INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI
Characterization of the Role of Domains and Subdomains of Aggrecan in Post-Translational Processing and Secretion

by

Gholam Chris Kiani

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Graduate Department of Laboratory Medicine and Pathobiology

University of Toronto

© Copyright by Gholam Chris Kiani 2002
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-69055-5
Characterization of the role of domains and subdomains of aggrecan in post-translational modifications and secretion

A thesis submitted in conformity with the requirements for the degree of Doctor of philosophy
Graduate Department of Laboratory Medicine and Pathobiology
University of Toronto, 2001
Gholam Chris Kiani

ABSTRACT

Aggrecan is the major proteoglycan in the articular cartilage. This molecule is crucial to the chondroskeletal morphogenesis during development and it is also important in the proper functioning of articular cartilage by providing hydrated gel structures (via its interaction with hyaluronic acid and link protein) which endow the cartilage with load-bearing properties. Aggrecan is a multimodular molecule expressed by chondrocytes. To determine the role of each aggrecan domain in its post-translational modification and secretion, we generated a number of recombinant DNA constructs bearing various aggrecan domains individually or in combinations with other domains. We observed that G1 and G2 domains inhibited product secretion and GAG chain attachment whereas, G3 domain promoted secretion and GAG chain attachment. Knowing that G1 and G2 inhibited secretion, we set to identify which region of G1 and G2 shared the inhibition of secretion properties. Utilizing recombinant genes, we demonstrated that the first tandem repeats of G1 and G2 inhibited product secretion if expressed alone or in combinations with other domains and subdomains. We next tried to isolate the amino acid residues in the first tandem repeat of G1 involved in the inhibition of secretion by mutagenesis experiments. C-terminal mutagenesis of the first tandem repeat of G1 did not promote the secretion, however, N-terminal mutagenesis (point mutations and deletions) gave rise to enhancement of secretion. We noticed that there were four N-terminal residues, FHYR, which seemed to be important in inhibition of secretion, however, we next realized that these four residues were part of a
stretch of 55 amino acids required for the inhibitory property. Yeast two-hybrid analyses also indicated that TR1 of G1 is interacting with ER-bound chaperones.
I would like to thank Dr. Burton Yang for giving me the opportunity to work in his laboratory as well as guiding and advising me throughout the course of this research. I consider myself extremely lucky to have had such a knowledgeable, kind and understanding supervisor during my Ph.D. tenure. I would like to express my appreciation to the members of my supervisory committee, especially the late Dr. M.E. Adams (may God bless him), Dr. J. Filmus, Dr. M. McGavin and Dr. R. Inman for their moral and scientific support as well as assistance upon preparing the thesis.

I would also like to extend my appreciation to the members of my lab, Dr. L. Cao, Dr. Y. Zhang, D. Wu, L. Chen and V. Lee for all the help and moral support that they have given me, it will not be forgotten.

Finally, I would like to take the opportunity to express my outmost appreciation to my family, Nathalie (my wife) and Heleni (my daughter) to out up with all the ordeals (inherent in doing graduate work) that I had to put them through in doing my Ph.D.
# Table of Contents

## CHAPTER 1: GENERAL INTRODUCTION .......................................................... 1

1.1 Overview .................................................................................................. 2

1.2 Structures and Functions of Proteoglycans .......................................... 6
  1.2.1 Proteoglycans and Their Entrapment in a Collagen Fibril Network .... 6
  1.2.2 Large Aggregating Proteoglycans .................................................. 9
    1.2.2.1 Aggrecan .................................................................................. 10
    1.2.2.2 Aggrecan Structure ................................................................. 11
    1.2.2.3 G1 Domain .............................................................................. 14
    1.2.2.4 IGD ......................................................................................... 16
    1.2.2.5 G2 Domain .............................................................................. 20
    1.2.2.6 KS Domain ............................................................................ 22
    1.2.2.7 CS Domain ............................................................................ 23
    1.2.2.8 G3 Domain ............................................................................ 24
    1.2.2.9 Post-Translational Modifications of Cartilage Aggrecan ......... 26
      1.2.2.9a GAG Chain Attachments .................................................. 26
      1.2.2.9b GAG Chain Synthesis and Molecular Structures and their comparison to Molecular Structures and Synthesis of N- and O-Linked Oligosaccharides 28
      1.2.2.9c Chondroitin Sulfates and Dermatan Sulfates .................... 29
      1.2.2.9d Keratan Sulfates ................................................................. 34
      1.2.2.9e GAG Synthesis .................................................................... 34
      1.2.2.9f Sulfations ........................................................................... 38
      1.2.2.10 Secretory Pathway of Aggrecan ......................................... 39
    1.2.2.11 Large Aggregating Chondroitin Sulfate proteoglycans, PG-M/Versican 40

1.3 Intracellular trafficking of Secretory Proteins ....................................... 42
  1.3.1 Targeting of Nascent Polypeptide Chains to the RER .................... 45
  1.3.2 Translocation of Polypeptide Chains to the RER ............................ 46
  1.3.3 Protein folding in the RER ............................................................ 48
  1.3.4 Consequences of the presence of signal sequences ....................... 50
CHAPTER 2: CHARACTERIZATION OF THE ROLE OF AGGREGAN DOMAINS IN POST-TRANSLATIONAL MODIFICATIONS AND SECRETION

2.1 Abstract ................................................................. 71
2.2 Introduction ............................................................ 72
2.3 Experimental Procedures ........................................... 74
   2.3.1 Materials ............................................................ 74
   2.3.2 Strategies for Cloning and Construction of Aggrecan Recombinant Genes ............... 74
   2.3.3 DNA Amplification, Purification, Ligation and Transformation .......................... 78
   2.3.4 Expression of Recombinant Constructs in COS-7 Cells ................................ 79
   2.3.5 Western Blot Assays ............................................. 80
   2.3.6 Chondroitinase ABC Treatment .................................. 80
   2.3.7 TCA Precipitations of Proteins .............................. 81
2.4 Results .................................................................. 83
   2.4.1 Expression of Recombinant Constructs ......................................................... 83
   2.4.2 G3 enhanced GAG modification of KS and CS and facilitated product secretion ........................................... 83
   2.4.3 G1's reduction of GAG modification, and inhibition of product secretion was partially abolished by CS and enhanced by G3 ........................................... 84
CHAPTER 3: THE FIRST TANDEM REPEAT IN AGGREGAN G1 DOMAIN CAN INHIBIT PRODUCT PROCESSING AND INTERACT WITH CALCIUM HOMEOSTASIS ENDOPLASMIC RETICULUM PROTEIN

3.1 Abstract ............................................................................................... 103
3.2 Introduction ............................................................................................ 104
3.3 Experimental Procedures ......................................................................... 106
  3.3.1 Materials .......................................................................................... 106
  3.3.2 Strategies for Gene Construction ............................................................... 106
    3.3.2.1 Cloning of TR G1, TR1 G1, TR2 G1 and IgG in pcDNA1 ................. 106
    3.3.2.2 Cloning of TR1 G2, TR2 G2 (or G2), TR1 G2 and TR2 G2 in pCR3.1
        (TA Cloning System) ................................................................................ 107
    3.3.2.3 Targeted Cloning of Different Aggrecan G1 Subdomains
        with Each Other and with the G3 Domain .............................................. 108
    3.3.2.4 Linkage of G2 Motifs with IgG Motif or the G3 Domain ................. 109
    3.3.2.5 Site-Directed Mutagenesis and Truncation ...................................... 110
    3.3.2.6 Addition of N-Terminal Fragments of TR1 G1 to IgG Motif ............ 110
    3.3.2.7 Generation of Yeast Two-Hybrid Binding Constructs .................... 110
    3.3.2.8 Expression of Recombinant Constructs in COS-7 Cells .................. 111
    3.3.2.9 Western Blot Assays ....................................................................... 111
    3.3.2.10 Immunoprecipitation .................................................................... 112
    3.3.2.11 Chondroitinase ABC Treatment ..................................................... 112
    3.3.2.12 Yeast Two-Hybrid Assay ............................................................... 112
3.4 Results ................................................................................................. 114
  3.4.1 Identification of the Motif in Aggrecan G1 Domain That Is Incapable
      of Secretion ............................................................................................. 114
  3.4.2 The G1 Domain Enhances GAG Addition ............................................. 116
3.4.3 TR1<sup>G1</sup> Binds to ER Protein and Affects Processing of Mini-aggrecan..................117

3.4.4 The Tandem Repeats of the G2 Domain Function

Similarly to G1 in Product Secretion.................................................................118

3.4.5 Identification of the Amino Acids in the TR1 Motif That

Account for the Incapability of Product Secretion.............................................119

3.5 Discussion...............................................................................................132

General Conclusion.........................................................................................139

References......................................................................................................142
List of Figures

Figure 1-1 .................................................................................................. 5
Illustration of Cartilage Matrix Composition

Figure 2-1 .................................................................................................. 8
The Combined Properties of Collagens and Proteoglycans in Articular Cartilage

Figure 3-1 ................................................................................................. 12
Electron Micrographs of Proteoglycan Aggregates

Figure 4-1 ................................................................................................. 15
Aggrecan Structure

Figure 5-1 ................................................................................................ 17
Sites of Cleavage of Aggrecan Core Protein by Tissue Proteinases

Figure 6-1 ................................................................................................. 31
A Schematic Diagram of the GAG Chain Tetrasaccharide Linkage Regions,
the Repeating Disaccharide Units and their Sulfation Patterns

Figure 7-1 ................................................................................................. 32
A schematic Structure of the N-Linked Core Oligosaccharide Added to Polypeptides
in the ER

Figure 8-1 ................................................................................................. 43
A Schematic Representation of the Secretory Pathway

Figure 9-1 ................................................................................................. 59
The Consecutive and Regulated Secretory Pathways

Figure 10-1 ............................................................................................. 62
Post-Translational Modifications of Proteins in the Golgi

Figure 1-2 ................................................................................................ 89
The structures of recombinant DNA constructs containing different
domains of aggrecan

Figure 2-2 ............................................................................................. 90
G3 enhances KS synthesis and secretion

Figure 3-2 ............................................................................................. 91
G3 enhances CS synthesis and secretion
Figure 4-2................................................................................................. 92
G1 inhibits G3 secretion

Figure 5-2................................................................................................. 93
G1 inhibits KS secretion

Figure 6-2................................................................................................. 94
G1 inhibition of secretion is overcome by CS and enhanced by G3

Figure 7-2................................................................................................. 95
The effects of G2 domain on GAG chain attachment and product secretion

Figure 1-3................................................................................................. 122
IgG, but not tandem repeats, is secreted to culture medium

Figure 2-3................................................................................................. 123
Amino acid homology of TR1 and TR2 motifs

Figure 3-3................................................................................................. 124
Products containing TR1G1 are not secreted

Figure 4-3................................................................................................. 125
The G1 domain enhances GAG modificatio

Figure 5-3................................................................................................. 126
Interaction of TR1G1 with ER protein

Figure 6-3................................................................................................. 127
Secretion of the mini-aggreican is inhibited by TR1

Figure 7-3................................................................................................. 128
G2 has similar function in product secretion

Figure 8-3................................................................................................. 129
Deletion of the N-terminal 9 amino acids allows product secretion

Figure 9-3................................................................................................. 130
Mutation of the amino acids H2R4 abolishes TR1’s effect on product secretion

Figure 10-3............................................................................................... 131
IgG-like motif linked to a fifty-five amino acid TR1 fragment is not secreted
List of Tables

Table 1-2..................................................................................................................88
Primers and Restriction Endonuclease Sites for Polymerase Chain Reactions of aggrecan
constructs (Chapter 2)

Table 1-3..................................................................................................................121
Sequences and restriction endonuclease sites of oligonucleotides (Chapter 3)
Abbreviations

A°: Angstrom
A: Alanine
ADAMTS: A Disintegrin and Metalloproteinase with Thromobospondin Motifs
APS: adenosine-phosphosulfate
ATP: Adenosine Triphosphate
bp: Base Pair
CAT: Chloramphenicol Acetyl-Transferase
CBP: Complement Binding Protein-Like Motif, also known as CRP
CHO: Chinese Hamster Ovary
CleNAc: N-acetyl-glucosamine
Cmd: Cartilage Matrix Deficiency
CPY: Carboxypeptidase Y
CRD: Carbohydrate Recognition Domain
CRP: Complement Regulatory Protein, also known as CBP
CS: Chondroitin Sulfate
CSPG: Chondroitin Sulfate Proteoglycans
CMV: Cytomegalovirus
D: Aspartate
DMEM: Dulbecco's Modified Eagle's Medium
dNTP: Dioxynucleotide Triphosphate
DS: Dermatan Sulfate
E: Glutamate
ECM: Extracellular Matrix

EDTA: Ethylene Diamine Tetra-Acetic Acid

EGF: Epidermal Growth Factor

F: Phenylalanine

FBS: Fetal Bovine Serum

FGF-1 and -2: Fibroblast Growth Factors 1 and 2

G: Glycine

G1: Globular Domain 1

G2: Globular Domain 2

G3: Globular Domain 3

GAA: Guanine-Adenine-Adenine

GAG: Glycosaminoglycan

Gal: Galactose

GalNAc: N-acetyl-galactosamine

GlcA: Glucuronic Acid

GlcAT I: Glucuronic Acid Transferase I

GT I: Galactosyl Transferase I

H: Histidine

h: Hour

HA: Hyaluronic Acid or hyaluronan

HS: Heparan Sulfate

HSPG: Heparan Sulfate Proteoglycan

HSP90: Heat Shock Protein 90

I: Isoleucine
IGD: Interglobular Domain

K: Lysine

kDa: kilo Dalton

KS: Keratan Sulfate

mDa: Mega Dalton

Mg: Magnesium

MHC: Major Histocompatibility Complex

MMP: Matrix Metalloproteinase

MRI: Magnetic Resonance Imaging

MTOC: Microtubule Organizing Center

N: Asparagine

NAC: Nascent Polypeptide associated Complex

NANA: N-acetyl Neuraminic Acid

OA: Osteoarthritis

P: Proline

PAPS: 3' Phosphoadenosine 5' Phosphosulfate

PBS: Phosphate Buffered Saline

PCR: Polymerase Chain Reaction

PDI: Protein Disulfide Isomerase

PG: Proteoglycan

PPα F: prepro-α-factor

PTR: Proteoglycan Tandem Repeats

Q: Glutamine
R: Arginine

RER: Rough Endoplasmic Reticulum

RPM: Rotation Per Minute

SDS-PAGE: Sodium Dodecyle Sulfate Polyacrylamide Gel Electrophoresis

SED: Spodylo-Epiphysael Dysplasia

SER: Smooth Endoplasmic Reticulum

SR: Signal Recognition Particle Receptor

SRP: Signal Recognition Particle

T: Threonine

TAA: Thymidine-Adenine-Adenine

TCA: Trichloroacetic Acid

TGN: Trans Golgi Network

TSG: Tumor associated Secreted Gene-6

UDP: Uridine Diphosphate

V: valine

XT: Xylosyl Transferase

Xyl: Xylose

Y: Tyrosine
CHAPTER I

GENERAL INTRODUCTION
1.1 Overview

Joint diseases are very prevalent in North America. One to five percent of the population under the age of 45 and 15-85% of older individuals suffer from some form of degenerative joint diseases, mainly osteoarthritis. Osteoarthritis is characterized by the progressive deterioration of articular cartilage over many years (Hardingham and Fosang, 1995; Mansson et al., 1995). In articular cartilage, synovial joints are covered with cartilage, a smooth, stiff and wear resistant material that provides a low friction surface that functions to facilitate movement and to distribute loads associated with body weight and muscular activity. The destruction of cartilage that occurs in various joint diseases leads to joint dysfunction and pain.

Based on the composition of the extracellular matrix, cartilagenous tissues are categorized (ECM) into four kinds: fibrocartilage, elastic (yellow cartilage), epiphyseal cartilage and hyaline cartilage. Fibrocartilage is a strong, tension-resistant, but flexible, transitional tissue without any perichondrium. Fibrocartilage is found in the intervertebral disk's annulus fibrosus, pubic symphysis, femoral ligamentum teres and meniscus. Elastic cartilage is found in external ear, pharngotympanic tube, epiglottis and larynx. The matrix in this kind of cartilage is also prone to calcification or degeneration and is also permeated by many elastic fibers. Epiphyseal cartilage is found at the end of long bone and is mainly responsible for longitudinal bone growth. Hyaline cartilage is found in the articular surfaces of all of the synovial joints as well as in costal, nasal and respiratory tract cartilages. This cartilage occurs fused with bone or as a discrete entity. This tissue looks hyaline or translucent to the unaided eye. Hyaline cartilage is covered with a perichondrium in most cases (except joints) where chondrocytes are large and rounded, each lying in a space
enclosed by the matrix. Articular cartilage is the most widely studied hyaline cartilage. This type of cartilage is an aneural, avascular (nourishment of this tissue is made possible by the process of diffusion), alymphatic and relatively acellular hyaline cartilage that covers the articulating surface of all diarthrodial joints. Cartilage is also unique in having only one cell type, chondrocytes. Because articular cartilage does not possess any vascular network, the chondrocytes in this tissue have a very limited repair capability. Chondrocytes make up 1-12% of the tissue volume in cartilage; the rest consists of water (up to 75%) and ECM molecules synthesized and secreted by these cells. Articular cartilage possesses ECM molecules such as collagen type II, IX, X and XI (endowing the articular cartilage with its tensile strength) as well as proteoglycans such as aggrecan, biglycan, decorin and fibromodulin (Watanabe et al., 1998). Among the cartilage proteoglycans, aggrecan is most crucial to the proper functioning of articular cartilage. Aggrecan is a multimodular proteoglycan secreted by the chondrocytes to the ECM in hyaline (articular) cartilage. Aggrecan interacts with hyaluronic acid and link protein through its G1 domain (Lee et al., 1992; Yu and Toole, 1995) and in so doing, it forms huge aggregates (hence the name aggrecan) in cartilage which provide the template for endochondral ossification and are crucial for determining the length and width of the skeleton (Yu and Toole, 1995). The CS domain of aggrecan is capable of binding up to 130 glycosaminoglycan (GAG) chains (Calabro and Hascall, 1994). GAGs have the ability to imbibe huge amounts of water molecules. The combination of huge aggregate formation and entrapment of water by GAGs gives rise to the hydrated gel structure of the articular cartilage, which is able to withstand compressive forces and repeated loading. It can also be appreciated that the ternary complexes formed by the co-operative association of aggrecan,
link protein and hyaluronic acid are constrained in a rather strong and stiff network of collagen molecules and their associated proteins (Fig. 1-1).

Human articular cartilage tends to deteriorate with advancing age. This deterioration is partly due to the chondrocytes’ limited capacity to keep up with the demand of tissue for new ECM molecules. There are variabilities in glycosylation, sulfation and other post-translational modifications of glycoproteins made by chondrocytes with age. These changes in cartilage may adversely affect its structure and function. Cartilage function is also likely to be affected by the ECM milieu and the way these molecules interact with one another which can also have a great impact on the overall functioning of cartilage. In order to formulate innovative treatment modalities for joint diseases, we have to comprehend firstly, what molecules are present in articular cartilage; secondly how these molecules are regulated, synthesized and secreted; and lastly, how these molecules interact with each other as well as with chondrocytes to bring about a smooth and frictionless functioning of the articular cartilage.
Figure 1-1. Illustration of cartilage matrix composition

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

1.2.1 Proteoglycans and their Entrapment in a Collagen Fibril Network

Proteoglycans belong to a special class of glycoproteins with attached long unbranched and highly charged GAG chains (Hardingham and Fosang, 1992). These chains are strongly hydrophilic and they dominate the physical properties of proteoglycans. Cartilage and bone are tissues that contain large expanded ECMs loaded with various proteoglycan components, although they are of quite different structures and have different biological functions.

Cartilage contains 70-75% water and it has a high proteoglycan content (up to 10%), predominantly containing chondroitin sulfate, but also containing a considerable amount of keratan sulfate. Proteoglycans make a major contribution to the biomechanical properties of cartilage. This is particularly important at the articular surfaces, where the tissue has a load bearing function. It is the structure and organization of the macromolecules in cartilage ECM that provide these properties (Hardingham and Fosang, 1992) and they result from the composite structure of cartilage and the contribution made by fibrillar and nonfibrillar components. The collagen fibrillar network is made up of fine fibers which have no preferred orientation in the mid zone of cartilage. At the articular surface, the fibers are parallel to the surface in a single orientation, and they are more perpendicular to the surface in deeper zones. The collagen provides a framework for the tissue that gives it shape and form. The triple helical collagen molecules are organized into fibrils with overlapping and cross-linking adjacent molecules, and the fibrils laterally associate with longer fibrils. The structure of collagen gives it impressive tensile properties and this is utilized in cartilage in a special way to produce a tissue that is not only strong in tension, but also resistant to compression. This is
achieved by filling the interfibrilar matrix with a very high content of proteoglycans, primarily aggrecan. At high concentrations, aggrecan creates a large osmotic swelling pressure and draws water into the tissue (Fig. 2-1). This occurs because all of the negatively charged anionic groups on aggrecan carry with them mobile counter ions such as Na\(^+\). This creates a large difference in the concentration of ions inside cartilage compared with outside and an imbalance amongst the freely diffusible anions and cations. Water is drawn into the tissue because of this osmotic imbalance and because aggrecan is too large and immobile to redistribute itself. The water thus swells and expands the aggrecan-rich matrix. Another feature of the composite collagen/aggrecan organization is also important.
Figure 2-1. The combined properties of collagens and proteoglycans in articular cartilage
Not only is aggrecan greatly restricted in its ability to move within the matrix, and the collagen/aggrecan network stiff and resistant to deformation, but aggrecan also offers great resistance to any fluid flow and redistribution of water. Thus, cartilage behaves like a stiff elastic polymer resistant to sudden impact loading, and it also shows some slow inelastic deformation with sustained loads (Mow et al., 1989).

It has been also reported that aggrecan is located in some highly differentiated systems such as brain and notochord. In chick cartilage, aggrecan expression begins at the embryonic day 5 in limb rudiments, continues through the entire period of chondrocyte development, and remains a biochemical marker of the cartilage phenotype thereafter. In brain, aggrecan has a very low level of expression beginning at day 7, increases up to day 13, markedly decreases after day 16, and it is not expressed post hatching.

1.2.2 Large Aggregating Proteoglycans

Large aggregating chondroitin sulfate proteoglycan (CSPG/aggrecan) is one of the major ECM components in cartilage. It has been shown that one molecule of hyaluronic acid (HA) can interact with numerous aggrecan molecules. Proteoglycans of the aggrecan family include other members such as human versican (Zimmermann and Ruoslahti, 1989) and its avian homologue PG-M (Shinomura et al., 1993), neurocan, brevican (Yamada et al., 1994) and the cell surface hyaluronan receptor CD44 (Goldstein et al., 1989). They are modular proteoglycans containing combinations of structural domains, such as epidermal growth factor-like domains, lectin-like domains, complement regulatory protein-like domains, immunoglobulin folds and proteoglycan tandem repeats (Perkins et al., 1989). Several other proteins are related to this family of molecules, including link protein (Deak et al., 1999; Neame and Barry, 1994) and TSG-6 (Lee et al., 1992). They contain some of the same highly
conserved structural motifs, but lack GAG chains. Among the aggrecan family members, neurocan (expressed in nerve tissues), brevican (expressed in nerve tissues), CD44 (present on all cells), TSG and link protein are able to bind to HA, but unable to form large aggregates on their own.

1.2.2.1 Aggrecan

Aggrecan is expressed in cartilage where it is the most abundantly expressed proteoglycan, and there is evidence for aggrecan expression in tendons (Vogel et al., 1994). Aggrecan is found as huge multimolecular aggregates comprised of numerous monomers non-covalently bound to HA (Fig. 3-1). A small glycoprotein, link protein, which is homologous to the N-terminus of aggrecan, helps stabilize aggregate formation (Fig. 1-1). Almost 90% of aggrecan mass is comprised of substituted GAG chains which are mostly chondroitin sulfate chains, but also keratan sulfate chains with N- and O-linked oligosaccharides. The complete cDNA sequences of human (Doege et al., 1991), rat (Doege et al., 1987), mouse (Walcz et al., 1994), chicken (Chandrasekaran and Tanzer, 1992; Li et al., 1993) (Schwartz et al., 1999) and bovine (Hering et al., 1997) aggrecan are known and partial sequences for pig (Barry et al., 1992a) have also been reported. The human aggrecan gene has been mapped to chromosome band 15q26 (Just et al., 1993; Korenberg et al., 1993). Aggrecan core protein is encoded by a single gene that is mapped to chromosome 7 in mouse (Watanabe et al., 1994) and chromosome 15 in humans (Just et al., 1993; Korenberg et al., 1993). Aggrecan core protein is also encoded by a single gene in chicken that is mapped to macrochromosome 3 (Jones et al., 1997). The genomic structures of rat, chicken, mouse and human aggrecans have been reported, showing a high degree of similarity among species.
Aggrecan has three globular domains (G1, G2 and G3) and three extended domains (IGD, KS and CS). Aggrecan is highly glycosylated where one molecule may contain up to 100 chondroitin sulfate chains. These are typically ~20 kDa each, and are either 4-sulfated, 6-sulfated or usually both. There are fewer keratan sulfate chains (up to 60) and they are usually of shorter length (5-15 kDa). The chondroitin sulfate chains are all attached to the long extended domain between globular domains 2 and 3, but the keratan sulfate is more widely distributed. The latter is most abundant in a keratan sulfate-rich region just C-terminal of the G2 domain. Keratan sulfate is attached elsewhere on the globular domains G1 and G2 as well as the non-globular IGD region. Aggrecan also contains a variable number of O-linked and N-linked oligosaccharides. The O-linked oligosaccharides have a linkage to protein similar to keratan sulfate and it appears that, during biosynthesis, some O-linked oligosaccharides are extended and sulfated to form keratan sulfate chains. Variations in the proportions subject to extension and the length of the chains synthesized are thus likely to account for the large differences in the content of keratan sulfate among aggrecan homologs from different species.
Figure 3-1. Electron micrographs of proteoglycan aggregates

(A) Large aggregate: the aggregate consists of multiple closely spaced monomers bound to a central hyaluronic acid filament. Free monomers and small clusters of monomers surround the aggregate. (B) Small aggregate: monomers consist of a thin segment (G1-IGD-G2) and a thick segment (KS-CS-G3). The thin segment attaches directly to the central hyaluronic acid filament, while the thick segment extends peripherally (adapted from Buckwalter, et al., 1985).
All three of the globular domains of the aggrecan protein contain sequences that are highly conserved amongst aggrecans from different species, but the extended domains are less conserved. The mouse aggrecan gene spans at least 61 kb and contains 18 exons. Exon 1 encodes 5'-untranslated sequence and exon 2 contains a translation start codon, methionine. The coding sequence is 6545 bp for a 2132-amino-acid protein with calculated molecular weight of 259,131 Da., including an 18-amino-acid signal peptide. There is a strong correlation between structural domains and exons. Notably, the chondroitin sulfate domain consisting of 1161 amino acids is encoded by a single exon of 3.6 kb. Although link protein has similar structural domains and subdomains, the sequence identity and the organization of exons encoding the subdomains B and B' of G1 and G2 domains in mouse aggrecan revealed a strong similarity of mouse aggrecan to both human versican and rat neurocan. Primer extension analysis identified four transcription start sites, which are close together. There are stretches of sequences similar to the promoter region of both the type-II collagen and link protein genes. These sequences may be important for cartilage gene expression (Watanabe et al., 1995). The human aggrecan cDNA is 7137 nucleotides in size, encoding 2316 amino acids. The human and rat aggrecan amino acid sequences are about 75% identical. The human sequence contains two regions of highly conserved repeats not found in rat aggrecan: 11 repeats of a hexameric sequence in the keratan sulfate attachment domain, E-E-P-(S,F)-P-S; and a 19-amino acid sequence reiterated 19 times, in the CS-1 portion of the serine-glycine-containing region. There are at least three forms of aggrecan transcripts, generated by alternative exon usage, and the form investigated in this study is the shortest and also the most prevalent, lacking both the epidermal growth factor-like motifs (Doege et al., 1991).
1.2.2.3 G1 domain

G1 domain is encoded by exons 3, 4 and 5. This domain is comprised of three modules: an immunoglobulin fold, and two copies of a hyaluronan-binding motif, or link module, also referred to as the proteoglycan tandem repeats (PTR) (Fig. 4-1). This module is present in tandem in all members of the hyaluronan-binding family of proteoglycans such as versican, neurocan and brevican, but it is also present as a single copy in the cell surface hyaluronan-binding receptor, CD44, and in TSG (tumor associated gene 6), a secreted matrix protein whose synthesis is induced by inflammatory cytokines. The immunoglobulin motif of aggrecan G1 corresponding to the hypervariable region of the immunoglobulins is about 100 amino acids long. This region is predicted to fold into two β sheets in a sandwich conformation (parallel orientation) stabilized by a conserved disulfide bond (Fig. 4-1). It is encoded in a single exon. Almost all of the proteins which contain immunoglobulin-related structures are cell surface molecules involved in cell recognition, cell adhesion or immune function. Their role in cell recognition is emphasized by the fact that many of these molecules interact with other members of the immunoglobulin superfamily, and indeed the interaction between aggrecan G1 and link protein is mediated through the immunoglobulin fold (Grover and Roughley, 1994). Tandem repeats (PTR) of G1 domain are comprised of two cysteine-rich motifs which are able to form disulfide bonds (two disulfide bonds per tandem) and are involved in aggrecan's interaction with hyaluronic acid which is essential for the aggregate formation (Fig. 4-1). The ternary structure of the PTRs formed by the disulfide bonds are of considerable importance since these tandem repeats do not bind to HA under reducing conditions (Watanabe et al., 1998).
Figure 4-1. Aggrecan structure
(A) Globular protein and attached glycosaminoglycan chain structure. (B) Protein domain structure. Folded modules: IgG, immunoglobulin fold; TR, tandem repeats; EGF, epidermal growth factor module; CRD, carbohydrate recognition domain; CRP, complement regulatory-protein like module. Extended domains: IGD, interglobular domain; KS, keratan sulfate attachment domain; CS-1 and CS-2, chondroitin sulfate attachment domains
1.2.2.4 IGD

The short extended region separating the G1 and G2 domains of aggrecan is known as the interglobular domain (IGD) and has a rod shape structure which contains proteolytic cleavage sites susceptible to a variety of proteinases such as matrix metalloproteinases (MMPs), serine proteinases such as plasmin and leukocyte elastase, and acid proteinases such as cathepsin B (cysteine protease) (Fig. 5-1)(Hardingham and Fosang, 1995; Mort and Buttle, 1997; Mort et al., 1998). This domain is encoded by exon 6 of aggrecan. IGD domain is unique to the aggrecan molecule since other members of the aggrecan family lack this region.

X-ray diffraction, MRI (magnetic resonance imaging) and rotary shadowing electron microscopic analyses have determined that the IGD is of constant length (25nm) and relatively stiff and inflexible (Dennis et al., 1990; Paulsson et al., 1987). This apparent stiffness of the IGD may be due to the density of keratan sulfate substitutions (Barry et al., 1992b; Fosang et al., 1992) which seems to be greater in the IGD region than the adjoining globular domain (G2). The functional, rather than structural, properties of IGD make it rather interesting. The IGD domain is the site for proteolytic attack on aggrecan during pathological cartilage degradation since cleavage of aggrecan molecules near its G1 domain results in rapid loss of the whole GAG-attachment region. IGD is also presumed to be involved in the physiological turnover of aggrecan. The IGD region seems to be a key site for proteolytic cleavage of aggrecan because it separates the major GAG-bearing portion of aggrecan from the G1 domain that anchors it in the matrix.
Figure 5-1. Sites of cleavage of aggrecan core protein by tissue proteinases.

Two major sites of cleavage are located in the short extended interglobular domain between the globular domains G1 and G2. The first one is generated by metalloproteinases (MMPs) and occurs between residues Asn341 and Phe342, whereas the second one generated by the so-termed aggrecanase is located between residues Glu373 and Ala374. Both cleavage sites separate the anchoring G1 domain from the bulk of the aggrecan molecule carrying GAG chains and thus deprive cartilage of its load-bearing properties. The N-terminal and C-terminal sequences of the neoepitopes resulting from the different cleavage sites as well as the nature of the proteinase acting at these sites are also indicated. The current size of different domains and the number of GAG chains are not respected.
The observations that fragments harboring a G1 domain bearing a VDIPEN-341 neoepitope (rather than a NITEGE-373 neoepitope) are present in human synovial fluids (Saxne and Heinegard, 1992) and appear rapidly in rabbit knee joint fluid after an intra-articular injection of MMP3 (Morales and Hascall, 1988) suggest that there is further processing of the HA-G1 complex after the large interglobular domain has been cleaved off by aggrecanase and has diffused out of the matrix. Recent work on the extent of the involvement of aggrecanase and MMPs in the cleavage of aggrecan at the IGD domain indicates that aggrecanase cleavage of the aggrecan by far exceeds that of MMPs under pathophysiological conditions (Little et al., 1999).

Two major cleavage sites have been identified in the IGD domain: one occurs between residues Asn 341 and Phe 342 and the other is located between residues Glu 373 and Ala 374. The first major cleavage site generates a G1-containing fragment with the C-terminal neoepitope VDIPEN-341 and a larger GAG-bearing fragment with the N-terminal neoepitope 342-FFGVGGE (Fosang et al., 1996a; Fosang et al., 1996b; Fosang et al., 1992; Lark et al., 1997; Singer et al., 1995). It is the major cleavage site of several metalloproteinases including stromelysin-1 (MMP-3), three collagenases (MMP-1, MMP-8 and MMP-13), and two gelatinases (MMP-2 and MMP-9). The results of many studies support the contention that MMPs are implicated in the breakdown of aggrecan molecules in vivo. First, cartilage and synovial fluid from patients with osteoarthritis (OA) and other inflammatory arthritis contain enhanced levels of MMPs. Second, retinoic acid and pro-inflammatory cytokines (IL-1 and TNF-α) known to upregulate the expression of several MMPs promote aggrecan loss from cartilage explants (Arner, 1994; Cawston, 1993;
Lewthwaite et al., 1994). Furthermore, this loss can be prevented by adding specific inhibitors of MMPs to the conditioned medium (Buttle et al., 1993; Seed et al., 1993). Third, N- and C-terminal neoepitopes generated by MMPs are present in both cartilage and synovial fluid. Finally, the demonstration that the concentration of the VDIPEN-341 neoepitope in normal human articular cartilage increases up to 25 years of age and then reaches an apparent steady state (representing 15-20% of the G1-containing molecules residing within the matrix) (Larsson et al., 1997) suggests that MMP-generated G1 fragments might account, at least in part, for the increase in cartilage content of aggrecan domains with growth and maturation.

Studies on the aggrecan fragments extracted from joint cartilage or synovial fluid suggested the presence of a proteinase that cleaves a specific site in IGD, which was named 'aggrecanase' (Sandy et al., 1992). Aggrecanase-1 is a member of the ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) protein family that cleaves aggrecan at the glutamic acid-373-alanine-374 bond (Tortorella et al., 1999). This cleavage produces a large GAG-rich aggrecan fragment with the N-terminal neoepitope 374-ARGSVI and a G1 fragment with the C-terminal neoepitope NITEGE-373 (Fig. 5-1). Aggrecanase-mediated cleavage occurs both in vivo and in vitro: the N-terminal neoepitope 374-ARGSVI has been found in the medium of cartilage explant cultures (Lark et al., 1995b) as well as in high density aggrecan fragments recovered from human synovial fluids (Hughes et al., 1995; Lohmander et al., 1993b; Sandy et al., 1992). The contention that MMPs and aggrecanase are both involved in the turnover of aggrecan molecules in normal and diseased cartilage is supported by the finding that the MMP-generated G1 fragment terminating in VDIPEN-341 and aggrecanase-generated G1 fragments terminating in NITEGE-373 are both detected not only in cartilage from joints with OA and RA (rheumatoid arthritis), two conditions
exhibiting quite contrasting pathological and clinical features, but also in cartilage from normal adult joints (Lark et al., 1997). The generation and/or turnover of these specific aggrecan fragments is not necessarily co-ordinated, since both the NITEGE-373 and VDIPEN-341 neoepitopes can be non-coincident within a single joint (Singer et al., 1995), and turnover of aggrecan by cultured rat chondrosarcoma cells and primary bovine chondrocytes can be mediated exclusively by aggrecanase (Lark et al., 1995a).

It has been reported that aggrecanase also cleaves off aggrecan at multiple sites in the CS-attachment domain (Mansson et al., 1995; Poole, 1994; Poole and Dieppe, 1994) generating GAG-rich fragments which are lost relatively rapidly from the cartilage matrix. These fragment diffuse into the synovial fluid, where they can be quantified by chemical assays of their sulfated GAGs (Ratcliffe et al., 1988) or by immunoassays capable of measuring specific protein or carbohydrate epitopes (Poole, 1994; Poole et al., 1994). Recently, it has been shown that cathepsin B (a cysteine protease with both endopeptidase and carboxypeptidase activities) is able to cleave aggrecan. Cleavage occurs at Asn341-Phe342, a site which corresponds to the cleavage sequence of other MMPs to yield the neoepitopes VDIPEN and FFGVG (Mort et al., 1998).

1.2.2.5 G2 Domain

This domain is encoded by exons 7 and 8 of the aggrecan gene and is the second globular aggrecan domain from the N-terminus. It consists of two tandem repeats (PTR) very similar to those of G1. The G2 domain is unique to aggrecan since other large aggregating chondroitin sulfate proteoglycan family members (such as PG-M/versican) lack this domain. It is noteworthy to add that aggrecan is also the only member of this family which is heavily glycosylated by ~ 100 chondroitin sulfate and up to 60 keratan sulfate chains postulating a
direct or indirect role for G2 domain in GAG chain attachment. The rationale for this postulation being that versican (another member of the large aggregating chondroitin sulfate proteoglycans) which is lacking a G2 domain is not GAG decorated to any extent as that of aggrecan.

The G2 domain shows an approximately 67% amino acid sequence similarity to that of the tandem repeats in the G1 domain. Whereas the PTR has been shown to be the major hyaluronate-binding element in link protein and G1, the G2 domain, either purified from cartilage after proteolytic digestion (Fosang and Hardingham, 1989) or the recombinant soluble G2 protein (Watanabe et al., 1997), shows no HA-binding function whatsoever. This surprising finding might be explained through a closer examination of the PTR loops of G2. Analysis of the structure of PTR loops in G2 indicates that N-linked oligosaccharides are likely to be attached close to the N-terminal cysteine in the exon 8 (coding for the second tandem repeat of G2). These sugar attachments might be responsible for disrupting the proper folding of tandem repeats in G2, abolishing its ability to interact with HA.

G2 has also been shown to contain keratan sulfate chains (Fosang and Hardingham, 1991), and while it is not clear where the KS is attached, it may significantly interfere with the ability of the G2 domain to interact with HA. The locations of cysteine residues essential for G2 ternary structure, and asparagine residues required for attaching carbohydrate in an N-linked fashion, are also well conserved. The G2 domain is also unable to bind to link protein, other aggregating proteoglycan monomers, collagen or other major components of the dissociate extract of cartilage (Fosang and Hardingham, 1989). However, it has been shown recently that the G2 domain inhibits product secretion, hinting that it may have a function in controlling the quality of the finished aggrecan products (Kiani et al., 2001).
1.2.2.6 KS Domain

The keratan sulfate (KS) domain is located at the C-terminus of the G2 domain, and is encoded by exon 11. There are about 30 KS chains on a mature aggrecan molecule. The amino acid sequence of this domain varies among different species. The potential consensus sequence for attachment of KS in human is E-(E,K)P-F-P-S or E-E-P-(S,F)-P-S (Doege et al., 1991). Mice and rats lack these sequences and also lack keratan sulfate chains in their articular aggrecan. This may be due to the absence of these sequences. In both bovine (Antonsson et al., 1989) and human aggrecans (Doege et al., 1991), the amino acid sequence of this region, revealed by cDNA sequence analysis, contains a striking repeating sequence. The sequence is a hexamer which contains a serine residue. It has been suggested that proline-serine and proline-threonine repeats in this region are good candidates for substitution by KS chains. The human cDNA encodes an additional 66 residues consisting of a highly conserved hexamer peptide motif repeated 11 times consecutively. This structure is consistent with the fact that human aggrecan is more heavily glycosylated in this region than that of rat aggrecan. Both bovine and rat aggrecan also contain this sequence; bovine has 23 copies and rat has one, slightly modified copy (Doege et al., 1991).

The KS region is not the only site of KS attachment in the protein core. KS is also distributed elsewhere, primarily in the CS region. Some of the KS chains are O-linked to threonine (Hopwood and Robinson, 1974). It is possible that the repeating Thr-Thr-Ala-Pro sequence in the N-terminal end of the CS region represents this attachment site.

KS chains from bovine aggrecan have a molecular weight of approximately 10 kDa (Hascall and Riolo, 1972). The KS chains vary in structure depending on the tissue source. KS from load-bearing tissues (articular cartilage and intervertebral disks) contain 1-3
fucose residues and 2-6 N-acetyl neuraminic acid residues, which are absent from non-load-bearing tissues (tracheal and nasal cartilage) (Nieduszynski et al., 1990). The KS-containing peptide isolated from chick aggrecan does not exhibit significant similarity to the human or bovine KS domain and may represent a further variation in this domain (Krueger et al., 1990). It has been reported that the concentration of keratan sulfate in human cartilage (especially menisci) increases with age as does the concentration of the 6-sulfated disaccharide of chondroitin sulfate (McNicol and Roughley, 1980).

1.2.2.7 CS Domain

The chondroitin sulfate (CS) domain is the largest domain of aggrecan and is decorated by approximately 100 chains of chondroitin sulfate. The domain is encoded by a single exon, exon 12, with a size of ~3.5 kbp. The CS domain consists of approximately 120 serine-glycine dipeptide repeats. There is often an acidic residue and a hydrophobic residue after the Ser-Gly dipeptide (Krueger et al., 1990). However, the glycine residue adjacent to the serine is probably the most important feature. Owing to the absence of a side chain on glycine, the protein in this region will be able to rapidly adopt a much greater number of configurations; in this region, indeed, this flexibility may be a requirement for the xylosyl transferase that adds the initiating xylose of the CS domain.

The possible recognition sequences for the attachment of CS chains have been proposed to be S-G-X-G (Bourdon et al., 1987) or (D-E)-X-S-G (Krueger et al., 1990). In addition to the primary sequence, molecular chaperone surveillance mechanisms and localization of enzymes for post-translational modifications may also be necessary for the recognition. The S-G pairs occur in two distinct patterns of repeating sequences designated the CS-1 and CS-2 repeat regions.
In proteoglycans, many, but not all, Ser-Gly dipeptides are substituted, often in regions containing flanking acidic residues. A comparison of chondroitin sulfate (CS) substitution sites in three proteoglycans, decorin (Brennan et al., 1984; Chopra et al., 1985), rat yolk sac tumor proteoglycan (Bourdon et al., 1985; Oldberg et al., 1981) and the invariant chain of human class II MHC complex molecules (Sant et al., 1985) has identified the tetrapeptide, -S-G-Xaa-G- (where X can be any amino acid) as a good substrate for xylosyltransferase (Bourdon et al., 1987). For example, aggrecan appears to have a consensus sequence with a decapeptide repeat containing two S-G pairs separated by a dipeptide (Doege et al., 1991; Upholt et al., 1993). Another sequence, acidic-G-S-G-acidic, is prominent in human versican and has also been identified as the CS attachment site in the α2 chain of chicken type IX collagen (McCormick et al., 1987). G-S-G triplet repeats and G-S pairs seem to be more common in chicken PG-M than are S-G pairs (Shinomura et al., 1993).

1.2.2.8 G3 Domain

The G3 domain, at the C-terminus of aggrecan, is a complex region that is modified by alternative splicing of exons during post-transcriptional processing of mRNA. This domain consists of three modules, epidermal growth factor (EGF)-like modules, C-type lectin-like modules also known as carbohydrate recognition domain (CRD), and complement regulatory protein-like module (CRP) as well as a short tail. The structure of the G3 domain is also found in other ECM molecules and proteoglycans including PG-M/versican, neurocan, and brevican. PG-M/versican and neurocan also have two EGF-like modules. In humans, variable alternative splicing results in about one quarter of aggrecan molecules containing an EGF-1 and a small proportion containing EGF-2 or both modules (Fulop et al.,
In mouse, rat, and dog, EGF-1 is a part of an intron and is not translated. In chicken, alternative splicing gives rise to aggrecan molecules with only one of the EGF modules encoded by exon 14. The three submodules of the C-lectin and the two CRP submodules in chicken are encoded by exons 14, 15, 16, 17 and 18 respectively.

The lectin module of aggrecan binds to fucose and galactose (Halberg et al., 1988). The recombinant C-terminal region (EGF-like module, C-type lectin module, and CRP-module) of PG-M/versican can bind to heparin and heparan sulfate (Ujita et al., 1994). The lectin module of versican binds to tenascin-R by protein-protein interactions (Aspberg et al., 1995). Comparison of tenascin-R interactions among a family of the G3 domain-containing proteoglycans revealed not only a carbohydrate-protein interaction, but also a distinct protein-protein interaction (Aspberg et al., 1997).

Aggrecan G3 domain is involved in secretion and GAG chain attachment, as well as ER quality control (through ubiquitin-proteosome dependent degradation pathway). The involvement of aggrecan G3 in secretion has been highlighted by two animal autosomal recessive mutations, nanomelia in chickens and cartilage matrix deficiency in mice.

In nanomelic aggrecan, a point mutation at position 4553 (G to T tranversion) at the end of exon 12 encoding the CS domain converts the codon GAA for glutamate at the amino acid position 1513 to TAA giving rise to a stop codon. This single mutation leads to a shortened core protein precursor with a calculated molecular weight of 158 kDa. The resulting phenotype, nanomelia, arises because the truncated core protein is neither processed nor secreted from the chondrocytes (Li et al., 1993). Homozygous nanomelic chickens die shortly after birth due to severe skeletal problems highlighting aggrecan function in skeletal morphogenesis.
Cmd, cartilage matrix deficiency, is an autosomal recessive mutation seen in mice. The aggrecan cmd gene has a single 7bp deletion in exon 5 which encodes the TR1 loop of the G1 domain. This deletion causes a frameshift resulting in the introduction of a termination codon in exon 6 (Watanabe et al., 1994).

The afflicted homozygous cmd animals die shortly after birth due to respiratory problems (stemming from malformations of the tracheal cartilage and rib cage), hinting again to the role of aggrecan in skeletal morphogenesis. The heterozygous animals die after 12-15 months, whereas wild type mice live for 2-2.5 years. Even though the cmd+/cmd- animals appear normal at birth, later in life they develop dwarfism, cleft palate (Kimata et al., 1981), herniation of the vertebral discs, deformation of the vertebral bodies, degenerative changes in the cervico-thoracic spine, and compression of the spinal cord by the herniated discs. The compression of the spinal cord by the herniated discs explains the spastic gait that these animals develop. These animals starve to death due to the spastic gait and moving difficulties (Watanabe et al., 1998). There are multiple reports on the role of G3 modules on aggrecan secretion and post-translational modifications. One study of the role of G3 in an aggrecan context suggests that G3 exerts its functions mainly through its lectin-like modules encoded by exon 15 in chicken (Domowicz et al., 2000b).

1.2.2.9 Post-Translational Modifications of Cartilage Aggrecan

1.2.2.9a GAG Chain Attachments

The feature that distinguishes proteoglycans as a special subset of glycoproteins is the glycosaminoglycan chains, which are covalently attached to the protein core as a post-translational modification. Proteoglycans (PGs) consist of a protein portion and long, unbranched polysaccharides (glycosaminoglycans or GAGs) attached covalently. The latter
have a high negative charge, owing to the presence of acidic sugar residues and/or modification by sulfate groups. The acidic sugar alternates with an amino sugar in repeated disaccharide units. The GAGs adopt an extended conformation, attract cations, and bind water. Hydrated GAG gels enable joints and tissues to absorb large pressure changes.

In addition to buffering pressure changes, PGs play important roles in control of growth and differentiation. Certain sulfation patterns in the GAG chains for example, allow ionic interactions with growth factors. Recent studies have identified ~30 PG protein cores. These cores are not just scaffolds for GAGs: they contain domains that have particular biological activities (Iozzo, 1998). Many PGs are thus multifunctional molecules that engage in several different specific interactions at the same time. After synthesis PGs are transported from the Golgi to their destinations: the ECM, the cell surface or intracellular organelles. Such vectorial transport requires mechanisms for recognition, sorting and delivery, which are especially important in cells such as epithelial cells and neurons where the cell membrane comprises separate domains. Recognition and sorting must require determinants in the GAG chains and/or in the PG protein cores.

The GAG chains may be chondroitin sulfate (or its epimerized homologue dematan sulfate), keratan sulfate, heparan sulfate or heparin. GAG chains are responsible for the extraordinarily poor resolution of proteoglycans by SDS-PAGE (sodium dodecyle sulfate polyacrylamide electrophoresis), a feature that, more than any other, hindered the identification of proteoglycans.
1.2.2.9b GAG Chain Synthesis and Molecular Structures and their Comparison to Molecular Structures and Synthesis of N- and O-Linked Oligosaccharides

GAG synthesis (CS, DS, HS and heparin) is initiated by sequential additions of four monosaccharides: xylose (Xyl), two galactoses (Gal) and glucuronic acid (GlcA) (Baker et al., 1975; Stein et al., 1982). From this linker tetrasaccharide, the sugar chains are extended by addition of two alternating monosaccharides, an aminosugar and a GlcA. In heparin and HS (heparan sulfate), the aminosugar is N-acetyl-glucosamine (GlcNAc) and in CS/DS it is N-acetyl-galactosamine (GalNAc). The extent of epimerization of GlcA to iduronic acid (IdoA) and the sulfation pattern of the disaccharide units distinguish heparin from HS, and DS from CS. In KS, the GAGs are initiated as N-linked or O-linked oligosaccharides and extended by addition of GlcNAc and Gal.

It is imperative to understand that GAGs are significantly different in terms of biogenesis and structures from those of N-linked and O-linked oligosaccharides. GAG synthesis is initiated as an O-linked addition of one xylose to a serine residue (on the hydroxyl group) on a protein core in the RER. The addition of the repeating disaccharide units occurs in the Golgi apparatus on the tetrasaccharide linkage unit (xylose, galactose, galactose and glucuronic acid).

N-linked oligosaccharides are synthesized as a 14-saccharide core unit (GlcNAc$_2$-Man$_9$-Glc$_2$ where Glc stands for glucose, Ac stands for acetyl and Man stands for mannose) that is assembled on both sides of the ER membrane (Burda and Aebi, 1999; Gahmberg and Tolvanen, 1996; Kornfeld and Kornfeld, 1985). The completed core oligosaccharide is transferred from the dolichylpyrophosphate carrier to a growing chain, nascent polypeptide chain, and is coupled through an N-glycosidic bond between the first GlcNAc and the side
chain of an asparagine residue (amide group) from the protein core. The oligosaccharyltransferase responsible for this transfer is a complex enzyme with its active site in the ER lumen (Silberstein and Gilmore, 1996). It recognizes a specific conformation of the glycosylation sequon (the sequence Asn-Xaa-Ser-Thr, where Xaa can be any amino acid except proline) transiently formed when the growing nascent polypeptide chain emerges from the translocon (Bause, 1983; Imperiali and Rickert, 1995). Of all the sequons, it has been estimated that 90% are glycosylated (Gavel and von Heijne, 1990). Immediately after coupling to the polypeptide chain, terminal glucose and mannose residues are removed by ER glucosidases and mannosidases (Kornfeld and Kornfeld, 1985; Moremen et al., 1994). In fact, the three glucose molecules are trimmed away by glucosidase I and II, and the terminal mannoses by one or more different ER mannosidases. When the glycoprotein moves to the Golgi complex, the glycan chains undergo further trimming. Due to these types of processing of N-linked oligosaccharides, a core of five sugar residues (Man-Man-Man-GlcNAc-GluNAc) is always present in all O-linked glycans. In many cases, new sugars are added to this pentasaccharide core during terminal glycosylation to produce various types of N-linked glycans such as high-mannose, complex and hybrid types. If the core structure is predominantly substituted by mannose, it is called high-mannose type N-glycan. When the core structure is substituted by one or more of the sugars such as N-acetylglucosamine, galactose, fucose or sialic acid, it is called a complex type. Hybrid type N-glycans have structural features of both the high-mannose and the complex type chains.

O-linked glycans are characterized by the presence of fucose, galactose, N-acetylglactosamine, N-acetylgalactosamine and may also have acidic sugars such as N-acetylgalactosamine (NANA), N-glycosylneuraminic acid and glucuronic acid. O-linked
glycans can also contain sulfate and phosphate residues. Mannose and glucose have been found as components of O-glycans, but they are not common. Unlike N-linked glycans, O-linked glycans are synthesized exclusively in the Golgi compartment and are manufactured one residue at a time. O-linked oligosaccharides are covalently linked to a core protein via O-glycosidic linkage to the hydroxyl groups of serine, threonine or hydroxylysine. The linkage of carbohydrate to hydroxylysine is generally found only in the collagens.
Glycosaminoglycans, repeating sequence and sulfation  
Linkage to core protein

<table>
<thead>
<tr>
<th>Chondroitin sulfate</th>
</tr>
</thead>
<tbody>
<tr>
<td>β1,4 β1,3 β1,4 β1,3 β1,4 β1,3 β1,4 β1,0</td>
</tr>
<tr>
<td>GlcUA → GalNAc → GlcUA → GlcUA → Gal → Gal → Xyl → Ser</td>
</tr>
<tr>
<td>4 or 6-SO³⁻</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dermatan sulfate</th>
</tr>
</thead>
<tbody>
<tr>
<td>β1,4 α1,3 β1,4 α1,3 β1,4 α1,3 β1,4 α1,3 β1,4 α1,3 β1,4 α1,3 β1,4 α1,3</td>
</tr>
<tr>
<td>IdUA → GalNAc → GlcUA → GalNAc → GlcUA → Gal → Gal → Xyl → Ser</td>
</tr>
<tr>
<td>+/- 2-SO³⁻  4-or 6-SO³⁻</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Haparan sulfate and Heparin</th>
</tr>
</thead>
<tbody>
<tr>
<td>α1,4 α1,4 β1,4 α1,4</td>
</tr>
<tr>
<td>IdUA → GlcN → GlcUA → GlcNAc → GalNAc → Gal → Gal → Xyl → Ser</td>
</tr>
<tr>
<td>2-SO³⁻  Ac⁺ or SO³⁻</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Keratan sulfate</th>
</tr>
</thead>
<tbody>
<tr>
<td>β1,4 β1,3 β1,4 β1,3</td>
</tr>
<tr>
<td>GlcNAc → Gal → GlcNAc → GlcNAc → GlcNAc → GalNAc → GalNAc → Ser(Thr)</td>
</tr>
<tr>
<td>Sa → Gal β1,3</td>
</tr>
<tr>
<td>6-SO³⁻</td>
</tr>
</tbody>
</table>

Figure 6-1. A schematic diagram of the GAG chain tetrasaccharide linkage regions, the repeating disaccharide units and their sulfation patterns

Abbreviations: GlcUA, glucuronic acid; GalNAc, N-acetyl galactosamine; Gal, galactose; Xyl, xylose; Ser, serine; IdUA, iduronic acid; GlcN, glucosamine; Sa, sialic acid; Thr, threonine; Man, mannose; Asn, asparagine

* Variable N-sulfation in heparan sulfate and extensive N-sulfation in heparin
Figure 7-1. A schematic structure of the N-linked core oligosaccharide added to polypeptides in the ER

Abbreviations: Glc, glucose; Man, mannose; GlcNAc, N-acetyl glucosamine; GI, glucosidase I; GII, glucosidase II; ERMI, endoplasmic reticulum mannosidase I; ERMII, endoplasmic reticulum mannosidase II; Asn, asparagine
The linkage of carbohydrate to 5-hydroxylysine is either via the single sugar galactose or the disaccharide glucosylgalactose. In serine and threonine type O-linked glycoproteins, the carbohydrate directly attached to the protein is N-acetylgalactosamine (GalNAc), whereas in N-linked glycoproteins, it is N-acetylglucosamine (GlcNAc).

1.2.2.9c Chondroitin Sulfates and Dermatan Sulfates

Chondroitin sulfate, CS, and dermatan sulfate, DS, chains are GAGs that derive from the polymer [D-glucuronic acid β (1→3) D-N-acetyl galactosamine β (1→4)]_n. Both may be modified by sulfation at position 2 of uronic acid and positions 4 or 6 of the amino sugar. N-sulfation does not occur. The distinction between the chondroitin sulfate and dematan sulfate involves epimerization of glucuronic acid to iduronic acid: if this modification is found with some frequency, the GAG is called DS; if not, it is called CS.

The termination of chain elongation in CS may be achieved by 4-sulfation of terminal hexosamine residues since this, together with selective 4- and 6- sulfation, is a feature of newly synthesized CS chains (Otsu et al., 1985); CS chains ending in N-acetylgalactosamine 4-sulfate are unable to accept a glucuronic acid (Silbert, 1978; Stevens et al., 1984). This linkage is characterized as being alkali labile as at high pH, there is a β elimination reaction that effectively cleaves the xylose-serine bond and converting serine to dehydroalanine. This releases xylose, which remains attached to the chondroitin sulfate/dermatan sulfate chain. Xylose survives intact and is unmodified by the treatment with alkali.

There is also local variability in the epimerization and sulfation in each GAG chain. Studies of these patterns have defined the motifs required for specific interactions with growth factors, cytokines, matrix components, enzymes and other proteins (Salmivirta, M., et al., 1996). The minimal requirement for binding to GAGs may differ from protein to protein.
For instance fibroblast growth factors FGF-1 and FGF-2 are recognized by different HS (heparan sulfate) structures expressed in discrete domains of the HS polymers (Kreuger et al., 1999). Chlorate treatment of MDCK cells reduces 6-O-sulfation; higher concentrations also reduce 2-O-sulfation, but N-sulfation of HSPG (heparan sulfate proteoglycan) is not affected (Safaiyan et al., 1999). In parallel with the dose-dependent loss of 6-O-sulfation, there is a reduction in binding to FGF-1, whereas binding to FGF-2 is essentially unchanged (Kreuger et al., 1999).

1.2.2.9d Keratan Sulfates

Keratan sulfate chains are synthesized attached to two different types of linkages to protein, one involving an O-linkage between serine/threonine and N-acetylgalactosamine (KS-II or skeletal KS), and the other an N-linkage between N-acetylglucosamine and asparagine (KS-I or corneal KS). Both types of linkages are also found with some nonsulfated oligosaccharides of glycoproteins. They are distinguished from each other by their sensitivity to alkaline cleavage, as the O-linkage is labile and N-linkage is stable. Both forms have the same disaccharide repeat of galactose linked β(1→4) to N-acetylglucosamine, these disaccharides being linked β(1→3) to one another. Sulfation occurs at the C-6 position of most of the N-acetylglucosamine residues and of some of the galactose residues.

1.2.2.9e GAG Synthesis

The cell takes up the building blocks for GAG synthesis, monosaccharides and sulfate, through specialized transporter complexes in the plasma membrane. Sugars (with a few exceptions) and sulfate are then activated by nucleotide consumption in the cytosol to form UDP-sugars and 3'phosphoadenosine 5'-phospho-sulfate (PAPS), respectively.
Specific transporters then translocate UDP-sugars and PAPS (Mandon et al., 1994) into the endoplasmic reticulum (ER) and Golgi lumens (Hirschberg, (Hirschberg et al., 1998; Hirschberg and Snider, 1987). Glycoproteins and glycolipids are also often sulfated. PAPS is the universal donor of sulfate to all sulfotransferases, both in the Golgi and the cytosol.

Although the lumen of the Golgi apparatus is the main site for GAG synthesis, the formation of the linker tetrasaccharide might start earlier in the secretory pathway. In chicken chondrocytes, xylosylation clearly takes place in a pre-Golgi compartment (Kearns et al., 1991; Kearns et al., 1993; Vertel et al., 1993a). Xylosylation was more efficiently catalyzed by detergent-treated Golgi fractions than by ER fractions from rat liver. Nuwayhid and Lohmander have also put forward kinetic arguments for Golgi localization of xylosyl transferase (XT) in rat chondrosarcoma cells (Lohmander et al., 1980; Lohmander et al., 1989b; Nuwayhid et al., 1986). CHO cells lacking either XT (Esko et al., 1985), galactosyl transferase I (GT I) (Esko et al., 1987) or glucuronic acid transferase I (GlcAT I) (Bai et al., 1999) can synthesize neither HS nor CS, which indicates that the pathways for synthesis of the linker tetrasaccharides of these GAG types share enzymes. Cells lacking GT I still synthesize GAGs on xylosides (Xylosides: compounds in which xylose is coupled to a hydrophobic group). Xylosides cross membranes and initiate GAG synthesis, bypassing the need for xylosylated core proteins. The GAGs initiated on xylosides are in almost all cases of the CS type, however (Esko et al., 1987). Actually, decorin protein cores contain xylose and one or two Gal residues in the presence of xylosides (Moses et al., 1999). The competition between synthesis of xyloside-based GAGs and endogenous CSPG might thus be mainly at the level of chain polymerization. Recently, a cDNA encoding a novel GT was transfected into GT-I-deficient CHO cells, restoring the PG synthesis (Okajima et al., 1999).
Patients with Ehlers-Danlos syndrome have previously been shown to exhibit reduced GT I activity (Quentin et al., 1990), and two missense substitutions have now been identified in the GT I gene (Almeida et al., 1999).

A mutant MDCK cell line exhibiting reduced import of UDP-Gal into the Golgi lumen, has dramatically reduced synthesis of KS, but essentially normal CS and HS production (Toma et al., 1996). Gal is incorporated into the polymerizing GAG chain in KS, whereas in CS/DS and HS/heparin Gal is found only in the linker tetrasaccharide. GTs involved in KS chain polymerization (GT III and IV) and linker tetrasaccharide synthesis (GT I and II) might have different Kms for UDP-Gal, but another possibility is that linker tetrasaccharide synthesis is localized to a compartment excluded from the MDCK cell Golgi fraction (Brandli et al., 1988). A pre-Golgi UDP-Gal transporter activity has not been demonstrated (Kawakita et al., 1998), but a functional GT has been localized to the ER (Sprong et al., 1998). UDP-Gal must therefore also be available in the lumen of this compartment.

Enzyme studies and a study using xylosides to prime GAG synthesis indicated that GT I and II localize to different sub-regions of rat liver Golgi (Etchison et al., 1995). The last enzyme needed for synthesis of the linker tetrasaccharide, GlcAT I, has a dual density distribution after gradient fractionation. The distribution resembles that of GlcAT II, which is involved in CS polymerization, but is clearly different from that of GT I and II (Sugumaran et al., 1998). Dual localization of enzymes involved in GAG synthesis could be interpreted in the context of the recently revived Golgi cisternal maturation model, in which a fraction of the transferases is constantly being transported retrogradely in vesicles from maturing Golgi
cisternae (Bonfanti et al., 1998; Glick, 2000; Glick and Malhotra, 1998; Mironov et al., 1998).

After completion of the linker tetrasaccharide, the addition of the fifth saccharide determines whether the GAG chain becomes CS/DS or HS/heparin. This sugar is GlcNAc in the case of HS/heparin and GalNAc in the case of CS/DS. GlcNAcT I (Fritz et al., 1997) and GalNAcT I mediate addition of these sugars (Sugumaran et al., 1998) and are postulated to be distinct from those enzymes used in elongation of the GAG chains of both CS (Rohrmann et al., 1985) and HS (Fritz et al., 1994). Nadanaka and co-workers (Nadanaka et al., 1999), however, provide evidence that the same enzyme catalyses the addition of the fifth sugar (GalNAc) to the linker tetrasaccharide and GalNAc addition during CS polymerization.

It has been shown that a GalNAcT is able to add GalNAc to the linker tetrasaccharide to produce structures that have 'unproductive' pentasaccharides because a \( \beta \) -linkage is required for further elongation of CS chains (Kitagawa et al., 1995). The same group has recently shown that this enzyme also catalyses the addition of a GlcNAc residue (Kitagawa et al., 1999; Kitagawa et al., 1998), which is the initial sugar in HS chain synthesis. This indicates that the enzyme is the GlcNAcT I active in HS and heparin synthesis. This enzyme could be important for regulation, capping some linker tetrasaccharides with a GalNAc residue and preventing their elongation. Several other factors might also regulate GAG-type synthesis.

Formation of CS/DS and HS/heparin requires UDP-Xyl, UDP-Gal, UDP-GlcA, UDP-GlcNAc and UDP-GalNAc in the Golgi and/or ER lumens. Uptake of UDP-sugars, ATP and PAPS into Golgi vesicles in vitro increases the concentration 50- to 100-fold in comparison with the incubation medium (Hirschberg and Snider, 1987). The UDP-Gal transporter (Ishida
et al., 1996; Miura et al., 1996) and the UDP-GlcNAc transporter (Guillen et al., 1998) from mammalian species have been characterized and cloned, whereas the UDP-GalNAc transporter from rat liver has been characterized but not yet cloned (Puglielli and Hirschberg, 1999; Puglielli et al., 1999).

Reduced levels of UDP-Gal in the Golgi lumen in intact MDCK cells give reduced levels of KS (Toma et al., 1996). In an in vitro system, the chain length of heparin GAGs produced is determined by the ratio of UDP-GlcNAc to UDP-GlcA (Lidholt et al., 1997). In a similar way, the ratio of UDP-GalNAc and UDP-GlcNAc could influence the extent of CS/DS versus HS/heparin synthesis. Normally, UDP-GalNAc is formed by epimerization of UDP-GlcNAc, which is catalyzed in the cytosol by UDP-GlcNAc-4-epimerase, an enzyme that also catalyses epimerization of UDP-glucose to UDP-Gal (Piller et al., 1983). A kidney cell line deficient in this enzyme has defects in the synthesis of glycolipids and N-linked and O-linked carbohydrate chains of glycoproteins. The defect can be corrected by exogenous Gal and GalNAc (Kingsley et al., 1986). Interestingly, a UDP-GlcNAc pyrophosphorylase isolated from kidney can use GalNAc-1-P and UTP to catalyze the formation of UDP-GalNAc (Szumilo et al., 1996). Thus, alternative pathways for formation of UDP-Gal and UDP-GlcNAc seem to exist. The enzymatic systems producing the different UDP-sugars, and the translocators, are thus likely to influence the concentration of UDP-sugars in the Golgi lumen and therefore also GAG-chain synthesis.

1.2.2.9f Sulfation

Sulfation is a necessary modification of proteins, carbohydrates, and lipids. One of the most prominent roles of sulfation in vertebrates is the post-translational modification of
glycosamino-glycans, the repeating disaccharide chains that are covalently linked to protein cores to constitute the family of proteoglycans. Sulfation requires active transport of sulfate into the cell, conversion into the "active" high energy form of phosphoadenosine-phosphosulfate (PAPS), translocation of PAPS across the Golgi membrane, and transfer of sulfate from PAPS, via a multitude of sulfotransferases, to the recipient biomolecules (Schwartz et al., 1998). The sulfate activation pathway consists of two activities, ATP-sulfurylase, which catalyzes synthesis of adenosine-phosphosulfate (APS) from ATP and $\text{SO}_4^{2-}$, and APS kinase, which phosphorylates APS in the presence of another molecule of ATP to form PAPS. In simpler organisms these activities are catalyzed by separate enzymes (Jain and Leustek, 1994; Korch et al., 1991; Leyh and Suo, 1992). In contrast, when these two sulfate-activating activities were purified from rat chondrosarcoma, they were found to exist as a bifunctional enzyme (Lyle et al., 1994), which uses a channeling mechanism to transfer the intermediate APS efficiently from the sulfurylase to the kinase active site (Lyle et al., 1994; Lyle et al., 1995). A mouse cDNA encoding a fused ATP-sulfurylase/APS kinase has also been isolated. This mammalian PAPS synthetase (referred to as SK1) is a bifunctional polypeptide that catalyzes both ATP-sulfurylase and APS-kinase reactions (Li et al., 1995).

1.2.2.10 Secretory Pathway of Aggrecan

The feature that distinguishes proteoglycans as a special subset of glycoproteins is the GAG chains, which are covalently attached to the protein core as a post-translational modification. Proteoglycans (PGs) consist of a protein portion and long, unbranched polysaccharides (glycosaminoglycans or GAGs). The latter have a high negative charge, owing to the presence of acidic sugar residues and/or modification by sulfate groups. The
acidic sugar alternates with an amino sugar in repeated disaccharide units. The GAGs adopt an extended conformation, attract cations, and bind water. Hydrated GAG gels enable joints and tissues to absorb large pressure changes.

Aggrecan is synthesized and secreted continuously by chondrocytes (via the constitutive secretory pathway). It follows the same intracellular pathway of synthesis as any other secretory molecule (see the section on the intracellular protein trafficking) (Hardingham and Fosang, 1992). The mRNA is translated on the membrane-bound ribosomes into the RER, followed by translocation to the Golgi for the main steps of glycosylation and glycosaminoglycan chain synthesis. The GAG chains appear to be synthesized rather rapidly on aggrecan as part of a highly concerted process that occurs just before secretion. There is no intracellular storage of the finished molecules prior to their release. Link protein, although it is less glycosylated and has no GAG chains attached, is also synthesized along the same intracellular pathway. However, HA is not synthesized within these same intracellular compartments, but it is formed by synthase enzymes that are located in the plasma membrane, such that HA is synthesized and secreted directly into the extracellular matrix (Spicer and McDonald, 1998). It is interesting to know that link protein is able to bind HA and aggrecan shortly after its secretion from the chondrocytes, whereas aggrecan (the G1 domain) needs to undergo further maturation in the ECM by forming disulfide bonds. The formation of these disulfide bonds is the last step in the aggrecan maturation which enables this molecule to bind HA.

1.2.2.11 Large Aggregating Chondroitin Sulfate Proteoglycans, PG-M/Versican

PG-M (chick) and its human homologue, versican, belong to the aggrecan family of CSPGs. The structural organizations of PG-M and versican are very similar to that of
aggrecan except that PG-M and versican lack IGD and G2 domains. PG-M is one of the major ECM molecules in the mesenchymal cell condensation regions of developing limb buds (Kimata et al., 1986). The gene is large, with 15 exons distributed among its 100 kbp. The CS attachment region is encoded by two large exons, VII and VIII, spanning ~3 and 5 kbp respectively. These exons are alternatively spliced out without frameshift, giving rise to four different isoforms, PG-M (V0), containing all domains; PG-M (V1), missing the first CS attachment domain; PG-M (V2), lacking the second CS domain; and PG-M (V3), lacking all the CS domains. The expression, however, is regulated in an inverse relationship to that of aggrecan, in that PG-M disappears after cartilage development (Shinomura et al., 1990). Transient expression of this proteoglycan is also seen in various embryonic tissues during morphogenesis and differentiation (Yamagata et al., 1993a; Yamagata et al., 1993b). Therefore PG-M can be expected to play a regulatory role in such biological events.

Versican G1 domain is structurally similar to aggrecan G1 and mediates interaction with hyaluronic acid, possibly through its tandem repeats. It has also been shown that versican G1 domain cannot be secreted efficiently if expressed on its own and this inhibition of secretion can be traced to the actions of the tandem repeats (Yang et al., 2000). Versican CS domains are in GAG chain attachment and it is speculated that the different amounts of GAG chains in different isoforms could point to the different roles these isoforms may have in various tissues. Versican G3 domain consists of two EGF-like modules, one CRD domain and CRP domain. It has been demonstrated that EGF-like modules of versican can promote cell proliferation (Zhang et al., 1998). The EGF-like domain has also been implicated in modulating cell morphology (Zhang et al., 2001). Versican interacts with tenascin-R via its
CRD motif. Finally, versican can enhance product secretion and GAG modifications via its CBP motif.

1.3 Intracellular Trafficking of Secretory Proteins

Eukaryotic cells are characterized by the existence of a number of membrane-bound subcellular secretory systems, which perform a variety of specialized functions. The secretory membrane system allows cells to regulate delivery of newly synthesized proteins, carbohydrates, and lipids to various cellular compartments as well as to the cell surface, which is a necessity for growth and homeostasis. The system is made up of distinct organelles including the endoplasmic reticulum (ER), Golgi complex, and plasma membrane as well as tubulo-vesicular transport-intermediates that mediate intracellular membrane transport between them (Fig. 8-1). Membrane traffic within this system flows along highly organized directional routes. Secretory cargo is synthesized and assembled in the ER and then transported to the Golgi complex for further processing and maturation. Upon arrival at the trans Golgi network (TGN), it is sorted and packaged into post-Golgi carriers that move through the cytoplasm to fuse with the cell surface. The translocation of proteins across biological membranes is a fundamental process of intracellular trafficking and organelle biogenesis. Entry into the secretory pathway occurs by translocation of proteins into or across the membrane of the ER. The ER is the starting point of the secretory pathway. It is the largest intracellular compartment, with an extensive array of interconnecting membrane tubules and cisternae that extend throughout the cell including the nuclear envelope. ER membranes are physiologically active, interact with the cytoskeleton, and contain differentiated domains specialized for distinct functions.
Secretory cargo destined to be secreted or to arrive at the plasma membrane (PM) leaves the ER via distinct exit sites that bud and translocate as tubular-vesicular structures (pre-Golgi) toward the (-) end of microtubules. Here they merge with Golgi membranes (Golgi), which in many mammalian cells are located near the microtubule organizing center (MTOC). After passing through the Golgi complex, secretory cargo is packed into post-Golgi transport intermediates (post-Golgi), which translocate plus-end directed along the microtubules to the plasma membrane.
These functions include protein folding, assembly and degradation, lipid metabolism, detoxification, regulation of calcium ion gradients, and membrane transport. ER membranes are differentiated into rough and smooth regions (RER and SER, respectively), depending on whether or not ribosomes are associated with their cytoplasmic surfaces. Whereas the RER is the site of co-translational membrane insertion of proteins, the SER is thought to be the site of lipid biosynthesis, detoxification, and calcium regulation. The protein translocation process involves two distinct steps which are dependent on the orchestrated action of several proteins. The initial step of targeting involves recognition of the signal sequences and delivery of the protein precursor to the ER in a translocation-competent conformation. The subsequent translocaional event is characterized by interaction of the preprotein with the translocation channel followed by unidirectional movement across the lipid bilayer of the ER membrane into the lumenal space. Therefore, upon post- or co-translational insertion into ER membranes, newly synthesized proteins encounter lumenal sugar transferases and chaperones resident in the ER (including BiP, calnexin, calreticulin, and protein disulfide isomerase) whose role is to facilitate some post-translational modifications (glycation and folding reactions are mainly necessary for protein maturation and oligomerization Helenius, 1992). The early post-translational glycation of proteins is of significance since it is known that glycans play major roles in promoting protein folding (by keeping the proteins in a soluble state), quality control (Hammond & Helenius 1995), and certain sorting events (Varki, 1998). Proteins which are incorrectly folded are degraded or retained in this compartment highlighting the ER’s important quality control role in protein transport into the secretory pathway. The finished products then bud off the RER and find their way into the Golgi apparatus where they undergo further post-translational modifications (mainly sugar
attachment) by the Golgi membrane-bound enzymes such as glycosidases, glycosyltransferases, proteolytic enzymes and permeases. Unlike the ER, the Golgi complex does not have a rigorous system for controlling the fidelity of its biosynthetic processes. For example, when cells are treated with glycosidase inhibitors, or when specific glycosyltransferases are mutated, glycoproteins carrying defective glycan moieties are readily exported (Eibein, 1991; Stanley, 1984). Even severe glycosylation defects or deficiencies are, as a rule, tolerated by the Golgi without retention or degradation. Once perfected in the Golgi, the protein products bud off from the Golgi and find their way to the plasma membrane via microtubules in the form of secretory vesicles.

1.3.1 Targeting of Nascent Polypeptide Chains to the RER

Proteins entering the secretory pathway contain an ER targeting signal, commonly referred to as the signal peptide sequence. This signal peptide is usually located at the amino terminus and consists of a continuous stretch of hydrophobic residues (6-20 aa) flanked by one or more basic residues to the N-terminal side of the hydrophobic core (von Heijne, 1990). In most cases, the signal sequence is cleaved during the translocation of the nascent chain and the remainder of the polypeptide is translocated into the lumen of the ER. Integral membrane proteins are inserted into the lipid bilayer of the ER by use of a signal-anchor sequence. Mammalian microsomes are able to translocate secretory proteins across the lipid bilayer through a co-translational mechanism. This cotranslational mechanism involves a cytosolic ribonucleoprotein complex, signal recognition particle (SRP), and its cognate SRP receptor (SR) in the ER membrane. SRP binds to the nascent polypeptide signal sequence as it emerges from the translating ribosome. Subsequent interaction of SRP with the ribosome results in an inhibition of translation elongation. This inhibition is relieved when SRP
dissociates from the nascent polypeptide chain and the ribosome upon interaction with SR. The ribosome then interacts directly with the ER membrane and components of the translocation machinery, effectively coupling translation and translocation. For the vast majority of secretory proteins, translocation across the ER membrane in mammals has a strict requirement for SRP and therefore occurs cotranslationally. However, the exact role of SRP in precursor targeting has been complicated by the discovery of the nascent-polypeptide-associated complex, NAC (Wiedmann et al., 1994). In the absence of additional factors, eukaryotic ribosomes have an intrinsic ability to bind the ER membrane (Jungnickel and Rapoport, 1995; Lauring et al., 1995). NAC appears to block the ER membrane binding site on the ribosome and also prevents binding of SRP to non-signal sequences (Wickner, 1995). Some functions attributed solely to SRP now appear to depend on the interplay between SRP and NAC. SRP homologues have been identified amongst distantly related organisms such as the yeast *Saccharomyces cerevisiae* suggesting that the SRP dependent pathway is ubiquitous amongst eukaryotes. Surprisingly, mutant yeast cells lacking SRP are viable albeit with a severe growth defect and defects in the translocation of some proteins (Brown et al., 1994; Hann and Walter, 1991; Stirling et al., 1992). This observation implies that targeting to the ER can occur by redundant pathways *in vivo*. The nature of the SRP-independent mechanism is at present unclear. One possibility is that it represents a post-translational mechanism which has been demonstrated *in vitro* for precursors as large as 250 residues (Hansen and Walter, 1988). In this case precursors that are released from the ribosome after translation termination are maintained in a "translocation competent" state by interaction with HSP70 molecules. Evidence indicates that the targeting of the vacuolar protease Carboxypeptidase Y (CPY) occurs independently of SRP and HSP70, suggesting the existence of an alternative
pathway which may involve novel cytosolic targeting components (Caplan et al., 1992; Hann and Walter, 1991; Hansen and Walter, 1988).

1.3.2 Translocation of Polypeptide Chains to the RER

In eukaryotes, the vast majority of secreted and integral membrane proteins are targeted to the membrane of the endoplasmic reticulum early during translation. These polypeptides are then either transported across or inserted into the ER membrane at sites termed translocons. Translocons are composed of several ER membrane proteins that associate to form an aqueous pore through which secretory proteins and lumenal domains of membrane proteins pass from the cytoplasm to the ER lumen. The pore of the translocon (sec61 complex) is the largest pore in the ER membrane, with an estimated diameter of 40-60 Å in the ribosome-bound state and a smaller diameter of 9-15 Å in the ribosome-free state (Hamman et al., 1998). When bound by a ribosome, the pore of the translocon is aligned with the peptide exit tunnel in the large subunit of the ribosome (Beckmann et al., 1997), and the average diameter of this tunnel is about 20 Å (Morgan et al., 2000). It should be noted that these pores are not passive holes in the membrane bilayer, but instead are quite dynamic both structurally and functionally. Translocons cycle between ribosome-bound and ribosome-free states, and convert between translocation and integration modes of operation. In translocons, polypeptides are translocated via an aqueous channel formed, at least in part, by integral membrane proteins of the ER (High, 1995; Martoglio et al., 1995). Genetic studies in the budding yeast Saccharomyces cerevisiae led to the identification of the Sec61p, a multi-spanning integral membrane protein of the ER required for the translocation of both secretory and membrane proteins (Stirling and Hewitt, 1992; Wilkinson et al., 1996). Crosslinking experiments have shown Sec61p to be in intimate contact with prepro-α-factor (pp α F)
during translocation across the bilayer suggesting that this protein corresponds to a
compartment of the translocon per se (Musch et al., 1992; Sanders et al., 1992). Substantive
evidence that Sec61p/Sec61alpha is a core component of ER translocon has been provided by
the ability of the purified proteins to promote translocation reactions when reconstituted into
liposomes. The cotranslational translocation of proteins into mammalian-derived
proteoliposomes requires only SRP, SR and the Sec61-complex (Gorlich and Rapoport,
1993). Although it remains to be formally proven, an increasing body of evidence strongly
supports the possibility that the nascent chain-ribosome-regulated pores (Simon and Blobel,
1991) are formed at least in part by the Sec61complex.

1.3.3 Protein Folding in the RER

The ER is the major protein folding compartment for secreted, plasma membrane and
organelle proteins. Each of the newly synthesized polypeptides folds in a determined process,
affected by the unique conditions that exist in the ER. An understanding of protein folding in
the ER is a fundamental biomolecular challenge at two levels. The first level tries to address
how the amino acid sequence programs that polypeptide to efficiently arrive at a particular
fold out of a multitude of alternatives, and how different sequences obtain similar folds. At
the second level are the issues introduced by folding not in the cytosol, but in the ER,
including the risk of aggregation in a molecularly crowded environment, accommodation of
post-translational modifications and the compatibility with subsequent intracellular
trafficking. ER molecular chaperones are key to overcoming both the physicochemical and
cell biological constraints of folding.

Many features of protein folding in the RER actually reflect features common to
intracellular folding in general. Perhaps the most significant of these is that folding in vivo
takes place in a complex and dense solution of proteins. The concentration of protein in the ER lumen is estimated to reach 100 mg/ml, a concentration at which aggregation is clearly promoted. The danger of aggregation is even greater for newly synthesized polypeptides, which expose hydrophobic patches that would later be buried in the native structure. A second ‘folding unfriendly’ aspect is that \textit{in vivo} a polypeptide folds concurrently with thousands of other different polypeptides, further increasing the likelihood of co-aggregation. Therefore, minimizing aggregation is one of the main and common roles of molecular chaperones, in the ER as well as in the cytosol. In both of the major folding compartments in the cell, the cytosol and the ER, ionic conditions are thought to be generally similar, except for the differences in Ca$^{2+}$ levels. The pH of the ER is approximately 7, as assessed by electron microscopy probes and pH-sensitive dyes. Thus, in terms of titration of charges on side chains, folding in the two compartments is likely to be equivalent. The requirement for subunit assembly also poses special problems for folding \textit{in vivo}. Whether the assembly is homotypic, as in the case of trimeric viral proteins, or heterotypic, as in the cases of cytokine receptors and ATPases, the assembly process has to be executed efficiently to be compatible with cell growth and function.

While the aspects of folding discussed above are common, several conditions make the ER unique as a folding compartment. They arise from the different nature of proteins that fold in the ER, from the nature of the machinery that segregates them across a membrane from cytosolic proteins, and from the unique physiological functions of the ER. The ER differs from other folding compartments (e.g. the cytosol and mitochondria) in its high oxidizing potential (Hwang \textit{et al.}, 1992), its (occasionally) high Ca$^{2+}$ concentration (Montero \textit{et al.}, 1995), and the presence of carbohydrates and glycosylation machinery. These
conditions impact the energetics and kinetics of protein folding, as discussed below. These unique aspects also provide the evolutionary pressure for the presence of enzymes and chaperones to facilitate folding in the ER. Indeed, the most abundant lumenal proteins BiP (Grp78), GRP94, PDI (protein disulfide isomerase) and calreticulin all function in protein folding.

1.3.4 Consequences of the Presence of Signal Sequences

Since most proteins that fold in the RER are targeted there by an N-terminal signal sequence, this extension of the nascent chain has several ramifications for folding. First, as long as the signal sequence is not cleaved, the N-terminus of the mature protein is constrained differently. In many cases this means a delay in folding. In some cases the presence of the signal sequence is essential for the choice of the proper folding pathway. For example, a signal sequence that is not cleaved immediately prolongs the association of HIV-1 gp120 with ER chaperones (Li et al., 1996). Second, because of the vast discrepancy between the folding and translation rates (translation usually lags behind folding), a mechanism for delaying folding is necessary. Translation proceeds at $4 \pm 2$ residues per second, while initial folding steps such as a hydrophobic collapse can occur within nanoseconds. Thus, as soon as a typical 20 amino acid signal sequence of a secretory protein protrudes from the large subunit of a ribosome, it is bound by the Signal Recognition Particle, which prevents premature folding and targets the ribosome to the ER membrane. A delay in folding is also necessary for the translocation across the membrane, in analogy to the unfolding of mitochondrial precursors before their import into that organelle (Horst et al., 1997). Randall et al. provided strong evidence for kinetic competition between productive translocation across the membrane and folding of preproteins into stable conformations which are not
compatible with translocation (Randall, 1986). Processing of signal sequences can occur either co- or post-translationally, but even for those preproteins that are processed cotranslationally, cleavage does not occur until the nascent polypeptide reaches a minimal size of 80% of the final length. Thus, in both co- and post-translational modes, there is a time window when the presence of the signal sequence can profoundly affect folding of the mature portion of the polypeptide.

1.3.5 Consequences of Translocation of Polypeptides into RER

A major difference between folding in the ER and refolding studies in vitro is that the polypeptide chain starts folding in the lumen well before its synthesis is complete. Due to this feature, the folding opportunities for α helices and β sheets are vastly different. Helix forming amino acids are translocated sequentially and therefore are available immediately to form all the interactions that stabilize a helix. Formation of β sheets, on the other hand, must be severely delayed by the translocation process, since each peptide that would assume an extended strand conformation must ‘wait’ until the other strands are synthesized before hydrogen bonds can be satisfied. It is therefore easy to understand why the peptide-binding chaperones of the HSP70 family evolved to bind extended β strand peptides (Landry et al., 1992) enabling them to prevent inappropriate folding until the partner amino acids become available in the lumen. An implication of this difference is that if such secondary structure elements form the initial folding intermediates, then helical proteins have a kinetic advantage in the ER.

A similar folding constraint exists for those Cys residues which form disulfide bonds and which are often separated by many amino acids. Indeed, Pluckthun et al showed that the rate-limiting step in the formation of functional Ig molecules in bacteria is the shuffling of
disulfide bonds to create the proper ones (Worn and Pluckthun, 1998). In contrast, disulfide bond formation of either hCG β or influenza hemagglutinin HA in the ER occurs without a large population of incorrect S-S bonds (Ruddon et al., 1996; Segal et al., 1992). Such high fidelity suggests that binding of molecular chaperones is used to limit possible disulfide errors, thus simplifying and directing the folding pathways in vivo (Hendershot et al., 1996).

1.3.6 Consequences of Glycosylation of the Polypeptides

Carbohydrates serve several major roles in protein folding. First, because of their hydrophilic nature, carbohydrates tend to increase the solubility of the glycoprotein. Second, they generally mark the surface of folding modules and are not buried within them. Third, they make the process of translocation across the ER membrane less reversible by increasing the energy barrier to back-translocation. Apart from these general roles, however, carbohydrates also affect protein folding in more specific ways. Detailed studies with viral glycoproteins whose Asn-linked glycosylation sites were mutated systematically showed that in many cases glycosylation is needed for proper folding: under-glycosylated proteins form intracellular aggregates and are retained in the ER. Somewhat paradoxically, however, no individual glycan is necessary for folding of such viral glycoproteins (Gallagher et al., 1992). Even when highly related proteins are compared, such as MHC class I molecules, some are sensitive to the presence of the glycans, whereas others fold equivalently in the presence or absence of the carbohydrates. Furthermore, even proteins that are normally not glycosylated can sometimes benefit from the inclusion of ectopic glycans (Jost et al., 1994). Thus, the 'rules' that govern the interplay between glycosylation and folding are at present still poorly understood.
1.3.7 Folding of Proteins in the Oxidizing Environment of ER

A general property distinguishing proteins that fold in the ER from cytosolic proteins is the presence of disulfide bonds. Cys residues in newly synthesized secretory polypeptides tend to oxidize in the ER lumen because of its high oxidative redox potential. The major redox buffer is the Cys-containing tripeptide glutathione, and the ratio of reduced to oxidized glutathione ranges from 1:1 to 3:1, whereas the overall cellular ratio ranges from 30:1 to 100:1. How such an oxidative environment is maintained is not clear. However, preferential transport of oxidized glutathione from the cytosol into the ER lumen is likely a contributing factor. Disulfide bonds are very important in dictating protein folding pathways, because they form covalent folding intermediates (Creighton, 1997) and severely restrict the landscape of available conformations. Formation of disulfide bonds begins very early in the life of the protein, sometimes as soon as the required Cys residues are available in the lumen (Marquardt and Helenius, 1992). On the other hand, some disulfide bonds do not form until much later in the folding process. This dichotomy suggests that in vivo there are mechanisms that selectively delay some oxidation steps.

Furthermore, proteins can undergo post-translational oxidation and achieve the same native structure as with co-translational oxidation. Addition of reducing agents to the medium of live cells prevented disulfide bond formation in newly synthesized influenza hemagglutinin (HA0) or asialoglycoprotein receptor and reduced the proportion of already-oxidized glycoproteins inside the ER. When the reductant was washed out, the reduced proteins rapidly oxidized, folded correctly and assembled (Braakman et al., 1992). These examples show that apparently oxidation follows the same pathway whether it occurs post- or co-translationally. Both of the above paradoxes can be accounted for if ER chaperones
could selectively bind near specific Cys residues and delay their oxidation until other local folding steps can be completed. According to this view, disulfide bonds in most cases serve to stabilize a local fold, achieved by multiple co-operative interactions, rather than to initiate the folding. In support of this view, mutagenesis experiments show that lysozyme can be reasonably well folded even without some disulfide bonds (Taniyama et al., 1992) and that if an Ig is made stable enough via genetic engineering, it can withstand the removal of its disulfide bonds.

These observations also indicate the need for precise control over protein oxidation in the ER. Indeed, several lumenal enzymes are capable of shuffling disulfide bonds, directing the formation of proper intermediates, and at the same time, resolving aberrant disulfide bonds. There is currently no evidence for the existence of disulfide isomerases in any other organelle. The most abundant of the ER disulfide isomerases are the 60 kDa protein disulfide isomerase (PDI) and two structurally-related proteins, ERp72 and ERp58. Each contains two or more thioredoxin-like domains which in vitro can perform some of the thiol exchange reactions ascribed to the entire enzyme (Darby and Creighton, 1995; Fullekrug et al., 1994).

1.3.8 Folding Factors and Chaperones in the RER  

Given the complex demands of folding secreted and membrane proteins, the ER contains at least 10 general chaperones, as well as some protein-specific chaperones whose expression is tissue-specific. Several of the ER chaperones are members of the main families of molecular chaperones. The primary peptide-binding chaperone is BiP/GRP78, a member of the HSP70 family. It binds early folding intermediates of a vast array of proteins (Melnick
and Argon, 1995) because of its ability to recognize relatively short peptides in their folded state (Gething et al., 1995). The inherent affinity of BiP for peptides is low, in the range of 1-100 mM. BiP binding and release allow the substrate to bury its hydrophobic sequences as its folding progresses. Another representative of the HSP70 family in the ER is GRP170, which is associated with Ig in B cells, but whose role is not yet defined (Lin et al., 1993). A yeast homologue of GRP170, Lhs1p, cannot substitute most mutants of yeast BiP (kar2p) and seems to have a distinct function in refolding heat-damaged proteins (Saris and Makarow, 1998). Both BiP and GRP170 are thought to interact, at least for some of their functions, with the dnaJ-like ER homologues that are membrane-bound proteins (Feldheim et al., 1992; Silberstein et al., 1998). Only one HSP90 family representative is known in the ER-GRP94. It also recognizes peptides, but apparently a different subset than BiP. This distinction presumably enables GRP94 to associate with later folding intermediates than BiP, thus chaperoning different phases of the folding process (Melnick and Argon, 1995). The immunophilin/ PPI family of chaperones is represented in the ER by FKBP13, S-cyclophilin and FKBP65 (Zeng et al., 1998). Each of these proteins can catalyze isomerization of prolines, but their mode of substrate recognition, and indeed the importance of their PPI activity in facilitating protein folding, is still poorly understood, despite the ability of the inhibitor cyclosporin A to block an early stage in transferrin biosynthesis (Lodish and Kong, 1991).

The ER has some unique types of lectin-binding chaperones that bind to glycoproteins. Three such chaperones are known: calreticulin, which is a luminal protein, calnexin, the only membrane-spanning chaperone, and ERp57, which is a member of PDI (protein disulfide isomerase) family. These three chaperones have been experimentally
defined as interacting specifically with newly synthesized glycoproteins. This specificity for glycoproteins is particularly apparent, since none of the other known ER folding factors select their substrates purely on the basis of whether or not they are N-glycosylated.

1.3.9 Quality Control System in RER

The efficiency of folding in the ER can vary dramatically, from almost 100% for molecules such as Ig or viral glycoproteins to only approximately 30% for proteins such as the cystic fibrosis transporter (Ward et al., 1981). It is now well established that if folding is incomplete, transport of cargo proteins out of the ER is inhibited. The set of proteins that distinguishes between native and incompletely folded cargo proteins and which either prevents their traffic or enhances their degradation, is known collectively as "the ER quality control system". Chaperones and folding enzymes are clearly one arm of this system and they are involved in determining the fate of cargo proteins at several levels.

In its capacity as a peptide-binder, BiP is often found to associate preferentially with non-secretable or misfolded mutants, presumably because these molecules continue to expose binding sites that wild type molecules do not. In this fashion, BiP can either retard the transport of substrates (Dorner et al., 1992) or target them for degradation (Knittler et al., 1995). With its ability to both bind substrates in the ER lumen and to associate with the translocon, (Hamman et al., 1998), BiP is well-positioned to target misfolded proteins back through the Sec61p channel to the proteasome machinery.

The lectin chaperones are also well positioned to determine the fate of misfolded glycoproteins. Calreticulin and calnexin serve as sensors of the glycosylation state via their specific binding to the monoglucosylated intermediate in the processing of high mannose oligosaccharides. They cooperate with glycoprotein glucosyltransferase, the enzyme that
reattaches the terminal glucose, UDP-Glc, and can distinguish between folded and unfolded polypeptides. In this way, the glycoprotein is either recognized as immature, allowing for further folding, retained in the ER, funneled to the degradation pathway via the interactions with BiP (Hebert et al., 1995; Jethmalani and Henle, 1998) or recognized as mature enough to be transported.

1.3.10 Folding and Targeting Information

In addition to forming structures that are recognized by the quality control system, folding in the ER must be compatible with targeting information on cargo proteins which determines their sites of residence within the cell. Such signals should be exposed on the surface of the native structure, so they are available for interactions with the proteins that react to them. One example is the targeting of most lysosomal hydrolases, where the carbohydrate moiety serves as a targeting signal. In these proteins, one or more glycan chains/residues is usually modified to contain mannose-6-phosphate, which is recognized later in the secretory pathway by specialized receptors that divert lysosomal protein traffic from the constitutive secretory traffic. The Golgi enzymes that phosphorylate the mannose recognize the appropriate target protein via a 'landing site' that is not adjacent to the glycan and is formed by non-contiguous amino acids. Mutations that affect the conformation of this landing site lead to mislocalization of lysosomal enzymes and are associated with lysosomal storage diseases (Tikkanen et al., 1997). This example illustrates how folding governs the formation of a conformation-dependent transport signal. Another type of signal is the "KDEL" peptide, which must be exposed at the C-terminus in order to restrict lumenal proteins to the ER. Yet a third type of signal is the C-terminal peptide of rhodopsin, necessary for its transport from the trans-Golgi network to the rhodopsin-rich disks in rod
outer segments (Deretic et al., 1998). Evolution may have placed such traffic signals at or near the end of polypeptide chains because the N- and C-termini are often solvent exposed and are therefore available for interacting proteins.

1.3.11 The Golgi Complex

The Golgi complex occupies a central position in the secretory pathway. It seems that the Golgi complex controls two kinds of secretion pathways, the constitutive and regulated secretory pathways. The constitutive pathway allows for the secretion of proteins that are needed outside the cell, such as proteoglycans in the ECM. The regulated pathway is a pathway for proteins that require a stimulus or trigger to elicit secretion. This pathway is active in exocrine, endocrine, and neuroendocrine cells which store prohormones and neuropeptides and secretion occurs in response to an appropriate stimulus (Fig. 9-1). The Golgi complex is the major site in the cell where proteins and lipids are modified and sorted, and it acts as a filter to segregate proteins and lipids to be retained in the ER/Golgi system from those to be delivered to the plasma membrane. Transport intermediates carrying cargo derived from the ER deliver their contents uniquely to the cis face of the Golgi complex, which exists as an elaborate tubular network. The cargo molecules then move through polarized stacks of flattened cisternae enriched in glycoprotein and glycolipid-processing enzymes to the TGN where they are packaged into membrane-bound carriers destined for the plasma membrane. The forward flow of secretory cargo through the Golgi is balanced by backward flow of selected components to the ER.

It seems that the Golgi apparatus has two key functions, which could be seen as a combined assembly line and logistics center.
Figure 9-1. The consecutive and regulated secretory pathways

This figure shows the two types of secretory pathways. The regulated secretory pathway, as its name implies, is a pathway for proteins that requires a stimulus or signal to elicit secretion. Some stimuli regulate synthesis of the protein as well as its release. The constitutive pathway allows for secretion of proteins that are needed outside the cell, like in the extracellular matrix. It does not require stimuli, although growth factors may enhance the process.
The Golgi complex houses the enzymes responsible for the synthesis of complex carbohydrate structures found on many proteins and lipids, while both the cis- and trans-faces of the Golgi apparatus are important sites for the sorting of proteins and lipids for delivery to specific subcellular destinations. Complex carbohydrate modifications in eukaryotes are compartmentalized into the ER and Golgi apparatus. Proteins (as can be recalled from previous sections) can be glycosylated in two different ways, N-linked glycosylation on asparagine residues, and O-linked glycosylation on serine or threonine residues. N-linked glycosylation is initiated in the ER, with the transfer of a branched sugar structure from the isoprenoid lipid dolichol to specific asparagine residues, while O-linked glycosylation is initiated in the Golgi. Complex carbohydrate processing continues in the Golgi apparatus, where sugar transferases and sugar-trimming enzymes add or remove a variety of sugars at various stereospecific positions (Fig. 10-1). These enzymes are localized to discrete cisternae within the Golgi apparatus, in an order that corresponds to their sequence of action. The Golgi apparatus could therefore be viewed as an assembly line for the production of correctly glycosylated proteins (Fig. 10-1). The first, or cis-Golgi, cisterna is a major site for the sorting of secretory and ER-derived proteins. Many ER proteins contain the carboxyl terminus sequence “KDEL”, which is recognized by an integral membrane receptor in the Golgi and results in the packaging of these molecules in vesicles for transport back to the ER (Fig. 10-1). Other signals are found in the cytoplasmic domains of transmembrane proteins. The di-lysine motif is one such signal found at the carboxyl terminus of membrane proteins. It binds directly to the proteinaceous vesicle coat termed COP-I, causing proteins that bear di-lysine to become concentrated in vesicles destined for
the ER. The exact mechanism by which COP-I recognizes and decodes sorting signals of proteins is highly complex, as it appears to discriminate between related signals specifying either forward or retrograde transport through the Golgi apparatus (Fig. 10-1). Finally, at the trans-face of the Golgi, is the trans-Golgi network (TGN) where proteins are sorted for delivery to their final subcellular destination (Fig. 10-1). Proteins destined for the endocytic pathway, such as lysosomal enzymes, are bound by a receptor in the TGN that recognizes a specific sugar modification mannose-6-phosphate on these proteins, and are segregated into forming vesicles covered with the clathrin coat protein. Other transport events at the TGN are less well understood, the exact nature of the transport intermediates formed and the coat proteins, if any, that cover them remaining mysterious (Farquhar and Palade, 1998).
Figure 10-1. Post-Translational Modifications of Proteins in the Golgi

The new vesicles from the ER enter the cis Golgi network and retrograde vesicles (bearing COP-I) coats move to merge with the cis region cisternae. These carry Golgi complex processing enzymes and their targeting to this region may be dependent on the low concentration of these processing enzymes. Then, as processing continues, the medial cisternae contain more mature product and lower amounts of the enzymes needed in the beginning. Finally, the trans region is specialized for addition of N-acetyl neuraminic acid (NANA) to the final protein products and sorting the finished products to their specific destinations.
1.4 General Hypotheses

We have hypothesized that apart from extracellular functions, aggrecan domains also play certain functions in post-translational processing and intracellular trafficking of aggrecan, contrary to the common view that aggrecan G3 domain could be the sole player in product secretion and quality control. Based on the considerable amino acid similarities between the first tandem repeat of aggrecan G1 and G2 globular domain, we have also hypothesized that these tandem repeats perform a common function in post-translational processing and export of aggrecan.

1.5 Significance of Broadening the Field

The significance of aggrecan post-translational modification and secretion has been highlighted by the occurrence of two inherited autosomal recessive mutations in chickens and mice, giving rise to nanomelia and cmd phenotypes, respectively. Nanomelia and cmd are both lethal in a homozygous form due to chondroskeletal dysplasia. However, the heterozygous animals for the above-mentioned phenotypes appear to be normal at birth, but they develop skeletal problems later in their life. This is specially true for the cmd heterozygotes where they exhibit dwarfism by the first year of their life and soon after they show misalignment of the cervical and thoracic spine followed by hyperlordosis. These animals also develop herniation of the vertebral disc followed by spastic gait. The above-mentioned phenotypes indicate that how crucial aggrecan secretion is to the proper development of skeletal systems as well as the functioning of various joints in the body.

The occurrence of joint diseases (especially osteoarthritis) in the general population and the increase in the prevalence of these disorders with aging provide strong rationale for
advancing the field on aggrecan secretion. It has been shown that in age related osteoarthritis there is a reduction in aggrecan biosynthesis and secretion. Therefore, it can be appreciated how a comprehensive study of aggrecan biology in terms of its biosynthesis, intracellular trafficking and secretion would benefit us in a long term to design therapeutic strategies to treat the condition.

It is the intention of this study to broaden the knowledge of the role of aggrecan domains in processing and secretion. This will in turn enable us to design custom-made aggrecan recombinant genes which are able to be properly GAG decorated, secreted and be able to function as efficiently as the wild type molecule in the ECM.

1.6 General Summary

Aggrecan is the shortened name of the large aggregating chondroitin sulfate proteoglycan. Aggrecan, which is one of the most widely studied proteoglycans, is abundant; it represents up to 10% of the dry weight of cartilage (articular cartilage is up to 75% water). An aggrecan monomer consists of a protein backbone of 210-250 kDa to which is attached both chondroitin sulfate (100-150 per monomer) and keratan sulfate (30-60 per monomer) as well as many O- and N-linked oligosaccharides. Up to 100 individual aggrecan monomers, interact with hyaluronic acid to form an aggregate of very high molecular weight (100-200 mDa). A short molecule called link protein, which interacts with both hyaluronic acid and aggrecan, further stabilizes this interaction. The primary role of aggrecan in articular cartilage appears to be a physical one, as it brings about an osmotic swelling and maintains the high levels of hydration in the cartilage extracellular matrix. Aggrecan also plays a crucial role in chondroskeletal morphogenesis.
Aggrecan is a multimodular molecule composed of globular and non-globular domains. There are three globular domains, G1, G2 and G3 and three non-globular domains, IGD, KS and CS (CS1 and CS2). G1 and G2 are located at the N-terminus of aggrecan and are separated by IGD. G3 domain is located at the extreme C-terminus of aggrecan and KS and CS domains are located between G2 and G3.

Primarily, I wished to study the effects of aggrecan G1 and G2 domains on post-translational modifications and intracellular trafficking of aggrecan products by using whole aggrecan molecules mutated in targeted fashions. In the pursuit of doing that, I generated a number of chicken aggrecan recombinant DNA molecules such as wild type aggrecan, aggrecan lacking the G1 domain, aggrecan lacking the G2 domain and aggrecan carrying a double copy of the G2 domain. I tried to conduct transfection assays using the above-mentioned recombinant constructs in chicken chondrocytes and COS-7 cell systems. However, to my surprise, the full length aggrecan cDNA as well as other mutant aggrecan cDNAs were not translated fully in either chicken chondrocytes or COS-7 cells. My transfection assay, for instance, in the case of the wild type chicken aggrecan gave rise to an aggrecan protein product of about 140 kDa in the lysate and a protein product of about 45 kDa in the medium. Having achieved these results, I decided to change strategy and study the function of aggrecan domains in processing and secretion in chondrocytes and COS-7 cells by utilizing aggrecan recombinants which were relatively smaller than the wild type molecule in order to prevent any possible degradation of the protein products. Chicken chondrocytes were initially my model system to conduct such studies on aggrecan, however, these cells proved to be difficult to be transfected and once transfected they did not produce a sufficient amount of products to be used for any downstream manipulations. Having said
that, I obtained similar results in these studies utilizing chicken chondrocytes and COS-7 cells. COS-7 cell system loomed to be a better model for these studies since these cells are robust enough to be easily transfected with all kinds of recombinant genes. On the other hand, these cells produce an ample amount of recombinant protein products which make these types of studies possible. Another feature of the COS-7 cells which make them suitable for these studies stem from the fact that these cells are SV40 transformed cells of the green African monkey kidney cells. Being already infected with the SV40 virus, these cells carry large T antigens. All of the aggrecan recombinant genes that I have used throughout these studies have been cloned in mammalian expression vectors carrying the SV40 origin of replication which allows these genes to be perpetuated continuously giving rise to a good level of protein expression.

The following studies were undertaken to elucidate how each aggrecan domain (globular and non-globular) and subdomain can affect the way the aggrecan molecule is post-translationally modified and secreted, hereby shedding some light on the function of aggrecan G2 domain.

In the first phase of these studies, we generated a variety of aggrecan recombinant genes in which the aggrecan domains were cloned separately or in combination into mammalian expression vectors. The results of this work indicated that the G3 domain promoted aggrecan secretion and GAG modification, whereas the G1 domain seemed to inhibit product secretion and GAG attachment. G2 domain acted very similar to that of G1 in terms of secretion and GAG attachment. We also noticed that the inhibition of secretion imposed by G1 could be overcome to a certain degree if we used a full length CS2 rather than a truncated one.
In the second phase of these studies, based on the high degree of amino acid similarity between G1 and G2, we concentrated on the identification of the subdomains of aggrecan G1 and G2 which exerted a negative effect on product secretion. In doing so, we generated recombinant genes containing G1 and G2 tandem repeats alone or in combination. The results of this study conclusively demonstrated that the first tandem repeats of aggrecan G1 and G2 shared a common function, which was inhibition of product secretion.

In the final stage of these studies, we set out to identify those amino acids in the first tandem repeats of G1 and G2, which inhibited product secretion. In doing so, we introduced selective C-terminal and N-terminal point mutations and deletion mutations. We noticed that the C-terminal mutations or deletions had little effect on promoting product secretion; however, the point mutations or deletions of the N-terminus of the first tandem repeat led to enhanced secretion. Deletions of nine N-terminal amino acids of TR1 of G1 gave rise to enhanced secretion. When we deleted just four of those nine amino acids, we again noticed product secretion, indicating that those four residues, FHYR, could be important in retention of the product in a subcellular compartment. Furthermore, FHYR→FLYS and FHYR→ILNS mutations in the TR1 of G1 led to product secretion. The above-mentioned point mutations and deletions of those four residues of TR1 in the G1 domain gave rise to the secretion of this domain which otherwise cannot be secreted in a wild type format. However, the N-terminal tagging of a well-secreted gene product with those four residues (FHYR) did not inhibit secretion to a significant degree. We reasoned that these four amino acids may function only in a context of a larger set of amino acids. We subsequently generated some targeted recombinant genes to show how many N-terminal residues of the TR1 of G1 were sufficient
to inhibit secretion. The results of these studies indicated that the N-terminal half of TR1 of G1, a stretch of about 55 amino acids, could negatively affect product secretion.

We utilized a yeast two-hybrid system assay to screen a human brain cDNA library in order to isolate TR1 binding protein partners. We cloned at least two groups of genes (coding for TR1 binding protein partners), a group of DnaJ/HSP40 proteins and a group of ER proteins involved in intracellular calcium homeostasis. The former is a protein chaperone and the function of the latter is as yet unknown. Investigation of the role of the former group of proteins, DnaJ/HSP40 proteins, in aggrecan processing is currently underway. The latter clone was sequenced and the sequence obtained was put into NCBI Blast search analysis. The results of the analysis indicated that this clone cDNA sequence was identical to that of cDNA of 2655 bp encoding calcium homeostasis endoplasmic reticulum protein (GenBank nucleotide access number NM_006387). Cloned from human erythroleukaemia cells, this ER protein is located at the ER membrane where it modulates intracellular calcium mobilization (Laplante et al., 2000). In addition, this protein has ER retention signals at the C-terminus and several potential domains for macromolecular assembly, such as the histidine repeats and serine/arginine-rich region for protein-protein interactions (30). Our clone has an insert of 2 kb. Interestingly, data from our laboratory has also indicated that it is the TR1 that binds to the ER protein. Last, but not least, I have demonstrated that the G3 domain of aggrecan enhances GAG chain attachment and product secretion. To understand how G3 functions, our laboratory has used a G3 domain as bait in a yeast two-hybrid system to screen a human cDNA library. A full length cDNA encoding a protein named CHIP has been isolated. Studies from other laboratories have shown that CHIP negatively regulates the expression of HSP70 and plays an important function in protein processing. The way by
which CHIP regulates aggrecan processing by binding to the G3 domain awaits further investigation.
CHAPTER II

Characterization of the Role of Aggrecan Domains in Post-Translational Modifications and Secretion
2.1 ABSTRACT

Aggrecan is a member of the chondroitin sulfate proteoglycan family which also includes versican/PG-M, neurocan and brevican. Members of this family exhibit structural similarity: a G1 domain at the amino terminus and a G3 domain at the carboxyl terminus with a central sequence for chondroitin sulfate (CS) modification. A unique feature of aggrecan is the insertion of three additional domains, IGD, G2 and KS, between G1 and CS. The G1 and G3 domains have been implicated in product secretion, but G2, though structurally similar to the tandem repeats of G1, performs an unknown function. To define the functions of each aggrecan domain in product processing, we cloned and expressed these domains in various combinations in COS-7 cells. The results indicated that G3 enhanced product secretion, alone or in combination with KS or CS, and promoted glycosaminoglycan (GAG) chain attachment. Constructs containing the G1 domain were not secreted. Addition of a CS sequence to G1 reduced this inhibition, but GAG chain attachment was still reduced. The potential GAG chain attachment site in IGD was occupied by glycosaminoglycans and IGD product was efficiently secreted. KS was modified by GAG chains and secreted. Finally, G2 was expressed but not secreted, and inhibited secretion of the IGD when expressed as an IGDG2 combination.
2.2 INTRODUCTION

Proteoglycans are a family of glycoconjugates with a central core protein to which glycosaminoglycan (GAG) side chain(s) are covalently linked post-translationally. The majority of proteoglycans’ functions are mediated by these GAG chains (Goetinck, 1991; Watanabe et al., 1995), which are polymers of repeated disaccharide units consisting of an uronic acid and a hexosamine. Biosynthesis of all GAGs, except hyaluronic acid, is initiated from a core protein. GAGs are O-linked to serine residues through a trisaccharide linkage sequence at their reducing ends: serine-xylose-galactose-galactose. These GAG chains are acidic molecules and they participate in a wide variety of binding interactions with other matrix macromolecules, cations and water (Toole, 1990; Vilim and Fosang, 1994). They can concentrate secretory products surrounding cells and sequester a variety of extracellular proteins at cell surfaces. In cartilage, the matrix molecules that make up the extracellular matrix (ECM) include proteoglycans, hyaluronic acid, type II collagen, glycoproteins and various mixtures of elastic fibers. Aggrecan is the major structural proteoglycan in cartilage and is responsible for cartilage’s resilience and load-bearing properties. Loss of aggrecan is a major feature of cartilage degradation associated with arthritis (Lohmander et al., 1989a; Lohmander et al., 1993a).

The core protein of aggrecan is composed of three globular domains (G1, G2, G3) with one inter-globular domain (IGD) linking G1 and G2, and two exons for keratan sulfate chain attachment (KS) and for chondroitin sulfate chain attachment (CS) situated between G2 and G3. Attachment of these glycosaminoglycan (GAG) chains occurs on the serine of serine-glycine dipeptide sequence present in this region, and one molecule of aggrecan can contain up to 100 chondroitin sulfate chains, 30 keratan sulfate chains and many O- and N-
linked oligosaccharides. G1 comprises the amino terminus of the core protein. This domain has the same structural motifs as link protein (Deak et al., 1986). G2 is homologous to the tandem repeats found in G1 and link protein. G3, which makes up the carboxyl terminus of the core protein, is composed of alternatively-spliced EGF-like domains, a carbohydrate recognition domain (CRD), a complement binding protein-like (CBP) domain and a short tail (Margolis and Margolis, 1994; Sai et al., 1986). Recent studies indicated that G1 domain is poorly secreted and the G3 domain plays a role in the secretion of recombinant products (Zheng et al., 1998). However, the effects of the other domains on product secretion are not known.

Given that there is up to 62.1% homology between G1 and G2 tandem repeat motifs (Watanabe et al., 1995), we reasoned that G2 might also play a role in product secretion. To test this, we have performed a comprehensive study of the role of the six domains of aggrecan in product processing using recombination and gene expression techniques. Our results indicated that G3’s promotion of product secretion was inhibited by G1 and G2. This inhibition was overcome in the presence of G3 and CS since G3 stimulated GAG chain attachment to CS, and the GAG-modified product is effectively secreted. IGD was modified by GAG and the product was also well secreted.

In a limited parallel study done on another member of the large aggregating chondroitin sulfate proteoglycans, versican, we show that versican G1 domain retard secretion whereas versican G3 domain promotes secretion and GAG chain attachment.
2.3 EXPERIMENTAL PROCEDURES

2.3.1 Materials—RT-PCR mRNA amplification kit, Taq DNA polymerase, T4 DNA ligase and restriction endonucleases were purchased from Boehringer Mannheim. Mammalian expression vectors (pcDNA1, pcDNA3 and pcR3.1), Unidirectional Mammalian TA Cloning Kit and *E. coli* strains MC1061 and TOP10F’ were purchased from Invitrogen. Bacterial growth medium was from Difco. Prep-A-Gene DNA purification kit and prestained protein markers were from Bio-Rad and New England Biolabs (NEB). DNA mini-prep kit was from Bio/Can Scientific. Lipofectin, DMEM growth medium, fetal bovine serum (FBS) and trypsin/EDTA were from GIBCO BRL. ECL Western Blot Detection kit was from Amersham. Horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG was from Sigma. DNA Midi-prep kit was from Qiagen Inc. Tissue culture plates (6-well and 100 mm) were from Nunc Inc. All chemicals were from Sigma.

2.3.2 Strategies for Cloning and Construction of Aggrecan Recombinant Genes—

In order to generate expression constructs containing various regions of mature aggrecan (Fig. 1a-2), we used the following schemes: using chicken cDNA (synthesized from chicken mRNA) as template and two primer sets, IGDNXhol / G2MRSpfl and KSCXbal / G2MFSpfl. Sequences for all primers used in this study are shown in Table I according to the published sequence (Chandrasekaran and Tanzer, 1992; Li et al., 1993). IGD-G2 (the N-terminal half of G2) was amplified using the first primer set, and G2-KS (the C-terminal half of G2) was amplified using the second primer set in PCR. These PCR products were agarose-gel-purified and doubly digested with their respective restriction enzymes. The restricted PCR fragments were further gel purified and ligated into *Xhol / Xbal*-restricted mammalian expression vector pcDNA3 carrying link peptide leading peptide which is recognized by the
monoclonal antibody 4B6. The cDNA of link protein signal peptide was synthesized using LPN\textit{EcoRI} and LPC\textit{BamHI} as primers and the link protein cDNA as template in a PCR reaction. The PCR products were then gel purified and doubly digested and ligated into \textit{EcoRI} / \textit{BamHI}-restricted pcDNA3. The ligation mixture was then used as template to PCR amplify IGDG2KS using the primers: LP\textit{EcoRI} / K\textit{SCXbal} utilizing the Invitrogen Unidirectional TA Cloning Kit (Invitrogen, San Diego). 2 \mu l of the ligation mix was used to transform the TOP10F' cells according to the manufacturer's instruction. The transformation mix was plated on LB plate containing 50 \mu g/ml kanamycin. The resultant construct, IGDG2KS was confirmed using restriction digests and sequencing.

IGDG2KS was then used as template to amplify IGD, G2, KS and IGDG2 fragments using the following primer sets: IG\textit{DNXhoI} / IG\textit{DRXbaI} for IGD amplification, G2\textit{FXhoI} / G2\textit{RXbaI} for G2, KS\textit{NSall} / K\textit{SCXbal} for KS and IG\textit{DNXhoI} / G2\textit{RXbaI} for IGDG2 amplification. The PCR products were gel purified and doubly digested overnight at 37\degree C with their respective enzymes. The digested fragment were further gel purified and used in ligation reactions (16\degree C overnight) with X\textit{hoI} / X\textit{baI}-restricted pcDNA3 harboring a 5' LP60 sequence. In order to increase the cloning efficiency, the ligation reaction mixtures were further used as templates to amplify IGD, G2, KS and IGDG2 using the following primer sets: LP\textit{EcoRI} IG\textit{DRXbaI} for IGD, LP\textit{EcoRI} / G2\textit{RXbaI} for G2, LP\textit{EcoRI} / K\textit{SCXbal} for the amplification of KS and LP\textit{EcoRI} / G2\textit{RXbaI} for IGDG2. The amplified products were cloned using Invitrogen Unidirectional TA Cloning Kit.

To generate the expression construct G1KS, we first PCR-amplified the G1 fragment from a recombinant construct containing G1, which was cloned and expressed by us previously (Cao and Yang, 1999; Cao \textit{et al}., 1998; Yang \textit{et al}., 1998), using the primer set:
LPNEcoRI / AG1CSall. KS was also amplified using the primers: KSNSall / KSCXbal and IGDG2KS as a template. This PCR product was gel-purified and doubly digested overnight with Sall and Xbal. The digested materials were again gel purified and ligated into EcoRI / Xbal-digested pcDNA3. The ligation mix was further used as template to PCR-amplify the G1KS fragment using the primer set: LPNEcoRI KSCXbal and cloned utilizing the Invitrogen Unidirectional TA Cloning Kit. The identity of the resultant recombinant DNA was confirmed by restriction digests and sequencing.

To generate G3KS, we PCR amplified the G3 moiety, which was originally cloned by Stirpe et al. (Stirpe et al., 1987), from the templates G3 cDNA in pcDNA3 using the primers: LPNEcoRI / G3CXhol. KS was also PCR amplified from the IGDKS construct using the primers: KSNSall / KSCSphl. The PCR products were gel purified and doubly digested with Sall and Sphl. The digested products were purified and eventually ligated into EcoRI / Sphl-restricted pcDNA1. The ligation mix was used in the transformation of E. coli strain MC1061 and the transformed cells were plated onto the LB plates containing 50 µg/ml ampicillin and 10 µg/ml tetracycline.

The construction of expression plasmid G1G2 was accomplished as follows: the G1 moiety was PCR amplified from the G1 construct using the primers: LPNEcoRI / AG1CSall and doubly digested with EcoRI and Sall. The G2 moiety was isolated from the Xhol / Xbal-digested G2 construct. These fragments were ligated into the EcoRI / Xbal-digested pcDNA3. The ligation mix was used to PCR amplify the whole fragment, G1G2, using the primer set: LPNEcoRI / G2RXbal. The PCR products were directly cloned into the pCR3.1.

To generate the G1G3 construct, the primers: LPNEcoRI / G1CSall were used to amplify the G1 fragment. G1 was ligated (after double digestion with EcoRI and Sall and gel
purification) into EcoRI/Xhol-digested G3 construct (in pcDNA3). The identity of resultant recombinant G1G3 was further confirmed by restriction digestion and sequencing.

To construct another set of the recombinant genes, CS, G1CS and CSG3, our strategy was to PCR amplify the following DNA fragments: link protein leading sequence, G1, CS and G3. The DNA fragments were then ligated together in various combinations in pcDNA3 and their identities were confirmed by restriction digests and sequencing. Briefly, an EcoRI site and a BamHI site were created at the 5' and 3' end of link protein that was previously cloned. This was done using LPNEcoRI and LPCBamH1 as primers and the link protein cDNA as template in a PCR reaction. The PCR products were gel purified and were digested with EcoRI and BamHI. Using the primer set: CSNBamH1 / CSCXbal and also using the chicken genomic DNA as template. The CS PCR products were purified and digested with BamHI and XbaI. The digested link protein leading sequence and CS fragment were ligated into EcoRI/XbaI-digested pcDNA3. After an overnight incubation of the ligation mix at 16°C, E. coli strain TOP10F' was transformed with the ligation mix.

In order to generate CSG3, we PCR-amplified the G3 domain using the primer set: G3NXhol / G3CXbal and amplified the CS fragment from the CS construct using primers LPNEcoRI / ACSCXhol. The PCR fragments were digested and ligated into EcoRI/XbaI-digested pcDNA3. The ligation mix was used in the transformation of TOP10F' cells.

G1CS construct was generated in the same manner. Using the primer sets LPNEcoRI / AG1CSall and CSNXhol / CSCXbal, we amplified G1 and CS fragments respectively. These PCR products were digested and ligated into EcoRI/XbaI-restricted pcDNA3 followed by transformation of E. coli strain TOP10F'.
All recombinant genes generated here for the current study carry an N-terminal secretion signal, LP60, which is derived from a 60-amino acid peptide of chicken link protein. This short peptide can be recognized by a monoclonal antibody, 4B6, which we have been using routinely in our western blot analyses.

2.3.3 DNA Amplification, Purification, Ligation and Transformation—DNA was amplified in a PCR reaction using pairs of appropriate primers. The reaction mixture (total final volume of 100 µl) contained 200 µM dNTPs, 0.2 µg of each primer, 50 ng template DNA, 2 units of Vent or Taq DNA Polymerase and the Mg-containing buffer (Boehringer Mannheim). The reactions were carried out at 94 °C for 5 minutes for one cycle, 94 °C (60 s), 55 °C (60 s) and 72 °C (60-120s depending on the size of DNA amplified) for 25 cycles and a final extension at 72 °C for 10 minutes.

The DNA products from PCR reactions were purified using a Prep-A-Gene DNA purification kit following manufacturer's instructions. The purified DNA was doubly digested with two appropriate restriction endonucleases, purified and eluted into 40 µl H2O. The DNA was then ligated into the appropriate plasmids (pcDNA1, pcDNA3 or pCR 3.1), which has been linearized with appropriate restriction enzymes. A ligation mixture typically contained 1 µl of ligation buffer, 1 µl of DNA ligase, 3 µl of plasmid vector (50 ng) and 5 µl of insert (150 ng). The ligation reaction was carried out at 14 °C (for pCR 3.1, Invitrogen) or 16 °C (for pcDNA1 and pcDNA3) overnight. Two µl of the ligation mixture were used to transform competent E. coli strain MC1061 (pcDNA1 vector backbone) or TOP 10F' (for pcDNA3 and pCR3.1 vector backbones).

To prepare electro-competent bacteria, E. coli cells were grown in 1 liter of LB medium until the density (O.D.590) reached 0.8. The cells were pelleted for 10 min at 10,000
xg, washed twice with H₂O, resuspended in 4 ml H₂O containing 10% glycerol, aliquoted
(100 μl each) and stored at -70°C. In a transformation reaction, 2- 5μl of DNA ligation mix
were combined with 100 μl of competent bacteria in a 0.2 cm cuvette and electroporated at
2.5 kv volts in a Bio-Rad electroporator. The settings for the gene pulser and capacitance
extender were: capacitance set at 25 μF; capacitance extender set at 960 μF and resistance set
at 200Ω. The mixture was transferred to 0.5 ml SOC medium, agitated at 230 RPM for 45
min at 37°C and spread onto regular LB agar plates containing appropriate antibiotics and
cultured at 37°C overnight.

2.3.4 Expression of Recombinant Constructs in COS Cells—COS-7 cells were
transiently transfected with recombinant constructs using Lipofectin (GIBCO) according to
the manufacturer’s instructions. Briefly, the cultured COS-7 cells were seeded onto a six-well
plate (1.5 x 10⁵ cells/well). The cells were allowed to attach and grow overnight in DMEM
supplemented with 5% FBS. The following day the COS-7 cells reached 70% confluence at
which time the cells were ready for transfection. Lipofectin (2 μl) was incubated with
plasmid DNA (5 μg) for 15 min in 200 μl DMEM followed by addition of 800 μl DMEM.
During the incubation, COS-7 cell culture was rinsed with 2 ml DMEM. The Lipofectin-
DNA mixture was applied to the rinsed cultures and incubated for 10 h. The DNA /
Lipofectin mixture was replaced with 1 ml of DMEM supplemented with 5% FBS. Three
days after transfection, growth medium was collected in a microfuge tube and spun down at
5000 RPM and the supernatant was transferred to a new tube and kept frozen until use. Cell
lysate was prepared by lysing the cells in 1 ml of the lysis buffer (150 mM NaCl, 25 mM
TRIS-HCl pH 8.0, 0.5 M EDTA, 20% Triton X-100, 8 M Urea and 1X protease inhibitor
cocktail). The genomic DNA in the lysate was subsequently removed using unscented
toothpicks. The final lysate product was spun down at 10000 RPM and the supernatant was collected and kept frozen until use. The same protocol for the collection of cell media and lysate has been used throughout all the pertinent studies in following chapters of this writing. For time course analysis of product expression, transfixion of COS-7 cells with G1, G2 or G3 construct was carried out as above. Culture medium and cell lysate were harvested 1, 2 and 3 days after DMEM supplemented with 5% FBS was added to the cultures, and expression of these constructs and secretion of the products were analyzed on Western blot.

**2.3.5 Western Blot Assays**—Cell lysate and culture medium were subjected to SDS-PAGE electrophoresis at equal protein concentrations utilizing Bio-Rad protein kit assay. Proteins separated in SDS-PAGE were transblotted onto a nitrocellulose membrane (Bio-Rad) in 1X TG buffer (Amresco) containing 20% methanol at 60 v for 2 h in a cold room. The membrane was blocked in TBST (10 mM Tris-Cl, pH 8.0, 150 mM NaCl, 0.05% Tween 20) containing 10% non-fat dry milk powder (TBSTM) for 1 h at room temperature, and then incubated at 4°C overnight with monoclonal antibody 4B6 diluted in TBSTM. The membranes were washed with TBST (3 x 30 min) and then incubated for 1 h with HRP-conjugated goat anti-mouse antibody diluted (1:1,000) in TBSTM. After washing as above, the bound antibody was visualized with chemiluminescence (ECL kit, Amersham).

**2.3.6 Chondroitinase ABC Treatment**—Protein A beads (50 μl of gel slurry) were incubated with excess amount of 4B6 antibody at room temperature for 2 h. The unbound antibody was recovered, and the gel beads were washed with 1x PBS extensively for three times. Culture medium from COS-7 cells transfected with mini-aggrecan construct was mixed with 2x PBS buffer in a 1:1 ratio, followed by an incubation with the antibody bound gel beads at 4°C overnight. The gel beads were extensively washed with 1x PBS and
resuspended in 20 µl of 1x PBS. Chondroitinase ABC (0.25 units) was added to digested GAG chains at 37°C for 2 h. The digested product was recovered with 1x protein loading dye, and analyzed on Western blot probed with 4B6 as above.

2.3.7 TCA Precipitations of Proteins

During the course of this study and the subsequent studies in the other chapters included in this writing, we performed protein TCA (trichloroacetic acid) precipitation whenever the results of the western blotting analyses from conditioned media had indicated that expression of a particular recombinant gene was not accompanied by protein secretion. In these cases, we usually took a significant portion of the conditioned medium (~0.5 ml) and TCA precipitated it. The protein precipitate was again subjected to western blotting analysis using the monoclonal antibody, 4B6, to reconfirm the results obtained from the western blot analysis of the unconcentrated conditioned media.

TCA precipitation was accomplished by mixing the condition media with 50%(w/v) TCA to reach a 10% final TCA concentration. The mixture was incubated on ice for 30' followed by 30' spin in a refrigerated centrifuge at about 15000 RPM (rotation per minute). The protein precipitate was washed once with 10%(w/v) TCA followed by another wash with 80% acetone. The pellet was resuspended in 50 µl of 50mM TRIS-HCl, pH 8.0 and then was mixed with 5X SDS-PAGE loading dye for western blot analysis.

Since our studies relied mainly on the secretion of the protein products from our recombinant genes, we presumed a protein to be secreted if our time course studies of the transfections as well as TCA precipitation of the conditioned media indicated the presence of protein products on the western blot analysis consistently.
For the time course studies of protein secretion, COS-7 cells were transiently transfected with the recombinant DNAs of the interest and cells were incubated at the appropriate conditions for tissue culture purposes. Conditioned culture media were sampled (40 µl each) and were mixed with 10 µl of 5X SDS-PAGE protein-loading buffer. The mixture was boiled for 5 minutes and subsequently up to 20 µl of the boiled mixture was run on an appropriate protein gel.
2.4 Results

2.4.1 Expression of Recombinant Constructs—To study the roles of each aggrecan domain in product biosynthesis, GAG chain modification and product secretion, we generated a number of recombinant constructs carrying different domains of aggrecan: G1, G2, G3, IGD, KS, CS, G1-G2, G1-G3, G1-CS, G1-KS, IGD-G2, CS-G3 and G3-KS. The structures of these constructs are shown in detail in Figure 1-2. To allow direct comparison of expression and secretion, each construct contained the same signal peptide, LP60, for product targeting. This leading peptide, originally obtained from link protein, contains an epitope recognized by a monoclonal antibody 4B6 (Binette et al., 1994). The mammalian expression vectors, pcDNA1, pcDNA3 and pCR3.1, are driven by a CMV promoter. These constructs were used in transfection of COS-7 cells using Lipofectin as described under Experimental Procedures. The growth medium and cell lysate were harvested and analyzed on Western blot probed with 4B6 and visualized with a ECL kit according to the manufacturer’s instructions.

2.4.2 G3 enhanced GAG modification of KS and CS and facilitated product secretion—It has been reported that an aggrecan G3 construct is synthesized and secreted into the culture medium while a G1 construct was poorly synthesized and weakly secreted to the culture medium in studies using CHO cells (Yang et al., 1998). Therefore, in our study of the role of this and other aggrecan domains in product biosynthesis, modification by GAG chains and secretion, we first confirmed that an aggrecan G3 construct was well synthesized and secreted to the culture medium when expressed in COS-7 cells (Fig. 2A-2). When the KS construct was expressed in the same way, it was weakly synthesized and secreted to the culture medium, and was also weakly modified by glycosaminoglycan chains (Fig. 2B-2). When the KS domain was linked to the G3 construct, the resulting G3KS construct was
synthesized and strongly modified by GAG chains. A high proportion of products appeared in the culture medium (Fig. 2C-2).

We further tested the role of the G3 domain in GAG modification of the CS sequence. A construct containing the G3 domain and a fragment of the CS domain (nucleotides 3838-5580, equal to 51% of the entire CS domain of aggrecan, Chandrasekaran and Tanzer, 1992; Li et al., 1993), was expressed in COS-7 cells. The CS construct containing the leading peptide and the CS fragment was used as a control. The experiment indicated weak synthesis and secretion of CS construct (Fig. 3A-2), while the CSG3 construct was well synthesized and secreted (Fig. 3B-2). When a G1G3 construct was expressed in the same way, the construct produced a noticeable amount of protein in the cell lysate (Fig. 4-2). Surprisingly, not that much of products was observed in the medium even though we loaded the same amounts of protein from the media as those of lysate, suggesting that the G1 domain is capable of exerting an inhibitory effect on secretion even in the presence of G3 domain.

2.4.3 G1's reduction of GAG modification, and inhibition of product secretion was partially abolished by CS and enhanced by G3—To further test the hypothesis concerning the effect of G1 on secretion, we performed a product secretion assay using a G1 construct. The assay revealed that the G1 products were synthesized but the product was hardly detected in the culture medium (Fig. 5A-2). It is worth mentioning that whenever we came across gene products not secreted into the culture media, we performed a TCA precipitation assay using the entirety of conditioned media followed by western blot analysis as it has been referred to in the “Material and Method Section “of this chapter. We then tested the effect of the G1 domain on GAG modification to KS sequence. COS-7 cells were transiently transfected with the construct G1KS. Cell lysate and culture medium were harvested and
equal amounts of protein from lysate and media analyzed on Western blot probed with 4B6. The presence of G1 inhibited addition of GAG chain to the G1KS core protein, and completely prevented secretion of the G1KS product (Fig. 5B-2).

Further, we tested if G1 could inhibit GAG modification to the CS sequence and inhibit product secretion. To do so, we expressed a G1CS construct in COS-7 cells. We observed that the G1 domain inhibited GAG modification of the CS sequence and partially reduced product secretion as compared with the control construct CS (Fig. 6A-2). G1's inhibitory effect on product secretion was partially overcome by the CS sequence. The G1 domain was then linked to a small fragment of CS sequence (CSD) to obtain G1CSD recombinant gene. G1CSD cDNA was also linked to G3 to obtain G1CSDG3 construct. These constructs were subsequently expressed in COS-7 cells. The core protein of G1CSD was not modified by GAG chains, and the product was not secreted (Fig. 6B-2), while the core protein of G1CSDG3 was modified by GAG chains and the products were secreted (Fig. 6C-2). Treatment of the mini-aggrecan product with chondroitinase ABC resulted in a core protein with a size of 130 kDa.

2.4.4 G2 inhibited product secretion — Knowing that G1 has inhibitory effect on product secretion and G2 exhibits extensive homology with the tandem repeats of G1, we reasoned that G2 might also inhibit product secretion. The G2 construct was expressed in COS-7 cells, and equal amounts of protein from cell lysate and culture medium were analyzed on Western blot probed with 4B6. The product was synthesized, but not observed in the culture medium even after TCA precipitation and western blotting (Fig. 7A-2). A time course study of G2 expression (Fig. 7B-2), with the controls of G1 (Fig. 7C-2) and G3 (Fig. 7D-2), indicated that the products of G1 and G2 accumulated in cell lysate and reached
equilibrium after two days of transfection, while the product of G3 reached equilibrium one
day later. When the G2 domain was linked to the G1 t producing G1G2 recombinant
construct and expressed in COS-7 cells, not to our surprise, no product could be detected in
the culture media, whereas lysate phase of the cells indicated proper gene expression for this
construct. (Fig. 7E-2).

Finally, we investigated the role of the IGD domain in product biosynthesis and
secretion. COS-7 cells were transiently transfected with the construct IGD. Analysis of the
cell lysate and culture medium indicated that not only were the IGD products well produced
and secreted to the culture medium, but also the core proteins were heavily modified by GAG
chains (Fig. 7F-2). A search of the sequence of the IGD domain revealed a unique pair of SG
residues, a potential sugar attachment site, coded by nucleotides 1055-1060. This result
suggests that this site is modified by GAG chain attachment. In light of the proximity of IGD
to G2 (G2 is directly C-terminus to IGD), we chose to investigate the potential effect of the
G2 domain on IGD biosynthesis, GAG chain modification and product secretion. An IGDG2
construct was generated and expressed as described above. Western blot analysis indicated
that the G2 domain completely prevented secretion of IGD (Fig. 7G-2). However, we
anticipated seeing a protein product in the cell lysate of about 68-70 kDa. But to our surprise
the size of the product seemed to be less than projected. In order to address this issue, we first
looked at the sequencing results for this construct (IGDG2) and found that the construct was
correct. We then went on to recover the cDNA from the cell lysate using PCR and various
primer sets to check the presence of different aggrecan domains on the above mentioned
construct. PCR products were of expected sizes. The discrepancy between the anticipated
and observed molecular weights of IGDG2 recombinant product could be explained in two
ways.. The first possibility is the conformational changes that this recombinant protein might have endured. Such conformational and folding changes might be responsible for the concealment of potential glycosylation sites and hence the lack of such modifications. This might explain the difference between the anticipated and observed molecular weights. The second possibility is the degradation of the protein products. Such unexpected degradation might give rise to truncated protein products visualized on our western blot assays. To determine that if we had a full length protein generated for this cDNA (IGDG2), we could have opted to generate another IGDG2 construct harboring an N-terminal LP60 signal peptide and a C-terminal hexahistidine tag. However, we chose not to that since we reasoned that even if there had been a degradation for the IGDG2 products at the C-terminal end of IGD domain, this moiety should have had no difficulty to be exported out since it requires no folding and therefore it is not subjected to RER quality control. It seems that G2 has somehow trapped the molecule in RER and therefore has prevented its secretion. Protein products trapped in RER eventually if not folded properly or glycosylated properly undergo degradation via the ubiquitin-mediated pathway.
Table 1-2. Primers and Restriction Endonuclease Sites for Polymerase Chain Reactions of *aggrecan* constructs

<table>
<thead>
<tr>
<th>primer</th>
<th>sequence</th>
<th>sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1N</td>
<td>5’AAA CTC GAG CTG GGA AGC TCC CTG AAC 3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>G1C</td>
<td>5’AAA CGG GAT CCC ATC ACC ACT GTA ACA GAT GGC ATC GTA 3’</td>
<td>BamH.</td>
</tr>
<tr>
<td>G1C</td>
<td>5’AAA GTC GAC ATC ACC ACT GTA ACA GAT 3’</td>
<td>SalI</td>
</tr>
<tr>
<td>G1C</td>
<td>5’AAA GCA TGC ATC ACC ACT GTA ACA GAT 3’</td>
<td>SphI</td>
</tr>
<tr>
<td>IGDN</td>
<td>5’AAA CTC GAG GAC TTC GAG GCT CTG GTC CCA 3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>IGDR</td>
<td>5’AAA TCT AGA GGG GGA AAT GGG CTG 3’</td>
<td>XbaI</td>
</tr>
<tr>
<td>G2MF</td>
<td>5’AAA GCA TGC GTA GGA GAC AAA GAG AGC 3’</td>
<td>SphI</td>
</tr>
<tr>
<td>G2MR</td>
<td>5’AAA GCA TGC GCT GCG GGG ATT CAC AAT 3’</td>
<td>SphI</td>
</tr>
<tr>
<td>G2F</td>
<td>5’AAA CTC GAG GGT GTG GTG TIC CAC TAC 3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>G2R</td>
<td>5’AAA TCT AGA TCT GAA GCA GAA AGC GTG 3’</td>
<td>XbaI</td>
</tr>
<tr>
<td>KSN</td>
<td>5’AAA GTC GAC GCT CTG CCA TCC GTA GTG 3’</td>
<td>SalI</td>
</tr>
<tr>
<td>KSC</td>
<td>5’AAA GCA TGC TGA GAG CAG CAG TAC GTG 3’</td>
<td>SphI</td>
</tr>
<tr>
<td>KSC</td>
<td>5’AAA TCT AGA TGA GAG CAG TGA TAC GTG 3’</td>
<td>XbaI</td>
</tr>
<tr>
<td>CSN</td>
<td>5’AAA AAA GGA TCC ATC CCG TAT TTC AGC GGA GAC 3’</td>
<td>BamH.</td>
</tr>
<tr>
<td>CSN</td>
<td>5’AAA AAA CTC GAG ATC CCG TAT TTC AGC GGA GAC 3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>CSC</td>
<td>5’AAA GAA TCC AGA AGT GCC AGC AGT GTG GCT 3’</td>
<td>EcoRI</td>
</tr>
<tr>
<td>CSC</td>
<td>5’AAA AAA TCT AGA AGT GCC AGC AGT GTG GCT 3’</td>
<td>XbaI</td>
</tr>
<tr>
<td>CSC</td>
<td>5’AAA AAA CTC GAG AGT GCC AGC AGT GTG GCT 3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>G3N</td>
<td>5’AAA GAA TCC AGT GTG GGT AGT GGT AGT GGT GCC CCG CCT GGT 3’</td>
<td>EcoRI</td>
</tr>
<tr>
<td>G3N</td>
<td>5’AAA CTC GAG AGT GTG GGT GCC CCT GGT GCT CCT GAG 3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>G3C</td>
<td>5’AAA GCA TGC AGT GCC CCC CCG CTG ATG GGT G 3’</td>
<td>SalI</td>
</tr>
<tr>
<td>G3C</td>
<td>5’AAA CTC GAG AGA TCC GTC CAT GCA TGC ATG GGT GGG TCT GTG CAC 3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>G3CHIS</td>
<td>5’AAA TCT AGA GTG ATG GTA ATG GTG ATG GTG GTG GGG TCT GTG CAC 3’</td>
<td>XbaI</td>
</tr>
<tr>
<td>LPN</td>
<td>5’AAA AAA GAA TTC CTA AGT CTA TCT CTT CTG GTG CTG 3’</td>
<td>EcoRI</td>
</tr>
<tr>
<td>LPC</td>
<td>5’ATA CGG GAT CCC GTT GTA AGC TCT GAA ACA GTA 3’</td>
<td>BamH.</td>
</tr>
</tbody>
</table>

Sequences underlined represent restriction enzyme sites denoted in the primer nomenclatures.
Figure 1-2. The structures of recombinant DNA constructs containing different domains of aggrecan. The G1 domain includes nucleotides 144-1038 of mature aggrecan; the IGD domain, nucleotides 1041-1554; the G2 domain, nucleotides 1557-2145; the KS domain, nucleotides 2154-2409; the truncated CS domain, nucleotides 3795-5226 and the G3 domain, nucleotides 5679-6327. IgG, immunoglobulin-like motif; TR, tandem repeat; G1, globular domain 1; G2, globular domain 2; G3, globular domain 3; IGD, inter-globular domain; KS, sequence modified by keratan sulfate chains; CS, sequence attached by chondroitin sulfate chains; CRD, carbohydrate recognition domain; CBP, complement binding protein. The leading peptide added to all constructs was obtained from link protein (nucleotides 1-180). Numbers above schematic correspond to nucleotides in the sequence of full-length aggrecan.
Figure 2-2. G3 enhances KS synthesis and secretion.

COS-7 cells were transiently transfected with the recombinant constructs G3 (A), KS (B) and G3KS (C). The cultures were maintained in DMEM supplemented with 5% FBS at 37°C in an incubator for 3 days. Cell lysate and culture media were subjected to SDS-PAGE (1, 2, 5 and 10 μl per well) in a 10% gel. The separated proteins were then transblotted onto a nitrocellulose membrane and probed with the monoclonal antibody 4B6. Products of G3 (~48 kDa) were synthesized and secreted to the culture media. Products of KS (~30-35 kDa) were synthesized, secreted to the culture media and modified by glycosaminoglycan resulting in smear bands in the gel. Products of G3KS (~50-70 kDa) were heavily modified by glycosaminoglycans.
Figure 3-2. G3 enhances CS synthesis and secretion.

COS-7 cells were transiently transfected with the recombinant constructs CS and CSG3. Cell lysate (A) and culture media (B) were analyzed on Western blot (2, 5, 10 and 20 µl per well) probed with 4B6. Products of CSG3 (~180 kDa) were synthesized and secreted to the culture media (B), while the products of CS (~90 kDa) were expressed, but weakly secreted (B). The products of CSG3 were heavily modified by GAG chains resulting in strong smearing bands on the blot (~180-200 kDa), while the products of CS were weakly modified by GAG chains.
Figure 4-2. G1 inhibits G3 secretion.

Cells were transiently transfected with the recombinant constructs G1G3. Cell lysate and culture media were analyzed on Western blot (1, 2, 5 and 10 μl per well) probed with 4B6. The products (~72 kDa) were synthesized but poorly secreted to the culture medium.
### Figure 5-2. G1 inhibits KS secretion.

Cells were transiently transfected with the G1 and G1KS constructs. Cell lysate (L1, 1 μl) and culture media (M10, 10 μl) were analyzed on Western blot probed with 4B6. Products of the G1 construct (55 kDa) were synthesized but not secreted to the culture media (A). G1KS construct was expressed (detected in cell lysate) but poorly modified by GAG chains and could not be detected in the culture medium even after TCA precipitation of the conditioned medium (B).
Figure 6-2. G1 inhibition of secretion is overcome by CS and enhanced by G3.

Cells were transiently transfected with constructs G1CS (A), G1CSD (B) and mini-aggrecan (C, or G1CSDG3). Cell lysate and culture media were analyzed on Western blot probed with 4B6. For comparison, the lysate (20 µl) and culture medium (20 µl) of CS products were loaded as controls in Panel A. The products of G1CS (~180-200 kDa) were well-synthesized, weakly modified by GAG chains and weakly secreted to the culture media as compared with the products of CS. The products of G1CSD were synthesized but hardly detected in culture medium. The products of mini-aggrecan (G1CSDG3) were well synthesized, modified by GAG chains and secreted to the medium. The product of mini-aggrecan in culture medium was also purified using 4B6 and treated with chondroitinase ABC, and the product with (ABC) or without (control) chondroitinase ABC treatment was analyzed on Western blot. The core protein of the mini-aggrecan was detected with a size of 130 kDa.
Figure 7-2. The effects of G2 domain on GAG chain attachment and product secretion.

Cells were transiently transfected with the G2, G1, G3, G1G2, IGD and IGDG2 as above. Cell lysate (L1 and L10, 1 and 10 μl) and culture medium (M10, 10 μl) were analyzed on Western blot and probed with 4B6. Products of G2 (~22 kDa) were synthesized, and were not secreted to the culture media (A). The time course expression of G2 was also analyzed (B), while expression of G1 (C) and G3 (D) served as controls for the time course study (10 μl). Products of G1G2 were synthesized but were not secreted to the culture medium (E). Products of IGD were synthesized, modified by GAG chains resulting in smear bands in the gel and secreted to the culture media. The sizes of protein in the lysate were approximately 40-45 kDa while they were approximately 45-65 kDa in the culture medium (F). The products of IGDG2 were synthesized, but they were not secreted (G).
2.5 DISCUSSION

Our work reported here represents the first comprehensive study of the role of each aggregcan domain in product processing. We demonstrated that the G1 domain reduces GAG modification to the core proteins and G1 and G2 domains are able to inhibit product secretion. There are also two components which stimulate product secretion: the G3 domain and the GAG chain-modified sequences (CS, KS and IGD). The G3 domain not only stimulates product secretion, but also promotes the attachment of GAG chains. The KS and CS sequences are poorly modified by GAG chains in the absence of the G3 domain in the same construct. The G3 domain, however, is not able to sufficiently stimulate the secretion of G1. For instance, the products of G1G3 were hardly detected in the culture medium. Only in the presence of a fragment of CS sequence (e.g., the CSD fragment in the G1CSDG3 construct), the G3 domain facilitates GAG chain attachment to the CS fragment resulting in the presence of two positive components for product secretion, can the inhibitory effects of G1 be abolished. The fact that the endogenous aggregcan, which contains two components (G1 and G2 domains) for inhibition of product secretion and two components (G3 domain and modified GAG chains) for product secretion, is secreted to the ECM implies that the G3 domain and CS fragment, by working together, play a predominant role in product secretion. In our experience, we never observed secretion of the core proteins of mini-aggregcan, mini-versican, aggregcan G1CS construct and versican G1CS construct. These results indicate that GAG chain attachment is imperative for secretion of these proteoglycans. Further support came from our result that the inhibition of product secretion by G1 was overcome by a larger CS sequence (G1CS) but not by a smaller CS sequence (G1CSD). It has been reported that
the enhancement by G3 domain of product secretion occurs through its interaction with HSP25 (Luo et al., 1996). Each protein module also enhances product secretion (Day et al., 1999; Domowicz et al., 2000). The mechanism of our observation that GAG chain modification stimulates product secretion awaits future investigation.

An interesting question then arises: why are chondrocytes programmed to produce two components for inhibition of product secretion? Our finding that G3, without the presence of a CS fragment, was not able to stimulate G1 secretion in the construct G1G3, and that the G1CS core protein was weakly modified by GAG chains and the GAG modified products were secreted, seems to indicate that the cells are programmed to produce “perfect” aggrecan molecules. Aggrecan core protein cannot be mistakenly secreted to the matrix following mRNA translation as the G1 and G2 domains will inhibit the secretion of aggrecan core protein. Only after the aggrecan core protein is correctly modified by GAG chains could the proteoglycan be secreted. This programming is very important for cartilage development and maintenance since the GAG chains play important roles in retaining H2O and negative charges and producing a load-bearing cartilage.

The presence of the G2 domain may be essential since cells have to completely block secretion of any imperfect aggrecan. Indeed, without the G2 domain, the G1CS core protein was weakly modified and the product was weakly secreted. One may imagine that the products of such under modification by GAG chains would be harmful in cartilage development and maintenance. This hypothesis is not only supported by our experimental results but also supported by literature on the other members of this chondroitin sulfate proteoglycan family, which indicated that these proteoglycans are essential for tissue development and growth (Kerr and Newgreen, 1997; Mjaatvedt et al., 1998; Vertel et al.,
Aggrecan contains a G2 domain, and its core protein is heavily modified by GAG chains, while other family members (i.e., versican, neurocan and brevican) do not contain a G2 domain and their core proteins are not heavily modified by GAG chains. This should not lead to an incorrect assumption that G2 enhances GAG chain modification. It is the G3 domain, but not the G2 domain, that promotes GAG chain attachment. The G2 domain not only inhibits product secretion but also reduces GAG chain modification. Its effect on product secretion seems to be predominant in order to inhibit secretion of insufficiently modified aggrecan. As a result, only those molecules sufficiently modified by GAG chains are secreted to the ECM. The G3 domain, on the other hand, enhances GAG chain attachment and stimulate product secretion. The importance of the G3 domain is seen in a fatal genetic disease of the chicken (nanomelia). Nanomelic aggrecan contains a premature stop codon at the 3' end of the CS sequence (5' to G3 domain), and the mutant aggrecan core protein is neither modified by GAG chains nor secreted to the ECM (Primorac et al., 1994; Vertel et al., 1994). Another fatal genetic disease to demonstrate the importance of the CS sequence and G3 domain is the cartilage deficiency in mice whose aggrecan contains a 7-bp deletion in exon 5 (TR1 motif of G1 domain). As a result, the truncated aggrecan can no longer be secreted to the matrix (Wai et al., 1998; Watanabe et al., 1994).

It has long been known that the G2 domain is structurally similar to the tandem repeats of the G1 domain. While the G1 domain binds to link protein via the IgG-like motif and binds hyaluronan via the tandem repeats, forming ternary complexes to maintain a stable matrix network in cartilage, the G2 domain does not bind to hyaluronan (Watanabe et al., 1997) and so far, no molecule has been found to interact with G2. The function of the G2
domain remains unknown. Our report on the role of the G2 domain in aggrecan processing has, for the first time, shed some light on the biology of this domain.

Another novel finding in our studies is the modification of the IGD domain by GAG chain. There is only one potential site on the IGD domain for GAG chain attachment. Our result that the IGD construct is synthesized, modified by GAG chain and the product was secreted to the culture medium implies that the potential site is actually modified by GAG chain in the COS-7 cell line. This finding can only be achieved by using small recombinant construct. In a large construct, which may contain many sites for GAG chain attachment, one cannot distinguish the presence or absence of one particular GAG chain. It is not yet known whether this site is modified by GAG chain in endogenous aggrecan.

Thus, given the diversity of effects exerted by aggrecan domains on product processing, it is likely that each domain is involved in the processing and secretion of the mature aggrecan product. In the model we propose, the G1 and G2 domains might have a stalling effect on products carrying these two domains keeping the proteins in the RER longer than the proteins which lack these two domains. By stalling the protein products in the RER, G1 and G2 may ensure that GAG chain initiation has properly completed. However, G1 and G2 do not seem to enhance GAG chain attachment since this role is fulfilled by the action of the G3 domain. G3 domain promotes intracellular trafficking by taking aggrecan from RER to the Golgi where the majority of GAG chain attachment is to occur. Hence G3 (in the construct G1CSDG3) is able to overcome the inhibition of product secretion imposed by G1 (in the construct G1CSD), but keep in mind that GAG chain attachment seems to promote secretion (by providing proper folding and solubility) since the product of G1G3 recombinant gene is neither glycosylated nor secreted. So, it can be concluded that G3,
through promotion of GAG chain attachment, overcomes the inhibitory effect of G1 and G2 on product secretion. The attachment of GAG chains to IGD, KS and CS domains which allows aggrecan to generate stable matrix meshwork, is also a prerequisite for aggrecan secretion.

It is interesting to know that the G1 and G3 domains of versican, another member of the large aggregating chondroitin sulfate proteoglycan family, behave in the same way as to those of aggrecan in terms of secretion and GAG chain attachment with versican G1 domain reducing secretion and the versican G3 domain promoting secretion and GAG chain attachment.

The last salient point which should be mentioned at this point regarding our studies (in this chapter and the next chapter) with these recombinant aggrecan genes is the discrepancies over the anticipated and observed molecular weights of the recombinant gene products. Throughout our studies with aggrecan cDNAs, we noticed that the molecular weights of protein products of recombinant genes having multiple domains did not equal the sums of the molecular weights of the single domains or subdomains used in their constructions. The molecular weights of the recombinant products with multiple domains tended to be less than those of the sums of single domains. For example, recombinant IGD and G2 protein products appeared to be about 43 and 25 kDa respectively, however, recombinant IGDG2 appeared to be around 43 kDa. In the case of IGDG2, one might expect to see a protein product of about 75 kDa. However, this was not the case on our western blot analysis. IGDG2 on our western assays appeared to be around 43 kDa. One can explain this discrepancy by proposing two scenarios. In the first scenario, one might expect some sorts of conformational change giving rise to the generation of smaller molecule. The second scenario puts forth the concept of intracellular degradation, which might account for the
production of smaller molecules. We noticed the same phenomenon (in terms of
discrepancies between the predicted and observed molecular weights of protein products) in
our studies with G3, KS and G3KS as well as G1, G3 and G1G3. In the case of G3, KS and
G3KS, the molecular weights of G3 and KS are about 50 and 32 kDa respectively. Therefore,
we anticipated observing G3KS protein to be around 80 kDa. The observed molecular weight
for G3KS was about 60 kDa, which is less than the projected value of 80 kDa. This
difference can be due to changes in the conformational or folding properties of the
recombinant protein with multiple domains. To elaborate more on this possibility, we have to
remember that the predicted sizes of G3 and KS core proteins are about 26 and 8 kDa,
however, G3 and KS protein products run at 50 and 32 kDa on our western blots. The
predicted molecular weight of G3KS protein products (based on the sums of the observed G3
and KS molecular weights) is about 80 kDa, however, the observed value for G3KS is 60
kDa. The discrepancy in molecular weight can be due to some conformational changes in the
way that G3KS protein folds. This folding might bury some potential glycosylation sites in
this molecule and hence give rise to a protein product, which appears to be smaller. The same
sort of explanation can be applied in the case of G1, G3 and G1G3 protein products. G1 and
G3 protein products on our western blots appear at about 55 and 50 kDa respectively. One
can expect a molecular weight of about 100 kDa for G1G3. However, this is not the case.
G1G3 protein products run at about 72 kDa on our western blots. The main reason for this
difference can be attributed once again to conformational changes that the G1G3 protein
products have endured. These conformational changes might hinder some post-translational
modifications of this recombinant protein and make it smaller than predicted.
CHAPTER III

The First Tandem Repeat in Aggrecan G1 Domain Can Inhibit Product Processing and Interact with Calcium Homeostasis Endoplasmic Reticulum Protein
3.1 ABSTRACT

Members of the large aggregating chondroitin sulfate proteoglycans are characterized by an N-terminal fragment known as G1 domain, which is composed of an immunoglobulin (IgG)-like motif and two tandem repeats (TR). Previous studies have indicated that the expressed product of aggregan G1 domain was not secreted. Here we demonstrated that the inability of G1 secretion was associated with the tandem repeats but not the IgG-like motif, and specifically with TR1 of aggregan. We also demonstrated that the G2 domain, a domain unique to aggregan, had a similar effect on product secretion. In a yeast two-hybrid assay, TR1 interacted with the calcium homeostasis endoplasmic reticulum protein. Product analysis indicated that the G1 domain while retarding product secretion, enhanced GAG modification to the miniaggrecan. The sequence of TR1 of G1 is highly conserved across species, which suggested that they play similar functions. Mutagenesis experiments indicated that the N-terminal fragment of the first tandem repeat, in particular, the amino acids H2R1 of this motif were important to its effect on inhibition of product secretion. However, these four amino acids seemed to function in a larger stretch of amino acid residues. I demonstrated that the N-terminal 55 amino acids were required to exert this function. Thus, this study reports for the first time a novel function for the tandem repeats of aggregan, reveals a likely candidate partner for this effects, and sheds light on a possible molecular mechanism.
3.2 INTRODUCTION

The large aggregating chondroitin sulfate proteoglycans include aggrecan, versican (also known as PG-M), neurocan and brevican (Li et al., 1993; Rauch et al., 1992; Shinomura et al., 1993; Yamada et al., 1994; Zimmermann and Ruoslahti, 1989). Members of this family share many structural features (Barry et al., 1994; Margolis and Margolis, 1994). These proteoglycans contain a central core protein to which glycosaminoglycan (GAG) chains are covalently linked post-translationally. The core protein of each molecule contains a signal peptide, an amino-terminal globular domain known as G1 domain, a carboxyl-terminal globular domain known as the G3 domain (or selectin-like domain) and a large sequence situated between G1 and G3 for modification by GAG chains, mainly chondroitin sulfate (CS) chains (Walcz et al., 1994; Watanabe et al., 1995). The G1 domain is composed of one IgG-like motif and two proteoglycan tandem repeats. This domain has the same structural motif as link protein (Barry et al., 1992a; Deak et al., 1986). G3 is composed of alternatively spliced epidermal growth factor (EGF)-like motif(s), a carbohydrate recognition domain (CRD), a complement binding protein-like (CBP) motif and a short tail.

Aggrecan, in addition to these features, has some extra motif structures (Chandrasekaran and Tanzer, 1992; Li et al., 1993). It contains a G2 domain, which is structurally similar to the two tandem repeats of the G1 domain and link protein. An inter-globular domain (IGD) is situated between the G1 and G2 domains. Situated between G2 and G3 is a large extended region for GAG chain attachment, which occurs on the serine of serine-glycine dipeptides. This extended region is further divided into two domains: one for attachment of keratan sulfate (KS) and one for attachment of chondroitin sulfate (CS) chains.
(Barry et al., 1994; Li et al., 1993). Compared to other members of this family, aggrecan contains significantly more potential sites for GAG chain attachment, and thus, its molecular mass is much greater. These GAG chains account in large part for the unique properties of aggrecan in cartilage development (Goetinck, 1991; Vertel et al., 1993b; Watanabe et al., 1994). The importance of GAG modification in cartilage development was initially observed in the chicken mutant nanomelia, a lethal genetic disease. In this disease, the chicken aggrecan gene contains a point mutation resulting in a premature stop codon on the N-terminal side of the G3 domain (Li et al., 1993). The truncated aggrecan core protein is neither modified by GAG chains nor secreted to the cartilage matrix (Li et al., 1993; Vertel et al., 1994; Vertel et al., 1993b). These studies suggested that aggrecan G3 domain was involved in GAG modification and product secretion (Luo et al., 1996). Since then, the role of the G3 domain in product processing has been extensively studied. It has been reported that the G3 domain enhances GAG modification and product secretion (Day et al., 1999; Domowicz et al., 2000; Kiani et al., 2001; Luo et al., 1996; Zheng et al., 1998), and that the G1 domain product accumulated in the cytoplasm and is not secreted (Kiani et al., 2001; Luo et al., 1996). Recently, we also demonstrated that a construct containing the G2 domain was not secreted (Kiani et al., 2001). However, the way by which G1, and G2, affects product processing is not clear. It is likely that the G1 domain interacts with unknown cellular components, thus inhibiting product secretion. This study was designed to characterize in detail how the G1 and G2 domains affect product secretion, and to investigate the biological function of this effect on product processing.
3.3 EXPERIMENTAL PROCEDURES

3.3.1 Materials—Taq DNA polymerase, T4 DNA ligase and restriction endonucleases were purchased from Boehringer Mannheim and New England Biolabs (NEB). Mammalian expression vectors (pcDNA1, pcDNA3 and pcR3.1), Unidirectional Mammalian TA Cloning Kit and Escherichia coli strains MC1061 and TOP10F' were purchased from Invitrogen. Bacterial growth medium was from Difco. Prep-A-Gene DNA purification kits were from Bio-Rad, and prestained protein markers were from NEB. DNA mini-prep kit was from Bio/Can Scientific. Lipofectin, DMEM growth medium, fetal bovine serum (FBS) and trypsin/EDTA were from GIBCO BRL. ECL Western Blot Detection kit was from Amersham. Horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG was from Sigma. DNA Midi-prep kit was from Qiagen Inc. Tissue culture plates (12-well, 6-well and 100 mm) were from Nunc Inc. Human cDNA library and supplies for yeast two-hybrid assays were from Clontech. All chemicals were from Sigma.

3.3.2 Strategies for Gene Construction — In this study, a total of 42 constructs were used: 38 constructs were used for gene expression and the rest 4 constructs were used for yeast two-hybrid binding assays. Four constructs — G1, G3, link protein and a mini-aggrecan (or G1CSG3) — have been described by us previously (Cao and Yang, 1999; Cao et al., 1998; Kiani et al., 2001; Yang et al., 1998; for structures of aggrecan and the G1 domain see Fig. 1A-3 and Fig. 1 B-3). A chimeric construct LPCSG3 was generated by replacing the G1 domain in the G1CSG3 construct with link protein. Production of the remaining 33 constructs is described in detail below. All new constructs generated by PCR were verified by sequencing.
3.3.2.1 Cloning of $\text{TR}^{G1}$, $\text{TR1}^{G1}$, $\text{TR2}^{G1}$ and IgG in pcDNA1 — Aggrecan G1 subdomains were PCR-amplified and cloned in the vector pcDNA1 via the following cloning strategy: using the ABN/G1C primer set (sequences for all primers used in this study are shown in Table I) and G1 cDNA as template, we PCR-amplified aggrecan G1 tandem repeats ($\text{TR}^{G1}$ or $\text{TR1}^{G1}\text{TR2}^{G1}$). $\text{TR1}^{G1}$ and $\text{TR2}^{G1}$ were also PCR-amplified individually using G1 cDNA as the template and two primer sets: ABN/ABC and AB'N/G1C.

The reaction mixture (total volume of 100 µl) typically contained 200 µM dNTPs, 0.2 µg of each primer, 50 ng template DNA, 2 units of Taq DNA polymerase and Mg$^{2+}$-containing buffer. The reactions were carried out at 94°C for 5 min for one cycle, 94°C (60 s), 55°C (60 s) and 72°C (60-120s depending on the size of DNA amplified) for 25 cycles and a final extension at 72°C for 10 min. The protocol was standard for all reactions. Products were agarose gel-purified and digested with two appropriate restriction endonucleases at 37°C overnight. The digested PCR products were purified and ligated into a plasmid vector. Upstream (5') to the inserts, all constructs contain the link protein leading peptide sequence (LP60), which harbors the signal sequence and an epitope recognized by a monoclonal antibody 4B6 (Binette et al., 1994).

Similarly, the IgG-like motif of aggrecan G1 was also PCR-amplified using the primers AIGN/AAC and G1 cDNA as a template. The PCR products were gel-purified and digested with BamHI/SphI at 37°C overnight. The digested PCR products were purified from agarose gel and ligated into a BamHI/SphI-digested pcDNA1 vector carrying the LP60 signal sequence 5' to the insert. pcDNA1 constructs were used to transform E. coli strain MC1061.

3.3.2.2 Cloning of $\text{TR1}^{G2}\text{TR2}^{G2}$ (or G2), $\text{TR1}^{G2}$ and $\text{TR2}^{G2}$ in pcR3.1 (TA Cloning System) — Aggrecan G2 domain and its subdomains were cloned as follows: using the
primer sets G2F/G2R, G2F/G2TR1C and G2TR2N/G2R, we PCR amplified aggrecan G2 (TR1G2-TR2G2), TR1G2 and TR2G2 using aggrecan G2 domain as the template. All PCR fragments were gel-purified and digested with their respective restriction enzymes. The digested PCR products were purified and ligated into an XhoI/XbaI-restricted pcDNA3 vector harboring LP60 5' to the inserts. The ligation mixes were used as templates in three separate PCR reactions using the primer sets LPNKozak/G2R (for the constructs G2 and TR2G2) and LPNKozak/G2TR1C (for the construct TR1G2). We were then able to PCR amplify G2, TR2G2 and TR1G2 to be used in the Unidirectional TA Cloning Kit. These fragments were ligated into the vector pCR3.1 at 14°C overnight. The ligation mixes were used in transformation of E. coli strain TOP10F'. The resultant recombinant constructs (G2, TR2G2 and TR1G2 in pCR3.1) were digested with EcoRI to confirm the proper orientation of the inserts in the vector.

3.3.2.3 Targeted Cloning of Different Aggrecan G1 Subdomains with Each Other and with the G3 Domain — Using the primer sets G1N/ABC and AB'N/G3C, we PCR amplified IgGTR1G1 and TR2G1G3 using G1 cDNA and G2G3 cDNA (already cloned by our group in the vector pCR3.1) as templates. The PCR products were gel-purified, digested with XhoI/SphI, re-purified and ligated into XhoI/SphI-restricted pcDNA1. 3' of the LP60 sequence.

We also cloned TR1G1G3 and IgGTR2G1 in pcDNA3 and pCR3.1, respectively. Using the primer set LPNKozak/G1TR1C and G1 cDNA as a template, we PCR amplified the TR1G1 fragment. The PCR fragments were gel-purified and digested with the appropriate restriction enzymes. The digested PCR products were purified and ligated into a HindIII/EcoRI-restricted pcDNA3 harboring the aggrecan G3 domain 3' to the ligated PCR
In order to generate IgGTR2G1 construct, we first PCR-amplified TR2G1 using the primer set G1TR2N/G1C and G1 cDNA as the template. The PCR products were gel-purified and digested with Sphl/XbaI. IgG fragment was obtained from the IgG-pcDNA1 construct (described earlier) by digestion with EcoRI/Sphl, and these two fragments, IgG (EcoRI/Sphl) and TR2G1 (Sphl/XbaI), were ligated into pcDNA3 vector digested with EcoRI/XbaI. Ligation was performed at 16°C overnight. The ligation mix was then used in a PCR reaction to amplify IgGTR2G1 using the primer set LPNKozak/G1C. The PCR products were cloned into pCR3.1 utilizing the Unidirectional TA Cloning Kit. The resultant construct was digested with EcoRI to confirm the proper orientation of the inserted PCR product.

### 3.3.2.4 Linkage of G2 Motifs with IgG Motif or the G3 Domain

To generate the construct IgGTR1G2, the LP60-IgG fragment was obtained from the construct G1N4mu by digestion with EcoRI/XhoI, while the TR1G2 fragment was obtained from the construct TR1G2 by digestion with XhoI/XbaI. These two fragments were inserted into EcoRI/XbaI-digested pcDNA3 vector. To generate the construct IgGTR2G2, the LP60-IgG fragment was obtained as above, and the TR2G2 fragment was obtained from the TR2G2 construct by digestion with XhoI/XbaI. These two fragments were cloned as above producing IgGTR2G2. To construct TR1G2G3, a LP60-TR1G2 fragment was obtained from the TR1G2 construct by digestion with EcoRI/XbaI, and the G3 fragment was synthesized in a PCR using two primers, G3Nxbal and G3CAPal, followed by digestion with XbaI/Apal. These two fragments were cloned into the pcDNA3 vector, which had been digested with EcoRI/Apal.
To generate construct TR2G2G3, the TR2G2G3 fragment was amplified by PCR using two primers, G2TR2NXhol and G3CApal, with the G2G3 construct as a template. The PCR products were digested and cloned into XhoI/Apal-digested pcDNA3 vector harboring the LP60 moiety.

3.3.2.5 Site-Directed Mutagenesis and Truncation—Generation of these types of constructs is more complex, and it is described in the figure legends (Fig. 8-3 and Fig. 9-3). These constructs include TR \(^{G1}K\rightarrow Q\), TR \(^{G1}mu1\), TR \(^{G1}mu2\), TR \(^{G1}mu3\), TR \(^{G1}C15del\), TR \(^{G1}N9del\), TR \(^{G1}Δ381\), G1N4del, G1N4mu, TR \(^{G1}N4del\), TR \(^{G1}N4mu\), and TR \(^{G1}HRmu\).

3.3.2.6 Addition of N-Terminal Fragments of TR \(^{G1}\) to IgG Motif—Constructs IgG10aa, IgG20aa, IgG35aa and IgG45aa were generated by PCR using the construct IgGTR1\(^{G1}\) as a template. Primers used were LPNKozak and one of the following: TR110aaR, TR120aaR, TR135aaR or TR145aaR. The construct IgG45aamu (containing a point mutation at amino acid C55 of TR1\(^{G1}\)) was PCR-generated using the same template and the primers LPNKozak and TR145aaRmu. As the primer LPNKozak contains an EcoRI site and the other primers contain an Xbal site, the PCR products were digested with EcoRI/Xbal and cloned into EcoRI/Xbal-digested pcDNA3 vector. Due to the presence of a KpnI site at amino acids 55-56 of TR1 motif, the construct IgG55aa was generated by releasing this fragment from TR1 construct by EcoRI/KpnI digestion.

3.3.2.7 Generation of Yeast Two-Hybrid Binding Constructs—One or both tandem repeats of the G1 domain were inserted into the plasmids pGBK7 or pGBDC1 to generate four constructs for yeast two-hybrid studies: pGBK7-TR1TR2, pGBDC1-TR1TR2, pGBDC1-TR1, and pGBDC1-TR2. Each bait fragment was amplified by PCR using two primers: TRIN and AG1C amplifying TR1TR2; TRIN and TR1C amplifying TR1; and
TR2N and AG1C amplifying TR2. All fragments were inserted into the cloning sites BamHI and SalI, producing four constructs named above.

**3.3.2.8 Expression of Recombinant Constructs in COS-7 Cells**—COS-7 cells were transiently transfected with the recombinant constructs using Lipofectin. Briefly, COS-7 cells were seeded to a six-well plate at $1.5 \times 10^5$ cells per well. The cells were allowed to attach and grow overnight in DMEM supplemented with 5% FBS and 100 units penicillin/100μg streptomycin/2.5μg amphotericin (per ml). Cells were transfected at 70% confluence. Lipofectin (1 μl) was incubated with plasmid DNA (5 μg) for 15 min in 200 μl DMEM followed by addition of 800 μl DMEM. During the incubation, COS-7 cell cultures were rinsed with 2 ml DMEM. The Lipofectin-DNA mixture was applied to the rinsed cultures and incubated for 10 h. The DNA/Lipofectin mixture was replaced with 1 ml of DMEM supplemented with 5% FBS. The growth medium and cell lysate were harvested separately after 3 days.

In co-expression studies, COS-7 cells prepared as above were co-transfected with the mini-aggreccan and TR1G1, TR2G1 or the control vector. Briefly, plasmid DNA (2 μg) containing mini-aggreccan or LPCSG3 was mixed with the plasmid DNA (3 μg) containing TR1G1, TR2G1 or the control vector, followed by the incubation and transfection as described above.

**3.3.2.9 Western Blot Assays**—Cell lysate and culture medium were subjected to SDS-PAGE electrophoresis as described earlier. Proteins separated in SDS-PAGE were transblotted onto a nitrocellulose membrane (Bio-Rad) in 1x TG buffer (Amresco) containing 20% methanol at 60 volt for 2 h in a cold room. The membrane was blocked in TBST (10 mM Tris-Cl, pH 8.0, 150 mM NaCl, 0.05% Tween 20) containing 10% non-fat dry milk
powder (TBSTM) for 1 h at room temperature, and then incubated at 4°C overnight with monoclonal antibody 4B6 diluted in TBSTM. The membranes were washed with TBST (3 x 30 min) and then incubated for 1 h with HRP-conjugated goat anti-mouse antibody (diluted 1:50,000) in TBSTM. After washing as above, the bound antibodies were visualized with chemiluminescence (ECL kit).

3.3.2.10 Immunoprecipitation — Protein G beads (25 μl of gel slurry equal to 10 μl of gel bed) were incubated with excess 4B6 antibody in a buffer containing 20 mM Tris-HCl, pH 8.0, and 150 mM NaCl (THS buffer) at room temperature for 2 h. The unbound antibody was removed, and the gel beads were washed three times with THS buffer. Culture medium or cell lysate from COS-7 cells transfected with different constructs was mixed with THS buffer in a 1:1 ratio, and added at saturating amounts for antibody binding to the antibody-bound gel beads, followed by incubation at 4°C overnight. The gel beads were extensively washed with THS and resuspended in 100 μl THS. Loading dye was added to the purified gel beads, followed by incubation at 100°C for 5 min to release bound proteins for Western blot analysis.

3.3.2.11 Chondroitinase ABC Treatment — The gel beads containing purified proteins indicated above were incubated with chondroitinase ABC (0.2 units) at 37°C for at least 3 h to cleave GAG chains from the products. The digested product was recovered with protein-loading dye and analyzed on Western blot probed with 4B6 as above.

3.3.2.12 Yeast Two-Hybrid Assay — In order to isolate proteins which bind to the tandem repeats, we employed Clontech Gal4 yeast two-hybrid system with chicken aggrecan TR1TR2G1 as a bait. Yeast strain JP-69-4A was co-transfected with pGBKTK7-TR1TR2 and human brain cDNA library generated in the plasmid pACT2. Library screening and colony
identification were performed according to the manufacturer's instructions. Sequencing of positive clones revealed a number of genes which may interact with pGBK7-TR1TR2. One of them was calcium homeostasis endoplasmic reticulum protein (ER protein), named pACT2-ER.

The tandem repeats were cloned into plasmid pGBDC1, producing pGBDC1-TR1, pGBDC1-TR2, pGBDC1-TR1TR2, which were then co-expressed with pACT2-ER in JP-69-4A. The transfected yeast cells were grown on the SD/-Trp/-Leu agar plates at 30°C for 4-5 days. Ten yeast colonies were picked up from each co-transfection assay and inoculated onto agar plate of minimal SD base/-Trp/-Leu/his plus 5 mM 3-amino-1,2,4-triazole. After 3-4 days, cells on the agar plates were transferred onto a filter. Color development was performed with β-galactosidase to detect protein-protein interaction according to the manufacturer's instruction.
3.4 RESULTS

3.4.1 Identification of the Motif in Aggrecan G1 Domain That Inhibits Secretion—

In order to dissect the motif in aggrecan G1 domain that is responsible for the inhibition of product secretion, we generated a number of recombinant constructs carrying either IgG or the tandem repeats: G1, IGD, TR\textsuperscript{G1} (or TR\textsuperscript{G1}TR\textsuperscript{G1}). Link protein is structurally similar to the G1 domain and was thus used as a control. The structures of these constructs are shown in Fig. 1A-3. In order to compare amounts of secreted products, we transfected COS-7 cells with these constructs. Western blot analysis indicated that the G1 product was synthesized but not secreted to culture medium (Fig. 1D-3). This confirms previous findings (Kiani et al., 2001; Luo et al., 1996). As a positive control, link protein, which is structurally similar to the aggrecan G1 domain, was also expressed and analyzed. High levels of this product were observed in the culture medium. The IgG motif, the largest motif in the aggrecan G1 domain, was also highly expressed and secreted (Fig. 1D-3). The products observed in the culture medium were larger than those in the cell lysate, indicating glycosylation of the product in the medium, as aggrecan is known to contain many N- and O-glycosylation sites (Chandrasekaran and Tanzer, 1992; Li et al., 1993). In contrast, the tandem repeat domains (TR\textsuperscript{G1}) were not secreted (Fig. 1D-3).

Our results indicated that, in spite of the resemblance between the G1 domain and link protein, the two peptides have different effects: aggrecan G1 product is not secreted, while link protein is well secreted. Since our evidence indicated that the inability of product secretion is related to the tandem repeats, we investigated these motifs in more detail. Alignment of amino acid sequences from different species indicates that the sequences of each motif are highly conserved between species. For example, TR1 and TR2 motifs of the
G1 domain (i.e. TR1\textsuperscript{G1} and TR2\textsuperscript{G1}), TR1 and TR2 motifs of the G2 domain (TR1\textsuperscript{G2} and TR2\textsuperscript{G2}) and TR1 motif of link protein (TR1\textsuperscript{LP}) exhibit 95.7, 94.2, 88.8, 88.2 and 97.2% homology, among human, rat, bovine, murine and chicken sequences respectively (Fig. 2A-3). Notably, human, rat, cow and mouse share 100% homology in the TR1 motif of the G1 domain. However, the homology between different motifs is relatively low in each species, and the patterns of homology are very similar (Fig 2B-F-3). Interestingly, among all species, the TR1 motif of G1 domain and TR1 motif of link protein, both of which bind hyaluronan and were thus expected to be more homologous, share, in fact, low degrees of homology (42.9-45.9%). In addition, the TR1 and TR2 motifs of the G1 domain, both of which are known to share the same function in hyaluronan-binding, exhibit the lowest level of homology (25.8-27.6%). On the other hand, the TR1 motifs in G1 and G2, which have no known functional similarity, share the highest degree of homology (63.3-66.3%) among the 5 motifs compared (Fig 2B-F-3). This suggests that the TR1 motifs might share some similar functions, while the TR1 and TR2 in the same domain may play different roles in spite of their hyaluronan binding activity.

These data have led to the speculation that individual tandem repeats of the G1 and G2 domains may play different roles in product secretion. In light of this, we designed experiments to determine which tandem repeats affect product secretion. Recombinant constructs IgGTR1\textsuperscript{G1}, IgGTR2\textsuperscript{G1}, TR1\textsuperscript{G1}, TR2\textsuperscript{G1}, TR1\textsuperscript{G1}G3 and TR2\textsuperscript{G1}G3 were generated (Fig. 3A-3). As a control, we used a G3 construct generated previously (Yang et al., 1998; Cao and Yang, 1999). These constructs were expressed in COS-7 cells, and the products were analyzed as above. The TR1\textsuperscript{G1} product was not secreted (Fig. 3B-3). Similar results were observed when the TR1\textsuperscript{G1} motif was combined with other domains, whose products
were otherwise well secreted: the IgGTRIG1 and TR1G1G3 products were not secreted. As a
close, G3 was expressed and secreted as expected (Fig. 3B-3). The TR2G1 product was
observed in the medium at high levels. The observed sizes of the TR2G1 (33 kDa in cell
lysate and 38 kDa in medium) were greater than the sequence of the TR2G1 peptide backbone
would predict (~12 kDa). This strongly suggested glycosylation had occurred, as this motif
contains many potential sites for glycosylation. The TR2G1 motif was then linked with the
IgG-like motif or the G3 domain of aggrecan, producing the IgGTR2G1 and TR2G1G3
constructs. Immunoblotting showed that both products of IgGTR2G1 and TR2G1G3 were well
secreted to culture medium (Fig. 3B-3).

3.4.2 The G1 Domain Enhances GAG Addition — We reasoned that G1’s ability to
retard product secretion must be of particular importance to the aggrecan processing and
secretion. Since GAG chain modification is an important step in aggrecan processing, we
investigated if the G1 domain had any effect on GAG modification. In order to do so, two
constructs, the mini-aggrecan (G1CSG3) and a chimeric construct LPCSG3, were generated
(Fig. 4A-3). In the latter, a link protein (LP) moiety, which we had shown to be well secreted
(Fig. 1D-3), replaced the G1 domain. Product analysis indicated that both constructs were
well expressed, but LPCSG3 was secreted at much greater levels than was G1CSG3 (Fig.
4B-3). However, it was also evident that G1CSG3 was subjected to higher levels of GAG
chain addition after secretion, as indicated by more diffusion at greater molecular mass. To
confirm this, products of G1CSG3 and LPCSG3 in culture medium were purified with equal
amounts of antibody 4B6, to isolate equal amounts of products (Fig. 4C-3). Western blot
analysis indicated that the G1CSG3 sample did exhibit more diffusion in the upper portion of
the band, suggesting higher levels of GAG addition. To confirm that the diffusion was the
result of GAG modification, products of these constructs were immuno-precipitated and treated with chondroitinase ABC. We observed a single band in all samples, indicating removal of GAG chains from all products (Fig. 4D-3).

3.4.3 TR1\textsuperscript{G1} Binds to ER Protein and Affects Processing of Mini-aggrecan—The finding that G1 retards product secretion but on the other hand facilitates greater GAG modification suggests that this domain may interact with some cellular components. This interaction might in turn affect the mechanics of aggrecan processing and secretion. To investigate this, we used the tandem repeats (TR1TR2\textsuperscript{G1}) as a bait in pGBK7 plasmid (producing a construct pGBK7-TR1TR2; Fig. 5A-3) to screen a human brain cDNA library in a yeast two-hybrid experiment. After clone selection and DNA sequencing, we obtained a clone (pACT2-ER) with sequence identical to the cDNA of 2655 bp encoding calcium homeostasis endoplasmic reticulum protein (GenBank nucleotide access number NM_006387). Cloned from human erythroleukaemia cells, the ER protein is located at the ER membrane where it modulates intracellular calcium mobilization (Laplante \textit{et al.}, 2000).

In addition, this protein has ER retention signals at the C-terminus and several potential domains for macromolecular assembly, such as the histidine repeats and serine/arginine-rich region for protein-protein interactions. Our clone has an insert of 2 kb. It contains a C-terminal coding region of 660 bp and a 3' non-coding region of \~{}1.3 kb. The serine/arginine-rich region (Ser\textsuperscript{705} to Ser\textsuperscript{785}) is located within the clone pACT2-RT we isolated.

We further confirmed the binding of pGBK7-TR1TR2 with pACT2-ER using yeast two-hybrid color development assay. This assay used the plasmid pGBDC1, another plasmid in yeast two-hybrid system. Three constructs, pGBDC1-TR1TR2, pGBDC1-TR1, and pGBDC1-TR2, were made to examine which tandem repeat binds to pACT2-ER. Color
development indicated that pACT2-ER interacted with pGBDC1-TR1TR2 and pGBDC1-TR1 but not with pGBDC1-TR2 compared to vector control (Fig. 5B-3).

We then investigated whether the interaction of the tandem repeats (particularly TR1) with a cellular component such as the calcium homeostasis ER protein is essential for product processing. If so, over-expression of TR1, which interacts with the ER protein and accumulates in the cytoplasm, would affect processing of a product containing TR1. The mini-aggrecaen was thus co-expressed with TR1G1, TR2G1 or the control vector pcDNA3. As predicted, the experiment indicated that the mini-aggrecaen was well expressed, but its secretion was inhibited when co-expressed with TR1G1 but not with TR2G1 (Fig. 6-3).

3.4.4 The Tandem Repeats of the G2 Domain Function Similarly to G1 in Product Secretion—The second globular domain in aggrecan, G2, differs from G1 in that it is composed only of two tandem repeat motifs and lacks an IgG-like motif. As shown by the homology analysis, the G2 tandem repeats are conserved across species but are not homologous to the TR’s of G1 or to each other. Only the TR1 motif of the G2 domain shares high degrees of homology with the TR1 motif of the G1 domain. Seven constructs, G2, TR1G2, TR2G2, TR1G2G3, IgGTR1G2, TR2G2G3 and IgGTR2G2 were generated in order to test their effects on product secretion (Fig. 7A-3). These constructs were expressed in COS-7 cells, and cell lysate and culture medium were analyzed on Western blot. Two products, TR1G2G3 and IgGTR1G2, were not secreted (Fig. 7B-3). Consistent with the results from G1 domain, the TR2G2 product and TR2G2-associated products (TR2G2G3 and IgGTR2G2) were secreted. Thus, in both globular domains, the first tandem repeat, TR1G1 and TR1G2, seems to incapacitate product secretion.
3.4.5 Identification of the Amino Acids in the TR1 Motif That Account for the Incapability of Product Secretion —To further investigate how the TR1 motif of the G1 domain failed to be secreted, seven constructs shown in Fig. 8A-3 (aligned with the construct TR1<sup>G1</sup>) were generated. Among these seven constructs, four of them (TR<sup>G1</sup>K→Q, TR<sup>G1</sup>mu1, TR<sup>G1</sup>mu2 and TR<sup>G1</sup>mu3) contain amino acids mutations. These amino acids are 100% conserved in the TR1 motifs of the G1 and G2 domains across all species analyzed, and we expected that these amino acids might be important in product processing. Two constructs were truncated: one lacked the N-terminal 9 amino acids (TR1<sup>G1</sup>N9del), and the other lacked the C-terminal 15 amino acids (TR1<sup>G1</sup>C15del). The seventh construct TR<sup>G1</sup>Δ381 contains a deletion in the middle of the two tandem repeats due to the presence of two KpnI restriction endonuclease sites in the same reading frame. COS-7 cells were transiently transfected with these seven constructs. Analysis of culture medium and cell lysate on Western blot probed with 4B6 indicated that only the products of TR1<sup>G1</sup>N9del were secreted, suggesting that the N-terminal 9 amino acids played an important function in product secretion (Fig. 8B-3).

To further narrow down the N-terminal sequence of amino acids affecting product secretion, five constructs G1N4del, G1N4mu, TR1<sup>G1</sup>N4del, TR1<sup>G1</sup>N4mu and TR1<sup>G1</sup>HRmu were generated. There are 4 amino acids (FHYR), which are identical in the TR1 motifs of the G1 and G2 domains across all five species analyzed. In G1N4del and TR1<sup>G1</sup>N4del, we deleted these 4 amino acids (FHYR) from the N-terminus of the TR1 motif. In G1N4mu and TR1<sup>G1</sup>N4mu, these 4 amino acids were mutated to ILNS, and in TR1<sup>G1</sup>HRmu, H and R were mutated to LS as shown in Fig. 9A-3. These truncation/mutation constructs were transiently expressed in COS-7 cells. Product analysis on Western blot indicated that all constructs were
well-expressed and secreted to culture medium (Fig. 9B-3). In order to test if the 4 amino
acids FHYR are responsible for the inability of product secretion, they were linked to the
aggrecan G3 construct. Product analysis indicated that this chimeric product was secreted
(data not shown).

In order to examine the effect of the TR1 N-terminal sequence on G1 secretion, we
generated six constructs containing the IgG motif linked to the N-terminal fragments of the
TR1 motif of increasing sizes (from 10 to 55 amino acids; Fig. 10A-3). A 45-amino acid
fragment of the TR1 N-terminal which contained a point mutation (C55→P45) was also
generated and linked to IgG (IgG45amu). These constructs were transiently expressed in
COS-7 cells as above. Analysis of these products indicted that IgG linked to the N-terminal
10 or 20 amino acids of TR1G1 was well secreted. The constructs containing the 35 or 45
amino acid fragments were partially secreted, and mutation of the cysteine apparently had no
effect. IgG linked to the N-terminal 55 amino acid fragment was not secreted (Fig. 10B-3).
Table 1-3. Sequences and restriction endonuclease sites of oligonucleotides

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
<th>Restriction Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIGN</td>
<td>5' AAAGGATCCCTGGAGGCTCCCTG3'</td>
<td>BamHI</td>
</tr>
<tr>
<td>ABN</td>
<td>5' AAACCTGGGAGCGCTGTTGATCCACCACTAC3'</td>
<td>XhoI</td>
</tr>
<tr>
<td>AAC</td>
<td>5' AAAAGGATCCCTGGAGGCTCCCTGACCTCCTAT3'</td>
<td>SphI</td>
</tr>
<tr>
<td>G1C</td>
<td>5' AAAAGGATCCCTGGAGGCTCCCTGAAAGAT3'</td>
<td>SphI</td>
</tr>
<tr>
<td>G3C</td>
<td>5' AAAAGGATCCCTGGAGGCTCCCTGACCTCCTGACAT3'</td>
<td>XhoI</td>
</tr>
<tr>
<td>ABC</td>
<td>5' AAAAGGATCCCTGGAGGCTCCCTGACCTCCTGACAT3'</td>
<td>SphI</td>
</tr>
<tr>
<td>AB'N</td>
<td>5' AAACCTGGGAGCGCTGTTGATCCACCACTAC3'</td>
<td>XhoI</td>
</tr>
<tr>
<td>G2F</td>
<td>5' AAACCTGGGAGCGCTGTTGATCCACCACTAC3'</td>
<td>XhoI</td>
</tr>
<tr>
<td>G2R</td>
<td>5' AAAAGGATCCCTGGAGGCTCCCTGAAAGAT3'</td>
<td>XhoI</td>
</tr>
<tr>
<td>G2TR1C</td>
<td>5' AAAAGGATCCCTGGAGGCTCCCTGAAAGAT3'</td>
<td>XhoI</td>
</tr>
<tr>
<td>G2TR2N</td>
<td>5' AAAAGGATCCCTGGAGGCTCCCTGAAAGAT3'</td>
<td>XhoI</td>
</tr>
<tr>
<td>G1TR2N</td>
<td>5' AAAAGGATCCCTGGAGGCTCCCTGAAAGAT3'</td>
<td>XhoI</td>
</tr>
<tr>
<td>G1TRIC</td>
<td>5' AAAAGGATCCCTGGAGGCTCCCTGAAAGAT3'</td>
<td>XhoI</td>
</tr>
<tr>
<td>LPNKozak</td>
<td>5' AAAAGGATCCCTGGAGGCTCCCTGAAAGAT3'</td>
<td>XhoI</td>
</tr>
<tr>
<td>5' AAAGGATCCCTGGAGGCTCCCTGAAAGAT3'</td>
<td>XhoI</td>
<td></td>
</tr>
<tr>
<td>TR1(mu (K→Q))</td>
<td>5' CTGCCTGGGAGCGCTGTTGATCCACCACTAC3'</td>
<td>XmaI</td>
</tr>
<tr>
<td>TR1Cdel15</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>TR1Ndel4</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>TR1N2mu</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>KpnRmu1</td>
<td>5' GATGGGGATCCACAGAGACGCTCGAG3'</td>
<td>KpnI</td>
</tr>
<tr>
<td>KpnRmu2</td>
<td>5' GATGGGGATCCACAGAGACGCTCGAG3'</td>
<td>KpnI</td>
</tr>
<tr>
<td>TR145aaR</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>TR145aaRmu</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>TR135aaR</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>TR120aaR</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>TR110aaR</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>G3C</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>TR1Cmu</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>TR1N4mu</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>G3N</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>G1C</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>TR1Ndel9</td>
<td>5' GGGCTAGCTTCACCCAGACAGGATGACT3'</td>
<td>XbaI</td>
</tr>
<tr>
<td>pcDNA1re</td>
<td>5' ATTATAGCTGCAATAGATAG3'</td>
<td>BamHI</td>
</tr>
<tr>
<td>TR1N</td>
<td>5' CCGGGATCCCTGGAGGCTCCCTGAAAGAT3'</td>
<td>BamHI</td>
</tr>
<tr>
<td>AG1C</td>
<td>5' AAAGGATCCACAGGACGCTGTTGATCCACCACTAC3'</td>
<td>SalI</td>
</tr>
<tr>
<td>TR2N</td>
<td>5' AAAGGATCCACAGGACGCTGTTGATCCACCACTAC3'</td>
<td>SalI</td>
</tr>
<tr>
<td>TR1C</td>
<td>5' CCGGGATCCCTGGAGGCTCCCTGAAAGAT3'</td>
<td>SalI</td>
</tr>
</tbody>
</table>
Figure 1-3. IgG, but not tandem repeats, is secreted to culture medium. A, diagram of the structures of aggrecan (wild-type and nanomelic). IgG, immunoglobulin-like motif; TR, tandem repeat; IGD, inter-globular domain; KS, the fragment of the core protein modified by keratan sulfate chains; CS, the fragment modified by chondroitin sulfate chains; CRD, carbohydrate recognition domain; CBP, complement binding protein. The signal peptide added to all constructs was obtained from link protein (nucleotides 1-180). Numbers above schematic correspond to nucleotides in the sequence of full-length aggrecan. B, structure of the aggrecan G1 domain. Each circle represents one amino acid residue. The IgG-like motif and the second tandem repeat (TR2) are shown as open circles, while the first tandem repeat (TR1) is shown as closed circles, where C represents a cysteine residue, and S-S represents a disulfide bond. C, four recombinant constructs were generated and used in COS-7 cell transfection. D, cell lysate (l) and culture medium (m) from COS-7 cells transfected with link protein, G1, IgG and TR1<sup>G1</sup> (TR<sup>G1</sup>G1TR2<sup>G1</sup>) were subjected to SDS-PAGE on a 10% gel. The separated proteins were then transblotted onto a nitrocellulose membrane and probed with the monoclonal antibody 4B6. The products of IgG were synthesized (32 kDa in lysate) and were secreted to the culture media (40 kDa) at high levels. The core protein of link protein was also well secreted to the culture medium, but neither the G1 product nor the TR1<sup>G1</sup>TR2<sup>G1</sup> (47 kDa) were secreted.
Figure 2-3. Amino acid homology of TR1 and TR2 motifs. A, sequence analysis indicated that each motif is highly conserved across species. The sequence of each species was compared to the human sequence; the total homology 95.7%, 94.2%, 88.8%, 88.2% and 97.2% is the homology of all species. Among human (B), cow (C), rat (D), mouse (E) and chicken (F), the patterns of conservation are very similar in the order of (from high homology to low homology) TR1G1 > TR1L > TR2G1 > TR1G2 > TR2G2 > TR1G1, TR1G2 and TR1L > TR1L, TR2G2 and TR2G1 > TR1G1 and TR1L > TR1G2 and TR2G1 > TR1G1 and TR2G1.
Figure 3-3. Products containing TR1\(^{G1}\) are not secreted. \(A\), six constructs containing one of the G1 tandem repeats were transiently expressed in COS-7 cells. Aggrecan G3 construct was used as a control. \(B\), cell lysate and culture medium were prepared and analyzed on Western blot probed with 4B6. G3 (48 kDa), TR2\(^{G1}\) (33 kDa), IgGTR2\(^{G1}\) (38 kDa), and TR2\(^{G1}\)G3 (50 kDa) were well synthesized and secreted. However, neither the products of TR1\(^{G1}\) (25 kDa), nor the TR1-associated products, TR1\(^{G1}\)G3 (42 kDa) and IgGTR1\(^{G1}\) (40 kDa), were secreted to the culture medium.
Figure 4-3. The G1 domain enhances GAG modification. A, the mini-aggrecan (G1CSG3) construct and the chimeric construct LPCSG3, in which the G1 domain of G1CSG3 was replaced with link protein. B, cell lysate (l) and culture medium (m) of G1CSG3 and LPCSG3 were analyzed on Western blot probed with 4B6 (sample volume as shown). Both constructs were well expressed, but the levels of secreted LPCSG3 were much higher than those of G1CSG3. C, the secreted products of G1CSG3 and LPCSG3 were purified using equal amounts of antibody 4B6 and analyzed on Western blot. G1CSG3 product exhibited a much higher level of GAG addition indicated by diffusion at higher molecular mass. D, treatment with chondroitinase ABC resulted in a single major band for all samples.
Figure 5-3. Interaction of TR1\textsuperscript{G1} with ER protein. 

A, generation of yeast two-hybrid constructs pGBDC1-TR1TR2\textsuperscript{G1}, pGBDC1-TR1\textsuperscript{G1} and pGBDC1-TR2\textsuperscript{G1}, and schematic of ER protein clone, pACT2-ER, isolated using TR1TR2\textsuperscript{G1} as the bait. The ER protein clone (2 kb insert) contained 660 bp coding sequence encoding the C-terminal 220 amino acids and \textasciitilde1.3 kb 3' non-coding sequence. 

B, in yeast two-hybrid assay with color development, the ER protein clone interacted with pGBDC1-TR1TR2\textsuperscript{G1} and pGBDC1-TR1\textsuperscript{G1}, but not with pGBDC1-TR2\textsuperscript{G1}, compared to pGBDC1 vector control.
Figure 6-3. Secretion of the mini-aggre can is inhibited by TR1. The mini-aggre can was co-expressed with TR1G1, TR2G1 or a control vector in COS-7 cells. Culture medium (m) and cell lysate (l) were subjected to electrophoresis in 7% gel, followed by Western blot analysis probed with 4B6. Co-expression with TR1G1 inhibited the secretion of the mini-aggre can
Figure 7-3. **G2 has similar function in product secretion.** *A,* seven constructs were generated as shown in the figure and transiently expressed in COS-7 cells. *B,* cell lysate (l) and culture medium (m) were analyzed on Western blot as above. All constructs were well expressed. Although TR2G2 and TR2G2-associated products (TR2G2G3 and IgGTR2G2) were well secreted, TR1G2 and TR1G2-associated products (TR1G2G3 and IgGTR1G2) were not secreted to culture medium.
Figure 8-3. The N-terminal is responsible for the inhibitory effect on secretion. A, seven constructs (aligned with the construct TR1G) were produced to study the effect of the C- and N-terminal fragments on product secretion. The strategy and primers for the construction was shown in the figure. For example, construct TR1GK→Q was produced by linking two fragments in EcoRI/XbaI-digested pcDNA3: fragment A was obtained by digesting the TR1G construct with EcoR1/Xmal, and the other fragment was synthesized with two primers TRimuXmal and GICXbaI in a PCR. TRGID381 was generated by deletion of the KpnI-digested fragment, which does not change the reading frame of the construct. The construct TR1Gmu3 was synthesized with two primers LPNKozak and TRICmuXbaI, and so on. B, COS-7 cells were transiently transfected with these constructs. Culture medium and cell lysate were analyzed on Western blot probed with 4B6. Among these products, only the products of TR1G N9del were secreted to the culture medium.
Figure 9-3. The amino acids H^2R^4 in the TR1 of G1 domain are essential for the inhibition of product secretion.

A, five constructs G1N4del, G1N4mu, TR1^G1N4del, TR1^G1N4mu and TR1^G1HRmu were generated, which contain deletion of 4 amino acids (FHYR) from the N-terminus of the TR1 motif, mutation of these 4 amino acids or mutation of H and R as shown in the figure. All primers used for PCR are shown in the figure, where pcDNA1re is located beyond the last cloning site XbaI of pcDNA1 vector. Thus, all PCR products were digested and inserted into XhoI/XbaI-digested vector containing LP60 or LP60-IgG fragment. B, these deletion/mutation constructs were transiently expressed in COS-7 cells. Product analysis on Western blot probed with 4B6 indicated that all constructs were well-expressed and secreted to culture medium. Some secreted products were larger in sizes than the products in cell lysate suggested glycosylation of those products.
Figure 10-3. Fifty-five amino acids at the amino terminus of TR1 are sufficient to inhibit product secretion.

A, six constructs containing the IgG motif linked with different sizes (10, 20, 35, 45 and 55 amino acids) of the N-terminal fragments of the TR1 motif or a point mutation (C<sup>45</sup>→X<sup>45</sup>) of the TR1 were generated and transiently expressed in COS-7 cells. B, analysis of these products on Western blot probed with 4B6 indicted that attachment of the N-terminal 20 amino acids did not inhibit IgG product secretion. Addition of 35 or 45 amino acids partially inhibited product secretion, while mutation of the cysteine had no additive effect. Addition of the N-terminal 55 amino acids completely inhibited product secretion.
**3.5 DISCUSSION**

The large aggregating chondroitin sulfate proteoglycans are a family of glycoconjugates with a central core protein to which GAG side chains are covalently linked post-translationally. Aggrecan, a structural proteoglycan of this family, is expressed in cartilage and is responsible for its resilience and load-bearing properties. Since its partial cloning, the structure and expression of aggrecan has been extensively studied (Fosang et al., 1992; Goetinck et al., 1987; Hascall and Heinegard, 1974; Sandy et al., 1992; Watanabe et al., 1994). The core protein of aggrecan is composed of three globular domains (G1, G2, G3). A large extended region (CS) is situated between G2 and G3 for GAG chain attachment that occurs on the serine of serine-glycine dipeptides present in this region (Chandrasekaran and Tanzer, 1992; Li et al., 1993). As aggrecan is such a large proteoglycan, it would be difficult to study its processing with full-length aggrecan. In fact, studies have indicated that each aggrecan domain has a unique functional role in product secretion (Day et al., 1999; Domowicz et al., 2000; Kiani et al., 2001; Luo et al., 1996; Zheng et al., 1998). In this report, we demonstrated that the function of each motif is consistent in all types of combinations, and changing restriction endonuclease sites in the ligation linkage (creating some non-native amino acids) did not alter its property. It seems likely that amino acids in the domain junctions do not change the folding of the domain nor the interactions with intracellular proteins such as chaperones. This property also allowed us to test the role of each motif in product synthesis and secretion in different types of cells, most of which would otherwise not be able to produce a native full-length aggrecan. These domain constructs have been successfully expressed in chondrocytes, mesenchymal cells, fibroblasts, CHO cells and COS-7 cells (Cao and Yang, 1999; Domowicz et al., 2000; Kiani et al., 2001; Luo et al.,
In our studies, we have not observed obvious differences among cell types in product properties such as product synthesis, secretion, GAG chain attachment and glycosylation (Cao and Yang, 1999; Cao et al., 1998; Kiani et al., 2001; Wu et al., 2001; Yang et al., 1998; Yang et al., 2000; Zhang et al., 1998).

We have previously observed that aggrecan GI and G2 domain constructs were expressed but not secreted (Kiani et al., 2001). This observation was confirmed in this study, and we further characterized these domains to determine which motifs could account for these effects. We demonstrated that, expressed individually, the product of the IgG-like motif of GI was well secreted. Only in the presence of the two tandem repeats was the GI product not secreted. Interestingly, when the IgG-like motif and the tandem repeats were expressed individually, the products of both constructs were heavily modified by sugar moieties, as indicated by the increase in the size of the products. When these motifs were expressed as one construct, the GI construct, the product was only slightly modified by glycosylation. One explanation may be that, when the peptides are small, potential sites for GAG chain attachment and glycosylation are exposed to modifying enzymes. It is not clear how these sites are modified on native aggrecan expressed by chondrocytes. As aggrecan contains a large number of potential sites for GAG chain attachment and glycosylation, and mutation of one potential site would not change the molecular mass significantly, the large size of aggrecan means it is impossible to examine how each potential site is modified.

However, using smaller constructs consisting of one or more motifs will provide valuable insights on GAG chain attachment and glycosylation for a particular site.

Further examination indicated that only one of the two tandem repeats of the GI domain was involved in modulating product secretion. These two tandem repeats are
structurally similar and both are known to bind hyaluronan (Goetinck et al., 1987; Hascall and Heinegard, 1974). No other function has been claimed for these two tandem repeats. Our study has revealed a novel functional role for the tandem repeats: we have shown that aggrecan TR1\textsuperscript{G1}, and not TR2\textsuperscript{G1}, retards secretion. The function of the two tandem repeats in the processing of other members of the large aggregating chondroitin sulfate proteoglycan family (including versican, neurocan and brevican) has not been tested. However, we speculate that the first tandem repeat in the G1 domain of these proteoglycans may have a similar effect on product secretion, as the TR\textsuperscript{1} sequences of these molecules are highly conserved (data not shown). Indeed, we have shown that the tandem repeats of versican also retarded product secretion, although they have not yet been individually tested (Yang et al., 2000). The amino acid sequences of the tandem repeats exhibit 100% homology across all species examined except chicken, which is only 86.3% homologous to the other species. The conservation was observed not only among the G1 domains of these large aggregating sulfate proteoglycans, but also between the G1 and G2 domains of aggrecan.

We hypothesized that the amino acid residues involved in hyaluronan binding must be conserved among aggrecan, versican and link protein, and is likely to be conserved between the two tandem repeats of the G1 domain. On the other hand, the amino acid residues involved in product processing must be conserved among the first tandem repeat of G1 and G2 domains of aggrecan and versican. Based on these hypotheses, we designed a series of point mutations to pin-point the amino acid residues involved in product processing. With this strategy, we were able to identify two amino acid residues (His and Arg) in the N-terminal of the first tandem repeat that play important role in retarding product secretion.
However, for complete inhibition of product secretion, the N-terminal 55 amino acid residues seem to be required.

Our studies provide the first proof of a biological activity for the aggrecan G2 domain, a domain which contains two tandem repeat motifs. G2 is unique to aggrecan and not found in any other chondroitin sulfate proteoglycan. It is homologous to the TR motifs of aggrecan G1 domain and also to link protein but, though the tandem repeats of G1 have been shown to bind to hyaluronan, G2 is incapable of this (Hascall and Heinegard, 1974; Lohmander et al., 1989b; Sandy et al., 1992). Using expression and immunoblotting techniques, our study has shown for the first time that G2, like G1, fails to be secreted in the presence of the first tandem repeat.

Why would aggrecan contain two elements (G1 and G2) that hinder product secretion? It is highly probable that they function as double check-points in the processing of aggrecan. As aggrecan contains significantly more GAG chains than the other chondroitin sulfate proteoglycans, and aggrecan contains an extra G2 domain, it is logical to suggest that their effects on product secretion are related to GAG chain attachment. Retarding product secretion may allow sufficient GAG chain attachment, and this may be required in order to produce mature aggrecan. If this cannot be achieved, the immature products will be subjected to degradation for recycling. Working together, G1 and G2 would ensure that only mature proteoglycan would be secreted into the extracellular space. To support this notion, we expressed a mini-aggrecan and a chimeric construct, in which the G1 domain of the mini-aggrecan was replaced with link protein (LPCSG3). Link protein is well secreted, and the product of the chimeric construct was secreted at much higher levels than the mini-aggrecan. However, in the presence of the G1 domain, the mini-aggrecan was modified by GAG chains
at a higher level, strong evidence of G1's role in proteoglycan maturation. Under normal physiological conditions, the combined retarding effect of G1 and G2 on product secretion is overcome by the combination of a CS sequence and the G3 domain (Kiani et al., 2001), allowing secretion of mature aggrecan. The G3 domain also promotes the attachment of GAG chains to the CS sequence, and this is essential for product processing.

It may seem contradictory that link protein, which is structurally similar to the G1 domain of aggrecan, is secreted to the extracellular space at high levels. Our analysis of link protein IgG-like motifs and the tandem repeats indicated that both products were well secreted to the culture medium. We speculate that link protein gene has evolved so that its products are secreted to serve as a matrix molecule. Consistent with this, the amino acid sequences of link protein are highly conserved among different species, while these sequences are not very conserved between link protein and the G1 domain nor between link protein and the G2 domain. These differences may alter link protein's secretory property but still allow it to retain hyaluronan-binding ability.

Although the mechanism by which TR1 affects product secretion is still not fully understood, our finding that the calcium homeostasis endoplasmic reticulum protein interacted with the TR1 motif shed some light on how TR1 affects aggrecan processing. This interaction may be key to aggrecan processing. There are at least two potential hypotheses to account for this: (i) The ER protein may bind mutated or misfolded core proteins and retain them within the cell for the purpose of quality control. (ii) The interaction of the ER protein with the TR1 motif may allow sufficient GAG modification to the CS sequence of the core protein. After being sufficiently modified by GAG chains, the product is released from the ER protein for further processing. There are numerous reports to support the former
explanation. For example, BiP is often associated with nonsecretable or misfolded mutants, and either retards substrate transportation or targets substrates for degradation through the Sec61p channel (Hamman et al., 1998; Knittler et al., 1995; Ward et al., 1981). The lectin chaperones are also well positioned to determine the fate of proteoglycans; a proteoglycan is either recognized as immature (to be retained in the ER allowing for further folding or directed to degradation pathway through binding to BiP) or recognized as mature and transported (Hebert et al., 1995). This is also supported by our observation that some products, if not secreted, were degraded without retention, and only trace amounts of degraded products could be detected in the cell lysate (data not shown). On the other hand, some other products, although not secreted (e.g., G1 product), could be detected at levels as high as those that were secreted (e.g., G3 product). This may be explained by the structures of the products: folded products may be more resistant to degradation. However, two pieces of evidence seem to favor the latter. (i) In the presence of the G1 domain, the mini-aggrecaen exhibited higher levels of GAG modification, suggesting a function for G1 in product processing. (ii) Over-expression of the TR1 construct reduced GAG modification and product secretion of mini-aggrecaen, suggesting the presence of a cellular component that binds to TR1, and this binding exerts a limiting effect on product processing. It should be noted that the mechanism by which G1 enhances GAG modification is not clear. Its ability to bind ER protein is only one explanation. As well, the fact that over-expression of TR1 reduced GAG modification and product secretion does not exclude the possibility that TR1 also binds to other cellular components (such as protein chaperons), and thereby interferes with the processing of the mini-aggrecaen. It should be noted that this type of interaction — the interaction of TR1 of G1 with the ER protein — in protein processing has never been
reported. Although the cDNA sequence of the ER protein has been deposited in GenBank, its function in protein processing is completely unknown. It is possible that the processing mechanism reported here is unique to the large aggregating chondroitin sulfate proteoglycan family.

Quality control proteins are known to recognize signals on the surface of a protein. This includes the carbohydrate moieties attached to the core proteins of proteoglycans, the KDEL peptide, detected at the C-terminus of many ER-retained proteins, and the C-terminal peptide of rhodopsin (Deretic et al., 1998). Evolution may have placed such trafficking signals on the surface or at the N- and C-terminal fragments so that they are easily accessible to interacting proteins. The motif that binds to the ER protein is located in the middle of the G1 domain. This represents an entirely novel mechanism for protein trafficking. Investigation of the biological functions of the ER protein may elucidate new information on aggregcan processing.
GENERAL CONCLUSION

Joint diseases can be broadly categorized into early onset and late onset disorders of the articular surfaces. Early onset joint diseases can have a number of etiologies such as infectious agents, genetic predisposition, autoimmunity and so on. Late onset joint diseases are usually due to the advancing age and the inability of the articular cartilage to maintain adequate levels of self-repair. Late onset joint disorders are becoming more prevalent in the North American populations which is partly due to the “Baby Boomers Era” and the aging of this rather large portion of the population. In order to remedy these joint diseases, one has to have a comprehensive understanding of the articular cartilage composition is required in terms of cells, chondrocytes, and the macromolecules secreted by these cells as well as deciphering how these molecules interact with chondrocytes and each other to bring about a smooth functioning of the articular surfaces. Articular cartilage consists of chondrocytes and their secreted macromolecules including a variety of collagen molecules, large aggregating chondroitin sulfate proteoglycans, small leucine rich proteoglycans, hyaluronic acid, link protein and cartilage oligomeric matrix protein. It can be appreciated that any malfunctioning of these proteins (which may well arise from point mutations, frame shift mutations and so on) can have a tremendously negative impact on the functioning of the articular cartilage. For instance, mutations in collagen molecules have been linked to a variety of chondro-skeletal dysplasias such as achondrogenesis and spondylo-epiphyseal dysplasia (SED) as well as Stickler’s syndrome, which is a form of early onset osteoarthritis (OA) compounded with myopia and retinal detachment. The importance of another type of ECM molecule, link protein, has been highlighted by the fact...
that the majority of the link protein knock-out mice harbored gross skeletal deformities which were lethal. The same scenario holds true for the aggrecan molecule where frame shift or point mutations have given rise to a truncated aggrecan which cannot be properly secreted to the ECM.

Aggrecan is a multimodular protein which is heavily decorated by GAG chains as well as N- and O-linked oligisaccharides. Previous studies on aggrecan have shown that the G3 domain, by virtue of its interaction with HSP25, is able to transport the mature aggrecan molecule out of the cells. However, our work with different domains of aggrecan showed that G3 was not the only participant in promoting aggrecan secretion. Our data indicate that GAG chain attachment is the other participant in promoting secretion as shown in our experiments with two recombinant genes, G1G3 and G1CSG3. The fact that G1G3 recombinant products are expressed but poorly secreted indicates that G3 needs some auxiliary factor to promote secretion and this factor comes in the form of GAG chain attachment since the recombinant G1CSG3 products are expressed and secreted. Our findings have demonstrated that each aggrecan domain and subdomain can impact the way aggrecan is processed and secreted. Furthermore, our findings suggest that aggrecan molecule might be using some of its domains (as quality control checkers) to regulate its secretion in such a way that only properly mature aggrecan molecules find their way out of the cells, since the presence of defective aggrecan macromolecules might adversely affect the formation of the gel-like structures in the articular cartilage. We have shown that G1 and G2 could negatively regulate aggrecan secretion and suggest that these two domains might cause a delay in aggrecan transit from the RER to the Golgi. We have demonstrated that the inhibitory effects of G1 and G2 on secretion is inherent in the TR1 of these two domains where four N-terminal
amino acids of TR1 of G1 are implicated in this effect. My studies not only provide the first evidence for the biological function of aggrecan G2 domain in product processing, but also shed some light on the mechanism by which aggrecan is produced. My results that TR1 motifs of G1 and G2 domains inhibit product secretion suggest that these motifs must interact with some as yet unidentified cellular component (cellular chaperones) in order to exert such an effect. Using a yeast two-hybrid system assay, our laboratory has recently cloned at least two groups of genes, a group of DnaJ/HSP40 proteins and a group of ER proteins involved in calcium homostasis. The former is a protein chaperone and the function of the latter is as yet unknown. Investigation of the role of DnaJ/HSP40 proteins in aggrecan processing is currently underway. The latter is an ER membrane-bound protein involved in intracellular calcium homostasis.
REFERENCES


alpha 1,4-N-acetylgalactosaminyltransferase from a human sarcoma cell line.

*Glycobiology.* 9:697-703.


