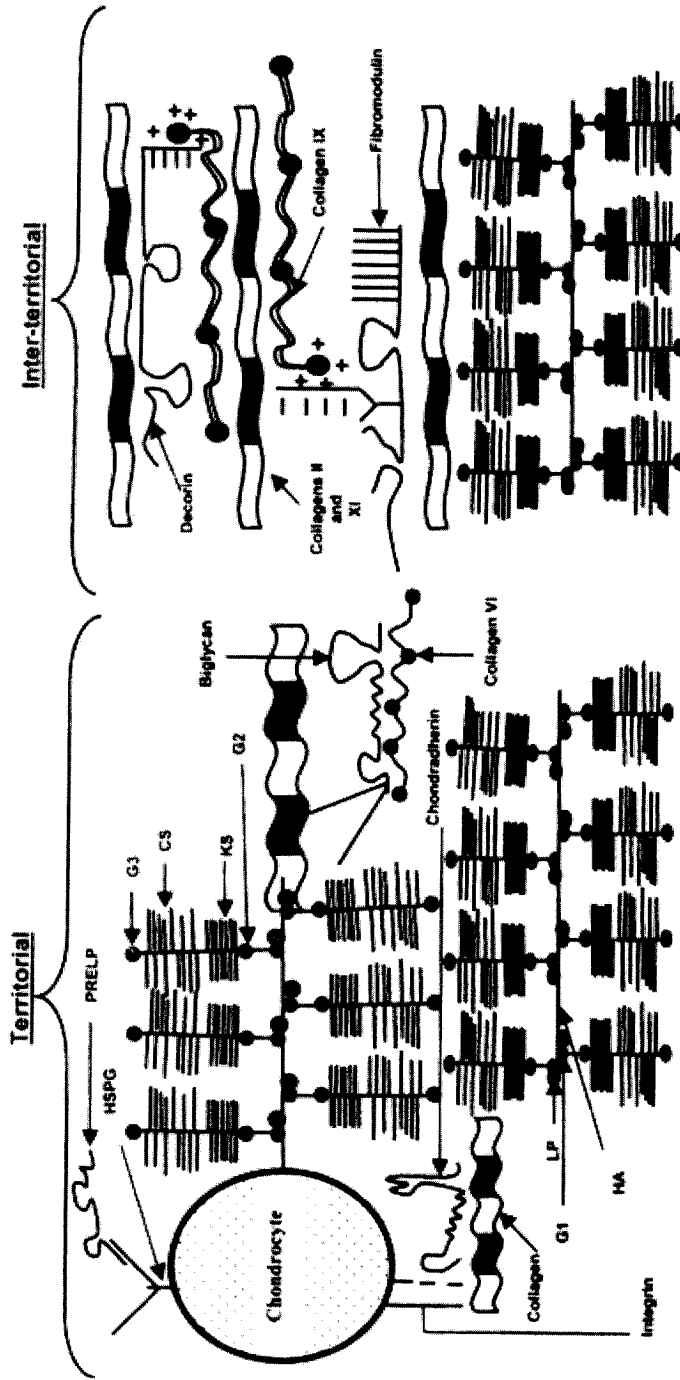


Fig 2. Illustration of cartilage matrix composition
 The different molecular organization of territorial (close to cells) and inter-territorial (distant from cells) matrices is depicted. Major constituents are proteoglycans and the collagen-based network, where collagen fibers contain numerous bound molecules that play roles in regulating assembly and maintaining function of the network. Interactions at the surface of the chondrocytes are likely to play roles in providing cells with information on the matrix properties.

pression of the genes. The addition of exogenous HA to the growth medium, or treatment of cells with HA, also decreased cell adhesion, indicating that HA also plays a role in the cell-substratum adhesion. The presence of aggrecan seems to increase the amount of link protein on the cell surface. Chondrocytes expressing high concentrations of aggrecan and link protein were maintained within a matrix network and were able to survive in suspended culture. Imbalances in aggrecan or link protein concentrations, or degradation of HA, disrupted the network and caused the chondrocytes to aggregate or

adhere to the plates[16, 17]. Digestion of HA also induces chondrocyte aggregation. The effect of HA on cell aggregation and cell attachment seems to predominate over aggrecan, type II collagen and link protein[17]. Another member of the large aggregating chondroitin sulfate proteoglycans, versican, also plays a role in mediating cell adhesion and migration[18]. As versican is expressed in a variety of tissues, its functions are also diverse. Versican has been shown to be able to promote the growth of NIH3T3 fibroblasts and chicken chondrocytes[19]. It also inhibits mesenchymal chondrogenesis of chicken limb

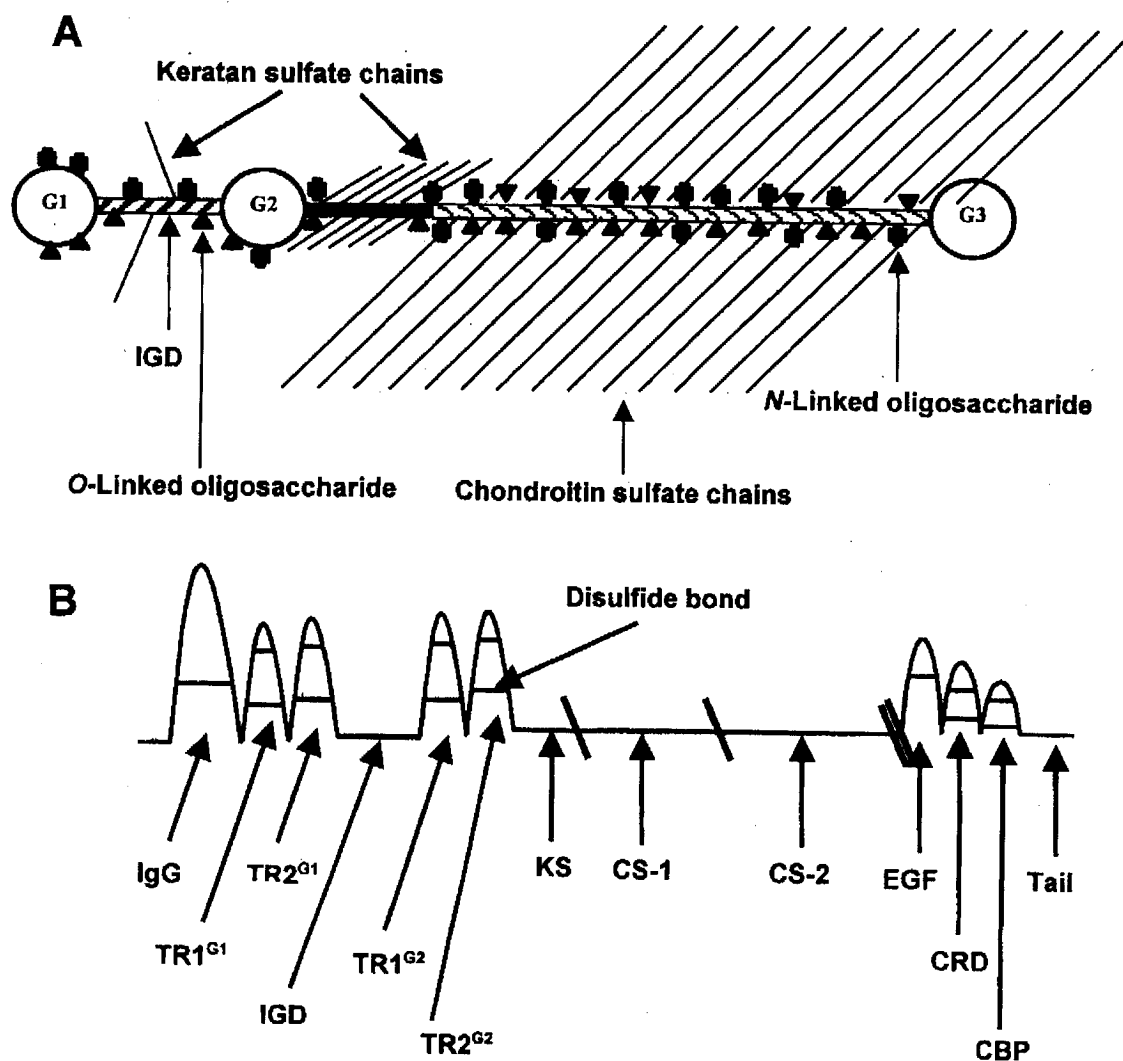


Fig 3. Aggrecan structure

(A) Globular protein and attached GAG chain structure. (B) Protein domain structure. Folded modules: IgG, immunoglobulin fold; TR, tandem repeats; EGF, epidermal growth factor-like module; CRD, carbohydrate recognition domain; CBP, complement binding protein-like module. Extended domains: IGD, interglobular domain; KS, keratan sulfate attachment domain; CS-1 and CS-2, chondroitin sulfate attachment domains.

buds[20]. Expression of a mini-versican construct alters chondrocyte morphology[21]. However, it is not clear whether aggrecan has similar effects on cell activities.

The G1 Domain

G1 domain is encoded by exons 3, 4, 5 and 6. This domain is comprised of three modules: an immunoglobulin fold, and two copies of a HA-binding motif, or link module, also referred to as the proteoglycan tandem repeat (PTR) (Fig 3). The immunoglobulin fold motif of aggrecan G1 corresponds to the hypervariable region of the immunoglobulins and is about 100 amino acids long. This region is predicted to fold into two β sheets in a sandwich conformation (parallel orientation) stabilized by a conserved disulfide bond (Fig 3). This motif is encoded by a single exon. Almost all of the proteins that contain immunoglobulin-related structures are cell surface molecules involved in cell recognition, cell adhesion or immune reaction. Their roles in cell recognition are corroborated by the fact that many of these molecules interact with other members of the immunoglobulin superfamily, and indeed the interaction between aggrecan G1 domain and link protein is mediated via the immunoglobulin fold[22]. The PTR is present in duplicate in all members of the HA-binding family of proteoglycans including versican, neurocan and brevican. However, it is present as a single copy in the cell surface HA-binding receptor, CD44, and in TSG, a secreted matrix protein whose synthesis is induced by inflammatory cytokines. The PTR of the G1 domain are comprised of two cysteine-rich motifs which are able to form disulfide bonds (four cysteine residues or two disulfide bonds per tandem). The tandem repeats, and particularly the disulfide bonds, are involved in aggrecan's interaction with HA, forming the large complexes (Fig 3). These tandem repeats do not bind to HA under reducing conditions[23].

In addition to its interaction with HA, the G1 domain also interacts with link protein, and thus forms the "glue" holding together the ternary complex in the matrix. The ternary complex plays a critical role in maintaining a stable matrix network in cartilage. Due to its binding activities, the G1 domain mediates interactions between chondrocyte and the matrix network. Overexpression of the G1

domain has been shown to reduce the adhesion of chicken chondrocytes[24]. The reduced cell adhesion enhances cell apoptosis[25]. The G1 domain of versican also reduces NIH3T3 fibroblast adhesion[26]. Recently, it was observed that the G1 domain of aggrecan regulates product processing[27]. The same effect was also observed for versican G1 domain[28]. In both cases, the G1 domains inhibit GAG modification and product secretion. These functions appear to play a key role as a checkpoint for proteoglycan quality control.

The inter-globular domain (IGD)

The short extended region separating the G1 and G2 domains of aggrecan is known as the interglobular domain (IGD). This domain has a rod-shaped structure in which contains proteolytic cleavage sites susceptible to a variety of proteinases such as matrix metalloproteinases (MMPs), serine proteinases such as plasmin and leukocyte elastase, and acid proteinases such as cathepsin B (cysteine protease) (Fig 4) [1, 29 30]. The IGD domain is encoded by exon 7 of aggrecan. It is unique to the aggrecan molecule since other members of the aggrecan family lack this region. X-ray diffraction, NMR (nuclear magnetic resonance imaging) and rotary shadowing electron microscopic analyses have determined that the IGD is of constant length (25 nm) and relatively stiff and inflexible[31, 32]. This apparent stiffness of the IGD may be due to the density of keratan sulfate substitution[33, 34], which seems to be greater in the IGD region than the adjoining globular domain (G2). The functional, rather than structural, properties of IGD make it interesting. The IGD domain is the site of proteolytic attack on aggrecan during pathological cartilage degradation. Cleavage of aggrecan molecules in the IGD region near its G1 domain results in rapid loss of the whole GAG attachment region. As such, IGD appears to be involved in the physiological turnover of aggrecan.

Key observations-that fragments harboring a G1 domain bearing a VDIPEN-341 neopeptide (rather than a NITEGE-373 neopeptide) are present in human synovial fluids[35] and appear rapidly in rabbit knee joint fluid after an intra-articular injection of MMP3[36]-suggest that there is further processing of the HA-G1 complex after the IGD has been cleaved off by aggrecanase and has diffused out of the matrix.

Recent work on the extent of the involvement of aggrecanase and MMPs in the cleavage of aggrecan at the IGD domain indicates that aggrecanase cleavage of aggrecan by far exceeds cleavage by MMPs under patho-physiological conditions[37].

Two major cleavage sites have been identified in the IGD domain: one occurs between residues Asn 341 and Phe 342 and the other is located between residues Glu 373 and Ala 374. The first major cleavage site generates a G1-containing fragment with the C-terminus neoepitope VDIPEN-341 and a larger GAG-bearing fragment with the N-terminus neoepitope 342-FFGVGGE[34, 38-40]. The results of many studies support the contention that MMPs are involved in the breakdown of aggrecan molecules *in vivo*. First, cartilage and synovial fluid from patients with osteoarthritis and other inflammatory arthritis types contain enhanced levels of MMPs. Second, retinoic acid and pro-inflammatory cytokines (IL-1 and TNF- α) are known to upregulate the expression of several MMPs, and this promotes aggrecan loss from cartilage explants[41-43]. This loss of aggrecan can be prevented by adding specific inhibitors of MMPs to the conditioned medium[44, 45]. Third, the N- and C-terminal neoepitopes generated by MMPs are present in both cartilage and synovial fluid. Finally, it has been demonstrated that the concentration of the VDIPEN-341 neoepitope in normal human articular cartilage increases up to 25 years of age and then reaches an apparent steady state (representing 15-20% of the G1-containing molecules residing within the matrix)[46]. This suggests that MMP-generated G1 fragments might account, at least in part, for the growth- and maturation-related increase in aggrecan domains in the cartilage.

Studies on the aggrecan fragments extracted from articular cartilage or synovial fluid suggested the presence of a proteinase that cleaves a specific site in IGD, which was named "aggrecanase"[47]. Aggrecanase-1 is a member of the ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) protein family that cleaves aggrecan at the Glu373-Ala374 bond[48]. This cleavage produces a large GAG-rich aggrecan fragment with the N-terminus neoepitope 374-ARGSVI, and a G1 fragment with the C-terminus neoepitope NITEGE-373 (Fig 4). Aggrecanase-mediated

cleavage occurs both *in vivo* and *in vitro*: the N-terminus neoepitope 374-ARGSVI has been found in the medium of cartilage explant cultures[49] as well as in high density aggrecan fragments recovered from human synovial fluids[47, 50, 51]. It is believed that MMPs and aggrecanase are both involved in the turnover of aggrecan molecules in normal and diseased cartilage. This is supported by the finding that the MMP-generated G1 fragment terminating in VDIPEN-341 and aggrecanase-generated G1 fragments terminating in NITEGE-373 are both detected in cartilage from joints with osteoarthritis and rheumatoid arthritis[40]. The generation and/or turnover of these specific aggrecan fragments is not necessarily co-ordinated, since both the NITEGE-373 and VDIPEN-341 neoepitopes can be non-coincident within a single joint[38]. Turnover of aggrecan in cultured rat chondrosarcoma cells and primary bovine chondrocytes can be mediated exclusively by aggrecanase[52].

It has been reported that aggrecanase also cleaves off aggrecan at multiple sites in the CS-attachment domain[2, 53, 54] generating GAG-rich fragments, which are lost relatively rapidly from the cartilage matrix. These fragments diffuse into the synovial fluid, where they can be quantified by chemical assays of their sulfated GAG chains[55] or by immunoassays capable of measuring specific protein or carbohydrate epitopes[53, 54]. Recently, it has been shown that cathepsin B (a cysteine protease with both endopeptidase and carboxypeptidase activities) is able to cleave aggrecan. Cleavage occurs at Asn341-Phe342, to yield the neoepitopes VDIPEN and FFGVGG[30].

The G2 domain

This domain is encoded by exons 8, 9 and 10 of the aggrecan gene and is the second globular domain from the N-terminus. It consists of two proteoglycan tandem repeats, which are very similar to those of the G1 domain. Aggrecan is the only molecule of the aggregating chondroitin sulfate proteoglycan family to contain this domain. It is noteworthy that aggrecan is also the only member of this family that is heavily glycosylated by \sim 100 chondroitin sulfate and up to 60 keratan sulfate chains. There may be a direct or indirect role for the G2 domain in GAG chain attachment.

The G2 domain shows an approximately 67% amino acid sequence similarity to that of the tandem repeats in the G1 domain. Whereas the tandem repeats have been shown to be the major HA-binding element in link protein and the G1 domain, the G2 domain, either purified from cartilage after proteolytic digestion[56] or the recombinant soluble G2 protein[57], shows no HA-binding function whatsoever.

The G2 domain has also been shown to contain keratan sulfate side chains[58]. While it is not clear where the KS is attached, it may also significantly interfere with the ability of the G2 domain to interact with HA. The locations of cysteine residues essential for maintaining the three-dimensional structure of the G2 domain, and the locations of asparagine residues required for attaching carbohydrate in an N-linked fashion, are well conserved. The G2 domain is not able to bind to link protein, other aggregating proteoglycan monomers, collagen or other major components of the dissociate extract of cartilage[56]. Nevertheless, we have recently shown that the G2 domain inhibits product secretion[27]. The product of a recombinant G2 construct is not secreted, even when linked to a signal peptide. Secretion of other recombinant products containing the G2 domain is either inhibited or retarded. This is so far the only function assigned to the G2 domain, and may be associated with product quality control in order to produce a mature functional aggrecan molecule.

The KS domain

Following the G2 domain is the keratan sulfate (KS) domain, which is encoded by exon 11. There are about 30 KS chains attached to the mature aggrecan molecule. The amino acid sequence of this domain varies among different species. The potential consensus sequence for the attachment of KS chains in the human is E-(E/K)P-F-P-S or E-E-P-(S/F)-P-S[15]. Mice and rats lack these sequences and also lack keratan sulfate chains in their articular aggrecan. In both bovine[59] and human aggrecan[15], the amino acid sequence of this region, revealed by cDNA sequence analysis, contains a striking repeating sequence. The sequence is a hexamer, which contains a serine residue. It has been suggested that proline-serine and proline-threonine repeats in this region are good candidates for substitution by KS

chains. The human aggrecan cDNA encodes an additional 66-residue-sequence consisting of a highly conserved hexamer peptide motif repeated 11 times consecutively. This structure is consistent with the fact that human aggrecan is more heavily glycosylated in this region than is rat aggrecan. Both bovine and rat aggrecan also contain this sequence, but contain fewer repeats[15].

The KS region is not the only site of KS attachment in the protein core. KS is also distributed elsewhere, primarily in the CS region. Some of the KS chains are O-linked to threonine[60]. It is possible that the repeating Thr-Thr-Ala-Pro sequence in the N-terminal end of the CS region represents this attachment site. KS chains from bovine aggrecan have a molecular weight of approximately of 10 kDa[61]. The KS chains vary in structure depending on tissue source. KS from load-bearing tissues (articular cartilage and intervertebral disks) contain 1-3 fucose residues and 2-6 N-acetyl neuraminic acid residues, which are absent from non-load-bearing tissues (tracheal and nasal cartilage)[62]. The KS-containing peptide isolated from chick aggrecan does not exhibit significant similarity to the human or bovine KS domain and may represent a further variation in this domain[63]. It has been reported that the concentration of keratan sulfate in human cartilage (especially menisci) increases with age as does the concentration of the 6-sulfated disaccharide of chondroitin sulfate does[64].

The function of this domain is not very clear. It may be involved in tissue distribution of aggrecan. The keratan sulfate chains may contribute to tissue development. Recently, we observed that the KS domain plays a role in product processing[27]. Secretion of some recombinant products was enhanced in the presence of a KS domain. Last, but not least, the keratan sulfate chains in this domain hold water within aggrecan molecule, and thus enhance the load-bearing capability of aggrecan in cartilage.

The CS domain

The chondroitin sulfate (CS) domain is the largest domain of aggrecan and is decorated by approximately 100 chains of chondroitin sulfate. This domain is encoded by a single exon, exon 12, with a size of ~3.5 kb. The CS domain consists of approximately 120 Ser-Gly dipeptide repeats. The dipep-