BIOSYNTHESIS AND REGULATION OF EXPRESSION OF PROTEOGLYCANS

Nancy B. Schwartz

Departments of Pediatrics and Biochemistry and Molecular Biology, University of Chicago, 5841 S. Maryland Avenue, MC 5058, Chicago, IL 60637

TABLE OF CONTENTS

1. ABSTRACT

Proteoglycans are a family of complex macromolecules characterized by the presence of one or more glycosaminoglycan chains covalently linked to a polypeptide backbone. Although originally named and categorized on the basis of the glycosaminoglycan substituent, increasingly they are being identified as members of gene families that encode their different core proteins. Proteoglycans are found predominantly in the extracellular matrix (ECM) or associated with the cell surface of most eucaryotic cells where they bind to other matrix- and cell-associated components. Their ability to be so interactive stems in large part from their structural diversity, which arises from variations in polysaccharide type, size and composition as well as core protein primary sequence, domain arrangement, degree of substitution and distribution of polysaccharide chains. Considering the complexity of proteoglycan molecules, often having modular core protein domains and posttranslational modifications that vary with developmental setting, the various steps of synthesis and processing are most likely highly regulated. Furthermore, regulation of proteoglycan expression is even more complex as they frequently are expressed transiently by multiple cell types and in different developmental time frames. Elucidation of cell- and developmental-specific control elements which regulate the expression of these complex macromolecular families are only beginning.

2. INTRODUCTION

In the past several years there has been a significant increase in the molecular characterization of proteoglycans as well as the identification of new family members, new localizations and new biological functions. Because the proteoglycans remain such important components of the extracellular matrix and cell surface milieu, as illustrated by the drastic change in expression during development of several tissue systems and in certain disease processes, a thorough understanding of the mechanisms which control the expression of these fascinating macromolecules remains a fundamental problem in cell biology. In order to fully appreciate the functional roles of proteoglycans as relate to changing patterns of expression in development and disease, detailed knowledge of their biosynthetic processing and structural organization is necessary. Although this has been a daunting task due to their size and the complexity of carbohydrate composition and core protein domains, significant progress has been made. With respect to transcriptional regulation, which is necessary for expression of the proteoglycan genes in multiple differentiating tissue systems and at different times over the developmental life of an organism, elucidation of transcriptional elements responsible for regulation of those genes is required. In these areas, studies on the transcriptional regulation of proteoglycan genes that examine tissue-specific or developmental patterns are only

beginning. Current information on structure, synthesis and regulation of expression is summarized herein, with particular emphasis on the prototypical proteoglycan, aggrecan.

3. OVERVIEW PG STRUCTURE/FUNCTION

3.1. Glycosaminoglycan structure

The proteoglycans encompass a group of complex macromolecules in which one or more polysaccharide chains are convalently linked to a central protein core. The predominant polysaccharides are known as glycosaminoglycans and consist of repeated disaccharides, usually containing a sulfated hexosamine and uronic acid. Proteoglycan structures are diverse in type, size and composition of polysaccharide attached as well as in primary sequence, domain arrangement, degree of substitution and distribution of the polysaccharide chains along the protein core. Furthermore, N- and O-linked glycoprotein-type oligosaccharides or more than one type of glycosaminoglycan chain may be attached to the same core protein, creating hybrid molecules and enhancing the diversity of structure.

Although there are common features among the glycosaminoglycans, six distinct classes are recognized based on fine structure differences due to specificity in composition and sulfation, epimerization or N-acetylation modifications. Four of these are linked to serines of the protein core via a common tetrasaccharide (xylose-galactose-galactose-galactose) and include chondroitin sulfate, dermatan sulfate, heparin and heparan sulfate. Chondroitin sulfate consists of a series of 30-50 repeating disaccharides of N-acetylgalactosamine and glucuronic acid in a beta 1-3 linkage, either unsulfated or sulfated on the 4 or 6 position of the galactosamine residue. Dermatan sulfate differs by containing iduronic acid in place of glucuronic acid residues, which originates by epimerization of the C5 carboxyl group and sulfation at the C2 position of glucuronic acid.

Complexity increases in heparin and heparan sulfate which consist of glucosamine and either glucuronic or iduronic acid in the repeating disaccharide and much more sulfation. Both glycosaminoglycans may have N-acetylglucosamine acetyl groups which are N-sulfated, often within block regions. Heparin has a higher sulfate content than heparan sulfate, with N- or O-sulfation of the hexosamine C6 or C3 positions and O-sulfation of the uronic acid C2 position. In contrast, heparan sulfate has more N-acetyl groups, and fewer N-sulfate or O-sulfate groups overall.

Aberrations in the common glycosaminoglycan structural features of a repeating hexosamine/uronic acid disaccharide is found in the last two glycosaminoglycan types, keratan sulfate and hyaluronate. Keratan sulfate differs from all the other glycosaminoglycans by not having uronic acid, being composed instead of a repeating disaccharide unit of galactose and N-acetylglucosamine in beta1,4 linkage. Keratan sulfate covalent linkages to core proteins are of two types: i) N-linked through N-acetyl-glucosamine to asparagine, which is typical of glycoproteins, and ii) O-linked through N-acetylgalactosamine to serine or threonine. The latter are often found attached to the same core protein as chondroitin sulfate. The keratan sulfates also contain mannose, fucose and sialic acid; sulfate content is variable and may occur on either or both, galactose or hexosamine, in the C6 position. Hyaluronate is also very different from the other five types of glycosaminoglycans in that it is not sulfated or found covalently linked to a protein core. It is structurally similar to the other glycosaminoglycans and consists solely of repeating disaccharides of N-acetylglucosamine and glucuronic acid. Although hyaluronate can be considered to have the least complex structure of all the glycosaminoglycans, the polysaccharide chains can reach molecular weights of $10^5 - 10^7$, a feature important to the biological function of hyaluronate.

3.2. Proteoglycan families

A better understanding of the diversity of proteoglycan structure and function is developing from the recent cloning of cDNAs encoding proteoglycan core proteins. More than forty full-length cDNAs of proteoglycan have been sequenced thus far, allowing the emergence of a system for classifying proteoglycans into gene families that encode the different core proteins.

The concept of modular proteoglycans composed of discrete structural and functional domains, including both carbohydrate attachment and carbohydrate-free regions, evolved from the examination of deduced primary structures as verified by limited protein sequencing of isolated peptides. This is especially true of the aggrecan gene family, which consists of four distinct proteoglycans, aggrecan, versican, neurocan and brevican. Comparison of domain organization has led to a structural model consisting of two N-terminal globular domains (G1 and G2), one of which binds hyaluronate, and a C-terminal multi-functional binding domain (G3) (part of which is lectin-like), separated by a variable length carbohydrate-rich domain. Even though their size and sequence varies, all four members share this general organization, leading to the designation as hyalectins, proteoglycans with hyaluronate- and lectin-interacting domains (1).

Several other families of proteoglycans are known, e.g. the cell-associated proteoglycans comprising the integral membrane syndecan family which have a short C-terminal cytoplasmic domain and a large extracellular domain substituted with heparan and chondroitin sulfate chains. Serglycin is another small proteoglycan containing serine/glycine dipeptide tandem repeats which are sites of glycosaminoglycan attachment. The basement membrane proteoglycan, perlecan, also has a large and modular structure consisting of five major domains with multiple functions and a single N-terminal heparan sulfate attachment domain. By rotary shadowing, the perlecan proteoglycan resembles five/six beads on a string with thin thread of heparan sulfate emanating from one end. The small leucine-rich proteoglycans are

typified by the dermatan/chondroitin sulfate-substituted decorin and biglycan and the keratan sulfate-substituted fibromodulin, lumican and several others. They are all characterized by a central leucine-rich repeat domain. The variation in these distinct gene families, based on modular core protein organization and diversity in glycosaminoglycan type, provides a vast combinatorial potential for functional specificity that has been exploited by nature. (2,3).

3.3. Function

Proteoglycans fulfill a variety of biological functions, i.e. , growth factor concentration, growth modulation, ionic filtration, biological lubrication, matrix organization, cell-adhesion and as structural scaffolds. Spatial immobilization of growth factors and cytokines may be one of the most important functions of proteoglycans. In this role, cell surface heparan sulfate proteoglycans bind growth factors like FGF, serve to protect the growth factor from degradation in the extracellular milieu, sequester a concentrated surface reservoir of growth factor which is released only by degradation of the proteoglycan, or act as co-receptor to alter the conformation of the growth factor, thereby facilitating binding to its receptor and triggering signal transduction pathways.

Certain of the surface-associated proteoglycans exhibit dynamic cellular functions. Thrombomodulin is important in inhibiting thrombin-induced clotting and the inactivation of thrombin by antithrombin III. Perlecan, found predominantly in basement membrane, functions in regulating the permeability of the glomerular basement membrane and as a modulator of FGF-2 signalling in vasculogenesis. The membrane-intercalated syndecans play a role in cell growth and transformation, which can be modulated during development by cytokines and growth factors.

Most importantly, proteoglycans act as molecular organizers of the extracellular matrix and promoters of cell adhesion. Examples of this important role are numerous and include the large electron-dense aggregates characteristic of cartilage extracellular matrix. The functional interactions that lead to these multimolecular aggregates involve unique terminal domains of the aggrecan core protein that interact noncovalently with other matrix constituents (i.e. hyaluronate, type II collagen), thereby interconnecting the extracellular matrix and constituents of the cell surface. Members of the low molecular weight leucine-rich proteoglycan family (i.e. decorin and fibromodulin) also participate in organizing the extracellular matrix by binding types I and II collagen.

4. BIOSYNTHESIS OF PROTEOGLYCANS

4.1. Glycosaminoglycan chains

Glycosaminoglycans (GAG), with the exception of hyaluronate, are all synthesized as components of proteoglycans. However early studies focussed predominantly on the polysaccharide and only later was the contribution of the protein sequence to selection of GAG type and placement elucidated. Cell-free studies of individual glycosyltransferase reactions revealed the importance of nucleotide sugars which provide energy for the transfer reaction and confer specificity on the enzymes catalyzing the individual reactions. A typical glycosyltransferase reaction involves the transfer of a sugar residue from a donor nucleotide sugar to the nonreducing end of an acceptor sugar, forming a glycosidic bond. The nature of this bond is determined by the specificity of each glycosyltransferase (4).

The linkage region sugar sequence which links the protein backbone to the repeating disaccharide structure is identical for chondroitin sulfate, dermatan sulfate, heparan sulfate and heparin, and thus chondroitin sulfate serves as the prototype for GAG biosynthesis. Chain initiation begins with addition of xylose to a single serine hydroxyl embedded in a specific peptide sequence, catalyzed by the chain-initiating xylosyltransferase. Elongation of the linkage region (galactose-galactose-glucuronic acid) is catalyzed by three distinct glycosyltransferases, each specific with respect to acceptor, donor and linkage formed. The subsequent characteristic repeating polymer of chondroitin sulfate is synthesized by the concerted action of an N-acetylgalactosaminyltransferase and a glucuronosyltransferase. Concomitant with or shortly following polymerization, the GAG chains are sulfated in either the 4 or 6 position of the hexosamine (5).

The relative simplicity of this biosynthetic scheme is contrasted with that of heparin and heparan sulfate which requires the concerted action of several additional modifying enzymes, many of which have now been cloned. These include the first modification enzyme, the N-deacetylase/N-sulfotransferase (NDST), the glucuronic acid C-5 epimerase, the iduronic acid 2-0-sulfotransferase, and the concluding modification enzymes, the glucosamine 6-0- and 3-0- sulfotransferases. Presumably, coordination between chain elongation and modification reactions leads to the regulated diversity of the heparan sulfates synthesized by different cells and tissues (6).

4.2. Core protein processing

Glycosaminoglycan synthesis occurs while the core protein substrate is traversing the intracellular secretory pathway. Most insights into the dynamic and topological aspects of synthesis have resulted from studies on the aggrecan system (5). In the ER N-linked oligosaccharides are added co-translationally to the nascent core protein, while chondroitin sulfate chains are initiated by xylose addition after complete extrusion into the lumen of the ER. The xylosylated precursor core protein is translocated to early compartments of the Golgi for further modification reactions and then moved through the secretory pathway, yielding a fully glycosylated and sulfated aggrecan molecule. Studies using semipermeable cells labeled directly with nucleotide sugar precursors (UDP-xylose, UDP-galactose and UDP-glucuronic acid and the sulfate donor, PAPS), support this biosynthetic model. More recently, the model was confirmed by studies of the processing of the truncated core protein precursor

synthesized by the nanomelic mutant, which is neither processed to a mature proteoglycan nor secreted from the cell; rather it is retained in the ER and becomes xylosylated only (7). While biosynthetic studies for other members of the aggrecan gene family (or other types of proteoglycans) have not been extensive, it is assumed that many aspects of glycosaminoglycan synthesis and assembly onto the various core proteins are similar to those elucidated for aggrecan.

5. REGULATION OF PROTEOGLYCAN GENE EXPRESSION

5.1. Modulation of proteoglycan expression in development

In addition to the dynamic and topological aspects of synthesis and secretion of proteoglycans, it is well established that many proteoglycans are developmentally expressed, during periods where they have been implicated in the regulation of cell migration and pattern formation. One of the most striking examples is the expression of aggrecan concomitant with the onset and establishment of the chondrogenic phenotype. In embryonic chick limb aggrecan begins to be expressed at embryonic day 5, is maximally expressed during the period of chondrocyte differentiation and maturation and remains a biochemical marker of the cartilage phenotype throughout life (8). This phenotypic modulation can be studied *in vitro* as well, where stage-24 limb mesenchyme cells are grown under conditions which promote chondrogenesis. As in the *in ovo* limb, condensation, aggregate formation and nodule accumulation is observed over approximately eight days in culture concomitant with a significant (>50 fold) increase in production of aggrecan mRNA or protein, and an approximate 36 fold increase in type II collagen. The patterns of expression suggest that these two markers of the chondrocyte phenotype may be coordinately controlled, presumably at the level of transcriptional regulation (8).

Most likely, regulation of the aggrecan gene is extremely complex since it is expressed transiently but prominently in other cell types, not of the chondrocyte lineage and in a different developmental time frame. Most interestingly, aggrecan is elsewhere developmentally expressed in patterns clearly distinct from that of cartilage, indicating tissue-specific regulatory mechanisms. In chick brain, aggrecan has a low level of expression at day 7, increases substantially to day 13, markedly decreases by day 16 and is not detected post-hatching (9). This expression pattern coincides with a period of active migration and establishment of neuronal nuclei in the telencephalon. The role of aggrecan in this process, i.e. facilitating or arresting migration, has not been clearly established but cultures derived from the nanomelic chick, whose aggrecan mutation creates a functional knockout, exhibit difficulty establishing neuronal aggregates *in vitro*. Very early in development aggrecan is expressed in the notochord between days 2-5, which correlates temporally with the period of active neural crest cell migration and the onset of sclerotomal differentiation (9). Studies using the nanomelic mutant also correlate aggrecan expression with the notochord's ability to inhibit neural crest cell migration.

Clearly, expression of the aggrecan gene in several unrelated highly differentiated tissues and at different times over the course of development suggests modulation of that gene at the transcriptional level (9).

5.2. Regulation of synthesis and processing

Regulation of expression is also exerted at the level of synthesis and processing of the gene product. Considering the complexity of the proteoglycan molecules, with modular core protein domains and numerous posttranslational modifications that vary with developmental setting, the various steps of synthesis and processing would be expected to be highly regulated as well. Because the series of glycosylation reactions that lead to a mature proteoglycan begin in the ER and continue through the Golgi, compartmentalization is most likely an important aspect of biosynthetic regulation. Since only a few of the CS glycosyltransferases and HS modifying enzymes have been purified, little information is available on properties and cellular localization. However permeabilizing cells, which allow access of nucleotide sugars directly to the site of their utilization, has refined the localization of each step in CS synthesis by defining the compartment where each glycosyltransferase reaction occurs (5).

The site of sugar interconversion has also been identified for one critical set of glycosaminoglycan sugar precursors. In addition to demonstrating production of UDP-xylose from UDP-glucuronic acid and its subsequent incorporation into core protein, UDP glucuronate carboxylyase, the only enzyme known to produce UDP-xylose, has been shown to have a lumenal orientation and co-localize with xylosyltransferase (7). Lumenal generation of UDP-xylose is ideal for a regulatory role in its own production via feedback inhibition of the carboxylyase. Regulation of UDP-xylose levels, critical for initiation of CS, DS, HS and heparin and UDP-glucuronic acid, necessary for the polymerization of these same glycosaminoglycans is envisioned to occur as follows: transport of UDP-glucuronic acid into the ER where it is irreversibly decarboxylated to UDP-xylose, which is subsequently used to xylosylate the proteoglycan core proteins. When xylosylation is complete, the unused UDP-xylose inhibits its further production, allowing an increase in UDP-glucuronic acid levels necessary for subsequent polymerization reactions. Thus there is strict regulation of the levels of both sugar nucleotides and a means to shift precursor levels as needed (7). Most likely similar mechanisms exist for regulation of other GAG precursors.

5.3. Proteoglycan gene mutations

At present, little is known about the regulation of folding of the core proteins or about the surveillance system that recognizes and deals with defective products. A fruitful area for study is genetic disorders of man or animal models which exhibit defects in synthesis of proteoglycans. One in particular is the molecular basis for the truncated version of the aggrecan core protein synthesized by the nanomelic chick. A single base change in the aggrecan gene forms a premature stop codon and results in production of a truncated aggrecan

precursor (10). Most interesting is why this mutation, which produces a core protein precursor missing only the C-terminal G3 domain, does not result in the further processing and secretion of the truncated version of aggrecan, rather than the complete absence of any form of aggrecan in the extracellular matrix. Some insights into this unexpected phenotype came from studies which indicated that the truncated precursor remains in the ER and is not translocated to the Golgi for further processing like its normal counterpart (11). While residing in the ER the nanomelic core protein is fully competent to become xylosylated and N-glycosylated, but unless it is introduced to Golgi functions no further carbohydrate addition or processing occurs. These studies suggest a previously unrecognized role for the C-terminal globular domain in containing possible recognition or retention signals, or that proper folding of this specific domain is necessary to effect exit of the entire core protein precursor from the ER. The latter mechanism would place the nanomelic mutant in the important category of folding abnormalities, which are increasingly being found to be responsible for genetic diseases such as cystic fibrosis (12).

With respect to posttranslational modifications, no naturally occurring mutations in glycosylation are known , while several have been identified for the sulfation process. One of the more interesting phenotypes is the brachymorphic mouse which results from a defect in synthesis of phosphoadenosylphosphosulfate (PAPS) (13). Since PAPS is the universal sulfate donor for all naturally occurring sulfated compounds, it is surprising that the defect in PAPS synthesis in the bm mouse is restricted to a skeletal disorder and does not result in a more severe phenotype. There are other tissues where a significant demand for PAPS might be predicted: i.e. liver which is rich in heparan sulfate and uses sulfoconjugation for detoxification, skin which is rich in dermatan sulfate, or kidney and brain which have high concentrations of sulfatides. This tissue-specific distribution of the defect correlates with the tissue-specific localization of the PAPS synthetase isozymes, with the S/K 2 form , which bears the bm mutation being more highly expressed in cartilage. Thus another level of control is exerted by tissue-specific, developmentally-regulated posttranslational mechanisms, functioning within the context of the requirement for specific products.

Mutations have also been found which affect proteoglycans other than aggrecan. There has been a recent report of a mutation in a heparan sulfate proteoglycan (HSPG) gene that is associated with a multi-faceted, X-linked syndrome. At least two families of cell surface HSPGs exist; the syndecan-like core protein which spans the membrane and the glypican-like proteoglycans which are linked to the cell surface via glycosylphospatidyl inositol (14). Mutations in GPC3, a glypican gene, have recently been shown to be responsible for the Simpson-Golabi-Behmel syndrome, but the mechanism underlying the overgrown phenotype in this syndrome is poorly understood. Interestingly, the Drosophila mutant gene *dally* encodes a protein belonging to the glypican family of cell surface HSPGs (15), which is required for proper

morphogenesis of several tissues. Both phenotypes, in fly and man, suggest a derangement of cellular growth control and illustrate another important level of regulation.

Some intriguing mutants that alter glycosaminoglycan biosynthesis have also been identified in Drosophila. *Sugarless* (*sgl*) encodes a protein homologous to vertebrate UDP glucose dehydrogenase which generates UDP-glucuronic acid used for synthesis of both chondroitin sulfate and heparan sulfate chains. Mutations in *sgl* suggest a role for sgl in Wg and Dpp signalling pathways (16-18). Another gene which disrupts Wg signalling, *sulfateless* (*sfl*) encodes a protein homologous to the HS modifying enzyme deacetylase/sulfotransferase (19). Lastly, mutants with defects in the Drosophila gene *tout-velu (ttv)* exhibit reduction in HS but not CS synthesis, and these mutations affect embryonic patterning by interfering with hedgehog signaling. *Ttv* was shown to encode an EXT1-homologue, which is a known tumor suppressor gene that severely affects cell growth and mutations which may lead to malignancy. EXT-1 and EXT-2 can restore co-polymerase activity, the enzyme that catalyzes transfer of glucuronic acid and N-acetylglucosamine to growing HS chains, in deficient cell lines (20).

5.4. Growth factor regulation

A large number of studies have focussed on the cellular and molecular factors that control the synthesis of aggrecan, in order to understand its role as a component of the cartilage extracellular matrix in such pathological conditions as osteoarthritis and rheumatoid arthritis. For instance TGF-beta appears to promote the chondrocytic phenotype by stimulating aggrecan and type II collagen expression, as do members of the bone morphogenetic protein (BMP) family (21).

Although information is far from complete, most of the other aggrecan family members as well as those of the other proteoglycan families, have been shown to be regulated in developmentally-specific or tissue-specific manners or modulated in response to growth- or differentiation-related factors. For instance, the effects of TGF-beta on biosynthesis of the membrane-anchored syndecans, which exhibit a broad tissue distribution, vary with the cell type examined. TGF-beta treatment causes cells of mesenchymal origin (e.g. 3T3) to increase syndecan-1 expression, while TGF-beta has no effect on syndecan-1 expression in endothelial cells, keratinocytes or mammary epithelial cells, which predominantly synthesize syndecan-1 (22). Thus much is yet to be elucidated about regulatory pathways controlling expression of the proteoglycans.

6. REGULATORY ELEMENTS AND TRANSCRIPTIONAL CONTROL

Although the number of proteoglycan genes identified and sequenced has increased significantly over the past decade, there is still limited information on the genomic organization or regulatory elements of the proteoglycan genes. The structural features of several

proteoglycan genes have been determined, but neither the mechanism of functional promoter activity nor the transcription factors involved in their regulation have been fully elucidated for a single proteoglycan gene. Surveys of the current status of knowledge concerning transcriptional regulation of proteoglycan gene expression have recently appeared (1,2). Our studies in the chick have provided the strongest evidence that the aggrecan gene is expressed spatially and temporally during development, and in cell-specific and exquisitely timed modes (8,9). Our findings have naturally led us to seek elucidation of the temporal and cell-specific control elements that regulate expression of the aggrecan gene during development. Although mouse, rat and human aggrecan promoters have been identified, no functional analysis of developmental tissue-specificity of expression has been forthcoming for the mammalian aggrecan genes.

We cloned the 5'-flanking region of the chick aggrecan gene and found that a 1.8 kb genomic fragment of the coding region is able to drive expression of a pGL2-Basic plasmid-borne luciferase reporter gene which contains the aggrecan promoter and proximal enhancer regions (23). Using transient transfections of chick chondrocytes, fibroblasts and neurons with reporter plasmids bearing progressive deletions of the aggrecan promoter and enhancer region, tissue-specific promoter activity was demonstrated. This analysis elucidated a 400 bp region in the genomic 5' flanking sequence responsible for negative regulation of the aggrecan gene. Sequencing of the complete 1.8 kb region identified *cis* elements of interest with respect to control of aggrecan expression. At two positions 5' to the transcription start site lie copies of the CACCTCC (CIIS2) sequence which has been suggested to be a silencer motif in the COL2A1 promoter and thereby inhibits transcription of the type II collagen promoter in fibroblasts but not chondrocytes. Deletion of these two motifs from the aggrecan promoter reduced cell-type specificity of reporter expression while overall promoter activity increased. A second silencer consensus sequence, ACCTCTCT (CIIS1), present in the COL2A1 promoter is also found in the aggrecan promoter.

Numerous other potential *cis* elements have also been identified. Of particular interest is the CACACA motif present at four positions in the chick aggrecan promoter. The proximal promoter region of COL2A1, which is responsible for regulating expression in hypertrophic chondrocytes, contains a novel transcription factor binding sequence, ACACACAGA. Footprinting analysis suggests it acts as a silencer in the regulation of COL10A1 transcription. The CACACA motif is also interesting as repeats of (CA)n are markers for Z-DNA formation, contribute to secondary structure and are potential hot spots for recombination. We have established that the chick aggrecan 5' flanking sequence contains three major transcription start sites in addition to multiple cis elements. Clustering of these putative binding sites near the transcriptional start sites may contribute to transcriptional regulation by altering DNA secondary structure. Although far from complete, these studies serve as a model for studying up-regulation of aggrecan in cartilage development and down-regulation in certain disease states like osteo- or rheumatoid arthritis (1,23).

With respect to the other classes of proteoglycans, i.e. cell-associated, basement membrane and small leucine-rich, gene and promoter elements have been characterized to a variable degree (2). As well, the effects of growth factors and cytokines on expression patterns and the relationship of proteoglycan synthesis to cell cycle regulation and cell growth properties are beginning to be addressed for certain of these proteoglycan family members. Clearly these will continue to be active areas of investigation for the forseeable future.

Another fruitful area that is just beginning to be addressed is the role of transcription factors that control proteoglycan expression by studying models with mutations in transcription factors. One candidate that falls in this realm may be campomelic dysplasia, a congenital skeletal abnormality in humans with characteristics similar to those observed in the nm and cmd animal mutants (12). This disorder appears to be caused by mutations in the SRY-related gene, SOX9. SOX genes constitute a large family of developmentally regulated genes coding for transcription factors, and SOX9 in particular is expressed in limb mesenchymal condensations prior to cartilage formation and its expression is maintained in perichondrium and chondrocytes of the resting, proliferative and upper hypertrophic zones (24). Thus this transcriptional activator may play a role in establishing and maintaining the chondrocytic phenotype, by perhaps controlling cartilage-specific genes like those for the type II and XI collagens and aggrecan. Clearly these examples, as well as the HSPG Drosophila mutants previously discussed, indicate that proteoglycans are key players in cell-cell signalling and morphogenesis.

7. PERSPECTIVE

As this brief treatise implies, we still have an incomplete picture of the patterns and modes of expression of the various proteoglycan gene families. While information in these areas is growing, we are also beginning to decipher the various levels and mechanisms by which the synthesis and expression of proteoglycans are regulated. This is an especially daunting task for these glycan-containing molecules, because a non-template mechanism involving a battery of biosynthetic enzymes is used for determining the structure and composition of the glycosaminoglycan portion, while the protein backbone is derived via a genetic template mechanism. For overall successful generation of proteoglycans *in vivo*, both paradigms must be considered simultaneously.

From the genetic disorders increasingly being identified due to mutations in proteoglycan molecules, as well as the numerous critical functions ascribed to these complex, versatile molecules during development and maintenance of all tissue and organ systems of higher organisms, methods for controlling the gene expression of proteoglycans are desired. Accomplishing this last frontier will require a thorough understanding of transcriptional control mechanisms of the core protein genes, the glycosyltransferase genes and the carbohydrate-modifying enzyme genes, as well as the complete elucidation of the

transducing components, i.e. those receptors and signalling molecules that regulate the transcriptional activity of all of these genes. There is much exciting work yet to be done.

8. ACKNOWLEDGMENTS

9. REFERENCES

1. Schwartz, N. B., E. W. Pirok, 3rd, J. R. Mensch, Jr. & M. S. Domowicz: Domain organization, genomic structure, evolution, and regulation of expression of the aggrecan gene family. *Prog Nucleic Acid Res Mol Biol* 62, 177-225 (1999)

2. Iozzo, R. V. & K. G. Danielson: Transcriptional and posttranscriptional regulation of proteoglycan gene expression. *Prog Nucleic Acid Res Mol Biol* 62, 19-53 (1999)

3. Schwartz, N.: Proteoglycans In: Encyclopedia of Life Sciences. Eds: Nature Publishing Group, London (1999)

4. Schwartz, N. B.: Carbohydrate Metabolism II and Special Pathways In: Textbook of Biochemistry. Eds: T. Devlin, John Wiley and Son, New York (1997)

5. Schwartz, N. B.: Xylosylation, the first step in the synthesis of proteoglycans. *Trends Glycosci. Glycotechnol.* 7, 429-445 (1995)

6. Lindahl, U., M. Kusche-Gullberg & L. Kjellen: Regulated diversity of heparan sulfate. *J Biol Chem* 273, 24979-82 (1998)

7. Kearns, A. E., B. M. Vertel & N. B. Schwartz: Topography of glycosylation and UDP-xylose production. *J. Biol. Chem.* 268, 11097-11104 (1993)

8. Schwartz, N. B., A. K. Hennig, R. C. Krueger, M. Krzystolik, H. Li & D. Mangoura: Developmental expression of S103L cross-reacting proteoglycans in embryonic chick In: Limb Development and Regeneration. Eds: J. F. Fallon, P. F. Goetinck, R. O. Kelley and D. L. Stocum, Wiley-Liss, Inc., New York (1993)

9. Schwartz, N. B., M. Domowicz, R. K. Krueger, H. Li & D. Mangoura: Brain Aggrecan. *Perspectives in Dev. Neurobiol.* 3, 291-306 (1996)

10. Li, H., N. B. Schwartz & B. M. Vertel: 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 (1993)

11. Vertel, B. M., B. L. Grier, H. Li & N. B. Schwartz: The chondrodystrophy, Nanomelia: biosynthesis and processing of the defective aggrecan precursor. *Biochem J.* 301, 211-216 (1994)

12. Schwartz, N. B. & M. Domowicz: Proteoglycan gene mutations and impaired skeletal development In: Skeletal Growth and Development. Eds: J. A. Buckwalter, M. G. Ehrlich, L. J. Sandell and S. B. Trippel, American Association of Orthopedic Surgeon Publications, Rosmont, IL (1998)

13. Kurima, K., M. L. Warman, S. Krishnan, M. Domowicz, R. C. Krueger, Jr., A. Deyrup & N. B. Schwartz: A member of a family of sulfate-activating enzymes causes murine brachymorphism. *Proc Natl Acad Sci U S A* 95, 8681-5 (1998)

14. David, G.: Integral membrane heparan sulfate proteoglycans. *FASEB J.* 7, 1023-1030 (1993)

15. Nakato, H., T. A. Futch & S. B. Selleck: The division abnormally delayed (dally) gene: a putative integral membrane proteoglycan required for cell division patterning during postembryonic development of the nervous system in Drosophila. *Development* 121, 3687-702 (1995)

16. Binari, R. C., B. E. Staveley, W. A. Johnson, R. Godavarti, R. Sasisekharan & A. S. Manoukian: Genetic evidence that heparin-like glycosaminoglycans are involved in wingless signaling. *Development* 124, 2623-32 (1997)

17. Hacker, U., X. Lin & N. Perrimon: The Drosophila sugarless gene modulates Wingless signaling and encodes an enzyme involved in polysaccharide biosynthesis. *Development* 124, 3565-73 (1997)

18. Haerry, T. E., T. R. Heslip, J. L. Marsh & M. B. O'Connor: Defects in glucuronate biosynthesis disrupt Wingless signaling in Drosophila. *Development* 124, 3055-64 (1997)

19. Lin, X. & N. Perrimon: Dally cooperates with Drosophila Frizzled 2 to transduce Wingless signalling [see comments]. *Nature* 400, 281-4 (1999)

20. Lind, T., F. Tufaro, C. McCormick, U. Lindahl & K. Lidholt: The putative tumor suppressors EXT1 and EXT2 are glycosyltransferases required for the biosynthesis of heparan sulfate. *J Biol Chem* 273, 26265-8 (1998)

21. Hogan, B. L. M.: Bone morphogenetic protein, multifunctional regulators of vertebrate development. *Genes Dev* 10, 1580-1591 (1996)

22. Bernfield, M., R. Kokenyesi, M. Kato, M. T. Hinkes, J. Spring, R. L. Gallo & E. J. Lose: Biology of the syndecans: a family of transmembrane heparan sulfate proteoglycans. *Annu Rev Cell Biol* 8, 365-93 (1992)

23. Pirok, E. W., H. Li, J. R. Mensch, J. Henry & N. B. Schwartz: Structural and functional analysis of the chick CSPG (Aggrecan) promoter and enhancer region. *J. Biol. Chem.* 272, 11566-11574 (1997)

24. Wagner, T., J. Wirth, J. Meyer, B. Zabel, M. Held, J. Zimmer, J. Pasantes, E. Hustert, U. Wolf, N. Tommerup, W. Schempp & G. Scherer: Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the *SRY*-related gene *SOX9. Cell* 79, 1111-1120 (1994)

Send correspondence to: Dr Nancy B. Schwartz, University of Chicago, Department of Pediatrics, 5841 S. Maryland Avenue, MC 5058, Chicago, IL 60637, Tel: 773-702-6426, Fax: 773-702-9234, E-mail: n-schwartz@uchicago.edu

This manuscript is available on line at:

http://www.bioscience.org/2000/d/schwartz/fulltext.htm