J. Biochem. 124, 687-693 (1998)

Roles of Aggrecan, a Large Chondroitin Sulfate Proteoglycan, in Cartilage Structure and Function

Hideto Watanabe,*,[1] Yoshihiko Yamada,* and Koji Kimata†

**Craniofacial Developmental Biology and Regeneration Branch, National Institute of Dental Research, National Institutes of Health, Bethesda MD 20892; and †Institute for Molecular Science of Medicine, Aichi Medical University, Aichi 480-1195*

Received for publication, June 17, 1998

Aggrecan, a large aggregating proteoglycan, is one of the major structural components of cartilage. Its core protein contains three glubular domains and two glycosaminoglycan-attachment domains. These domains play various roles to maintain cartilage structure and function. An N-terminal globular domain binds hyaluronan and link protein to form huge aggregates. The chondroitin sulfate (CS) chains attach to the CS domain and provide a hydrated, viscous gel that absorbs compressive load. Two autosomal recessive chondrodysplasias, cartilage matrix deficiency (*cmd*) in mice and nanomelia in chicken are both caused by aggrecan gene mutations. *Cmd* homozygotes die shortly after birth, while the heterozygotes are born normal. However, *cmd* heterozygotes develop late onset of spinal disorder, which suggests aggrecan as a candidate gene predisposing individuals to spinal problems. Nanomelia is a useful model to elucidate intracellular trafficking of proteoglycans. Further studies on aggrecan will lead to prophylaxis and treatment of joint destructive diseases such as osteoarthrosis and to elucidation of cartilage development, which is essential for skeletal formation.

Key words: cartilage matrix deficiency, G1 domain, G3 domain, link protein, nanomeria.

Cartilage is a highly specialized tissue with an important function: bearing compressive load in joints. Since cartilage has limited repair capacity, its destruction results in major problems in joint diseases such as rheumatoid arthritis and osteoarthrosis. Cartilage also serves as the precursor for most bone tissues during development. Therefore, the study of cartilage formation is important for elucidation of the mechanisms of skeletal development and joint diseases. Cartilage has at least three characteristics. First, it has no blood vessels or innervation. Without blood vessels, this tissue cannot be a primary region of inflammation. Secondly, it consists of a single cell type, chondrocyte, which undergoes various stages in differentiation. Because of a single cell type, the concept of parenchyma or intima does not hold true to cartilage tissue. Thirdly, it consists of unique extracellular matrix molecules. For example, type II, IX, X, and XI collagens are unique to cartilage. They form collagen fibrils and fibers that endow cartilage with tensile strength. Cartilage also contains a variety of proteoglycans such as aggrecan, decorin, biglycan, PG-Lb, and fibromodulin. Among them, aggrecan and PG-Lb are rather unique to cartilage. Aggrecan is one of the major components of cartilage, and binds to hyaluronan (HA) and link protein to form huge aggregates. These aggregates lead to a hydrated gel-like structure of cartilage and resistibility to compression and deformation in joints. In this review, we describe structures and function of aggrecan, and its genetic disorders.

[1] To whom correspondence should be addressed. Tel: +1-301-402-0512, Fax: +1-301-402-0897, E-mail: watanabe@yoda.nidr.nih.gov

Abbreviations: *cmd*, cartilage matrix deficiency; CRP, complement regulatory proteins; CS, chondroitin sulfate; EGF, epidermal growth factor; Gn, guanidine; HA, hyaluronan; IGD, interglobular domain; KS, keratan sulfate; MMP, matrix metalloproteinase; OA, osteoarthrosis; PTR, proteoglycan tandem repeat; RT, reverse transcriptase; VNTR, variable number of tandem repeat.

Molecular and gene structure

Proteoglycan is defined as a molecule composed of a protein (core protein) and glycosaminoglycan chains that covalently attached to the core protein (*1*). The core protein of aggrecan has a molecular mass of ~230 kDa. With many glycosaminoglycan chains (up to ~130 chains) attached to the core protein, the total molecular mass can reach ~2,200 kDa. Complete coding sequences for aggrecan have been determined for the rat (*2*), mouse (*3*), human (*4*), and chicken (*5*). The core protein consists of three globular domains, G1, G2, and G3, and two glycosaminoglycan-attachment domains, KS and CS domains, located between the G2 and G3 domains. Between G1 and G2, there is another rod-shaped domain called the interglobular domain (IGD). Similar domain structures are also found in other proteoglycans such as PG-M/versican (*6, 7*), neurocan (*8*), and brevican (*9*) (Fig. 1). Therefore, these extracellular matrix proteoglycans, sharing similar domain structures, form a gene family (*10*). The coding sequence of the three globular domains shows high homology among species, while that of the KS and CS attachment domains is

relatively diverse.

Aggrecan core protein is encoded by a single gene that is mapped to chromosome 7 in mouse (*11*) and chromosome 15 in humans (*12*). The genomic structure of rat, chicken, mouse, and human aggrecan has been reported, showing a high similarity among species (*13-16*). The aggrecan gene contains 18 exons: Exon 1 is untranslated and exon 2 contains a translation initiation codon. The organization of exons is strongly correlated to the specific domains, as shown in Fig. 1. Comparison of the exon organization among species reveals that the aggrecan gene family is evolved distinctly from the link protein gene, which encodes a globular structure similar to the G1 domain.

G1 and G2 domains

Interactions of aggrecan with HA and link protein are essential for the aggregate formation. Biochemical studies revealed important properties of the aggregates. Cesium chloride density-gradient ultracentrifugation in 0.4 M guanidine hydrochloride (Gn-HCl) precipitates the aggregates, and in 4 M Gn-HCl, separates the three molecules (*17*). Mild tryptic digestion of the aggregates produces a fragment that still retains HA-binding function and contains the G1 and G2 domains, called hyaluronan-binding region (HABR). Extensive tryptic digestion revealed that aggrecan interacts with HA and link protein *via* the G1 domain. The G1 domain contains three looped subdomains: A, B and B′. Both B and B′ loops form a similar disulfide bonded double loop structure called proteoglycan tandem repeat (PTR) (*18*) (Fig. 2). Several studies have suggested that a PTR loop acts as a functional site of the interaction with HA. Like aggrecan, other HA-binding extracellular matrix proteoglycans and link protein contain the G1 domain, composed of the same three loops. The HA receptor CD44 (*19*) and the arthritis-associated protein tumor necrosis factor-stimulated gene-6 (TSG-6) (*20*), which have a single PTR loop in a molecule, also have HA-binding function, suggesting that a single PTR module may be enough for interaction with HA. Since HABR does not bind to HA under reducing condition, the ternary structure of PTR formed by two disulfide bonds is clearly essential (*21*). Previous studies using site-specific antibodies and oligopeptides showed that tip portions of the B or B loops of link protein are sites for HA interaction (*18*), which later proved to be inconclusive by NMR studies (*22*). Recent NMR studies of a PTR segment of TSG-6 show structural resemblance to the C-type lectin domain (*23*).

Previous studies with tryptic digestion of cartilage aggregates showed that a region of B and/or B′ loop(s) interacts with HA with a minimal length of hyaluronan decasaccharide (HA10). On the other hand, a minimal length of hexasaccharides (HA6) was required for CD44 interaction. Studies on the interaction with HA by equilibrium dialysis in dilute solution, together with those by ultracentrifugal methods, showed that a dissociation constant of the interaction with a hyaluronate of M_r 670,000 was about 2×10^{-7} M (*24*). Recombinant aggrecan fragments bound to HA at KD $\sim2\times10^{-7}$ M (*25*), which is consistent with the previous report using native aggrecan. We have shown using recombinant aggrecan domains that the minimal segment of aggrecan for HA-binding is the B-B′ segment of the G1 domain and that a single PTR has no HA-binding activity (*25*). Carbohydrate chains of the G1 domain is also important for HA-binding as is the case with CD44 (*26*). Although CD44 contains only one PTR, it may

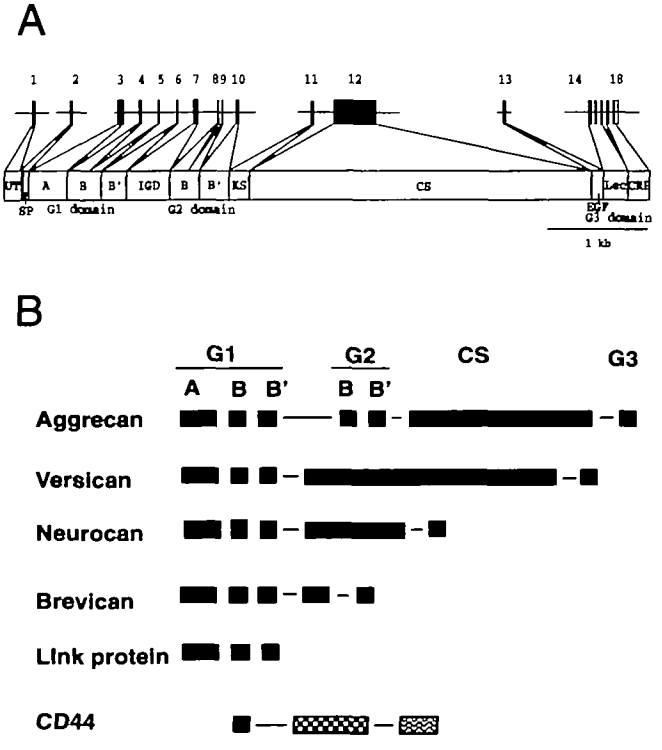

Fig. 1. **Genomic and domain structures of aggrecan and related proteoglycans.** A: Correlation of exon organization and domain structure of aggrecan. UT, untranslated region; SP, splicing signal; IGD, interglobular domain; KS, keratan sulfate domain; CS, chondroitin sulfate domain; EGF, epidermal growth factor-like subdomain; Lec, lectin-like subdomain; CRP, complement regulatory protein-like subdomain. B: Comparison of domain structure of aggrecan family proteoglycans. Only aggrecan contains the G2 domain.

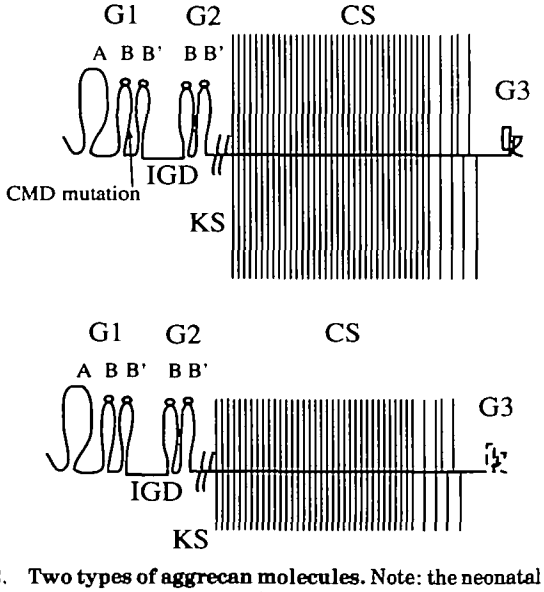

Fig. 2. **Two types of aggrecan molecules.** Note: the neonatal type with longer CS chains (upper) and adult type with shorter CS chains that lack the G3 domain (lower).

form a dimer on the cell surface to interact with HA. An N-terminal loop, the A loop of the G1 domain forms an immunoglobulin fold (Ig-fold) (*27*). This loop region, since it was protected from proteolytic digestion of a conjugate of link protein and aggrecan, is supposed to interact with link protein (*27, 28*). Like immunoglobulin, the regions that correspond to hypervariable regions, may interact with each other (*29*). The A loop also enhances the interaction of the B-B′ segment with HA, and its ternary structure is critical for the enhancement (*25*).

The G2 domain is unique to aggrecan. This domain shows approximately 67% amino acid sequence identity to that of the B-B′ subdomain of the G1 domain. However, the G2 domain either purified from cartilage after proteolytic digestion (*30*) or the recombinant G2 domain (*25*) has no HA-binding function. The locations of cysteine residues essential for its ternary structure and asparagine residue attaching carbohydrate in an *N*-linked manner are well conserved among link proteins, the G1 domain of aggrecan, PG-M/versican, and neurocan. The critical difference in HA-binding activity from the G1 domain has not yet been found. The function of the G2 domain also remains to be elucidated.

Keratan sulfate (KS) domain

The keratan sulfate (KS) domain is located at the C-terminus of the G2 domain, and is encoded by exon 11. This domain varies in amino acid sequences among species. The possible consensus sequence for attachment of KS in human is E-(E,K)-P-F-P-S or E-E-P-(S,F)-P-S (*4, 31*). These sequences are not found in mice and rats. Rodents do not have keratan sulfate chains in their aggrecan. This may be due to the absence of these sequences.

Chondroitin sulfate (CS) domain

The chondroitin sulfate (CS) domain is the largest domain of aggrecan, attached by approximately a hundred chains of chondroitin sulfate. This domain is encoded by a single large exon, exon 12, with a size of ~3.5 kb. The CS domain consists of approximately 120 serine-glycine (S-G) repeats. The possible recognition sequence for the attachment of CS chains have been proposed to be S-G-X-G (*32*) or (D,E)-X-S-G (*33*). In addition to the primary sequence, molecular chaperon surveillance mechanisms and localization of enzymes for post-translational modifications may also be necessary for the recognition. Recent recombinant study has revealed that a truncated CS-G3 segment of aggrecan is substituted with CS. On the other hand, that of G1-CS is not (*34*). The S-G pairs occur in two distinct pattern of repeating sequences designated the CS-1 and CS-2 repeat regions. Human cDNA contains a sub region of the CS-1 repeat that was remarkably conserved in repeats of 19 amino acids. Recently, the presence of variable number of tandem repeats (VNTR) in human aggrecan CS domain was identified, ranging from 13 to 33 (*35*). This variation may affect the CS content. Each repeat contains two possible attachment sites for CS, so that the range of VNTR may vary as many as 40 CS chains per core protein monomer. Assuming full substitutions, there would be a range of 172 to 132 CS chains between these two alleles, respectively, representing a 30% variation. Since CS chains contribute to a hydrated gel-like structure unique to cartilage, such VNTR may influence cartilage function and may explain the fragility of some diseases such as juvenile onset of osteoarthrosis.

A highly sensitive fluorotag HPLC method to quantitate products in chondroitinase digests of human aggrecan isolated from cartilages of individuals at various ages indicated age-related changes in the structure of CS chains (*36*). A decrease in the average chain size occurred between fetal and early postnatal ages. At skeletal maturity, the average aggrecan CS chain size decreases from 20 kDa to approximately 8 kDa, and the ratio of 6- to 4-sulfation on interior disaccharide is increased from ~0.77 to ~23 in the adult, as shown in Fig. 2. These age-related changes in CS chain of aggrecan molecule may affect not only its hydrated size but also the interactions with other cartilage molecules, which are important for cartilage functions.

G3 domain

The G3 domain, located at the C-terminus, consists of three modules: epidermal growth factor (EGF)-like module, C-type lectin-like module, and complement regulatory proteins (CRP)-module (*2*). The structure of the G3 domain is also found in other extracellular matrix molecules and proteoglycans such as PG-M/versican, neurocan, and brevican. PG-M/versican and neurocan have two EGF-like modules. Similarly, aggrecan has two EGF-like modules, EGF-1 (*37*) and EGF-2 (*38*). However, their expression varies among species (*39*). For example, both modules are found in human, but bovine EGF1 expresses very little. In mouse, rat, and dog, EGF1 is a part of intron and is not translated.

The lectin module of aggrecan binds to fucose and galactose (*40*). The recombinant C-terminal region (EGF-like module, C-type lectin module, and CRP-module) of PG-M/versican has the activity to bind to heparin and heparan sulfate as well as to these simple carbohydrates (*41*). The lectin module of PG-M/versican binds to tenascin-R by carbohydrate-protein interactions (*42*). Comparison of interactions with tenascin-R among a family of the G3 domain-containing proteoglycans revealed not only a carbohydrate-protein interaction but also a distinct protein-protein interaction (*43*). Aggrecan G3 domain may interact with tenascin in the cartilage. However, *in situ* hybridization data indicates that mRNAs of both genes are expressed in a mutually exclusive manner (*44*).

With aging, the population of aggrecan without the G3 domain increases compared to that with full length (*36, 45*). This is probably due to proteolytic cleavage in the cartilage matrix rather than alternative splicing or other intracellular modification.

Interglobular domain (IGD)

The interglobular domain (IGD) between the G1 and G2 domains has a rod-shape, and contains proteolytic cleavage sites susceptible to a variety of proteinases. For example, matrix metalloproteinases (MMPs) such as MMP-1, 2, 3, and 8, serine proteinases such as plasmin and leukocyte elastase, and acid proteinases like cathepsin B cleave their specific sites in IGD (*46*). Since cleavage of IGD can be related to the turnover of aggrecan, identification of the enzyme(s), which are involved in degradation of aggrecan *in vivo*, is important. Studies on the aggrecan fragments extracted from joint cartilage or synovial fluid suggested the presence of a proteinase that cleaves a specific site in

IGD, which was named "aggrecanase" (*47*). In the articular cartilage of normal adults, at least eleven aggrecan fragments have been found with the molecular weight ranging from 300,000 to 43,000 and degradation seems to start at the C-terminus (*48*). Aggrecanase has been implicated in joint destructive diseases such as osteoarthrosis (OA) (*47, 49*). The pathogenesis of OA is presently focused on proteolytic degradation of cartilaginous molecules. Based on the size variation of aggrecan fragments extracted from articular cartilage and synovial fluid of OA patients, aggrecanase seems to cleave not only the specific site in the IGD but also four similar sites in the CS domain, which are well conserved among species. As well as articular cartilage, meniscus and ligaments may also serve as source of aggrecan fragments (*50*). Interestingly, the G1 domain of aggrecan is more abundant in these tissues.

Transcriptional regulation

Since aggrecan expression is quite specific to cartilage, studies of the transcription of aggrecan gene may lead to elucidation of the mechanisms of chondrocyte differentiation. In rat (*13*), mice (*15*), and chicken (*51*), the transcriptional start sites of the aggrecan gene were determined. S1 nuclease protection and/or primer extension revealed two, four, and three transcription start sites in rats, mice, and chicken, respectively. The aggrecan promoter in any species has no TATA sequence. The mouse promoter contains two glucocorticoid receptor-binding sequences (TGTTCT/C), one GGGCGG sequence (Sp-1 site), and several homologous direct repeat sequences. In addition, a region between −54 to −111 shows sequence homology to a sequence of the rat type II collagen promoter (−103 to −132). This sequence is highly conserved in both the rat and mouse type II collagen genes (*52*) and is important for type II collagen gene promoter activity. Another stretch of a sequence from −287 to −259 shows homology to a sequence of the rat link protein promoter (−82 to −60). These sequences may play a role in cartilage-specific gene expression. In rats, a 922-bp fragment containing 640-bp 5′-flanking DNA and 282-bp exon 1 shows high promoter activity in transfected chondrocytes than in fibroblasts. In chicken, a 1.8 kb genomic fragment from the 5′ end of the aggrecan gene is able to drive expression in a tissue-specific manner. A 1.8 kb segment of chicken aggrecan promoter and exon 1 contains four CACACA motif. The mouse aggrecan exon 1 contains two E-box motifs. Overexpression of scleraxis, a helix-loop-helix transcription factor, enhances aggrecan gene expression of osteoblastic ROS cells through their sites (*53*).

Animal models of aggrecan gene defects

Two autosomal recessive chondrodysplasias, *cartilage matrix deficiency* (*cmd*) in mice and nanomelia in chicken were shown to be caused by mutations in the aggrecan gene. The *cartilage matrix deficiency* (*cmd/cmd*) is the first example of the mutation of a proteoglycan gene identified in mammals. The homozygotes (*cmd/cmd*) are characterized by dwarfism, short snout, and cleft palate (*54*). Heterozygous mice (*cmd/+*) are born normal, while homozygous mice (*cmd/cmd*) die just after birth due to respiratory failure. The cartilage of homozygous mice appears as tightly packed chondrocytes with little matrix, unlike the extensive matrix seen in normal mice. The *cmd* aggrecan gene has a single 7 bp deletion in exon 5 which encodes the B loop of the G1 domain. This deletion causes a frameshift resulting in the appearance of a termination codon in exon 6 (*11*). The potentially truncated 32 kDa polypeptide created by this mutation contains an incomplete PTR structure incapable of HA-binding. Biochemical and immunological studies have demonstrated the absence of aggrecan in the cartilage matrix of *cmd* mice, although normal levels of link protein and type II collagen were detected (*55*). *Cmd* heterozygotes (*56*) are apparently normal at birth. Since heterozygous *cmd* mice have only one normal allele of the aggrecan gene, some metabolic differences in aggrecan of the heterozygotes may occur and create abnormal phenotypic changes after birth. Quantitative reverse transcriptase (RT)-PCR studies revealed reduced levels of aggrecan mRNA in *cmd* heterozygotes and in homozygotes to 81 and 41%, respectively, compared to that in the wild type mice. Those of type II collagen mRNA of both the heterozygote and of the homozygote were similar to that in the wild type mice. The levels of chondroitin sulfate in cartilage from 90-day-old *cmd* heterozygous mice was reduced to 87% of the wild type. Since aggrecan is important in cartilage development, the reduced level of aggrecan is likely to affect the growth of the heterozygotes. Indeed, *cmd* heterozygotes show two abnormal phenotypes; a slight dwarfism and late onset spinal misalignment. Approximately 28 days after birth, the heterozygotes are noticeably smaller than the wild type mice. The dwarfism of the heterozygotes is proportional. The most notable abnormality of the *cmd* heterozygotes is a misalignment of the cervical and thoracic spine which develops about one year after birth (Fig. 3). In mice, the cervical spine is particularly susceptible to gravitational loading because the animals have to support their head. The heterozygotes die after 12-15 months, while the wild type mice live for 2-2.5 years. They develop a marked lordosis of the cervical spine and kyphosis of the thoraco-lumbar spine. Mice with spinal distortions suddenly acquire a spastic gait disturbance and showed decreased movement. They were unable to eat and starved to death within one month following acquisition of the gait disturbance. From histological examinations, the heterozygotes showed herniation of the vertebral disc, deformation of the vertebral bodies and degenerative changes of the cervico-thoracic spine. Compression of the spinal cord by the herniated disc was also observed in these mice, explaining their spastic gait. Disappearance of the apophysis of the vertebral bodies was also observed. In contrast, such specific changes are hardly found in the spine from one-year old wild type mice. Alcian blue and toluidin blue stainings in the heterozygotes show a reduction of glycosaminoglycans. Alcian blue stained tissues surrounding the chondrocytes in the disc cartilage of the heterozygotes, while diffuse staining was observed in the wild type. General features of spinal degeneration are accompanied by pathological changes in both the intervertebral discs and the facet joints. In the *cmd* heterozygotes, pathological changes were found entirely in the intervertebral discs, while the facet joints were apparently normal.

These histological findings indicate that the primary lesion lies in the disc and that degeneration characteristic of reactive bone growth did not occur. The knee joint and other cartilaginous tissues were apparently normal in the het-

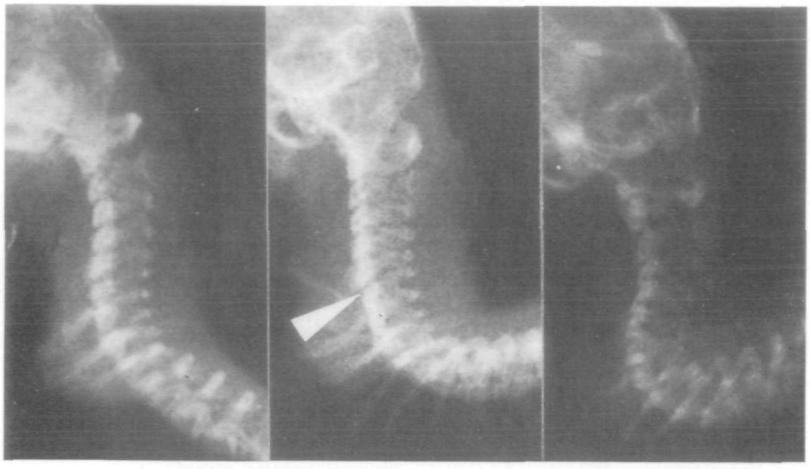

Fig. 3. **Radiographs of 1-year-old wild type and heterozygous *cmd* mice.** (Left) Control wild type mouse. (Center and right) Two different *cmd* heterozygotes. Arrow indicates misalignment at C7-Th1. In the right panel, vertebral bodies show deformation.

erozygotes. The turnover of collagen within the disc is estimated to be very slow (>100 years), while the aggrecan turnover is more rapid, with the half-life of 8-300 days in rabbits (*57*). Because of the relatively rapid turnover of aggrecan, decreased synthesis of aggrecan is likely to cause significant reduction of its deposition in the tissue. A certain level of aggrecan deposition in the disc may be critical for maintenance of disc function. For instance, aggrecan has been shown to play a key role in maintaining the collagen network. Electron microscopy of the cartilage of the *cmd* homozygotes shows abnormal collagen fibrils which display an increase in the diameter, the appearance of periodic banding patterns, and bundling formation (*58*). These results also suggest a role of aggrecan in collagen fibrillogenesis. Similar changes, such as rough fiber distribution with concentric patterns, are found in the disc of the *cmd* heterozygotes likely due to a reduced deposition of aggrecan in the disc. Chondrocytes in the affected cervical spine of the heterozygotes are abnormally packed together, containing degenerative vacuoles in their cytoplasm, and rough fibers in the matrix are organized in concentric circles which surrounded the chondrocytes. Control chondrocytes in the wild type mice are distributed as individual cells in a fine extracellular matrix. It is interesting to note that the pathological changes in the tissues of the *cmd* heterozygotes were found mainly in the specific portions of the spine which are most susceptible to gravitational load.

Cmd has been classified as an autosomal recessive disorder. Autosomal recessive inheritance is defined as inheritance in which a clinical phenotype occurs only when both alleles are defective. However, heterozygotes of some recessive disorders may have subtle differences in phenotype which may be accentuated by environmental factors. The pathologies reported here for *cmd* heterozygote mice suggest that defects in one of alleles of this gene can have clinical manifestations. Although several genetic disorders of collagens, such as osteogenesis imperfecta and chondrodysplasia (*59-61*) have been reported, human genetic diseases caused by a defect of the aggrecan gene have not yet been identified. The *cmd* heterozygous mice show a high incidence of spinal misalignment and movement problems which develop with age, primarily involving spastic paralysis of the hind limbs. This resembles spinal paralysis in humans. It is conceivable that an analogous aggrecan gene defect causes spinal disc herniation or spondylo-myelopathy, which are well known as diseases typical of older humans. The phenotype of the heterozygous *cmd* mice may provide a useful clue for linkage studies of aggrecan gene defects in suspected human patients. Our findings support aggrecan as a candidate gene predisposing individuals to spinal problems. The other aggrecan gene defect is nanomelia in chicken, which bears a phenotype similar to the *cmd* mouse. A point mutation in the aggrecan gene was found at the end of exon 12 encoding CS domain in the aggrecan gene of chick nanomelia (*5*). In nanomelia, a truncated protein is localized in the endoplasmic reticulum (*62*), suggesting that G3 is important for intracellular trafficking.

Conclusions

Aggrecan is one of the major structural macromolecules of cartilage. Studies on the structure and function of aggrecan may lead to prophylaxis and treatment of joint destructive diseases such as osteoarthrosis. Since cartilage serves as a precursor tissue of bone formation, aggrecan plays a critical role in skeletal formation. Studies on animal models of aggrecan gene defects such as *cmd* and nanomelia will provide us with clues to identification of human genetic diseases caused by aggrecan gene defects.

REFERENCES

1. Hascall, V.C., Heinegard, D.K., and Wight, T.N. (1991) Proteoglycans: metabolism and pathology in *Cell Biology of Extracellular Matrix* (Hay, E.D., ed.) pp. 149-175, Plenum Press, New York
2. Doege, K., Sasaki, M., Horigan, E., Hassell, J.R., and Yamada, Y. (1987) Complete primary structure of the rat cartilage proteoglycan core protein deduced from cDNA clones [published erratum appears in *J. Biol. Chem.* 1988 Jul 15; 263(20): 10040], *J. Biol. Chem.* **262**, 17757-17767
3. Walcz, E., Deak, F., Erhardt, P., Coulter, S.N., Fulop, C., Horvath, P., Doege, K.J., and Glant, T.T. (1994) Complete coding sequence, deduced primary structure, chromosomal localization, and structural analysis of murine aggrecan. *Genomics* **22**, 364-371
4. Doege, K.J., Sasaki, M., Kimura, T., and Yamada, Y. (1991) Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. Human-specific repeats, and additional alternatively spliced forms. *J. Biol. Chem.* **266**, 894-902
5. Li, H., Schwartz, N.B., and Vertel, B.M. (1993) cDNA cloning of

chick cartilage chondroitin sulfate (aggrecan) core protein and identification of a stop codon in the aggrecan gene associated with the chondrodystrophy, nanomelia. *J. Biol. Chem.* **268**, 23504-23511
6. Shinomura, T., Nishida, Y., Ito, K., and Kimata, K. (1993) cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan expressed during chondrogenesis in chick limb buds. Alternative spliced multiforms of PG-M and their relationships to versican. *J. Biol. Chem.* **268**, 14461-14469
7. Zimmermann, D.R. and Ruoslahti, E. (1989) Multiple domains of the large fibroblast proteoglycan, versican. *EMBO J.* **8**, 2975-2981
8. Rauch, U., Karthikeyan, L., Maurel, P., Margolis, R.U., and Margolis, R.K. (1992) Cloning and primary structure of neurocan, a developmentally regulated, aggregating chondroitin sulfate proteoglycan of brain. *J. Biol. Chem.* **267**, 19536-19547
9. Yamada, H., Watanabe, K., Shimonaka, M., and Yamaguchi, Y. (1994) Molecular cloning of brevican, a novel brain proteoglycan of the aggrecan/versican family. *J. Biol. Chem.* **269**, 10119-10126
10. Margolis, R.U. and Margolis, R.K. (1994) Aggrecan-versican-neurocan family proteoglycans. *Methods Enzymol.* **245**, 105-126
11. Watanabe, H., Kimata, K., Line, S., Strong, D., Gao, L.Y., Kozak, C.A., and Yamada, Y. (1994) Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene. *Nat. Genet.* **7**, 154-157
12. Just, W., Klett, C., Vetter, U., and Vogel, W. (1993) Assignment of the human aggrecan gene AGC1 to 15q25→q26.2 by in situ hybridization. *Hum. Genet.* **92**, 516-518
13. Doege, K.J., Garrison, K., Coulter, S.N., and Yamada, Y. (1994) The structure of the rat aggrecan gene and preliminary characterization of its promoter. *J. Biol. Chem.* **269**, 29232-29240
14. Li, H. and Schwartz, N.B. (1995) Gene structure of chick cartilage chondroitin sulfate proteoglycan (aggrecan) core protein. *J. Mol. Evol.* **41**, 878-885
15. Watanabe, H., Gao, L., Sugiyama, S., Doege, K., Kimata, K., and Yamada, Y. (1995) Mouse aggrecan, a large cartilage proteoglycan: protein sequence, gene structure and promoter sequence. *Biochem. J.* **308**, 433-440
16. Valhmu, W.B., Palmer, G.D., Rivers, P.A., Ebara, S., Cheng, J.F., Fischer, S., and Ratcliffe, A. (1995) Structure of the human aggrecan gene: exon-intron organization and association with the protein domains. *Biochem. J.* **309**, 535-542
17. Hascall, V.C. and Kimura, J.H. (1982) Proteoglycans: isolation and characterization in *Methods in Enzymology* (Cunningham, L.W. and Frederikson, D.W., eds.) Vol. 82, pp. 769-800, Academic Press, New York
18. Goetinck, P.F., Stirpe, N.S., Tsonis, P.A., and Carlone, D. (1987) The tandemly repeated sequences of cartilage link protein contain the sites for interaction with hyaluronic acid. *J. Cell Biol.* **105**, 2403-2408
19. Stamenkovic, I., Amiot, M., Pesando, J.M., and Seed, B. (1989) A lymphocyte molecule implicated in lymph node homing is a member of the cartilage link protein family. *Cell* **56**, 1057-1062
20. Lee, T.H., Wisniewski, H.G., and Vilcek, J. (1992) A novel secretory tumor necrosis factor-inducible protein (TSG-6) is a member of the family of hyaluronate binding proteins, closely related to the adhesion receptor CD44. *J. Cell Biol.* **116**, 545-557
21. Yu, Q. and Toole, B.R. (1995) Biotinylated hyaluronan as a probe for detection of binding proteins in cells and tissues. *Biotechniques* **19**, 122-128
22. Horita, D.A., Hajduk, P.J., Goetinck, P.F., and Lerner, L.E. (1994) NMR studies of peptides derived from the putative binding regions of cartilage proteins. No evidence for binding to hyaluronan. *J. Biol. Chem.* **269**, 1699-1704
23. Kohda, D., Morton, C.J., Parkar, A.A., Hatanaka, H., Inagaki, F.M., Campbell, I.D., and Day, A.J. (1996) Solution structure of the link module: a hyaluronan-binding domain involved in extracellular matrix stability and cell migration. *Cell* **86**, 767-775
24. Nieduszynski, I.A., Sheehan, J.K., Phelps, C.F., Hardingham, T.E., and Muir, H. (1980) Equilibrium-binding studies of pig laryngeal cartilage proteoglycans with hyaluronate oligosaccharide fractions. *Biochem. J.* **185**, 107-114
25. Watanabe, H., Cheung, S.C., Itano, N., Kimata, K., and Yamada, Y. (1997) Identification of hyaluronan-binding domains of aggrecan. *J. Biol. Chem.* **272**, 28057-28065
26. Bartolazzi, A., Nocks, A., Aruffo, A., Spring, F., and Stamenkovic, I. (1996) Glycosylation of CD44 is implicated in CD44-mediated cell adhesion to hyaluronan. *J. Cell. Biol.* **132**, 1199-1208
27. Perkins, S.J., Nealis, A.S., Dudhia, J., and Hardingham, T.E. (1989) Immunoglobulin fold and tandem repeat structures in proteoglycan N-terminal domains and link protein. *J. Mol. Biol.* **206**, 737-753
28. Grover, J. and Roughley, P.J. (1994) The expression of functional link protein in a baculovirus system: analysis of mutants lacking the A, B and B′ domains. *Biochem. J.* **300**, 317-324
29. Lesk, A.M. and Chothia, C. (1982) Evolution of proteins formed by beta-sheets. II. The core of the immunoglobulin domains. *J. Mol. Biol.* **160**, 325-342
30. Fosang, A.J. and Hardingham, T.E. (1989) Isolation of the N-terminal globular protein domains from cartilage proteoglycans. Identification of G2 domain and its lack of interaction with hyaluronate and link protein. *Biochem. J.* **261**, 801-809
31. Antonsson, P., Heinegård, D., and Oldberg, A. (1989) The keratan sulfate-enriched region of bovine cartilage proteoglycan consists of a consecutively repeated hexapeptide motif. *J. Biol. Chem.* **264**, 16170-16173
32. Bourdon, M.A., Krusius, T., Campbell, S., Schwartz, N.B., and Ruoslahti, E. (1987) Identification and synthesis of a recognition signal for the attachment of glycosaminoglycans to proteins. *Proc. Natl. Acad. Sci. USA* **84**, 3194-3198
33. Krueger, R.C., Jr., Fields, T.A., Hildreth, J. IV, and Schwartz, N.B. (1990) Chick cartilage chondroitin sulfate proteoglycan core protein. I. Generation and characterization of peptides and specificity for glycosaminoglycan attachment. *J. Biol. Chem.* **265**, 12075-12087
34. Luo, W., Kuwada, T.S., Chandrasekaran, L., Zheng, J., and Tanzer, M.L. (1996) Divergent secretory behavior of the opposite ends of aggrecan. *J. Biol. Chem.* **271**, 16447-16450
35. Doege, K.J., Coulter, S.N., Meek, L.M., Maslen, K., and Wood, J.G. (1997) A human-specific polymorphism in the coding region of the aggrecan gene. Variable number of tandem repeats produce a range of core protein sizes in the general population. *J. Biol. Chem.* **272**, 13974-13979
36. Plaas, A.H., Wong-Palms, S., Roughley, P.J., Midura, R.J., and Hascall, V.C. (1997) Chemical and immunological assay of the nonreducing terminal residues of chondroitin sulfate from human aggrecan. *J. Biol. Chem.* **272**, 20603-20610
37. Baldwin, C.T., Reginato, A.M., and Prockop, D.J. (1989) A new epidermal growth factor-like domain in the human core protein for the large cartilage-specific proteoglycan. Evidence for alternative splicing of the domain. *J. Biol. Chem.* **264**, 15747-15750
38. Fulop, C., Walcz, E., Valyon, M., and Glant, T.T. (1993) Expression of alternatively spliced epidermal growth factor-like domains in aggrecans of different species. Evidence for a novel module. *J. Biol. Chem.* **268**, 17377-17383
39. Fulop, C., Cs-Szabo, G., and Glant, T.T. (1996) Species-specific alternative splicing of the epidermal growth factor-like domain 1 of cartilage aggrecan. *Biochem. J.* **319**, 935-940
40. Halberg, D.F., Proulx, G., Doege, K., Yamada, Y., and Drickamer, K. (1988) A segment of the cartilage proteoglycan core protein has lectin-like activity. *J. Biol. Chem.* **263**, 9486-9490
41. Ujita, M., Shinomura, T., Ito, K., Kitagawa, Y., and Kimata, K. (1994) Expression and binding activity of the carboxyl-terminal portion of the core protein of PG-M, a large chondroitin sulfate proteoglycan. *J. Biol. Chem.* **269**, 27603-27609
42. Aspberg, A., Binkert, C., and Ruoslahti, E. (1995) The versican C-type lectin domain recognizes the adhesion protein tenascin-R. *Proc. Natl. Acad. Sci. USA* **92**, 10590-10594
43. Aspberg, A., Miura, R., Bourdoulous, S., Shimonaka, M., Heinegard, D., Schachner, M., Ruoslahti, E., and Yamaguchi, Y. (1997) The C-type lectin domains of lecticans, a family of

aggregating chondroitin sulfate proteoglycans, bind tenascin-R by protein-protein interactions independent of carbohydrate moiety. *Proc. Natl. Acad. Sci. USA* **94**, 10116-10121
44. Glumoff, V., Savontaus, M., Vehanen, J., and Vuorio, E. (1994) Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues by Northern and in situ hybridization using species specific cDNA probes. *Biochim. Biophys. Acta* **1219**, 613-622
45. Dudhia, J., Davidson, C.M., Wells, T.M., Vynios, D.H., Hardingham, T.E., and Bayliss, M.T. (1996) Age-related changes in the content of the C-terminal region of aggrecan in human articular cartilage. *Biochem. J.* **313**, 933-940
46. Hardingham, T.E. and Fosang, A.J. (1995) The structure of aggrecan and its turnover in cartilage. *J. Rheumatol. Suppl.* **43**, 86-90
47. Sandy, J.D., Flannery, C.R., Neame, P.J., and Lohmander, L.S. (1992) The structure of aggrecan fragments in human synovial fluid. Evidence for the involvement in osteoarthritis of a novel proteinase which cleaves the Glu 373-Ala 374 bond of the interglobular domain. *J. Clin. Invest.* **89**, 1512-1516
48. Sandy, J.D., Plaas, A.H., and Koob, T.J. (1995) Pathways of aggrecan processing in joint tissues. Implications for disease mechanism and monitoring. *Acta Orthop. Scand. Suppl.* **266**, 26-32
49. Lark, M.W., Bayne, E.K., and Lohmander, L.S. (1995) Aggrecan degradation in osteoarthritis and rheumatoid arthritis. *Acta Orthop. Scand. Suppl.* **266**, 92-97
50. Lark, M.W., Bayne, E.K., Flanagan, J., Harper, C.F., Hoerrner, L.A., Hutchinson, N.I., Singer, I.I., Donatelli, S.A., Weidner, J.R., Williams, H.R., Mumford, R.A., and Lohmander, L.S. (1997) Aggrecan degradation in human cartilage. Evidence for both matrix metalloproteinase and aggrecanase activity in normal, osteoarthritic, and rheumatoid joints. *J. Clin. Invest.* **100**, 93-106
51. Pirok, E.W. III, Li, H., Mensch, J.R., Jr., Henry, J., and Schwartz, N.B. (1997) Structural and functional analysis of the chick chondroitin sulfate proteoglycan (aggrecan) promoter and enhancer region. *J. Biol. Chem.* **272**, 11566-11574
52. Metsaranta, M., Toman, D., de Crombrugghe, B., and Vuorio, E. (1991) Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and alternative splicing. *J. Biol. Chem.* **266**, 16862-16869
53. Liu, Y., Watanabe, H., Nifuji, A., Yamada, Y., Olson, E.N., and Noda, M. (1997) Overexpression of a single helix-loop-helix-type transcription factor, scleraxis, enhances aggrecan gene expression in osteoblastic osteosarcoma ROS17/2.8 cells. *J. Biol. Chem.* **272**, 29880-29885
54. Rittenhouse, E., Dunn, L.C., Cookingham, J., Calo, C., Spiegelman, M., Dooher, G.B., and Bennett, D. (1978) Cartilage matrix deficiency (cmd), a new autosomal recessive lethal mutation in the mouse. *J. Embryol. Exp. Morphol.* **43**, 71-84
55. Kimata, K., Barrach, H.J., Brown, K.S., and Pennypacker, J.P. (1981) Absence of proteoglycan core protein in cartilage from the cmd/cmd (cartilage matrix deficiency) mouse. *J. Biol. Chem.* **256**, 6961-6968
56. Watanabe, H., Nakata, K., Kimata, K., Nakanishi, I., and Yamada, Y. (1997) Dwarfism and age-associated spinal degeneration of heterozygote cmd mice defective in aggrecan. *Proc. Natl. Acad. Sci. USA* **94**, 6943-6947
57. Mankin, H.J. and Lippiello, L. (1969) The turnover of adult rabbit articular cartilage. *J. Bone Joint Surg. Am.* **51**, 1591-1600
58. Kobayakawa, M., Iwata, H., Brown, K.S., and Kimata, K. (1985) Abnormal collagen fibrillogenesis in epiphyseal cartilage of CMD (cartilage matrix deficiency) mouse. *Coll. Relat. Res.* **5**, 137-147
59. Olsen, B.R. (1995) The roles of collagen genes in skeletal development and morphogenesis. *Experientia* **51**, 194-195
60. Sakai, L.Y., Burgeson, R.E., Olsen, B.R., Rowe, D.W., and Gordon, S.L. (1996) Current knowledge and research directions in heritable disorders of connective tissue. *Matrix Biol.* **15**, 211-229
61. Li, Y. and Olsen, B.R. (1997) Murine models of human genetic skeletal disorders. *Matrix Biol.* **16**, 49-52
62. Vertel, B.M., Grier, B.L., Li, H., and Schwartz, N.B. (1994) The chondrodystrophy, nanomelia: biosynthesis and processing of the defective aggrecan precursor. *Biochem. J.* **301**, 211-216