

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Chondroitin proteoglycans in Caenorhabditis elegans development

Permalink

<https://escholarship.org/uc/item/8pt667hf>

Author

Olson, Sara Kathryn

Publication Date

2005

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Chondroitin Proteoglycans

in

Caenorhabditis elegans Development

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Sara Kathryn Olson

Committee in charge:

Professor Jeffrey D. Esko, Chair
Professor Raffi V. Aroian
Professor Richard L. Gallo
Professor Karen Oegema
Professor Ajit P. Varki

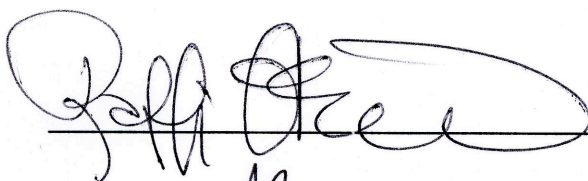
2005

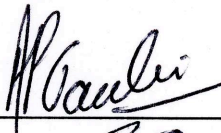
Copyright

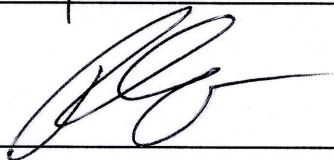
Sara Kathryn Olson, 2005

All rights reserved

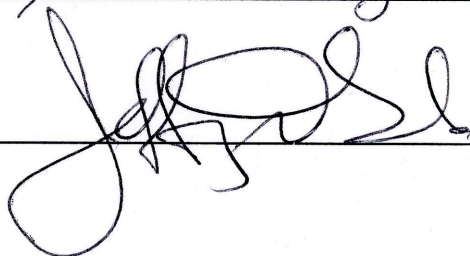
The dissertation of Sara Kathryn Olson is approved, and it is acceptable in quality and form for publication on microfilm.











Chair

University of California, San Diego

2005

Dedication

I would like to dedicate this dissertation to:

My parents...for challenging me from a young age, encouraging me to aim high, understanding the decisions and choices I have made, and providing unwavering love and support. You have brought me to where I am today.

Jeff Esko...for being a better mentor than I ever could have hoped for. Thank you for having faith in me, even during the “Dark Ages.” Your love of science is so pure and complete, but what is even more impressive is your ability to compliment it with a true passion for life.

Nancy Wall and Elizabeth De Stasio...for the memorable years at Lawrence University. You fostered my interest in developmental biology and provided me with the tools to succeed. The two of you truly are “The Lawrence Difference.”

Scott Selleck...for inspiring me to pursue a career in research. Thank you for giving me the opportunity to spend a summer in your lab. You opened my eyes to the joy and excitement of benchwork, and made glycobiology seem not quite so scary.

All the wonderful people I met at UCSD...you made graduate school much more fun than it probably should have been.

Table of Contents

Signature Page.....	iii
Dedication.....	iv
Table of Contents.....	v
List of Figures and Tables.....	xi
Acknowledgements.....	xiii
Vita.....	xiv
Abstract.....	xv

CHAPTER 1:

Introduction.....	1
A. Chondroitin sulfate proteoglycan structure.....	2
1. Chondroitin glycosaminoglycan chains.....	2
2. Protein cores.....	5
B. Function of chondroitin sulfate proteoglycans.....	10
1. Mutations resulting in cartilage defects.....	10
2. Mutations resulting in connective tissue disorders.....	11
3. Mutations resulting in central nervous system defects.....	12
4. The role of hybrid Syndecan family members in development.....	13
C. Chondroitin sulfate biosynthesis.....	16
1. Nucleotide sugar building blocks.....	16

2. Linkage region biosynthesis.....	17
3. Polymerization of chondroitin sulfate.....	19
4. Chondroitin sulfate modification.....	21
D. <i>C. elegans</i> as a model system to study chondroitin proteoglycans.....	21
1. Advantages of the <i>C. elegans</i> model.....	21
2. Genetic screen for squashed vulva (<i>sqv</i>) mutants.....	23
3. <i>sqv</i> genes encode components of the GAG biosynthesis pathway.....	30
E. Concluding remarks.....	37
F. References.....	39

CHAPTER 2:

The <i>C. elegans</i> genes <i>sqv-2</i> and <i>sqv-6</i> , which are required for vulval morphogenesis, encode glycosaminoglycan galactosyltransferase II and xylosyltransferase.....	49
A. Summary.....	49
B. Introduction.....	50
C. Experimental Procedures.....	52
1. <i>C. elegans</i> maintenance.....	52
2. Molecular biology.....	52
3. Rescue of <i>C. elegans sqv-2</i> and <i>sqv-6</i> mutants.....	53
4. SQV-2 galactosyltransferase II assay.....	53
5. Rescue of CHO pgsA-745 cells by <i>sqv-6</i>	54
6. SQV-6 xylosyltransferase assay.....	54

D. Results and Discussion.....	55
1. Molecular identification of <i>sqv-2</i>	55
2. <i>sqv-2</i> encodes a protein similar to galactosyltransferase II.....	56
3. SQV-2 has galactosyltransferase II activity.....	56
4. Molecular identification of <i>sqv-6</i>	60
5. <i>sqv-6</i> encodes a protein similar to xylosyltransferases.....	61
6. <i>sqv-6</i> can correct a xylosyltransferase defect in CHO cells.....	63
7. The <i>sqv-2</i> and <i>sqv-6</i> genes act in the <i>C. elegans</i> chondroitin and heparan sulfate biosynthesis pathway.....	65
E. Acknowledgements.....	67
F. References.....	68

CHAPTER 3:

<i>C. elegans</i> early embryogenesis and vulval morphogenesis depend on chondroitin glycosaminoglycan biosynthesis.....	72
A. Summary.....	72
B. Introduction.....	72
C. Results.....	75
1. Cloning of <i>sqv-5</i>	75
2. SQV-5 is similar to the human chondroitin sulfate synthase.....	78
3. SQV-5 has chondroitin synthase activity.....	81
4. <i>sqv-5(n3611)</i> mutants have decreased chondroitin levels.....	84

5. SQV-5 localization.....	85
6. The Sqv phenotype results from loss of chondroitin, not heparan sulfate.....	87
D. Methods.....	90
1. <i>sqv-5</i> mapping.....	90
2. <i>sqv-5</i> cDNA.....	90
3. Deletion allele of <i>sqv-5</i>	91
4. Glycosyltransferase assays.....	91
5. Chondroitin and heparan sulfate characterization.....	93
6. Anti-SQV-5 antibodies.....	93
E. Acknowledgements.....	94
F. References.....	95

CHAPTER 4:

Novel chondroitin proteoglycans mediate early embryonic cell division in <i>Caenorhabditis elegans</i>	99
A. Summary.....	99
B. Introduction.....	100
C. Materials and methods.....	102
1. <i>C. elegans</i> maintenance.....	102
2. <i>In silico</i> analysis.....	102
3. Biochemical purification.....	103

4. Mass spectrometry.....	104
5. Recombinant protein expression.....	104
6. RNAi.....	105
7. Western blotting.....	106
D. Results.....	106
1. <i>C. elegans</i> expresses multiple novel CPGs.....	106
2. CPG-1 and CPG-2.....	117
3. CPG-1 and CPG-2 are required for cytokinesis and embryonic viability.....	121
E. Discussion.....	127
1. A proteomic approach for identifying novel proteoglycans.....	127
2. CPGs play essential roles in embryonic cell division in <i>C. elegans</i>	129
F. Acknowledgements.....	131
G. References.....	133

CHAPTER 5:

Perspective and Future Directions.....	139
A. Evolution of chondroitin proteoglycan components.....	139
1. Vertebrate chondroitin synthesis is conserved in <i>C. elegans</i>	139
2. <i>C. elegans</i> expresses a novel set of chondroitin core proteins.....	143
B. A proteomics approach to proteoglycan identification.....	144
1. Universal applicability.....	144

2. Advantages.....	145
3. Limitations.....	146
C. Functions of <i>C. elegans</i> chondroitin proteoglycans.....	148
1. Functional redundancy.....	148
2. Temporal and spatial expression.....	150
D. Model for CPG function in embryonic cell division.....	151
1. Site of action.....	152
2. Role in embryogenesis.....	152
E. Model for CPG function in vulval invagination.....	157
1. Generation of a hydrated matrix by an osmotic gradient.....	157
2. Epithelial invagination in sea urchin development.....	157
F. Concluding remarks.....	158
G. References.....	159

APPENDICIES

Appendix A. Cloning, Golgi localization, and enzyme activity of the full-length heparin/heparan sulfate-glucuronic acid C5-epimerase.....	164
Appendix B. Enzyme interactions in heparan sulfate biosynthesis: uronosyl 5-epimerase and 2-O-sulfotransferase interact <i>in vivo</i>	172

List of Figures and Tables

Chapter 1

Table 1. Modification of the chondroitin backbone.....	4
Table 2. Proteoglycan families.....	9
Table 3. <i>In vivo</i> functions of chondroitin sulfate proteoglycans.....	15
Figure 1. Nucleotide sugar transport and linkage tetrasaccharide synthesis.....	18
Figure 2. Multiple enzymes polymerize the chondroitin backbone.....	20
Figure 3. Vulval development and the squashed vulva (Sqv) phenotype.....	25
Figure 4. <i>C. elegans</i> early embryogenesis.....	29
Figure 5. The <i>sqv</i> genes encode components of the GAG biosynthetic pathway.....	36

Chapter 2

Figure 1. SQV-2 is similar to galactosyltransferase II.....	58
Table 1. The SQV-2 fusion protein has acceptor specificity consistent with its being galactosyltransferase II.....	59
Figure 2. SQV-6 is similar to xylosyltransferases.....	62
Figure 3. <i>sqv-6</i> rescues a xylosyltransferase-deficient CHO cell line.....	64
Figure 4. Model for the role of seven <i>sqv</i> genes in GAG biosynthesis.....	66

Chapter 3

Figure 1. Cloning of <i>sqv-5</i>	77
---	----

Table 1. Amino acid sequence identities.....	79
Figure 2. Sequence alignment of SQV-5 and homologs.....	80
Figure 3. SQV-5 has chondroitin synthase activity.....	83
Figure 4. SQV-5 localization.....	86
Figure 5. Model.....	89

CHAPTER 4

Figure 1. <i>C. elegans</i> expresses multiple CPGs.....	109
Figure 2. Material released by BEMAD is ABC-sensitive.....	110
Figure 3. <i>C. elegans</i> CPGs identified by mass spectrometry.....	114
Figure 4. CPG-1 through CPG-6 are CPGs.....	116
Figure 5. CPG-1 and CPG-2 protein sequences resemble proteoglycans, and CPG-2 acts as a CPG <i>in vivo</i>	120
Figure 6. CPG-1 and CPG-2 are required for embryonic cell division.....	124
Table 1. 24-hour brood size and viability in RNAi-treated animals.....	126

CHAPTER 5

Figure 1. Chondroitin biosynthetic enzymes are conserved in <i>C. elegans</i>	142
---	-----

Acknowledgements

Chapter 1, in part, includes material as it appeared in the Encyclopedia of Biological Chemistry, Olson, S.K. and Esko, J.D. (2004). The dissertation author was the primary author and the co-author listed in this publication directed and supervised the writing which forms the basis for this chapter.

The text of Chapter 2 is a reprint of the material as it appeared in the Journal of Biological Chemistry, Hwang, H.Y., Olson, S.K., Brown, J.R., Esko, J.D. and Horvitz, H.R. (2003). The dissertation author was a secondary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

The text of Chapter 3 is a reprint of the material as it appeared in the journal Nature, Hwang, H.Y., Olson, S.K., Esko, J.D. and Horvitz, H.R. (2003). The dissertation author was a secondary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

The text of Chapter 4 is a manuscript in preparation, Olson, S.K., Bishop, J.R., Yates, J.R. 3rd, Oegema, K. and Esko, J.D. (2005). The dissertation author is the primary researcher and author and the co-authors listed in this manuscript directed and supervised the research which forms the basis for this chapter.

The text of Appendix A is a reprint of the material as it appeared in the Journal of Biological Chemistry, Crawford, B.E., Olson, S.K., Esko, J.D. and Pinhal, M.A. (2001). The dissertation author was a secondary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

The text of Appendix B is a reprint of the material as it appeared in the Proceedings of the National Academy of Sciences USA, Pinhal, M.A., Smith, B., Olson, S., Aikawa, J., Kimata, K. and Esko, J.D. (2001). The dissertation author was a secondary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

Vita

- 1998 Phi Beta Kappa
- 1999 Bachelors Degree, Magna Cum Laude in course, Lawrence University, WI
- 2001 NIH Genetics Training Grant
- 2005 Ph.D., University of California, San Diego

PUBLICATIONS

Tsuda, M., Kamimura, K., Nakato, H., Archer, M., Staatz, W., Fox, B., Humphrey, M., Olson, S., Futch, T., Kaluza, V., Siegfried, E., Stam, L. and Selleck, S.B. (1999). The cell-surface proteoglycan Dally regulates Wingless signalling in *Drosophila*. *Nature* **400**: 276-80.

Crawford, B.E., Olson, S.K., Esko, J.D. and Pinhal, M.A. (2001). Cloning, Golgi localization, and enzyme activity of the full-length heparin/heparan sulfate-glucuronic acid C5-epimerase. *J Biol Chem* **276**: 21538-43.

Pinhal, M.A., Smith, B., Olson, S., Aikawa, J., Kimata, K. and Esko, J.D. (2001). Enzyme interactions in heparan sulfate biosynthesis: uronosyl 5-epimerase and 2-O-sulfotransferase interact *in vivo*. *Proc Natl Acad Sci USA* **98**: 12984-9.

Hwang, H.Y., Olson, S.K., Brown, J.R., Esko, J.D. and Horvitz, H.R. (2003). The *Caenorhabditis elegans* genes *sqv-2* and *sqv-6*, which are required for vulval morphogenesis, encode glycosaminoglycan galactosyltransferase II and xylosyltransferase. *J Biol Chem* **278**: 11735-11738.

Hwang, H.Y., Olson, S.K., Esko, J.D. and Horvitz, H.R. (2003). *Caenorhabditis elegans* early embryogenesis and vulval morphogenesis require chondroitin biosynthesis. *Nature* **423**: 439-443.

Olson, S.K. and Esko, J.D. (2004) Proteoglycans, In Encyclopedia of Biological Chemistry, Lennarz, W. and Lane, M.D. (Eds.) Elsevier, Oxford, Vol. 3, pp.549-555.

Olson, S.K., Bishop, J.R., Yates, J.R., Oegema, K. and Esko, J.D. (2005). Novel chondroitin proteoglycans mediate early embryonic cell division in *C. elegans* (manuscript in preparation).

ABSTRACT OF THE DISSERTATION

Chondroitin Proteoglycans

in

Caenorhabditis elegans Development

by

Sara Kathryn Olson

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2005

Professor Jeffrey D. Esko, Chair

The role of non-sulfated chondroitin in developmental events of the nematode *Caenorhabditis elegans* has been investigated. This work took advantage of a class of mutants defective in glycosaminoglycan biosynthesis that show perturbations in vulval invagination and early embryogenesis. It was demonstrated that all components of the mammalian chondroitin biosynthetic machinery are conserved in *C. elegans*, and that

the phenotypic defects observed in the mutants result from loss of chondroitin, not heparan sulfate. A biochemical screen identified nine novel chondroitin core proteins that carry chondroitin chains, none of them common to mammalian chondroitin sulfate core proteins. Two of these, chondroitin proteoglycans-1 and -2, are functionally redundant and required for the same set of early embryonic events affected in mutants lacking chondroitin biosynthetic enzymes. We hypothesize that the chondroitin proteoglycans play biophysical or structural roles that direct vulval invagination and embryogenesis in the worm.

CHAPTER 1

Introduction

Since its discovery in the late 1800's, the importance of chondroitin sulfate (CS) has been demonstrated in many aspects of mammalian biology. Historically, studies of CS and the core proteins that harbor these sugar chains, called proteoglycans (PGs), have centered on biochemical purification and chemical analyses of the CSPG components. A large number of core proteins have been identified, and their expression patterns suggest specific sites of action. Additionally, the structures of different sulfated forms of chondroitin have been determined, as have many of the enzymes that catalyze biosynthesis of the chains. The relatively recent advent of reverse genetic technology enables us to study specific chondroitin sulfate biosynthetic enzymes and protein cores in an *in vivo* setting. Gene knockout technology provides valuable information about genes that are known to be involved in generating CSPGs, but analysis has proven to be complicated at times by apparent functional redundancy in the system. The complexity of vertebrate systems can be circumvented by use of model organisms with a more rudimentary system. Specifically, the use of *Caenorhabditis elegans* has afforded insight not only into identification of chondroitin biosynthesis machinery, but also into the evolution of chondroitin sulfate proteoglycans as a whole.

A. Chondroitin sulfate proteoglycan structure

Proteoglycans are a particular class of glycoproteins that are comprised of two main structural components - long unbranched sugar chains called glycosaminoglycans (GAGs) and the protein cores that serve as their scaffold. There are five main classes of GAGs – chondroitin sulfate (CS), a derivative of CS called dermatan sulfate (DS), heparan sulfate (HS), hyaluronan (HA) and keratan sulfate (KS). GAG chains show variable degrees of complexity and modification, resulting in heterogeneous molecules that are able to interact with a wide array of ligands (Esko and Selleck, 2002). The chondroitin sulfate proteoglycan (CSPG) family comprises a large spectrum of structurally diverse proteins, with the presence of variably modified CS chains on the cores being the common thread that links them together.

1. Chondroitin glycosaminoglycan chains

The chondroitin backbone is a linear polysaccharide composed of repeating, alternating units of N-acetyl-galactosamine (GalNAc) and glucuronic acid (GlcA). Linkage of the sugars is highly structured and invariable, consisting of repeating GalNAc β 1-4GlcA β 1-3 disaccharide units. Various modifications of the chondroitin backbone increase the complexity of the molecule, generating variations of the chain (Table 1). Sulfotransferase enzymes catalyze the addition of sulfate groups to either the 4-O or 6-O position of GalNAc (chondroitin sulfate A and C, respectively). Chondroitin sulfate B (CS-B), more commonly known as dermatan sulfate (DS), is characterized by the epimerization of a subset of GlcA units to IdoA. CS-D is

similar to CS-C with the additional modification of sulfate at the 2-O position of GlcA. CS-E is di-sulfated at both the 4-O and 6-O positions of GalNAc, but is a rare modification. Heterogeneous sulfation patterns at these various positions help generate binding sites for CS ligands, which include collagen, laminin, and other extracellular matrix components, as well as various growth factors. CS-A through -E are commonly found in more complex organisms; interestingly, non-sulfated chondroitin is rarely, if ever, seen in vertebrates and its biological role is unknown. Chapter 3 addresses this question and demonstrates that a non-sulfated form of chondroitin plays an essential role in the development of the nematode *Caenorhabditis elegans*.

Table 1. Modification of the chondroitin backbone

Chondroitin type	Disaccharide Unit
Unmodified	GalNAc β 1-4GlcA β 1-3
A	GalNAc4S-GlcA
B	GalNAc4S-IdoA2S
C	GalNAc6S-GlcA
D	GalNAc6S-GlcA2S
E	GalNAc4,6diS-GlcA

2. Protein cores

Chondroitin sulfate does not exist as a free polymer, but rather is attached to a core protein. In many instances the CS chain is the source of biological activity and the core acts as a passive scaffold, but recent evidence suggests that some proteoglycan cores themselves serve functional roles. The number of GAG chains carried by a proteoglycan is highly variable, ranging from a single chain (decorin) to more than 100 chains (aggrecan). Of the potential glycosylation sites, the extent of modification is also variable, with any or all glycosylation sites being utilized. It is unknown whether this variability is spatially and temporally regulated. A loose consensus sequence for GAG attachment sites on the core protein has been identified in vertebrates (Zhang and Esko, 1994). GAG chains are initiated at serine (Ser) residues that are followed glycine (Gly). These sites are typically flanked by the acidic amino acids aspartate (Asp) and glutamate (Glu). It is unknown whether this consensus sequence is universal or specific to vertebrates. Chapter 4 demonstrates that the amino acid composition of GAG attachment sites is maintained even in the more simplistic nematodes.

The majority of work on chondroitin proteoglycans has been performed with human and mouse samples. More than 20 chondroitin sulfate proteoglycans (CSPGs) have thus far been identified in vertebrate systems and are expressed in a wide array of tissues (Table 2). They can carry solely CS or exist as hybrids carrying CS/HS or CS/KS. CSPGs can be secreted into the extracellular matrix (ECM), embedded in the cell membrane, or stored in secretory granules. Some

CSPGs are large globular molecules reaching masses of 400 kD while others have tiny masses around 10 kD. While structurally diverse, the mammalian CSPGs can be classified into the following families.

a. Aggrecan

The Aggrecan family of proteoglycans, also known as the Hyallectins, includes aggrecan, versican, brevican, and neurocan (Table 2). All four members share the following common features. The N-terminal domain is able to bind HA, another class of GAG that exists in ECM as a free polymer not attached to a protein core. The protein-GAG interaction between aggrecan and HA, modulated by the “link protein,” is particularly strong and non-dissociating under physiological conditions (Knudson and Knudson, 2001). The central region of the family members contains variable numbers of GAG attachment sites. Aggrecan can carry more than 100 CS chains, while neurocan as few as one. The C-terminal region contains a C-type lectin domain, which is defined as a protein domain able to bind carbohydrates in a calcium-dependent manner. Few binding partners are known to interact with the C-type lectin domain, but they include tenascin-C, tenascin-R, fibulin-1, and sulfated glycolipids (Aspberg et al., 1999; Aspberg et al., 1997; Miura et al., 1999; Rauch et al., 1997). All members of the Aggrecan family are expressed in the central nervous system, though individual members have additional distinct localization patterns including cartilage, heart, tendon, hepatocytes, and fibroblasts (Table 2).

b. Small Leucine-Rich Proteoglycan (SLRP), or Small Interstitial

The SLRP family of secreted proteoglycans includes members that contain either CS or KS chains. The CS family members are listed in Table 2. The hallmark of SLRPs is the presence of leucine-rich repeats (LRRs) in the central domain (reviewed in Iozzo, 1999). The number of LRRs varies between SLRP classes. For example, Class I members decorin and biglycan both contain ten LRRs, while Class III epiphygan contains only six. The N-terminal region of SLRPs contains a cluster of four cysteine residues that are able to form disulfide bridges. Spacing of cysteines in characteristic patterns is conserved between classes and is one determinant of class assignment. Many SLRPs are expressed in connective tissues and are thought to bind collagen and inhibit fibrillogenesis.

c. Miscellaneous

More than half of the mammalian CSPGs do not belong to the Aggrecan or SLRP families and make up a diverse set of glycoproteins (Table 2). Some of these are secreted into the ECM of their respective tissues and may serve structural or biophysical roles, while others are membrane bound, likely interacting with ligands at the cell surface or promoting cell-cell interactions. Type IX collagen is comprised of three subunits, only one of which ($\alpha 2$) carries a single CS chain. Its expression in cartilage suggests structural functions. The Syndecan family of proteoglycans contains four members that carry long HS chains; only two of these, syndecan-1 and syndecan-3, have been shown to contain short CS chains at the base of the extracellular domain (reviewed in Bellin et al., 2002). All four

members are transmembrane proteins. The short cytoplasmic tail contains a PDZ binding motif suggesting involvement in signal transduction pathways or interaction with cytoskeletal components.

NG2, expressed on the surface of immature precursor cells, contains a collagen binding site in the extracellular domain, as well as a PDZ binding motif in the cytoplasmic domain. Its CS chains reside in the central region of the protein. CD44 is occasionally modified with CS, though its interaction with HA is the more dominant feature. It is expressed by lymphocytes, but also by chondrocytes where it interacts with the HA-aggrecan complex. Serglycin is the main PG localized to intracellular secretory granules of hematopoietic and connective tissue mast cells. Upon stimulation, granule contents are released into the blood stream. Serglycin is modified with both CS and heparin.

The structural diversity of the CSPGs is quite amazing. Such diversity would suggest that these molecules play defined roles in different biological systems.

Table 2. Proteoglycan Families

Class	Family	Proteoglycan	Core mass (kD)	Chain #	Chain Type	Tissue expression
Matrix (Secreted)	Aggrecan (Hyalactins)	Aggrecan	208-220	~100	CS/KS	cartilage, central nervous system (CNS), tendon, heart
		Vertecan	265	12 to 15	CS	fibroblasts, CNS, chondrocytes, hepatocytes, smooth muscle
		Neurocan	145	1 to 2	CS	CNS
		Brevican	96	0 to 4	CS	CNS
Small Leucine Rich (Small Interstitial)		Decorin	36	1	CS	connective tissue cells, cartilage
		Biglycan	38	1 to 2	CS	connective tissue cells, cartilage, CNS
		Endocan	15	1	CS/DS	vascular endothelial cells
		Epiphysean	35	2 to 3	CS	epiphyseal cartilage (bone)
Miscellaneous		α 2(IX) collagen	68	1	CS	cartilage, vitreous humor
		Testicans 1-3 (SPOCK)	47-49	2	CS/HS	testicles, CNS, vascular endothelium, cartilage
		Leprecan	82	2 to 3	CS	vasculature and smooth muscle of many tissues
		Serglycin	10 to 19	10 to 15	CS/HS	hematopoietic cells
Membrane Bound	Miscellaneous	Syndecan-1 and -3	31-45	1 to 3	CS/HS	epithelial cells, CNS
		NG2 (MCSP)	251	2 to 3	CS	immature cells, CNS
		Phosphocan (RPTP β)	176	4	CS/KS	CNS
		Neuroglycan C	57	1	CS	CNS
		Betaglycan	110	1 to 2	CS/HS	fibroblasts
		CD44	37	1 to 4	CS	lymphocytes, chondrocytes
		Thrombospondin	58	1	CS	endothelial cells
		Invariant chain	31	1	CS	antigen-processing cells

B. Function of chondroitin sulfate proteoglycans

The CSPGs outlined in Table 2 represent a diverse set of proteins with a wide range of expression patterns. The advent of targeted gene knockout technology in the past decade has made it possible to study the functions of the CSPGs *in vivo*, rather than rely upon localization patterns, cell culture, or *in vitro* assays to suggest biological function. Several of the knockout mice show more subtle defects than one would predict, suggesting functional redundancy between CSPGs expressed in the same tissue (Table 3).

1. Mutations resulting in cartilage defects

Aggrecan is a major component of cartilage. When complexed with HA, the high concentration of negative charge (imparted by both GlcA and sulfated GalNAc) is balanced by counterions, which help to draw water into cartilage, thereby providing the ability to resist compressive forces characteristic of this tissue (reviewed in Iozzo, 1998). Mutations in the aggrecan gene have been identified in mouse, chick, and humans (Table 3). The cartilage matrix deficiency (*cmd*) mouse strain contains a truncated aggrecan protein lacking most of the CS attachment sites (Krueger et al., 1999; Watanabe et al., 1994). *cmd*^{-/-} mice die perinatally as dwarfed animals with shortened limbs and cleft palates (Rittenhouse et al., 1978). Cartilage from these mice contains mostly chondrocytes, with little matrix deposition. The chick *nanomelia* aggrecan mutant shows a very similar phenotype, with shortened limbs and cellularized cartilage (Li et al., 1993). The human aggrecan gene has been mapped and shown to segregate with a familial form of spondyloepiphyseal

dysplasia (Gleghorn et al., 2005). Human subjects also exhibit shortened limbs and trunk, as well as early-onset osteoarthritis. All of these phenotypes are consistent with the role of aggrecan in generating a hydrated, viscoelastic cartilage matrix.

While aggrecan has historically been understood to be the major CSPG in cartilage, other CSPGs are also expressed in cartilage and exhibit similar mutant phenotypes (Table 3). The SLRP family members decorin and biglycan, as well as type IX collagen, are expressed by chondrocytes. These proteins have been shown to associate with cartilage collagen fibrils. Biglycan and type IX collagen knockout mice develop osteoarthritis and osteoporosis, as well as chondrodysplasia phenotypes (Ameys et al., 2002; Hagg et al., 1997; Muragaki et al., 1996; Nakata et al., 1993; Xu et al., 1998). A mutation in the human type IX collagen gene results in multiple epiphyseal dysplasia, a condition characterized by short stature and limbs, as well as painful joints resulting from articular cartilage irregularities (Briggs et al., 1994; Holden et al., 1999; Spayde et al., 2000). Mutational analysis has thus shown that the presence of CSPGs in cartilage of the joints and bones is essential for proper matrix deposition and hydration.

2. Mutations resulting in connective tissue disorders

Decorin and biglycan are the two main CSPGs expressed in connective tissue and interact with collagen in the matrix. Much work to date has concentrated on understanding the role of these PGs in stabilizing collagen fiber assembly, likely by contributing to the highly ordered orientation of the fibrils. In the case of decorin, this action appears to be dictated by the proteoglycan core since removal of the GAG

chains of decorin fails to disrupt fibrillogenesis (Vogel et al., 1987). Decorin knockout mice are viable but show a fragile skin defect (Table 3) (Danielson et al., 1997). High magnification studies of decorin mutants revealed that collagen formation was disrupted in the skin; fibrils varied in diameter and were irregularly spaced. In addition to defects in loose connective tissue, decorin and biglycan knockouts show fibrotic defects in tendon, bone, lung, and heart (Corsi et al., 2002; Fust et al., 2005; Robinson et al., 2005; Weis et al., 2005). These findings demonstrate that CSPGs act not only in cartilage, but in other connective tissues as well to establish the architecture of the tissue.

3. Mutations resulting in central nervous system defects

Chondroitin sulfate has been known to play a role in development, maturation, and remodeling of the central nervous system (CNS) for nearly 15 years. Quite a large number of CSPGs are expressed in components of the CNS, but only recently have they been studied *in vivo*. All members of the Aggrecan family (Table 2) are localized to neural tissue. No reports of the aggrecan *cmd* knockout mouse have yet addressed CNS phenotypes, and the versican knockout is embryonic lethal (Mjaatvedt et al., 1998). Neurocan and brevican knockout mice show brain-specific defects in long-term potentiation (Brakebusch et al., 2002; Zhou et al., 2001). Additionally, brevican mutants show a reduction in formation of perineuronal nets, structures that arise to restrict plasticity of the maturing brain. Brevican is also thought to inhibit neuronal regeneration following CNS injury (Morgenstern et al., 2002). Neurocan mutants show plasticity and remodeling defects, and neurocan is

thought to inhibit neurite outgrowth *in vitro* (Rauch et al., 2001). Several other CSPGs are expressed in the CNS including biglycan, NG2, phosphacan, neuroglycan C, and the testicans. Either phenotypes have not yet been described in the CNS for these mutants, or gene knockouts have not yet been generated. CSPGs have thus been found to play inhibitory or barrier roles in the brain *in vivo*, comparable to what was seen in *in vitro* studies.

4. The role of hybrid Syndecan family members in development

Gene inactivation of the four syndecan hybrid proteoglycans has not lent much insight into the *in vivo* function of these molecules in mouse model systems. Potential functional redundancy among the multiple isoforms may explain the lack of profound phenotypes. Mouse syndecan-2 and -4 will not be discussed since they carry only HS, and not CS. The lack of syndecan-1 in mice inhibits Wnt-1 induced mammary tumor formation (Alexander et al., 2000). These animals are resistant to *Pseudomonas* bacterial infection as well (Park et al., 2001). Syndecan-3 null mice show a dampened response to food deprivation, while overexpression of transgenic syndecan-1 results in obese mice that feed continuously (Reizes et al., 2001). The precise role CS might be playing remains to be determined. In *Xenopus*, inhibition of syndecan-2 by morpholinos results in a striking defect in left-right axis formation in the early embryo (Kramer and Yost, 2002), though it is not known whether this effect is attributable to lack of HS or CS. Studies of *Drosophila* and *C. elegans* syndecan homologs suggest that axonal guidance defects are related more to the loss of HS, rather than CS, due to their interactions with the heparan sulfate-binding Slit

family of guidance molecules (Johnson et al., 2004; Minniti et al., 2004; Rawson et al., 2005; Rhiner et al., 2005; Spring et al., 1994; Steigemann et al., 2004). It has not yet been shown that worm and fly syndecans carry chondroitin (sulfate) chains. In all systems studied to date, the precise role, if any, of syndecan CS is unknown.

Table 3. *In vivo* functions of chondroitin sulfate proteoglycans

Proteoglycan	Organism	Mutant phenotype	Reference
Aggrecan	Mouse Human Chick	Perinatal lethal dwarfism, Craniofacial abnormalities, Cartilage matrix deficiency Spondyloepiphyseal dysplasia (shortening of trunk and limbs, premature osteoarthritis) Chondrodyostrophy	(Watanabe, 1994; Krieger, 1999) (Gleghorn, 2005) (Li, 1993)
Vertecan	Mouse	Defects in heart formation, embryonic lethal	(Mjaatvedt, 1998)
Neurocan	Mouse	Mild defects in synaptic plasticity and long-term potentiation	(Zhou, 2001)
Brevican	Mouse	Reduced perineuronal nets, Defects in long term potentiation	(Brakebusch, 2002)
Decorin	Mouse	Abnormal collagen fibrillogenesis, Skin fragility Fibrotic defects in heart, tendon, lung	(Danielson, 1997) (Weis, 2005; Robinson, 2005; Fust, 2005)
Biglycan	Mouse	Osteoporosis Osteoarthritis Fibrotic defects in tendon, bone, dermis	(Xu, 1998) (Ameje, 2002) (Robinson, 2005; Corsi, 2002)
$\alpha 2(\text{X})$ collagen	Mouse Mouse Human	Multiple epiphyseal dysplasia Osteoarthritis, Mild chondrodysplasia Multiple epiphyseal dysplasia	(Muragaki, 1996) (Nakata, 1993; Hagg, 1997) (Briggs, 1994; Holden, 1999; Spayde, 2000)
Syndecan-1 Syndecan-3 Syndecan	Mouse Mouse <i>Drosophila</i> <i>C. elegans</i>	Reduced susceptibility to Wnt-induced tumorigenesis and bacterial infection Altered feeding behavior Axon guidance defects Axon guidance defects, Egg laying defects	(Alexander, 2000; Park, 2001) (Reizes, 2001) (Stegemann, 2004; Johnson, 2004) (Rhiner, 2005; Mimiti, 2004)
Serglycin	Mouse	Lack of mast cells (maturation defective)	(Abrink, 2004)
NG2 (MCS'P)	Mouse	Reduction in pathological angiogenesis	(Ozerdem, 2004)
CD44	Mouse	Several subtle phenotypes, role of CS unclear	(Gee, 2004)
Thrombospondulin	Mouse	Embryonic consumptive coagulopathy, Juvenile onset thrombosis	(Isermann, 2001a; Isermann, 2001b)

Chondroitin sulfate proteoglycans are expressed in a number of tissues and play distinct roles in several physiological systems in vertebrates. Many of the sites of CSPG action, such as cartilage, tendons, bone, the circulatory system, and the central nervous system with a true brain, are lacking in lower animals such as *C. elegans*, an organism known to express a non-sulfated version of chondroitin (Toyoda et al., 2000; Yamada et al., 1999). It is unknown which proteoglycan cores harbor *C. elegans* chondroitin chains, and it is interesting to speculate about whether differences in body complexity would necessitate use of a different set of chondroitin core proteins. Chapter 4 addresses this question and shows that *C. elegans* does in fact utilize a completely different set of core proteins that are not expressed by mammals.

C. Chondroitin sulfate biosynthesis

Assembly of CS chains is a dynamic process that involves more than ten different enzymes. An overview of vertebrate CS synthesis is provided below to demonstrate the complexity of the system.

1. Nucleotide sugar building blocks

Sugars used in CS biosynthesis include xylose (Xyl), galactose (Gal), GlcA, and GalNAc. These sugars are converted into activated donors to be used in chain synthesis by conjugation with uridine diphosphate (UDP). With the exception of UDP-Xyl, which is derived from UDP-GlcA in the endoplasmic reticulum, all of the other sugar nucleotides are assembled in the cytosol and shuttled into the Golgi

apparatus by nucleotide sugar transporters (Figure 1). UDP-sugar donors are utilized in the Golgi by specific glycosyltransferases to initiate and elongate the chondroitin backbone.

2. Linkage region biosynthesis

CS is synthesized while attached to a core protein. The chains initiate through β -linkage of Xyl to specific serine (Ser) residues of the protein. This activity is accomplished by the xylosyltransferase, XT-1. A second enzyme, XT-2, exists but its function is unknown (Götting et al., 2000; Kuhn et al., 2000). Xyl initiation is followed by addition of two Gal residues by the β 1-4 galactosyltransferase-I (GalT-I, β 4GalTX) and β 1-3 GalT-II (β 3GalT6). The linkage tetrasaccharide is completed by addition of GlcA by β 1-3 glucuronosyltransferase-I (Figure 1). The resultant $\text{GlcA}\beta$ 3 $\text{Gal}\beta$ 3 $\text{Gal}\beta$ 4 $\text{Xyl}\beta$ -O-Ser precursor can undergo additional modifications, including sulfation of the Gal residues and phosphorylation of xylose, which may play a regulatory role in the assembly process (Sugahara and Kitagawa, 2002).

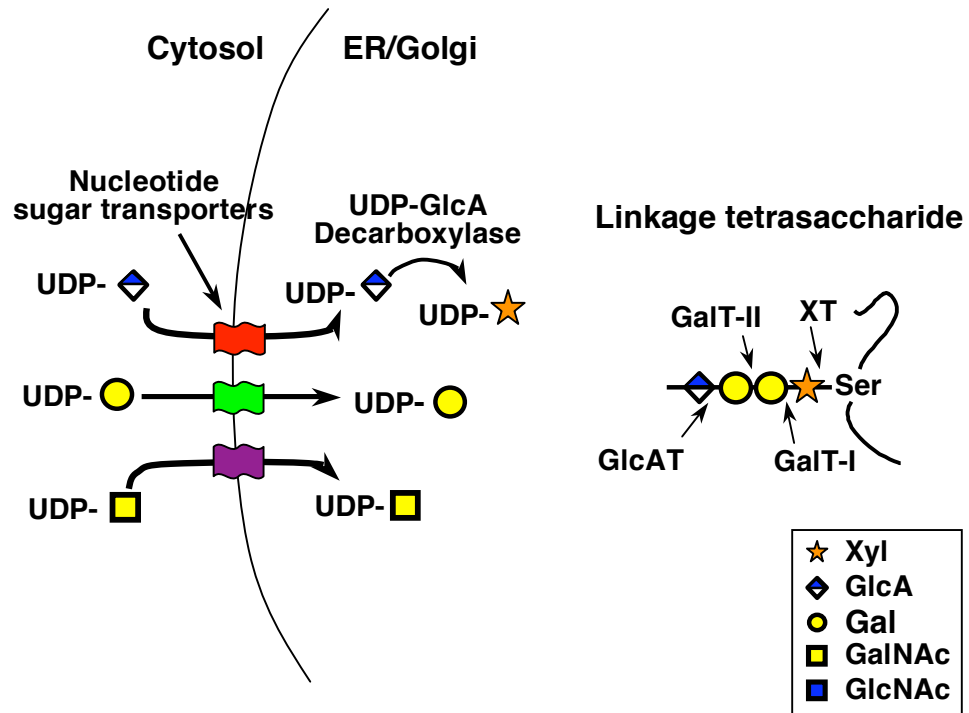


Figure 1. Nucleotide sugar transport and linkage tetrasaccharide synthesis

Activated UDP-sugar nucleotide donors are transported from the cytosol into the endoplasmic reticulum (ER) and Golgi apparatus. Further modification of UDP-sugars is possible in the ER. Separate glycosyltransferases catalyze formation of the linkage tetrasaccharide precursor from the UDP-sugar donors. The boxed legend shows symbols representing various sugar moieties.

3. Polymerization of chondroitin sulfate

The linkage tetrasaccharide is a common precursor that can generate either CS or HS, depending on which sugar is added next (Figure 2). The growing chain is committed to become chondroitin after addition of GalNAc β 1-4 by the initiating activity of GalNAcT-I, accomplished by the human gene GalNAcT-1 (Gotoh et al., 2002; Uyama et al., 2002). A recently identified family of chondroitin sulfate synthases (CSS) polymerize the chondroitin backbone by the stepwise assembly of repeating GalNAc β 4GlcA β 3 disaccharide units (Figure 2). CSS-1, -2, and -3 possess synthase activity based on the ability of each to processively transfer both GalNAc and GlcA onto a synthetic chondroitin acceptor (GalNAcT-II and GlcAT-II activity, respectively) (Kitagawa et al., 2001; Yada et al., 2003a; Yada et al., 2003b). This dual functionality is accomplished by the presence of two catalytically active sites in a single enzyme. In addition to the synthases, CS-GlcAT and CS-GalNAcT-2 contain elongation activity that is much greater than their ability to initiate the CS chain (Gotoh et al., 2002; Uyama et al., 2003). More recently, the identification of a chondroitin polymerizing factor (ChPF) provided evidence that the CSS enzymes require activation for their polymerizing activity (Kitagawa et al., 2003). On average, a chondroitin chain consists of ~40 disaccharides (~20 kDa), but the size varies in different tissues and can reach over 100 disaccharide units (~50 kDa) (Silbert and Sugumaran, 2002).

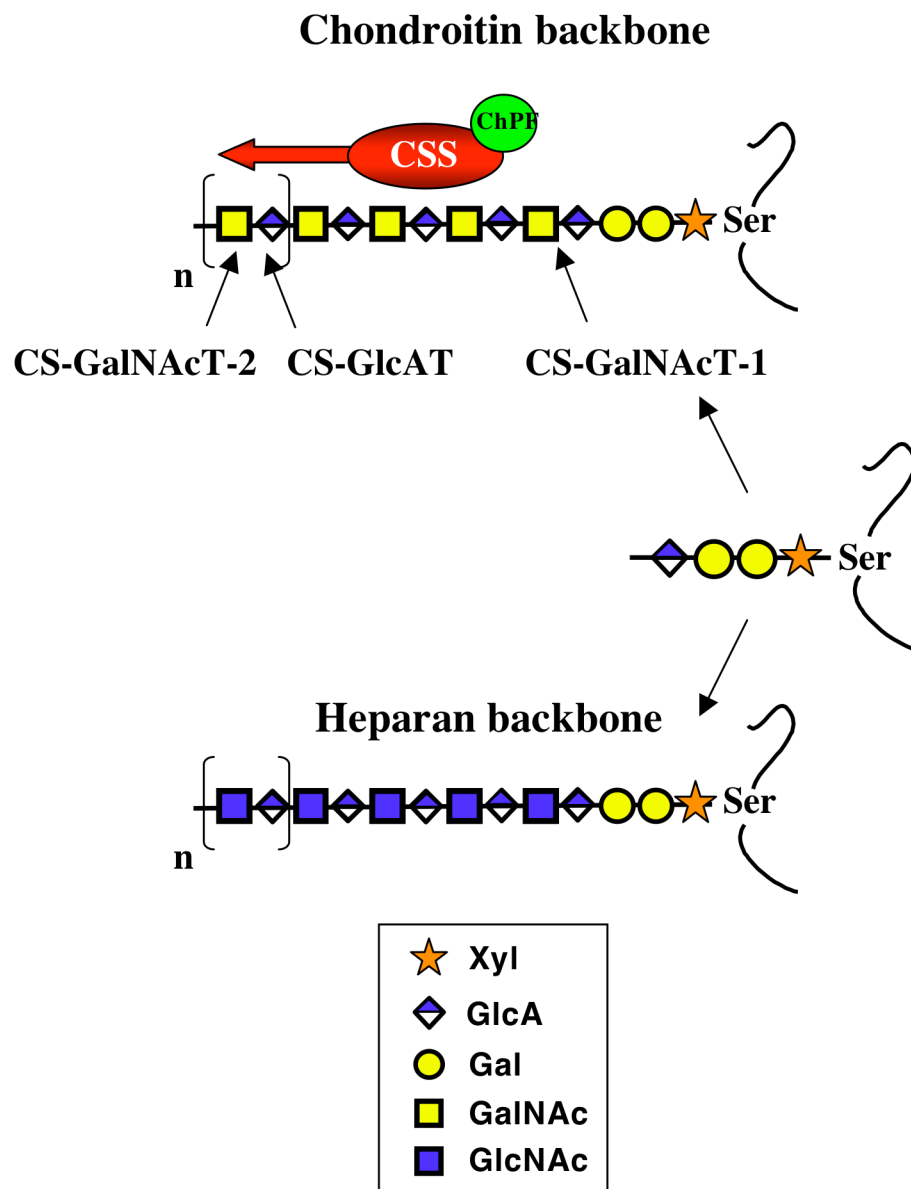


Figure 2. Multiple enzymes polymerize the chondroitin backbone

The linkage tetrasaccharide is a precursor able to generate both chondroitin sulfate and heparan sulfate. Addition of GalNAc by CS-GalNAcT-1 initiates chondroitin biosynthesis. The chondroitin chain is polymerized by the three CSS enzymes, which are activated by ChPF, and add repeating GlcA and GalNAc units. CS-GlcAT and CS-GalNAcT-2 also help synthesize the chondroitin chain.

4. Chondroitin sulfate modification

Following polymerization, the chondroitin backbone can subsequently be modified by sulfation of GalNAc units at the C-4 and/or C-6 positions. These reactions are carried out by distinct sulfotransferases using the phosphoadenosine-5' phosphosulfate (PAPS) sulfate donor as a substrate. In DS, a subset of CS D-GlcA units epimerize to L-IdoA, which can then be sulfated at C-2. Epimerization of GlcA to IdoA occurs by stereochemical inversion at C5 (the proton is axial in D-GlcA but planar in L-IdoA).

Each step of the chondroitin biosynthetic pathway is catalyzed by a specific enzyme. Homologs of these enzymes have been identified in mouse, zebrafish, and other vertebrate genomes. However, it is not known whether an organism such as *C. elegans*, which expresses a simplistic non-sulfated chondroitin variation, assembles the chain in an identical manner. Chapters 2 and 3 show that, indeed, the chondroitin biosynthetic machinery is completely conserved between humans and worms. Enzyme conservation suggests that chondroitin is an ancient molecule that serves a fundamental role in animal biology.

D. *C. elegans* as a model system to study chondroitin proteoglycans

1. Advantages of the *C. elegans* model

C. elegans is an ideal model organism in which to study complex systems in a simplistic manner. Major advantages include, but are not limited

to, a transparent body that enables observation of developmental events, a fully sequenced genome, an invariant cell lineage that has been fully mapped, ease of both forward and reverse genetics, and susceptibility to RNA interference (RNAi). Additionally, families comprised of many members in vertebrate systems often have only a single counterpart in the worm. For example, vertebrate signal transduction systems contain more than 20 fibroblast growth factor (FGF) ligands and four different FGF receptors. It is difficult to study the system by gene inactivation, due to redundancy of related family members. *C. elegans*, on the other hand, expresses only two FGF ligands (EGL-17 and LET-756) and a single receptor (EGL-15). Lack of redundancy enables one to tease out gene function much more easily than in a complicated system.

As described in the previous sections, vertebrate CSPGs are a complex class of molecules. Not only does assembly of the chondroitin backbone require numerous enzymes, some of these with multiple isozymes, but redundancy and overlapping expression patterns are also seen with the core proteins. The study of chondroitin in vertebrate systems might prove problematic due to multiple isoforms performing similar functions. It may be necessary to remove multiple proteins before a true phenotypic effect can be seen. A model system with fewer isoforms would simplify matters greatly. The nematode *C. elegans* is just such a system in which to study the role of chondroitin proteoglycans in development.

A genetic screen designed to identify worms with defects in vulval morphogenesis identified a class of mutants with a “squashed vulva” phenotype (*Sqv*) (Herman et al., 1999). The *sqv* mutants displayed both vulval invagination defects and maternal effect lethality. Analysis of each of the eight identified genes demonstrated that they encode components of the chondroitin biosynthetic pathway. These studies are described in Chapters 1, 2 and 3.

2. Genetic screen for squashed vulva (*sqv*) mutants

a. Vulval invagination

The worm vulva is a passageway between the uterus and the external environment. It serves to accept sperm from males and expel embryos from the body. Vulval development (Figure 3A-E) (reviewed in Greenwald, 1997) begins as early as the first larval stage (L1) with the generation of six ventrally located vulval precursor cells (VPCs P3.p through P8.p) that are competent to give rise to the vulva. During the early L3 larval stage the anchor cell, lying just dorsal to the VPCs, sends a LIN-3/EGF signal that is received by LET-23/EGFR receptors on the VPCs. The cell closest to the AC, P6.p, receives the highest level of signal and becomes a 1^o vulval cell. The P5.p and P7.p cells on each side of the 1^o cell receive slightly lower amounts of signal and become 2^o vulval cells. The remaining three VPCs are not exposed to enough signal and eventually fuse with the hypodermis, not contributing to the vulva. A number of lateral and hypodermal signaling events, regulated by the LIN-12/Notch pathway, also occur to further establish the identity of the vulval lineage.

The 1^o and 2^o vulval cells undergo a number of cell divisions and fusions, eventually resulting in 22 cells that will form the vulva, 11 on each side of a “midline” located near the center of the body (reviewed in Greenwald, 1997). Cells on each side of the midline migrate toward it, eventually contacting their respective partner and fusing to form a multinucleate torroidal ring. The seven torroids invaginate into the body during the L4 stage. The hollow space within the stack of rings is known as the vulval lumen or extracellular space. As the worm progresses to the young adult stage the invaginated vulva everts to form a mature structure that is capable of laying eggs.

The majority of the work in the field of vulval morphogenesis has concentrated on the early signaling pathways that establish VPC specification and identity. Much less is known about the later steps of vulval invagination and eversion. Herman et al. (1999) performed a genetic screen to identify mutants defective in the L4 invagination step. These mutants exhibited a “squashed vulva” (Sqv) phenotype characterized by a reduced lumen where the vulva appears collapsed (Figure 3F). Twenty-five mutants identified in the initial screen fell into eight complementation groups and were named *sqv-1* through *sqv-8*. Strong loss of function or null alleles of all genes showed identical vulval phenotypes, including an egg-laying defect that resulted in accumulation of eggs in the uterus of all homozygous mutant adults, except those of *sqv-5(n3039)*. The *sqv* genes are thus excellent candidates to use as a model to study the more generalized process of epithelial invagination.

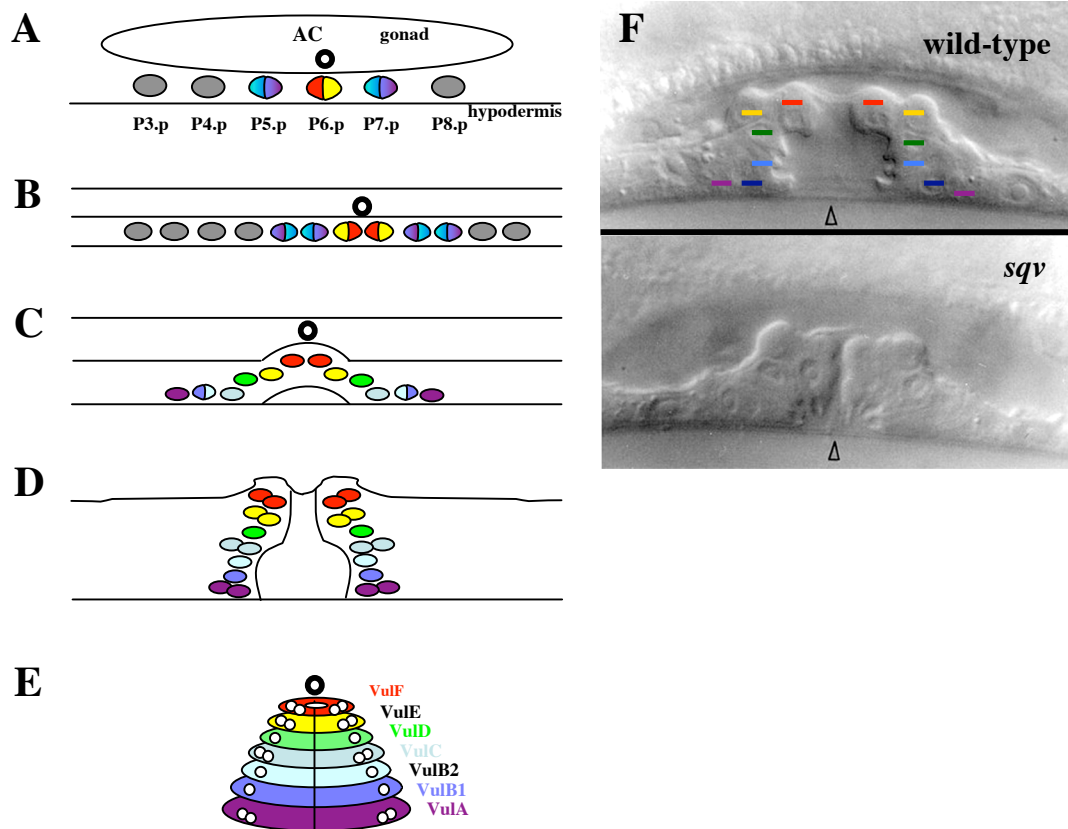


Figure 3. Vulval development and the squashed vulva (*Sqv*) phenotype

Vulval development involves a number of cell signaling, division, migration, and fusion events. Figure adapted from (Greenwald, 1997) and (Sharma-Kishore, 1999). (A) The anchor cell (AC, dark outlined circle) signals from the somatic gonad to the underlying vulval precursor cells, P5.p, P6.p, and P7.p, instructing them to become 1^o (red/yellow) or 2^o (green/blue/purple) vulval cells. The remaining VPCs (grey) eventually fuse with the ventral hypodermis. (B,C) 1^o and 2^o cells divide twice and begin to invaginate. (D,E) All cells, except those that will generate VulB2 and VulD torroids, divide once more, migrate toward the midline, and fuse with their corresponding partners on the other side of the midline. (E) Fusion creates multi-nucleate torroidal rings with a hollow luminal center. Nuclei, small white circles. (F) DIC Nomarski image of L4 staged vulvae from wild-type (top panel) and *sqv* mutant (bottom panel) worms. Outlined arrowheads indicates the vulval lumen, which is reduced in *sqv* mutants. Wild-type vulval cells are color coded according to lineages described in (A-E).

b. Maternal effect lethality

In addition to the vulval phenotype described above, *sqv* mutants also exhibit maternal effect lethality (Herman et al., 1999). Progeny of homozygous null *sqv* mutants arrested as single-celled embryos, suggesting an early defect in cell division. *C. elegans* early embryogenesis has been studied in great detail. Events from fertilization leading up to the first division are well characterized and highly reproducible (Figure 4) (Cowan and Hyman, 2004; Kemphues, 1997). As the proximal oocyte of the gonad passes through the spermatheca it becomes fertilized by sperm stored by the hermaphrodite during the L4 larval stage. Immediately following fertilization, chitin and other extracellular components are secreted to form the eggshell that surrounds the embryo throughout embryogenesis. Also following fertilization, the zygote completes both meiosis I and meiosis II by extruding two polar bodies from the maternal chromatin.

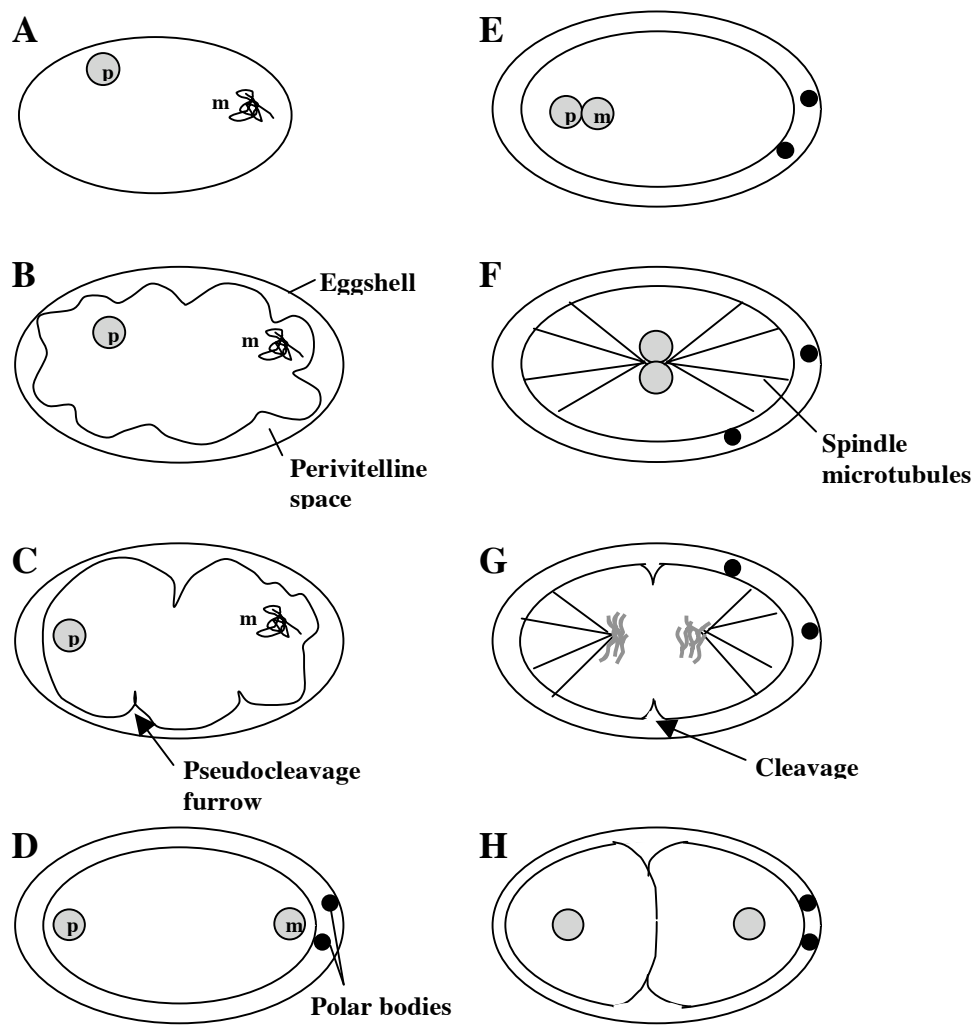
While the paternal centrosome becomes established, the cortex of the embryo undergoes rearrangement of polarity determinants, which initially are localized throughout the embryonic cortex (reviewed in Cowan and Hyman, 2004). One visible landmark of this rearrangement is cortical membrane ruffling, a phenomenon resulting from contraction of the actomyosin network. Studies have shown that ruffling of the membrane is not necessary for establishment of polarity, but is a hallmark of it. Smoothing of the ruffling cortex eventually begins in the posterior and spreads to the anterior of the embryo.

As membrane smoothening progresses, the maternal pronucleus migrates toward the paternal pronucleus. The two fuse and move toward the center of the embryo, where they rotate 90° so the centromeres align with the anterior-posterior long axis of the embryo. Nuclear division occurs, with the two nuclei separating into the anterior and posterior halves. A cleavage furrow forms between the two nuclei, ingresses into the embryo, and the membrane pinches off to form two daughter cells.

Nomarski DIC images of *sqv-7(n3789)* mutants suggest that the maternal effect defect is a result of cytokinesis failure (Hwang and Horvitz, 2002b). The early events of embryogenesis occur normally: *sqv-7(n3789)* embryos are fertilized, the pronuclei meet and fuse, and nuclear division occurs as expected. However, the embryo fails to generate a cleavage furrow and daughter cells never form. Instead, the embryo re-attempts nuclear synthesis and division, resulting in a multinucleated, single-celled dead embryo. It was inferred that the cell division defects seen in the other seven *sqv* mutants is identical in cause to that of *sqv-7(n3789)*.

Figure 4. *C. elegans* early embryogenesis

(A) Embryogenesis begins with fertilization of the oocyte. (B) Soon after fertilization a chitinous eggshell is secreted around the embryo, forming a perivitelline space. The plasma membrane begins to ruffle, indicating that cortical rearrangement is occurring inside the embryo. (C) As the paternal pronucleus (p) becomes established at the posterior cortex, smoothing of the membrane is occurs. Pseudocleavage is observed at this time. (D) As cortex smoothing progresses toward the anterior end of the embryo, two polar bodies (black circles) are extruded from the maternal nuclear material into the perivitelline space. (E) The maternal pronucleus (m) migrates toward the paternal pronucleus (p). (F) The nuclei migrate toward the center of the embryo and rotate 90° to align with the anterior-posterior axis of the embryo. (G) Spindle microtubules separate sister chromatids at anaphase, and the cleavage furrow begins to ingress into the embryo. (H) Cytokinesis is completed, forming two daughter cells. Anterior is to the right.



3. *sqv* genes encode components of the GAG biosynthesis pathway

a. *sqv-3*, *-7*, and *-8* are GAG synthesis genes

In a manuscript accompanying the original identification of the *sqv* mutants, three of the *sqv* genes, *sqv-3*, *sqv-7*, and *sqv-8*, were cloned (Herman and Horvitz, 1999). *sqv-7* showed homology to a *Leishmania* GDP-mannose transporter, suggesting that SQV-7 was a nucleotide sugar transporter that shuttles sugar donors from the cytoplasm, where most are synthesized, into the Golgi apparatus, where they are utilized by glycosyltransferases. *sqv-3* showed homology to vertebrate β 1-4 galactosyltransferases, including identity at the “DXD” catalytic domain, providing evidence that SQV-3 was likely a glycosyltransferase of some variety. *sqv-8* was found to resemble human and rat β 1-3 glucuronosyltransferases and also included the catalytic DXD domain. The authors proposed a model in which SQV-7 transports an unknown nucleotide sugar donor, possibly UDP-Gal or UDP-GlcA, into the Golgi apparatus. It was additionally hypothesized that SQV-3 transfers a Gal unit onto a GlcNAc moiety of a glycoprotein, followed by SQV-8 addition of GlcA onto the recently added Gal. The authors suggested that SQV-3 and SQV-8 ultimately help generate an HNK epitope with the structure GlcA β 3Gal β 4GlcNAc-protein, though the possibility of alternate glycan structures was raised.

The identification of SQV-3 and SQV-8 as putative Gal and GlcA transferases, respectively, raised the possibility that these enzymes could be components of the GAG biosynthetic pathway (Figure 5). A second group showed

that this was in fact the case (Bulik et al., 2000). A combination of *in vitro* and *ex vivo* approaches demonstrated that SQV-3 contained GalT-I enzymatic activity. The ability of a *sqv-3*⁺ transgene to correct a Chinese hamster ovary (CHO) cell line (pgsB-618) defective in GalT-I activity (Esko et al., 1987) demonstrated the functional orthology between *sqv-3* and GAG GalT-I.

Recombinant SQV-8 was shown to have GAG GlcAT-I activity in an *in vitro* assay (Bulik et al., 2000). Additionally, Western blots showed that *sqv-8(mn63)* mutants had dramatically reduced levels of chondroitin proteoglycans compared to wild-type controls. Evidence that SQV-7 is a nucleotide sugar transporter was lacking, but its role in GAG synthesis was confirmed by chemical analysis. *sqv-7(n2839)* mutants had significantly reduced levels of HS disaccharides compared to wild-type controls. If SQV-7 were to function as a transporter, loss of transporting activity in mutants would reduce the pools of nucleotide sugar donors, thereby reducing GAG (and other glycan) content in the animal.

b. SQV-7 is a multi-substrate nucleotide sugar transporter

SQV-7 was clearly identified as a nucleotide sugar transporter by a set of experiments by Berninsone et al. (Berninsone et al., 2001). Golgi-enriched vesicles of yeast cells transfected with *sqv-7*⁺ were tested for the ability to transport a set of eight radiolabeled nucleotide sugars *in vitro*. Of these sugar donors, UDP-GlcA, UDP-Gal, and UDP-GalNAc demonstrated increased transport activity over vector control vesicles. Mutation of the *sqv-7* construct abrogated its ability to transport all three nucleotide sugars. Transport is

competitive, with increasing concentrations of UDP-GlcA or UDP-GalNAc inhibiting transport of UDP-Gal, suggesting that the same active site is used for each UDP-sugar. All three sugars transported by SQV-7 are utilized during GAG biosynthesis (Figure 5). UDP-Gal is a donor for linkage tetrasaccharide formation. UDP-GlcA is a component of both the linkage tetrasaccharide and the repeating disaccharide units of the HS and CS backbone. UDP-GalNAc can be used directly for the polymerization of the chondroitin backbone; alternatively, it can be converted to UDP-GlcNAc by the Gal-4-epimerase and used as a donor in HS backbone formation. The identification of SQV-7 as a multi-substrate transporter was only the second instance this phenomenon had been seen in nature, the first identification being in *Leishmania*. At the time, all evidence in vertebrate systems suggested that sugar transporters were specific for a single UDP-sugar. However, since publication of this observation (Berninsone et al., 2001), multi-substrate nucleotide sugar transporters have been identified in humans, zebrafish, and *Drosophila* (Goto et al., 2001; Selva et al., 2001; Suda et al., 2004).

c. SQV-4 is the UDP-glucose dehydrogenase

The cloned *sqv-4* gene showed high homology to human and *Drosophila* UDP-glucose dehydrogenase (UGDH) enzymes (Hwang and Horvitz, 2002a). UGDH converts UDP-glucose to UDP-GlcA by oxidation of C6-OH to a carboxylate group, forming a hexuronic acid (Figure 5). Cell culture lysates containing recombinant SQV-4 showed 20-fold more UGDH activity than control lysates, and introduction of the *sqv-4(nn2827)* and *sqv-4(2840)* missense mutations

resulted in loss of catalytically active SQV-4. Antibodies generated against SQV-4 showed it to be localized to the cytoplasm of several different cell types, including vulval cells, oocytes, and the uterus. The cytoplasmic localization is typical of many nucleotide sugar modification enzymes; the activated sugar donors are transported (by SQV-7 in *C. elegans*) from their site of synthesis in the cytoplasm to their site of action in the Golgi. One would hypothesize that reduction of UDP-GlcA by mutation of *sqv-4* would impact more glycosylation pathways than just CS and HS.

d. *sqv-1* encodes a UDP-GlcA decarboxylase, and co-localizes with *sqv-7*

sqv-1 was cloned next and found to show generic homology to nucleotide sugar modifying enzymes (Hwang and Horvitz, 2002b). Another nucleotide sugar modification that occurs prior to GAG synthesis is the decarboxylation of UDP-GlcA to form UDP-Xyl, the donor transferred to Ser residues of core proteins to initiate GAG synthesis (Figure 5). HPLC and mass spectrometry analysis demonstrated that recombinant SQV-1 was able to convert UDP-GlcA to UDP-Xyl. A SQV-1 antibody reacted with epitope present in many cell types of the worm, including the vulva, uterus, and oocytes. Comparison to antibody localization of SQV-7 was made, and the two proteins were found to co-localize in punctate staining patterns in the three most proximal oocytes, suggesting Golgi localization.

It is interesting that SQV-1 is localized to Golgi, while the other nucleotide sugar modifying enzyme SQV-4 is present in the cytosol. This suggests that UDP-

GlcA is synthesized in the cytosol and transported into the Golgi, where SQV-1 then converts a subset of it to UDP-Xyl to be used in GAG initiation (Figure 5). Evidence suggests that mammalian UDP-GlcA decarboxylase activity is localized to the endoplasmic reticulum in mammals, which is the site of XT initiation activity (Hoffmann et al., 1984). Closer analysis of the maternal effect lethality phenotype was also presented in this manuscript (Hwang and Horvitz, 2002b). Embryos of *sqv-7(3789)* homozygous mutants not only fail to initiate cytokinesis, as originally reported, but they also show defects in polar body extrusion and separation of the embryonic plasma membrane from the eggshell.

Figure 5. The *sqv* genes encode components of the GAG biosynthetic pathway

The SQV-1 UDP-Glc dehydrogenase and the SQV-4 UDP-GlcA decarboxylase generate sugar nucleotide precursors that are used during GAG assembly. The SQV-7 nucleotide sugar multi-transporter shuttles UDP-GlcA, UDP-Gal, and UDP-GalNAc from their site of synthesis in the cytosol to their site of action in the Golgi. SQV-3 is galactosyltransferases-I, which adds galactose on to xylose. The SQV-8 glucuronosyltransferase also contributes to formation of the linkage tetrasaccharide by adding GlcA.

E. Concluding remarks

The *Sqv* genetic screen identified eight genes that were essential for vulva morphogenesis and early embryonic cell division. Three of these genes, *sqv-1*, *-4*, and *-7*, are involved in synthesis and transport of nucleotide sugar donors used in synthesis of many different classes of glycans, including GAGs. Two others, *sqv-3* and *-8*, were identified as glycosyltransferases that specifically contribute to formation of the GAG linkage tetrasaccharide (Figure 5). At the time I began my dissertation work, it was unknown whether the remaining *sqv* genes were also involved in biosynthesis of chondroitin and heparan sulfate. Chapter 2 shows that this is in fact the case, and that *sqv-2* and *sqv-6* contribute to linkage tetrasaccharide formation. The GAG biosynthetic machinery is conserved between worms and humans. Seven of the eight *sqv* genes had thus been shown to contribute toward the linkage tetrasaccharide, which is competent to generate both CS and HS chains. Chapter 3 reports the identification of the final gene, *sqv-5*, as the worm ortholog of the chondroitin synthase, the enzyme that polymerizes the chondroitin backbone. This finding demonstrates that chondroitin is the essential polymer involved in vulval development and early embryogenesis. Conservation of the chondroitin assembly apparatus in worms suggests that chondroitin is an ancient molecule that has been retained by selective pressure. What is unknown is whether the protein scaffolds are concomitantly conserved. It is conceivable that a non-sulfated chondroitin may have different functions in a simple organism than a fully sulfated version in a more complex organism. Chapter 4 shows this to be the case. *C. elegans* expresses an entirely

different repertoire of chondroitin proteoglycans whose sites of action may not be relevant in mammals. The implications of these findings will be discussed in Chapter 5, as well as future avenues of research.

F. References

Abrink, M., Grujic, M. and Pejler, G. (2004). Serglycin is essential for maturation of mast cell secretory granule. *J Biol Chem* **279**, 40897-905.

Alexander, C. M., Reichsman, F., Hinkes, M. T., Lincecum, J., Becker, K. A., Cumberledge, S. and Bernfield, M. (2000). Syndecan-1 is required for Wnt-1-induced mammary tumorigenesis in mice. *Nat. Genet.* **25**, 329-332.

Ameye, L., Aria, D., Jepsen, K., Oldberg, A., Xu, T. and Young, M. F. (2002). Abnormal collagen fibrils in tendons of biglycan/fibromodulin-deficient mice lead to gait impairment, ectopic ossification, and osteoarthritis. *Faseb J* **16**, 673-80.

Aspberg, A., Adam, S., Kostka, G., Timpl, R. and Heinegard, D. (1999). Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and versican. *J Biol Chem* **274**, 20444-9.

Aspberg, A., Miura, R., Bourdoulous, S., Shimonaka, M., Heinegård, 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.

Bellin, R., Capila, I., Lincecum, J., Park, P. W., Reizes, O. and Bernfield, M. R. (2002). Unlocking the secrets of syndecans: Transgenic organisms as a potential key. *Glycoconj J* **19**, 295-304.

Berninsone, P., Hwang, H. Y., Zemtseva, I., Horvitz, H. R. and Hirschberg, C. B. (2001). SQV-7, a protein involved in *Caenorhabditis elegans* epithelial invagination and early embryogenesis, transports UDP-glucuronic acid, UDP-N-acetylgalactosamine, and UDP-galactose. *Proc Natl Acad Sci U S A* **98**, 3738-3743.

Brakebusch, C., Seidenbecher, C. I., Asztely, F., Rauch, U., Matthies, H., Meyer, H., Krug, M., Bockers, T. M., Zhou, X., Kreutz, M. R. et al. (2002). Brevican-deficient mice display impaired hippocampal CA1 long-term potentiation but show no obvious deficits in learning and memory. *Mol Cell Biol* **22**, 7417-27.

Briggs, M. D., Choi, H., Warman, M. L., Loughlin, J. A., Wordsworth, P., Sykes, B. C., Irlen, C. M., Smith, M., Wynne-Davies, R., Lipson, M. H. et al. (1994). Genetic mapping of a locus for multiple epiphyseal dysplasia (EDM2) to a region of chromosome 1 containing a type IX collagen gene. *Am J Hum Genet* **55**, 678-84.

Bulik, D. A., Wei, G., Toyoda, H., Kinoshita-Toyoda, A., Waldrip, W. R., Esko, J. D., Robbins, P. W. and Selleck, S. B. (2000). *sqv-3* -7, and -8, a set of genes affecting morphogenesis in *Caenorhabditis elegans*, encode enzymes required for glycosaminoglycan biosynthesis. *Proc.Natl.Acad.Sci.USA* **97**, 10838-10843.

Corsi, A., Xu, T., Chen, X. D., Boyde, A., Liang, J., Mankani, M., Sommer, B., Iozzo, R. V., Eichstetter, I., Robey, P. G. et al. (2002). Phenotypic effects of biglycan deficiency are linked to collagen fibril abnormalities, are synergized by decorin deficiency, and mimic Ehlers-Danlos-like changes in bone and other connective tissues. *J Bone Miner Res* **17**, 1180-9.

Cowan, C. R. and Hyman, A. A. (2004). Asymmetric cell division in *C. elegans*: cortical polarity and spindle positioning. *Annu Rev Cell Dev Biol* **20**, 427-53.

Danielson, K. G., Baribault, H., Holmes, D. F., Graham, H., Kadler, K. E. and Iozzo, R. V. (1997). Targeted disruption of decorin leads to abnormal collagen fibril morphology and skin fragility. *J.Cell Biol.* **136**, 729-743.

Esko, J. D. and Selleck, S. B. (2002). ORDER OUT OF CHAOS: Assembly of ligand binding sites in heparan sulfate. *Annu.Rev.Biochem.* **71**, 435-471.

Esko, J. D., Weinke, J. L., Taylor, W. H., Ekborg, G., Rodén, L., Anantharamaiah, G. and Gawish, A. (1987). Inhibition of chondroitin and heparan sulfate biosynthesis in Chinese hamster ovary cell mutants defective in galactosyltransferase I. *J.Biol.Chem.* **262**, 12189-12195.

Fust, A., LeBellego, F., Iozzo, R. V., Roughley, P. J. and Ludwig, M. S. (2005). Alterations in lung mechanics in decorin-deficient mice. *Am J Physiol Lung Cell Mol Physiol* **288**, L159-66.

Gee, K., Kryworuchko, M. and Kumar, A. (2004). Recent advances in the regulation of CD44 expression and its role in inflammation and autoimmune diseases. *Arch Immunol Ther Exp (Warsz)* **52**, 13-26.

Gleghorn, L., Ramesar, R., Beighton, P. and Wallis, G. (2005). A mutation in the variable repeat region of the aggrecan gene (AGC1) causes a form of spondyloepiphyseal dysplasia associated with severe, premature osteoarthritis. *Am J Hum Genet* **77**, 484-90.

Goto, S., Taniguchi, M., Muraoka, M., Toyoda, H., Sado, Y., Kawakita, M. and Hayashi, S. (2001). UDP-sugar transporter implicated in glycosylation and processing of Notch. *Nat. Cell Biol.* **3**, 816-822.

Gotoh, M., Yada, T., Sato, T., Akashima, T., Iwasaki, H., Mochizuki, H., Inaba, N., Togayachi, A., Kudo, T., Watanabe, H. et al. (2002). Molecular cloning and characterization of a novel chondroitin sulfate glucuronyltransferase that transfers glucuronic acid to N-acetylgalactosamine. *J Biol Chem* **277**, 38179-38188.

Götting, C., Kuhn, J., Zahn, R., Brinkmann, T. and Kleesiek, K. (2000). Molecular cloning and expression of human UDP-D-xylose:proteoglycan core protein β -D-xylosyltransferase and its first isoform XT-II. *J.Mol.Biol.* **304**, 517-528.

Greenwald, I. (1997). *Development of the Vulva*. In *C. elegans II*, (ed. B. T. Riddle DL, Meyer BJ, and Priess JR, Eds.), pp. 519-541. New York: Cold Spring Harbor Laboratory Press.

Hagg, R., Hedbom, E., Mollers, U., Aszodi, A., Fassler, R. and Bruckner, P. (1997). Absence of the alpha1(IX) chain leads to a functional knock-out of the entire collagen IX protein in mice. *J Biol Chem* **272**, 20650-4.

Herman, T., Hartwig, E. and Horvitz, H. R. (1999). sqv mutants of *Caenorhabditis elegans* are defective in vulval epithelial invagination. *Proc.Natl.Acad.Sci.U.S.A.* **96**, 968-973.

Herman, T. and Horvitz, H. R. (1999). Three proteins involved in *Caenorhabditis elegans* vulval invagination are similar to components of a glycosylation pathway. *Proc.Natl.Acad.Sci.USA* **96**, 974-979.

Hoffmann, H. P., Schwartz, N. B., Rodén, L. and Prockop, D. J. (1984). Location of xylosyltransferase in the cisternae of the rough endoplasmic reticulum of embryonic cartilage cells. *Connect.Tissue Res.* **12**, 151-163.

Holden, P., Canty, E. G., Mortier, G. R., Zabel, B., Spranger, J., Carr, A., Grant, M. E., Loughlin, J. A. and Briggs, M. D. (1999). Identification of novel pro-alpha2(IX) collagen gene mutations in two families with distinctive oligo-epiphyseal forms of multiple epiphyseal dysplasia. *Am J Hum Genet* **65**, 31-8.

Hwang, H. Y. and Horvitz, H. R. (2002a). The *Caenorhabditis elegans* vulval morphogenesis gene *sqv-4* encodes a UDP-glucose dehydrogenase that is temporally and spatially regulated. *Proc Natl Acad Sci U S A* **99**, 14224-9.

Hwang, H. Y. and Horvitz, H. R. (2002b). The SQV-1 UDP-glucuronic acid decarboxylase and the SQV-7 transporter may act in the Golgi apparatus to affect *Caenorhabditis elegans* vulval morphogenesis and embryonic development. *Proc Natl Acad Sci U S A* **99**, 14218-23.

Iozzo, R. V. (1998). Matrix proteoglycans: From molecular design to cellular function. *Annu.Rev.Biochem.* **67**, 609-652.

Iozzo, R. V. (1999). The biology of the small leucine-rich proteoglycans - Functional network of interactive proteins. *J.Biol.Chem.* **274**, 18843-18846.

Isermann, B., Hendrickson, S. B., Hutley, K., Wing, M. and Weiler, H. (2001a). Tissue-restricted expression of thrombomodulin in the placenta rescues thrombomodulin-deficient mice from early lethality and reveals a secondary developmental block. *Development* **128**, 827-38.

Isermann, B., Hendrickson, S. B., Zogg, M., Wing, M., Cumiskey, M., Kisanuki, Y. Y., Yanagisawa, M. and Weiler, H. (2001b). Endothelium-specific loss of murine

thrombomodulin disrupts the protein C anticoagulant pathway and causes juvenile-onset thrombosis. *J Clin Invest* **108**, 537-546.

Johnson, K. G., Ghose, A., Epstein, E., Lincecum, J., O'Connor, M. B. and Van Vactor, D. (2004). Axonal heparan sulfate proteoglycans regulate the distribution and efficiency of the repellent slit during midline axon guidance. *Curr Biol* **14**, 499-504.

Kemphues, K. a. S. S. (1997). Fertilization and Establishment of Polarity in the Embryo. In *C. elegans II*, (ed. B. T. Riddle DL, Meyer BJ, and Priess JR, Eds.). New York: Cold Spring Harbor Laboratory Press.

Kitagawa, H., Izumikawa, T., Uyama, T. and Sugahara, K. (2003). Molecular cloning of a chondroitin polymerizing factor that cooperates with chondroitin synthase for chondroitin polymerization. *J Biol Chem* **278**, 23666-71.

Kitagawa, H., Uyama, T. and Sugahara, K. (2001). Molecular cloning and expression of a human chondroitin synthase. *J.Biol.Chem.* **276**, 38721-38726.

Knudson, C. B. and Knudson, W. (2001). Cartilage proteoglycans. *Semin Cell Dev Biol* **12**, 69-78.

Kramer, K. L. and Yost, H. J. (2002). Ectodermal syndecan-2 mediates left-right axis formation in migrating mesoderm as a cell-nonautonomous Vgl cofactor. *Dev Cell* **2**, 115-24.

Krueger, R. C., Jr., Kurima, K. and Schwartz, N. B. (1999). Completion of the mouse aggrecan gene structure and identification of the defect in the cmd-Bc mouse as a near complete deletion of the murine aggrecan gene. *Mamm Genome* **10**, 1119-25.

Kuhn, J., Götting, C., Schnolzer, M., Kempf, T., Brinkmann, T. and Kleesiek, K. (2000). First isolation of human UDP-D-Xylose: proteoglycan core protein β -D-xylosyltransferase secreted from cultured JAR choriocarcinoma cells. *J Biol Chem* **276**, 4940-4947.

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.

Minniti, A. N., Labarca, M., Hurtado, C. and Brandan, E. (2004). Caenorhabditis elegans syndecan (SDN-1) is required for normal egg laying and associates with the nervous system and the vulva. *J Cell Sci* **117**, 5179-90.

Miura, R., Aspberg, A., Ethell, I. M., Hagihara, K., Schnaar, R. L., Ruoslahti, E. and Yamaguchi, Y. (1999). The proteoglycan lectin domain binds sulfated cell surface glycolipids and promotes cell adhesion. *J.Biol.Chem.* **274**, 11431-11438.

Mjaatvedt, C. H., Yamamura, H., Capehart, A. A., Turner, D. and Markwald, R. R. (1998). The Cspg2 gene, disrupted in the hdf mutant, is required for right cardiac chamber and endocardial cushion formation. *Dev Biol* **202**, 56-66.

Morgenstern, D. A., Asher, R. A. and Fawcett, J. W. (2002). Chondroitin sulphate proteoglycans in the CNS injury response. *Prog Brain Res* **137**, 313-32.

Muragaki, Y., Mariman, E. C., van Beersum, S. E., Perala, M., van Mourik, J. B., Warman, M. L., Hamel, B. C. and Olsen, B. R. (1996). A mutation in COL9A2 causes multiple epiphyseal dysplasia (EDM2). *Ann N Y Acad Sci* **785**, 303-6.

Nakata, K., Ono, K., Miyazaki, J., Olsen, B. R., Muragaki, Y., Adachi, E., Yamamura, K. and Kimura, T. (1993). Osteoarthritis associated with mild chondrodysplasia in transgenic mice expressing alpha 1(IX) collagen chains with a central deletion. *Proc Natl Acad Sci U S A* **90**, 2870-4.

Ozerdem, U. and Stallcup, W. B. (2004). Pathological angiogenesis is reduced by targeting pericytes via the NG2 proteoglycan. *Angiogenesis* **7**, 269-76.

Park, P. W., Pier, G. B., Hinkes, M. T. and Bernfield, M. (2001). Exploitation of syndecan-1 shedding by Pseudomonas aeruginosa enhances virulence. *Nature* **411**, 98-102.

Rauch, U., Clement, A., Retzler, C., Frohlich, L., Fassler, R., Gohring, W. and Faissner, A. (1997). Mapping of a defined neurocan binding site to distinct domains of tenascin-C. *J Biol Chem* **272**, 26905-12.

Rauch, U., Feng, K. and Zhou, X. H. (2001). Neurocan: a brain chondroitin sulfate proteoglycan. *Cell.Mol.Life Sci.* **58**, 1842-1856.

Rawson, J. M., Dimitroff, B., Johnson, K. G., Rawson, J. M., Ge, X., Van Vactor, D. and Selleck, S. B. (2005). The heparan sulfate proteoglycans Dally-like and Syndecan have distinct functions in axon guidance and visual-system assembly in *Drosophila*. *Curr Biol* **15**, 833-8.

Reizes, O., Lincecum, J., Wang, Z., Goldberger, O., Huang, L., Kaksonen, M., Ahima, R., Hinkes, M. T., Barsh, G. S., Rauvala, H. et al. (2001). Transgenic expression of syndecan-1 uncovers a physiological control of feeding behavior by syndecan-3. *Cell* **106**, 105-116.

Rhiner, C., Gysi, S., Frohli, E., Hengartner, M. O. and Hajnal, A. (2005). Syndecan regulates cell migration and axon guidance in *C. elegans*. *Development* **132**, 4621-33.

Rittenhouse, E., Dunn, L. C., Cookingham, J., Calo, C., Spiegelman, M., Doohar, 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.

Robinson, P. S., Huang, T. F., Kazam, E., Iozzo, R. V., Birk, D. E. and Soslowky, L. J. (2005). Influence of decorin and biglycan on mechanical properties of multiple tendons in knockout mice. *J Biomech Eng* **127**, 181-5.

Selva, E. M., Hong, K., Baeg, G. H., Beverley, S. M., Turco, S. J., Perrimon, N. and Häcker, U. (2001). Dual role of the *fringe* connection gene in both heparan sulphate and *fringe*-dependent signalling events. *Nat. Cell Biol.* **3**, 809-815.

Sharma-Kishore, R., White, J.G., Southgate, E., and Podbilewicz, B. (1999). Formation of the vulva in *C. elegans*: a paradigm for organogenesis. *Development* **126**, 691-699.

Silbert, J. E. and Sugumaran, G. (2002). Biosynthesis of chondroitin/dermatan sulfate. *IUBMB Life* **54**, 177-186.

Spayde, E. C., Joshi, A. P., Wilcox, W. R., Briggs, M., Cohn, D. H. and Olsen, B. R. (2000). Exon skipping mutation in the COL9A2 gene in a family with multiple epiphyseal dysplasia. *Matrix Biol* **19**, 121-8.

Spring, J., Paine-Saunders, S. E., Hynes, R. O. and Bernfield, M. (1994). *Drosophila* syndecan: Conservation of a cell-surface heparan sulfate proteoglycan. *Proc.Natl.Acad.Sci.USA* **91**, 3334-3338.

Steigemann, P., Molitor, A., Fellert, S., Jackle, H. and Vorbruggen, G. (2004). Heparan sulfate proteoglycan syndecan promotes axonal and myotube guidance by slit/robo signaling. *Curr Biol* **14**, 225-30.

Suda, T., Kamiyama, S., Suzuki, M., Kikuchi, N., Nakayama, K., Narimatsu, H., Jigami, Y., Aoki, T. and Nishihara, S. (2004). Molecular cloning and characterization of a human multisubstrate specific nucleotide-sugar transporter homologous to *Drosophila* fringe connection. *J Biol Chem* **279**, 26469-74.

Sugahara, K. and Kitagawa, H. (2002). Heparin and heparan sulfate biosynthesis. *IUBMB Life* **54**, 163-175.

Toyoda, H., Kinoshita-Toyoda, A. and Selleck, S. B. (2000). Structural analysis of glycosaminoglycans in *Drosophila* and *Caenorhabditis elegans* and demonstration that *tout-velu*, a *Drosophila* gene related to EXT tumor suppressors, affects heparan sulfate *in vivo*. *J.Biol.Chem.* **275**, 2269-2275.

Uyama, T., Kitagawa, H., Tamura, J. and Sugahara, K. (2002). Molecular cloning and expression of human chondroitin *N*-acetylgalactosaminyltransferase - The key enzyme for chain initiation and elongation of chondroitin/dermatan sulfate on the protein linkage region tetrasaccharide shared by heparin/heparan sulfate. *J.Biol.Chem.* **277**, 8841-8846.

Uyama, T., Kitagawa, H., Tanaka, J., Tamura, J., Ogawa, T. and Sugahara, K. (2003). Molecular cloning and expression of a second chondroitin *N*-

acetylgalactosaminyltransferase involved in the initiation and elongation of chondroitin/dermatan sulfate. *J Biol Chem* **278**, 3072-8.

Vogel, K. G., Koob, T. J. and Fisher, L. W. (1987). Characterization and interactions of a fragment of the core protein of the small proteoglycan (PGII) from bovine tendon. *Biochem Biophys Res Commun* **148**, 658-63.

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**, 154-7.

Weis, S. M., Zimmerman, S. D., Shah, M., Covell, J. W., Omens, J. H., Ross, J., Jr., Dalton, N., Jones, Y., Reed, C. C., Iozzo, R. V. et al. (2005). A role for decorin in the remodeling of myocardial infarction. *Matrix Biol* **24**, 313-24.

Xu, T., Bianco, P., Fisher, L. W., Longenecker, G., Smith, E., Goldstein, S., Bonadio, J., Boskey, A., Heegaard, A. M., Sommer, B. et al. (1998). Targeted disruption of the biglycan gene leads to an osteoporosis-like phenotype in mice. *Nat Genet* **20**, 78-82.

Yada, T., Gotoh, M., Sato, T., Shionyu, M., Go, M., Kaseyama, H., Iwasaki, H., Kikuchi, N., Kwon, Y. D., Togayachi, A. et al. (2003a). Chondroitin sulfate synthase-2. Molecular cloning and characterization of a novel human glycosyltransferase homologous to chondroitin sulfate glucuronyltransferase, which has dual enzymatic activities. *J Biol Chem* **278**, 30235-47.

Yada, T., Sato, T., Kaseyama, H., Gotoh, M., Iwasaki, H., Kikuchi, N., Kwon, Y. D., Togayachi, A., Kudo, T., Watanabe, H. et al. (2003b). Chondroitin sulfate synthase-3. Molecular cloning and characterization. *J Biol Chem* **278**, 39711-25.

Yamada, S., Van Die, I., Van den Eijnden, D. H., Yokota, A., Kitagawa, H. and Sugahara, K. (1999). Demonstration of glycosaminoglycans in *Caenorhabditis elegans*. *FEBS Lett.* **459**, 327-331.

Zhang, L. and Esko, J. D. (1994). Amino acid determinants that drive heparan sulfate assembly in a proteoglycan. *J.Biol.Chem.* **269**, 19295-19299.

Zhou, X. H., Brakebusch, C., Matthies, H., Oohashi, T., Hirsch, E., Moser, M., Krug, M., Seidenbecher, C. I., Boeckers, T. M., Rauch, U. et al. (2001). Neurocan is dispensable for brain development. *Mol Cell Biol* **21**, 5970-8.

Chapter 1, in part, includes material as it appeared in the Encyclopedia of Biological Chemistry, Olson, S.K. and Esko, J.D. (2004). The dissertation author was the primary author and the co-author listed in this publication directed and supervised the writing which forms the basis for this chapter.

CHAPTER 2

The *C. elegans* genes *sqv-2* and *sqv-6*, which are required for vulval morphogenesis, encode glycosaminoglycan galactosyltransferase II and xylosyltransferase

A. Summary

In mutants defective in any of eight *C. elegans sqv* (squashed vulva) genes, the vulval extracellular space fails to expand during vulval morphogenesis. Strong *sqv* mutations result in maternal-effect lethality, caused in part by the failure of the progeny of homozygous mutants to initiate cytokinesis and associated with the failure to form an extracellular space between the egg and the eggshell. Recent studies have implicated glycosaminoglycans in these processes. Here we report the cloning and characterization of *sqv-2* and *sqv-6*. *sqv-6* encodes a protein similar to human xylosyltransferases. Transfection of *sqv-6* rescued the glycosaminoglycan biosynthesis defect of a xylosyltransferase mutant hamster cell line. *sqv-2* encodes a protein similar to human galactosyltransferase-II. A recombinant SQV-2 fusion protein had galactosyltransferase II activity with substrate specificity similar to that of human galactosyltransferase-II. We conclude that *C. elegans* SQV-6 and SQV-2 likely act in concert with other SQV proteins to catalyze the step-wise formation of the

proteoglycan core protein linkage tetrasaccharide GlcA β 1,3Gal β 1,3Gal β 1,4Xyl β -O-(Ser), which is common to the two major types of glycosaminoglycans in vertebrates, chondroitin and heparan sulfate. Our results strongly support a model in which *C. elegans* vulval morphogenesis and zygotic cytokinesis depend on the expression of glycosaminoglycans.

B. Introduction

Glycosaminoglycans (GAGs) are important in animal development, and defects in GAGs are responsible for certain human disorders. For example, mutations in the *Drosophila melanogaster* genes *tout velu* (Bellaiche et al., 1998) and *sulfateless* (Lin and Perrimon, 1999), which encode homologs of heparan sulfate copolymerase and heparan sulfate N-deacetylase/N-sulfotransferase, respectively, cause zygotic lethality and defects in segmentation. Mutations in the mouse *tout-velu* homolog EXT1 disrupt gastrulation and the generation of mesoderm (Lin et al., 2000), while mutations in human EXT1 and EXT2 have been associated with Hereditary Multiple Exostoses (reviewed in Zak et al., 2002). Mutations in the human galactosyltransferase I have been associated with a progeroid variant of the connective-tissue disorder Ehlers-Danlos Syndrome (EDS) (Quentin et al, 1990; Almeida et al., 1999; Okajima et al., 1999a). EDS is a group of heritable disorders characterized by hyperelasticity of the skin and hypermobile joints. Tout-velu, EXT-1, EXT-2 and Sulfateless affect the biosynthesis of heparan sulfate specifically, while

galactosyltransferase-I deficiency affects the biosynthesis of both chondroitin and heparan sulfate.

The backbones of chondroitin and heparan sulfate consist of repeating disaccharide units: GlcA β 1,3GalNAc β 1,4 for chondroitin and GlcA β 1,4GlcNAc α 1,4 for heparan sulfate (reviewed in Esko and Selleck, 2002). Their polymerization occurs on a tetrasaccharide primer (GlcA β 1,3Gal β 1,3Gal β 1,4Xyl β -) that is linked to the protein core of a proteoglycan. The addition of these four sugars is catalyzed stepwise in the lumen of the Golgi apparatus and requires three nucleotide sugars, UDP-Xyl, UDP-Gal and UDP-GlcA, and four glycosyltransferases.

Eight *sqv* (squashed vulva) genes were genetically identified in a screen for *C. elegans* mutants defective in vulval morphogenesis (Herman et al., 1999). All *sqv* mutants fail to form a large fluid-filled vulval extracellular space and have a reduced separation of the anterior and posterior halves of the vulva from the early to middle phases of L4 larval development. Strong mutant alleles of all eight *sqv* genes also cause maternal-effect lethality. Most progeny of mothers homozygous for a strong *sqv* mutant allele arrest at the one-cell stage (Herman et al., 1999). The nuclei of the arrested progeny divide normally, but the extrusion of the polar bodies and the initiation of cytokinesis are impaired (Hwang and Horvitz, 2002a). These mutant eggs fail to form the normal fluid-filled extracellular space between the membrane of the egg and the eggshell. We have postulated that the *sqv* genes control the biosynthesis of GAGs that are secreted and become hydrated to form fluid-filled extracellular spaces (Hwang and Horvitz, 2002a; Herman and Horvitz, 1999).

The molecular identification of five *sqv* genes has led to a model implicating the biosynthesis of chondroitin and/or heparan sulfate in *C. elegans* development. *sqv-1*, *-3*, *-4*, *-7* and *-8* encode UDP-GlcA decarboxylase (Hwang and Horvitz, 2002a), galactosyltransferase I (Bulik et al., 2000), UDP-glucose dehydrogenase (Hwang and Horvitz, 2002b), UDP-GlcA/UDP-Gal/UDP-GalNAc transporter (Berninsone et al., 2001), and glucuronosyltransferase I (Bulik et al., 2000), respectively. *sqv-3* was used to identify the human galactosyltransferase I, which has been implicated in the progeroid variant of EDS (Okajima et al., 1999a; Okajima et al., 1999b). In this paper, we show that *sqv-6* encodes the xylosyltransferase that adds Xyl to the protein core, thus initiating GAG biosynthesis. *sqv-2* encodes a galactosyltransferase that adds the second Gal residue to the linkage tetrasaccharide.

C. Experimental Procedures

1. *C. elegans* maintenance

Strains were cultured as described (Brenner, 1974) and were grown at 20 to 22°C unless indicated otherwise.

2. Molecular biology

Standard molecular biology techniques were used (Sambrook et al., 1989). The sequences of all PCR-amplified DNAs used for cloning were confirmed to exclude unintended mutations. Oligonucleotide sequences used for amplification or mutagenesis of DNA are shown in Supplementary Experimental Procedures.

3. Rescue of *C. elegans sqv-2* and *sqv-6* mutants

For germline rescue, we injected cosmids carrying genomic DNA into *sqv-2(n2821)* and *sqv-6(n2845) unc-60(e677)/unc-34(s138)* animals with the dominant roller marker pRF4, as described by (Mello et al., 1991). Rol lines were established, and Rol animals and Unc-60 Rol animals were examined for rescue of the *sqv-2* and the *sqv-6* mutant phenotype, respectively. We injected *sqv-2(n2821)* hermaphrodites with plasmids containing the *sqv-2* ORF under the control of the *C. elegans* heat-shock promoters (Stringham et al., 1992) and pRF4 as the coinjection marker. We injected *sqv-6(n2845)/nT1(n754)* hermaphrodites with plasmids containing the *sqv-6* ORF under the control of the *C. elegans* heat-shock promoters (Stringham et al., 1992) and pRF4. Rol lines were established, and Rol (non-Unc) animals were examined for rescue of the *sqv-2* and *sqv-6* mutant phenotype following induction of *sqv-2* and *sqv-6* expression by 30 minutes of heat-shock treatment at 33°C.

4. SQV-2 galactosyltransferase II assay

A sequence encoding amino acids 25 to 330 of SQV-2, thus lacking the presumptive transmembrane domain at the amino terminus, was cloned into pDEST-CMV-protA. This plasmid was designed to express a secreted fusion protein containing protein A and SQV-2 amino acids 25-330. COS-7 cells were transiently transfected with pDEST-CMV-protA-*sqv-2* using LipofectAMINE (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. After 72 hours of incubation, the fusion protein was recovered from the cell culture supernatant by

affinity chromatography using IgG-agarose (Wei et al., 1993). Galactosyltransferase II activity was assayed as described by Bai et al. (2001).

5. Rescue of the xylosyltransferase defect in CHO pgsA-745 cells by *sqv-6*

The xylosyltransferase-deficient CHO pgsA-745 cells (Esko et al., 1985) were transfected with *sqv-6* ORF, which was cloned into pcDNA3.1. Stable transfectants were selected with 400 $\mu\text{g/ml}$ geneticin (Invitrogen). Several drug-resistant colonies were isolated and screened by flow cytometry for *sqv-6* expression based on binding of biotinylated FGF-2 as described (Bai et al., 1999). Incorporation of $^{35}\text{SO}_4$ into GAG chains of wild-type CHO or pgsA-745 cells with or without *sqv-6* was assayed essentially as described by Bame and Esko (1989), labeling cells overnight at 30°C with 50 $\mu\text{Ci/ml}$ $^{35}\text{SO}_4$ (NEN, Boston, MA).

6. SQV-6 xylosyltransferase assay

Cell extracts of wild-type CHO, pgsA-745, and *sqv-6* or empty vector stable transfectants of pgsA-745 were prepared as described (Esko et al., 1985). Xylosyltransferase activity was assayed essentially as described (Esko et al., 1985), by incubating 25 μg crude cell extract with 50 μg soluble silk acceptor and 6×10^5 cpm UDP-[1- ^3H]xylose (NEN, 8.9 Ci/mmol) at 26°C for 5 hours.

D. Results and Discussion

1. Molecular identification of *sqv-2*.

sqv-2 was previously mapped to the left of *lin-31* on LGII (Herman and Horvitz, 1997). We further mapped *sqv-2* to an interval between *sup-9* and *lin-31*. We assayed 27 cosmids in this interval for the ability to rescue the *sqv-2* mutant phenotype, but none rescued (Fig. 1A).

We examined the DNA sequence corresponding to the gaps between the cosmids in this interval and found a predicted gene, Y110A2AL.14, that is weakly similar to galactosyltransferases. Because all previously cloned *sqv* genes to date are implicated in the biosynthesis of chondroitin and/or heparan sulfate, we suspected that *sqv-2* also encodes a protein involved in GAG biosynthesis. Specifically, it seemed plausible that Y110A2AL.14 encodes the galactosyltransferase II involved in the formation of the protein core linkage tetrasaccharide and that had not been identified molecularly in any organism at the time.

We identified three molecular lesions corresponding to three of the four identified alleles of *sqv-2* in the open reading frame (ORF) of Y110A2AL.14 (Fig. 1B). The two stronger alleles of *sqv-2*, *n3037* and *n3038*, cause a maternal-effect lethal phenotype and are an opal nonsense mutation at arginine 225 and a methionine-to-isoleucine missense mutation of the predicted start codon, respectively. A weak allele, *n2826*, that results in live progeny is a missense mutation causing a glycine-to-arginine substitution at amino acid position 99. The molecular lesion of the weakest allele, *n2840*, has not been identified yet.

We determined the sequences of two cDNA clones, yk94e4 and yk292g2, that correspond to Y110A2AL.14. The yk292g2 clone contains 990 bases of ORF, 17 bases of 5' untranslated region (UTR), and 121 bases of 3' UTR. The 5' end contains three bases that correspond to the sequence of 5' SL1 *trans*-spliced leader, which is found at the 5' end of many *C. elegans* transcripts (Krause and Hirsh, 1987). The 3' end contains a poly-A sequence. The longest ORF in this cDNA is identical to Y110A2AL.14 and is predicted to encode a protein of 330 amino acids. The yk94e4 clone lacks the 5' end of Y110A2AL.14. Expression of the longest ORF in yk292g2 under the control of the *C. elegans* heat-shock promoters (Stringham et al., 1992) rescued the defect in *sqv-2* vulval morphogenesis in all five isolated lines.

2. *sqv-2* encodes a protein similar to galactosyltransferase II.

The predicted SQV-2 protein contains a putative transmembrane domain near the amino terminus, suggesting it may be a type II transmembrane protein (Fig. 1B). All known glycosyltransferases that act in the lumen of the ER and the Golgi apparatus are type II transmembrane proteins. Of 330 amino acids of SQV-2, 93 (28%) and 133 (40%) are identical to the *Drosophila* and human homologs, respectively (Fig. 1B). Recently, the human homolog of SQV-2 was identified as the galactosyltransferase II (Bai et al., 2001).

3. SQV-2 has galactosyltransferase II activity.

We assayed recombinant Protein A-SQV-2 fusion protein expressed in COS-7 cells for galactosyltransferase II activity (see Experimental Procedures). The SQV-2

fusion protein specifically catalyzed the addition of galactose to a disaccharide acceptor, Gal β 1,4Xyl β -O-benzyl, that had been used to demonstrate the acceptor substrate specificity of the human galactosyltransferase II (Bai et al., 2001) (Table 1).

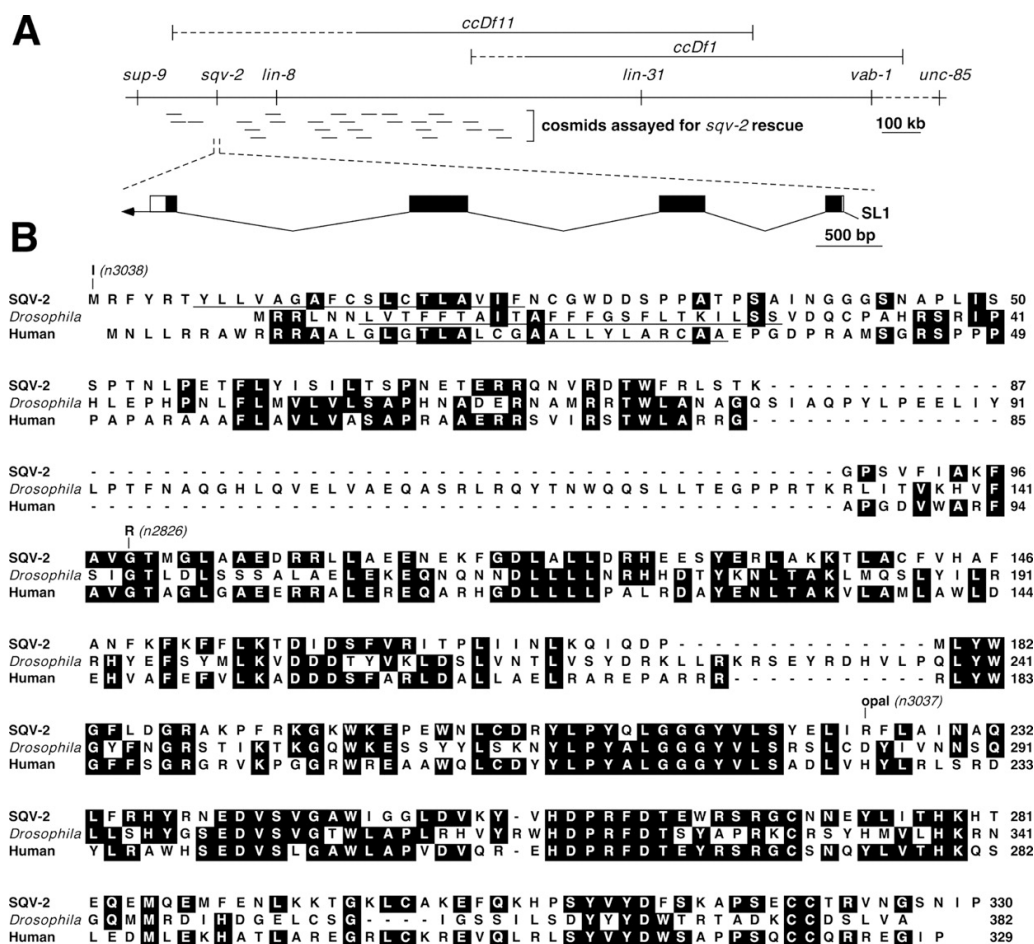


Figure 1. SQV-2 is similar to galactosyltransferase II.

(A) Genetic and physical maps showing *sqv-2*. The dashed horizontal lines depicting *ccDf11* and *ccDf1* indicate the possible extents of the left end points of these deletions. Short solid lines represent cosmid clones that were assayed in germline transformation experiments. Below is the structure of the *sqv-2* gene as deduced from genomic and cDNA sequences. Solid boxes indicate exons, and open boxes indicate untranslated sequences. The *trans*-spliced leader SL1 is indicated, and the arrow indicates the poly(A) tail. (B) Alignment of SQV-2, the *Drosophila* homolog, and human galactosyltransferase II. Identities between at least two proteins are shaded in black. The predicted transmembrane domains are underlined. The three *sqv-2* mutant alleles are indicated. The numbers on the right indicate amino acid positions.

Table 1. The SQV-2 fusion protein has acceptor specificity consistent with its being galactosyltransferase II

Galactosyltransferase activity was assayed *in vitro* using UDP- ^3H galactose together with various acceptor substrates. No substrate controls ranged from 339 to 357 cpm. The range for all substrates shown as "0" activity was 137-619 cpm. The range for the substrate Gal1,4Xyl-*O*-Bn was 141,000-142,000 cpm. Bn, benzyl; NM, naphthalenemethanol; C₁₀, *O*-decenyl (CH₂)₈CHCH₂

Acceptor substrates	Enzyme activity
Monosaccharides (5 mM)	<i>pmol/h/ml medium</i>
Xyl β - <i>O</i> -Bn	0
Xyl β - <i>O</i> -naphthol	0
Gal β - <i>O</i> -NM	0
GalNAc α - <i>O</i> -Bn	0
GlcNAc β - <i>O</i> -NM	0
Disaccharides (5 mM)	
Gal β 1,4Xyl β - <i>O</i> -Bn	2660
Gal β 1,3GalNAc α - <i>O</i> -NM	1
Gal β 1,3Gal β - <i>O</i> -NM	3
Gal β 1,4GlcNAc β - <i>O</i> -NM	0
Gal β 1,3GlcNAc β - <i>O</i> -NM	0
GlcNAc β 1,3Gal β - <i>O</i> -NM	6
Man α 1,6Man α - <i>O</i> -C ₁₀	0

4. Molecular identification of *sqv-6*

sqv-6 was previously mapped to the left of the polymorphism *stP3* on LGV (Herman et al., 1999). We further mapped *sqv-6* to the left of the cosmid W07B8 and within about 0.2 map units of *unc-34*. We assayed 11 cosmids to the right of *unc-34* for the ability to rescue the *sqv-6* mutant phenotype, but none rescued (Fig. 2A).

We examined the DNA sequences in the gaps in the cosmid coverage near the cosmid W07B8 and *unc-34* and found a gene, Y50D4C.d, that is similar to two recently identified human xylosyltransferases (Gotting et al., 2000). In the only allele of *sqv-6*, *n2845*, we identified in the ORF of Y50D4C.d an amber nonsense mutation causing a deletion of the last 42 amino acids of the predicted protein product (Fig. 2B).

We determined the sequence of PCR-amplified cDNA and 5'-RACE products corresponding to Y50D4C.d. We found that this cDNA contains a 5' SL1 *trans*-spliced leader, 23 bases of 5' UTR and 2418 bases of ORF, including two additional 5' exons not in Y50D4C.d. The longest ORF in this cDNA including the additional exons is predicted to encode a protein of 806 amino acids. Expression of this ORF under the control of the *C. elegans* heat-shock promoters (Stringham et al., 1992) prior to the start of vulval morphogenesis rescued the *sqv-6* vulval morphogenesis defect in all animals (n=13) and the maternal-effect lethality of the progeny of *sqv-6* homozygotes generated by *+/sqv-6* heterozygous parents for three of 13 *sqv-6* homozygotes studied.

5. *sqv-6* encodes a protein similar to xylosyltransferases.

Of the 806 amino acids of the SQV-6 protein, 182 (23%) and 193 (24%) are identical to human xylosyltransferases I and II, respectively (Fig. 2B). Both the predicted SQV-6 protein and the human xylosyltransferase II contain a putative transmembrane domain near the amino terminus and are likely to be type II transmembrane proteins. Neither the start codon nor a presumptive transmembrane domain has been defined for human xylosyltransferase I (Gotting et al., 2000).

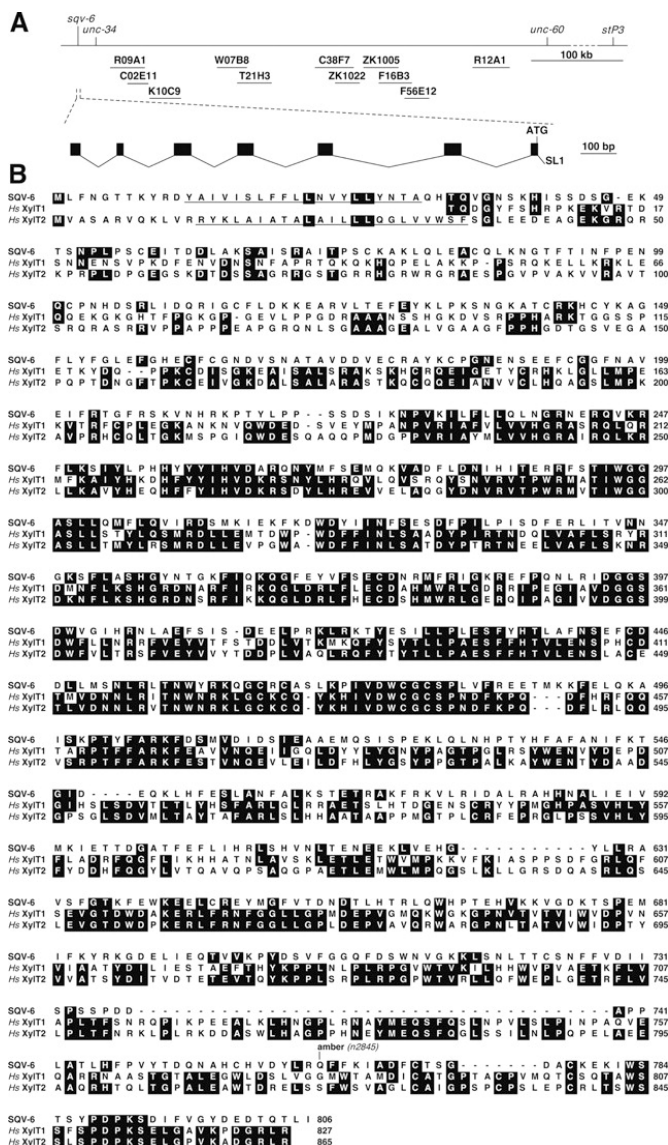


Figure 2. SQV-6 is similar to xylosyltransferases.

(A) Genetic and physical maps showing *sqv-6*. Short solid lines represent cosmid clones that were assayed in germline transformation experiments. Below is the structure of the *sqv-6* gene as deduced from genomic and cDNA sequences. Solid boxes indicate exons. The *trans*-spliced leader SL1 and the start codon (ATG) are indicated. (B) Alignment of SQV-6 and two human xylosyltransferases. Identities between at least two proteins are shaded in black. The predicted transmembrane domains are underlined. The single *sqv-6* nonsense allele is indicated. The numbers on the right indicate amino acid positions.

6. *sqv-6* can correct a xylosyltransferase defect in CHO cells.

We tested the ability of *sqv-6* to act as a xylosyltransferase by testing its ability to complement GAG-deficient Chinese hamster ovary (CHO) mutant cells lacking this enzymatic activity (Esko et al., 1985). Mutant pgsA-745 cells were transiently transfected with a plasmid containing *sqv-6* under the control of a cytomegalovirus (CMV) promoter. These cells showed partial rescue of the defect, as assayed by the ability to incorporate $^{35}\text{SO}_4$ into GAGs (16-27% that of the wild type) and by binding of biotinylated FGF-2 to cell surface heparan sulfate by flow cytometry (data not shown). From these transiently transfected cells, we obtained a clonal cell line stably expressing *sqv-6*. This cell line showed full restoration of FGF-2 binding to heparan sulfate on the cell surface (Fig. 3A). Stable expression of *sqv-6* in pgsA-745 enhanced the incorporation of $^{35}\text{SO}_4$ into GAGs to approximately 50% of wild-type levels, compared to 1% for the untreated mutant or mutant transfected with empty vector (Fig. 3B). Of the $^{35}\text{SO}_4$ incorporated into GAGs, 30-40% was released by treatment with chondroitinase ABC and 55-65% by a heparin lyase mixture, indicating that the composition of chondroitin and heparan sulfate was comparable in wild-type CHO cells and pgsA-745 cells transfected with *sqv-6*. Expression of *sqv-6* also resulted in restoration of xylosyltransferase activity, as measured by the transfer of xylose from UDP-xylose to a soluble silk acceptor, whereas pgsA-745 cells transfected with empty vector had virtually no activity (Fig. 3C).

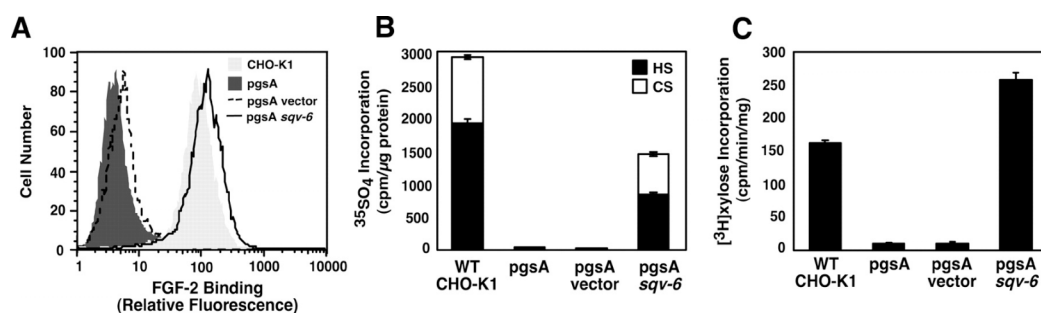


Figure 3. *sqv-6* rescues a xylosyltransferase-deficient CHO cell line.

(A) FGF-2 binding to cell-surface heparan sulfate as assayed by flow cytometry (Bai et al., 1999). Light gray shading, wild-type CHO-K1. Dark gray shading, mutant pgsA-745. Dashed line, pgsA-745 with empty vector. Solid line, pgsA-745 with *sqv-6*. (B) ³⁵SO₄ incorporation into GAGs (see Experimental Procedures). Black bars, [³⁵S]heparan sulfate (HS). White bars, [³⁵S]chondroitin sulfate (CS). The average values ± standard deviations of means (n=3) are shown. (C) Xylosyltransferase activity in crude cell extracts (see Experimental Procedures). Average incorporation of [³H]xylose from UDP-[1-³H]xylose into soluble silk acceptor ± standard deviations of means (n=3) are shown.

7. The *sqv-2* and *sqv-6* genes act in the *C. elegans* chondroitin and heparan sulfate biosynthesis pathway.

Our findings indicate that *sqv-2* and *sqv-6* encode galactosyltransferase II and xylosyltransferase, respectively. With the previously identified *sqv-3* galactosyltransferase I and *sqv-8* glucuronosyltransferase I, all four *C. elegans* genes responsible for the biosynthesis of the proteoglycan core protein linkage tetrasaccharide of chondroitin and heparan sulfate have now been defined (Fig. 4). Three previously identified genes, *sqv-4* UDP-glucose dehydrogenase, *sqv-1* UDP-GlcA decarboxylase and *sqv-7* UDP-GlcA/UDP-Gal/UDP-GalNAc transporter, act in earlier steps of GAG biosynthesis. All *sqv* genes identified to date affect the biosynthesis of both chondroitin and heparan sulfate. Based upon these observations, we conclude that in *C. elegans* early embryonic cytokinesis and epithelial invagination during vulval development depend on the expression of GAGs.

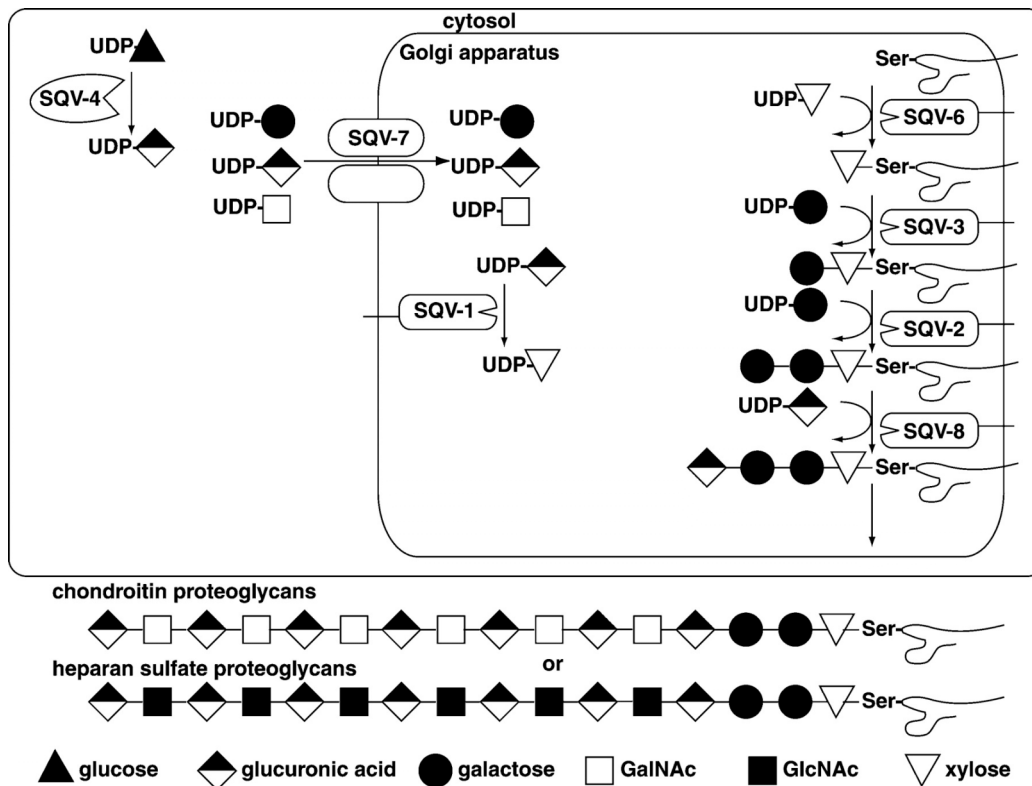


Figure 4. Model for the role of seven *sqv* genes in glycosaminoglycan biosynthesis.

SQV-4 converts UDP-glucose to UDP-GlcA (Hwang and Horvitz, 2002b). SQV-7 transports UDP-GlcA, UDP-Gal and UDP-GalNAc from the cytoplasm to lumen of the Golgi apparatus (Berninsone et al., 2001). SQV-1 converts UDP-GlcA to UDP-Xyl in the lumen of the Golgi apparatus (Hwang and Horvitz, 2002a). SQV-6 is xylosyltransferase (this study). SQV-3 is galactosyltransferase I (Bulik et al., 2000). SQV-2 is galactosyltransferase II (this study). SQV-8 is glucuronosyltransferase I (Bulik et al., 2000). In other organisms, two additional sets of glycosyltransferases act in later steps of the biosynthesis of chondroitin and heparan sulfate (Esko and Selleck, 2002).

E. Acknowledgements

We thank Beth Castor for help with DNA sequence determination, Ewa Davison and Ignacio Perez de la Cruz for communicating the locations of the *lin-8* and *sup-9* loci, respectively, and Mark Alkema and Melissa Harrison for reading this manuscript. This work was supported by NIH grant GM24663 to H.R.H and NIH grant GM33063 to J.D.E. H.R.H. is an Investigator of the Howard Hughes Medical Institute.

F. References

Almeida, R., Levery, S.B., Mandel, U., Kresse, H., Schwientek, T., Bennet, E.P., and Clausen, H. (1999). Cloning and expression of a proteoglycan UDP-galactose:b-xylose b,4-galactosyltransferase I. *J. Biol. Chem.* **274**, 26165-71.

Bai, X., Wei, G., Sinha, A., and Esko, J. D. (1999). Chinese hamster ovary cell mutants defective in glycosaminoglycan assembly and glucuronosyltransferase I. *J. Biol. Chem.* **274**, 13017-24.

Bai, X., Zhou, D., Brown, J. R., Crawford, B. E., Hennet, T., and Esko, J. D. (2001). Biosynthesis of the linkage region of glycosaminoglycans: cloning and activity of galactosyltransferase II, the sixth member of the beta 1,3-galactosyltransferase family (beta 3GalT6). *J. Biol. Chem.* **276**, 48189-95.

Bame, K. J., and Esko, J. D. (1989). Undersulfated heparan sulfate in a Chinese hamster ovary cell mutant defective in heparan sulfate N-sulfotransferase. *J. Biol. Chem.* **264**, 8059-65.

Bellaïche, Y., The, I., and Perrimon, N. (1998). Tout-velu is a *Drosophila* homologue of the putative tumour suppressor EXT-1 and is needed for Hh diffusion. *Nature* **394**, 85-8.

Berninsone, P., Hwang, H. Y., Zemtseva, I., Horvitz, H. R., and Hirschberg, C. B. (2001). SQV-7, a protein involved in *Caenorhabditis elegans* epithelial invagination and early embryogenesis, transports UDP-glucuronic acid, UDP-N-acetylgalactosamine, and UDP-galactose. *Proc. Natl. Acad. Sci. USA* **98**, 3738-43.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71-94.

Bulik, D. A., Wei, G., Toyoda, H., Kinoshita-Toyoda, A., Waldrip, W. R., Esko, J. D., Robbins, P. W., and Selleck, S. B. (2000). *sqv-3*, *-7*, and *-8*, a set of genes affecting morphogenesis in *Caenorhabditis elegans*, encode enzymes required for glycosaminoglycan biosynthesis. *Proc. Natl. Acad. Sci. USA* **97**, 10838-43.

Esko, J. D., and Selleck, S. B. (2002). ORDER OUT OF CHAOS: Assembly of ligand binding sites in heparan sulfate. *Annu. Rev. Biochem.* **71**, 435-71.

Esko, J. D., Stewart, T. E., and Taylor, W. H. (1985). Animal cell mutants defective in glycosaminoglycan biosynthesis. *Proc. Natl. Acad. Sci. USA* **82**, 3197-201.

Gotting, C., Kuhn, J., Zahn, R., Brinkmann, T., and Kleesiek, K. (2000). Molecular cloning and expression of human UDP-d-Xylose:proteoglycan core protein beta-d-xylosyltransferase and its first isoform, XT-II. *J. Mol. Biol.* **304**, 517-28.

Herman, T., Hartweg, E., and Horvitz, H. R. (1999). *sqv* mutants of *Caenorhabditis elegans* are defective in vulval epithelial invagination. *Proc. Natl. Acad. Sci. USA* **96**, 968-73.

Herman, T., and Horvitz, H. R. (1997). Mutations that perturb vulval invagination in *C. elegans*. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 353-9.

Herman, T., and Horvitz, H. R. (1999). Three proteins involved in *Caenorhabditis elegans* vulval invagination are similar to components of a glycosylation pathway. *Proc. Natl. Acad. Sci. USA* **96**, 974-9.

Hwang, H.-Y., and Horvitz, H. R. (2002a). The SQV-1 UDP-glucuronic acid decarboxylase and the SQV-7 nucleotide-sugar transporter may act in the Golgi apparatus to affect *C. elegans* vulval morphogenesis and embryonic development. *Proc. Natl. Acad. Sci. USA* **99**, 14218-14223.

Hwang, H.-Y., and Horvitz, H. R. (2002b). The *C. elegans* vulval morphogenesis gene *sqv-4* encodes a UDP-glucose dehydrogenase that is temporally and spatially regulated. *Proc. Natl. Acad. Sci. USA* **99**, 14224-14229.

Krause, M., and Hirsh, D. (1987). A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* **49**, 753-61.

Lin, X., and Perrimon, N. (1999). Dally cooperates with *Drosophila* Frizzled 2 to transduce Wingless signalling. *Nature* **400**, 281-4.

Lin, X., Wei, G., Shi, Z., Dryer, L., Esko, J. D., Wells, D. E., and Matzuk, M. M. (2000). Disruption of gastrulation and heparan sulfate biosynthesis in EXT1-deficient mice. *Dev. Biol.* **224**, 299-311.

Mello, C. C., Kramer, J. M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959-3970.

Okajima, T., Fukumoto, S., Furukawa, K., and Urano, T. (1999a). Molecular basis for the progeroid variant of Ehlers-Danlos syndrome. *J. Biol. Chem.* **274**, 28841-4.

Okajima, T., Yoshida, K., Kondo, T., and Furukawa, K. (1999b). Human homolog of *Caenorhabditis elegans sqv-3* gene is galactosyltransferase I involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans. *J. Biol. Chem.* **274**, 22915-8.

Quentin, E., Gladen, A., Roden, L., and Kresse, H. (1990). A genetic defect in the biosynthesis of dermatan sulfate proteoglycan: galactosyltransferase I deficiency in fibroblasts from a patient with a progeroid syndrome. *Proc. Natl. Acad. Sci. USA* **87**, 1342-6.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Plainview, New York

Stringham, E. G., Dixon, D. K., Jones, D., and Candido, E. P. (1992). Temporal and spatial expression patterns of the small heat shock (hsp16) genes in transgenic *Caenorhabditis elegans*. *Mol. Biol. Cell* **3**, 221-33.

Wei, Z., Swiedler, S. J., Ishihara, M., Orellana, A., and Hirschberg, C. B. (1993). A single protein catalyzes both N-deacetylation and N-sulfation during the biosynthesis of heparan sulfate. *Proc. Natl. Acad. Sci. USA* **90**, 3885-8.

Zak, B. M., Crawford, B. E., and Esko, J. D. (2002). Hereditary multiple exostoses and heparan sulfate polymerization. *Biochim Biophys Acta* **1573**, 346-55.

The text of Chapter 2 is a reprint of the material as it appeared in the Journal of Biological Chemistry, Hwang H.Y., Olson S.K., Brown J.R., Esko J.D. and Horvitz H.R. (2003). The dissertation author was a secondary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

CHAPTER 3

***C. elegans* early embryogenesis and vulval morphogenesis depend on chondroitin glycosaminoglycan biosynthesis**

A. Summary

Defects in glycosaminoglycan (GAG) biosynthesis disrupt animal development and cause human disease. To date, much of the focus on GAGs has been on heparan sulfate (HS). Mutations in eight *sqv* (squashed vulva) genes in *C. elegans* cause defects in cytokinesis during embryogenesis and in vulval morphogenesis during postembryonic development. Seven of the eight *sqv* genes have been shown to control the biosynthesis of the GAGs chondroitin and HS. Here we present the molecular identification and characterization of the eighth gene, *sqv-5*. The *sqv-5* gene encodes a bifunctional glycosyltransferase that probably is localized to the Golgi apparatus and is responsible for the biosynthesis of chondroitin but not of HS. This finding reveals that chondroitin plays a crucial role in cytokinesis and morphogenesis during *C. elegans* development.

B. Introduction

GAGs or mucopolysaccharides have been of great interest to biologists for decades (Comper et al., 1978; Ruoslahti, 1988; Perrimon and Bernfield, 2000; Esko

and Selleck, 2002; Schwartz and Domowicz, 2002; Silbert and Sugumaran, 2002; Zak et al., 2002). Recent analyses of mutations that cause developmental defects in *Drosophila melanogaster* revealed that GAG biosynthesis, in particular HS synthesis, is important for intercellular signaling mediated by the *wingless*, *hedgehog*, and FGF pathways (Perrimon and Bernfield, 2000). In addition, mutations in GAG biosynthesis have been implicated in human diseases, including a progeroid variant of the connective tissue disorder Ehlers-Danlos Syndrome (EDS) (Quentin et al., 1990) and Hereditary Multiple Exostoses (HME) (Zak et al., 2002), which is characterized by inappropriate chondrocyte proliferation and bone growth. Much of the focus on GAGs has been on HS, in part because of the many ligands it binds, its action in growth factor signaling, and its role in *Drosophila* development (Perrimon and Bernfield, 2000; Esko and Selleck, 2002). Studies of chondroitin sulfate (CS), another major class of GAGs in vertebrates, have focused on the development of cartilage, tendon and bone (Schwartz and Domowicz, 2002).

The *C. elegans* genes *sqv-1* to *-8* are important for both embryonic development and postembryonic vulval morphogenesis (Herman et al., 1999). The progeny of mutants homozygous for strong loss-of-function *sqv* mutations die during embryogenesis, with most arresting at the one-cell stage. This arrest is caused by a defect in the initiation and completion of cytokinesis, which may be caused by a failure to form a fluid-filled extracellular space between the plasma membrane and the eggshell (Hwang and Horvitz, 2002a). During the L4 larval stage, *sqv-1* to *-8* mutants fail to expand the extracellular space of the vulva, which is the opening through which

sperm and eggs pass in adult hermaphrodites. These mutants form a partially functional vulva but are normal in vulval cell proliferation, migration and fusion. We proposed that in *C. elegans* during both embryogenesis and vulval morphogenesis, GAGs added to extracellular matrices drive the formation of fluid-filled extracellular spaces (Hwang and Horvitz, 2002a). In sea urchin fertilization, the secretion of GAGs has long been thought to cause the swelling of a fluid-filled space between the vitelline envelope and the plasma membrane (Austin, 1965).

The molecular identities of seven of the *sqv* genes indicate a defect in the biosynthesis of two types of GAGs present in *C. elegans*, HS and a non-sulfated chondroitin (Hwang and Horvitz, 2002a; Herman and Horvitz, 1999; Bulik et al., 2000; Berninsone et al., 2001; Hwang and Horvitz, 2002b; Hwang et al., 2003). SQV-4 (UDP-glucose dehydrogenase) synthesizes UDP-glucuronic acid (UDP-GlcA) in the cytoplasm (Hwang and Horvitz, 2002b), which is translocated into the lumen of the Golgi apparatus by the SQV-7 nucleotide-sugar transporter (Berninsone et al., 2001). SQV-7 also translocates UDP-galactose and UDP-N-acetylgalactosamine (Berninsone et al., 2001). SQV-1 catalyzes the decarboxylation of UDP-GlcA (Hwang and Horvitz, 2002a), forming the first nucleotide-sugar donor required for GAG biosynthesis, UDP-xylose. In the lumen of the Golgi apparatus, UDP-xylose, UDP-galactose, and UDP-GlcA are used as substrates of the SQV-6 xylosyltransferase (Hwang et al., 2003), the SQV-3 galactosyltransferase I (Herman and Horvitz, 1999; Bulik et al., 2000), the SQV-2 galactosyltransferase II (Hwang et al., 2003) and the SQV-8 glucuronosyltransferase I (Herman and Horvitz, 1999; Bulik et al., 2000) to build the

protein core-GAG linkage tetrasaccharide (GlcA β 1,3Gal β 1,3Gal β 1,4Xyl β -O-serine) on which GAG backbones polymerize. Evidence that four such glycosylation reactions are essential for mammalian GAG polymerization has been obtained from studies of mutant hamster cell lines (Esko and Selleck, 2002). Mutation in the human homolog of the *sqv-3* galactosyltransferase I gene has been implicated as the cause of a progeroid variant of the connective-tissue disorder Ehlers-Danlos Syndrome (Almeida, 1999; Okajima et al., 1999), which is characterized by loose skin and hypermobile joints. It seems likely that homologs of other *sqv* genes are involved in similar disorders.

C. Results

1. Cloning of *sqv-5*

By physically mapping chromosomal deletions, we localized *sqv-5* to a roughly 200-kilobase region between *fog-3* and the left endpoint of *qDf10* (Fig. 1). An 18,448 base *Bam*HI-*Pst*I fragment of cosmid K09A8, containing a single complete predicted gene, T24D1.1, rescued the *sqv-5* mutant phenotype. Introducing a nonsense or frameshift mutation in T24D1.1 eliminated this rescuing activity. The *sqv-5 n3039* allele is a nonsense mutation in the T24D1.1 open reading frame (ORF). We isolated a mutation (*n3611*) that deletes most of the T24D1.1 ORF and causes the same Sqv mutant phenotype as that of *sqv-5(n3039)* animals. We conclude that *sqv-5* corresponds to T24D1.1. We found three discrepancies between our DNA sequencing results and those of the *C. elegans* Sequencing Consortium,

one of which caused us to modify the predicted gene structure of T24D1.1 to that depicted in Fig. 1. From the sequences of *sqv-5* cDNA clones and 5' RACE products (see Methods), we identified two alternatively spliced forms of *sqv-5* cDNAs, which encode predicted proteins of 734 and 736 amino acids. We detected a size 3.6 kb transcript on a northern blot (data not shown), consistent with the size predicted by our cDNA and 5' RACE results.

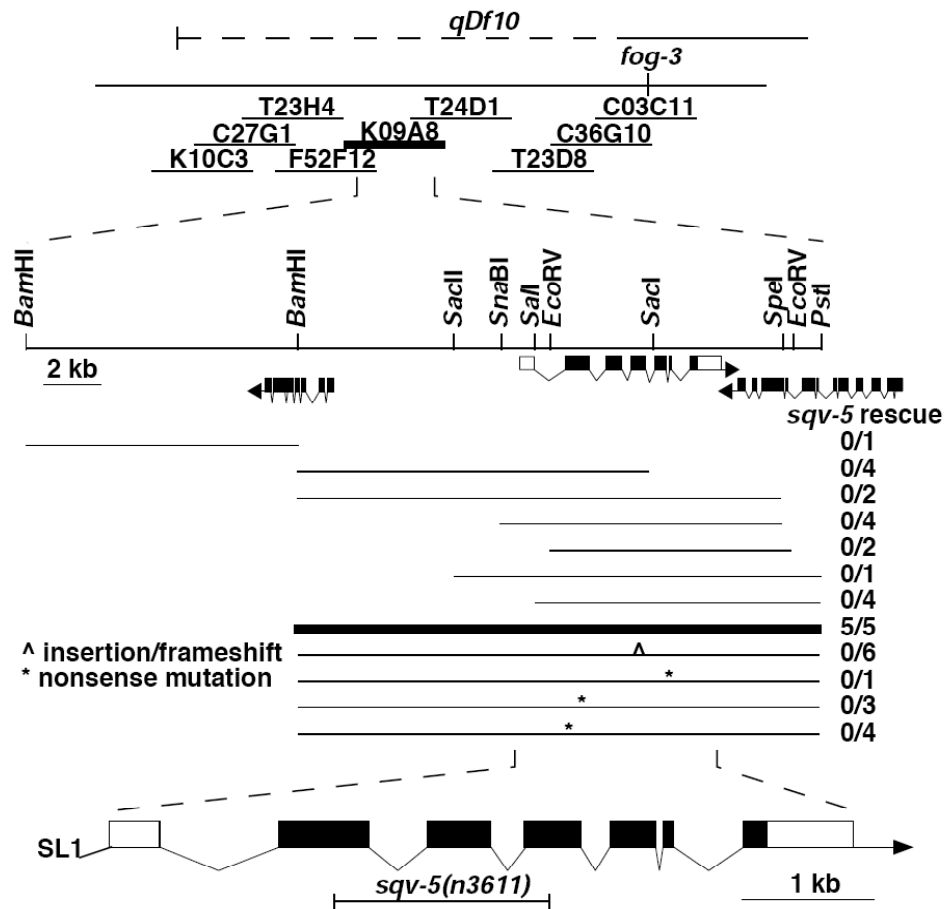


Figure 1. Cloning of *sqv-5*

Top, Genetic and physical maps of *sqv-5*. The dashed horizontal line depicting *qDf10* indicates the left deletion end point between cosmids K10C3 and C03C11. Cosmid clones assayed in germline transformation experiments are shown. Cosmid K09A8, in bold, rescued the *sqv-5* mutant phenotype. Middle, Subclones of K09A8 assayed for rescue are shown. Predicted genes in the minimal rescuing fragment (in bold) are shown. Solid boxes indicate exons, open boxes indicate untranslated sequences, and the arrow indicates 3' poly A sequence. Rescue data are shown as the number of transformed lines that rescued/total. The caret (^) indicates introduction of a four-base (CGCG) addition/frameshift (after A451) and the asterisk (*) indicates the introduction of a nonsense codon (W664opal, Y160amber or G21opal) into the *sqv-5* coding sequence. Bottom, The structure of the *sqv-5* gene as deduced from genomic and cDNA sequences. The *trans*-spliced leader SL1 is indicated. The extent of the *sqv-5*(n3611) deletion is indicated by a horizontal line.

2. SQV-5 is similar to the human chondroitin sulfate synthase

Of the 734 amino acids in the short form of the SQV-5 protein, 270 (37%) are identical to a recently cloned human CS synthase (Kitagawa et al., 2001) (Fig. 1, Table 1). We also identified and determined the sequences of a cDNA from a *Drosophila melanogaster* gene that is predicted to encode an 832 amino acid protein with 270 amino acid identities (37%) with SQV-5. SQV-5 is less similar to the human CS N-acetylgalactosaminyltransferase I (GalNAcT-I) (Uyama et al., 2002), with which it shares 109 amino acid identities (20%). Interestingly, SQV-5 is the only protein in the *C. elegans* genome with extensive similarity to the human CS GalNAcT-I. By contrast, the *Drosophila* genome contains a second gene with strong similarity to and that likely encodes the ortholog of the human CS GalNAcT-I (37% identity) (Table 1, Fig. 2). All five proteins contain a single predicted transmembrane domain near the N-terminus, consistent with a type II transmembrane topology. Such N-terminal transmembrane domains are typical of glycosyltransferases, which have active sites that face the lumen of the Golgi apparatus.

Table 1. Amino acid sequence identities

Percentages of amino acid sequence identities are shown, as calculated by dividing the number of identical amino acids (shown in parenthesis) divided by total number of amino acids of the smaller protein. The proteins shown are SQV-5, Human CS synthase (*Hs CS synth*), *Drosophila* CS synthase candidate (*Dm CS synth*), Human GalNAcT I (*Hs GalNAcT-I*), and *Drosophila* CS GalNAcT I candidate (*Dm GalNAcT-I*).

	<i>Hs CS synth</i>	<i>Dm CS synth</i>	<i>Hs GalNAcT-I</i>	<i>Dm GalNAcT-I</i>
SQV-5	37% (270)	37% (270)	20% (109)	19% (103)
<i>Hs CS synth</i>	-	41% (326)	25% (134)	23% (124)
<i>Dm CS synth</i>	-	-	23% (124)	22% (117)
<i>Hs GalNAcT</i>	-	-	-	37% (196)

3. SQV-5 has chondroitin synthase activity

In vertebrates, CS synthase catalyzes the alternating, stepwise addition of GlcA and N-acetylgalactosamine (GalNAc) to the nascent chain, resulting in the polymerization of the CS backbone. Protein extracts prepared from whole animals that were homozygous for the *sqv-5(n3611)* null allele, heterozygous for *sqv-5(n3611)*, or wild-type were assayed for glucuronosyltransferase (GlcAT-II) and N-acetylgalactosaminyltransferase (GalNAcT-II) activities using chemically desulfated CS as the acceptor. Significant enzymatic activities were observed in the extracts of the wild-type animals in both assays (Fig. 3a-b). Animals heterozygous for *sqv-5(n3611)* contained slightly over one-half the enzymatic activity observed in wild-type animals, and no significant enzymatic activity above the negative control was observed in the extracts of the homozygous mutant animals. About half the enzymatic activity was observed in assays consisting of half wild-type and half *sqv-5* null extracts (w.t.=1466±144, *sqv-5*=105±35, w.t. + *sqv-5*=809±57, ctl.=68±24 cpm (s.d.)), indicating that the mutant animals do not contain an inhibitor of the enzyme. These findings establish that *sqv-5* controls *C. elegans* chondroitin synthase activity and in combination with our sequence data show that the SQV-5 protein has synthase activity. By contrast, we observed no difference in N-acetylglucosaminyltransferase II (GlcNAcT-II) activity, which is required for HS synthesis, in extracts of *sqv-5(n3611)* mutants (Fig. 3c), suggesting that chondroitin but not HS biosynthesis is disrupted in *sqv-5* mutants.

The addition of the first GalNAc residue to the linkage tetrasaccharide is thought to be catalyzed by an enzyme different from the one involved in CS backbone polymerization (Kitagawa et al., 2001; Uyama et al., 2002). This reaction can be assayed using glucuronic acid β 1,3galactose-O-naphthalenemethanol as the acceptor instead of desulfated CS. Using this assay, we observed significant enzymatic activity in extracts of wild-type animals (Fig. 3d). Approximately one-half of the enzymatic activity above that of the negative control was present in extracts of animals heterozygous for *sqv-5(n3611)*, whereas homozygous null animals lacked activity. These findings indicate that SQV-5 acts in both the initiation and the elongation of chondroitin chains. We find it interesting that SQV-5, which shares a similar degree of amino acid sequence identity to the human GalNAcT-I as to the human CS synthase (Table 1), has this additional GalNAcT-I activity. The human CS synthase apparently lacks this activity (Kitagawa et al., 2001).

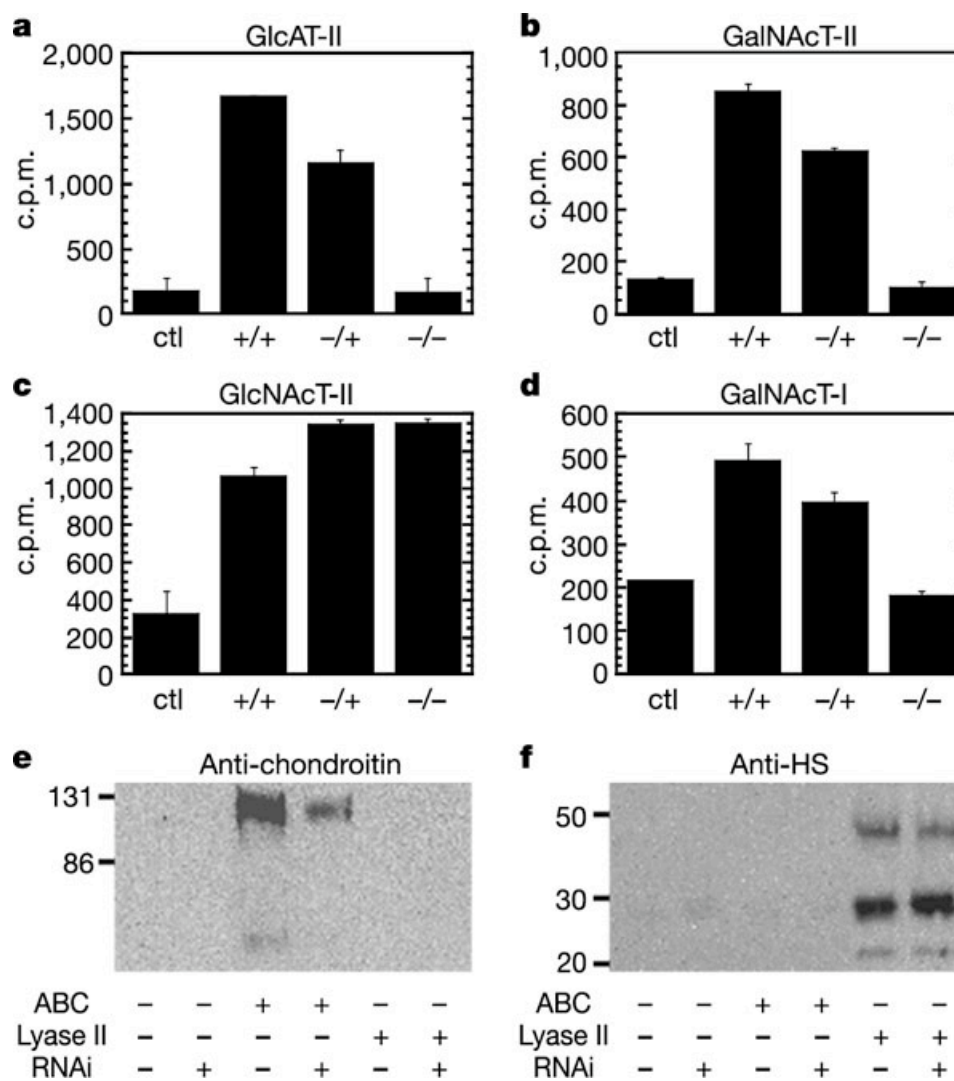


Figure 3. SQV-5 has chondroitin synthase activity

Glycosyltransferase activities were assayed (see Methods) using cell-free extracts prepared from wild-type animals (w.t.) and *sqv-5(n3611)* heterozygotes (*sqv-5/+*) and homozygotes (*sqv-5*). Activities measured without an acceptor (ctl.) are also shown. Means and standard errors from representative experiments are shown. **a**, GlcAT-II assay. **b**, GalNAcT-II assay. **c**, GlcNAcT-II assay. **d**, GalNAcT-I assay. **e** and **f**, Intact proteoglycans were purified from vector (+) or *sqv-5* (-) RNAi-treated animals and digested with chondroitinase ABC and/or heparin lyase II or buffer. Western blot analysis was done using **(e)** anti-chondroitin or **(f)** anti-HS mAbs and visualized using HRP-conjugated secondary antibodies.

4. *sqv-5(n3611)* mutants have decreased chondroitin levels

To assay changes in chondroitin and HS content in *sqv-5* mutants, we isolated GAGs from wild-type and *sqv-5(n3611)* adult hermaphrodites. The amount of chondroitin was reduced from 182 ± 52 (s.d.) fmol/worm in wild-type animals to 24 ± 20 fmol/worm in mutant animals. The level of HS was below the limits of detection, because the relative amount of HS is 150- to 250-fold less than chondroitin in *C. elegans* (Yamada et al., 1999; Toyoda et al., 2000). We also suppressed *sqv-5* function by RNA-mediated interference (RNAi) by ingestion of dsRNA (Timmons and Fire, 1998). *sqv-5* RNAi-treated L4 larvae had reduced vulval extracellular spaces reminiscent of mutants homozygous for a weak loss-of-function mutation in other *sqv* genes (data not shown). The *sqv-5* RNAi-treated adults were reduced in brood size (mean= 29 ± 23 (s.d.), n=47) compared to vector RNAi-treated adults (mean= 264 ± 33 (s.d.), n=31). These observations suggest that *sqv-5* function was incompletely suppressed in these animals, since animals homozygous for either *sqv-5* mutant allele have average brood sizes of zero. Intact proteoglycans were isolated from L4 larvae that had been treated with *sqv-5*- or vector-RNAi and digested with either chondroitinase ABC or heparin lyase II. These enzyme treatments of chondroitin and HS result in a “stub” oligosaccharide consisting of the linkage tetrasaccharide and one disaccharide repeat containing a terminal 4,5-unsaturated uronic acid, which is recognized by an anti-chondroitin monoclonal antibody (mAb) or an anti-HS mAb, respectively. Western blot analysis using the anti-chondroitin mAb revealed a major proteoglycan of approximately 120

kDa and several minor bands. This major band of 120 kDa was reduced 5-8 fold in *sqv-5* RNAi-treated animals compared to vector RNAi-treated animals (Fig. 3e-f). Western blot analysis using the anti-HS mAb detected three proteoglycans of 24, 30 and 48 kDa, which did not vary in level between vector and *sqv-5* RNAi-treated animals. Thus, loss of *sqv-5* function selectively reduced chondroitin levels.

5. SQV-5 localization

To study the expression and subcellular localization of the SQV-5 protein, we generated affinity-purified rabbit polyclonal antibodies against a SQV-5-GST fusion protein. Anti-SQV-5 antibodies stained multiple punctate foci in the cytoplasm of the vulva, the uterus and oocytes (Figs. 4a-c). This punctate staining was not seen in animals homozygous for the *sqv-5(n3611)* null allele (data not shown). Previously, we observed a similar punctate staining pattern using antibodies specific for the SQV-7 nucleotide sugar transporter and the SQV-1 UDP-GlcA decarboxylase (Hwang and Horvitz, 2002a). We used anti-SQV-1 rat antibodies and anti-SQV-5 rabbit antibodies to show that SQV-1 and SQV-5 proteins co-localize and hence appear to be present in the same cytoplasmic compartment (Figs. 4c-e). We suggest that the site of nucleotide-sugar biosynthesis by SQV-1, nucleotide-sugar transport by SQV-7 (Hwang and Horvitz, 2002a) and elongation of the disaccharide region of chondroitin by SQV-5 are all catalyzed in the same subcellular compartment, presumably the Golgi apparatus. Most glycosyltransferases involved in GAG biosynthesis have been shown to localize to the Golgi apparatus in vertebrates (Silbert and Sugumaran, 2002).

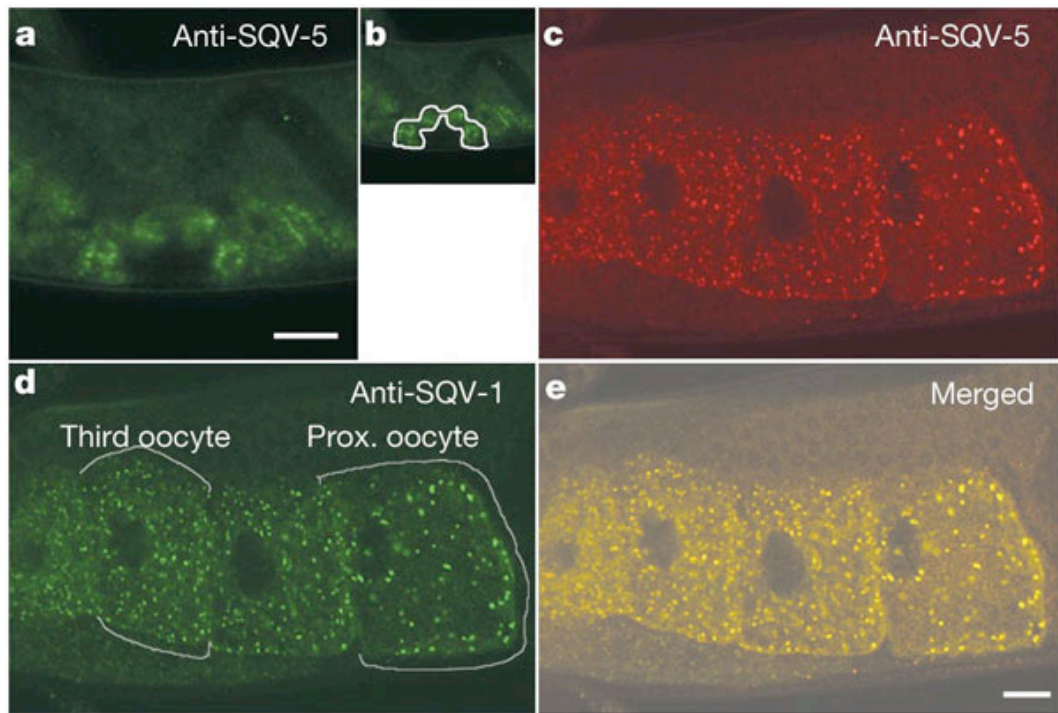


Figure 4. SQV-5 localization

Wild-type animals were stained with anti-SQV-5 antibodies (see Methods) and/or anti-SQV-1 antibodies (Hwang and Horvitz, 2002a). **a**, L4 vulva stained with SQV-5 rabbit antibodies. **b**, Diagram of vulva in **(a)**. **c**, Adult oocytes stained with SQV-5 rabbit antibodies. **d**, Adult oocytes stained with SQV-1 rat antibodies. Staining was visualized using FITC-conjugated and/or Texas Red-conjugated secondary antibodies. **e**, Merged image of panels **(c)** and **(d)** showing that the SQV-1 and SQV-5 proteins colocalize in oocytes.

6. The *Sqv* phenotype results from loss of chondroitin, not heparan sulfate

Our molecular identification of *sqv-5* defines the last step in the *C. elegans* biosynthetic pathway for chondroitin and suggests that defects in the biosynthesis of chondroitin account for the embryonic and vulval defects caused by mutations in all *sqv* genes (Fig. 5). By contrast, defects in the biosynthesis of HS leading to abnormalities in cell signaling pathways, such as the *wingless* and *hedgehog* pathways, are implicated as the cause of many developmental defects in *Drosophila*, including those defects caused by mutations that are predicted to disrupt both CS and HS biosynthesis (Perrimon and Bernfield, 2000). Unlike HS, which is known to bind a large number of ligands involved in morphogenesis, wound healing, host defense and energy metabolism (Esko and Selleck, 2002), non-sulfated chondroitin is not known to bind to specific ligands. In vertebrates, large amounts of CS are secreted into the extracellular matrix, where it plays a structural role and binds to ligands such as type I collagen (Ruoslahti, 1988). The ability of chondroitin to interact with water, cause swelling and generate osmotic pressure on its surroundings could be responsible for its biological effects (Comper, 1978), including the expansion of the extracellular spaces of the *C. elegans* embryo and vulva. Studies of sea urchin gastrulation have led to the proposal that the secretion of CS proteoglycan can result in the hydration of the extracellular matrix and cause epithelial invagination (Lane et al., 1993). Our studies provide support for such a mechanism. However, we cannot exclude a variety of possible alternatives, e.g., mechanisms involving adhesion, cytoskeletal rearrangement or intercellular

signaling (Herman et al., 1999; Hwang and Horvitz, 2002a). Whatever the mechanism of chondroitin action, our findings demonstrate the importance of chondroitin in cytokinesis, early embryogenesis and epithelial morphogenesis in *C. elegans* and support the hypothesis that CS, like HS, has a broad and major role in development and disease.

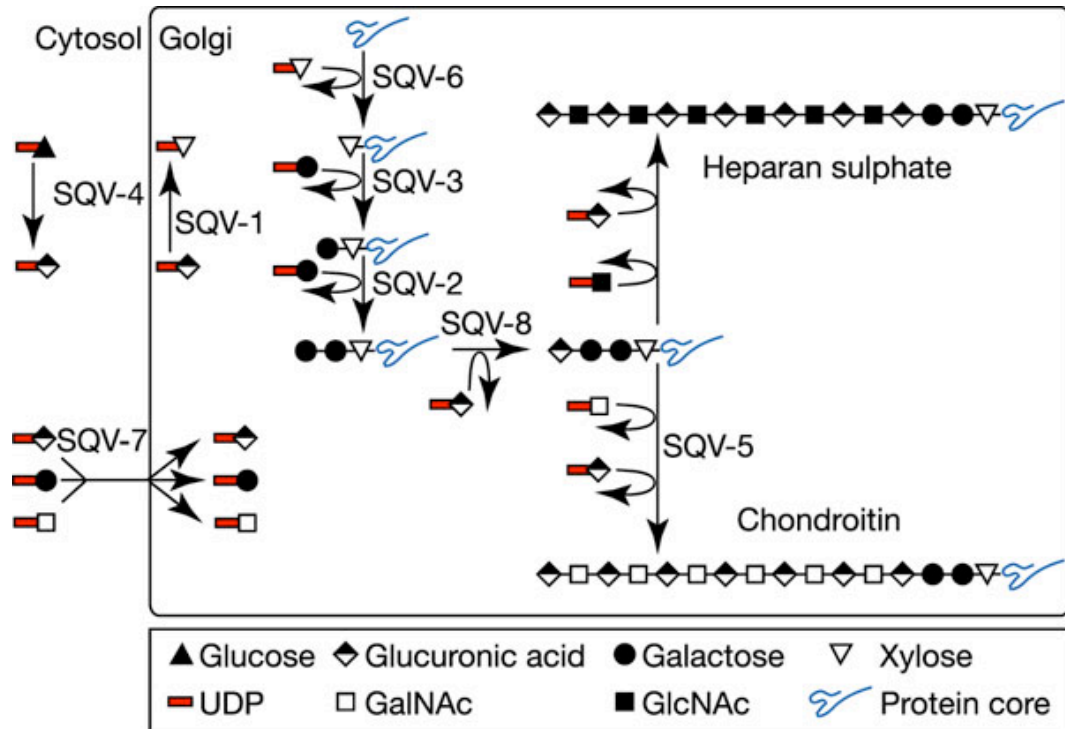


Figure 5. Model

The SQV proteins act in the biosynthesis of heparan and chondroitin backbones. SQV-4 converts UDP-glucose to UDP-glucuronic acid (Hwang and Horvitz, 2002b). SQV-7 transports UDP-glucuronic acid, UDP-galactose and UDP-N-acetylgalactosamine (UDP-GalNAc) from the cytoplasm to lumen of the Golgi apparatus (Berninsone et al., 2001). SQV-1 converts UDP-glucuronic acid to UDP-xylose in the lumen of the Golgi apparatus (Hwang and Horvitz, 2002a). SQV-6 is the GAG xylosyltransferase (Hwang et al, 2003), SQV-3 is the GAG galactosyltransferase I (Bulik et al., 2000), SQV-2 is the GAG galactosyltransferase II (Hwang et al., 2003), and SQV-8 is the GAG glucuronosyltransferase I (Bulik et al., 2000). SQV-5 is required for the synthesis of the repeating disaccharide region of chondroitin (this study).

D. Methods

1. *sqv-5* mapping

We obtained Unc non-Vul and Vul non-Unc progeny from *unc-29(e1072) lin-11(n566)/sqv-5(n3039)* hermaphrodites. Five of 10 Unc non-Vul progeny carried *sqv-5(n3039)*, and three of 10 Vul non-Unc progeny carried *sqv-5(n3039)*, indicating *sqv-5(n3039)* is located between *unc-29* and *lin-11* and to the left of *lin-11*. We examined the vulval phenotype of animals of the genotype *ces-1(n703) Df/sqv-5(n3039)*, where the *Df*'s used were *qDf5*, *qDf7*, *qDf8*, *qDf9* and *qDf10*. All animals were Sqv, except for *ces-1(n703) qDf5/sqv-5(n3039)*. Because *qDf5*, *qDf7*, *qDf8*, *qDf9* and *qDf10* delete *fog-3* but only *qDf5* and *qDf7* delete *lin-11* (Ellis and Kimble, 1995), *sqv-5* maps to the left of *fog-3*. Using single *qDf10* eggs, we amplified genomic DNA sequence corresponding to the cosmids K10C3 and C03C11. A PCR product of expected length was amplified for K10C3 but not for C03C11 (n=10), thus placing the left end point of *qDf10* between K10C3 and C03C11 (Fig. 1).

2. *sqv-5* cDNA

We determined the sequences of two cDNA clones, yk20d7 and yk21g9, corresponding to T24D1.1, and of six 5'-rapid amplification of cloned ends (RACE) products derived from mixed stage RNA. The 5' RACE products contained a 5' SL1 *trans*-spliced leader, which is found at the 5' end of many *C. elegans* transcripts. The *sqv-5* cDNA contained a 417 base 5' UTR, a 2202 base ORF and a 657 base 3' UTR sequence. We identified two alternatively spliced forms of the transcript by 5'

RACE. Two of six cloned 5' RACE products represented a longer spliced-form containing six additional bases at the 5' end of the second exon.

3. Deletion allele of *sqv-5*

We isolated the deletion mutation *sqv-5(n3611)* from a library of animals mutagenized with UV illumination and trimethylpsoralen (Jansen et al., 1997). Mutant animals containing *sqv-5(n3611)* were backcrossed to the wild-type strain N2 six times. The 1641 bp deletion in *sqv-5(n3611)* removes bases “6124” to “7767” of the cosmid T24D1. The discrepancy of two bases reflects a DNA sequencing error by the *C. elegans* Sequencing Consortium. *sqv-5(n3611)* is predicted to encode a truncated SQV-5 missing 385 amino acids (amino acids 130 to 447) in the middle of SQV-5 and an alanine-to-phenylalanine substitution at amino acid 129.

4. Glycosyltransferase assays

sqv-5(n3611), *sqv-5(n3611)/hT2*, and wild-type N2 hermaphrodites were picked as L4 larvae by visual examination of the vulva using a dissecting microscope. The worms were allowed to grow for 23 to 27 hours at 22°C, then frozen in 50 mM Tris, pH 7.5 and stored at -70°C. Samples were sonicated in 0.05% Triton-X-100, 50 mM Tris and centrifuged at 15,000 g for 10 minutes. The protein content of the cleared supernatant was assessed by the Bradford assay and portions of the extracts were used for the following assays. The chondroitin acceptor was prepared by desulfation of shark cartilage chondroitin-4-sulfate (Nagasawa et al., 1977). N-acetylheparosan was prepared from *E. coli* K5 (Fritz et al., 1994) and the

disaccharide GlcA β 1,3Gal β -O-naphthalenemethanol was synthesized (Fritz et al., 1994).

The GlcAT-II activity of chondroitin synthase was measured by mixing 1.3×10^5 cpm UDP-[1- 3 H]glucuronic acid donor (20 Ci/mmol), 6 μ g β -glucuronidase-treated chondroitin acceptor, and 3 μ g worm extract in a 25 μ l reaction volume containing 0.05% Triton X-100, 10 mM MnCl₂, 100 μ M ATP, and 25 mM MES, pH 6.5. The GalNAcT-II activity of chondroitin synthase was detected by mixing 3×10^5 cpm UDP-[1- 3 H]GalNAc donor (38.5 Ci/mmol), 12 μ g chondroitin acceptor, and 15 μ g worm extract in a 25 μ l reaction volume containing 0.05% Triton X-100, 10 mM MnCl₂, and 25 mM MES, pH 6.5. The GlcNAcT-II activity of HS polymerase was assayed by mixing 5 mCi UDP-[6- 3 H]GlcNAc, 12 mg N-acetylheparosan acceptor, and 10 mg worm extract in 25 μ l reaction volume containing 20 mM MnCl₂, 0.45% Triton X-100, and 25 mM MOPS, pH 6.5. GlcAT-II, GlcNAcT-II and GalNAcT-II reactions were incubated 2 hours at 25°C, and products were separated from free nucleotide sugars by DEAE-Sephacel (Wei et al., 2000). GalNAcT-I activity was assayed as described for GalNAcT-II except that 5×10^5 cpm UDP-[1- 3 H]GalNAc donor and 12 mM GlcA β 1,3Gal β -O-naphthalenemethanol acceptor were used. Reactions were incubated 3 hours at 25°C, and products were separated from free nucleotide sugars using a Sep-Pak C18 cartridge as described (Fritz et al., 1994). Reactions were linear with time and amount of protein.

5. Chondroitin and HS characterization

Between 220 to 250 *sqv-5(n3611)* and wild-type animals were collected as described for glycosyltransferase assays, except the animals were lyophilized and homogenized in acetone. Free GAGs were isolated by alkali extraction as described (Esko, 1993), except that they were extracted overnight in 0.5 M NaOH, 1 M NaBH₄ at 4°C, and neutralized with 1 M HCl. Chondroitin was digested with 20 mU chondroitinase ABC (Seikagaku) and analysed by HPLC with post-column derivatization of the disaccharides (Toyoda et al., 2000).

RNAi was performed essentially as described (Timmons and Fire, 1998), except L1-stage hermaphrodites were placed onto Petri plates containing bacteria and grown for 44-48 hours at 20°C before being collected as L4-stage hermaphrodites. Intact proteoglycans were isolated by anion-exchange chromatography as described (Esko, 1993), except the animals were sonicated in 0.5% Triton X-100 and protease inhibitor cocktail (Sigma). For western blots, intact proteoglycans were digested with chondroitinase ABC or heparin lyase II and incubated with anti-chondroitin mAb (1-B-5, Seikagaku) or anti-HS mAb (F69-3G10, Seikagaku).

6. Anti-SQV-5 antibodies

The *sqv-5* ORF was cloned into vectors pGEX-4T3 and pMAL-c2 to generate GST-SQV-5 and MBP-SQV-5 fusion proteins, respectively. The GST-SQV-5 and MBP-SQV-5 fusion proteins were purified by isolating insoluble proteins from inclusion bodies followed by SDS-PAGE and electro-elution. GST-SQV-5 was

injected into two rabbits (Covance). Anti-SQV-5 antibodies were affinity purified by binding to and eluting from MBP-SQV-5 fusion protein, as described (Hwang and Horvitz, 2002b).

E. Acknowledgements

We thank Beth Castor for help with DNA sequencing, Yuji Kohara for the cDNA clones yk20dy and yk21g9, Alan Coulson for the cosmids, Jillian Brown for GlcA β 1,3Gal-O-NM acceptor, Beverly Zak for N-acetylheparosan acceptor, the Glycotechnology Core at UCSD (supported by NIH grant R24GM61894) for the preparation of UDP-[1-³H]GlcA, and Brendan Galvin and Ignacio Perez de la Cruz for critical reading of this manuscript. This work was supported by NIH grant GM24663 to H.R.H and NIH grant GM33063 to J.D.E. S.K.O. was supported by NIH training grant GM08666. H.R.H. is an Investigator of the Howard Hughes Medical Institute.

F. References

Almeida, R., Levery, S.B., Mandel, U., Kresse, H., Schwientek, T., Bennett, E.P., and Clausen, H. (1999). Cloning and expression of a proteoglycan UDP-galactose:b-xylose b,4-galactosyltransferase I. *J. Biol. Chem.* **274**, 26165-71.

Austin, C. R. *Fertilization* (Prentice Hall, Englewood Cliffs, New Jersey, 1965).

Berninsone, P., Hwang, H. Y., Zemtseva, I., Horvitz, H. R., and Hirschberg, C. B. (2001). SQV-7, a protein involved in *Caenorhabditis elegans* epithelial invagination and early embryogenesis, transports UDP-glucuronic acid, UDP-N-acetylgalactosamine, and UDP-galactose. *Proc. Natl. Acad. Sci. USA* **98**, 3738-43.

Bulik, D.A., Wei, G., Toyoda, H., Kinoshita-Toyoda, A., Waldrip, W.R., Esko, J.D., Robbins, P.W., and Selleck, S.B. (2000). *sqv-3*, *-7*, and *-8*, a set of genes affecting morphogenesis in *Caenorhabditis elegans*, encode enzymes required for glycosaminoglycan biosynthesis. *Proc. Natl. Acad. Sci. USA* **97**, 10838-43.

Comper, W. D. and Laurent, T. C. (1978). Physiological function of connective tissue polysaccharides. *Physiol. Rev.* **58**, 255-315.

Ellis, R. E. and Kimble, J. (1995). The *fog-3* gene and regulation of cell fate in the germ line of *Caenorhabditis elegans*. *Genetics* **139**, 561-77.

Esko, J. D. (1993). In *Current protocols in molecular biology* (ed. Ausubel, F., et al.) 17.2.1-17.2.9 (John Wiley and Sons, Inc., New York).

Esko, J. D. and Selleck, S. B. (2002). ORDER OUT OF CHAOS: Assembly of ligand binding sites in heparan sulfate. *Annu. Rev. Biochem.* **71**, 435-71.

Fritz, T. A., Gabb, M. M., Wei, G., and Esko, J. D. (1994). Two N-acetylglucosaminyltransferases catalyze the biosynthesis of heparan sulfate. *J. Biol. Chem.* **269**, 28809-14.

Herman, T., Hartwieg, E., and Horvitz, H. R. (1999). *sqv* mutants of *Caenorhabditis elegans* are defective in vulval epithelial invagination. *Proc. Natl. Acad. Sci. USA* **96**, 968-73.

Herman, T. and Horvitz, H. R. (1999). Three proteins involved in *Caenorhabditis elegans* vulval invagination are similar to components of a glycosylation pathway. *Proc. Natl. Acad. Sci. USA* **96**, 974-9.

Hwang, H.-Y. and Horvitz, H. R. (2002a). The SQV-1 UDP-glucuronic acid decarboxylase and the SQV-7 nucleotide-sugar transporter may act in the Golgi apparatus to affect *C. elegans* vulval morphogenesis and embryonic development. *Proc. Natl. Acad. Sci. USA* **99**, 14218-14223.

Hwang, H.-Y., and Horvitz, H. R. (2002b). The *C. elegans* vulval morphogenesis gene *sqv-4* encodes a UDP-glucose dehydrogenase that is temporally and spatially regulated. *Proc. Natl. Acad. Sci. USA* **99**, 14224-14229.

Hwang, H.-Y., Olson, S. K., Brown, J. R., Esko, J. D., and Horvitz, H. R. (2003). The *C. elegans* genes *sqv-2* and *sqv-6*, which are involved in vulval morphogenesis, encode glycosaminoglycan galactosyltransferase II and xylosyltransferase. *J. Biol. Chem.* **278**, 11735-8.

Jansen, G., Hazendonk, E., Thijssen, K. L., and Plasterk, R. H. (1997). Reverse genetics by chemical mutagenesis in *Caenorhabditis elegans*. *Nat. Genet.* **17**, 119-21.

Kitagawa, H., Uyama, T., and Sugahara, K. (2001). Molecular cloning and expression of a human chondroitin synthase. *J. Biol. Chem.* **276**, 38721-6.

Lane, M. C., Koehl, M. A., Wilt, F., and Keller, R. (1993). A role for regulated secretion of apical extracellular matrix during epithelial invagination in the sea urchin. *Development* **117**, 1049-60.

Nagasawa, K., Inoue, Y., and Kamata, T. (1977). Solvolytic desulfation of glycosaminoglycuronan sulfates with dimethyl sulfoxide containing water or methanol. *Carbohydr. Res.* **58**, 47-55.

Okajima, T., Fukumoto, S., Furukawa, K., and Urano, T. (1999). Molecular basis for the progeroid variant of Ehlers-Danlos syndrome. *J. Biol. Chem.* **274**, 28841-4.

Perrimon, N. and Bernfield, M. (2000). Specificities of heparan sulphate proteoglycans in developmental processes. *Nature* **404**, 725-8.

Quentin, E., Gladen, A., Roden, L., and Kresse, H. (1990). A genetic defect in the biosynthesis of dermatan sulfate proteoglycan: galactosyltransferase I deficiency in fibroblasts from a patient with a progeroid syndrome. *Proc. Natl. Acad. Sci. USA* **87**, 1342-6.

Ruoslahti, E. (1988). Structure and biology of proteoglycans. *Annu. Rev. Cell Biol.* **4**, 229-55.

Schwartz, N. B. and Domowicz, M. (2002). Chondrodysplasias due to proteoglycan defects. *Glycobiology* **12**, 57R-68R.

Silbert, J. E. and Sugumaran, G. (2002). Biosynthesis of chondroitin/dermatan sulfate. *IUBMB Life* **54**, 177-86.

Timmons, L. and Fire, A. (1998). Specific interference by ingested dsRNA. *Nature* **395**, 854.

Toyoda, H., Kinoshita-Toyoda, A., and Selleck, S. B. (2000). Structural analysis of glycosaminoglycans in *Drosophila* and *Caenorhabditis elegans* and demonstration that *tout-velu*, a *Drosophila* gene related to EXT tumor suppressors, affects heparan sulfate *in vivo*. *J. Biol. Chem.* **275**, 2269-75.

Uyama, T., Kitagawa, H., Tamura Ji, J., and Sugahara, K. (2002). Molecular cloning and expression of human chondroitin N-acetylgalactosaminyltransferase. *J. Biol. Chem.* **277**, 8841-6.

Wei, G., Bai, X., Gabb, M.M., Bame, K.J., Koshy, T.I., Spear, P.G., and Esko, J.D. (2000). Location of the glucuronosyltransferase domain in the heparan sulfate

copolymerase EXT1 by analysis of Chinese hamster ovary cell mutants. *J. Biol. Chem.* **275**, 27733-40.

Yamada, S., Van Die, I., Van den Eijnden, D.H., Yokota, A., Kitagawa, H., and Sugahara, K. (1999). Demonstration of glycosaminoglycans in *Caenorhabditis elegans*. *FEBS Lett.* **459**, 327-31.

Zak, B. M., Crawford, B. E., and Esko, J. D. (2002). Hereditary multiple exostoses and heparan sulfate polymerization. *Biochim Biophys Acta* **1573**, 346-55.

The text of Chapter 3 is a reprint of the material as it appeared in the journal Nature.

Hwang, H.Y., Olson, S.K., Esko, J.D., and Horvitz, H.R. (2003). The dissertation author was a secondary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

CHAPTER 4

Novel chondroitin proteoglycans mediate early embryonic cell division in *Caenorhabditis elegans*

A. Summary

Vertebrates express several types of sulfated chondroitin chains covalently bound to the core proteins of proteoglycans. Lower organisms like *C. elegans* also express chondroitin, but the chains lack sulfate groups and none of the core proteins have been identified. *In silico* analysis of the *C. elegans* genome did not reveal any obvious homologs of mammalian chondroitin sulfate proteoglycans. Here, we have employed a biochemical purification scheme, western blotting, and mass spectrometry that led to the identification of nine novel chondroitin proteoglycan core proteins in *C. elegans*, none of which are present in the mammalian genome. Recombinant protein expression in mammalian cells demonstrated that six of the core proteins can carry chondroitin sulfate chains. Single gene RNAi depletion experiments showed no effect on embryonic viability or tissue morphogenesis, but simultaneous silencing of two of the proteoglycans (*cpg-1/cej-1* and *cpg-2*) resulted in dead multinucleated single-cell embryos. This embryonic lethal phenotype resembles that of *squashed vulva (sqv)* mutants defective in chondroitin assembly, suggesting that the chondroitin chains on these two proteoglycans are required for cytokinesis.

B. Introduction

Chondroitin sulfate proteoglycans (CSPGs) consist of a protein core and one or more covalently attached glycosaminoglycan chains. The chains assemble step-wise while attached to specific serine residues of the protein cores, starting with the synthesis of a tetrasaccharide of xylose, galactose and glucuronic acid (-GlcA β 3Gal β 3Gal β 4Xyl β -O-Ser) and ending with alternating addition of N-acetylgalactosamine (GalNAc) and GlcA units ([GlcA β 3GalNAc β 4]_n) (Roden et al., 1972). In mammals, chondroitin chains undergo further modification, in which sulfate residues are added, and in dermatan sulfate, a portion of GlcA residues undergo epimerization to L-iduronic acid (IdoA) followed by sulfation. The presence of sulfate groups and charged carboxyl groups of GlcA and IdoA units allows the chains to interact with extracellular matrix proteins (e.g. fibrillar collagens) (Iozzo and Murdoch, 1996) and growth factors (e.g. FGF) (Sugahara et al., 2003). Furthermore, secreted chondroitin sulfate proteoglycans create a hydrated matrix allowing for tissue expansion and the capacity to absorb compressive loading, e.g. in cartilage. The different proteoglycan core proteins help organize the extracellular matrix and determine the concentration and distribution of the chondroitin sulfate chains.

Vertebrates make more than 20 CSPGs based on the primary sequence of the core protein (Olson and Esko, 2004). Interestingly, very little is known about CPGs present in invertebrates, even in the well studied nematode, *C. elegans* (Schimpf et al., 1999; Yamada et al., 1999; Bulik et al., 2000; Toyoda et al., 2000; Berninsone et al., 2001; Beeber and Kieras, 2002). Nematodes make chondroitin chains in a manner

identical to vertebrates (Yamada et al., 2002; Hwang and Horvitz, 2002a; Hwang and Horvitz, 2002b; Hwang et al., 2003a; Hwang et al., 2003b; Izumikawa et al., 2004), but lack the sulfotransferases that add sulfate to GalNAc and uronic acids and the epimerase that converts GlcA to IdoA. In spite of the simplicity of the chains, genetic experiments demonstrate a crucial role for chondroitin in embryonic cell division and vulval morphogenesis (Hwang and Horvitz, 2002a; Hwang and Horvitz, 2002b; Hwang et al., 2003a; Hwang et al., 2003b; Izumikawa et al., 2004; Herman et al., 1999; Bulik and Robbins, 2002; Mizuguchi et al., 2003). A genetic screen for mutations in vulval development (*squashed vulva*, *sqv*) led to the discovery that all of the components of the chondroitin biosynthetic machinery are completely conserved. Additionally, progeny of *sqv* mutant animals showed maternal effect lethality due to defective cytokinesis of single-celled embryos during the first cell division (Hwang and Horvitz, 2002a; Herman et al., 1999; Mizuguchi et al., 2003). Although all of the enzymes required for chondroitin synthesis were identified in the *Sqv* screen, no protein cores that harbor chondroitin chains were found.

To identify chondroitin proteoglycans (CPGs) in *C. elegans*, we conducted BLAST searches with known mammalian CSPG core protein sequences but failed to identify any obvious homologs. We therefore pursued a biochemical approach taking advantage of the chemical properties of the long, negatively charged chondroitin chains. Western blotting using antibodies that recognize stubs of the chondroitin chains (Christner et al., 1980), a tagging method to modify glycosylation sites on the proteins (Wells et al., 2002), and mass spectrometry (Washburn et al., 2002)

uncovered nine novel CPG proteins, none of which show homology to mammalian CSPGs. Characterization of two of these proteins, CPG-1/CEJ-1 and CPG-2, shows that they function redundantly in early embryonic development, like *sgv-5* chondroitin synthase.

C. Materials and Methods

1. *C. elegans* maintenance

The OD58 strain (a generous gift from Anjon Audhya, University of California, San Diego) carries a plextrin-homology (PH) membrane domain of phospholipase-C 1 α fused in-frame to a GFP reporter (*unc119; pAA1 pie-1::GFP-PH (PLC1 α PH); unc-119*) under the control of the *pie-1* promoter. All other strains were obtained from the CGC stock center (St. Paul, MN) and cultivated as described (Brenner, 1974).

2. *In silico* analysis

BLAST searches were performed in the NCBI database (<http://www.ncbi.nlm.nih.gov>) or WormBase (<http://wormbase.org>) using annotated human or mouse CSPG core protein sequences. The presence of signal peptides was confirmed with PSORT II (<http://psort.nibb.ac.jp/form2.html>) and SignalP 3.0 Server (<http://www.cbs.dtu.dk/services/SignalP>).

3. Biochemical Purification

Worm extracts were prepared from 20 g batches of a mixed stage N2 (Bristol) population by sonication in 50 mM sodium acetate buffer, pH 6. Material was extracted for 48 hr at 4°C in 3 volumes of solution containing 4 M guanidine-HCl, 0.1 M NaCl, 0.3% 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonic acid (CHAPS), 50 mM sodium acetate, pH 6, and protease inhibitors (1 µg/ml leupeptin, 1 mM PMSF, and 1 µg/ml pepstatin A) (Esko, 1993). The extract was dialyzed against 6 M urea, 0.1 M NaCl, 50 mM sodium acetate, pH 6, with 3 changes of buffer. Insoluble material was removed by low speed centrifugation and filtration through Whatman No. 1 filter paper. The concentration of protein was assayed by the Bradford method (BioRad).

Worm extract (250 mg protein) was purified by anion exchange chromatography (DEAE-Sephacel, Pharmacia) and desalted by gel filtration (PD-10, Pharmacia) as described (Bame and Esko, 1989). Partially purified material was treated with trypsin (Washburn et al., 2002). Following proteolysis, glycopeptides were further purified either by an additional pass over DEAE-Sephacel as described above or by gel filtration HPLC (TSK-2000 column, TOSOH Biosciences) using a buffer of 1 M NaCl in 10 mM KH₂PO₄, pH 6. Uronic acids in glycosaminoglycans were quantitated by the carbazole method (Esko and Manzi, 1996). Fractions containing glycosaminoglycan were pooled and desalted by PD-10 chromatography.

Glycosaminoglycan chains were removed by the β-elimination followed by Michael addition with dithiothreitol (DTT) method (BEMAD, (Wells et al., 2002)).

Briefly, glycopeptides were incubated for 3 hr at 50°C in 20% ethanol, 1% triethylamine, 10 mM DTT, 0.1% NaOH. The reaction was quenched by adjusting the sample to 0.1% trifluoroacetic acid (Sigma) and the peptides were purified by reverse-phase chromatography on a C18 Sep-Pak cartridge (Waters Corporation) using 70% acetonitrile in 0.1% trifluoroacetic acid. Samples were dried before further analysis.

4. Mass Spectrometry

Tagged peptides were identified by a multi-dimensional peptide identification technique (MudPIT) (Washburn et al., 2002; Wolters et al., 2001; Washburn et al., 2001; MacCoss et al., 2002). Peptides were eluted stepwise from a bi-phasic capillary column made of strong cation exchange resin coupled to a reverse phase resin directly into a tandem mass spectrometer. The identities of the DTT-tagged peptides were determined by searching the tandem mass spectra against a *C. elegans* proteome database using SEQUEST software and a computer array. The unique mass signature imparted by DTT (+167 Da) and unmodified dehydroalanine residues (-18 Da) were used to determine the sites of glycosylation (Wells et al., 2002).

5. Recombinant Protein Expression

C. elegans cDNA was prepared from total RNA with the SuperScript III First-Strand kit (Invitrogen). *cpg-1* through *cpg-9* were amplified from *C. elegans* cDNA with primers that included the start codon and the penultimate codon and restriction

sites to subclone the products in-frame into the pcDNA3.1(-)MycHis B vector (Invitrogen). Expression constructs were transfected into COS-7 cells with Lipofectamine (Invitrogen) following the manufacturer's instructions. Media was harvested 48 hr later and purified over DEAE-Sephacel as described (Bame and Esko, 1989). COS-7 cells were cultured in Dulbecco's Modified Eagle's Medium (CellGro) supplemented with 10% fetal bovine serum, 100 U/ml penicillin G, and 100 µg/ml streptomycin sulfate.

6. RNAi

C. elegans cDNA was amplified by PCR with primers engineered to contain T7 (forward primer) or T3 (reverse primer) bacterial promoter sequence and base pairs 1-169 for *cpg-1*, and 1-775 and 754-1572 for *cpg-2*. dsRNA was generated with the Megascript T7 and T3 transcription kits (Ambion) according to the manufacturer's instructions. OD58 L4 or young adult worms were injected with dsRNA and allowed to recover at 16°C or 20°C. To count brood sizes, worms were transferred to individual plates 24 hr post-injection (Maddox et al., 2005). After 24 hr, the number of eggs and hatched L1 larvae were counted as a measure of brood size. Viability was measured 24-36 hr later by counting the number of hatched larvae and unhatched embryos. Percent viability was calculated as number of hatched progeny divided by total number of eggs laid per 24 hr period.

Early embryonic cell division was assessed 24 hr post-injection. Embryos were filmed *in utero* since *cpg-1/cpg-2(RNAi)* and *sqv-5(RNAi)* embryos were fragile and osmotically sensitive. Injected animals were anesthetized with 1 mM

levamisole in M9 buffer, mounted on an agarose pad, and filmed as described (Maddox et al., 2005).

To determine the effect of RNAi on proteoglycan expression, cDNA sequences described above for *cpg-1* and *cpg-2* were subcloned into vector pL4440, which carries dual T7 promoter sites (Fire Lab Vector Kit) and drives the formation of dsRNA. The expression vector was transformed into the HT115 bacterial strain, which was then fed to the worms as described (Kamath et al., 2001). After 48-60 hr, protein extracts were prepared by sonication of either whole worms or embryos and subjected to SDS-PAGE and Western blotting as described below.

7. Western Blotting

Protein was digested with 1 mU chondroitinase ABC (Seikagaku) and/or 1 mU heparin lyase II (Sigma) for 3-5 hr at 37°C and analyzed by SDS-PAGE after reduction with β -mercaptoethanol and alkylation with iodoacetamide. Recombinant proteins expressed in animal cells were Western blotted with either an anti-Myc mAb (1:5000 dilution, Invitrogen) or the 1-B-5 chondroitin stub mAb (1:1000, Seikagaku) followed by goat anti-mouse secondary antibody (1:2000, BioRad). Blots were visualized with a WestPico Chemiluminescent Kit (Pierce).

D. Results

1. *C. elegans* expresses multiple novel CPGs

To identify CPGs in *C. elegans*, BLAST and PSI BLAST searches using all known mammalian CSPG sequences against the *C. elegans* genome were performed.

No obvious homologs emerged from this analysis, with the exception of bamacan/SMC3 (Y47D3A.26). However, the coding sequence for bamacan lacks a signal peptide present in all proteoglycans, consistent with the idea that most of the protein is found in the nucleus (Ghiselli and Iozzo, 2000). To identify *C. elegans* CPGs, worm extracts were digested with chondroitinase ABC, analyzed by SDS-PAGE after reduction and alkylation, followed by Western blotting using the 1-B-5 mAb, which recognizes a neo-epitope generated by chondroitinase digestion (Fig. 1). Numerous protein bands were detected in whole worm (Lanes 2-3) and embryo extracts (Lane 4), and some of the bands in the embryos were unique. Omission of chondroitinase ABC did not yield any of the bands (Lane 1). Digestion with the combination of chondroitinase ABC and heparin lyase II produced an identical pattern, suggesting that none of the chondroitin bearing proteoglycans existed as hybrid molecules containing both heparan sulfate and chondroitin chains (data not shown). Treatment of samples with PNGase F, which removes Asn-linked glycans, also had no effect on the pattern (data not shown). The mass of the individual proteoglycans varied from ~10 kDa to >200 kDa.

To identify the individual proteoglycan core proteins, CPGs were purified from adult worms by anion exchange chromatography using DEAE-Sephacel (0.2 M – 1 M NaCl). Samples were reduced, alkylated, and treated with trypsin. The glycopeptides were repurified by a second round of anion-exchange chromatography. In some experiments, the glycopeptides were further purified by gel filtration chromatography, which took advantage of the large hydrodynamic

volume imparted by the chondroitin chains to separate the glycopeptides from contaminating peptides rich in acidic amino acids. Treatment of the glycopeptides with alkali resulted in β -elimination of the chains. The resulting free polysaccharides eluted from a gel filtration column near the void volume based on carbazole reaction (Fig. 2). This material was entirely sensitive to chondroitinase ABC, which digests chondroitin into disaccharides. Thus, the glycopeptide fraction obtained in this way contained very little heparan sulfate or Asn-linked oligosaccharides.

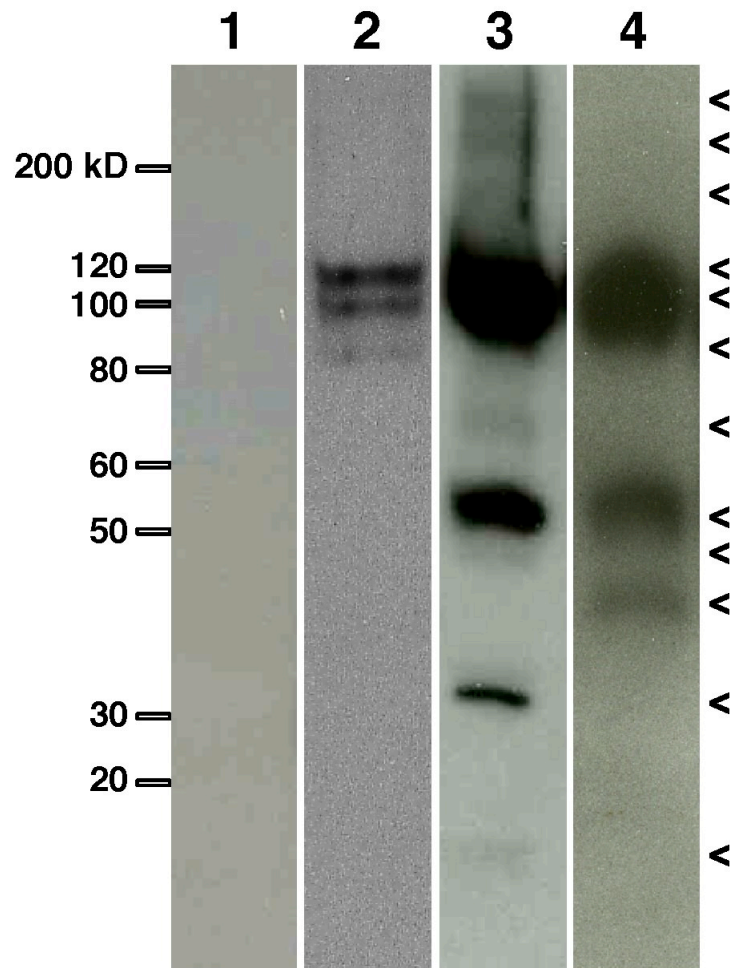


Figure 1. *C. elegans* expresses multiple CPGs

Crude extracts from whole worms (lanes 1-3) or embryos (lane 4) were digested with chondroitinase ABC (lanes 2-4) and separated by SDS-PAGE. Western blotting was carried out with mAb 1-B-5, which recognizes a neo-epitope generated by chondroitinase digestion. The undigested control (lane 1) demonstrates the specificity of the antibody for chondroitin stubs. Lane 2 and 4, 10 µg protein; lane 1 and 3, 40 µg protein. The carats indicate the position of protein bands that arose only after chondroitinase ABC digestion.

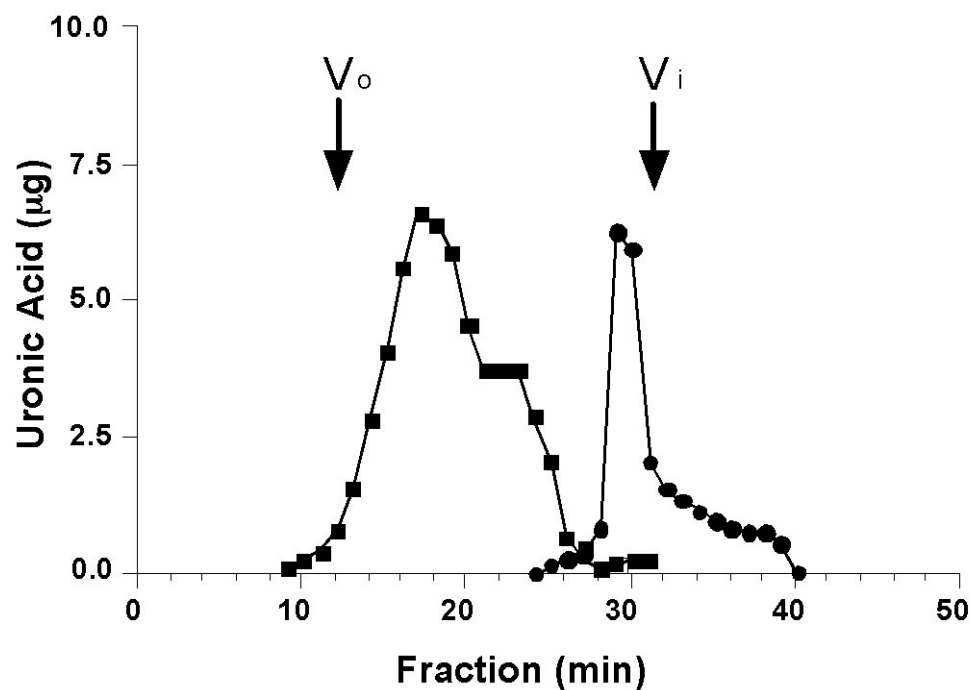


Figure 2. Material released by BEMAD is ABC-sensitive

Material released by β -elimination was analyzed by gel filtration HPLC and carbazole reaction to measure uronic acids (Materials and Methods). Untreated sample eluted near the void volume (V_o , elution position of blue dextran) (●). Sample digested with chondroitinase ABC prior to HPLC analysis eluted near the included volume (V_i , elution position of [$1\text{-}^3\text{H}$]glucose) where disaccharides elute (■).

β -elimination results in the formation of dehydroalanine from Ser, which is susceptible to Michael addition with a nucleophile like dithiothreitol (DTT) (Wells et al., 2002). This reaction was inefficient, resulting in some peptides containing DTT (mass increase of 167 Da) and some containing dehydroalanine (loss of 18 Da). Peptides that fit these mass criteria were then used to search the *C. elegans* proteome database using SEQUEST software. In some cases peptides were found in which cleavage had occurred either proximal to or directly at putative Ser attachment sites, possibly due to β -elimination or susceptibility of the bonds near the glycosylation site to ionization in the mass spectrometer. Additional selection criteria required that the putative proteoglycan contain a hydrophobic signal peptide, which would direct the protein into the secretory pathway where glycosylation occurs. Additionally, each protein had to contain at least one putative glycosaminoglycan attachment consensus sequence, which consists of a Ser residue flanked on its C-terminal side by Gly and one or more Asp or Glu residues near the glycosylation site (Esko and Zhang, 1996). Nine independent purifications of *C. elegans* CPGs were analyzed in this way, which yielded nine putative CPG core proteins designated CPG-1 through CPG-9 (Fig. 3).

CPG-1 (C07G2.1a) and CPG-2 (B0280.5) have predicted masses of 62 and 54 kDa, respectively (Fig. 3). CPG-1 was identified previously as CEJ-1 based on its cross reactivity with an antibody to a mammalian tight junction protein. It contains five putative glycosylation sites (1 confirmed by mass spectrometry), whereas CPG-2 has as many as 34 sites (4 confirmed by mass spectrometry). CPG-1 and CPG-2

also contain 3 or 6 peritrophin-A chitin-binding motifs, respectively, defined by the arrangement of six cysteine residues that can form three disulfide bridges in a characteristic pattern (Wright et al., 1991; Beintema, 1994; Venegas et al., 1996; Merzendorfer and Zimoch, 2003). CPG-1 contains two chitin-binding domains in the N-terminal half of the protein and one in the C-terminus, whereas in CPG-2 the chitin-binding domains lie between sites predicted to be glycosylated. Additionally, the C-terminal half of CPG-1 contained many Thr, Val and Pro residues, resembling a segment of Muc-2, a membrane mucin present in vertebrates (Fig. 5A). CPG-3 (R06C7.4) has a predicted mass of 30 kDa and potentially carries 15 chondroitin chains, but exhibits no other conserved protein folds. It shows 21% identity to CPG-2, mainly in the putative glycosylated region. No glycosylation sites were identified by tagging, as the recovered peptides flanked the putative glycosaminoglycan attachment sites but did not include them. This finding suggested that either the modified peptide was weakened at these sites during MS/MS fragmentation, or that β -elimination resulted in peptide cleavage at the glycosylation sites. Expression of *cpg-1*, *cpg-2*, and *cpg-3* is enriched in the *C. elegans* germline, suggesting possible activity in germline development and progression, spermatogenesis, oogenesis, or embryogenesis (Reinke et al., 2000).

CPG-4 (C10F3.1) has the largest predicted mass of the *C. elegans* CPGs at 84 kDa and has the majority of its 35 predicted glycosaminoglycan attachment sites in the C-terminal half of the protein. CPG-5 (C25A1.8) and CPG-6 (K10B2.3a) are highly related, showing 67% identity and 90% similarity to each other. CPG-5 and

CPG-6 contain only one putative chondroitin site towards the N-terminus and a C-type lectin domain in the C-terminal half of the protein, like the aggrecan family of vertebrate CSPGs (Zelensky and Gready, 2003). *cpg-5* and *cpg-6* transcripts are also enriched in the germline.

CPG-7 and CPG-8 have predicted masses of only ~12 kDa. CPG-7 contains 11 putative glycosylation sites, five of which were confirmed by mass spectrometry, whereas CPG-8 contained 6 putative glycosylation sites, 5 confirmed by mass spectrometry. Of the multiple peptides identified in CPG-8 samples, some contained a single modified site at either residue 61 or 63, while others were modified at both positions. The same was true of residues 81 and 88. It is not known whether site-specific glycosylation occurred in a tissue-specific or developmentally regulated manner. CPG-9 was only 7 kDa. Two of four glycosylation sites were confirmed by mass spectrometry.




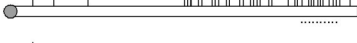
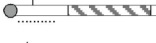
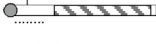



Protein Name (Accession Number)	Independent Verification (9)	Predicted (Apparent) Molecular Weight (kD)	Putative (Identified) GAG Attachment Sites	Germline Enriched?	Protein Structure
CPG-1 (CEJ-1 (C07G2.1a))	2	62 (150)	5 (1)	Yes	
CPG-2 (B0280.5)	6	54 (100-120)	34 (4)	Yes	
CPG-3 (R06C7.4)	7	30 (60)	15 (0)	Yes	
CPG-4 (C10F3.1)	4	84 (220)	35 (4)	ND	
CPG-5 (C25A1.8)	3	27 (50)	1 (1)	Yes	
CPG-6 (K10B2.3a)	2	28 (40)	1 (1)	Yes	
CPG-7 (K09F4.6)	1	12 (ND)	11 (6)	ND	
CPG-8 (K03B4.7a)	5	12 (ND)	6 (5)	No	
CPG-9 (Y67D8C.8)	5	7 (ND)	4 (2)	ND	

Figure 3. *C. elegans* CPGs identified by mass spectrometry

CPGs were purified from worm extracts and identified by mass spectrometry analysis. Nine independent purifications resulted in identification of nine CPGs. The predicted sizes are based on amino acid sequence, and the apparent sizes are based on SDS-PAGE migration of recombinant proteins expressed in COS-7 cells after digestion with chondroitinase ABC and reduction. Putative glycosylation sites consist of Ser-Gly dipeptides flanked by one or more acidic amino acids. Identified sites are serine residues modified with DTT by the BEMAD method (Materials and Methods). Genes enriched for germline expression were identified by Reinke et al. (2000). ND = not determined. Schematic drawings of each CPG are shown. Gray circles indicate signal peptides, black ovals are peritrophin-A chitin binding domains, diagonally hatched boxes identify C-type lectin domains, and dotted lines below the protein indicate peptide coverage discovered by mass spectrometry. Short vertical lines identify putative Ser-Gly glycosylation sites, whereas the tall vertical lines show Ser-Gly sites modified with DTT.

To confirm that the *C. elegans* CPGs identified by mass spectrometry could serve as a scaffold for chondroitin synthesis, the cDNA sequences were cloned into a mammalian expression vector with a C-terminal Myc-tag and expressed in mammalian COS-7 cells. Conditioned media was harvested 48 hr post-transfection, purified by anion exchange chromatography, digested with chondroitinase ABC, separated by SDS-PAGE and reacted with an antibody to Myc. CPG1-6 yielded reactive protein bands that were not present when chondroitinase ABC digestion was omitted (Fig. 4). Attempts to express CPG7-9 have thus far been unsuccessful, possibly due to the small size of the protein core. Interestingly, several of the proteoglycans did not migrate at the molecular mass predicted by their primary sequence, suggesting tertiary or quaternary structures or that the chondroitin stub oligosaccharides altered their migration. Syndecan proteoglycans expressed by animal cells also migrate aberrantly, often at twice their predicted size (Bellin et al., 2002).

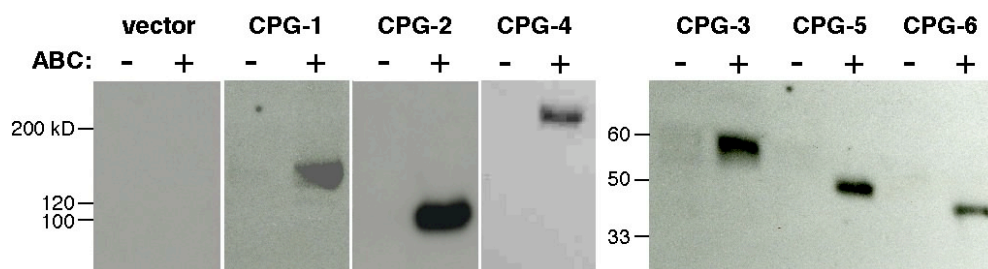


Figure 4. CPG-1 through CPG-6 are CPGs

CPG-1 through CPG-6 were expressed as myc-tagged recombinant proteins in COS-7 cells. Conditioned media was purified over DEAE-Sephacel. Bound material was digested with chondroitinase ABC (+), or left untreated (-), and separated by SDS-PAGE. Western blotting was performed with an anti-myc mAb.

2. CPG-1 and CPG-2

Genome-wide RNAi screens have not identified phenotypes for any of the *cpg* genes, suggesting either functional redundancy or subtle phenotypes (www.wormbase.org). However, simultaneous RNAi inhibition of *cpg-1/cej-1* and *cpg-2* resulted in early embryonic lethality, with accumulation of multi-nucleated single-cell embryos suggesting a defect in cytokinesis (Lee and Schedl, 2001). Because this phenotype resembles *sqv-5(n3611)* chondroitin synthase mutants or wildtype worms treated with *sqv-5* RNAi, we decided to analyze these proteoglycans in greater detail.

To examine whether CPG-1 and CPG-2 behaved as CPGs *in vivo*, worms were fed bacteria expressing dsRNA directed against *cpg-1* or *cpg-2* mRNA and samples were analyzed by Western blotting. Western blotting of *cpg-1* depleted extracts did not show any differences from extracts of worms fed on empty vector, suggesting that CPG-1 may have been expressed at low levels or that additional modifications to the chain (e.g., O-linked oligosaccharides in the mucin-like domain) obscured its resolution by SDS-PAGE (Fig. 5C). When expressed in COS-7 cells, CPG-1 migrated as ~150 kDa protein, but worms did not show an obvious band at this M_r . In contrast, *cpg-2* depleted extracts showed a striking reduction of the major band at 100-120 kDa, while the other major core protein band at 60 kDa remained unaltered (Fig. 5C). Reduction of the bands at ~50 kDa (the predicted mass of the protein was 54 kDa, Fig. 3) and at ~85 kDa was also observed, suggesting that most of the protein bore multiple chondroitin chains that shifted the M_r to higher values. Similar

results were observed in extracts generated from eggs of dsRNA-fed animals, suggesting that embryos also express CPG-2 (data not shown). Extracts depleted of both *cpg-1* and *cpg-2* resemble those depleted of *cpg-2* alone. These findings confirmed that CPG-2 was expressed as a CPG in worms, but left open the question as to extent of modification of CPG-1. Both CPG-1 and CPG-2 have the hallmarks of a proteoglycan based on the requisite Gly residue C-terminal to each Ser attachment site and flanking Asp or Glu residues (Fig. 5A,B). The observation that CPG-1 was tagged by DTT addition and was expressed efficiently as a CSPG in COS-7 cells suggest that it most likely contains at least one chondroitin chain *in vivo*.

Figure 5. CPG-1 and CPG-2 protein sequences resemble proteoglycans, and CPG-2 acts as a CPG *in vivo*.

Amino acid sequences of CPG-1 (A) and CPG-2 (B) are shown. The signal peptide is indicated by a solid underline. Peptide sequences identified by mass spectrometry are designated by dashed lines. Putative glycosylation sites are identified by an asterisk (*). Utilized glycosylation sites, as determined by DTT addition, are boxed. Numbers indicate amino acid position. (C) Western blot of extract from worms fed with the indicated dsRNA. Samples were digested with chondroitinase ABC and blotted with the 1-B-5 mAb that recognizes the chondroitin stub remaining after enzyme digestion. Numbers on the left side of the gel indicate molecular weights (kDa).

3. CPG-1 and CPG-2 are required for cytokinesis and embryonic viability

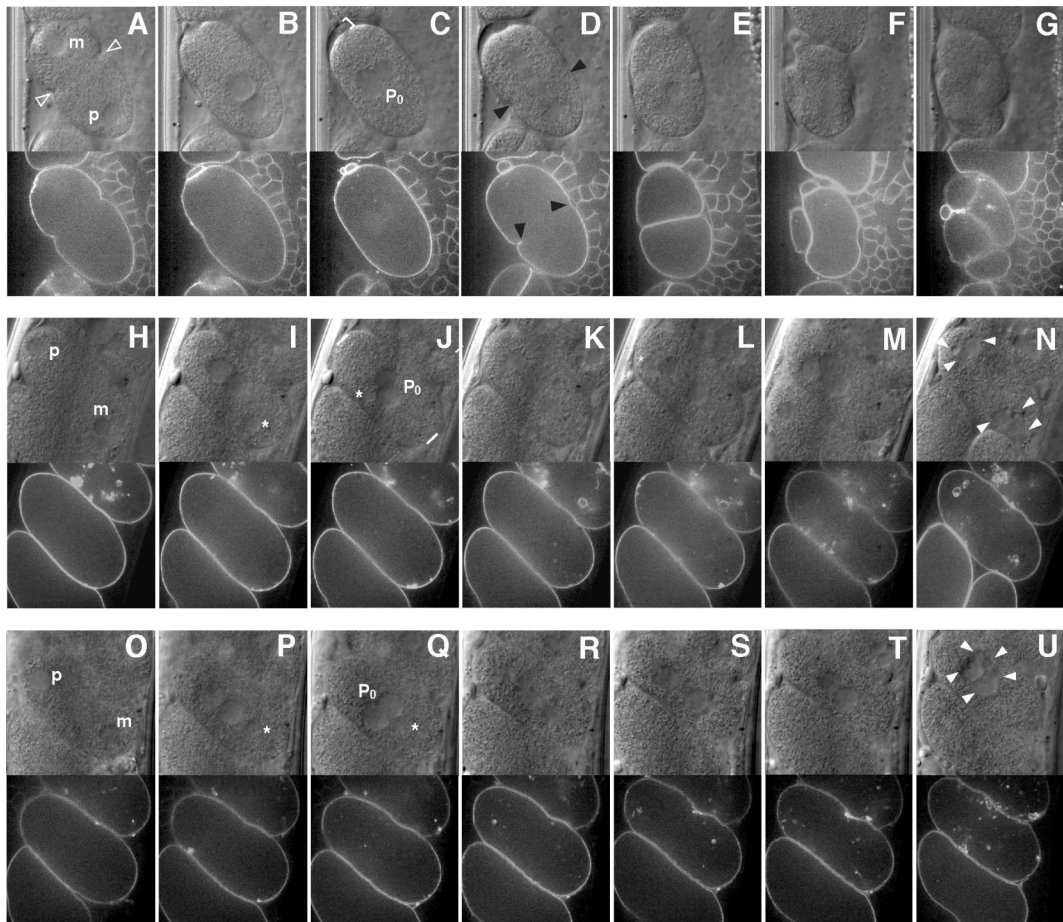
To further understand the role of CPG-1 and CPG-2 in embryonic development, we looked more closely at the first cell division where defects have been reported in *sqv-5* mutants (Hwang et al., 2003b). The OD58 strain was used in these studies since it expresses in the germline a GFP-tagged phospholipase-C plextrin homology domain, which binds to PIP₃, thus rendering the plasma membrane fluorescent green. L4 larvae or young adults were injected with dsRNA targeting *cpg-1*, *cpg-2*, *cpg-1/cpg-2*, or *sqv-5* and analyzed by Nomarski DIC microscopy (Fig. 6, upper panels) or fluorescence microscopy (Fig. 6, lower panels). Depletion of *cpg-1* or *cpg-2* alone had no effect on fertilization, membrane ruffling, pseudocleavage, pronuclear meeting and rotation, karyokinesis, and cytokinesis. Initiation of the cleavage furrow was evident (Fig. 6D, bottom panel), and two separate daughter cells were routinely seen (Fig. 6E). The embryo rotated during filming in the example, but a four-celled embryo was distinctly seen at the end, demonstrating the fidelity of the second round of cell division (Fig. 6G).

Simultaneous depletion of *cpg-1* and *cpg-2* resulted in severe perturbations in embryonic development. Fertilization, pronuclear meeting and rotation occurred normally, but polar body extrusion failed and the nuclear content fused with that of the pronucleus (Fig. 6I,J). Additionally, membrane ruffling normally seen in wild-type embryos was absent, as was the space between the embryonic plasma membrane and the eggshell present during normal cell division (compare Fig. 6C, bracket and Fig. 6J, solid white lines). Spindle formation appeared normal and the

chromosomes attempted segregation during nuclear division, but the cleavage furrow failed to initiate (Fig. 6K,L) and cytokinesis stalled (Fig 6L). Without the barrier of a new cell membrane, the daughter nuclei fused (Fig. 6M) and repeatedly attempted cell division without cytokinesis (Fig. 6N). These data are similar to that observed in *sqv-5(RNAi)* experiments (Fig. 6O-U, also see (Hwang et al., 2003b)).

Figure 6. CPG-1 and CPG-2 are required for embryonic cell division.

The first two cell divisions of embryos depleted of *cpg-2* (A-G), *cpg-1/cpg-2* (H-N), and *sqv-5* (O-U) were analyzed by DIC Nomarski microscopy (upper panels) and membrane-GFP fluorescence (lower panels). (A, H, O) Pseudocleavage initiates in the *cpg-1* single RNAi (outlined arrowheads), but not in the *cpg-1/cpg-2* double or *sqv-5* RNAi embryos. The maternal (m) and paternal (p) pronuclei are indicated. (B, I, P) Pronuclear fusion proceeds normally in single, double, and *sqv-5* RNAi embryos. However, extrusion of the second polar body fails in doubly depleted and *sqv-5* embryos (*). (C, J, Q) Rotation of the fused nuclei. (D, K, R) Initiation of the first cell division. The cleavage furrow is seen in *cpg-1(RNAi)* embryos (black arrowheads). (E, L, S) Cytokinesis is complete in *cpg-1(RNAi)* embryos, forming two daughter cells. Nuclear division looks normal in *cpg-1/cpg-2(RNAi)* and *sqv-5* embryos but cytokinesis did not initiate, causing the daughter nuclei and polar body to rejoin (M, T). Cell division is attempted again (N, U) resulting in multiple nuclei in a single-celled embryo (white arrowheads). The second cell division (F, G) proceeds normally in *cpg-1(RNAi)* embryos.



Singly- and doubly-injected animals had brood sizes comparable to uninjected worms, as well as worms injected with buffer alone (Table 1). *cpg-2(RNAi)* and *sqv-5(RNAi)* worms had slightly smaller average brood sizes, but this difference was not statistically significant ($P > 0.05$ by one-way ANOVA). RNAi depletion of *cpg-1* or *cpg-2* alone also had no effect on viability, since 96-99% of the embryos hatched into healthy larvae that grew into normal fertile adult animals (Table 1). However, simultaneous depletion of *cpg-1* and *cpg-2* had a synergistic effect, resulting in failure of all embryos to hatch. *sqv-5(RNAi)* treatment also had no effect on egg production or laying, but had a similar hatching phenotype to *cpg-1/cpg-2(RNAi)* (95% of progeny failed to hatch). A few *sqv-5(RNAi)* escapers hatched and survived to the adult stage, but these animals failed to produce viable progeny (data not shown). Heterozygous *sqv-5(n3611) +/-* animals injected with *cpg-1* or *cpg-2* dsRNA had similar brood sizes and embryonic viability as their buffer injected counterparts (data not shown), suggesting that chondroitin levels were not reduced enough in the heterozygote to see an effect after depletion of a single CPG. *cpg-1/cpg-2(RNAi)* and *sqv-5(RNAi)* embryos were refractory due to buildup of multiple nuclei (data not shown). Together, these results show that inhibiting expression of CPG-1 and CPG-2 has the same biological effect as removing the SQV-5 chondroitin synthase.

Table 1. 24-hour brood size and viability in RNAi-treated animals

Number of embryos and hatched L1 larvae was assessed 24 hours after placing control worms or worms injected with dsRNA on individual plates. Viability was determined 24+ hours after the initial brood size was counted, and is calculated as number of hatched larvae divided by total number of embryos laid. n=6-11.

RNAi Treatment	24 hr Brood Size*	SD	Range	Viability
Uninjected	133	43	75-197	98.9%
Buffer	120	45	65-178	99.6%
<i>cpg-1</i>	95	34	64-184	95.9%
<i>cpg-2</i>	144	42	87-202	99.2%
<i>cpg-1/cpg-2</i>	129	56	62-203	0.0%
<i>sqv-5</i>	77	22	22-114	4.8%

* One-way ANOVA analysis showed no significant difference between brood sizes. $P > 0.05$

E. Discussion

1. A proteomic approach for identifying novel proteoglycans

We report here the identification of a novel class of CPGs in *C. elegans*. The identification scheme employed a combination of conventional methods to purify proteoglycans and glycopeptides based on the large mass and polyanionic character of the chondroitin chains coupled with Western blotting of core proteins after chondroitinase digestion. The additional step of tagging glycosylation sites after β -elimination, mass spectrometry, and the requirement that all hits had to have characteristic sequences flanking the putative sites and a signal peptide provided adequate restrictions to ensure that the majority of identified proteins had the properties of a proteoglycan. Although DTT addition was not stoichiometric, the β -elimination step introduced an alternate method to identify the protein via the dehydration of Ser to dehydroalanine (Wells et al., 2002). Attempts to optimize the DTT-addition reaction have not yet yielded higher efficiency, and other methods are needed to improve this step so that more detailed information about the sites of attachment of the chondroitin chains and relative stoichiometry might emerge. Regardless of its limitations, the current protocol led to the identification of at least nine proteins that had the hallmarks of CPGs, and six served as substrates for chondroitin sulfate assembly when expressed in mammalian cells. This method should be generally applicable to other organisms, tissues, isolated cells, and secretions for which little information about the proteoglycan composition exists.

It should also be possible to adapt the current method for the detection of heparan sulfate proteoglycans. Heparan sulfate proteoglycans were not detected in this study, probably because worms express ~250 times more chondroitin sulfate than heparan sulfate (Yamada et al., 1999; Toyoda et al., 2000). However, genomic analysis of *C. elegans* using vertebrate heparan sulfate proteoglycan core protein sequences revealed homologs of the membrane proteoglycans syndecan and glypican, as well as the secreted proteoglycans agrin, perlecan and collagen XVIII. Vertebrates also express multiple membrane CSPGs (e.g., NG2, CD44, and phosphacan) (Olson and Esko, 2004). However, all of the *C. elegans* CPGs detected to date appear to be secretory proteoglycans since they lacked any predicted membrane spanning segments or consensus sites for the attachment of glycosylphosphatidylinositol anchors.

The lack of homology between *C. elegans* CPGs and vertebrate CSPGs raises interesting questions about the evolution of these molecules. All core proteins regardless of source contain consensus sequence motifs required for initiation of glycosaminoglycan biosynthesis (Esko and Zhang, 1996). Furthermore, *C. elegans* expresses orthologs of all of the vertebrate enzymes required for assembly of the linkage region tetrasaccharide (xylosyltransferase, galactosyltransferases I and II, and glucuronyltransferase I, (Bulik et al., 2000; Hwang et al., 2003a)) as well as the chondroitin polymerizing system (Hwang et al., 2003b; Mizuguchi et al., 2003), including the newly discovered chondroitin polymerizing factor (Izumikawa et al., 2004). These findings indicate that the mechanism of chain initiation and

polymerization evolved early in metazoans and has been maintained by strong selection, whereas the proteins on which the chains assemble continue to evolve to serve specialized functions. Thus, CPG-1 and CPG-2 have binding domains that can interact with chitin in the nematode eggshell, whereas no vertebrate proteoglycans contain this motif. The presence of chitin in insect exoskeletons and peritrophic matrices that line the gut suggests that orthologs might exist in *Drosophila* (Merzendorfer and Zimoch, 2003). BLAST searches with CPG-2, -3, and -4 protein sequences identified CG6048 as a potential homolog in *Drosophila*, supporting this hypothesis. Interestingly, the evolutionary separation of organisms containing exoskeletons (in the protostome lineage) from those containing endoskeletons (in the deuterostome lineage) seems to correlate with the disappearance of chitin and appearance of hyaluronic acid (a copolymer of GlcNAc β 1,4GlcA β 1,3) and sulfated forms of chondroitin. During this period, new chondroitin core proteins evolved with hyaluronic acid binding motifs to stabilize the specialized extracellular matrix unique to cartilage, bone and tendons (Yamaguchi, 2000). Simultaneously, CPGs with chitin binding motifs disappeared. Current data regarding the distribution of different classes of CPGs in species is sparse, and additional studies are needed using techniques like the one reported here to identify the proteoglycans in different organisms.

2. CPGs play essential roles in embryonic cell division in *C. elegans*

Our preliminary analysis of RNAi depletion of either *cpg-1* or *cpg-2* suggested no effect on embryogenesis or morphogenesis, but depletion of both genes resulted

in a strong embryonic phenotype characterized by failure of the embryo to extrude the second polar body following fertilization, loss of membrane ruffling preceding pronuclear fusion, and failure to initiate the cleavage furrow prior to cytokinesis. All of these phenotypes may be related to loss of the extracellular space between the embryo plasma membrane and the eggshell since the various processes might be sterically inhibited by their apposition (Fig. 5). Chondroitin is present both on the embryonic cell surface as well as the eggshell, suggesting functions at both locations (Sugahara et al., 2003). One of the roles of proteoglycans may be to fill the space between the eggshell and the embryo, and the presence of a high concentration of polyanions and their counterions causes sufficient hydrostatic pressure to aid in formation of the extraembryonic space and subsequent ingress of the cleavage furrow. CPG-1 and CPG-2 could also act as structural elements bridging chitin polymers in the eggshell with other components of the embryonic plasma membrane. Regardless of which hypothesis is correct, the presence of the proteoglycans is crucial since depletion of both genes resulted in multinucleated single cell embryos.

Of interest is the finding that less severe inhibition of chondroitin allowed furrow formation to initiate and start ingress into the embryo, but completion of the furrow did not occur efficiently, resulting in regression of the embryo to a single cell. This phenotype was observed by Mizuguchi et al. (2003) at early time points after *sqv-5* RNAi treatment, in this study after *sqv-5* RNAi treatment and overnight recovery at 16°C (data not shown), and in a study by Wang et al. (2005) with partial

depletion of *sqv-2* and *sqv-6*, other enzymes involved in chondroitin synthesis. These findings suggest that CPGs may participate at more than one stage of cell division. Further studies are needed to determine the identity of CPGs required for each of these activities.

The original *sqv* mutants were isolated based on defective larval vulval invagination, but interestingly the loss of *cpg-1* and *cpg-2* by RNAi had no effect on vulval morphogenesis (data not shown). The identity of the relevant CPGs that act in vulval morphogenesis is unknown. CPG-8 represents a candidate gene since it is not enriched in the germline like the other CPGs, suggesting a different site of action, but the other proteoglycans described here might also play a role in vulval morphogenesis. Since cells may express multiple CPGs, the proteoglycans might have redundant properties, making it necessary to perform dual RNAi depletion studies and mutagenesis to elicit the squashed vulva phenotype. Understanding the cell-specific expression of individual proteoglycans may provide insights into the role these molecules play in early development as well as in the physiology of the adult worm.

F. Acknowledgements

We would like to thank Ian McLeod (The Scripps Research Institute) and Weidong Zhou and Huilin Zhou (University of California, San Diego) for helpful discussion regarding mass spectrometry, Lindsay Lewellyn, Amy Maddox, and Nathan Portier for assistance with worm imaging studies, and Raffi Aroian for many

helpful and inspiring discussions (University of California, San Diego). Anjon Audhya kindly provided the PH-domain worm strain. Other worm strains were provided by the CGC stock center in St. Paul, MN. Worm vectors were provided by Andrew Fire. This work was supported by NIH grant GM33063 (J.D.E.) and Genetics Training Grant GM08666 (S.K.O).

G. References

- Bame, K. J. and Esko, J. D. (1989). Undersulfated heparan sulfate in a Chinese hamster ovary cell mutant defective in heparan sulfate N-sulfotransferase. *J.Biol.Chem.* **264**, 8059-8065.
- Beeber, C. and Kieras, F. J. (2002). Characterization of the chondroitin sulfates in wild type *Caenorhabditis elegans*. *Biochem.Biophys.Res.Comm.* **293**, 1374-1376.
- Beintema, J. J. (1994). Structural features of plant chitinases and chitin-binding proteins. *FEBS Lett* **350**, 159-63.
- Bellin, R., Capila, I., Lincecum, J., Park, P. W., Reizes, O. and Bernfield, M. R. (2002). Unlocking the secrets of syndecans: transgenic organisms as a potential key. *Glycoconj J* **19**, 295-304.
- Berninsone, P., Hwang, H. Y., Zemtseva, I., Horvitz, H. R. and Hirschberg, C. B. (2001). SQV-7, a protein involved in *Caenorhabditis elegans* epithelial invagination and early embryogenesis, transports UDP-glucuronic acid, UDP-N-acetylgalactosamine, and UDP-galactose. *Proc Natl Acad Sci U S A* **98**, 3738-3743.
- Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71-94.
- Bulik, D. A. and Robbins, P. W. (2002). The *Caenorhabditis elegans sqv* genes and functions of proteoglycans in development. *Biochim Biophys Acta Gen Subj* **1573**, 247-257.
- Bulik, D. A., Wei, G., Toyoda, H., Kinoshita-Toyoda, A., Waldrip, W. R., Esko, J. D., Robbins, P. W. and Selleck, S. B. (2000). *sqv-3*, *-7*, and *-8*, a set of genes affecting morphogenesis in *Caenorhabditis elegans*, encode enzymes required for glycosaminoglycan biosynthesis. *Proc.Natl.Acad.Sci.USA* **97**, 10838-10843.
- Christner, J. E., Caterson, B. and Baker, J. R. (1980). Immunological determinants of proteoglycans. Antibodies against the unsaturated oligosaccharide products of chondroitinase ABC-digested cartilage proteoglycans. *J.Biol.Chem.* **255**, 7102-7105.

Esko, J. D. (1993) in *Current protocols in molecular biology*, eds. Ausubel, F., Brent, R., Kingston, B., Moore, D., Seidman, J., Smith, J., Struhl, K., Varki, A. & Coligan, J. (Greene Publishing and Wiley-Interscience, New York), pp. 17.2.1-17.2.9.

Esko, J. D. and Manzi, A. (1996) in *Current protocols in molecular biology*, eds. Ausubel, F., Brent, R., Kingston, B., Moore, D., Seidman, J., Smith, J., Struhl, K., Varki, A. and Coligan, J. (Greene Publishing and Wiley-Interscience, New York), pp. 17.9.8-17.9.11.

Esko, J. D. and Zhang, L. (1996). Influence of core protein sequence on glycosaminoglycan assembly. *Curr.Opin.Struct.Biol.* **6**, 663-670.

Ghiselli, G. and Iozzo, R. V. (2000). Overexpression of bamacan/SMC3 causes transformation. *J Biol Chem* **275**, 20235-8.

Herman, T., Hartweg, E. and Horvitz, H. R. (1999). *sqv* mutants of *Caenorhabditis elegans* are defective in vulval epithelial invagination. *Proc.Natl.Acad.Sci.U.S.A.* **96**, 968-973.

Hwang, H. Y. and Horvitz, H. R. (2002a). The SQV-1 UDP-glucuronic acid decarboxylase and the SQV-7 nucleotide-sugar transporter may act in the Golgi apparatus to affect *C. elegans* vulval morphogenesis and embryonic development. *Proc Natl Acad Sci U S A* **99**, 14218-23.

Hwang, H. Y. and Horvitz, H. R. (2002b). The *C. elegans* vulval morphogenesis gene *sqv-4* encodes a UDP-glucose dehydrogenase that is temporally and spatially regulated. *Proc Natl Acad Sci U S A* **99**, 14224-9.

Hwang, H. Y., Olson, S. K., Brown, J. R., Esko, J. D. and Horvitz, H. R. (2003a). The *C. elegans* genes *sqv-2* and *sqv-6*, which are involved in vulval morphogenesis, encode glycosaminoglycan galactosyltransferase II and xylosyltransferase. *J. Biol. Chem.* **278**, 11735-8.

Hwang, H. Y., Olson, S. K., Esko, J. D. and Horvitz, H. R. (2003b). *Caenorhabditis elegans* early embryogenesis and vulval morphogenesis require chondroitin biosynthesis. *Nature* **423**, 439-443.

- Iozzo, R. V. and Murdoch, A. D. (1996). Proteoglycans of the extracellular environment: clues from the gene and protein side offer novel perspectives in molecular diversity and function. *FASEB J.* **10**, 598-614.
- Izumikawa, T., Kitagawa, H., Mizuguchi, S., Nomura, K. H., Nomura, K., Tamura, J., Gengyo-Ando, K., Mitani, S. and Sugahara, K. (2004). Nematode chondroitin polymerizing factor showing cell-/organ-specific expression is indispensable for chondroitin synthesis and embryonic cell division. *J Biol Chem* **279**, 53755-61.
- Kamath, R. S., Martinez-Campos, M., Zipperlen, P., Fraser, A. G. and Ahringer, J. (2001). Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol* **2**, RESEARCH0002.1-0002.10.
- Lee, M. H. and Schedl, T. (2001). Identification of *in vivo* mRNA targets of GLD-1, a maxi-KH motif containing protein required for *C. elegans* germ cell development. *Genes Dev* **15**, 2408-20.
- MacCoss, M. J., McDonald, W. H., Saraf, A., Sadygov, R., Clark, J. M., Tasto, J. J., Gould, K. L., Wolters, D., Washburn, M., Weiss, A., Clark, J. I. and Yates, J. R., 3rd (2002). Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci U S A* **99**, 7900-7905.
- Maddox, A. S., Habermann, B., Desai, A. and Oegema, K. (2005). Distinct roles for two *C. elegans* anillins in the gonad and early embryo. *Development* **132**, 2837-48.
- Merzendorfer, H. and Zimoch, L. (2003). Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *J Exp Biol* **206**, 4393-412.
- Mizuguchi, S., Uyama, T., Kitagawa, H., Nomura, K. H., Dejima, K., Gengyo-Ando, K., Mitani, S., Sugahara, K. and Nomura, K. (2003). Chondroitin proteoglycans are involved in cell division of *Caenorhabditis elegans*. *Nature* **423**, 443-448.

Olson, S. K. and Esko, J. D. (2004). Proteoglycans, in *Encyclopedia of Biological Chemistry*, eds. Lennarz, W. & Lane, M. D. (Elsevier, Oxford), Vol. 3, pp. 549-555.

Reinke, V., Smith, H. E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S. J., Davis, E. B., Scherer, S., Ward, S. and Kim, S. K. (2000). A global profile of germline gene expression in *C. elegans*. *Mol Cell* **6**, 605-16.

Rodén, L., Baker, J. R., Helting, T., Schwartz, N. B., Stoolmiller, A. C., Yamagata, S. and Yamagata, T. (1972). Biosynthesis of chondroitin sulfate. *Meth.in Enzymol.* **28**, 638-676.

Schimpf, J., Sames, K. and Zwillig, R. (1999). Proteoglycan distribution pattern during aging in the nematode *Caenorhabditis elegans*: an ultrastructural histochemical study. *Histochem.J.* **31**, 285-292.

Sugahara, K., Mikami, T., Uyama, T., Mizuguchi, S., Nomura, K. and Kitagawa, H. (2003). Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate. *Curr Opin Struct Biol* **13**, 612-20.

Toyoda, H., Kinoshita-Toyoda, A. and Selleck, S. B. (2000). Structural analysis of glycosaminoglycans in *Drosophila* and *Caenorhabditis elegans* and demonstration that *tout-velu*, a *Drosophila* gene related to EXT tumor suppressors, affects heparan sulfate *in vivo*. *J.Biol.Chem.* **275**, 2269-2275.

Venegas, A., Goldstein, J. C., Beauregard, K., Oles, A., Abdulhayoglu, N. and Fuhrman, J. A. (1996). Expression of recombinant microfilarial chitinase and analysis of domain function. *Mol Biochem Parasitol* **78**, 149-59.

Wang, H., Spang, A., Sullivan, M. A., Hryhorenko, J. and Hagen, F. K. (2005). The terminal phase of cytokinesis in the *Caenorhabditis elegans* early embryo requires protein glycosylation. *Mol Biol Cell* **16**, 4202-13.

Washburn, M. P., Ulaszek, R., Deciu, C., Schieltz, D. M. and Yates, J. R., 3rd (2002). Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem* **74**, 1650-1657.

Washburn, M. P., Wolters, D. and Yates, J. R., 3rd (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242-7.

Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J. and Hart, G. W. (2002). Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* **1**, 791-804.

Wolters, D. A., Washburn, M. P. and Yates, J. R., 3rd (2001). An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73**, 5683-90.

Wright, H. T., Sandrasegaram, G. and Wright, C. S. (1991). Evolution of a family of N-acetylglucosamine binding proteins containing the disulfide-rich domain of wheat germ agglutinin. *J Mol Evol* **33**, 283-94.

Yamada, S., Okada, Y., Ueno, M., Deepa, S. S., Nishimura, S., Fujita, M., Van Die, I., Hirabayashi, Y. & Sugahara, K. (2002). Determination of the glycosaminoglycan-protein linkage region oligosaccharide structures of proteoglycans from *Drosophila melanogaster* and *Caenorhabditis elegans*. *J Biol Chem* **277**, 31877-31886.

Yamada, S., Van Die, I., Van den Eijnden, D. H., Yokota, A., Kitagawa, H. and Sugahara, K. (1999). Demonstration of glycosaminoglycans in *Caenorhabditis elegans*. *FEBS Lett.* **459**, 327-331.

Yamaguchi, Y. (2000). Lecticans: organizers of the brain extracellular matrix. *Cell.Mol.Life Sci.* **57**, 276-289.

Zelensky, A. N. and Gready, J. E. (2003). Comparative analysis of structural properties of the C-type-lectin-like domain (CTLCD). *Proteins* **52**, 466-77.

The text of Chapter 4 is a manuscript in preparation, Olson, S.K., Bishop, J.R., Yates, J.R., Oegema, K. and Esko, J.D. (2005). The dissertation author is the primary researcher and author and the co-authors listed in this manuscript directed and supervised the research which forms the basis for this chapter.

CHAPTER 5

Perspective and Future Directions

A. Evolution of chondroitin proteoglycan components

1. Vertebrate chondroitin synthesis is conserved in *C. elegans*

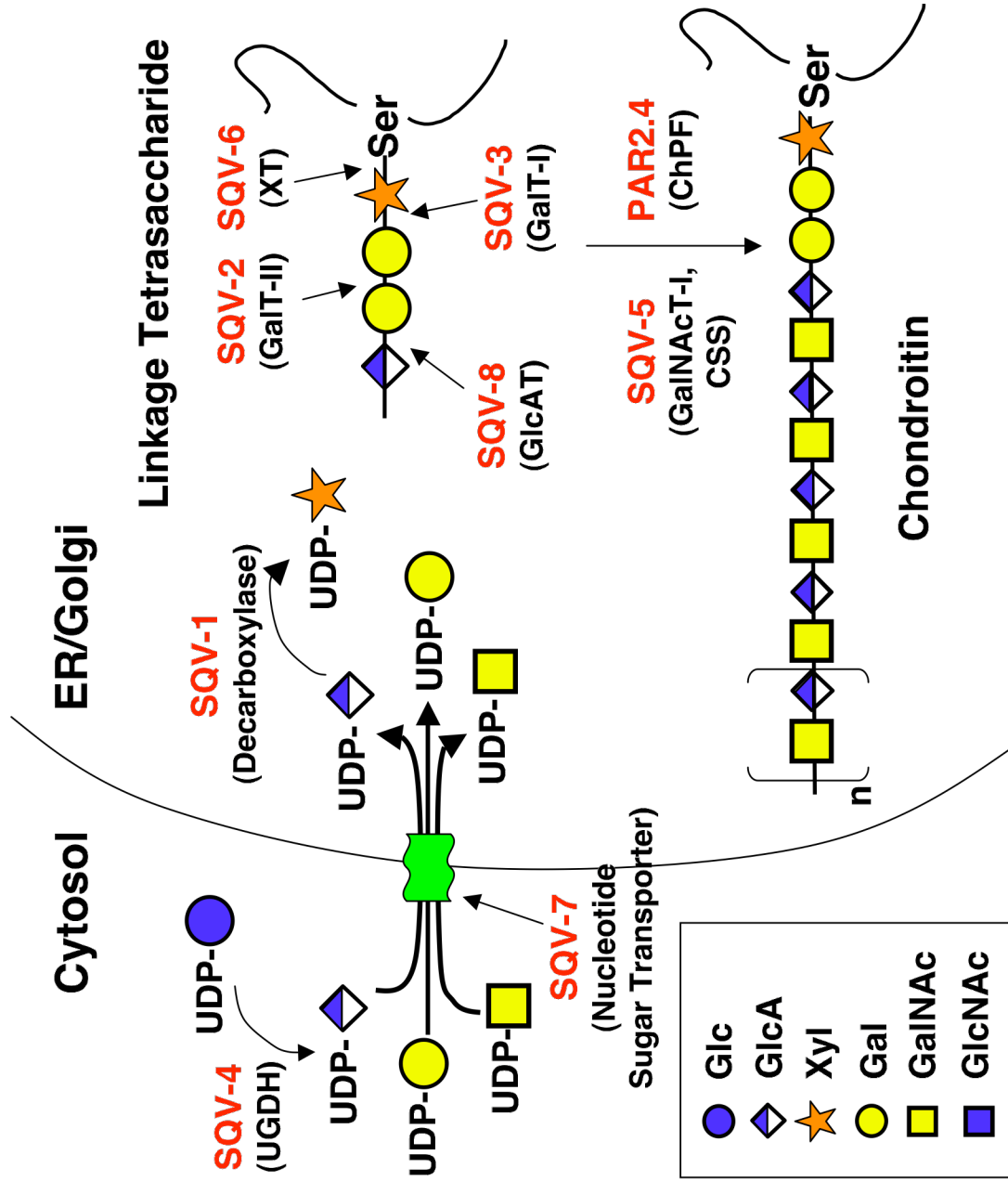
Vertebrate chondroitin sulfate (CS) is synthesized by several different enzymes, starting with a set of four distinct glycosyltransferases that catalyze each step of linkage tetrasaccharide assembly. Chondroitin is then initiated with addition of GalNAc by a separate enzyme (Gotoh et al., 2002; Uyama et al., 2002), and chain polymerization is carried out by five more (Gotoh et al., 2002; Kitagawa et al., 2001; Uyama et al., 2003; Yada et al., 2003a; Yada et al., 2003b). Activation of the synthase enzymes is accomplished by the chondroitin polymerizing factor (Kitagawa et al., 2003). In contrast to vertebrate CS, *C. elegans* chondroitin is a simplified, non-sulfated version of the vertebrate chain (Toyoda et al., 2000; Yamada et al., 1999). While modification by sulfation is absent in the worm, synthesis of the chain occurs in a manner identical to vertebrates. Chapter 2 demonstrated that SQV-6 is the *C. elegans* xylosyltransferase homolog, the enzyme that initiates GAG synthesis on protein cores. SQV-2 was found to have galactosyltransferase-II activity. These two enzymes, in conjunction with the SQV-3 galactosyltransferase-I and SQV-8 glucuronosyltransferase (Bulik et al., 2000), generate the linkage tetrasaccharide

precursor that is common to both chondroitin and heparan sulfate (Figure 1, See also Figure 2 of Chapter 1).

Chapter 3 demonstrated that SQV-5 is the chondroitin synthase, a dual-function enzyme that catalyzes polymerization of the chondroitin chain. SQV-5 also possessed chondroitin GalNAcT-I activity, a step catalyzed by a separate enzyme in vertebrates (Gotoh et al., 2002; Uyama et al., 2002). Figure 1 illustrates that each enzyme involved in vertebrate chondroitin backbone synthesis has a *C. elegans* counterpart. Vertebrates have evolved a multigene family for chondroitin polymerization, perhaps by gene duplication events, but the essential components of the biosynthetic pathway are conserved between vertebrates and *C. elegans*. This finding suggests that the mechanism of chondroitin assembly arose early during metazoan evolution and was maintained by strong selective pressure.

Figure 1. Chondroitin biosynthetic enzymes are conserved in *C. elegans*

C. elegans contains all proteins and enzymes necessary to generate a chondroitin chain. Synthesis of nucleotide sugar donors is performed by the SQV-4 UDP-Glc dehydrogenase (UGDH) and the SQV-1 UDP-GlcA decarboxylase. Sugar nucleotides are transported from the cytosol into the endoplasmic reticulum (ER) and/or Golgi apparatus by the SQV-7 nucleotide sugar transporter. The linkage tetrasaccharide is generated through the actions of the SQV-6 xylosyltransferase (XT), SQV-3 galactosyltransferase-I (GalT-I), SQV-2 GalT-II, and SQV-8 glucuronosyltransferase (GlcAT). The chondroitin chain is elongated by SQV-5, which contains both N-acetylgalactosaminyltransferase-I activity (GalNAcT-I) and chondroitin synthase activity (CSS). The PAR2.4 chondroitin polymerizing factor (ChPF) activates the synthase enzyme(s).



2. *C. elegans* expresses a novel set of chondroitin core proteins

While chondroitin biosynthesis is essentially identical in vertebrates and *C. elegans*, the core proteins on which the chains assemble are distinctly different. *In silico* analysis showed that no known CSPGs were present in the worm genome. A biochemical purification and proteomics approach identified nine novel chondroitin proteoglycans (CPGs) in the worm, none of which showed homology to any vertebrate proteins. One notable observation was that even though the main body of the protein was not similar to other known CSPGs, the consensus sequence for GAG attachment (Ser-Gly flanked by acidic amino acids) was conserved in all CPGs. Recognition of the attachment site is likely a function of the highly conserved xylosyltransferase, suggesting that GAG attachment sites have been maintained by selective pressure while the remainder of the protein scaffold is flexible.

It is interesting to speculate that a different set of proteoglycan core proteins have evolved to serve functions specific to the need of the organism. For example, *C. elegans* CPG-1 and CPG-2 are necessary for early cell division, possibly through interaction with chitin in the embryonic eggshell (see discussion below). Higher organisms do not synthesize chitin, but they do require chondroitin sulfate in tissues not present in the nematode, including cartilage, bone, and the circulatory system (See Tables 2 and 3 in Chapter 1). A more thorough understanding of how core protein choice ties to chondroitin (sulfate) function could be addressed by analyzing proteoglycans of organisms lying between nematodes and mammals on the evolutionary tree. Potential targets include *Drosophila melanogaster*, another

invertebrate with a chitinous exoskeleton that expresses a partially modified chondroitin with sulfate residues at the 4-O position. *D. melanogaster* lacks dermatan sulfate (DS) (Toyoda et al., 2000); the sea urchin, an invertebrate lacking chitin but shown to require chondroitin sulfate for embryogenesis (Lane et al., 1993; Lane and Solursh, 1988; Lane and Solursh, 1991; Solursh et al., 1986); and zebrafish, another organism with a more complicated body plan that expresses CS and DS. Few CSPG core proteins have been described in zebrafish, and none have been characterized in flies or sea urchins. It would be fascinating to determine which evolutionary events coincided with the development of modern mammalian CSPGs as core proteins. One possibility is the transition from an exoskeleton to an endoskeleton. This transition seems to correlate with the loss of chitin and expression of hyaluronan (HA) in higher organisms. One hallmark of vertebrate CSPGs is the presence of HA binding domains in the Aggrecan family of proteoglycans. Interaction between HA and CSPGs stabilizes the extracellular matrix of cartilage and other tissues not present in lower organisms.

B. A proteomics approach to proteoglycan identification

1. Universal applicability

The biochemical purification and proteomic identification of proteoglycan core proteins outlined in Chapter 4 can be applied to any system where knowledge of PG components is desirable. One application of this method is to identify novel core proteins in model organisms in which PG content is unknown, such as *Drosophila*

and sea urchin as described above. As evidenced by my work in *C. elegans*, the method is especially useful in identifying novel PGs that do not show sequence homology to known PGs, or in organisms whose genomes have not yet been sequenced. This approach can also be used to determine the spatial and temporal distribution patterns in order to study functional impact of PGs. The work described here identified CPGs from a mixed stage population of worms, but it should be possible to apply the method to determine PG expression at different stages of worm development. For example, comparing the PGs present in eggs versus L4 larvae might shed light on which PGs are relevant for embryogenesis versus vulval invagination. A similar approach makes it possible to compare CSPGs expressed in various mouse tissues, such as liver or brain.

2. Advantages

The combination of target protein enrichment and BEMAD tagging of glycosylation sites makes this method a powerful diagnostic technique. Historically, identification of an unknown protein by MALDI-TOF mass spectrometry necessitated purification of the protein of interest to homogeneity, a potentially daunting task that could exclude proteins expressed at low levels. The subproteomic approach of partial purification and analysis by the multi-dimensional peptide identification technique (MudPIT, a 2-dimensional liquid chromatography separation) makes it possible to identify proteins that are present in the sample at relative concentrations as low as 1 part in 1000 (Schirmer et al., 2003). This approach proved to be essential for our identification of *C. elegans* CPGs. Earlier

attempts at purifying CPGs to a point where they were pure and abundant enough to be visualized by silver or Coomassie stained gels were highly unsuccessful. With the MudPIT technique, the protein of interest need only be partially purified, reducing the effort and time required for identification.

The MudPIT subproteomic approach identified nine CPGs present amid extensive contamination. The BEMAD tagging technique excluded all proteins that were not modified with DTT at serine residues. The end result was not only ability to screen CPGs from contaminants, but also to identify sites of chondroitin attachment to the protein cores. Most PGs contain multiple putative GAG attachment sites. The BEMAD method makes it possible to identify which of these sites are modified *in vivo*.

3. Limitations

While the methods described above resulted in identification of nine novel *C. elegans* CPGs, the incomplete conditions used for tagging sites might have obscured the true number of worm CPGs. All of the identified CPGs are thought to be secreted proteins due to lack of transmembrane domains or consensus sites for the attachment of glycosylphosphatidylinositol anchors. However, many vertebrate CSPGs are embedded in the membrane, so it seems possible that *C. elegans* membrane CPGs were missed with this method. Isolation of *C. elegans* membrane CPGs, if they exist, may necessitate a different extraction procedure. Heparan sulfate proteoglycans also were not identified by this method. One explanation is the low abundance of heparan sulfate compared to chondroitin (Toyoda et al., 2000;

Yamada et al., 1999). Another explanation is that this method did not enrich for membrane PGs such as syndecan and glypican.

It may also be necessary to refine the BEMAD method to identify all sites of chondroitin attachment. β -elimination of the chain occurs efficiently but DTT addition was sub-stoichiometric, evidenced by peptides that included a mass consistent with dehydroalanine (-18 Da), the product of serine β -elimination. Fortunately it is possible to search the protein database for both gain of DTT (+167 Da) and loss of water, but it would be more desirable that all GAG attachment sites were labeled. Further work is needed to optimize the Michael addition reaction.

True tryptic peptides were observed less frequently than expected in most samples. Some peptides contained either arginine or lysine at one terminus of the peptide, but others ended in an unexpected amino acid. In some cases the peptide terminated in either a Gly or modified Ser, suggesting that the GAG attachment sites are susceptible to cleavage either during β -elimination or mass spectrometry fragmentation. In other cases the peptide seemed to have been cleaved at a random position. Analysis of expected tryptic peptides of CPG proteins showed that they could be extremely long, sometimes reaching ~100 amino acids. Optimal peptide length for MS/MS analysis is 10-20 amino acids. To accurately identify proteoglycans by this method, it may be necessary to find a combination of different proteases to optimally cleave proteins into optimal fragment sizes.

C. Functions of *C. elegans* chondroitin proteoglycans

1. Functional redundancy

It may be difficult to determine the role of individual CPGs in *C. elegans* development. Several genome-wide RNAi screens failed to identify phenotypes for any of the nine *cpg* genes (Gonczy et al., 2000; Hanazawa et al., 2001; Kamath et al., 2003; Piano et al., 2002; Rual et al., 2004; Sonnichsen et al., 2005). The wild-type phenotype might be explained by functional redundancy of the CPG proteins, which would be consistent with the fact that none of the *sqv* mutants had defects in a CPG core protein. Depletion of *cpg-1* and *cpg-2* alone had no phenotype, but simultaneous depletion of both genes resulted in a dramatic embryonic cell division defect (Chapter 4). Both CPG-1 and CPG-2 are putative chitin-binding proteins, and the synergy resulting from removal of both gene products suggests they are functionally redundant. It is possible that RNAi depletion of single *cpg* transcripts results in subtle phenotypes that were not picked up by the genome-wide RNAi screens, but chances are that multiple *cpg* transcripts will need to be depleted to see a phenotype.

One possible explanation for the apparent functional redundancy of the CPGs is similarity of protein structure. CPG-1 and CPG-2 both contain chitin-binding domains. CPG-3 and CPG-4 contain multiple Ser-Gly GAG attachment sites resembling those of CPG-2. CPG-5 and CPG-6 are 67% identical and 90% similar, so they might compensate for each other. CPG-7, -8, and -9 are all very small proteins with several GAG attachment sites. The fact that the nine *C. elegans* CPGs

seem to fall into “classes” suggests that the various class members may have redundant functions.

Future experiments will entail depleting the *cpgs* in combination with each other to determine which CPGs are functionally redundant in order to elucidate their biological functions. *cpg* depletion can initially be accomplished through RNAi experiments, but eventually null mutations in each gene will need to be generated (no mutants currently exist for any of the *cpgs*). Reverse genetic technology in the worm has advanced into a relatively routine method. These mutants can be generated either by the *C. elegans* Gene Knockout Consortium (<http://celeganskoconsortium.omrf.org>) or through an individual effort. While the null allele “sledgehammer” approach will demonstrate the function of the CPGs, it will also be interesting to separate chondroitin function from possible roles of the protein core itself. Site-directed mutagenesis can be used to generate a CPG construct that is essentially wild-type except for deletion of the chondroitin attachment sites. This will be an interesting experiment to perform with CPG-1 and CPG-2 to determine the importance of the chitin binding domains versus the chondroitin chains. It will hopefully also shed light on the relevance of the C-type lectin domains of CPG-5 and CPG-6. These two proteins contain only one putative glycosylation site, similar to the mammalian CSPG decorin. Evidence shows that the decorin core protein is the biologically relevant domain of the molecule, as removal of the CS chain has no effect on phenotype (Vogel et al., 1987). Whether CPG-5 and CPG-6 function in a similar manner remains to be determined.

2. Temporal and spatial expression

C. elegans CPGs were purified from a mixed-stage population of worms, so it is not possible to draw any definitive conclusions about their sites of action. It is likely that some, if not all, of the CPGs will function in vulval morphogenesis and early embryogenesis based on the phenotype of the *sqv-5* mutants (Herman et al., 1999; Hwang et al., 2003). Hints as to possible involvement come from a study of genes enriched for germline expression. Transcripts from wild-type and germline-deficient (*glp-1*) worms were compared by microarray analysis (Reinke et al., 2000). Genes not expressed in *glp-1* mutants but abundant in wild-type worms were inferred to be enriched for germline expression. *cpg-1*, -2, -3, -5, and -6 are enriched in the germline, while *cpg-8* is not (Figure 2 of Chapter 4). *cpg-4*, -7, and -9 were not analyzed. This study gives us candidate genes to test for embryogenesis and vulval phenotypes.

In addition to the candidate gene approach, it will also be necessary to directly establish the localization patterns of each CPG. The Serial Analysis of Gene Expression (SAGE) database (<http://tock.bcgsc.bc.ca/cgi-bin/sage140>), which analyzes mRNA expression at various stages of development, found that *cpg-1* through -6 transcripts are upregulated around the L4 larval stage and persist through the adult stage before decreasing during embryogenesis, consistent with their expression in the germline. Expression pattern information is not available for any of the *cpgs* through the Genome Sciences Centre that looks at promoter::GFP

constructs (<http://elegans.bcgsc.ca/perl/eprofile/index>) or the Nematode Expression Pattern Database (<http://nematode.lab.nig.ac.jp>).

Expression constructs for *cpg-1* and *cpg-2* using native genomic DNA promoter sequence to drive a translational GFP fusion product are under development. Localization of the GFP-tagged proteins will hopefully shed some light on the site(s) of action of each CPG. In the case of CPG-1 and CPG-2, it will be interesting to determine if the proteins are expressed at the embryonic plasma membrane, adjacent to the eggshell, and/or within the cleavage furrow (see discussion below in section D). One study found CPG-1/CEJ-1 to be expressed in only two cells near the pharynx (Mounsey et al., 2002), but the study by Reinke et al. (Reinke et al., 2000) and our RNAi results (Chapter 4) show that it is expressed and acts in the germline and embryos. This suggests *cpg-1* is either expressed at low levels, or that the 2000 base pairs of upstream promoter used by Mounsey et al. (Mounsey et al., 2002) did not contain enough of the endogenous 8000 base pair sequence to drive normal expression. We also plan to generate antibodies against each of the CPGs as an alternate method to study protein localization.

D. Model for CPG function in embryonic cell division

Chondroitin was identified as a molecule fundamental for vulval invagination and embryogenesis, the loss of which results in the squashed vulva phenotype. How might chondroitin and its associated proteoglycan cores direct these processes?

1. Site of action

Expression of *cpg-1/cej-1* and *cpg-2* is translationally regulated in the *C. elegans* gonad. Transcripts are present as early as the L4 stage and increase in abundance during the adult stage of development (SAGE database <http://tock.bcgsc.bc.ca/cgi-bin/sage140>). A screen performed by Lee et al. showed that *cpg-1* and *cpg-2* are two targets of the translational inhibitor GLD-1 (Lee and Schedl, 2001). GLD-1 is normally present in the distal gonad and is down-regulated at the gonad loop, the point where developing oocytes start to become cellularized (Jones et al., 1996). Loss of GLD-1 inhibition would allow *cpg-1/cej-1* and *cpg-2* to be translated, suggesting that the protein products are required in the oocyte and/or early embryo. *sqv-5* was not identified in the GLD-1 target screen, but shows a similar pattern of low expression in the distal gonad and increased expression in the proximal gonad (Ho-Yon Hwang, personal communication). It remains to be determined whether translation of *sqv-5* and *cpg-1/cej-1* and *cpg-2* are in fact co-regulated in the distal gonad. Nevertheless, specific upregulation of the CPG proteins at this point in time suggests they are required for early embryogenesis.

2. Role in embryogenesis

a. Creation of a perivitelline space

A number of important events occur following oocyte formation. First, the oocyte passes through the spermatheca and is fertilized. Following fertilization, the chitinous eggshell is secreted and surrounds the embryo to protect it from environmental challenges. After the eggshell is formed, the maternal nuclear

material divides and extrudes two polar bodies into the perivitelline space that has formed between the embryo and eggshell. Ruffling of the embryonic plasma membrane is also observed following fertilization, thought to be directed by the rearrangement of the actomyosin network within the embryo (reviewed in Cowan and Hyman, 2004). During ruffling, a pseudocleavage furrow forms, which is a transient indentation of the plasma membrane that resembles the early stages of furrow formation at the first cell division.

While fertilization and eggshell formation appear normal in *cpg-1/cpg-2(RNAi)* and *sqv-5(RNAi)* embryos, membrane ruffling and pseudocleavage do not occur, nor does initiation of the cleavage furrow following nuclear division (Chapter 4, Figure 6). One possible explanation for these observations is that secreted chondroitin proteoglycans are required to create the perivitelline space between the embryo and the eggshell. Notably, this extracellular space is lacking in both *sqv-5(RNAi)* and *cpg-1/cpg-2(RNAi)* embryos. It is possible to imagine that juxtaposition of the plasma membrane and eggshell would sterically inhibit all the described membrane events that lead up to the first cell division.

b. Perivitelline spaces in other organisms

Components of the eggshell and perivitelline space have not been thoroughly characterized in *C. elegans*, but in other parasitic nematode species these layers are comprised of various proteins, lipids, chitin, and mucoproteins (Burgwyn et al., 2003; Mariano, 1967; Wharton, 1980; Wharton, 1979a; Wharton, 1979b; Wharton, 1979c). More thorough analysis of events that follow fertilization has been

conducted in sea urchins. Cortical granules reside in the oocyte near the plasma membrane. Upon fertilization, the granules fuse with the membrane and empty their contents, effectively creating a block to polyspermy and forming a perivitelline space between the plasma membrane and vitelline layer. Cortical granules have been found to contain proteases which may cleave proteins linking the two layers, as well as sulfated mucopolysaccharides (GAGs) that might aid in enlarging the perivitelline space (Ishihara, 1968; Schuel et al., 1974; Vacquier et al., 1973). It is possible that similar events occur in *C. elegans*, and that vesicles containing chondroitin proteoglycans might be secreted into the extracellular space following fertilization. One way to address this question is to generate GFP-tagged CPG-1 and CPG-2, as described in section C of this chapter. Using time-lapse microscopy, the proteins can be followed from their point of synthesis to site of action, observing whether onset of expression correlates with formation of the perivitelline space.

c. Speculation on the role of CPGs in creating the perivitelline space

If the theory is correct about the requirement of CPGs in the perivitelline space, how might they result in its formation? The two distinct features of CPG-1 and CPG-2 are their chitin binding domains and abundance of glycosylation sites (especially true for CPG-2). At least two possibilities exist, neither of which is mutually exclusive. In vertebrates, highly anionic chondroitin sulfate chains lead to formation of a hydrated matrix (See Chapter 1); the same phenomenon may occur in *C. elegans*. Generation of hydrostatic pressure could maintain the fluid-

filled space lying between the embryonic membrane and the eggshell. Another possibility is that CPG-1 and CPG-2 interact with chitin polymers in the eggshell through their peritrophin-A chitin binding domains. The functionality of the chitin domains has not been established. If they are functional, it is conceivable that these proteins could form a network bridging the chitinous eggshell with components at the plasma membrane.

d. Possible secondary role of chondroitin in completion of cytokinesis

Of interest is the finding that less severe inhibition of chondroitin results in regression of cytokinesis. *sqv-5(RNAi)* experiments conducted at a lower temperature allowed cleavage furrow formation to occur. The furrow ingressed approximately halfway into the embryo, did not complete, and regressed back to the point of origin resulting in a single celled embryo (data not shown). This phenotype was also observed by Mizuguchi et al. at early time points after *sqv-5* RNAi treatment (Mizuguchi et al., 2003). The fact that the furrow was able to initiate in these studies suggests a second role for chondroitin in completion of cytokinesis, rather than initiation.

Chondroitin has been localized to both the embryonic cell surface, as well as the eggshell, providing circumstantial evidence for this theory of dual sites of action (Sugahara et al., 2003). Other evidence suggesting that chondroitin needs to be associated with the embryonic plasma membrane was shown in an experiment by Mizuguchi et al. (Mizuguchi et al., 2003). The eggshell of young embryos was removed with chitinase, and the embryonic blastomeres were cultured and treated

with chondroitinase ABC. Blastomeres lacking chondroitin attempted cell division, sometimes progressing to the 4- or 8-cell stage, but then regressed back to a 2- or 4-cell stage. It will be necessary to conduct the *cpg-1/cpg-2(RNAi)* experiments at the less stringent 16°C temperature to determine if the regression phenotype is present. It is also possible that other CPGs are involved in completion of cytokinesis.

e. Other possible mechanisms

While it is more likely that chondroitin and its associated CPGs are playing a biophysical role during embryogenesis, it is also possible that chondroitin may be functioning through a signaling pathway. Heparan sulfate is known to interact with a large number of growth factors, morphogens, and extracellular matrix components through variably sulfated domains on its chain. Chondroitin sulfate is known to bind a number of matrix components as well, though binding is less dependent on specific sulfation patterns. *C. elegans* chondroitin is nonsulfated, further reducing the likelihood of specific ligand binding. However, the nonsulfated HA polymer is known to bind a number of proteins, so it is possible that worm chondroitin may as well (Day and Prestwich, 2002; Day and Sheehan, 2001; Lee and Spicer, 2000). Regardless of whether any of these hypotheses are correct, it has been demonstrated that chondroitin proteoglycans are essential components of early embryogenesis.

E. Model for CPG function in vulval invagination

1. Generation of a hydrated matrix by an osmotic gradient

The possible biophysical role chondroitin might be playing during embryogenesis could also explain the squashed vulva defect seen in the *sgv* mutants. Both the perivitelline space and the vulval lumen are extracellular structures likely supported by a hydrated matrix. It is feasible that vulval epithelial cells could secrete CPGs into the lumen. Similar to aggrecan in cartilage, the negatively charged chondroitin chains in the lumen would be neutralized with counterions (e.g. Na^+), increasing the salt concentration and generating an osmotic gradient that would draw water into the space. It will be necessary to first identify which CPGs are involved in vulval invagination, and then either tag them with GFP or generate antibodies to determine if they are indeed secreted into the vulval lumen. It is also ultimately possible that CPGs function instead through signaling events, cell-cell interactions, or cytoskeletal rearrangement of vulval cells.

2. Epithelial invagination in sea urchin development

The sea urchin provides evidence that chondroitin sulfate can direct the process of epithelial invagination. Lane et al. (Lane et al., 1993) showed that gastrulation-stage epithelial invagination can be prematurely stimulated by a calcium ionophore. It was demonstrated that ionophore-dependent invagination requires sulfation and Golgi-mediated secretion, both common to GAG synthesis, and newly deposited extracellular matrix. Using a monoclonal antibody against chondroitin sulfate types A and C (Table 1, Chapter 1), they found that both the precocious and natural

invagination sites contain CS deposited in the extracellular matrix. The authors propose that the CSPGs cause swelling of the matrix, leading to inward bending of the epithelial layer into the gastrulating embryo.

F. Concluding remarks

This thesis has demonstrated the essential and fundamental requirement of nonsulfated chondroitin in *C. elegans* developmental events. Identification and characterization of three *sqv* genes showed that chondroitin was required for vulval invagination and early embryogenesis, and that chondroitin biosynthesis is conserved between worms and vertebrates. Interestingly, identification of nine novel chondroitin proteoglycans showed that core protein choice is not conserved, possibly explaining the difference in sites of chondroitin action of the two very distantly related organisms. The future of this project is wide open. One initial goal will be to more thoroughly characterize the *cpg-1/cpg-2* embryonic defect and hopefully shed light on the mechanism of cleavage furrow initiation. Another short-term goal is to identify the functions of the other CPGs by simultaneous RNAi depletion of multiple genes, in the hopes of finding the relevant proteins that contribute toward embryogenesis and/or vulval invagination. Understanding how the extracellular matrix regulates these fundamental and well-studied developmental events is a new approach not yet undertaken by the *C. elegans* community, suggesting that this will be an exciting and promising niche to investigate.

G. References

Bulik, D. A., Wei, G., Toyoda, H., Kinoshita-Toyoda, A., Waldrip, W. R., Esko, J. D., Robbins, P. W. and Selleck, S. B. (2000). *sqv-3 -7, and -8*, a set of genes affecting morphogenesis in *Caenorhabditis elegans*, encode enzymes required for glycosaminoglycan biosynthesis. *Proc.Natl.Acad.Sci.USA* **97**, 10838-10843.

Burgwyn, B., Nagel, B., Ryerse, J. and Bolla, R. I. (2003). Heterodera glycines: eggshell ultrastructure and histochemical localization of chitinous components. *Exp Parasitol* **104**, 47-53.

Cowan, C. R. and Hyman, A. A. (2004). Asymmetric cell division in *C. elegans*: cortical polarity and spindle positioning. *Annu Rev Cell Dev Biol* **20**, 427-53.

Day, A. J. and Prestwich, G. D. (2002). Hyaluronan-binding proteins: Tying up the giant. *J.Biol.Chem.* **277**, 4585-4588.

Day, A. J. and Sheehan, J. K. (2001). Hyaluronan: polysaccharide chaos to protein organisation. *Curr.Opin.Struct.Biol.* **11**, 617-622.

Gonczy, P., Echeverri, C., Oegema, K., Coulson, A., Jones, S. J., Copley, R. R., Duperon, J., Oegema, J., Brehm, M., Cassin, E. et al. (2000). Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331-6.

Gotoh, M., Yada, T., Sato, T., Akashima, T., Iwasaki, H., Mochizuki, H., Inaba, N., Togayachi, A., Kudo, T., Watanabe, H. et al. (2002). Molecular cloning and characterization of a novel chondroitin sulfate glucuronyltransferase that transfers glucuronic acid to N-acetylgalactosamine. *J Biol Chem* **277**, 38179-38188.

Hanazawa, M., Mochii, M., Ueno, N., Kohara, Y. and Iino, Y. (2001). Use of cDNA subtraction and RNA interference screens in combination reveals genes required for germ-line development in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **98**, 8686-91.

Herman, T., Hartweg, E. and Horvitz, H. R. (1999). sqv mutants of *Caenorhabditis elegans* are defective in vulval epithelial invagination. *Proc.Natl.Acad.Sci.U.S.A.* **96**, 968-973.

Hwang, H. Y., Olson, S. K., Esko, J. D. and Horvitz, H. R. (2003). *Caenorhabditis elegans* early embryogenesis and vulval morphogenesis require chondroitin biosynthesis. *Nature* **423**, 439-443.

Ishihara, K. (1968). An analysis of acid polysaccharides produced at fertilization of sea urchin. *Exp Cell Res* **51**, 473-84.

Jones, A. R., Francis, R. and Schedl, T. (1996). GLD-1, a cytoplasmic protein essential for oocyte differentiation, shows stage- and sex-specific expression during *Caenorhabditis elegans* germline development. *Dev Biol* **180**, 165-83.

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M. et al. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231-7.

Kitagawa, H., Izumikawa, T., Uyama, T. and Sugahara, K. (2003). Molecular cloning of a chondroitin polymerizing factor that cooperates with chondroitin synthase for chondroitin polymerization. *J Biol Chem* **278**, 23666-71.

Kitagawa, H., Uyama, T. and Sugahara, K. (2001). Molecular cloning and expression of a human chondroitin synthase. *J.Biol.Chem.* **276**, 38721-38726.

Lane, M. C., Koehl, M. A., Wilt, F. and Keller, R. (1993). A role for regulated secretion of apical extracellular matrix during epithelial invagination in the sea urchin. *Development* **117**, 1049-60.

Lane, M. C. and Solursh, M. (1988). Dependence of sea urchin primary mesenchyme cell migration on xyloside- and sulfate-sensitive cell surface-associated components. *Dev.Biol.* **127**, 78-87.

Lane, M. C. and Solursh, M. (1991). Primary mesenchyme cell migration requires a chondroitin sulfate/ dermatan sulfate proteoglycan. *Dev.Biol.* **143**, 389-397.

Lee, J. Y. and Spicer, A. P. (2000). Hyaluronan: a multifunctional, megaDalton, stealth molecule. *Curr.Opin.Cell Biol.* **12**, 581-586.

Lee, M. H. and Schedl, T. (2001). Identification of in vivo mRNA targets of GLD-1, a maxi-KH motif containing protein required for *C. elegans* germ cell development. *Genes Dev* **15**, 2408-20.

Mariano, M. (1967). Histochemical investigation on egg shell polysaccharides in *Capillaria hepatica* (Banfroft 1918, Capillaridae, Trichuroidea). *Acta Histochem* **26**, 144-50.

Mizuguchi, S., Uyama, T., Kitagawa, H., Nomura, K. H., Dejima, K., Gengyo-Ando, K., Mitani, S., Sugahara, K. and Nomura, K. (2003). Chondroitin proteoglycans are involved in cell division of *Caenorhabditis elegans*. *Nature* **423**, 443-448.

Mounsey, A., Bauer, P. and Hope, I. A. (2002). Evidence suggesting that a fifth of annotated *Caenorhabditis elegans* genes may be pseudogenes. *Genome Res* **12**, 770-5.

Piano, F., Schetter, A. J., Morton, D. G., Gunsalus, K. C., Reinke, V., Kim, S. K. and Kempthues, K. J. (2002). Gene clustering based on RNAi phenotypes of ovary-enriched genes in *C. elegans*. *Curr Biol* **12**, 1959-64.

Reinke, V., Smith, H. E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S. J., Davis, E. B., Scherer, S., Ward, S. et al. (2000). A global profile of germline gene expression in *C. elegans*. *Mol Cell* **6**, 605-16.

Rual, J. F., Ceron, J., Koreth, J., Hao, T., Nicot, A. S., Hirozane-Kishikawa, T., Vandenhaute, J., Orkin, S. H., Hill, D. E., van den Heuvel, S. et al. (2004). Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome Res* **14**, 2162-8.

Schirmer, E. C., Florens, L., Guan, T., Yates, J. R. r. and Gerace, L. (2003). Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science* **301**, 1380-2.

Schuel, H., Kelly, J. W., Berger, E. R. and Wilson, W. L. (1974). Sulfated acid mucopolysaccharides in the cortical granules of eggs. Effects of quaternary ammonium salts on fertilization. *Exp Cell Res* **88**, 24-30.

Solursh, M., Mitchell, S. L. and Katow, H. (1986). Inhibition of cell migration in sea urchin embryos by β -D-xyloside. *Dev.Biol.* **118**, 325-332.

Sonnichsen, B., Koski, L. B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A. M., Artelt, J., Bettencourt, P., Cassin, E. et al. (2005). Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* **434**, 462-9.

Sugahara, K., Mikami, T., Uyama, T., Mizuguchi, S., Nomura, K. and Kitagawa, H. (2003). Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate. *Curr Opin Struct Biol* **13**, 612-20.

Toyoda, H., Kinoshita-Toyoda, A. and Selleck, S. B. (2000). Structural analysis of glycosaminoglycans in *Drosophila* and *Caenorhabditis elegans* and demonstration that *tout-velu*, a *Drosophila* gene related to EXT tumor suppressors, affects heparan sulfate *in vivo*. *J.Biol.Chem.* **275**, 2269-2275.

Uyama, T., Kitagawa, H., Tamura, J. and Sugahara, K. (2002). Molecular cloning and expression of human chondroitin *N*-acetylgalactosaminyltransferase - The key enzyme for chain initiation and elongation of chondroitin/dermatan sulfate on the protein linkage region tetrasaccharide shared by heparin/heparan sulfate. *J.Biol.Chem.* **277**, 8841-8846.

Uyama, T., Kitagawa, H., Tanaka, J., Tamura, J., Ogawa, T. and Sugahara, K. (2003). Molecular cloning and expression of a second chondroitin *N*-acetylgalactosaminyltransferase involved in the initiation and elongation of chondroitin/dermatan sulfate. *J Biol Chem* **278**, 3072-8.

Vacquier, V. D., Tegner, M. J. and Epel, D. (1973). Protease released from sea urchin eggs at fertilization alters the vitelline layer and aids in preventing polyspermy. *Exp Cell Res* **80**, 111-9.

Vogel, K. G., Koob, T. J. and Fisher, L. W. (1987). Characterization and interactions of a fragment of the core protein of the small proteoglycan (PGII) from bovine tendon. *Biochem Biophys Res Commun* **148**, 658-63.

Wharton, D. (1980). Nematode egg-shells. *Parasitology* **81**, 447-63.

Wharton, D. A. (1979a). Oogenesis and egg-shell formation in *Aspicularis tetraptera* Schulz (Nematoda: Oxyuroidea). *Parasitology* **78**, 131-43.

Wharton, D. A. (1979b). The structure and formation of the egg-shell of *Syphacia obvelata Rudolphi* (Nematoda: Oxyurida). *Parasitology* **79**, 13-28.

Wharton, D. A. (1979c). The structure of the egg-shell of *Aspicularis tetraptera* Schulz (Nematoda: Oxyuroidea). *Parasitology* **78**, 145-54.

Yada, T., Gotoh, M., Sato, T., Shionyu, M., Go, M., Kaseyama, H., Iwasaki, H., Kikuchi, N., Kwon, Y. D., Togayachi, A. et al. (2003a). Chondroitin sulfate synthase-2. Molecular cloning and characterization of a novel human glycosyltransferase homologous to chondroitin sulfate glucuronyltransferase, which has dual enzymatic activities. *J Biol Chem* **278**, 30235-47.

Yada, T., Sato, T., Kaseyama, H., Gotoh, M., Iwasaki, H., Kikuchi, N., Kwon, Y. D., Togayachi, A., Kudo, T., Watanabe, H. et al. (2003b). Chondroitin sulfate synthase-3. Molecular cloning and characterization. *J Biol Chem* **278**, 39711-25.

Yamada, S., Van Die, I., Van den Eijnden, D. H., Yokota, A., Kitagawa, H. and Sugahara, K. (1999). Demonstration of glycosaminoglycans in *Caenorhabditis elegans*. *FEBS Lett.* **459**, 327-331.

APPENDIX A

**Cloning, Golgi localization, and enzyme activity of the full-length
heparin/heparan sulfate-glucuronic acid C5-epimerase**

Cloning, Golgi Localization, and Enzyme Activity of the Full-length Heparin/Heparan Sulfate-Glucuronic Acid C5-epimerase*

Brett E. Crawford, Sara K. Olson, Jeffrey D. Esko, and Maria A. S. Pinhal‡

From the Department of Cellular and Molecular Medicine, Glycobiology Research and Training Center, University of California, San Diego, La Jolla, California 92093-0887

Received for publication, January 30, 2001, and in revised form, March 7, 2001
Published, JBC Papers in Press, March 12, 2001, DOI 10.1074/jbc.M100880200

While studying the cellular localization and activity of enzymes involved in heparan sulfate biosynthesis, we discovered that the published sequence for the glucuronic acid C5-epimerase responsible for the interconversion of D-glucuronic acid and L-iduronic acid residues encodes a truncated protein. Genome analysis and 5'-rapid amplification of cDNA ends was used to clone the full-length cDNA from a mouse mastocytoma cell line. The extended cDNA encodes for an additional 174 amino acids at the amino terminus of the protein. The murine sequence is 95% identical to the human epimerase identified from genomic sequences and fits with the general size and structure of the gene from *Drosophila melanogaster* and *Caenorhabditis elegans*. Full-length epimerase is predicted to have a type II transmembrane topology with a 17-amino acid transmembrane domain and an 11-amino acid cytoplasmic tail. An assay with increased sensitivity was devised that detects enzyme activity in extracts prepared from cultured cells and in recombinant proteins. Unlike other enzymes involved in glycosaminoglycan biosynthesis, the addition of a c-myc tag or green fluorescent protein to the highly conserved COOH-terminal portion of the protein inhibits its activity. The amino-terminally truncated epimerase does not localize to any cellular compartment, whereas the full-length enzyme is in the Golgi, where heparan sulfate synthesis is thought to occur.

Heparan sulfate proteoglycans are located on the cell surface and in the extracellular matrix, where they play important roles in cell adhesion, differentiation, and growth *in vitro* and *in vivo* (1–3). To a large extent, these biological activities depend on the heparan sulfate chains attached to the core protein. Heparan sulfate, a type of glycosaminoglycan, initially assembles by the copolymerization of *N*-acetyl-D-glucosamine (GlcNAc) and D-glucuronic acid (GlcA). The backbone then undergoes extensive modification initiated by the *N*-deacetylation and *N*-sulfation of subsets of GlcNAc residues. Subsequently, D-GlcA residues adjacent to the *N*-sulfated sugars are converted to L-IdoUA¹ by a C5-epimerase and are sulfated at C-2

by a specific sulfotransferase. The glucosamine units also can be sulfated at C-6 and to a lesser extent at C-3. The blocklike arrangement of the modified residues confers specific binding properties to the chains for protein ligands, which in turn facilitate various biological activities.

Many of the enzymes involved in heparan sulfate and heparin formation seem to be members of multienzyme gene families. Two exceptions are the C5-epimerase that interconverts D-GlcA and L-IdoUA and the 2-*O*-sulfotransferase that adds sulfate to C-2 of IdoUA residues and to a lesser extent GlcA residues. The C5-epimerase has been partially purified from mouse mastocytoma (4) and purified to homogeneity from bovine liver (5). A bovine cDNA for the epimerase has been cloned as well (6). Kinetic studies have clarified the substrate specificity of the epimerase, showing that the enzyme will react with both D-GlcA (forward reaction) and L-IdoUA (reverse reaction) when these residues are located toward the reducing side of *N*-sulfated glucosamine residues, but it will not react with uronic acids that are *O*-sulfated or that are adjacent to *O*-sulfated glucosamine residues (7, 8). This specificity is consistent with the overall order of modification, suggesting that epimerization begins to occur after GlcNAc *N*-deacetylation and *N*-sulfation but before glucosamine residues undergo 6-*O*-sulfation and 3-*O*-sulfation (7, 9). The fact that the epimerase seems to be represented only once in vertebrate and invertebrate genomes suggests that the extent of uronic acid epimerization depends on the level of enzyme expression and production of the *N*-sulfated tracts.

In an attempt to study the cellular localization and potential interaction of the epimerase with other enzymes in the pathway, we discovered that the published bovine sequence encodes a truncated protein.² This report provides the full-length sequence from mouse and human, an improved set of conditions for assaying the epimerase in cell extracts, and a demonstration that the enzyme is localized to the Golgi in vertebrate cells.

EXPERIMENTAL PROCEDURES

Cell Culture—Chinese hamster ovary cells (CHO-K1) were obtained from the American Type Culture Collection (CCL-61, Manassas, VA). MST cells were derived from the Furth murine mastocytoma (10). CHO cells were grown in Ham's F-12 medium (Life Technologies, Inc.), and MST cells were grown in RPMI 1640 medium. Both media were supplemented with 10% (*v/v*) fetal bovine serum (Hyclone Laboratories), 100 μ g/ml streptomycin sulfate, and 100 units/ml penicillin G. The cells were cultured at 37 °C under an atmosphere of 5% CO₂ in air at 100% relative humidity.

Cloning the Murine C5-Epimerase—A murine epimerase cDNA frag-

reaction; YFP, yellow fluorescent protein; UTR, untranslated region; PBS, phosphate-buffered saline; BSA, bovine serum albumin; PIPES, piperazine-*N,N'*-bis(2-ethanesulfonic acid); MOPS, 3-(*N*-morpholino)propanesulfonic acid; MES, 2-(*N*-morpholino)ethanesulfonic acid.

* M. A. S. Pinhal, B. Smith, J. Aikawa, K. Kimata, and J. D. Esko, unpublished results.

* This work was supported by National Institutes of Health Grant R37GM33063 (to J. D. E.), a fellowship from the FEW Latin American Fellows Program in the Biomedical Sciences (to M. A. S. P.), and National Institutes of Health Training Grants CA67734 (to B. E. C.) and GM08666 (to S. K. O.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Universidade Federal De São Paulo, Vila Clementino, CEP 04044-020, São Paulo, Brazil. E-mail: mspinh@bioc@epm.br.

¹ The abbreviations used are: IdoUA, L-iduronic acid; CHO, Chinese hamster ovary; GFP, green fluorescent protein; PCR, polymerase chain

GlcA C5-Epimerase

ment corresponding to the published bovine sequence (GenBank™ accession number AF003927) was cloned from an MST cDNA library using PCR, the forward primer 5'-ATGTCCTTTGAAGGC TACAATGTGG-3', and the reverse primer 5'-CTAGTTGTGCTTTGCCCGCTGCC-3', which anneal with the first and last 24 bases of the partial bovine sequence (6). The PCR product was blunt-end cloned into pGEM (Promega) for sequencing. Subsequently, the primers XhoEpi5' (5'-CCCCGGCTCAGGCGCCATGCTCTTGAAGCCCTACAATG-3') and BamEpi8' (5'-CTGGATCCCTAGTTGTGCTTTGCCCGG-3') were used to amplify the cDNA from the pGEM epimerase clone for transfer into pCDNA3.1 (Invitrogen) using the XhoI and BamHI sites. To generate the YFP-tagged truncated epimerase, the primers XhoEpi5' and 3'Epi-GFPBam (5'-CTGGATCCCGCTTGTGCTTTGCCCGG-3') were used to amplify the epimerase cDNA, which was cloned into the XhoI and BamHI sites of pEYFP-N1.

The cDNA containing the full-length epimerase was cloned using a primer designed to the proposed 5' end deduced from the human genomic DNA sequence (GenBank™ accession number AC026992). This primer, 5EpiXhoI (5'-CTCAGCCATGCGTTGCTGGCAGCTCG-3'), was used with an internal reverse primer, 3EpiBam (5'-GGATCCGAGATCCATGCGCTGCTACAAG-3'), to amplify the 5' 900 base pairs of the cDNA from the murine MST cDNA library. The cDNA encoding the full-length epimerase was then constructed by digesting the truncated epimerase in pCDNA3.1 with XhoI and HindIII and then by inserting the extended 5' end amplified by PCR. The GFP-tagged full-length epimerase was generated by amplifying the full coding region from pCDNA3.1 with the primers 5EpiXhoI and 3'Epi-GFPBam and by cloning into the XhoI and BamHI sites of pEGFP-N1. The coding sequence was verified by directly sequencing PCR products from three independent amplifications from the MST cDNA library. All PCR amplifications were done using Vent DNA polymerase (New England Biolabs), and clones were sequenced on an ABI 373 DNA sequencer using dye terminator cycle sequencing.

Generation of Full-length Murine Epimerase—5' Rapid amplification of cDNA ends was performed according to the manufacturer's instructions (CLONTECH) with mRNA isolated from MST cells. Two gene-specific primers were used based on the murine sequence for the epimerase described above, 3EpiBam (5'-GGATCCGAGATCCATGCGCTGCTACAAG-3') and BC15 (5'-ACATGGTGGATCTAGACTT-3'). Analysis of five independent clones, each with the same 5' end, yielded a consensus sequence.

The sequence of the 3' end of the murine coding sequence was determined by amplifying the 3' end from an MST cDNA library using BC11 (5'-GGAGACACAGAAAAGAATC-3') and BC42 (5'-GGAGACACAGAAAAGAATC-3'). BC11 was designed to anneal between nucleotides 1164 and 1188 of the mouse epimerase cDNA, whereas BC42 was designed to anneal to the 3'-untranslated region (UTR) and was designed based on nucleotides 1892–1855 of the partial human sequence (GenBank™ accession number AB020648). The PCR product was cloned, and three independent isolates were sequenced. All three clones contained two silent changes from the published human sequence. The GenBank™ accession numbers for the murine cDNA and encoded protein are AF925582 and AAG42004, respectively.

Enzyme Localization—Epimerase constructs were generated with COOH-terminal *c-myc* and GFP tags by subcloning into pCDNA3.1 containing *mycHis₆* and pEGFP (CLONTECH), respectively. An amino-terminal *c-myc*-tagged clone was generated by annealing the oligonucleotides EC46 (5'-ATGTC TAGAGAACAAAACTCATCTCAG-AAGAGGATCTGTCTAGAGCA-3') and BC47 (5'-TGCTCTAGACAGATCCTCTTCTGAGATGAGTTTTTGTCTCTAGACAT-3'), which codes for a *myc* tag. The oligonucleotides were boiled for 1 min, cooled on ice, phosphorylated with polynucleotide kinase (New England Biolabs), and cloned in-frame into the *EcoRV* site in the polylinker region of pCDNA3.1 containing the full-length epimerase. The clone used in these experiments actually contained two *Myc* tags in a tandem repeat at the amino terminus.

CHO cells were transiently transfected with 2 μ g of plasmid DNA using LipofectAMINE according to the manufacturer's directions (Life Technologies, Inc.). Cells were grown on 24-well glass microcospe slides and were processed for enzyme localization studies 24–36 h after transfection. After the cells were fixed for 1 h with 2% paraformaldehyde in 75 mM phosphate buffer, pH 7.5, they were rinsed several times with phosphate-buffered saline (PBS) (11). Cells were then permeabilized with 1% Triton X-100 (*v/v*) and 0.1% bovine serum albumin (BSA) (*w/v*) in PBS. The primary antibody, mouse anti-Myc (Invitrogen) monoclonal antibody, and rabbit polyclonal anti- α -mannosidase II antiserum (a gift from Marilyn G. Farquhar, University of California, San Diego) were diluted 1:400 in PBS with 1% BSA and incubated for 1 h with the fixed

cells. To remove the unbound primary antibody, the cells were washed several times for 30 min with PBS containing 0.1% BSA. The samples were then incubated for 1 h with the secondary antibody, anti-rabbit Cy5 (Qcovate Chemicals, NT) or anti-mouse-TRITC (Sigma), diluted 1:200 in PBS containing 1% BSA. After several washes, the cells were mounted with Vectashield containing 4',6'-diamidino-2-phenylindole for nuclear staining (Vector Laboratories). The photomicrographs shown in Fig. 5, A–D, were captured with a Photometrics charge-coupled device mounted on a Nikon microscope adapted to a DeltaVision (Applied Precision, Inc.) deconvolution imaging system. The data sets were deconvolved and analyzed using SoftWorx software (Applied Precision, Inc.) on a Silicon Graphics Octane work station. The photomicrograph shown in Fig. 5E was captured with a Hamamatsu C5810 three-color chilled charge-coupled device camera mounted on a Zeiss Axiophot ($\times 100$ lens) microscope.

Epimerase Assay—Normal and transfected cells were washed twice with cold PBS and once with cold 0.25 M sucrose in 20 mM Tris, pH 7.4, and were then scraped with a rubber policeman into 100 μ l of cold buffer containing 0.25 M sucrose, 20 mM Tris-HCl, pH 7.4, 20 μ M phenylmethylsulfonyl fluoride, 0.5 μ g/ml leupeptin, and 0.5 μ g/ml pepstatin. Cells were lysed by sonication with a microtip sonicator, and the protein concentration was quantitated with the Bradford assay (Bio-Rad) using BSA as the standard. The extracts were stable when stored at -20°C .

The epimerase substrate consisted of modified *N*-acetylheparosan and was prepared as described (12). Briefly, *Escherichia coli* K5 capsular polysaccharide was labeled *in vivo* with D-[5- ^3H]glucose (PerkinElmer Life Sciences) and purified from the growth medium. The GlcNAc residues in the labeled polysaccharide were *N*-deacetylated to near completion with anhydrous hydrazine and hydrazine sulfate (Sigma) and were *N*-sulfated with trimethylamine sulfur trioxide complex (Sigma). The concentration of *N*-acetylheparosan was determined by a carbazol assay for uronic acids (18), which yielded a radioisotopic activity of 76 cpm/pmol GlcA (49 Ci/mol).

Detection of epimerase activity was based on the release of ^3H from [5- ^3H]GlcA units in the polysaccharide and recovery as $^3\text{H}_2\text{O}$ (12). Initial assays were set up according to the published reaction conditions ("original"), which contained 50 mM HEPES, 15 mM EDTA, 100 mM KCl, and 0.015% Triton X-100, pH 7.4 (12). Protein, substrate, and various ancillary factors were adjusted to maximize the activity detected in normal and transfected cells. The "revised" assay consisted of 25 mM HEPES, pH 7.0, 0.1% Triton X-100, 800 pmol of ^3H -sulfated heparosan substrate, and 2 μ g of cell protein in a total volume of 20 μ l. Some assays contained 40 mM CaCl_2 , but divalent cations were later found not to be required. The reactions were incubated for 2 h at 37 $^{\circ}\text{C}$ and halted by the addition of 50 μ l of cold 50 mM sodium acetate buffer, pH 4.0, containing 50 mM LiCl. The sample and a 100- μ l rinse of the tube with buffer (25 mM HEPES, pH 7.0, and 0.1% Triton X-100) were transferred to a 0.4-ml column of DEAE-Sephael (Amersham Pharmacia Biotech) that was equilibrated with the same buffer. The column was washed with 0.9 ml of assay buffer, and the $^3\text{H}_2\text{O}$ recovered in the flow-through fractions was counted by liquid scintillation spectrometry using Ultima Gold (Packard Instrument Co.). A reagent blank containing everything except a source of enzyme was included as a control. This yielded values of ~ 200 cpm, which were subtracted from the experimental values that ranged from 800 to 3000 counts. All assays were done in duplicate with comparable results from three or more independent experiments.

Western Blotting—Cells were harvested, and 25 μ g of protein for each sample was analyzed by SDS-polyacrylamide gel electrophoresis on a 10% gel. The samples were transferred to a nitrocellulose membrane using the Bio-Rad Mini Protean II system. The membrane was blocked at 4 $^{\circ}\text{C}$ overnight with 4% BSA in Tris-buffered saline (10 mM Tris, pH 8.0, and 150 mM NaCl) with 0.1% Tween 20 (TBST). Mouse anti-GFP (CLONTECH) and mouse anti-Myc (Invitrogen) were diluted 1:1000 and 1:5000 in TBST, respectively, and incubated for 1 h with the membrane at room temperature with shaking. The membrane was washed three times with TBST before the application of the secondary antibody, goat anti-mouse horseradish peroxidase (Bio-Rad) diluted 1:3000 in TBST. The membrane was washed six times with TBST and developed with SuperSignal West Pico chemiluminescent substrate (Pierce).

RESULTS AND DISCUSSION

The GlcA C5-epimerase was previously purified from murine mastocytoma and bovine liver (4, 5). Sequencing the purified protein yielded proline as the amino-terminal amino acid, sug-

GlcA C5-Epimerase

```

1 ACACACGAGCGCTTCCTCGCCGAGAGGCTGGAGCTCGGGCAGCGCAGGCGTGGGCGCGCG
61 TGCTCTCGCCGCTCGCTTCAGTTTCTCCTCAGGGCTCGGGGGCGCGCTCGTCCGCCGAGCG
121 GCCTCAGGAATTAACAAGAACTGAAGTTTGTGATCAAAATTTTGAATTGAAGCA
181 GAAATGTAAGATTTGATTTCTTTCATTTGATTAGGTATGGCTGAATATGCGTTGC
M R C 3
241 TTGGCAGCTCGGGTCAACTAAGACTTTGATTTATCATCTGTGGCGTATTCACTTTGGTC
L A A R V N Y K T L I I I C A L F T L V 23
ACAGTACTTTTGGGAATAAGTGTCCAGCGACAAAAGCAATCCAGTTTCTCGGCACTTG
T V L L W N K C S S D K A I Q F P R H L 43
361 AGTAGTGGATTGAGTGGATTAGAAAAAGATCAGCAGCATCTGAAAAGTAACCAC
S S G F R V D G L E K R S A A S E S N H 63
421 TATGCCAACACATAGCCAAACAGCAGTCAGAAGAGGCATTTCTCAGGAACAACAGAAG
Y A N H I A K Q Q S E E A F P Q E Q Q K 83
481 GCACCCCTGTTTGGGGGCTTCAATAGCAACGGGGGAAGCAAGGTGTTAGGGCTCAA
A P P V V G G F N S N G G S K V L G L K 103
541 TATGAAGAGATTGACTGTCTATAACGATGACACACCATTAAGGGGAGCAGAGAGGG
Y E E I D C L I N D E H T I K G R R E G 123
601 AATGAAGTTTCTCCTCATTCACTTGGGTAGAAAATCTTGTATTTATGAAAAGTG
N E V F L P F T W V E K Y F D V Y G K V 143
661 GTCCAGTATGACGGCTATGATGATTTGAATTTCTCATAGCTATTCCAAAGTCTATGCA
V Q Y D G Y D R F E F S H S Y S K V Y A 163
721 CAGAGATCCTTATCACCTGACGGTGTGTTTATGTCCTTTGAGGCTACAATGTGGAA
Q R S P Y H P D G V F M S F E G Y N V E 183
781 GTCCGAGACAGAGTCAAATGTATAGTGAAGTTGAAGGTGCTCATTTATCCAGTGG
V R D R V K C I S G V E G V P L S T Q W 203
841 GGGCTCAAGGTATTTTCAACCAATCCAGATGACACAGATATGGGCTAAGTCAATACAG
G P Q G Y F Y P I Q I A Q Y G L S H Y S 223
901 AAGAATCTAACCGAGAAAACCCCTCAGATAGAAATATGAAAACGACAGAGACAGGGAC
K N L T E K P P H I E V Y E T A E D R D 243
961 AGAATCATCAGACCTAATGAATGGACTGTGCCAAGGGGTGCTTCATGGCCAGTGTGGCA
R N I R P N E W T V P K G C F M A S V A 263
1021 GACAAGCTAGATCCACCAATGTTAAACAGTTTATGCTCCAGAAAACAGTGAAGTGTG
D K S R S T N V K Q F I A P E T S E G V 283
1081 TCTTTGAGCTGGGAAACAAAAGCTTCATTTATTTGACCTCAAGCTTTAAACA
S L Q L G N T K D F T I S F D L K L T 303
1141 AATGGAGTGTCTGTGCTGTGGAGCCAGAAAAGAAATCAAGCTTTCACGTGGAT
N G S V S V V L E T E T E K N Q L P T V H 323
1201 TATGCTCAACACCCAGCTGATGCTTTCAGAGCAGGGACATATACACGGGATGGG
Y V S N T Q L I A F R D R D I Y Y G I G 343
1261 CCCAGAATTCATGAGTACAGTTACAGAGACCTGGTCACTGACCTCAGAAAAGGAGTG
P R T S W S T V T R D L V T D L R K G V 363
1321 GGGCTTCAACACAAAAGCTGTCAGCCACCAAAATCACTGCCAAAAGGTGGTTAGG
G L S N T K A V K P T K I M P K K V R 383
1381 TTGATTCGAAAAGGGAAGGATTCCTGGACACATACCACTCAACACAGCCCAATG
L I A K G K G F L D N I T I S T T A H M 403
1441 GCTGATTTCTTGTCAAGTCACTGCTAGTGAAGAAACAGGATGAGAAAGTGGCTGG
A A F F A A S D W L V R N Q D E K G G W 423
1501 CCAATTTGTTGACCCGGAAGTTAGGGGAAGGGTTAAATCTTTAGAACAGGATGGTAC
P I M V T R K L G E G F K S L E P G W Y 443
1561 TCTGCCATGGCAAGGGCAAGCATCTTACCTTAGTCAGGGCTATCTTCAAGCAAA
S A M A Q G Q A I S T L V R A Y L L T K 463
1621 GACTATGATCTCAGTTCAGCTTTAAGGGCAACAGCCCAATCAAGTTCCCGTCAGAG
D Y V F L S S A L R A T A F Y K P S E 483
1681 CAGCATGAGTTAAGCCCTCTCATGATAAACAATGACTGGTATGAAGAATATCCAAC
Q H G V K A V F N N K H D W Y E E Y P T 503
1741 ACACCTAGCTCTTTGTTTAAATGGCTTTATGATTTCTTAAATGGGCTGATGACCTA
T P S S F V L N G F M Y S L I G L Y D L 523
1801 AAAGAACAGCAGGGGAGACACTTGGGAAGAAGCAGGCCCTTGTACGAGCGCGCATG
K E T A G E T L G K E A R P L Y E R G M 543
1861 GAATCTCTTAAAGCCATGCTGCCCTTGTATGATACCTGGCTCCGGGACCATCTATGACCT
E S L K A M L P L Y D T G S G T I Y D L 563
1921 CGCCACTTCACTTGGCATTGCTCCCACTGGCCGCTGGGACTATCACACCCACC
R H F M L G I A P N L A R W D Y H T T H 583
1981 ATTAACAGCTGACGCTGCTCAGCACCATGATGAGTCCCAATCTCAAGAATTTGTC
I N Q L Q L L S T I D E S P I F K E F V 603
2041 AAGAGGTGAAAAGCTACCTTAAAGCAGTAGGGCAAGCACAACCTAG
K R W K S Y L K G S R A K H N * 618

```

Fig. 2. Full-length cDNA of murine C5-epimerase. The isolation of the cDNA encoding the full-length epimerase is described in the text. The revised sequence adds another 522 bases of coding nucleotides that translate into 174 additional amino acids on the amino terminus. In addition, the cDNA contains a 281-base 5'-UTR. The transmembrane domain identified by PSORT is enclosed by a *box* extending from residues 12–28. The initiating methionine residue indicated in the previously published bovine cDNA is enclosed by a *circle* (residue 176), and the amino-terminal proline residue of the purified bovine epimerase is enclosed by a *box* at residue 248. Three potential Asn-linked glycans may be attached at residues 225, 304, and 394 (*shaded boxes*). The GenBank™ accession number for the murine cDNA is AF325532.

cations, such as Ca^{2+} or Mg^{2+} , and EDTA (up to 40 mM) had no effect (4, 7). The activity was highly dependent on detergent, even in sonicated extracts, with maximal effects obtained with 0.1% Triton X-100 (Fig. 3B). However, other detergents inhibited the reaction, suggesting that the effect was not merely because of solubilization of the protein from membranes. The pH optimum was ~ 7.0 , which is in general agreement with previous findings (Fig. 3C) (4, 5, 7), but the activity showed marked sensitivity to the type of buffer (Fig. 3D). HEPES was found to be optimal. Under the revised conditions, the reaction was proportional with time for over 2 h and with protein concentration in the range of 1–30 μg . The K_m of the enzyme for the *N*-deacetylated/*N*-sulfated heparosan was estimated to be

25 μM GlcA equivalents (~ 500 pmol of GlcA/assay). With 300 pmol of GlcA/assay, a 4.5-fold increase in epimerase activity in MST cell extracts and a 6.5-fold enhancement in CHO cell extracts were observed compared with the original conditions (Fig. 3E). At lower concentrations of substrate, the difference was even more dramatic (data not shown).

Transient transfection of CHO cells revealed 4-fold greater activity associated with the full-length protein compared with the truncated enzyme in the revised assay (Fig. 4). Increasing the substrate 10-fold did not enhance the rate of reaction for either recombinant enzyme (data not shown). These findings indicated that the natural amino terminus was not a prerequisite to detect activity, which is consistent with previous find-

GlcA C5-Epimerase

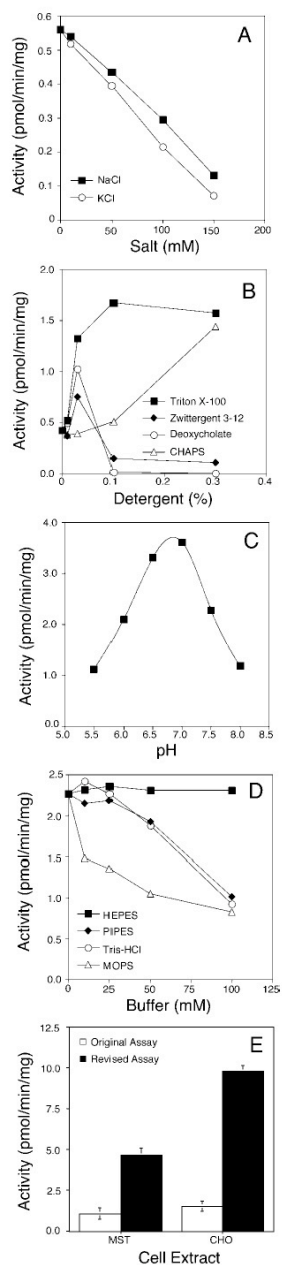


FIG. 3. Effects of salt, detergent, pH, and buffer on epimerase activity. Cell extracts prepared from MST cells (A–D) were assayed for

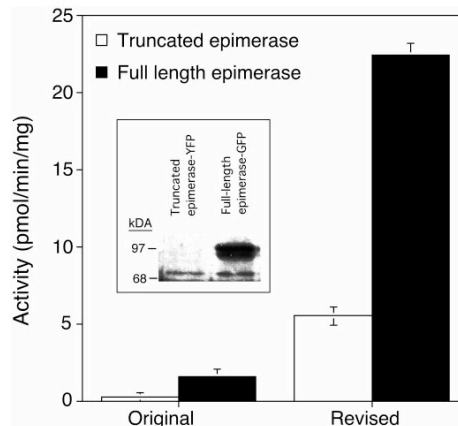


FIG. 4. Enzymatic activity of truncated and full-length epimerase. CHO cells were transiently transfected with a truncated epimerase construct representing either amino acids 175–618 or a full-length construct. After 48 h, cell extracts were prepared and assayed for enzymatic activity under original and revised reaction conditions (described under "Experimental Procedures"). Endogenous CHO C5-epimerase activities of 2.4 pmol/min/mg under the original assay conditions and 10.2 pmol/min/mg under the revised conditions were subtracted from the respective data points to obtain the activities of the transfected proteins. Error bars represent the standard error of the mean. Experiments were done in duplicate and were reproducible with independent transfections. Inset, Western blot analysis of full-length epimerase-GFP and truncated epimerase-YFP expressed transiently in CHO cells.

ings showing that the truncated protein purified from liver and mastocytoma had substantial activity (4, 5, 12, 17). Extracts prepared from cells transfected with epimerase containing a COOH-terminal GFP or YFP tag were analyzed by Western blotting with an anti-GFP monoclonal antibody. As shown in the inset of Fig. 4, the tagged full-length protein was present at higher levels than the truncated enzyme. Both forms were engineered into a near perfect Kozak sequence in the expression vector, suggesting that their expression was similar. Thus, we believe that the lower amount of the truncated enzyme was caused by decreased stability. As shown below, the truncated enzyme was also mislocalized, which may add to its instability. Thus, the amino-terminal domain does not seem to enhance the intrinsic activity of the enzyme.

Fusing *c-myc* or GFP to the COOH terminus resulted in a dramatic reduction of enzyme activity (<1 pmol/min/mg), but when a *c-myc* tag was placed at the amino terminus, enzyme activity was normal (23 pmol/min/mg versus 26 pmol/min/mg, respectively). These findings suggested that the highly con-

epimerase activity under various reaction conditions. Initial conditions consisted of 300 pmol of ^3H -substrate and 50 μg of cell extract in addition to the listed components. A, salt dependence was measured in a reaction of 50 mM PIPES, pH 6.5, 15 mM EDTA, 0.015% Triton X-100, and 0–150 mM salt. B, detergent preference was assayed in a reaction of 50 mM PIPES, pH 6.5, 15 mM EDTA, and 0–0.3% detergent. C, pH dependence was determined in the presence of 40 mM CaCl_2 and 0.1% Triton X-100 and buffer solutions of 25 mM MOES/25 mM MES at pH values ranging from 5.5 to 8.0. D, buffer specificity was measured in a reaction of 40 mM CaCl_2 , 0.1% Triton X-100, and 0–100 mM of various buffers at pH 7.0. E, parallel assays were run with MST and CHO cell extract under original conditions (12) and the revised conditions described under "Experimental Procedures." Error bars represent standard error of the mean. All assays were done in duplicate.

GlcA C5-Epimerase

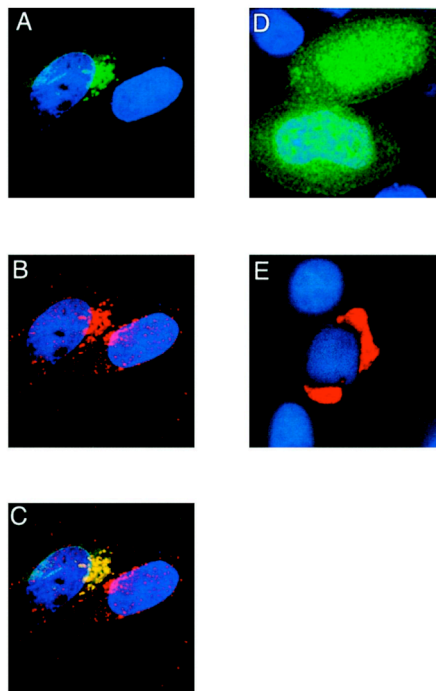


FIG. 5. **The full-length epimerase localizes to the Golgi.** CHO cells were transfected with GFP and Myc-tagged epimerase constructs. Two days later, the cells were processed for fluorescence microscopy (see "Experimental Procedures"). A, GFP (COOH-terminal)-tagged full-length epimerase. B, α -mannosidase, a marker of the medial Golgi. C, a digital overlay of A and B. D, cytoplasmic localization of the truncated epimerase with a COOH-terminal YFP tag. E, Golgi localization of the full-length epimerase containing an amino-terminal myc tag.

served COOH terminus plays an important role in binding, conformation, or catalysis. Recent investigations into the catalytic mechanism of the C5-epimerase implicated two polyprotic bases in the proton exchanges at C-5 (17) that are possibly mediated by two lysine residues. Interestingly, two lysine residues (amino acids 547 and 616) in the COOH-terminal domain of the epimerase are highly conserved across phylogeny (Fig. 1, *asterisks*). Adding GFP to the COOH terminus of other enzymes involved in heparin/heparan sulfate biosynthesis does not result in loss of activity (18).²

The Full-length Epimerase Localizes to the Golgi—To study the intracellular localization of epimerase, cDNAs encoding the truncated and full-length enzymes were fused to GFP or *c-myc* and expressed in CHO cells. Full-length epimerase was located in a juxtannuclear position, co-localizing with the Golgi marker, α -mannosidase II (Fig. 5, A–C). This localization was observed with tags on either the C or amino terminus, indicating that the location of the tag did not interfere with subcellular localization signals in the protein (Fig. 5E). When the truncated epimerase was expressed, it behaved as a soluble protein exhibiting diffuse cytoplasmic staining (Fig. 5D). This is not an unexpected result given that the protein lacks a signal peptide.

The mislocalization of the truncated enzyme may act to destabilize its structure and activity (Fig. 4, *inset*).

Future studies of the epimerase will be greatly expedited by having the full-length sequence. Interestingly, very little information is available about the function of IdoUA in the biological activity of heparin and heparan sulfate. In general, it is assumed that the greater conformational flexibility of IdoUA will enhance the binding opportunities for heparin and heparan sulfate (19). The best studied example is the interaction of antithrombin with a heparin pentasaccharide, in which a critical IdoUA residue located to the reducing side of a central 3-O-sulfated glucosamine unit confers high affinity binding to antithrombin (20). Fibroblast growth factor-2 also apparently requires at least one IdoUA unit for binding and activation (21, 22). In the former case, the addition of the 2-O-sulfate group to the IdoUA residue seems to be dispensable (20, 23), whereas in the latter it is essential for binding (24–26). These findings suggest that in some cases the IdoUA may play a direct role in binding to the ligand, whereas in others it may simply serve as a scaffold for placement of a critical sulfate residue. In both cases, the epimerase plays an essential role in creating the preferred binding site for the ligand. With full-length recombinant enzyme now available, it should be possible to engineer binding sites in isolated oligosaccharides and to explore the function of epimerase *in vivo* by creating mutants in cells and model organisms.

Acknowledgments—We thank James Feramisco and Brian Smith from the Digital Imaging Shared Resource at the University of California, San Diego Cancer Center for their help in the deconvolution microscopy.

REFERENCES

- Lindahl, U., Kusche-Gullberg, M., and Kjellén, L. (1998) *J. Biol. Chem.* **273**, 24979–24982
- Selleck, S. B. (2000) *Trends Genet.* **16**, 206–212
- Park, P. W., Reizes, O., and Bernfield, M. (2000) *J. Biol. Chem.* **275**, 29923–29926
- Malmström, A., Rodén, L., Feingold, D. S., Jacobsson, I., Bäckström, G., and Lindahl, U. (1980) *J. Biol. Chem.* **255**, 3878–3883
- Campbell, P., Hannesson, H. H., Sandback, D., Rodén, L., Lindahl, U., and Li, J. (1994) *J. Biol. Chem.* **269**, 26953–26958
- Li, J. P., Hagner-McWhirter, A., Kjellén, L., Palgi, J., Jalkanen, M., and Lindahl, U. (1997) *J. Biol. Chem.* **272**, 28158–28163
- Jacobsson, I., Bäckström, G., Hook, M., Lindahl, U., Feingold, D. S., Malmström, A., and Rodén, L. (1979) *J. Biol. Chem.* **254**, 2975–2982
- Jacobsson, I., Lindahl, U., Jensen, J. W., Rodén, L., Prihar, H., and Feingold, D. S. (1984) *J. Biol. Chem.* **259**, 1056–1063
- Lindahl, U., Jacobsson, I., Hook, M., Bäckström, G., and Feingold, D. S. (1976) *Biochem. Biophys. Res. Commun.* **70**, 492–499
- Montgomery, R. L., Lidholt, K., Flay, N. W., Liang, J., Vertel, B., Lindahl, U., and Esko, J. D. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 11327–11331
- Dulbecco, R., and Vogt, M. (1954) *J. Exp. Med.* **99**, 167–182
- Hagner-McWhirter, A., Hannesson, H. H., Campbell, P., Westley, J., Rodén, L., Lindahl, U., and Li, J. P. (2000) *Glycobiology* **10**, 159–171
- Bitter, T., and Muir, H. M. (1982) *Anal. Biochem.* **4**, 330–334
- Nakai, K., and Kanehisa, M. (1992) *Genomics* **14**, 897–911
- Kozak, M. (1996) *Mamm. Genome* **7**, 563–574
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402
- Hagner-McWhirter, A., Lindahl, U., and Li, J. P. (2000) *Biochem. J.* **347**, 69–75
- McCormick, C., Duncan, G., Goutsos, K. T., and Tufaro, F. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 668–673
- Mulloy, B., and Forster, M. J. (2000) *Glycobiology* **10**, 1147–1156
- Atha, D. H., Lormeau, J. C., Petitou, M., Rosenberg, R. D., and Choay, J. (1985) *Biochemistry* **24**, 6723–6729
- Maccarana, M., Casu, B., and Lindahl, U. (1993) *J. Biol. Chem.* **268**, 23898–23905
- Guimond, S., Maccarana, M., Olwin, B. B., Lindahl, U., and Rapraeger, A. C. (1993) *J. Biol. Chem.* **268**, 23906–23914
- Björk, I., and Lindahl, U. (1982) *Mol. Cell. Biochem.* **48**, 161–182
- Bai, X. M., and Esko, J. D. (1996) *J. Biol. Chem.* **271**, 17711–17717
- Pellegrini, L., Burke, D. F., Von Delft, F., Mulloy, B., and Blundell, T. L. (2000) *Nature* **407**, 1029–1034
- Schlessinger, J., Plotnikov, A. N., Ibrahim, O. A., Eliseenkova, A. V., Yeh, B. K., Yayon, A., Linhardt, R. J., and Mohammadi, M. (2000) *Mol. Cell* **6**, 743–750

Appendix A is a reprint of the material as it appeared in the Journal of Biological Chemistry, Crawford, B.E., Olson, S.K., Esko, J.D. and Pinhal, M.A. (2001). The dissertation author was a secondary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

APPENDIX B

**Enzyme interactions in heparan sulfate biosynthesis:
uronosyl 5-epimerase and 2-O-sulfotransferase interact *in vivo***

Enzyme interactions in heparan sulfate biosynthesis: Uronosyl 5-epimerase and 2-O-sulfotransferase interact *in vivo*

Maria A. S. Pinhal^{1*}, Brian Smith^{2*}, Sara Olson^{3*}, Jun-ichi Aikawa⁴, Koji Kimata⁵, and Jeffrey D. Esko^{1†}

¹Department of Cellular and Molecular Medicine, Glycobiology Research and Training Center, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0687; ²Cellular Biochemistry Laboratory, The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako-Shi, Saitama 351-0198 Japan; and ³Aichi Medical University, Institute for Molecular Science of Medicine, Yazako, Nagakute Aichi 480-1195, Japan

Edited by William J. Lennarz, State University of New York, Stony Brook, NY, and approved September 12, 2001 (received for review April 9, 2001)

The formation of heparan sulfate occurs within the lumen of the endoplasmic reticulum–Golgi complex–trans-Golgi network by the concerted action of several glycosyltransferases, an epimerase, and multiple sulfotransferases. In this report, we have examined the location and interaction of tagged forms of five of the biosynthetic enzymes: galactosyltransferase I and glucuronosyltransferase I, required for the formation of the linkage region, and GlcNAc *N*-deacetylase/*N*-sulfotransferase 1, uronosyl 5-epimerase, and uronosyl 2-*O*-sulfotransferase, the first three enzymes involved in the modification of the chains. All of the enzymes colocalized with the medial-Golgi marker α -mannosidase II. To study whether any of these enzymes interacted with each other, they were relocated to the endoplasmic reticulum (ER) by replacing their cytoplasmic N-terminal tails with an ER retention signal derived from the cytoplasmic domain of human invariant chain (p33). Relocating either galactosyltransferase I or glucuronosyltransferase I had no effect on the other's location or activity. However, relocating the epimerase to the ER caused a parallel redistribution of the 2-*O*-sulfotransferase. Transfected epimerase was also located in the ER in a cell mutant lacking the 2-*O*-sulfotransferase, but moved to the Golgi when the cells were transfected with 2-*O*-sulfotransferase cDNA. Epimerase activity was depressed in the mutant, but increased upon restoration of 2-*O*-sulfotransferase, suggesting that their physical association was required for both epimerase stability and translocation to the Golgi. These findings provide *in vivo* evidence for the formation of complexes among enzymes involved in heparan sulfate biosynthesis. The functional significance of these complexes may relate to the rapidity of heparan sulfate formation.

glycosaminoglycans | sulfation | epimerization | enzyme localization | Golgi complex

The biosynthesis of heparan sulfate initiates by the translation of a proteoglycan core protein and the assembly of the so-called linkage region tetrasaccharide on specific serine residues (–GlcA β 1,3Gal β 1,3Gal β 1,4Xyl β 1-O-Ser). The chain then polymerizes by the addition of alternating *N*-acetylglucosamine (GlcNAc α 1,4) and glucuronic acid (GlcA β 1,4) residues. A series of modification reactions takes place simultaneously that involves at least six enzymatic activities: (i) *N*-deacetylation of a portion of GlcNAc residues, (ii) *N*-sulfation of the resulting unsubstituted amino groups to form GlcNS units, (iii) 5-epimerization of adjacent D-GlcA residues to form L-iduronic acid (IdoA), (iv) 2-*O*-sulfation of L-IdoA and more rarely D-GlcA residues, (v) 6-*O*-sulfation of glucosamine units, and (vi) occasional 3-*O*-sulfation of glucosamine residues (Fig. 1A). A major question concerns how these enzymes orchestrate the formation of specific oligosaccharide sequences with unique binding properties for ligands (1, 2). Multiple isozymes exist for several of the transferases that differ in substrate specificity and temporal/spatial expression during development (3–5). The physical

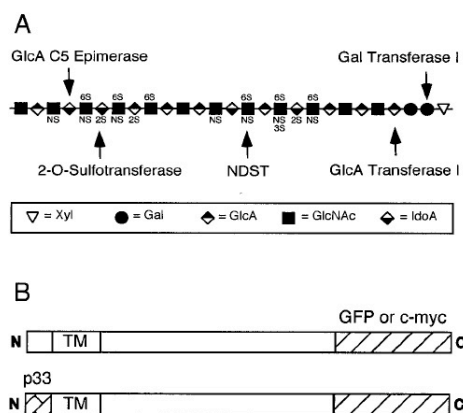


Fig. 1. Heparan sulfate and chimeric enzymes. (A) Schematic representation of heparan sulfate. The arrows indicate the sites of action for the enzymes examined in this study. NDST, *N*-deacetylase/*N*-sulfotransferase 1. (B) To localize the enzymes, green fluorescent protein (GFP) was spliced on the C termini of the enzymes. To relocate the enzymes, the cytoplasmic tail of the enzymes was replaced with the endoplasmic reticulum (ER)-retention domain from human invariant chain (p33). TM, transmembrane domain.

arrangement of the enzymes in the Golgi complex could also play a role.

With the exception of one glucosaminyl 3-*O*-sulfotransferase isozyme (6), all of enzymes involved in heparan sulfate synthesis are type II membrane proteins, with relatively short cytoplasmic tails, a single hydrophobic transmembrane segment, and a stem region that is thought to support a lumenally oriented globular catalytic domain. Like most enzymes involved in glycosylation,

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: CHO, Chinese hamster ovary; Epi, heparan sulfate uronosyl 5-epimerase; ER, endoplasmic reticulum; GalTI, xylose-UDP-Gal galactosyltransferase I; GFP, green fluorescent protein; GlcA, glucuronic acid; GlcATI, galactose-UDP-GlcA glucuronosyltransferase I; GlcNAc, *N*-acetylglucosamine; IdoA, iduronic acid; NDST, GlcNAc-*N*-deacetylase/*N*-sulfotransferase; 2OST, uronosyl/PAPS 2-*O*-sulfotransferase (PAPS, 3'-phosphoadenosine 5'-phosphosulfate); p33, the cytoplasmic domain of human invariant chain (hIIP33); TGN, trans-Golgi network.

^{*}Present address: Universidade Federal De São Paulo, Rua Três De Maio, Nº 100, 4º Andar—Biologia Molecular, Vila Clementino, São Paulo CEP. 04044020, Brazil.

[†]To whom reprint requests should be sent: E-mail: jesko@ucsd.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

all of the enzymes of heparan sulfate synthesis are thought to reside in elements of the ER–Golgi–trans-Golgi network (TGN) membrane system (7, 8). Evidence suggests that the transmembrane domains and/or the stem sequences play important roles in sorting and retention of the enzymes in their respective locations (9–11). Differential location of the enzymes and the potential for supramolecular complexes of multiple enzymes could help channel substrates in a selective manner and affect the structure of the chains (12).

In this report, we have examined the location and potential interaction of enzymes involved in heparan sulfate synthesis by using an *in vivo* approach pioneered by Nilsson and coworkers in their study of glycoprotein processing enzymes in the Golgi (13, 14). These investigators grafted an ER retention sequence on the N terminus of Golgi glucosaminyltransferase I and discovered that retention of the enzyme in the ER simultaneously relocated α -mannosidase II. This finding led to the kin-recognition hypothesis as a means for colocalization of Golgi enzymes involved in a common pathway (10, 13, 14). Using this technique, we now show that two of the enzymes that act early in heparan sulfate biosynthesis localize independently (galactosyltransferase I and glucuronosyltransferase I), whereas two of the modification enzymes that act later in the pathway physically interact [uronosyl 2-O-sulfotransferase (2OST) and 5-epimerase (Epi)]. Complex formation affects the location and activity of the epimerase and may play a functional role in chain modification by ensuring rapid and complete 2-O-sulfation of iduronic acid (IdoA) units generated by the epimerase.

Materials and Methods

Cell Culture. Chinese hamster ovary cells (CHO-K1) were obtained from the American Type Culture Collection (ATCC CCL61). CHO mutants pgsF17, deficient in 2OST, pgsG-224, deficient in glucuronosyltransferase I, and pgsB-761, deficient in galactosyltransferase I were described previously (15–17). The cells were grown under an atmosphere of 5% CO₂ in air and 100% relative humidity in Ham's F-12 growth medium or F12/DMEM (1:1, vol/vol) (Life Technologies) supplemented with 10% (vol/vol) FBS (HyClone), 100 μ g/ml streptomycin sulfate, and 100 units/ml penicillin G.

Recombinant DNA. Full-length cDNA encoding human invariant chain (hIip33) was obtained from Michael R. Jackson (Johnson Research Institute, La Jolla, CA). PCR fragments encoding the first 282 bp of hIip33 cDNA were generated with an *NheI* restriction site at the 5' end and *EcoRI* or *XhoI* restriction sites at the 3' end. After digestion with the appropriate restriction enzymes, the fragment was ligated to the N terminus of each specific enzyme, just before the first amino acid of the predicted transmembrane domain. cDNAs encoding murine GlcNAc *N*-deacetylase/*N*-sulfotransferase (NDST1) and galactosyltransferase I (GalTI) and CHO glucuronosyltransferase I (GlcATI), 2OST, and Epi were inserted in-frame with GFP (pEGFP-N1) or with a *c*-Myc-tag (pcDNA3.1) at the C terminus. Cells were transfected by using Lipofectamine (Life Technologies) in accordance with the manufacturer's instructions. In transient expression experiments, the cells were analyzed after 2 days. Stable cell clones were selected in 400 μ g/ml G418 and screened by flow cytometry to obtain strains with comparable levels of GFP or Myc-tag expression.

Fluorescence Microscopy. Cells grown on 24-well glass microscope slides were rinsed with PBS (18), fixed for 1 h with 2% paraformaldehyde in 75 mM phosphate buffer, and permeabilized with 0.1% Triton X-100 in PBS containing 0.1% BSA. Primary antibodies against α -mannosidase II (medial Golgi marker, rabbit polyclonal antibody, a gift from Marilyn G. Farquhar, Univ. of California at San Diego), calreticulin (ER

marker, rabbit polyclonal antibody, StressGen Biotechnologies, Victoria, Canada), and mouse monoclonal anti-Myc (Invitrogen) were diluted 1:400 in PBS containing 1% BSA and incubated with fixed cells in a humid chamber for 1 h. After several washes the cells were incubated with 4',6-diamidino-2-phenylindole (DAPI) to stain nuclei and a secondary antibody (anti-rabbit IgG coupled with CY5 from Accurate Chemicals, or anti-rabbit IgG coupled with Alexa from Molecular Probes) diluted 1:200 in PBS containing 1% BSA. Myc-tagged proteins were visualized with tetramethylrhodamine isothiocyanate (TRITC)-labeled anti-mouse IgG (Sigma). The coverslips were washed several times with PBS containing 0.1% BSA and mounted with Vectashield mounting medium (Vector Laboratories). Digital images were captured with a Hamamatsu C5810 three-color chilled charge-coupled device (CCD) camera mounted on a Zeiss Axiophot (100 \times lens) microscope. For deconvolution, images were captured with a Photometrics CCD mounted on a Nikon microscope adapted to a DeltaVision (Applied Precision, Issaquah, WA) deconvolution imaging system. The data sets were deconvoluted and analyzed by using SOFTWORKX software (Applied Precision) on a Silicon Graphics Octane workstation.

Detergent and Salt Extraction. As described in ref. 19, cells were harvested by scraping in 50 mM Mes buffer (2-*N*-morpholinoethane-sulfonic acid), pH 6.5, containing 0.5% Triton X-100 and protease inhibitors (0.1 mM phenylmethylsulfonyl fluoride, 10 μ M pepstatin A, 10 μ M leupeptin, and 25 μ g/ml aprotinin). The samples were incubated with 0–300 mM NaCl for 1 h with gentle rotation on ice, and then centrifuged for 15 min at 15,000 \times *g*. The supernatant was then centrifuged for 1 h at 100,000 \times *g*. The final supernatants were adjusted to the same volume in SDS sample buffer, separated by SDS/PAGE, and transferred electrophoretically to nitrocellulose membranes. Myc-tagged proteins were detected with anti-Myc mAb (Invitrogen), and chimeras containing GFP were detected with an anti-GFP mAb (JL-8; CLONTECH). The secondary antibody, goat anti-mouse IgG (H + L) conjugated to horseradish peroxidase (Bio-Rad) was detected by chemiluminescence.

Immunoprecipitation. pgsF17 cells stably transfected with 2OST-GFP were transiently transfected with Epi-Myc. After 2 days, the cells were lysed in 50 mM Tris-HCl pH 7.4/0.5% Triton X-100/50 mM NaCl/5 mM CaCl₂/5 mM MgCl₂/20 mM benzamidine/1 mM PMSF/50 mM aminocaproic acid/10 mM iodoacetamide. The lysates were centrifuged at 12,000 \times *g* for 20 min, precleared for 30 min with 30 μ l of staphylococcal protein A-Sepharose (Amersham Pharmacia) at 4°C, and then incubated with 0.5 mg of mouse anti-GFP mAb for 4 h, followed by incubation with 50 μ l of Protein A-Sepharose overnight. The beads were washed three times with PBS, boiled for 5 min in 50 μ l of SDS-PAGE sample buffer and analyzed by SDS-PAGE electrophoresis. Proteins were transferred to nitrocellulose membrane (Millipore), blocked with 4% albumin in PBS, containing 0.05% Tween-20 and developed with mouse anti-GFP mAb or anti-Myc mAb. Epi-Myc and 2OST-GFP were detected with calf anti-mouse IgG secondary antibody conjugated to horseradish peroxidase. The specific proteins were detected by using ECL Western blotting detection kit (Amersham Pharmacia) and exposed to x-ray film (Kodak) for 1 min.

Enzymatic Assays. Enzyme activities were measured in cell homogenates as described previously (4, 15–17, 20). Cells were grown to confluence in six-well plates, rinsed three times with cold PBS, and detached with a rubber policeman in 40 μ l of buffer containing 0.25 M sucrose, 50 mM Tris-HCl (pH 7.5), 1% (wt/vol) Triton X-100, 1 μ g/ml leupeptin, 1 μ g/ml pepstatin A, and 1 mM PMSF. Aliquots of the cell extracts were stored at –20°C and sonicated before assay.

Results

Heparan sulfate synthesis occurs in a sequential fashion by the transfer of single monosaccharides to a nascent chain attached to a proteoglycan protein core (Fig. 1A). The initial reactions consist of xylosylation at specific serine residues, followed by attachment of two galactose units and GlcA. The alternating addition of GlcNAc and GlcA units forms the backbone of the chain, and a series of modification reactions then ensues, in which subsets of GlcNAc residues undergo *N*-deacetylation and *N*-sulfation, adjacent D-GlcA units epimerize to L-IdoA, and the uronic acids and glucosaminyl residues undergo sulfation at various positions (Fig. 1A). Prior studies have shown that the assembly process is quite rapid (21), suggesting that the enzymes may be located in the same subcellular compartment, possibly in complexes, to facilitate channeling of reaction products to downstream enzymes in the pathway. Two of the enzymes actually exist as bifunctional proteins containing dual activities: NDST1 isozymes catalyze GlcNAc *N*-deacetylation and *N*-sulfation, and EXTs catalyze GlcNAc and GlcA copolymerization (2). Here, we have examined the location and interaction of the enzymes in greater detail by expressing recombinant forms of GalTI and GlcATI, which catalyze early reactions involved in the formation of the linkage region tetrasaccharide, and Epi and 2OST, which modify the GlcA residues after chain polymerization.

Location and Relocation of GalTI and GlcATI. Chimeric forms of GalTI and GlcATI containing C-terminal GFP were introduced into CHO cells (Fig. 1B), where they colocalized with the medial Golgi marker, α -mannosidase II (22) (Fig. 2A and C, respectively). Both of these chimeric enzymes were enzymatically active as judged by activity assays and their ability to correct heparan sulfate biosynthesis in corresponding mutants of CHO cells (16, 17). By deconvolution fluorescence microscopy, GlcATI-GFP completely colocalized with α -mannosidase II, whereas GalTI-GFP was located in compartments shared by a α -mannosidase II as well as a nearby compartment that does not contain the marker. Because GalTI action precedes GlcATI biosynthetically, this compartment may represent the *cis* Golgi network or the ER-Golgi intermediate compartment (23).

Since GalTI and GlcATI overlap in the same compartment, we next determined whether they could associate physically. In these experiments, we used a retargeting method described by Nilsson *et al.* (14), in which an ER retention peptide from the human invariant chain (hIip33) is spliced onto the N terminus of one of the transferases (Fig. 1B). The recombinant proteins, designated p33-GalTI-GFP and p33-GlcATI-GFP, were enzymatically active, but were located in the ER, colocalizing with the ER marker calreticulin (Fig. 2B and D, respectively). To test for interactions, a Myc-tagged form of p33-GalTI was prepared and cotransfected with GlcATI-GFP. These recombinant proteins migrated independently to the ER and Golgi compartments, respectively (Fig. 2E). Similar results were obtained when p33-GlcATI-Myc was cotransfected with GalTI-GFP (data not shown). Furthermore, the location of each chimeric enzyme was unaffected when expressed in mutant cell lines lacking endogenous GalTI (17) or GlcATI (16) (data not shown), suggesting that the native forms of the enzymes do not interact with the chimeras. Thus, GalTI and GlcATI overlap in location, but do not appear to physically interact.

Location and Relocation of 2OST and Epi. We next examined the location and interaction of Epi, which converts GlcA units in heparan sulfate to IdoA, and 2OST, which adds sulfate preferentially to the C2 position of the resulting IdoA units, preventing reversal of the epimerization reaction (24). Like GalTI and GlcATI, tagged forms of Epi and 2OST showed a typical

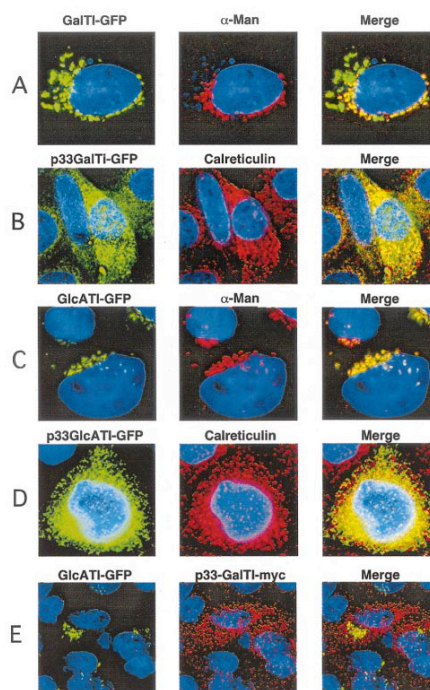


Fig. 2. Localization of chimeric enzymes involved in the formation of linkage region. Wild-type CHO cells were transiently transfected with the indicated chimeric enzymes and imaged by deconvolution microscopy (*Materials and Methods*). (A and C) Golgi localization of GalTI-GFP and GlcATI-GFP, respectively. (B and D) ER localization of p33-GalTI-GFP and p33-GlcATI-GFP, respectively. (E) Coexpression of GlcATI-GFP (green) and p33-GalTI-Myc (red). α -Man, α -mannosidase II, the medial Golgi marker; Calreticulin, ER marker; Merge, colocalization (yellow) of the chimeric enzymes with each other or the markers.

juxtannuclear Golgi position, colocalizing with α -mannosidase II (Fig. 3A and B). The p33 tagged forms of these enzymes also relocated to the ER, colocalizing with calreticulin (Fig. 3C shows p33-Epi-Myc). However, in contrast to the behavior of GalTI and GlcATI, 2OST-GFP in cells containing p33-Epi-Myc was found in the ER (Fig. 3C). This relocation of 2OST was selective because GFP-tagged forms of NDST1, GalTI, and GlcATI maintained their typical Golgi localization in cells containing p33-Epi-Myc (data not shown).

To exclude the possibility that the relocation of 2OST by p33-Epi was an artifact caused by protein overexpression or the introduction of the C-terminal tags, we examined the location of Epi-Myc in pgsF17 mutant CHO cells. This mutant lacks endogenous 2OST activity and mRNA, suggesting that it also does not contain 2OST protein (15). In these cells, 2OST-GFP and NDST1-GFP colocalized with α -mannosidase II (Fig. 4A and B), but transfected Epi-Myc localized to the ER, coinciding with calreticulin (Fig. 4C). The relocation of Epi to the ER in these cells suggested that the movement of Epi to the Golgi may depend on the presence of 2OST. To confirm this idea,

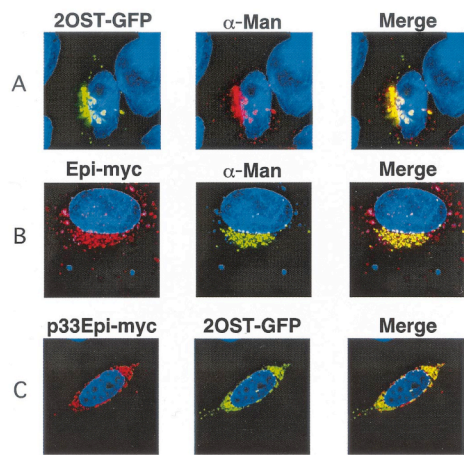


Fig. 3. ER-localized Epi relocates 2OST. (A and B) Wild-type CHO cells were transiently transfected with chimeric 2OST-GFP or Epi-Myc and their location was determined by deconvolution microscopy (*Materials and Methods*). (C) Cells were cotransfected with p33-Epi-Myc and 2OST-GFP. The two enzymes colocalized in the ER, in the compartments containing the calreticulin marker. α -Man, α -mannosidase II, the medial Golgi marker; Calreticulin, the ER marker; Merge, colocalization (yellow) of the chimeric enzymes with each other or the markers.

pgsF17 cells were transfected with a cDNA encoding wild-type 2OST. Reexpression of the enzyme reverted the location of Epi to the Golgi, and its new location completely coincided with 2OST-GFP (Fig. 4D). Introduction of p33-2OST in the mutant did not affect the location of Epi (data not shown). Together with the results shown in Fig. 3, these findings provided strong evidence that 2OST and Epi can form a complex and that the interaction is required for translocation of both Epi and 2OST to the Golgi.

Complexes of Golgi enzymes have been inferred to exist by a requirement for high salt for solubilization from detergent extracts of cell membranes (19). For example, glucosaminyltransferase I and α -mannosidase require high salt for solubilization and sediment as complexes (19). These enzymes also exhibit mutual relocation when one of the proteins is re-targeted by using the p33 grafting method (14). In contrast, late-acting Golgi enzymes involved in glycoprotein assembly (β 1,4-galactosyltransferase and α 1,2-fucosyltransferase) readily solubilize in low salt and migrate as monomers in sucrose density gradient centrifugation. When the solubilization scheme was applied to Epi and 2OST, we found that both enzymes required high salt for solubilization, consistent with the idea that they resided in the Golgi as a complex. In contrast, GlcATII-GFP solubilized readily in salt-free buffer (Fig. 5A).

To confirm the interaction, samples of pgsF17 cells, stably transfected with 2OST-GFP, were transiently transfected with Epi-Myc and reacted with an anti-GFP mAb. The immunoprecipitates were then analyzed by SDS/PAGE and Western blotting with anti-GFP mAb or anti-Myc mAb (Fig. 5B). The anti-Myc mAb reacted with single band that migrated at the expected molecular mass of the epimerase (20). The mAb against GFP reacted with a somewhat larger band at the expected mass of the chimera (\approx 70 kDa). The anti-GFP mAb

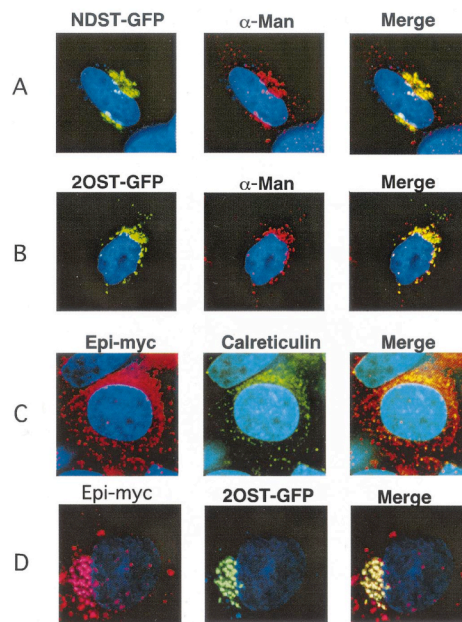


Fig. 4. Golgi localization of Epi depends on the presence of 2OST. pgsF17 cells lacking endogenous 2OST were transiently transfected with NDST1-GFP (A), 2OST-GFP (B), or Epi-Myc (C). NDST1 and 2OST colocalized with the α -mannosidase marker, whereas Epi-Myc was found in the ER. In D, the mutant was stably transfected with 2OST cDNA to correct the enzyme deficiency, and then transiently transfected with 2OST-GFP or Epi-Myc.

also reacted with proteins that varied in position and intensity in independent experiments, both in transfected and nontransfected cells, suggesting that these bands were nonspecific.

Location of Epi Alters Enzyme Activity. To determine whether the interaction of Epi and 2OST might affect enzyme activity, we measured epimerase in extracts prepared from pgsF17 and wild-type cells. As shown in Fig. 6, extracts of pgsF17 cells contained less Epi activity compared with the wild-type (5 ± 1 pmol/min per mg of protein versus 22 ± 2 pmol/min per mg, respectively). This difference was not due to altered kinetic properties of the enzyme, because the substrate was saturating in the *in vitro* reactions. Mixing experiments, in which equal amounts of pgsF17 and wild-type extracts were combined and assayed showed that the resultant activity was the arithmetic mean of the individual samples (Fig. 6). This finding suggested that the lower activity in pgsF17 cells was not due to the presence of an inhibitor or the absence of a soluble activator in the wild type. However, transfection of pgsF17 with 2OST restored Epi activity to wild-type levels (15 ± 3 pmol/min per mg). In contrast, introduction of p33-2OST did not (7 ± 1 pmol/min per mg). Thus, the restorative effect of 2OST was likely due to relocation of Epi to the Golgi (Fig. 4D), which may serve to stabilize the enzyme against degradative processes active in the ER (25).

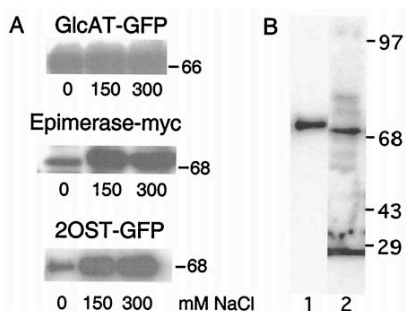


Fig. 5. Physical interaction of Epi and 2OST. (A) CHO cells expressing GlcAT-GFP, Epi-Myc, or 2OST-GFP were extracted in buffer containing 50 mM Mes (pH 6.5), 0.5% Triton X-100, and 0, 150, or 300 mM NaCl as indicated. The cell extracts were centrifuged, separated by SDS/gel electrophoresis, and analyzed by immunoblotting (*Materials and Methods*). (B) Samples of cells containing Epi-Myc and 2OST-GFP were immunoprecipitated with antibody to GFP and analyzed by SDS/PAGE. Lane 1, Western blot using mouse anti-Myc mAb. Lane 2, Western blot using mouse anti-GFP mAb.

Discussion

Nearly all of the enzymes involved in heparan sulfate biosynthesis have now been molecularly cloned and partially characterized, but very little is known about their organization and distribution in the ER-Golgi-TGN, where assembly of the chains takes place. Like glycosyltransferases required for glycolipid and glycoprotein assembly, the biosynthetic enzymes for heparan sulfate are likely to be arrayed across the secretory apparatus in the order in which the reactions occur (8). However, significant overlap exists, dependent in part on the size and number of cisternae, which can vary in different cell types. CHO cells have relatively simple Golgi, and as shown here all five of the enzymes appeared to locate in the compartment containing the α -man-

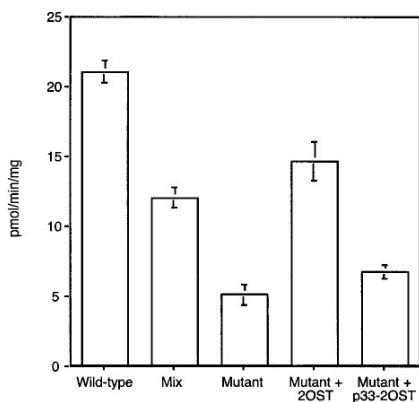


Fig. 6. Decreased epimerase activity in 2OST-deficient cells. Wild-type, pgsF17, and pgsF17 cells stably transfected with 2OST or p33-2OST were assayed for Epi activity. "Mix" refers to a sample containing equal amounts of wild-type and mutant extracts at half the level in samples assayed alone.

nosidase II marker. Prior studies have suggested that NDST1 may be located in the TGN in mouse LTA cells, probably reflecting differences in the arrangement of the Golgi in these cells (26).

Analysis of two early reactions in the pathway catalyzed by GalTI and GlcATI showed that these enzymes colocalize with the medial-Golgi marker, α -mannosidase II, but GalTI also appears in a compartment lacking α -mannosidase II. Although the identity of this compartment is not yet known, it may represent the transitional elements between the ER and Golgi. Xylosyltransferase (XyIT) is thought to be located in the ER and transitional elements possibly related to the ER-Golgi intermediate compartment based on electron microscopic autoradiography, immunocytochemistry, and localization of the decarboxylase that generates UDP-xylose (27-29). In addition, GalTI can physically interact with XyIT *in vitro* (30-32), suggesting that they may reside in the same compartment *in vivo*. Because XyIT precedes galactosyltransferase in the pathway, it is reasonable to assume that the association may occur in a more proximal biosynthetic compartment. The recent purification and cloning of XyIT should help clarify this issue (33, 34).

The formation of the linkage region is completed by the action of galactosyltransferase II (GalTII) and GlcATI. GalTII has been recently cloned and shown to also overlap with α -mannosidase II as well as CALNUP, a marker of the cis Golgi (35). The physical interaction of GalTII with GlcATI or GalTI has not yet been explored. GalTI apparently does not interact with GlcATI by the criteria of relocation in the p33 grafting technique and ease of extraction (Fig. 2). The ability to readily solubilize GlcATI from microsomal membranes suggests that complexes may not exist (Fig. 5), implying that none of these enzymes may be associated with one another.

The data presented here represent *in vivo* evidence for the formation of a complex between sequential enzymes in the pathway of heparan sulfate biosynthesis, Epi and 2OST. This conclusion is based on their mutual relocation when one of the proteins is relocated to the ER, on the mislocalization of a tagged form of Epi in a mutant cell line lacking 2OST, on the requirement for high salt to solubilize both enzymes from microsomal membranes, and coimmunoprecipitation. An important question concerns the functional significance of the Epi-2OST interaction. Substrate channeling through this enzyme complex would ensure a high degree of sulfation after epimerization has occurred, which is borne out by structural studies of heparan sulfate in CHO cells and other cell types (15, 36). A second function concerns the proper placement of the enzymes in the Golgi, either in the same compartment or distal to the site of chain polymerization. Chain polymerization occurs by the alternating addition of GlcNAc and GlcA residues, and these reactions are catalyzed by one or more enzymes (EXT1 and EXT2) having both transferase activities (37-41). It is interesting to note that EXT1 and EXT2 can form heterooligomers apparently required for maximal activity *in vitro* and translocation to the Golgi *in vivo* (42), reminiscent of the behavior of epimerase and 2OST. Whether a relationship exists between these two systems is unknown.

Unlike other pathways of glycosylation, glycosaminoglycan biosynthesis employs at least two enzymes that have more than one activity. The EXT isozymes (EXT1 and EXT2) each have the capacity to transfer GlcA and GlcNAc to a growing chain, and NDSTs have the capacity to remove acetyl groups and substitute the free amino groups with sulfate. Fused proteins with dual or multiple catalytic functions provide an appealing strategy to effect the rapid assembly of glycans on protein and lipid cores as they pass through the Golgi on their way to the plasma membrane or extracellular matrix. The physical interaction of separate enzymes achieves the same goal, but provides an

added degree of control in the system because the abundance of individual subunits can be altered independently.

With the evidence presented here, complexes have now been demonstrated between sequential enzymes involved in the formation of all major glycoconjugates (glycoproteins, glycolipids, and proteoglycans) (13, 43). Apparently enzyme–enzyme interactions can occur in different subcompartments of the Golgi, indicating that “kin-recognition” may be a general mechanism for ensuring that certain protein subsets reside in the appropriate compartments. These complexes may include other components involved in glycosylation, including nucleotide transporters (44) as well as chaperones that participate in quality control. Further

studies are needed to determine if such complexes affect the fidelity of the process, the fine structure of the carbohydrate chains, and thus the formation of binding sites for various carbohydrate-binding proteins.

We thank Jim Feramisco (Imaging Center, Univ. of California San Diego Cancer Center) for many helpful discussions, Joel Shaper (Johns Hopkins School of Medicine) for providing the cDNA clone for GalTI, Haroko Habuchi (Aichi Medical School) for providing the 2-OST cDNA, and G. Wei (Salk Institute) for providing GlcATI cDNA. A grant from The FEW Latin American Program (to M.A.S.P.) and Grant R37GM33063 from the National Institutes of Health (to J.D.E.) supported this work.

- Lindahl, U., Kusche-Gullberg, M. & Kjellén, L. (1998) *J. Biol. Chem.* **273**, 24979–24982.
- Esko, J. D. & Lindahl, U. (2001) *J. Clin. Invest.* **108**, 169–173.
- Slworak, N. W., Liu, J. A., Petros, L. M., Zhang, L. J., Kobayashi, M., Copeland, N. G., Jenkins, N. A. & Rosenberg, R. D. (1999) *J. Biol. Chem.* **274**, 5170–5184.
- Aikawa, J., Grobe, K., Tsujimoto, M. & Esko, J. D. (2001) *J. Biol. Chem.* **276**, 5376–5382.
- Habuchi, H., Tanaka, M., Habuchi, O., Yoshida, K., Suzuki, H., Ban, K. & Kinata, K. (2000) *J. Biol. Chem.* **275**, 2859–2868.
- Slworak, N. W., Liu, J., Fritz, L. M. S., Schwartz, J. J., Zhang, L. J., Logeart, D. & Rosenberg, R. D. (1997) *J. Biol. Chem.* **272**, 28008–28019.
- Fernández, C. J. & Warren, G. (1998) *J. Biol. Chem.* **273**, 19030–19039.
- Varki, A. (1998) *Trends Cell Biol.* **8**, 34–40.
- Colley, K. J. (1997) *Glycobiology* **7**, 1–13.
- Nilsson, T., Rabouille, C., Hui, N., Watson, R. & Warren, G. (1996) *J. Cell Sci.* **109**, 1975–1989.
- Munro, S. (1998) *Trends Cell Biol.* **8**, 11–15.
- Salmivirta, M., Lidholt, K. & Lindahl, U. (1996) *FASEB J.* **10**, 1270–1279.
- Nilsson, T., Hoe, M. H., Susarewicz, P., Rabouille, C., Watson, R., Hunte, F., Watzel, G., Berger, E. G. & Warren, G. (1994) *EMBO J.* **13**, 562–574.
- Nilsson, T., Susarewicz, P., Hoe, M. H. & Warren, G. (1993) *FEBS Lett.* **330**, 1–4.
- Bai, X. M. & Esko, J. D. (1996) *J. Biol. Chem.* **271**, 17711–17717.
- Bai, X. M., Wei, G., Sinha, A. & Esko, J. D. (1999) *J. Biol. Chem.* **274**, 13017–13024.
- Esko, J. D., Weinke, J. L., Taylor, W. H., Ekberg, G., Rodén, L., Anantharamah, G. & Gawish, A. (1987) *J. Biol. Chem.* **262**, 12189–12195.
- Dulbecco, R. & Vogt, M. (1954) *J. Exp. Med.* **99**, 167–182.
- Opat, A. S., Houghton, F. & Gleeson, P. K. (2000) *J. Biol. Chem.* **275**, 11836–11845.
- Crawford, B. E., Olson, S. K., Esko, J. D. & Pinhal, M. A. S. (2001) *J. Biol. Chem.* **276**, 21538–21543.
- Hök, M., Lindahl, U., Hallén, A. & Bäckström, G. (1975) *J. Biol. Chem.* **250**, 6065–6071.
- Velasco, A., Hendricks, L., Moremen, K. W., Tulsiani, D. R., Touster, O. & Farquhar, M. G. (1993) *J. Cell Biol.* **122**, 39–51.
- Lahtinen, U., Dahlhof, B. & Saraste, J. (1992) *J. Cell Sci.* **103**, 321–333.
- Jacobson, I., Lindahl, U., Jensen, J. W., Rodén, L., Prihar, H. & Feingold, D. S. (1984) *J. Biol. Chem.* **259**, 1056–1063.
- Lord, J. M., Davey, J., Frigenio, L. & Roberts, L. M. (2000) *Semin. Cell. Dev. Biol.* **11**, 159–164.
- Humphries, D. E., Sullivan, B. M., Aleixo, M. D. & Stow, J. L. (1997) *Biochem. J.* **325**, 351–357.
- Hoffmann, H. F., Schwartz, N. B., Rodén, L. & Prockop, D. J. (1984) *Connect. Tissue Res.* **12**, 151–163.
- Vertel, B. M., Walters, L. M., Flay, N., Keams, A. E. & Schwartz, N. B. (1993) *J. Biol. Chem.* **268**, 11105–11112.
- Keams, A. E., Vertel, B. M. & Schwartz, N. B. (1993) *J. Biol. Chem.* **268**, 11097–11104.
- Schwartz, N. B. & Rodén, L. (1974) *Carbohydr. Res.* **37**, 167–180.
- Schwartz, N. B., Rodén, L. & Dorfman, A. (1974) *Biochem. Biophys. Res. Commun.* **56**, 717–724.
- Schwartz, N. B. (1975) *FEBS Lett.* **49**, 342–345.
- Götting, C., Kuhn, J., Zahn, R., Brinkmann, T. & Kleesiek, K. (2000) *J. Mol. Biol.* **304**, 517–528.
- Kuhn, J., Götting, C., Schnolzer, M., Kempf, T., Brinkmann, T. & Kleesiek, K. (2000) *J. Biol. Chem.* **275**, 4940–4947.
- Bai, X., Zhou, D., Brown, J. R., Hennes, T. & Esko, J. D. (2001) *J. Biol. Chem.* **276**, in press.
- Bame, K. J., Lidholt, K., Lindahl, U. & Esko, J. D. (1991) *J. Biol. Chem.* **266**, 10287–10293.
- Lidholt, K., Weinke, J. L., Kiser, C. S., Lugenwa, F. N., Bame, K. J., Cheifetz, S., Massagué, J., Lindahl, U. & Esko, J. D. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2267–2271.
- Lind, T., Tufaro, F., McCormick, C., Lindahl, U. & Lidholt, K. (1998) *J. Biol. Chem.* **273**, 26265–26268.
- McCormick, C., Leduc, Y., Martindale, D., Mattison, K., Estford, L. E., Dyer, A. P. & Tufaro, F. (1998) *Nat. Genet.* **19**, 158–161.
- Wei, G., Bai, X. M., Gabb, M. M. G., Bame, K. J., Koshy, T. I., Spear, P. G. & Esko, J. D. (2000) *J. Biol. Chem.* **275**, 27733–27740.
- Senay, C., Lind, T., Muguruma, K., Tone, Y., Kitagawa, H., Sugahara, K., Lidholt, K., Lindahl, U. & Kusche-Gullberg, M. (2000) *EMBO Rep.* **1**, 282–286.
- McCormick, C., Duncan, G., Goutsos, K. T. & Tufaro, F. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 668–673.
- Giraud, C. G., Danotti, J. L. & Maccioni, H. J. F. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 1625–1630. (First Published January 23, 2001; 10.1073/pnas.031458398)
- Hirschberg, C. B., Robbins, P. W. & Abejon, C. (1998) *Annu. Rev. Biochem.* **67**, 49–69.

Appendix B is a reprint of the material as it appeared in the Proceedings of the National Academy of Sciences USA, Pinhal, M.A., Smith, B., Olson, S., Aikawa, J., Kimata, K. and Esko, J.D. (2001). The dissertation author was a secondary researcher and author and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.