

Molecular Studies of AggreCAN: Experiments and Simulations

by

Fei Liang

B.E., Dalian University of Technology (2006)

Submitted to the Department of Chemical Engineering
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Chemical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

© 2011 Massachusetts Institute of Technology. All rights reserved.

Signature of Author.....
Department of Chemical Engineering
May 6, 2011

Certified by
Alan J. Grodzinsky
Professor of Biological Engineering
Thesis Supervisor

Certified by
Paula T. Hammond
Bayer Professor of Biological Engineering
Thesis Supervisor

Certified by
Christine Ortiz
Professor of Materials Science and Engineering
Thesis Supervisor

Accepted by
William M. Deen
Graduate Officer, Department of Chemical Engineering

Molecular Studies of Aggrecan: Experiments and Simulations

by

Fei Liang

Submitted to the Department of Chemical Engineering
on May 6, 2011 in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Chemical Engineering

Abstract

Located in the extracellular matrix (ECM) of the cartilaginous tissues such as human intervertebral disc, the large self-assembling proteoglycans, aggrecan, are essential for the disc in resisting multiaxial compressive loads. Aggrecan degradation, a combinational result of abnormal and/or reduced cellular synthesis and proteolytic cleavage, leads to a reduction in the disc functionality. The knowledge of how aggrecan distributes at different stage of degradation on a molecular level can help reveal both the processes of aggrecan degradation and that of human intervertebral disc degeneration and repair. This thesis will cover both the experimental work and the simulation work in an attempt to fully understand aggrecan structure and the origin of its mechanical properties.

The experimental study utilizes high resolution atomic force microscopy (AFM) to directly visualize the single molecular conformation of aggrecan before and after removal of keratan sulfate (KS) or chondroitin sulfate (CS) glycosaminoglycans (GAGs) via enzymes Chondroitinase ABC (C'ase) and Keratanase II (K'ase) respectively. The aggrecan was extracted and purified from the intervertebral disc of a 24-year old human (obtained from the Shriners' Hospital, Montreal) and separated into two pools: the HA-associated pool is composed of aggrecan proteoglycans that were originally attached to hyaluronic acid (HA) *in vivo*; while the non-HA-associated pool constitutes the aggrecan that lack in G1 domains and were floating around the ECM. The aggrecan solutions were deposited on 3-aminopropyltriethoxysilane functionalized mica substrates and imaged via tapping mode AFM (tip radius <10 nm). The results have shown that aggrecan from normal human disc (24 yr) appears highly degraded, composed mainly of short fragments, and that the non-HA associated molecules are composed primarily of the degradation products and arise because of entrapment in the unique matrix organization of the disc. We can infer that the trace length of aggrecan core protein is extended by GAG chains, specifically by CS GAG which leads to study GAG specifically by Molecular Dynamic Simulations.

It has been found that the sulfation type (4- versus 6-sulfation) and sulfation pattern vary with age in articular cartilage and intervertebral disc. This study developed an implicit solvation model of three types of tetrasaccharides of both sulfated and unsulfated chondroitin, namely, chondroitin 4-sulfate (C4S), chondroitin 6-sulfate (C6S) and chondroitin (CN). In this model, water is treated implicitly as a continuum medium by incorporating its electrostatic screening and viscous effects. The structural analysis on the atomic level based on the temporal dynamics of these tetrasaccharides suggested a closer similarity between 4 and 6-sulfated chondroitin as compared to between sulfated and unsulfated chondroitin. To note, this implicit solvation model provides a consistent result with explicit solvation models of GAGs, indicated by the maintenance of 3-fold helix structure. We have also observed the intramolecular hydrogen bonds that are present intermittently on the sub-nanosecond time scale and cooperate with the glycosidic torsion angles. In future, a coarse-grained model based on this all-atom simulation will provide us with a better understanding of the mechanical properties of aggrecan proteoglycans.

Thesis Supervisor: Alan J. Grodzinsky
Title: Professor of Biological Engineering

Thesis Supervisor: Paula T. Hammond
Title: Bayer Professor of Biological Engineering

Thesis Supervisor: Christine Ortiz
Title: Professor of Materials Science and Engineering

Acknowledgements

My graduate school has not been a smooth one, and I guess it is true for most graduate students here at MIT. I would never come to this point in my life without the tremendous understanding and help I received from everyone around me.

I would like to thank my advisors first: Prof. Paula Hammond, Prof. Alan Grodzinsky and Prof. Christine Ortiz. Prof. Paula Hammond is a great professor with all these smart ideas and positive energies and I really hoped that I have had more interactions with her. Prof. Alan Grodzinsky taught me how to balance life and science and he is a role model for remaining passionate for science in all situations. Prof. Christine Ortiz demonstrates the great courage and how to deal with all different challenges in life, and I am glad to see she is the youngest Dean for Graduate Education at MIT, also as a WOMAN! The discussion with all three of them have brought me to think deeper of the problems and helped to solve the unexpected tasks, and I am really grateful. I also would like to convey my gratitude to Prof. Mark Bathe. He is my thesis committee member for my original doctorate committee, he had many interesting discussions together with me and he generously provided the computational cluster in his lab for my work.

Secondly, I would also thank all the people I have worked with in both Grodzinsky's lab and Ortiz's lab. I have spent most of time in the Grodzinsky's lab, and enjoyed all the lunch gatherings. Linda makes all the logistics possible, and Han-Hwa places orders for us and keeps the lab organized and safe, and I also liked her cakes! Thanks to people who have helped with some part of my experiments first: Lin, Hsu-Yi, Rachel, Paul and Eric. Then also thanks to other members who taught me their research to broaden my knowledge: Emily, Hadi, Ambika, Yi-hong, Yang female and Yang male. In Ortiz's lab, wow, everyone is so smart. As a chemical engineer, I would not have survived/enjoyed the weekly Monday group meetings composed mainly of mechanical engineers without the wonderful research they are discussing, and I would give many thanks to Juha, Steffen, Fevci, Matt, Elaine, Ashley, Lifeng and Haimin.

I would also thank all my friends who have been supporting me greatly throughout this process. Maokai Lin taught me driving and shared with me a lot fun, Jing Chen showed me the calmness, Ermin

Wei showed me how to maintain a child's smile, Jingqing Zhang showed me how industrious one can be, Dahua Lin and Yaodong Zhang are two smart and reliable computer scientists for me to learn all kinds of computer skills, Andrej Košmrlj spent hours with me on the tennis court, Beracah Yankama is a great listener and I enjoyed his laughter and many others who I haven't mentioned here but still share a lot of gratitude towards.

I would also thank everyone in my church, especially Cynthia and Irene. They have been extremely supportive and loving like my family. Coming from a country across the coast, I would have never been able to go through what I have encountered in life without their support, not to mention this thesis. No matter what I have achieved so far or where I will go forward next, I counted all onto God's Grace.

Contents

List of Figures	11
List of Tables	15
1. Introduction	17
1.1 Structure of aggrecan proteoglycans and glycosaminoglycans	17
1.2 Aggrecan's role in human intervertebral disc degeneration	20
1.4 Abbreviations	23
1.5 Thesis overview	24
2. Nanomolecular Architecture of Intervertebral Disc Aggrecan	25
2.1 Materials and methods	25
2.1.1 Aggrecan extraction and enzyme treatment	25
2.1.2 Aggrecan monolayer sample preparation monolayer	26
2.1.3 AFM imaging and post-image processing	27
2.2 Objectives	28
2.3 Results - Conformational properties of healthy intervertebral disc aggrecan	28
2.3.1 Overview of the distribution of the human IVD aggrecan molecule sizes from the biochemistry data and AFM measurement	28
2.3.2 HA-associated vs. Non-HA-associated	30
2.3.3 Effect of enzymatic treatments to remove specific GAG chains	33
2.3.4 Measurement of GAG chains of untreated aggrecan molecules	33
2.3.5 Direct visualization of keratan sulfate chains	35
2.4 Conclusions	36
3. Molecular Dynamic Simulation of Glycosaminoglycans	39
3.1 Literature review and motivation	39
3.1.1 Background and introduction	39
3.1.2 Coarse-grained molecular model of glycosaminoglycans	40

3.1.3 Continuum models.....	41
3.1.4 Molecular theory of tethered polymers to investigate two interacting aggrecan molecules...	43
3.1.5 Proposal of molecular dynamic simulation.....	44
3.2 Objectives	45
3.3 Methods	45
3.3.1 Model Set-up.....	45
3.3.2 Simulation procedures	46
3.3.3 Software and post-processing	47
3.3.4 Computational clusters	47
3.4 Results	47
3.4.1 Cremer-Pople puckering parameters.....	48
3.4.2 Glycosidic torsions	49
3.4.3 Hydrogen bond	51
3.4.4 Helix structure.....	52
3.4 Conclusions.....	52
4. Discussions and Future Directions	55
4.1 Study of disc aggrecan from different age groups.....	55
4.2 Identification of keratan sulfate chains.	55
4.3 Coarse-grained aggrecan model	57
4.3.1 Coarse-graining scheme for GAG chains.....	57
4.3.2 Coarse-graining scheme for aggrecan protein core.....	58
4.4 Inter-chain interactions of neighboring GAGs	59
4.4.1 Motivation and background.....	59
4.4.2 Model set up	60
4.4.3 Preliminary result.....	61
A. Protocol for Guanidine Extraction and Aggrecan Purification	63
B. Protocol for dimethylmethylene blue (DMMB) analysis	67
C. Protocol for Chondroitinase ABC Digestion of Aggrecan	69

D. Nanomechanics of human intervertebral disc	71
E. Explicit solvation model	73
F. LAMMPS input data files	75
Bibliography.....	77

List of Figures

Figure 1-1: The molecular organization of normal articular cartilage. (Cited from Heinegård 2010 [10]).....	18
Figure 1-2: (a) The molecular structure of aggrecan proteoglycans and the chemistry of the side chains, glycosaminoglycans. (b) Aggrecan cleavage sites by aggrecanases and matrix metalloproteinases (MMPs). (cited from reference [11])	19
Figure 1-3: (a) Intervertebral disc is the fibro-cartilaginous tissue between vertebrae in the spinal column, consisting of a hydrated central nucleus pulposus (NP) surrounded by the lamellar collagenous annulus fibrosus (AF). (Cited from reference [17]) (b) A healthy disc provides axial compression, flexion and extension under mechanical loadings. (c) The physiological changes in disc degeneration include decreased water of nucleus pulposus, loss of disc height, distortion of fibers and tears in lamellae. (Picture cited from http://www.coastalneurosurgery.com.au) (d)The involvement of aggrecan in intervertebral disc degeneration and repair (cited from ref. [14])	21
Figure 1-4: Variation in composition of the human nucleus pulposus with age. (Adapted from Roughly 2004[14]).....	22
Figure 2-1: Procedure of aggrecan extraction and enzyme digestion.....	26
Figure 2-2: AFM sample preparation. (Cited from reference [24]) Aggrecan monomers were deposited flatly onto a positively charged mica surface and air-dried.	27
Figure 2-3: Extraction of aggrecan structural parameters: (a) the trace length, L_t , and the end-to-end length, R_{ee} , of the aggrecan core protein; (b) the contour length of GAG chains	28
Figure 2-4: Biochemical analysis by agarose gel electrophoresis of six different specimens of aggrecan molecules extracted from a 24 year old human intervertebral disc (provided by Peter Roughley, Shriner's Hospital, Montreal).	29
Figure 2-5: Histogram of the trace length of aggrecan protein core, L_t , in the HA-associated (a) - (c) and non-HA-associated (d) – (f) pools: (a) and (d) for untreated aggrecan molecules; (b) and (e) for Keratanase treated aggrecan; (d) and (f) for Chondroitinase ABC treated aggrecan. Insets show the AFM height images of individual aggrecan molecules from each aliquot.....	31
Figure 2-6: Comparison of (a) means of trace length L_t , and (b) means of end-to-end distance R_{ee} , of aggrecan protein core measured by AFM imaging (error bars indicating the standard errors. Statistical test: 3-way ANOVA, * = $p < 0.01$)	32
Figure 2-7: Histogram of the trace length, $L_{t,GAG}$, of GAG chains attached to the aggrecan molecules extracted from the 24 year-old human intervertebral disc: (a) HA-associated and untreated; (b) non-HA-associated and untreated; (c) HA-associated and Keratanase treated; (d) non-HA-associated and Keratanase treated.....	34

Figure 2-8: (a) Some biochemistry experiments have already shown that a small portion of the total keratan sulfates were distributed along the chondroitin sulfate enriched region located by the C-terminus of the molecule; (b) AFM height images of C'ase treated and HA-associated aggrecan molecules.	35
Figure 3-1: Disaccharide repeat units of CN, C4S and C6S, alternatively connected with β (1 $\rightarrow$ 3) and β (1 $\rightarrow$ 4) glycosidic linkages.....	40
Figure 3-2: Definition of the coarse-grained model bonded backbone structure (thick solid lines) based on the all-atom disaccharide representation for the (A) β 1 $\rightarrow$ 3 and (B) β 1 $\rightarrow$ 4 linkages. (Cited from reference [36]).....	41
Figure 3-3: Schematic of (a) constant surface charge model, (b) constant volume charge model, and (c) charged rod model in plane parallel geometry. (d-f) Schematic of the models with tip approximated as a hemisphere that were used when comparing to experimental High-Resolution Force Spectroscopy (HRFS) data [40]. (Cited from reference [39]).....	42
Figure 3-4: Schematics of the theoretical model of two adjacent aggrecan molecules employed. D is the distance from the centers of the cylinders. The dots on the polymer chains represent dissociated groups. The radius is not to scale. (Cited from Nap and Szleifer 2008 [44])	43
Figure 3-5: Schematics of Cremer-Pople representation. (Cited from http://www.ric.hi-ho.ne.jp/asfushi/)	48
Figure 3-6: Energy contour plots of β (1 $\rightarrow$ 3) glycosidic torsions.	50
Figure 3-7: Appearance of hydrogen bonds in (a) C4S (b) C6S and (c) CN chains in the implicit solvation model. (d) The hydrogen bond types "A" (N-H-O), "B" (O-H-O), and "C" (O-H-O) are visualized by VMD. (This example is taken from C4S chain dynamics data post equilibration.)	51
Figure 4-1: Thought experiment scheme for identifying keratan sulfates along core protein.	56
Figure 4-2: The AFM height image of Keratanase treated aggrecan molecules extracted from 1 - 2 weeks old calf.....	57
Figure 4-3: Coarse-graining scheme for GAG chains. C_M represents the center of mass of one monosaccharide, C_Q represents the center of charge of one monosaccharide and O is the glycosidic oxygen.	58
Figure 4-4: Schematics of (a) two opposing aggrecan macromolecules undergoing interpenetration and entanglement in vivo; (b) network structure between aggrecan via self-adhesion due to Ca^{2+} -mediated ion-bridging and molecular entanglements between GAG chains, and energy dissipation of the Ca^{2+} -ion-bridges upon mechanically induced GAG molecular elongation (not drawn on scale); and (c) possible hydrogen bonding dashed arrows), hydrophobic interaction (could occur between the methyl groups and carbon rings), and	60

Ca ²⁺ -mediated ion-bridging between CS-GAG chains in the presence of water molecules. Water bridges could exist between the hydrogen bonding donors and acceptors. (Cited from Han <i>et. al.</i> 2008 [33])	60
Figure 4-5: Schematics of the inter-chain model to study the interactions between neighboring GAG chains: (a) the two GAG chains are aligned parallel to each other and (b) the two GAG chains are aligned anti-parallel to each other.	61
Figure 4-5: A test run of SMD simulation of two parallel-aligned C4S chains (disaccharide) at T = 100 K and pull together at centers of mass from 30 Å. (b), (c), (d) and (e) are the VMD visualization of the simulation chains at points B ,C ,D and E in plot (a). (b) When the two chains were far apart, they don't feel each other and the alignment is parallel. (c) The starting point that the two chains interact with each other with a tendency of chain twisting. (d) An attractive force was established by the two chains due to Van der Waals interactions, and the two chains align perpendicular to each other. (d) An high repulsion was observed, and the two planes of the chains were perpendicular to each other.	62
Figure A-1: DMMB analysis of the prepared standard solution and diluted extracted aggrecan solution.	65
Figure D-1: Nanomechanics results of disc aggrecan (24 year old human being, HA-associated, and untreated aggrecan) under the ionic strength (controlled by NaCl concentration) of 0.001 mol/L.	71

List of Tables

Table 2-1: Measured aggrecan structural parameters. (<i>n</i> is the sample number. “C’ase treated” stands for aggrecan treated by Chondroitinase ABC, and “K’ase treated” for aggrecan treated by Keratanase II)	30
Table 3-1: Summary of Cremer-Pople parameters and the percentage of ⁴ C ₁ chair conformation of three types of chondroitin chains. ^a Aqueous simulation of C4S using CHARMM force field (Almond <i>et. al.</i> 1998 [56]). ^b Experimental data of CN using NMR as reported in Protein Data Bank (http://www.pdb.org [57]).	49
Table 3-2: Summary of glycosidic torsion angles of three types of chondroitin chains. ^a Aqueous simulation of C4S using GLYCAM/TIP3 force field (Almond <i>et. al.</i> 2010 [58]). ^b Aqueous simulation of C6S from quantum mechanics (Cilpa <i>et. al.</i> 2010 [59]). ^c Experimental data of CN from NMR from Protein Data Bank.	50

Chapter 1

Introduction

1.1 Structure of aggrecan proteoglycans and glycosaminoglycans

Proteoglycans are glycosaminoglycan (GAG)-containing molecules, and exist extensively in mammalian tissues along with collagen fibers.[1] Aggrecan is the most abundant and important proteoglycan located in the extracellular matrix (ECM) of cartilaginous tissues such as articular cartilage [2] and intervertebral disc (IVD) [3] in human beings along with other mammalian animals (**Error! Reference source not found.**). A “full-length” aggrecan (prior to enzymatic degradation) monomer is cylindrical “bottle brush”-like molecule, consisting of a protein backbone of 220-250 kDa with three globular domains, namely, G1, G2 and G3 domains (Figure). [4] The G1 domain comprises the amino terminus of the core protein, while G3 carboxyl terminus. The “brushes” are negatively charged chains of glycosaminoglycan covalently attached to the serine sites along the protein core between G2 and G3 domains. [5]

Under the physiological conditions, aggrecan can form large aggregates by self-assembly. [6] Hyaluronic acid (HA) acts as a scaffold in aggrecan self-assembling process by binding non-covalently to the G1 domain of aggrecan molecules stabilized by the link protein (LP).[7] The aggrecan/HA aggregates are enmeshed in a network of collagen fibers. This secondary “bottle brush”-like architecture of aggregated aggrecan molecules along the central filament HA is thought to prevent aggrecan monomers from diffusing out of the ECM.

Glycosaminoglycans are unbranched polysaccharides consisting of a repeating disaccharide unit.[1] The class of glycosaminoglycans includes chondroitin 4-sulfate, chondroitin 6-sulfate, keratan sulfate, hyaluronic acid, heparan sulfate and dermatan sulfate etc.[8] Due to the large number of the carboxyl groups and sulfate groups or both, glycosaminoglycans are by nature polyanions. The high negativity of GAG chains is essential to the biochemical and biomechanical properties of the mammalian extracellular matrix, and renders the molecular origin of osmotic pressure inside the tissue.

Along the protein core of aggrecan molecules, the long linear region between the G2 and C-terminal G3 domains is the principal region of the molecule substituted with GAG chains. Chondroitin sulfate (CS, 100 - 150 per aggrecan monomer) is a major component found in aggrecan molecules, and it locates primarily near the C terminus of the aggrecan protein core. Aggrecan chondroitin sulfate has two sulfation types: chondroitin 4-sulfate (C4S) and chondroitin 6-sulfate (C6S), and it has been found that the sulfation type (4- versus 6-sulfation) and sulfation pattern vary with age and disease in human articular cartilage.[9] In contrary to chondroitin sulfate, keratan sulfate (KS, 30 - 60 per aggrecan monomer) is preferentially located towards the N terminus. [6]

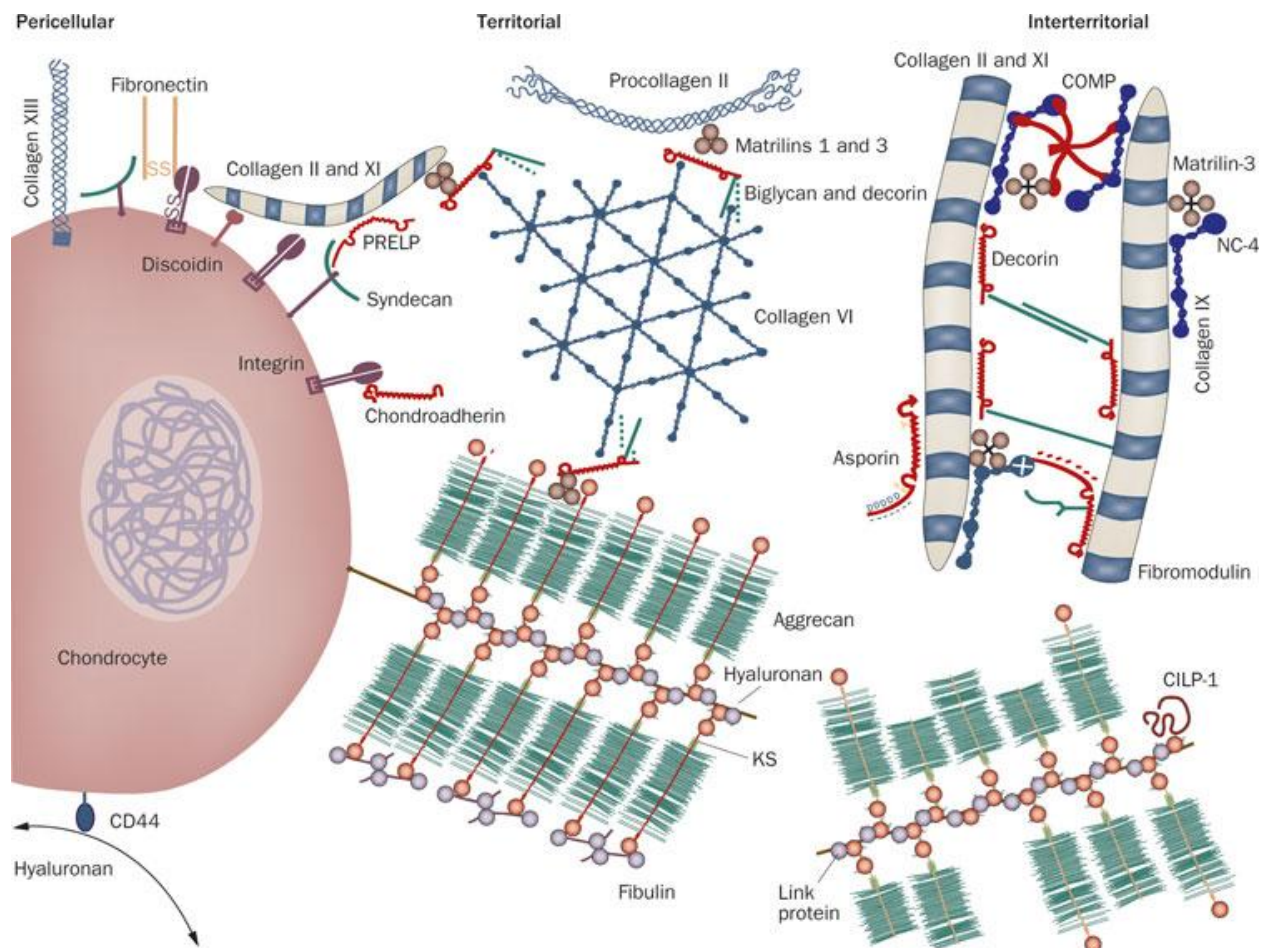


Figure 1-1: The molecular organization of normal articular cartilage. (Cited from Heinegård 2010 [10])

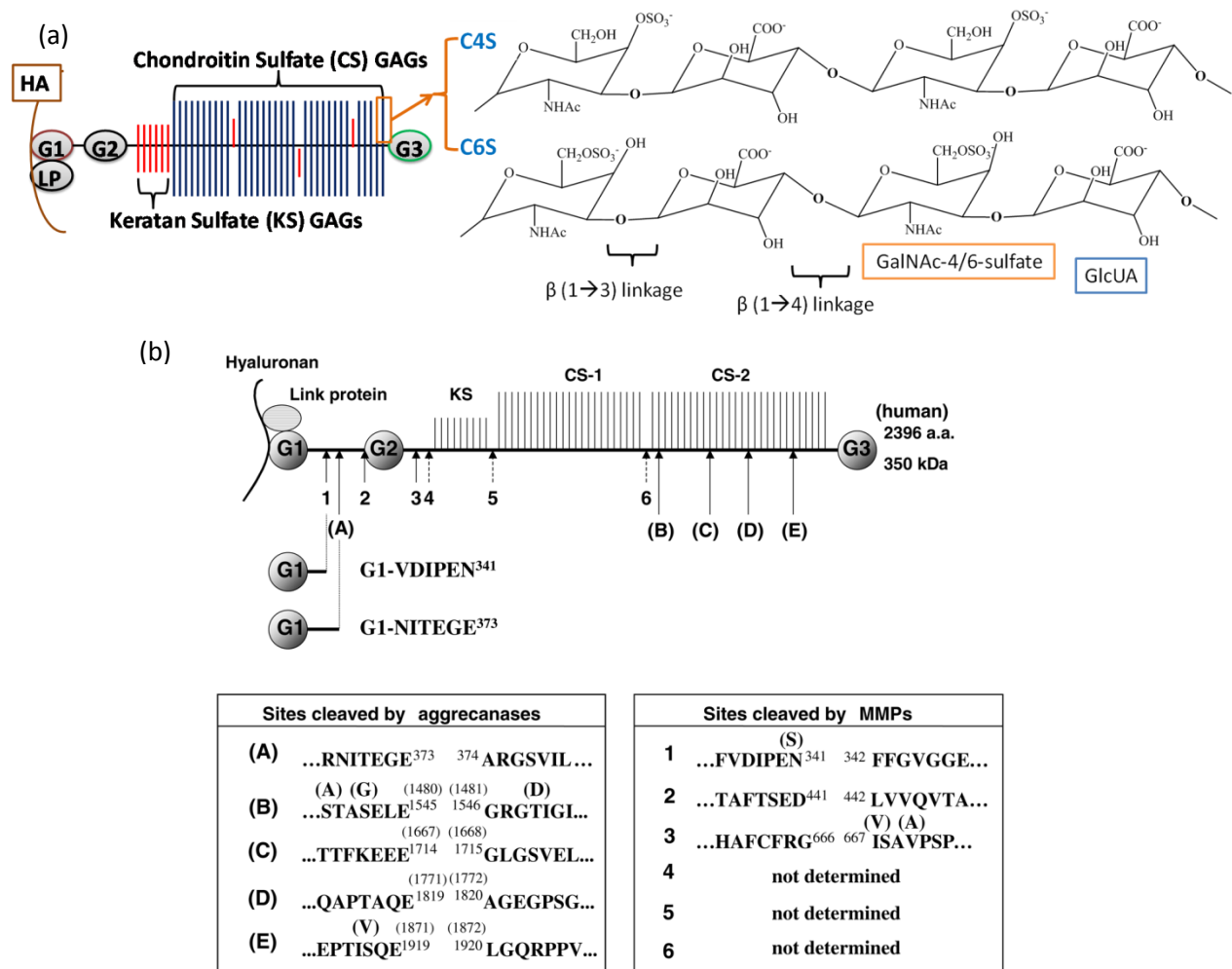


Figure 1-2: (a) The molecular structure of aggrecan proteoglycans and the chemistry of the side chains, glycosaminoglycans. (b) Aggrecan cleavage sites by aggrecanases and matrix metalloproteinases (MMPs). (cited from reference [11])

It has been shown that aggrecan proteoglycans undergo proteolytic cleavage at a number of different sites [12], among which the two most extensively characterized cleavage sites are located within the interglobular domain, the extended region between the G1 and G2 domain, and are targeted by metalloproteinases (MMPs) and aggrecanases, as shown in Figure (b). The degradation processes of aggrecan proteoglycans are thought to be modulated by the mechanical loads. The G1-containing degradation products of aggrecan by metalloproteinases and aggrecanases can still be detained inside the tissue by binding to HA. Despite the importance of both metalloproteinases and aggrecanase in the degradation of aggrecan in human intervertebral disc and articular cartilage *in vivo* [12], other agents may contribute as well. With age and disease, the functional properties of tissues are altered or reduced, and the molecular origin of these aging effects is decreased aggrecan synthesis, aggrecan

depletion from the tissue through diffusion and proteolytic degradation and the accumulation of degradation products that might compete with newly synthesized aggrecan molecules.

1.2 Aggrecan's role in human intervertebral disc degeneration

The role of aggrecan in human articular cartilage has been extensively reported and reviewed [13], and relatively little is known about the aggrecan degradation process in human intervertebral disc. This section will focus on discussing the structure and function of the human intervertebral disc, age or disease related disc degeneration process and the role of aggrecan degradation in disc degeneration.

The intervertebral disc is the fibrocartilaginous tissue consisting of a hydrated central nucleus pulposus (NP) surrounded by the lamellar annulus fibrosus (AF), as shown in Figure 1-. A healthy disc provides axial compression, flexion and extension under mechanical loadings.[14]

Spinal diseases, such as spine stiffness, low back pain and neck pain, are steadily increasing in USA. In the year of 2004, more than 44.6 million patients visited a physician complaining an occurrence of back pain throughout the country.[13] Despite decades of research to identify the causes of disc degeneration, such as declining nutrition, loss of viable cells, accumulation of degraded matrix molecules and post-translation modification of matrix proteins, a fundamental mechanistic understanding of disc degeneration is lacking.[15]

Aggrecan helps the disc to resist compressive loads. Consistently, aggrecan has been identified by far the most abundant proteoglycan on a weight basis among the various proteoglycans in the ECM of the disc.[16] Aggrecan proteoglycan concentration determines the matrix water concentration, and this effect in turn provides the tissue stiffness and resilience to compression. As shown in Figure 1, for adults, the nucleus pulposus is depicted as containing proteoglycan aggregates entrapped in a collagen fiber work. The hydration properties of the GAG chains of aggrecan cause the tissue to swell until equilibrium is reached, in which the swelling potential is balanced by tensile forces in the collagen network. Compression on the spine pushes some water out of the disc which increases the aggrecan concentration and its swelling potential in order to resist further compression. On removal of the compressive load, disc height is restored as water is drawn back and the original equilibrium condition is

restored. The whole process indicates that any parameter that decreases proteoglycan concentration, especially the concentration of aggrecan aggregates will reduce the function of the disc.

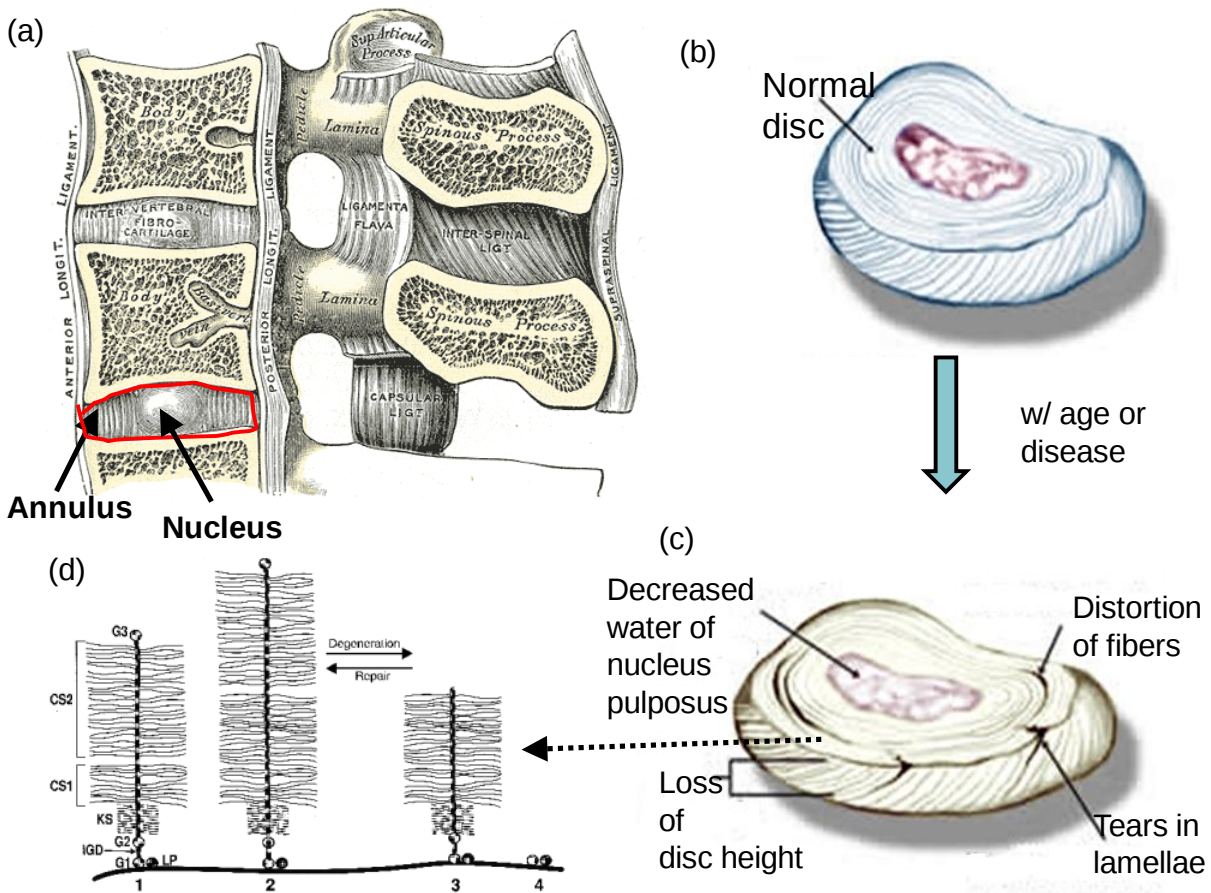


Figure 1-3: (a) Intervertebral disc is the fibro-cartilaginous tissue between vertebrae in the spinal column, consisting of a hydrated central nucleus pulposus (NP) surrounded by the lamellar collagenous annulus fibrosus (AF). (Cited from reference [17]) (b) A healthy disc provides axial compression, flexion and extension under mechanical loadings. (c) The physiological changes in disc degeneration include decreased water of nucleus pulposus, loss of disc height, distortion of fibers and tears in lamellae. (Picture cited from <http://www.coastalneurosurgery.com.au>) (d) The involvement of aggrecan in intervertebral disc degeneration and repair (cited from ref. [14])

The location and structure of aggrecan vary spatially and temporarily.[18] At birth, aggrecan is located mainly in the outer annulus [19] and substituted with chondroitin sulfate. With age, the proteoglycan content of the inner annulus, the portion of keratan sulfate and the length of keratan sulfate increase while the length of chondroitin sulfate decreases (age effect is shown in Figure 1). [14]

The heterogeneity of aggrecan proteoglycans [20] is further enhanced by the appearance of disease. Human IVD degeneration is a combined effect of aging and spinal diseases.

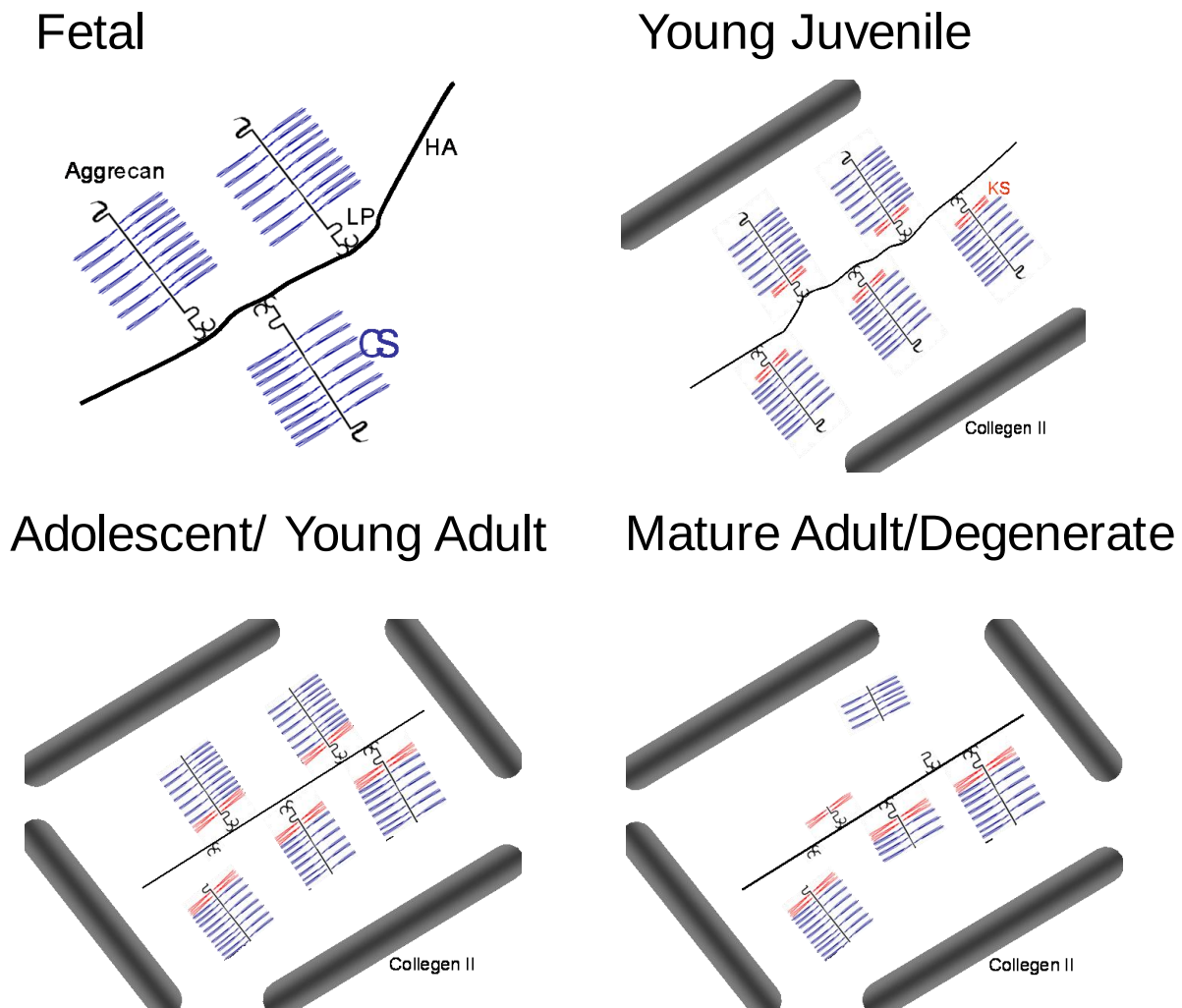


Figure 1-4: Variation in composition of the human nucleus pulposus with age. (Adapted from Roughly 2004[14])

Aggrecan structural degradation, which is thought to be caused by a combination of abnormal and/or reduced cellular synthesis and photolytic cleavage, leads to a reduction in IVD functionality.[21] As showed in Figure 1-(d), it was hypothesized that aggrecan proteoglycans aggregating with HA possess short fragments in the highly charged chondroitin sulfate domain that possess only the CS1 region, the shorter keratan sulfate GAGs, and the interglobular (IGD) domains containing only the binding region G. At the same time, free aggrecan proteoglycans that are proteolytically released from the aggregates become trapped within the tissue; they possess degraded fragments rich in CS2 and G3 domains, but

their functionality remains unknown as they are no longer attached to the aggregates. Based on this hypothesis, my research on the experimental part will focus on directly visualizing the structure of human disc aggrecan using Atomic Force Microscopy to correlate its degradation with disc degeneration.

1.4 Abbreviations

The following is an alphabetical list of abbreviations used throughout this thesis.

APTES	AminoPropylTriEthoxySilane
AF	Annulus Fibrosus
AFM	Atomic Force Microscopy
CN	Chondroitin
CS	Chondroitin Sulphate
C4S	Chondroitin 4-Sulphate
C6S	Chondroitin 6-Sulphate
C'ase	Chondroitinase ABC
ECM	Extracellular Matrix
HA	Hyaluronic Acid
IGD	Interglobular Domain
IVD	Intervertebral Disc
KS	Keratan Sulphate

K'ase	Keratanase
LP	Link Protein
GAG	Glycosaminoglycan
MD	Molecular dynamics
NP	Nucleus Pulposus
VMD	Visual Molecular Dynamics
SMD	Steered Motion Dynamics

1.5 Thesis overview

So far, I have introduced in this chapter the background of the structure of both aggrecan molecules and their constituent glycosaminoglycans, and discussed the important role of aggrecan degradation in disc degeneration. In Chapter 2, I will focus on the experimental studies of human intervertebral disc aggrecan by Atomic Force Microscopy imaging techniques to unveil its ultrastructure on the nano scale and the variation of structural properties by enzyme digestion. In Chapter 3, I will switch the gear to the simulation of glycosaminoglycans and discuss the simulation results which current experimental techniques are incapable of investigating. In the final chapter, I will conclude my thesis work to date, and provide suggestion for future direction in the molecular studies of aggrecan.

Chapter 2

Nanomolecular Architecture of Intervertebral Disc Aggrecan

This study applies Atomic Force Microscopy (AFM) technique to individually visualize the ultrastructure of disc aggrecan proteoglycans extracted from a 24 year-old human being. Conclusions have been drawn by comparing the HA-associated and non-HA-associated aggrecan proteoglycans and different enzyme treatment groups (Untreated, Keratanase treated and Chondroitinase ABC treated aggrecan). This study established the methodology of characterizing human disc aggrecan for further investigation of disc samples from different age groups.

2.1 Materials and methods

2.1.1 Aggrecan extraction and enzyme treatment

The specimens of aggrecan used in this study were provided by our collaborator Peter Roughley (Shriners Hospital for Children, Montreal, QC, Canada).

The aggrecan extraction and enzyme treatment procedures are summarized in Figure 2-1. First, aggrecan monomers were dissociatively extracted (with 4 M guanidinium chloride as described in Appendix A) and purified from the healthy intervertebral disc of a 24-year old human (with no separation of nucleus aggrecan and annulus aggrecan because the previous biochemical studies have shown that there is significant difference between the NP and AF aggrecan for young adults [22]). Then, the extract was supplemented with hyaluronic acid, dialyzed to associative conditions, and the proteoglycans purified by Caesium chloride (CsCl) density gradient centrifugation with A1-preparation (density greater than 1.55 g/ml). The mixture of proteoglycan aggregates and non-aggregated proteoglycans which were separated into two pools by gel filtration chromatography through Sepharose CL-2B: (1) a pool composed of aggrecan or its degradation products that were attached to hyaluronic acid via a functional G1 domain ("HA-associated") and (2) a pool consisting of aggrecan or its degradation products that lacked a functional G1 domains and were not able to interact with hyaluronic

acid and were assumed to be free in the ECM ("non-HA-associated"). Aggrecan in both pools was either left untreated or treated with the enzymes: Keratanase (K'ase) to remove the keratan sulfate GAG chains or Chondroitinase ABC (C'ase) to remove the chondroitin sulfate GAG chains. Aggrecan monomers were purified from all aggrecan digests under dissociative conditions. These purified aggrecan molecules allow me to perform the AFM imaging experiments discussed in the next session to study the individual molecular morphology.

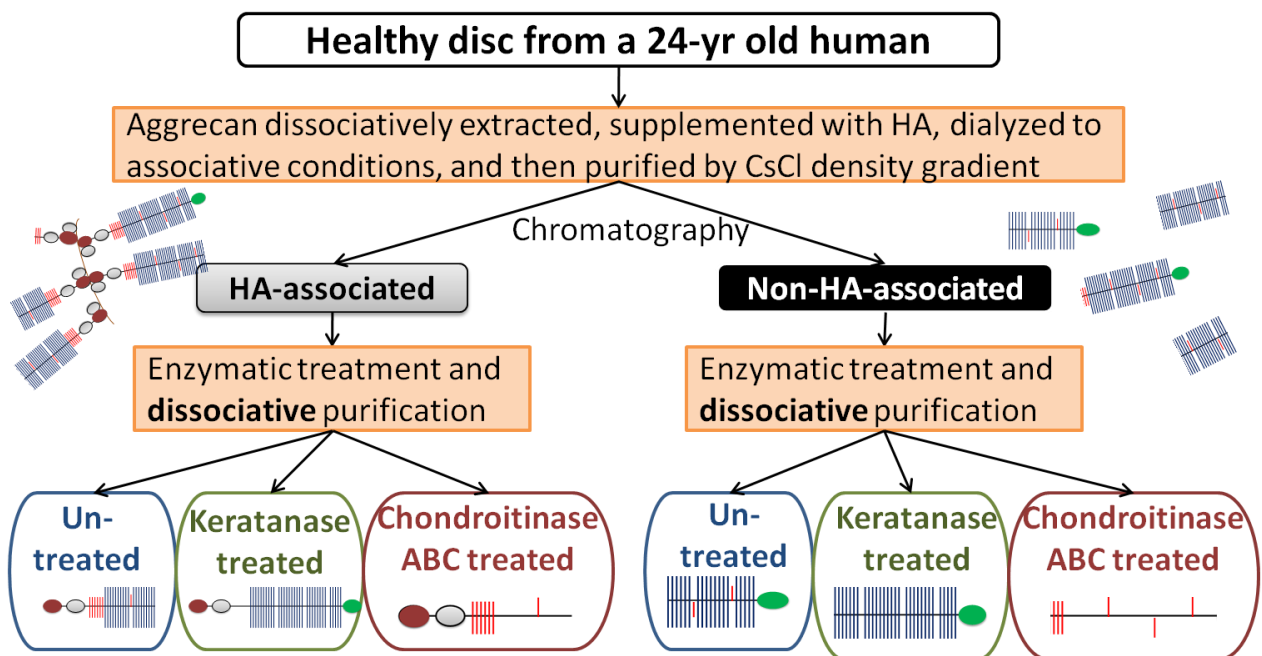


Figure 2-1: Procedure of aggrecan extraction and enzyme digestion.

2.1.2 Aggrecan monolayer sample preparation monolayer

To visualize aggrecan proteoglycans, we using the tapping mode of AFM, the *in vivo* three dimensional aggrecan monomers were deposited on the mica to obtain the two dimensional images. The experiments have followed the steps described by L. Ng and coworkers [23](as illustrated in **Error! Reference source not found.**): (1) muscovite mica surfaces (SPI Supplies, West Chester, PA, #1804 V-5) were treated with 0.01% 3-aminopropyltriethoxysilane (APTES; Sigma Aldrich Co., St. Louis, MO) v/v MilliQ water (18 M-cm resistivity, Purelab Plus UV/UF, US Filter, Low-ell, MA), and this positively-

charged APTES-mica surface facilitated electrostatic binding with the negatively-charged COO^- and SO_3^- groups on the GAG chains to hold the aggrecan non-covalently on the surface to form a monolayer of aggrecan molecules; (2) fifty microliters of aggrecan solution containing of 10 $\mu\text{g}/\text{ml}$ GAG (measured from DMMB) were deposited to the APTES treated mica surfaces for 30 min, and then rinsed with MilliQ water gently for 5 - 10 seconds; (3) after incubated in aggrecan monomer solutions and air-dried, the micas were imaged with the tapping mode of AFM under ambient conditions.

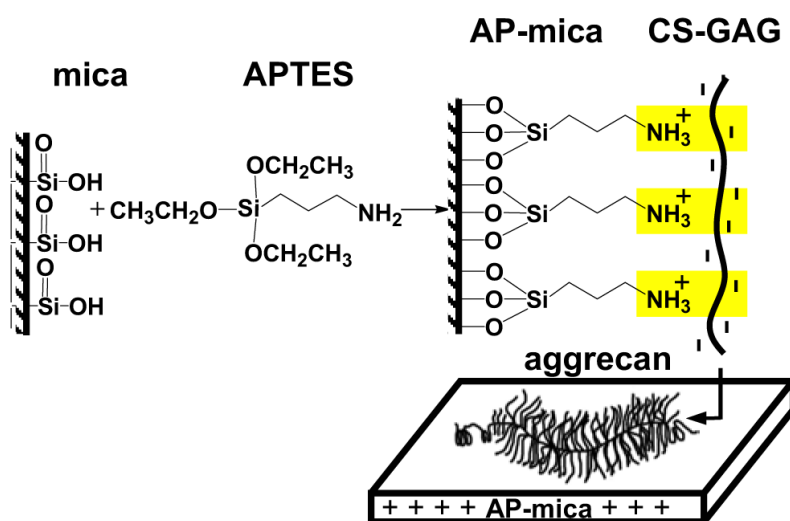


Figure 2-2: AFM sample preparation. (Cited from reference [24]) Aggrecan monomers were deposited flatly onto a positively charged mica surface and air-dried.

2.1.3 AFM imaging and post-image processing

All AFM images were taken by Nanoscope IIIa Multimode AFM (Digital Instruments, Santa Barbara, CA) and the tips used for all imaging are the rectangular Si cantilevers (AC240TS-2, Olympus, $k = 2 \text{ N/m}$, tip radius $< 10 \text{ nm}$). The height and deflection data was recorded and optimized for high resolution of the topology of aggrecan proteoglycans in the sample.

To date, I have directly visualized the individual aggrecan macromolecules from human intervertebral disc *in vitro* using the methodology introduced here. To extract the relevant structural parameters, height images of aggrecan proteoglycans were flattened to the first order for clearer resolution, and then the trace length, L_t , and the end-to-end length, R_{ee} , of the aggrecan core protein were traced by customized Matlab scripts (illustrated in Figure 2-3). Furthermore, an extension value

was calculated by **Error! Reference source not found..** The contour length of GAG chains, were also traced manually using commercial software “SigmaScan Pro.”

Equation 1: —

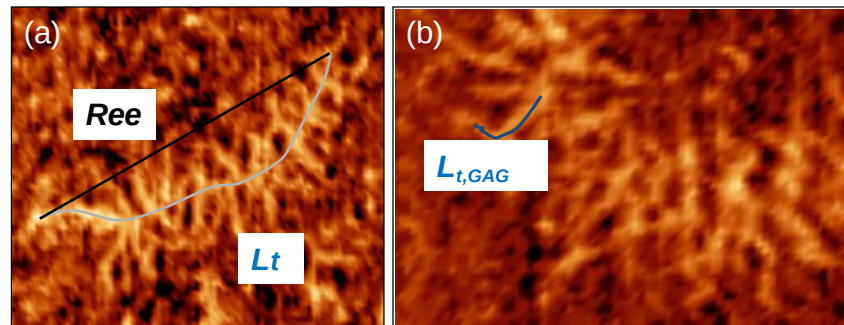


Figure 2-3: Extraction of aggrecan structural parameters: (a) the trace length, L_t , and the end-to-end length, R_{ee} , of the aggrecan core protein; (b) the contour length of GAG chains $L_{t,GAG}$.

2.2 Objectives

The goal of this project is three-folded:

1. Visualize the single molecular structure of aggrecan proteoglycans in human disc and quantify their structural parameters;
2. Investigate the structural difference of aggrecan proteoglycans between the HA-associated and non-HA-associated groups;
3. Study the structural change of aggrecan caused by enzymatic treatments to remove KS or CS GAG chains.

2.3 Results - Conformational properties of healthy intervertebral disc aggrecan

2.3.1 Overview of the distribution of the human IVD aggrecan molecule sizes from the biochemistry data and AFM measurement

Directly after the six specimens of aggrecan proteoglycans were treated and purified, biochemical analysis by agarose gel electrophoresis has been performed on them to get an overview of the composition of aggrecan molecules in the samples (these experiments were performed by Peter

Roughley) and the result is as shown in Figure 2-3. The fine difference in the aggrecan structural compositions among these six specimens is unveiled by AFM imaging discussed in later sections.

From Figure 2-3, we have two main impressions: (1) Non-HA-associated aggrecan molecules contain smaller fragments as a whole compared to the HA-associated aggrecan molecules. (2) The untreated molecules are more heterogeneous in size compared to the K'ase treated or C'ase treated ones in both the HA-associated and non-HA-associated pools. These findings are consistent with the structural properties by AFM imaging summarized in Section 2.2.4, and the details of AFM imaging results will be introduced as followed. Despite the apparent agreement of the data obtained by AFM imaging technique and gel electrophoresis method, AFM imaging technique provides a finer distribution and detailed analysis of aggrecan individual molecular morphology.

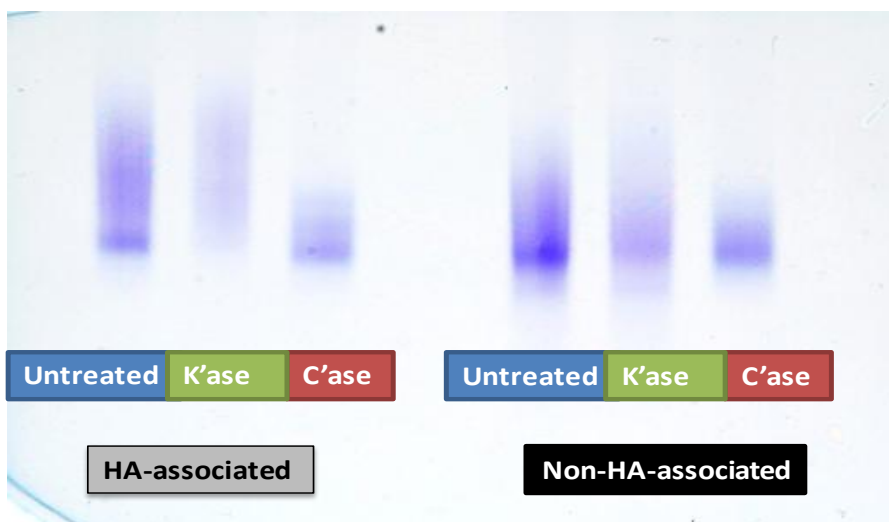


Figure 2-4: Biochemical analysis by agarose gel electrophoresis of six different specimens of aggrecan molecules extracted from a 24 year old human intervertebral disc (provided by Peter Roughley, Shriners' Hospital, Montreal).

From Table 2-1, we can see that human IVD aggrecan on average is much shorter (162.5 ± 10.3 nm, Mean $\pm$ SE) and more heterogeneous than those from the articular cartilage from a similar age group. A previous graduate student in Grodzinsky's lab has imaged and analyzed using the same protocols aggrecan molecules extracted 29 yr-old human knee articular cartilage [25], which study has reported a contour length of the protein core L_t to be 216 ± 10 nm (Mean $\pm$ SE). That study also showed

in human articular cartilage, the average core protein contour length L_t of the full length untreated aggrecan (with both the G1 and G3 domains visible) was measured as 474 ± 56 nm (19 monomers measured). The difference between proteoglycans in articular cartilage and those in the disc has been previously studied by electron microscopy [26], and our result is consistent with their conclusions. If the core protein in both tissues were synthesized to have the exactly same amino acid sequence of the protein core to start with [21], the difference of this distribution between the IVD aggrecan and the articular cartilage aggrecan indicates that the aggrecan in the disc is comparatively more degraded and that the disc has less capability for intrinsic repair and regeneration[27] due to the avascular nature of disc tissues.

Table 2-1: Measured aggrecan structural parameters. (n is the sample number. “C’ase treated” stands for aggrecan treated by Chondroitinase ABC, and “K’ase treated” for aggrecan treated by Keratanase II)

	Sample Type	Sample Size (n)	L_t /nm (Mean $\pm$ SE)	R_{ee} /nm (Mean $\pm$ SE)
HA-associated	Untreated	319	162.5 ± 10.3	42.5 ± 0.8
	K’ase treated	152	166.0 ± 10.0	38.2 ± 1.5
	C’ase treated	112	110.9 ± 7.6	18.8 ± 0.8
Non-HA-associated	Untreated	242	131.4 ± 8.6	76.6 ± 1.1
	K’ase treated	126	144.0 ± 8.6	16.4 ± 0.7
	C’ase treated	130	76.3 ± 5.6	22.3 ± 0.7

The histogram of the trace length of aggrecan protein core, L_t , in the HA-associated and non-HA-associated pools with three subgroups: (1) untreated aggrecan molecules; (2) Keratanase (K’ase) treated aggrecan molecules and (3) Chondroitinase ABC (C’ase) treated aggrecan molecules, is summarized in Figure 2-5. Further, the measurement of both the trace length, L_t , and the end-to-end distance, R_{ee} , from all six specimens are shown in Figure 2-6. The following sessions will focus on discussing the difference between the HA-associated and the non-HA-associated groups (Section 2.1.2) and the effects of enzyme digestion.

2.3.2 HA-associated vs. Non-HA-associated

From this study, non-HA-associated aggrecan was found to be significantly shorter ($p < 0.01$) than the HA-associated molecules based on the measurement of both the trace length, L_t , and the end-to-end distance, R_{ee} . In addition, very few full-length aggrecan molecules (possessing both G1 and G3 domains)

were observed in the disc specimens, further suggesting that the young adult IVD aggrecan is more extensively degraded compared to young adult articular cartilage aggrecan.

As shown in the insets of Figure 2-5 and other AFM images which are not reported here, aggrecan in the HA-associated and non-HA-associated pools have a very similar appearance even before or after K'ase and C'ase treatment. This suggests that many of the non-HA-associated aggrecan molecules are of a similar structure to the HA-associated aggrecan, but lack the G1 domain.

The wide distribution of the molecule size of human IVD aggrecan suggests a high heterogeneity due to the accumulation of aggrecan degradation products in the human disc. The vast accumulation of such degradation products that do not interact with HA does not occur in articular cartilage and arises because of entrapment of fragments in the unique matrix organization of the disc.

Regarding the standard deviations of the lengths L_c and R_{ee} , the non-HA-associated group shows a narrower variance. This may be biological understood as the process of aggrecan degradation is highly and systematically regulated by aggrecanases which primarily target the interglobular domains between the G1 and G2 domains. Once the aggrecan molecules are detached from the HA due to the degradation, they are not subjective to the enzyme degradation anymore.

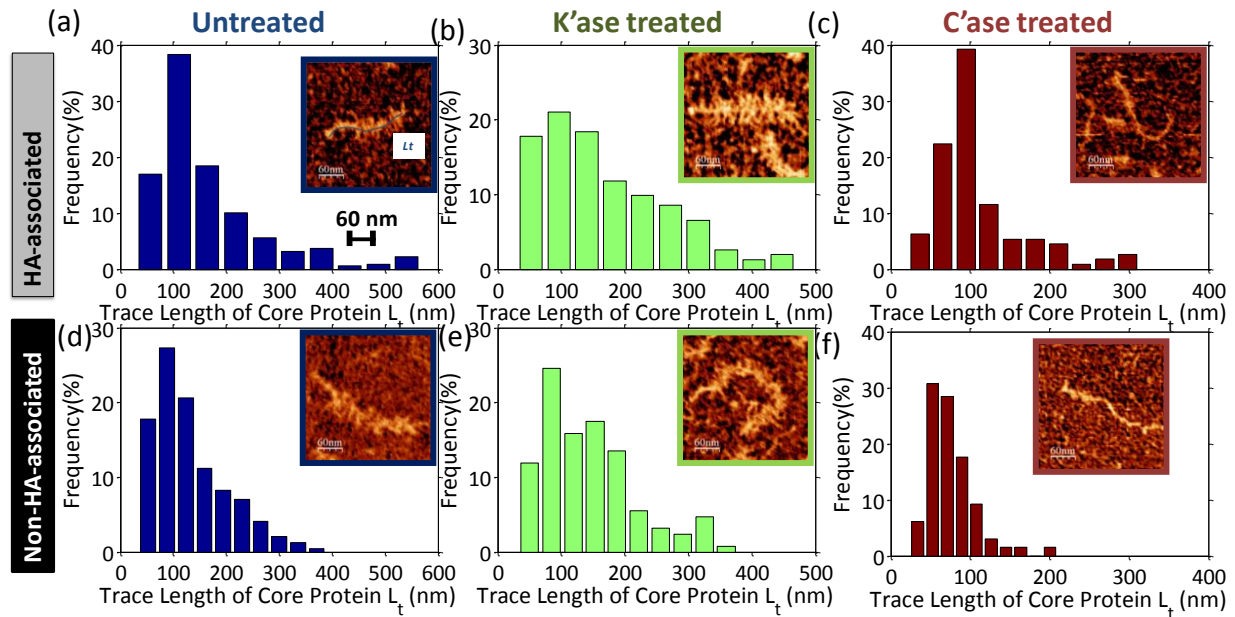


Figure 2-5: Histogram of the trace length of aggrecan protein core, L_c , in the HA-associated (a) - (c) and non-HA-associated (d) - (f) pools: (a) and (d) for untreated aggrecan molecules; (b) and (e) for Keratanase treated

aggrecan; (d) and (f) for Chondroitinase ABC treated aggrecan. Insets show the AFM height images of individual aggrecan molecules from each aliquot.

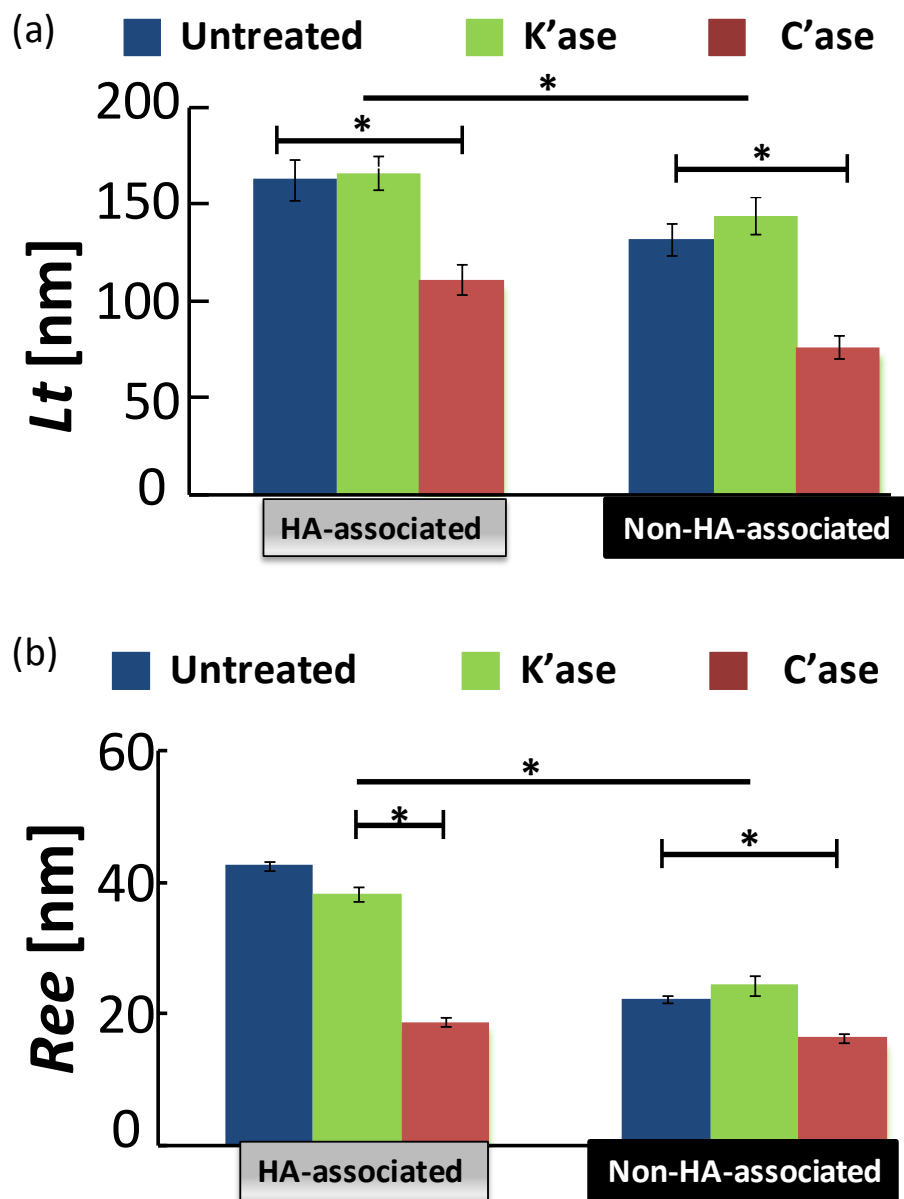


Figure 2-6: Comparison of (a) means of trace length L_t , and (b) means of end-to-end distance R_{ee} , of aggrecan protein core measured by AFM imaging (error bars indicating the standard errors. Statistical test: 3-way ANOVA, $* = p < 0.01$).

2.3.3 Effect of enzymatic treatments to remove specific GAG chains

AFM height images have shown that the core protein of the C'ase treated HA-associated molecules ($L_t = 110.9 \pm 7.6$ nm, $n = 112$) was significantly shorter ($p < 0.01$) than the untreated molecules ($L_t = 162.5 \pm 10.3$ nm, $n = 319$) (Figure 2-5(a) & (c)). On contrary to the C'ase treated aggrecan molecules, the contour length L_t for K'ase treated HA-associated aggrecan ($L_t = 166.0 \pm 10.0$ nm, $n = 152$) was not significantly different from untreated molecules. Similar trends were observed for the non-HA-associated pool (Figure 2-5 (d) – (f)).

These data suggest that the core protein of aggrecan is extended by the presence of the CS GAG chains, presumably due to repulsive electrostatic double layer and steric effects of intra- and/or intermolecular interactions. The variations of L_t among three subsets (untreated, K'ase treated, and C'ase treated) agree with the experimental work previously done on rat chondrosarcoma [28], and both studies have suggested that the only factor of the phenomenological polydispersity in proteoglycans is the variation of chondroitin sulfate chains. This may be due to the elastic stretching of the core protein by the mutual repulsion of the negatively charged CS chains. Physiologically, these electrostatic interactions between the neighboring CS GAG chains have the major contribution to maintain the tissue integrity by maintaining the osmotic pressure of the tissues and absorbing the mechanical shocks, while the interactions between KS GAG chains are comparably much weaker. As such, the specific contribution of KS GAG chains to the mechanical properties of the aggrecan molecules or the tissues as a whole has not been investigated much in the literature, and the biological functions of the KS GAG chains require further studies to be understood.

2.3.4 Measurement of GAG chains of untreated aggrecan molecules

The single molecule imaging by AFM allows us to individually trace the contour lengths of GAG chains without any distinguishment between the CS and KS GAGs. I have measured both the untreated and K'ase treated aggrecan molecules in the 24 year old human intervertebral disc (method illustrated in Figure 2-3(b)), and the results are summarized in Figure 2-7. Since the KS GAGs in C'ase treated aggrecan molecules (Figure 2-8(b)) are both sparse and short, they are not able to be identified individually and thus there is no report here for the GAG trace lengths of C'ase treated aggrecan.

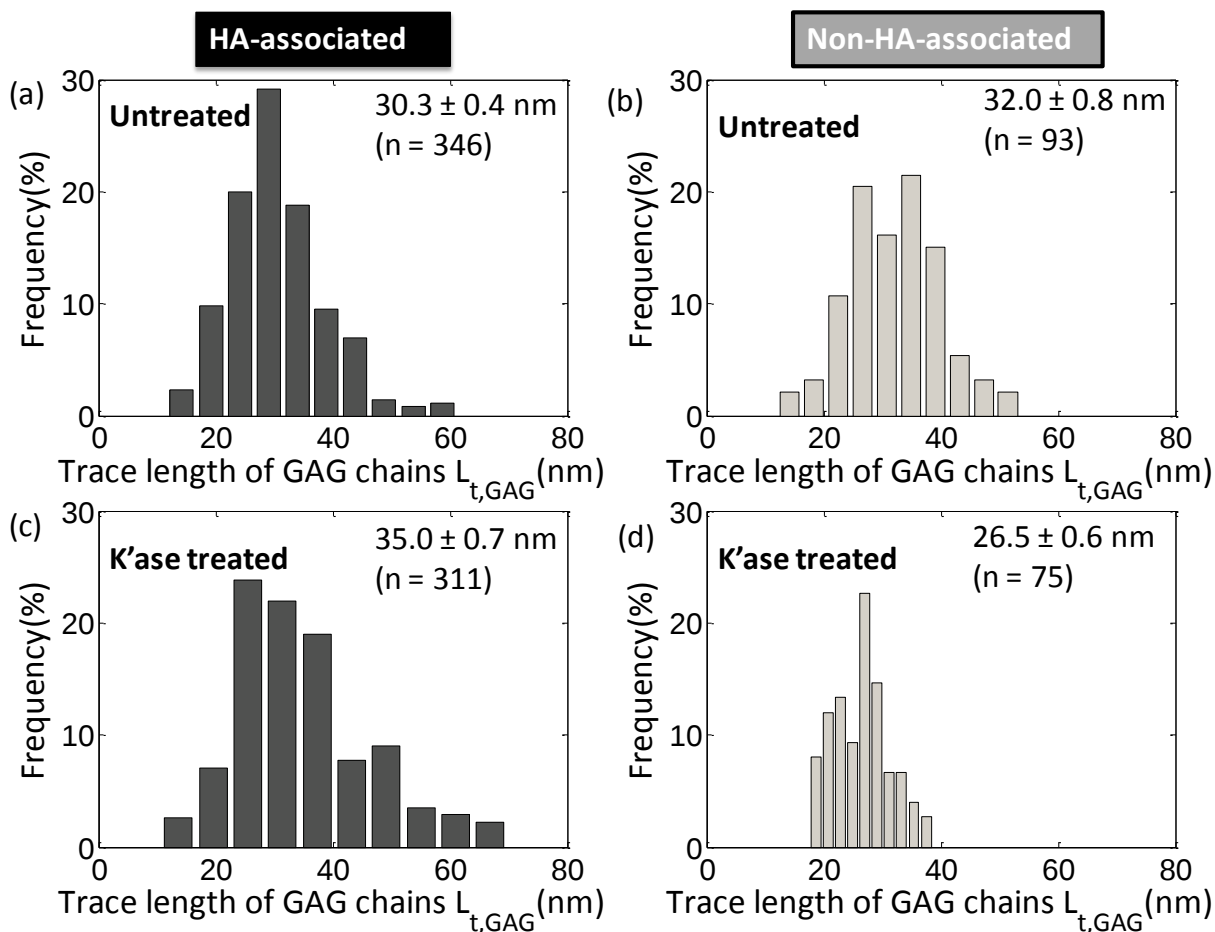


Figure 2-7: Histogram of the trace length, $L_{t,GAG}$, of GAG chains attached to the aggrecan molecules extracted from the 24 year-old human intervertebral disc: (a) HA-associated and untreated; (b) non-HA-associated and untreated; (c) HA-associated and Keratanase treated; (d) non-HA-associated and Keratanase treated.

Comparing subfigures (a) and (b) in Figure 2-7, non-HA-associated and untreated aggrecan molecules have marginally longer GAG chains ($L_{t,GAG} = 30.3 \pm 0.4$ nm), but not significantly different from those from the HA-associated and untreated aggrecan molecules ($L_{t,GAG} = 32.0 \pm 0.8$ nm). Statistical analysis was performed using the student's *t*-test assuming unequal variances and the calculated *p*-value is 0.07. The possible explanation is that HA-associated molecules consist of a higher portion of KS GAG chains, and at the same time, AFM imaging is not capable of distinguishing the “brushes” as KS or CS GAGs as their similar appearance. Since the KS-enriched region is located near the N-terminus of the molecules (G1), it is more probably for the HA-associated aggrecan to retain that region with a concentrated KS GAG attachment, and the KS GAG chains in aggrecan are known to be shorter than

aggrecan CS GAGs. However, the observed trend in the untreated aggrecan molecules is reversed in the K'ase treated aggrecan, and this may be contributed to the fact that all KS chains which may be mistaken as CS chains under AFM image have been removed by enzyme Keratanase.

2.3.5 Direct visualization of keratan sulfate chains

We have observed KS chains in both the KS-enriched region and CS-enriched region, while a previous study has only observed KS-GAGs in the KS-enriched region of C'ase treated aggrecan extracted from 29-yr old human articular cartilage [25]. This may suggest some different biosynthesis or post-translational processing of the intervertebral disc aggrecan compared articular cartilage aggrecan. Due to the limitation of AFM imaging resolution, we proposed another experiment introduced in Section 4.2 to confirm our conclusion that KS GAGs are also located in CS-enriched region along aggrecan protein core.

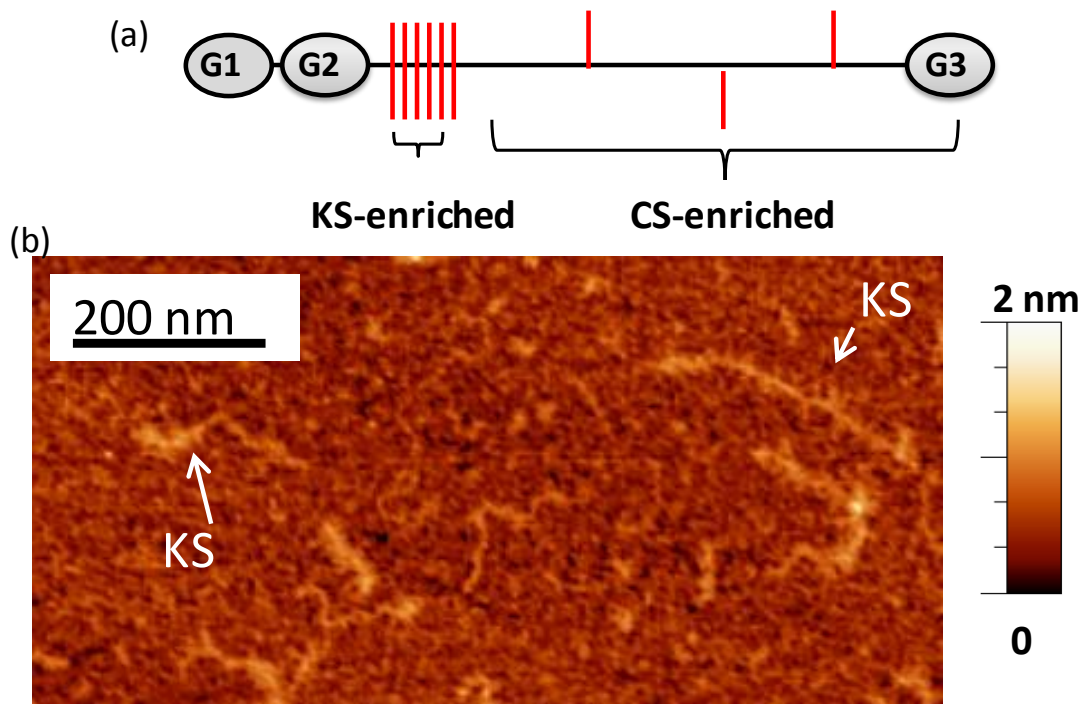


Figure 2-8: (a) Some biochemistry experiments have already shown that a small portion of the total keratan sulfates were distributed along the chondroitin sulfate enriched region located by the C-terminus of the molecule; (b) AFM height images of C'ase treated and HA-associated aggrecan molecules.

2.4 Conclusions

To conclude, I have imaged the single molecular structure of aggrecan from a healthy human intervertebral disc (24 year old), and studied the fine structural difference between the HA-associated and non-HA-associated aggrecan molecules and also investigated the effects of enzyme treatment, Keratanase and Chondroitinase respectively. Compared to the articular cartilage, the disc aggrecan appears highly degraded, composed mainly of short fragments, and this result is consistent with other biochemistry data reported so far [12,22]. The methodology applied in this study can be used in aggrecan extracted from other age groups to have a better mechanistic understanding of the correlation between aggrecan degradation and disc degeneration.

The non-HA-associated molecules are composed primarily of the degradation products and arise because of entrapment in the unique matrix organization of the disc. The human intervertebral disc is confined between the two neighboring vertebral end plates and circumferentially by the outer AF, and it is believed that aggrecan is synthesized in the central NP region and slowly diffuse out to the outer AF region with age. It is assumed that as long as the non-HA-associated aggrecan molecules remain in the disc, they can still aid in resisting the mechanical loadings since the fixed charged density of the tissue is primarily determined by the negatively charged GAG chains. The accumulation of the non-HA-associated aggrecan molecules does not happen in the articular cartilage, because the non-HA-associated aggrecan will diffuse into the synovial fluid and gets lost from the tissue soon after being generated by degradation.

The trace length of aggrecan core protein is extended by GAG chains, specifically by CS GAGs, due to the electrostatic interactions. There is also a possibility that the protein core without GAG substitution will form local knots on a length scale that is shorter than the current AFM imaging resolution. After the removal of attached CS GAG chains (KS GAG chains are fairly short and distributed sparsely), the aggrecan protein core is free from the steric and electrostatic interactions contributed from the GAGs and is essentially just a chain of amino acids. To maximize the entropy, the protein cores tends to go form a rigid-rod-like conformation to a random-coil-like one, and the change of conformation may also be accelerated by the positive charges on the APTES-mica surfaces. In other words, the trace lengths we measured by AFM may be shorter than the real contour length of aggrecan protein cores.

The AFM imaging technique also allows me to individually trace the GAG chain contour lengths. The difference between the non-HA-associated and HA-associated in terms of GAG chain length are not conclusive. It can be understood by the fact that aggrecan degradation process is mainly the proteolysis of the protein core while the attached GAG chains, either KS GAGs or CS GAGs were maintained intact during the aggrecan biosynthesis, transportation and proteolysis processes. The electrostatic interactions between the neighboring GAG chains are the major source of the tissue osmotic pressure under compressive loads. The quantitative analysis of these interactions may be achieved by the method of approach using molecular dynamic simulation discussed in Section 4.4.

Keratan sulfate GAG chains were observed both in the KS-enriched region and CS-enriched region of disc aggrecan. Studies from Roughley *et. al.* 2006 [22] have suggested that KS is localized in the KS domain of human intervertebral disc but no certain conclusion can be drawn from the experimental methods they applied. Currently, it is still not clear whether KS substitute is confined in the KS-enriched domain (near N-terminus) and CS substitute in CS-enriched region. This puzzle gives rise the experimental design proposed in Section 4.2 and hopefully the proposed experiment can address the puzzle one day. Then, the next question to ask is whether the confined or distributed substitution of different sulfated GAG types has any biological functionality or significance, and this may be answered by the coarse-grained aggrecan model proposed in Section 4.3.

Chapter 3

Molecular Dynamic Simulation of Glycosaminoglycans

This chapter will switch the gear from AFM experiments to the simulation part of this thesis. Molecular dynamic simulation is the perfect tool to study the length-scale or time-scale of biomolecules that are not measurable even by the most cutting-edge experimental techniques, such as the AFM imaging techniques introduced in Chapter 2. In this study, an implicit solvation model of three types of tetrasaccharides of both sulfated and unsulfated chondroitin, namely, chondroitin 4-sulfate (C4S), chondroitin 6-sulfate (C6S) and chondroitin (CN) was explored to study the dynamic properties of the single sugar rings, and these data can be built upon for future scale-up studies of aggrecan molecules (Section 4.3 and 4.4).

3.1 Literature review and motivation

3.1.1 Background and introduction

As we have discussed in Chapter 2, glycosaminoglycans (GAGs) play a central role in maintaining the molecular structure of aggrecan molecules and they are also the major source of the tissue's biomechanical properties such as maintaining the osmotic pressure under compression, due to the high negativity of the charges contributed from the sulfate and carboxyl groups in GAG chains.

As illustrated in Figure 3-1, chondroitin is an unbranched polysaccharide consisting of repeating disaccharide units of *D*-glucuronic acid (GlcUA) and *N*-acetyl-*D*-galactosamine (GalNAc), alternatively connected with β (1 $\rightarrow$ 3) and β (1 $\rightarrow$ 4) glycosidic linkages. Chondroitin can be sulfated at either 4- or 6-carbon of the GalNAc residue, denoted as C4S and C6S respectively. In aggrecan, the sulfation types of substituted chondroitin sulfates vary with age and species. For example, in human articular cartilage, aggrecan in the newborn is primarily composed of C4S while the portion of C6S substitution increases with age. [29] The same trend was observed in the other species as well. [30]

In literature, extensive studies have been attempted both experimentally [31-33] and theoretically to study the physical properties of the glycosaminoglycans or aggrecan molecules [34] and their role in extracellular matrix in cartilaginous tissues. Compared to the experimental studies, very little simulation work has been done specifically on aggrecan molecules. In the following sessions, I will first review the two main approaches in to simulate single GAG chains, the molecular approach and the continuum approach, and then introduce the only model in literature on aggrecan molecules. Motivated by the increasing interest in simulating aggrecan so far in literature, I will propose in Section 3.1.5 a Molecular Dynamics (MD) simulation approach to set up coarse-grained aggrecan model.

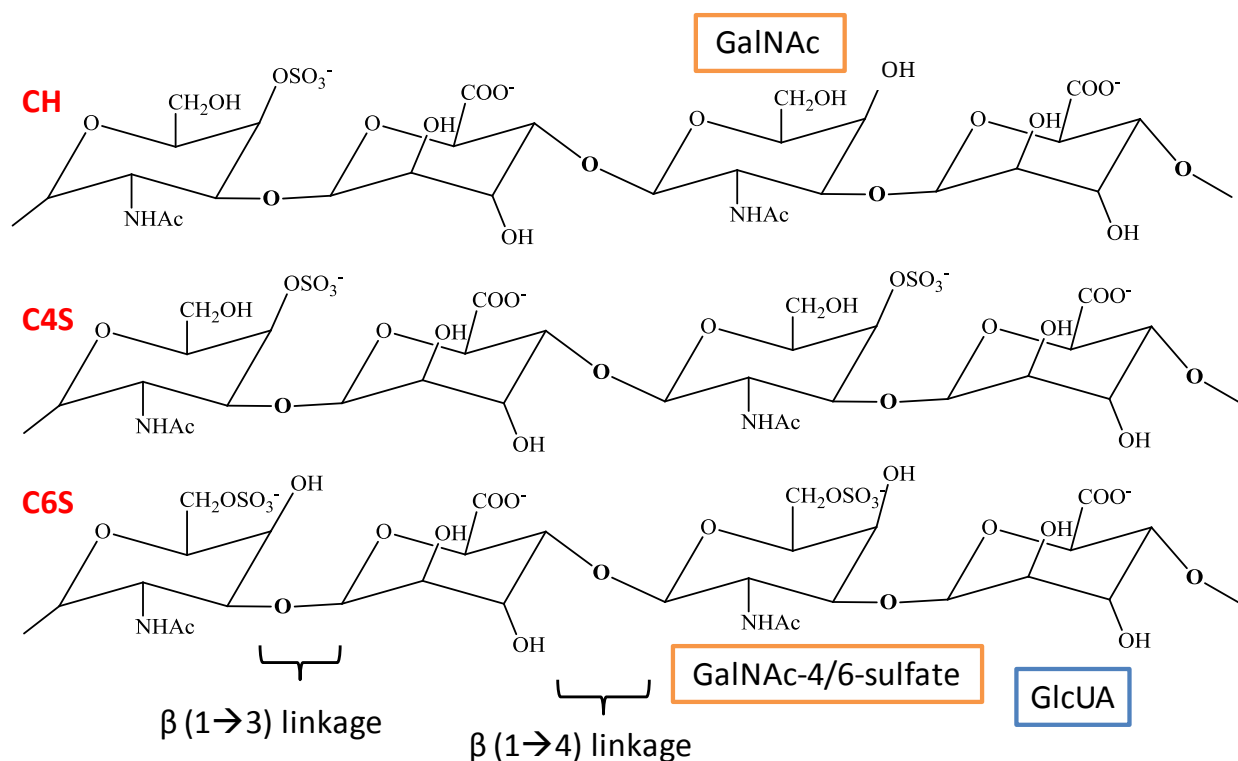


Figure 3-1: Disaccharide repeat units of CN, C4S and C6S, alternatively connected with $\beta(1 \rightarrow 3)$ and $\beta(1 \rightarrow 4)$ glycosidic linkages.

3.1.2 Coarse-grained molecular model of glycosaminoglycans

Bathe and co-workers [35-37] have applied a coarse-grained model for the study of the equilibrium conformation and titration behavior of chondroitin, chondroitin sulfate and hyaluronic acid. In this model, the backbone of GAG chains is coarse-grained into the sequence of chemical and virtual bond

depicted in Figure 3-2. The coarse-graining applied in this model significantly increase the computational efficiency, and this allowed a study of a GAG chain as long as hundreds of repeat units of disaccharides. By varying chains of simulated GAG chains, this study extracted the important properties of GAGs, such as characteristic ratio, persistence length, titration, effect of PH on conformation and also predication of osmotic pressure.

This model has captured most of the molecular structure of GAG chains, and is essentially a freely-jointed chain model. The Monte Carlo simulation results of this model were consistent with the experimental data available and also provided a conceptual insight into the chain conformation and behavioral responses to environment change of four different types of GAGs.

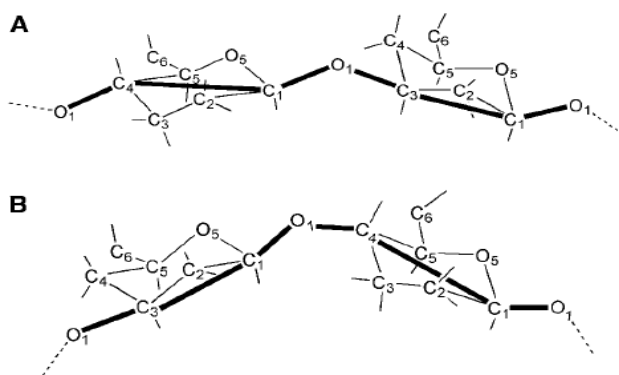


Figure 3-2: Definition of the coarse-grained model bonded backbone structure (thick solid lines) based on the all-atom disaccharide representation for the (A) β 1 $\rightarrow$ 3 and (B) β 1 $\rightarrow$ 4 linkages. (Cited from reference [36])

3.1.3 Continuum models

The continuum modeling approach towards aggrecan dated back two decades ago. Bushman and Grodzinsky [38] compared a macroscopic model based on Donnan equilibrium and a microstructural unit cell model represented by Poisson-Boltzmann (PB) equation, and showed that the latter model is more consistent with experimental data on the swelling pressure of articular cartilage. Dean and coworkers[39] expended the previous PB continuum approach to different geometry representation of the molecular structure (Figure 3-3), and this work was originally motivated to understand the experimental High-Resolution Force Spectroscopy (HRFS) data from the nanomechanic study of opposing layers GAG chains by Seog *et. al.*[40]. The rod model used by Dean and co-workers included

the aspects of polymer molecular geometry and non-uniform molecular charge distribution inside the GAG brushes and gave the best agreement to the available HRFS data.

Since the fixed charged density attributed from the GAG chains is the main source of the swelling pressure of the tissues, the continuum model predicts well the macroscopic behavior such as the repelling forces between the two opposing GAG surfaces. Geometrically speaking, the charged rod model employed by Dean and the co-workers has the most similarity to the molecular level structure of GAGs and also considers the non-uniform charge distribution inside the brush. As such, it provides consistent results with the HRFS data. However, since the set-up of this model was built from the experimental methods used in HRFS, it failed to consider the random orientation of the GAG chains (it assumed that all GAG chains are standing up straight) *in vivo* and is hard to be expanded further to model the dynamic properties of GAGs in the ECM.

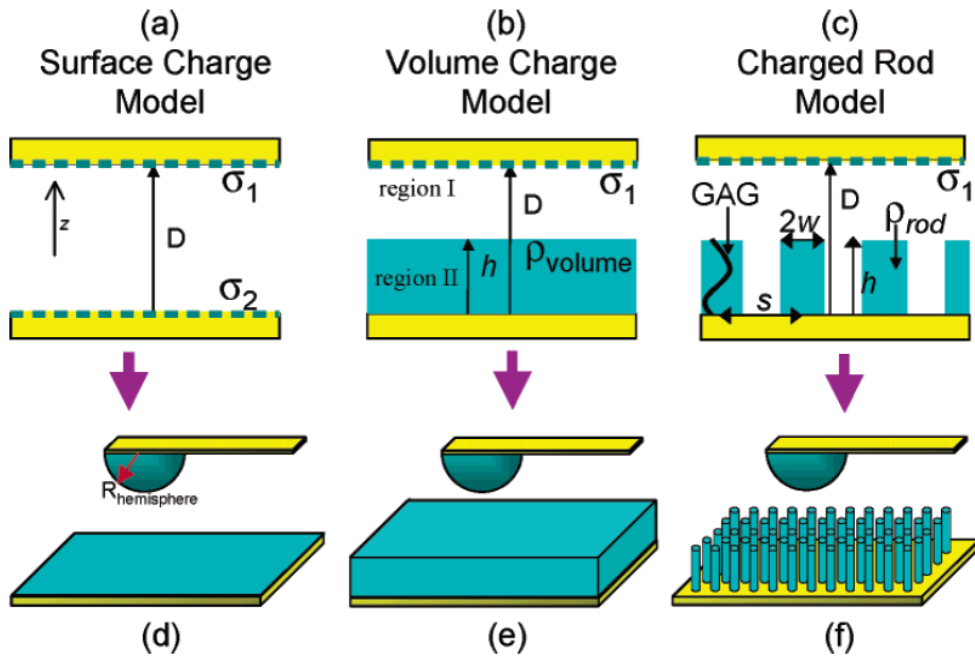


Figure 3-3: Schematic of (a) constant surface charge model, (b) constant volume charge model, and (c) charged rod model in plane parallel geometry. (d-f) Schematic of the models with tip approximated as a hemisphere that were used when comparing to experimental High-Resolution Force Spectroscopy (HRFS) data [40]. (Cited from reference [39])

3.1.4 Molecular theory of tethered polymers to investigate two interacting aggrecan molecules

Very few publications were specially on modeling aggrecan molecules. Though there have been quite a few studies on the bottle-brush copolymers [41-43], setting up a theoretical model of bottle-brush-like molecules with tethered weak polyelectrolytes like aggrecan is a fairly complex problem and computationally expensive. Nap and Szleifer [44] attempted to use a molecular density functional theory to model the two interactional aggrecan proteoglycans. Their work was motivated by the fact that aggrecan molecules were located next to each other by binding to the HA in cartilage with an average spacing of 15 - 40 nm [45], and this model was set up to understand how the interaction between two aggrecan molecules would be affected by the environment (PH and salt concentration) and the distance in between the two molecules.

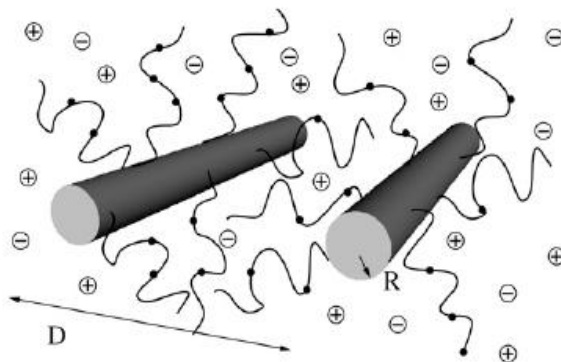


Figure 3-4: Schematics of the theoretical model of two adjacent aggrecan molecules employed. D is the distance from the centers of the cylinders. The dots on the polymer chains represent dissociated groups. The radius is not to scale. (Cited from Nap and Szleifer 2008 [44])

In this model as schemed in Figure 3-4, the protein core of aggrecan was modeled as a long cylinder with a fixed radius $R = 0.5$ nm, and the two rigid cylinders are placed parallel to each other with a certain distance D . The GAG chains were modeled as end-tethered weak polyelectrolytes. The molecular model is accomplished by representing all relevant contribution in a thermodynamic free energy equation and by sampling Monte Carlo algorithm.

The advantage of this model is that it explicitly takes into account the molecular details and the charge distribution of every molecular type including ions and this framework can be expanded to

model other end-tethered polyelectrolytes. On the other hand, Nap and Szleifer's model is an "aggrecan-like molecule" or "simple aggrecan" as they called it, and by simplifying the protein core as the rigid and infinite long rod, the single molecular structure of aggrecan won't be captured in this model.

3.1.5 Proposal of molecular dynamic simulation

No molecular-level based model of a complete aggrecan molecule has been reported in the literature. The complexity of an aggrecan model involves: (1) both the protein core and the attached GAG chains are flexible polymer chains, and the approximation of a rigid rod model won't accurately compute the single molecular conformation of aggrecan; (2) the tethered GAG chains are negatively charged polyelectrolytes, and this electrostatic term in free energy makes the computation fairly expensive; (3) the interest in understanding the molecules' response under compression requires a time-dependent dynamic simulation where a continuum model will fail to account for the transport of water molecules and ions and a model explicitly including all the water and ion molecules is high in computational loads.

Facing the above three challenges in modeling aggrecan, I propose here to simulate a coarse-grained aggrecan molecule by molecular dynamic simulation. Molecular dynamics (MD) is a form of computer simulation in which atoms and molecules are allowed to interact for a period of time by approximations of known physics, giving a view of the motion of the particles. MD simulations have been extensively applied in the theoretical studies of biological molecules to investigate their structures, dynamics and thermodynamics. An all-atom simulation of aggrecan can be computational expensive and unnecessary, and a careful coarse-graining [46] based on some initial all-atom simulation can dramatically increase the computational efficiency without losing any relative molecular details.

A question one can ask for this modeling scheme is that how much aggrecan can be coarse-grained in terms of length-scale. We can roughly estimate the length scale based on the work by Bathe *et. al.* described in Section 3.1.2. The intrinsic persistent length of GAG chains were found to be 7 - 10 nm, and the stretch of one disaccharide unit is typically around 1 nm. The comparison of these two numbers suggest that we can coarse grain a whole sugar ring as one entity and this will not have much effects on simulating the chain conformation.

The end goal of modeling aggrecan by MD will be accomplished by three steps. First, an all-atom simulation of tetrasaccharides (five repeating units of disaccharides) is performed, and the choice of a tetrasaccharides is that the length of the simulated chain is on the order of its persistent length. I didn't directly borrow the all-atom simulation results from Bathe *et. al.* (2005) since his coarse-graining is based on an all-atom simulation of one disaccharides and the simulated chain is not long enough to be further coarse-gained than reported. Second, the free energies of essential interactions are extracted as the basis for future coarse-grained modeling of aggrecan proteoglycans. Third, an aggrecan model is set up by coarse-graining every sugar ring into two "atoms", one representing the center of mass and one representing the center of charge (more detailed in Section 4.3). This thesis has accomplished the first step.

3.2 Objectives

In this chapter, I will report the MD simulation I have done so far on tetrasaccharides of chondroitin and chondroitin sulfates and the specific goals of this work are as following:

1. To study the conformation of both the unsulfated chondroitin (CN) and chondroitin sulfates (both C4S and C6S) at the all-atom level;
2. To investigate the applicability of the implicit solvation model to negatively charged polysaccharides such as aggrecan proteoglycans;
3. To extract the free energies of essential interactions as the basis for future coarse-grained modeling of aggrecan proteoglycans.

3.3 Methods

3.3.1 Model Set-up

Molecular dynamics simulations were performed on the tetrasaccharides of unsulfated condroitin (CN), 4-sulfated chondroitin (C4S) and 6-sulfated chondroitin (C6S) using AMBER force field (or GLYCAM06 [47] since our simulation system is comprised solely of carbohydrate). The original conformation was adapted from the Protein Data Bank (ID: 2KQO). A tetrasaccharides chain of C4S contains 248 atoms, 257 bonds, 476 angles and 853 dihedrals.

Atom-based Van der Waals and electrostatic interactions were calculated with a cutoff of 12 Å. An additional $1/r$ term is included in the Coulombic formula, thus the Coulombic energy varies as $1/r^2$. This is effectively a distance-dependent dielectric term which is a simple model for an implicit solvent with additional screening.[48] A minimum distance of 10 Å was kept between the chondroitin molecules and the edge of simulation box.[49] Periodic boundary conditions were applied. The partial charges for each atom was determined by Gaussian 03W (license owned by Chemical Engineering Department @MIT) using HF6-31g(d) basis. The dielectric constant was set to be 80 as water in room temperature.

Here, “Amber force field” refers to the functional form used by the family of AMBER force fields. This form includes a number of parameters; each member of the family of AMBER force fields provides values for these parameters and has its own name. The functional form of a general AMBER force field is summarized in Equation 3-1 with the four terms calculating the bond, the angle, the dihedral and the pair-wise interactions respectively. The subset for carbohydrates of AMBER force field is the also known as GLYCAM force field [47] (where I got the input parameters). AMBER was chosen specifically because it is reported to be more appropriate for protein-carbohydrate systems.[50] Aggrecan, the molecule of our research interest, is made up with a protein core and carbohydrates as the side chains, and in future one can easily add the protein core segments to the current model to build up the aggrecan molecules without change the force fields applied.

Equation 3-1:

$$E_{\text{total}} = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

3.3.2 Simulation procedures

Initial coordinates of tetrasaccharides were minimized using the steepest descent algorithm followed by Polak–Ribiere version of the conjugate gradient algorithm. After energy minimization, the temperature was gradually warmed up over a period of 200 ps from 0 K to 300 K by applying a Langevin thermostat to model an implicit solvent. Finally, the system would be equilibrated at 300 K in Langevin dynamics

after 500 ps. Coordinates of all atoms were written every 50 fs while an integration step was taken every 1 fs. The dynamics trajectory of 1 ns post equilibration (20000 sample data points in total) was used for data analysis.

3.3.3 Software and post-processing

All simulations were conducted using the open source codes LAMMPS [51], a molecular dynamic simulator with parallel communication. Conformation information was extracted by customized MATLAB scripts and/or Python scripts.

The visualization of molecules cited in this work was realized by Visual Molecular Dynamics (VMD), a molecular modeling and visualization computer program. [52]

3.3.4 Computational clusters

The initial work was performed on the High Performance Computing (HPC) core invested by Computational and systems biology (CSBi) at MIT with permission by Mark Bathe (mark.bathe@mit.edu). The cluster is located on the first floor of building NE47 at MIT (500 Tech Sq, Cambridge, MA, 02139) and has one head node, thirty-two compute nodes and four CPUs per node. The cluster encountered a fatal crash in June 2010. Most simulation done on this cluster was explicit solvation models [53], i.e. the water molecules were accounted explicitly.

Since June 2010, I have transferred my simulation work from HPC core to the computational cluster located on the fourth floor of building NE47. This computational cluster has only five compute nodes with 2 processors each, and the computational capacity is incomparable to HPC core. Thus, I was limited to perform the implicit solvation model only. I used one month to set up the MPI and Torque modules on this cluster to enable the parallel computing [54] and job queuing system. I will spare the cluster set-up procedures from this thesis since it is not directly related to my research goal, but feel free to contact me if you want a detailed procedure for your future reference.

3.4 Results

As described in the previous section, the dynamics trajectory of 1 ns post equilibration was recorded and analyzed to capture the detailed conformation of GAG chains on the molecular level. The

3.4.1 Cremer-Pople puckering parameters

The simulation result of cremer-pople parameters are summarized in Table 3-1 and compared with simulation results in literature (as in gray shading). Most rings in both unsulfated and sulfated chondroitin models are in 4C_1 chair conformation (defined as the azimuthal angle ϑ is smaller than 60°),

and chondroitin sulfates have more stable ring conformation than unsulfated chondroitin as shown in a higher occupation in the chair conformation over the sample space. The stability of the ring conformation is the basis of future coarse-graining, and if the ring puckers between different conformations such as chair and boat, the center of mass of the whole ring would jump in space which makes it hard to use a traditional translational mode to represent such movements.

Table 3-1: Summary of Cremer-Pople parameters and the percentage of 4C_1 chair conformation of three types of chondroitin chains. ^aAqueous simulation of C4S using CHARMM force field (Almond *et. al.* 1998 [56]).

^bExperimental data of CN using NMR as reported in Protein Data Bank (<http://www.pdb.org> [57]).

		C4S	C6S	CN	C4S ^a	CN ^b
GalNAc4S	Q	0.67	0.61	0.66	0.58	0.55
/ GalNAc6	ϑ	15.75	9.83	14.12	8.78	1.89
/GalNAc	4C_1	100%	100%	76.02%	100%	100%
GlcUA	Q	0.58	0.58	0.59	0.59	0.55
	ϑ	11.07	12.59	11.85	8.80	6.86
	4C_1	99.99%	98.57%	92.32%	100%	100%

3.4.2 Glycosidic torsions

Glycosidic bonds are formed in between the neighboring monosacchrides in chondroitin, and the dihedral angles to describe the glycosidic bonds are terms as glycosidic torsions. An analysis of the glycosidic torsions in GAG chains can help us quantitatively define how the monossacchrides are linked. The conformational exploration of disaccharide β (1 $\rightarrow$ 3) and β (1 $\rightarrow$ 4) glycosidic linkages in all three types of tetrasaccharides are summarized in Table 3-2. The contour plot of the β (1 $\rightarrow$ 3) glycosidic torsion energy is shown in Figure 3-6, and the contour plot of the β (1 $\rightarrow$ 4) glycosidic torsion energy shares the similar feature. This contour plot is obtained by splitting the dihedral angle space into small bins, and then calculating the probability of pairs of dihedral angles occurring at different grid points and the energy in each bin based on the logarithm of the probability.

Table 3-2: Summary of glycosidic torsion angles of three types of chondroitin chains. ^a Aqueous simulation of C4S using GLYCAM/TIP3 force field (Almond *et. al.* 2010 [58]). ^b Aqueous simulation of C6S from quantum mechanics (Cilpa *et. al.* 2010 [59]). ^c Experimental data of CN from NMR from Protein Data Bank.

	C4S	C6S	CN	C4S ^a	C6S ^b	CN ^c
$\varphi(\text{H1-C1-O1-C3})$	55 ± 18	51 ± 21	18 ± 76	40 ± 17		48
$\psi(\text{C1-O1-C3-H3})$	18 ± 60	-22 ± 63	51 ± 76	-11 ± 30	N/A	-12
$\varphi(\text{H1-C1-O1-C4})$	42 ± 15	45 ± 13	-69 ± 15			45
$\psi(\text{C1-O1-C4-H4})$	-14 ± 101	-15 ± 65	-25 ± 91	N/A	N/A	0
$\varphi(\text{O5-C1-O1-C4})$	-66 ± 18	-71 ± 21	-45 ± 69	-79 ± 16	-109	-73
$\psi(\text{C1-O1-C3-C4})$	48 ± 110	66 ± 88	10 ± 76	106 ± 29	110	106

Sulfated chondroitin chains share similar torsion angles compared to unsulfated chondroitin. This can be understood by the fact that hydrogen bonds were constantly forming intermittently between the neighboring monosaccharides as we will see in the following section. The distribution of the glycosidic torsions in the sulfated chondroitin is narrower compared to the unsulfated chondroitin, and it may be because the hydrogen bonds formed in the sulfated chondroitin helped to stabilize the chain structure. As such, the freedom of the torsion angle is more confined in the sulfated chondroitin chains.

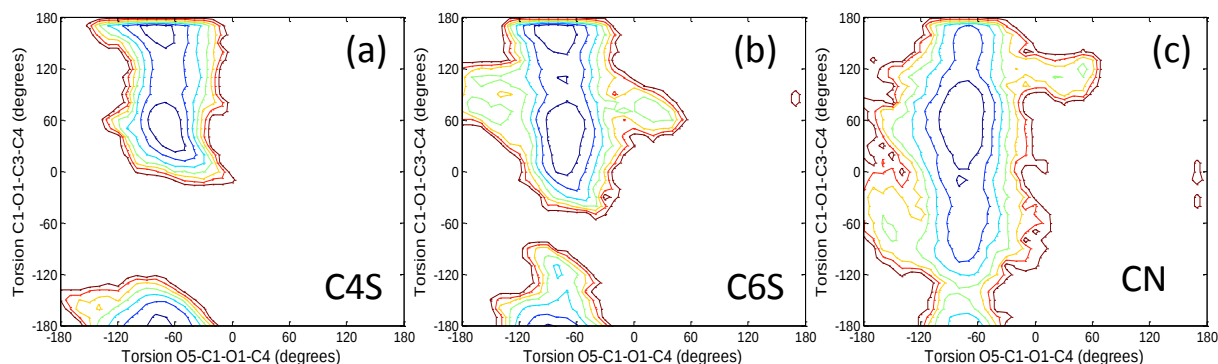


Figure 3-6: Energy contour plots of β (1 $\rightarrow$ 3) glycosidic torsions.

3.4.3 Hydrogen bond

Hydrogen bonds were formed between neighboring monosacchrides along the chain. To quantify the presence of hydrogen bonds, the distances and the bond angles were calculated between all hydrogen donors and acceptors, and a hydrogen bond is formed if the distance $D-A < 3.5 \text{ \AA}$ and the angle $A-D-H < 30^\circ$ (where D is the hydrogen donor, A is the acceptor and H is the hydrogen atom).

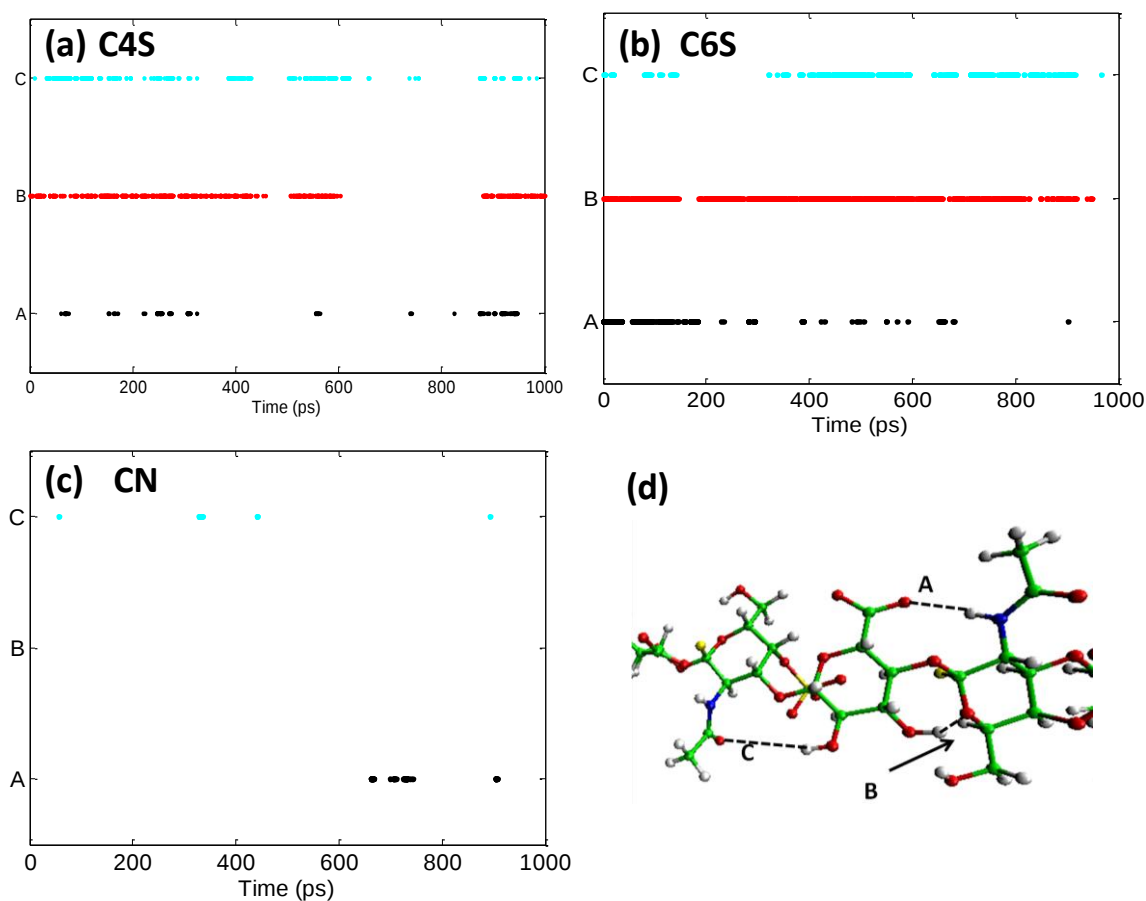


Figure 3-7: Appearance of hydrogen bonds in (a) C4S (b) C6S and (c) CN chains in the implicit solvation model. (d) The hydrogen bond types “A” (N-H-O), “B” (O-H-O), and “C” (O-H-O) are visualized by VMD. (This example is taken from C4S chain dynamics data post equilibration.)

Interestingly, previously reported hydrogen bonds in unsulfated chondroitin [58] were observed rarely in this implicit solvation model. The possible reason is that CN is less negatively charged (the total charge of a tetrasaccharide chain of CN is negative five while it is negative 10 for sulfated tetrasaccharides), and this fact may render the implicit solvation model less applicable.

For both 4- and 6-sulfated chondroitin chains, the hydrogen bonds were formed and broken intermittently. As illustrated in Figure 3-4, there are basically three types of hydrogen bonds: N-H-O, O-H-O, and O-H-O. Once the hydrogen bonds are formed, they stay bonded on the order of hundreds of picoseconds and the hydrogen bonds persists 60 - 80% of the total simulation time. The long-lasting hydrogen bonds formed in sulfated chondroitin, in addition to other chemical bonds, help to define the glycosidic torsion and as such the helix structure.

3.4.4 Helix structure

The helix structure of chondroitin was mainly maintained by the intra-residual hydrogen bonds across the glycosidic linkages. The helix structure of an elongated polymer chain, such as GAG chains can be quantified by applying the parameters of the helical fold, n , representing the number of disaccharides per turn and the helical rise, h , representing the axial rise per disaccharide.

In this study, the helical parameters of n and h were calculated using a general method described by Sugeta & Miyazawa 1967 [60]. Previous studies have shown that the sulfated chondroitin has a left-handed helix structure with $n = 3$ [56] or $n = 4$ [59]. Our implicit solvation model agrees with the previous studies by having an average of $n = 2.62$ for C4S, $n = 3.63$ for C6S and $n = 3.75$ for CN. The helical rise, h , varies from 7 to 11 Å over the dynamic simulation processes.

3.4 Conclusions

I have successfully set up a molecular dynamics model for the tetrasaccharides of the both the unsulfated and sulfated chondroitin. The dynamic trajectory produced by the simulation can be used to quantitatively analyze the single ring structure on the length scale of less than 1 nm.

In terms of all the single ring structural parameters I investigated in this chapter (Cremer-Pople puckering parameters, glycosidic torsions, hydrogen bond, and helix structure), the simulation results

provided by the introduced implicit solvation models agrees with both the available experimental data and explicit solvation models (Appendix E) to study highly negatively charged polysaccharides, namely 4- and 6-sulfated chondroitin by preserving the ring conformation of monosaccharides and the intermolecular hydrogen bonds. Similarity of both the monosaccharide ring conformation and the overall tetrasaccharide conformation has been observed between 4 and 6- sulfated chondroitin. Implicit solvation model can provide accurate results for highly charged system, and this result can provide some guidance for other simulation works of choosing between the implicit and explicit solvation models while the latter is highly computational expensive.

The current model can be served as the first step for the end goal of modeling aggrecan molecules. I will discuss in the following chapter more details of the possible future work: (1) Section 4.3 will discuss how to build a coarse-grained aggrecan model based on the current all-atom model, and the MD simulation has the advantage of investigating the dynamic responses of aggrecan upon external forces (to simulate the compression in the ECM); (2) Section 4.4 will introduce the methodology of exploring the inter-chain interactions between the neighboring GAG chains, and this may help us to understand the GAG spacing in nature (~ 4 nm [24]).

Chapter 4

Discussions and Future Directions

4.1 Study of disc aggrecan from different age groups

The study discussed in Chapter 2 has paved the approach to understand disc degeneration using the state-of-art nano-imaging technology. This kind of multidisciplinary approach to science problems is a general trend for a lot of research studies nowadays. The same experimental methods and statistical analysis combining the biochemistry, nanomechanics and statistical analysis can be applied to human disc aggrecan sample from other age group and disease groups. Our collaborator Peter Roughley was kind enough to provide some of these samples (a 37-year old human disc sample and a 15-year old human disc sample), while a more feasible direction for someone who wants to study disc degeneration can start with the relatively abundant porcine aggrecan.

4.2 Identification of keratan sulfate chains.

It was agreed among biologists that the keratan sulfate enriched region is located in the proximity of the N-terminus of aggrecan, but there lacked a quantitative analysis of the distribution of the keratan sulfates along the chondroitin sulfate enriched region located near the C-terminus of the molecule. Some biochemistry experiments have already shown that a small portion of the total keratan sulfates were distributed along the chondroitin sulfate enriched region located in the C-terminus of the molecule [61], which other biologists hold skepticism over. Our AFM image capable of resolving the single molecular structure of Chondroitinase ABC treated aggrecan molecules may aid in the understanding of the binding and the function of keratan sulfates. The difficulty lies in distinguishing keratan sulfate and degraded chondroitin sulfate based on the fact that both are of short lengths.

Our thought experiment, illustrated in Figure 4-1, is designed as an attempt to identify keratan sulfates specifically for the purpose of finalizing the debate over their distribution along aggrecan core protein. Firstly, we will digest the aggrecan proteoglycans (mainly “full-length”, obtained from calf

cartilage) Chondroitinase ABC to remove all the CS GAGs. Then, we will incubate the Chondroitinase ABC treated aggrecan molecules with KS antibodies, 5D4, provided by our collaborator Bruce Caterson [62]. The KS antibodies will specifically bind to the keratan sulfates, and this enables us to localize their distribution by visualizing little bulbs under AFM.

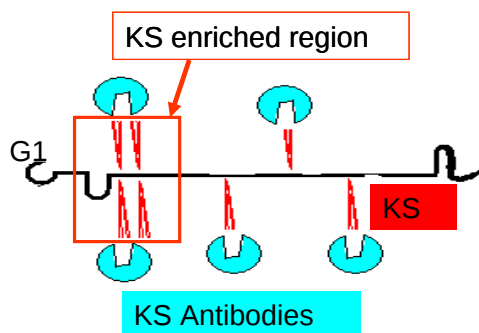


Figure 4-1: Thought experiment scheme for identifying keratan sulfates along core protein.

I have extracted aggrecan proteoglycans from 1-2 weeks old bovine calves (research 87, Marlborough, MA) and treated these aggrecan molecules with Keratanase with an effort to prepare the materials required for this thought experiments (Appendix A&C). A representative AFM image of the digested calf aggrecan molecules is shown in Figure 4-2. We can see clearly the CS GAG chains have been cleaved, leaving a flexible rod shape protein core, and the KS GAG chains are hard to be identified since they are very short (5-15 kDa). Judged from the contour length of the calf aggrecan molecules, we can tell there has been a low level of proteoglycan proteolytic cleavage as we expected. Further identification and localization of KS GAG chains can be realized by the proposed thought experiments. The anticipated difficulty will be (1) the fine tuning of relative portions of KS antibodies and the aggrecan solutions because the excess KS antibodies can contaminate the image qualities by AFM, (2) the separation between the aggrecan core proteins and Chondroitinase ABC due to their similar molecular weights (the bright dots in Figure 4-2 can be the unpurified Chondroitinase ABC).

Despite my insufficient literature review of the related protocols, I would like to propose here the tentative experimental protocol of this thought experiment, and I am still hoping one day some graduate student (or maybe undergraduate student) would make this interesting project work (This protocol was adapted from Roberts *et. al.* 1994 [63] and Sorrell *et. al.* 1988 [64]):

1. Make the buffer: 0.1 M TRIS-acetate (adjust to PH = 8.0).
2. Dilute the original KS antibody solution with 1: 3000 ratio (suggested).
3. Extract and purify aggrecan molecules. Treat the aggrecan with the Chondroitinase ABC. Adjust the concentration of aggrecan solution to 1 mg/mL (can be quantified by DMMB analysis).
4. Incubate the KS antibody together with aggrecan solution for 30 mins at room temperature.
5. Dilute the final solution to 10 - 100 $\mu\text{g}/\text{ml}$ for AFM imaging.

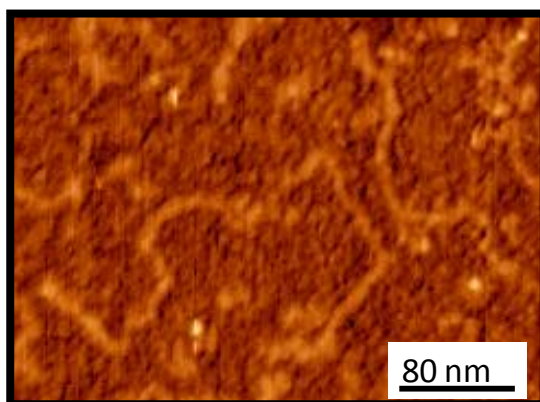


Figure 4-2: The AFM height image of Keratanase treated aggrecan molecules extracted from 1 - 2 weeks old calf.

4.3 Coarse-grained aggrecan model

4.3.1 Coarse-graining scheme for GAG chains

As discussed earlier, previous work has shown that one can further coarse grain [46] the aggrecan side chains based on the calculated persistence length of GAG chains (7~10 nm). Coarse-graining techniques are a well-applied method to replace a few atoms in a macromolecule with a heavy virtual “atom,” and the simulation efficiency is greatly enhanced. Here, I will propose a coarse-graining scheme by which each monosaccharide is replaced by two beads (shown in Figure 4-3): one bead representing the center of mass (C_M), and the other the center of charge (C_Q). Two bonds connecting the center of charge and the center of mass are rigid (this can be achieved by a high bonding coefficient in MD) to prevent high thermal fluctuation of the massless bead C_Q . The interactions coefficients between all the beads will be

calculated by the potential of mean force in the all-atom models.[65] The coefficients of interest include the bonding coefficients, the angle coefficients and the dihedral coefficients.

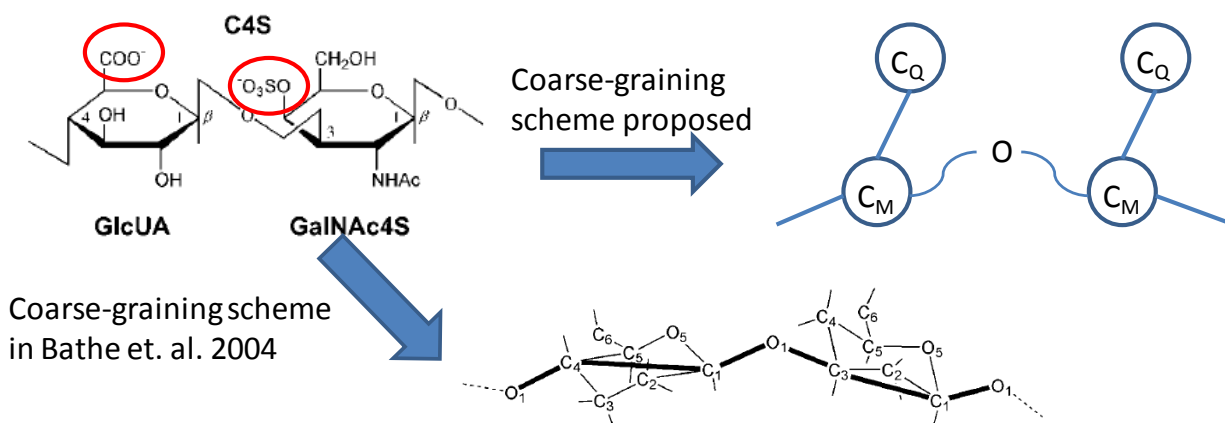


Figure 4-3: Coarse-graining scheme for GAG chains. C_M represents the center of mass of one monosachride, C_Q represents the center of charge of one monosachride and O is the glycosidic oxygen.

4.3.2 Coarse-graining scheme for aggrecan protein core

Regarding simulating the protein core of aggrecan proteoglycans, one simply applies a neutral freely-jointed chain model or worm-like chain model [66]. Otherwise, one can also adopt coarse-graining on aggrecan protein core in MD simulation.

According to Doege and coworkers [67], human aggrecan protein core has conservative structure similar to the repetitive disaccharide units in GAG chains, which may allow us to use similar coarse-graining scheme (to perform the MD simulation of a small repeating unit and then to coarse grain it to simulate an elongated chain). In details, human cartilage aggrecan protein core has 2415 amino acids, and the GAG attachment sites are serine. The dominant regions with most dense GAG attachment is chondroitin sulfate attachment domain 1 (CS-1), and this domain has repeating sequences with conservative structure of 19-20 amino acids.

One repeating unit in aggrecan protein core can be presented by 19 beads in MARTINI force field [68], a common coarse-graining force field for protein. Using the nomenclature of MARTINI force

field, the sequence of 19 beads is Qa-C2-C1-Qa-P1-P1-P4-P1-P5-C2-Qa-Qa-C1-P1-P5-C1-Na-**P1**-P5 (P: Polar, N: Non-polar, C: Apolar, Q: Charged) and the second to the last amino acid is serine, the binding site for GAG chains.

4.4 Inter-chain interactions of neighboring GAGs

4.4.1 Motivation and background

It has been observed in the experiment [33] that certain orientation or alignment of neighboring GAG chains can give rise to either repulsive or adhesive interactions as described in Figure 4-4. The adhesive interaction was hypothesized to be derived from the hydrogen bonds aided by the Ca^{2+} ions. This hypothesis can be further verified by the molecular dynamic simulation by adding specific ions.

There has been no simulation work done to study the inter-chain interactions between GAG chains while the intra-chain interactions have been extensively explored previously as shown in Chapter 3. The model to study the inter-chain interactions can be simply built from the single chain model with Steered Molecular Dynamics (SMD) [69].

The interaction between the two chains is the sum of the pair-wise interactions, and the pair-wise interactions consist of Van der Waals interactions and electrostatic interactions (sometimes a tail correction is added if a cut-off distance is applied to describe Van der Waals forces), i.e. $E_{pair} = E_{vdw} + E_{coul} (+ E_{tail})$.

Steered Molecular Dynamics (SMD) is a simulation method by applying an external force to one or more atoms with constant velocity or constant force. This method was based on the theory Jarzynski's equality [70] which relates free energy differences to the work done through non-equilibrium processes, i.e. $\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}$. SMD is an inherited feature in LAMMPS.

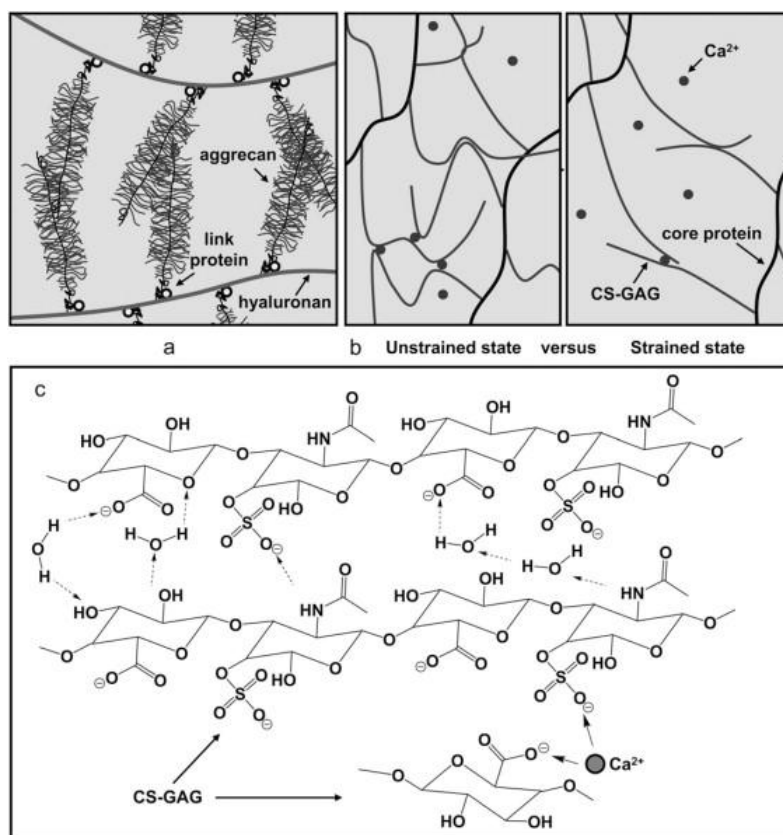


Figure 4-4: Schematics of (a) two opposing aggrecan macromolecules undergoing interpenetration and entanglement in vivo; (b) network structure between aggrecan via self-adhesion due to Ca^{2+} -mediated ion-bridging and molecular entanglements between GAG chains, and energy dissipation of the Ca^{2+} -ion-bridges upon mechanically induced GAG molecular elongation (not drawn on scale); and (c) possible hydrogen bonding (dashed arrows), hydrophobic interaction (could occur between the methyl groups and carbon rings), and Ca^{2+} -mediated ion-bridging between CS-GAG chains in the presence of water molecules. Water bridges could exist between the hydrogen bonding donors and acceptors. (Cited from Han *et. al.* 2008 [33])

4.4.2 Model set up

One can simply adopt the single-chain modeling introduced in Chapter 3 to study the inter-chain interactions. To study whether the sulfation type has any effect, I propose here to align the two sulfated GAG chains (of the same sulfation type or of the different sulfation type) either in the parallel direction or in the anti-parallel direction, as shown in Figure 4-5.

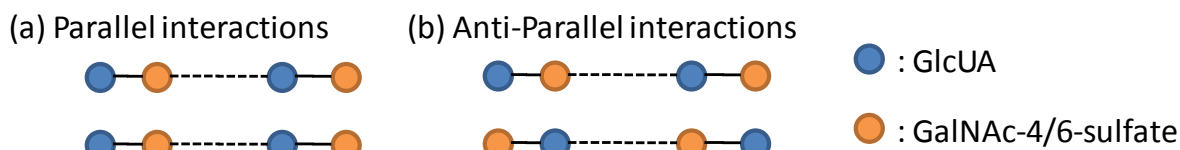


Figure 4-5: Schematics of the inter-chain model to study the interactions between neighboring GAG chains: (a) the two GAG chains are aligned parallel to each other and (b) the two GAG chains are aligned anti-parallel to each other.

The simulation procedures are described as following:

- Place two sulfated GAG chains with 30 Å apart (with either parallel alignment or anti-parallel alignment), and the left ends of both molecules (glycosidic oxygen O13) were fixed at the desired points before applying forces (to prevent the chains from flowing around making different simulation trials incomparable);
- Energy minimization;
- Gradually heat up from 0K to 300K ;
- Apply SMD coupling forces at both ends or at the centers of masses of two chains.
- Repeat the previous step till sufficient trials have been simulated for a meaningful average.[69]

4.4.3 Preliminary result

I did a test run of two tetra two parallel-aligned C4S chains (disaccharide) at $T = 100$ K and then pulled two chains together at centers of mass from 3 nm. As we expected, the Van der Waals interactions dominates at the atomic length scales (10 times higher than the Coulombic interactions), and we observed an attraction well when the chain-to-chain distance is 2.5 Å. When the two chains were far away, the inter-chain interactions are negligible.

A deterministic result can be derived by applying multiple SMD simulations and take average similar to the method described in Buehler 2006 [71].

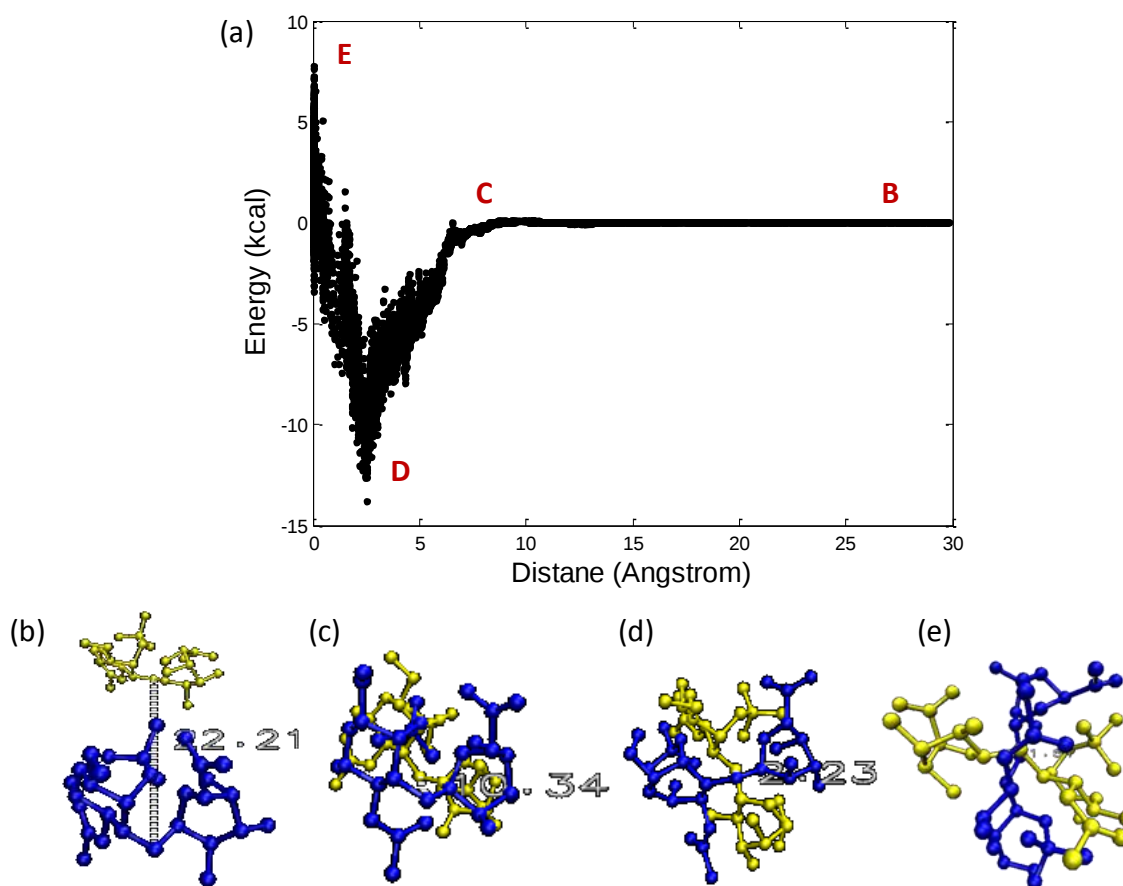


Figure 4-5: A test run of SMD simulation of two parallel-aligned C4S chains (disaccharide) at $T = 100$ K and pull together at centers of mass from $30 \text{ \AA}$. (b), (c), (d) and (e) are the VMD visualization of the simulation chains at points B, C, D and E in plot (a). (b) When the two chains were far apart, they don't feel each other and the alignment is parallel. (c) The starting point that the two chains interact with each other with a tendency of chain twisting. (d) An attractive force was established by the two chains due to Van der Waals interactions, and the two chains align perpendicular to each other. (e) An high repulsion was observed, and the two planes of the chains were perpendicular to each other.

Appendix A

Protocol for Guanidine Extraction and Aggrecan Purification

This protocol was written by Paul Kopesky on February 29, 2008 as an adaption from protocols developed by John Sandy and Anna Plaas. I have used this protocol to extract aggrecan proteoglycans from 1-2 weeks old bovine calves (research 87, Marlborough, MA), which proteoglycans I have used for control experiments to train my experimental skills. I would also highly recommend future students who have an interest in working on human aggrecan molecules to use the well-characterized calf aggrecan proteoglycans to train the experimental techniques in the starting stage. (There are more “full-length” aggrecan molecules in calves.)

Materials:

- Guanidine HCl (4M)
- Sodium Acetate (100mM)
- Protease Complete Tablets (Roche)

Protocols:

1. Make desired volume of 4M Guanidine HCl in 100mM Sodium Acetate buffer and adjust pH to ~7.2. Rough guide: For tissue engineered plugs, make 1mL for every ~40mg of tissue wet weight with sGAG content of ~120 μg (i.e. plug sGAG content is ~3 $\mu\text{g}/\text{mg}$ wet weight). I would double the amount of extraction solution if using native tissue instead of engineered samples.
2. Add appropriate fraction of protease complete tablet per manufacturer's instructions. (Inhibits a broad range of proteases).
3. Mechanically disrupt sample. May use a scalpel to dice (works for tissue engineered plugs). If using native cartilage tissue, may want to pulverize and homogenize.
4. Add appropriate volume of combined solution samples as per step #1 above.

5. Mix on shaker at 4C for 48hrs (probably excessive, but will maximize PG extraction)
6. Spin for 0.5-1hr at ~16,000 x g in microcentrifuge at 4C.
7. Transfer clarified supernatant to new tube and discard pellet.
8. Add CsCl powder until solution density is 1.58g/mL. Check density by weighing a known volume of solution.
9. Transfer sample to an appropriate tube for ultracentrifuge and balance ultracentrifuge rotor exactly!!
10. Spin as long and as hard as you can. 72hrs at ~500,000 x g should produce a good density gradient.
11. Immediately at the end of the spin, carefully pipette fractions of the sample. Splitting 1mL into ~10 x 100 μ L fractions works well. Start by pipetteing off the very top and working down the tube. Be aware that a hard pellet of CsCl may form at the bottom; the density gradient will likely still be sufficient for PG purification.
12. Measure the density of the fractions as in step #8.
13. Combine fractions based on density. $D1 > 1.54\text{g/mL}$; $1.46\text{g/mL} < D2 < 1.54\text{g/mL}$ (Roughly & White. JBC 255:217-224. 1980)
14. Dialyze with 500 volumes of 1M NaCl for 4 hours (change the solution in the middle).
15. Dialyze with 500 volumes of MilliQ filtered distilled H₂O overnight.

Important: Guanidine is a chaotropic agent and should be handled carefully.

Quantification:

For the final step of purification, half of the sample was purified by dialysis, and the other half was purified by the spinning (I was just curious which purification procedure will give me the final gain of purified products). Interestingly, even though the dialyzed sample appears much more viscous than the

spinned sample, they showed the similar concentration as shown in the Figure A-1 analyzed by colorimetric method with the substrate DMMB analysis (Appendix B).

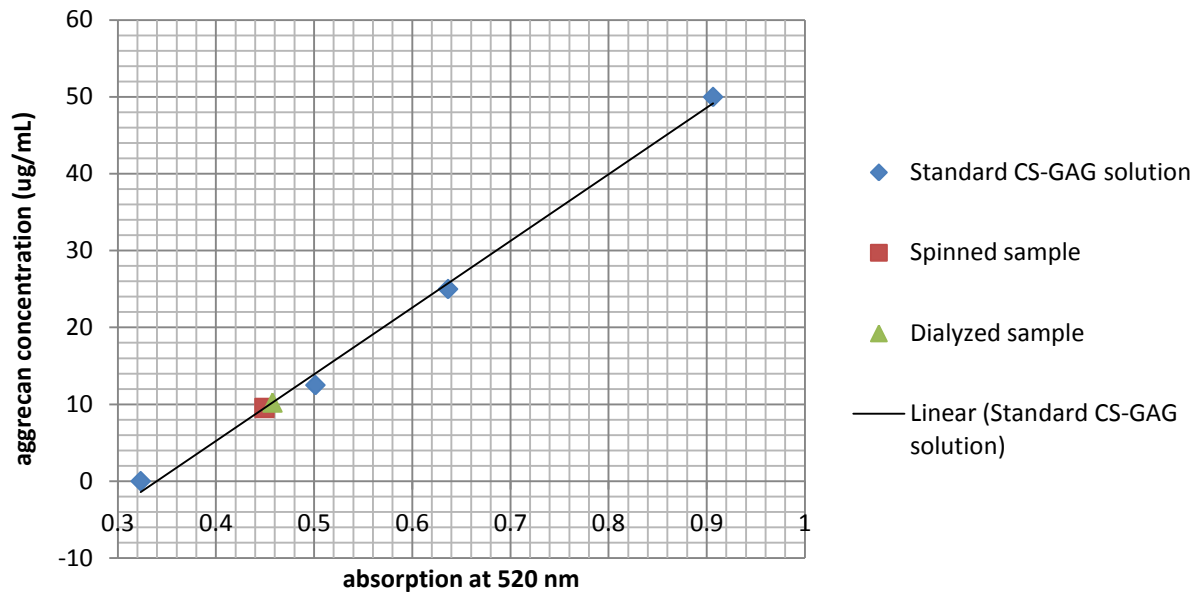


Figure A-1: DMMB analysis of the prepared standard solution and diluted extracted aggrecan solution.

The GAG concentration in the spinned sample was found to be 1.19 mg/mL , and the GAG concentration in the dialyzed sample was found to be 1.28 mg/mL .

Appendix B

Protocol for dimethylmethylene blue (DMMB) analysis

Dimethylmethylene is blue-colored dye that can react with GAG chains, the change of dye's color can be detected by the light absorbance. This absorbance is proportional to the substrate concentration (GAG concentration in this case). This protocol was orally taught by Paul Kopesky and recorded by me.

Materials:

- Comercial CS GAG powder
- Dimethylmethylene blue (DMMB) solution

Protocols of standard curve:

1. Prepare the CS GAG solutions. 10 mg/mL (~1mL, calculated by weight dissolved in water).
2. Two serial 1:10 dilution and make 100 ug/mL **stock** GAG solution (100 uL standard + 900 uL water).
3. Prepare the standard sample (200 uL in total for all standards)

Concentration ($\mu\text{g/mL}$)	0	12.5	25	50	75	100
Volume of stock (μL)	0	25	50	100	150	200
Volume of water (μL)	200	175	150	100	50	0

4. Dilution the aggrecan sample accordingly. Let the concentration of the diluted solution to be roughly in the middle of the standards (~50 $\mu\text{g/mL}$).
5. Add 10 μL of the standard solutions and 200 μL DMMB into the wells on the plate. (Repeat for several rows).
6. Read off 570 nm of absorbance of prepared solution.

Appendix C

Protocol for Chondroitinase ABC Digestion of Aggrecan

This protocol was provided by Rachel Miller. With her help, I prepared some Chondroitinase ABC treated aggrecan molecules extracted from 1 - 2 weeks calf (Figure 4-2).

Materials:

Chase Buffer (pH 7.6):

- 0.05 M Tris HCl (or Tris Base, corrected to pH 7.6)
- 0.05 M Na Acetate
- 0.01 M EDTA

[Note: To prepared 10 ml solution, use 60 mg Tris Base (MW 121.1), 41 mg Na acetate(MW 82.03), 37.2 mg EDTA (MW 372.2)]

Protocols:

1. Suspend 100 µg GAG lyophilized sample in 100 µl Chase Buffer:
2. Digest with protease-free Chondroitinase ABC (30 mU/ 100 µg GAG) (Warner) for 3 hours in a water bath of 37 °C.

Appendix D

Nanomechanics of human intervertebral disc

I have utilized the well-established protocols for micro-contact printing and high resolution force spectroscopy [32] to quantify the nanoscale deformability and compressibility of functionalized aggrecan. Aggrecan molecules are chemically end-grafted onto gold substrate and densely packed in a hexagon regions. A micro-sized gold coated tip is used to compress the aggrecan pattern and data analysis enables the calculation of an effective stress-strain curves. The measurements are obtained in a fully hydrated fluid medium (NaCl solution), such that the ionic strength can be varied to mimic the physiological environment. These experiments serve as testing the mechanical properties of the intervertebral disc aggrecan in the physiological medium.

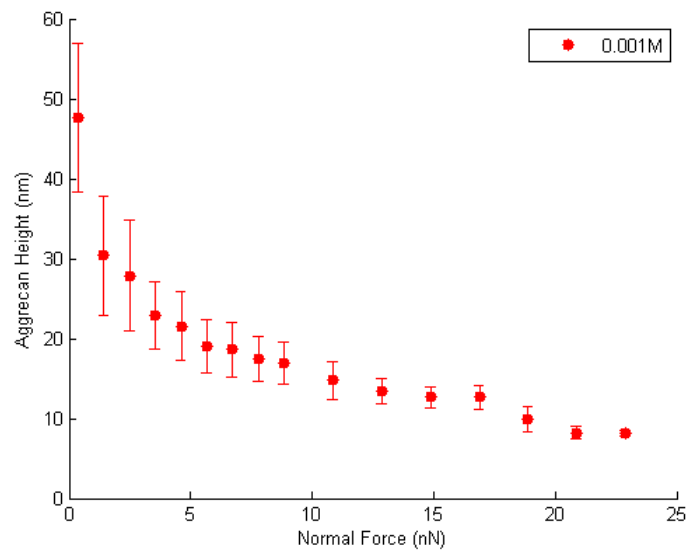


Figure D-1: Nanomechanics results of disc aggrecan (24 year old human being, HA-associated, and untreated aggrecan) under the ionic strength (controlled by NaCl concentration) of 0.001 mol/L.

The plan was to apply the above methodology to the untreated aggrecan in both the HA- and the non-HA-associated group (Chapter 2). I did a few trials of nanomechanics experiments on untreated

and HA-associated sample. Unfortunately, I haven't got any satisfactory results. Previous nanomechanics experiments [72] showed a height of aggrecan brushes by Pico-force is decreased to 200 – 300 nm (possibility due to a low friction force applied) even though a full-length aggrecan measured is around 400 nm (from AFM imaging). Taking into account that the contour length measured by AFM imaging of the human disc sample is less than 200 nm in average, measurement by AFM Pico-force may be applicable for disc aggrecan.

Appendix E

Explicit solvation model

As I mentioned in Section 3.3.4 Computational clusters, I have done some explicit solvation model on cluster HPC. Below I will just briefly explain what I have done, and the input files were generated by customized Python scripts.

The simulation box is 7 nm X 7 nm X 7 nm, and was filled in the beginning with TIP3 water molecules only ($\sim 10^5$ molecules). The system first went through energy minimization, then gradually warmed up to 300 K in NVT ensemble, and finally equilibrated at 300 K till the water density stabilized around 1 g/cm³ (1.021 ± 0.002 g/cm³ in my simulation to be exact).

Solvated ions and the chondroitin chain (in energy minimized conformation) were added to the equilibrated waterbox. To account for the physiological ionic strength (~ 0.15 mol/L) and the charge neutrality, forty sodium ions (Na⁺) and thirty chloride ions (Cl⁻) were added. The simulations steps that followed are the same as those described in Chapter 3 for implicit solvation model. I only simulated 500 ps for each system at that time to test the robustness of my simulation systems, and the results are consistent with the results as those in Chapter 3 in terms of Cremer-Pople parameters, glycosidic torsions and hydrogen bonds. To save the space, I won't list them here again redundantly.

Appendix F

LAMMPS input data files

The input data files for MD simulations in LAMMPS require the following information of the simulated molecules:

1. The initial coordinates and velocities of atoms;
2. The bonds and their coefficients;
3. The angles and their coefficients;
4. The dihedrals and their coefficients;
5. The impropers and their coefficients.

The initial coordinates of atoms, the bonds, the angles, the dihedrals and the impropers are solely depend on the conformation of the molecule one wants to start with, the velocities can be scaled by the temperature and all coefficients are determined by the force field applied. Writing the input files manually is a tedious job and error-prone for large molecules, and there are tools developed for LAMMPS users to generate the input files (http://lammps.sandia.gov/doc/Section_tools.html) from commercialized software AMBER, CHARMM, Materials Studio and etc (One should also check Protein Data Bank to see whether there is experimental data available to be used as the initial conformation). The basic idea is to use user-friendly software to create the conformation of simulated molecules and then convert to the LAMMPS input files with appropriate syntax. This way one can have the ease of creating the input files with a large number of atoms while enjoy the vast simulation capacity and features provided by LAMMPS.

Chemical Engineering Department at MIT has licensed Accelrys Materials Studio software installed in its Athena Cluster (66-080). I have tried the “msi2lmp” tool to convert msi files of a C4S chain created by Materials Studio to the LAMMPS input file, but the following challenges were encountered:

1. msi2lmp does not always output the cross-terms correctly

2. It does not handle the failsafe parameters (with 'auto' suffix) in Force Fields (e.g.cff91,cvff ...)

3. It doesn't support all force fields as LAMMPS does, the conversion is possible only if one chooses to use the force fields of CFF91, CVFF and PCFF.

Therefore, in order to get exactly the same single point energy as that obtained from MS, in LAMMPS we have to do two things:

1. To visually check the parameters already generated by msi2Imp.

2. To add additional terms to some pair interactions using '***_style hybrid' to include the fail safe terms.

Bibliography

- [1] J.F. Kennedy, *Proteoglycans: biological and chemical aspects in human life*, Elsevier Scientific Pub. Co., 1979.
- [2] P.J. Basser, R. Schneiderman, R.A. Bank, and E. and M. Wachtel A., "Mechanical properties of the collagen network in human articular cartilage as measured by osmotic stress technique," *Archives of Biochemistry and Biophysics*, vol. 351, 1998, pp. 207-219.
- [3] H. Watanabe, Y. Yamada, and K. Kimata, "Roles of aggrecan, a large chondroitin sulfate proteoglycan, in cartilage structure and function," *Journal of Biochemistry*, vol. 124, 1998, pp. 687-693.
- [4] R.V. Iozzo, *Proteoglycans: structure, biology, and molecular interactions*, CRC, 2000.
- [5] B.M. Vertel, "The ins and outs of aggrecan," *Trends in Cell Biology*, vol. 5, 1995, pp. 458-464.
- [6] C. Kiani, C. Liwen, W.U. Yao Jiong, A.J. YEE, and B.Y. Burton, "Structure and function of aggrecan," *Cell Research*, 2002.
- [7] R.V. Iozzo, "Matrix proteoglycans: From molecular design to cellular function," *Annu. Rev. Biochem.*, vol. 67, 1998, p. 609.
- [8] B. Chakrabarti and J.W. Park, "Glycosaminoglycans: structure and interaction.,", *CRC critical reviews in biochemistry*, vol. 8, 1980, p. 225.
- [9] M.P. Brown, L. West, K. Merritt, and A.H.K. Plass, "Changes in sulfation patterns of chondroitin sulfate in equine articular cartilage and synovial fluid in response to aging and osteoarthritis," *American Journal of Veterinary Research*, vol. 59, 1998, pp. 786-791.
- [10] D. Heinegård and T. Saxne, "The role of the cartilage matrix in osteoarthritis.,", *Nature reviews. Rheumatology*, vol. 7, Nov. 2010, pp. 50-56.
- [11] H. Nagase and M. Kashiwagi, "Aggrecanases and cartilage matrix degradation," *Arthritis Research & Therapy*, vol. 5, 2003, p. 94.
- [12] R. Sztrolovics, M. Alini, P.J. Roughley, and J.S. Mort, "Aggrecan degradation in human intervertebral disc and articular cartilage.,", *The Biochemical journal*, vol. 326 (Pt 1, Aug. 1997, pp. 235-41.

- [13] Jacobs, *The Burden of Musculoskeletal Diseases in the United State*, Rosemont, IL: 2008.
- [14] P.J. Roughley, "Biology of intervertebral disc aging and degeneration," *Spine*, vol. 29, 2004, pp. 2691-2699.
- [15] J.P. Urban and S. Roberts, "Degeneration of the intervertebral disc," *Arthritis Research and Therapy*, vol. 5, 2003, pp. 120-138.
- [16] M.A. Adam and P.J. Roughley, "What is intervertebral disc degeneration, and what causes it?," *Spine*, vol. 31, 2006, pp. 2151-2161.
- [17] others, "Magnetic resonance imaging of the cervical spine: frequency of degenerative changes in the intervertebral disc with relation to age," *Clinical radiology*, vol. 58, 2003, pp. 980-984.
- [18] J. Melrose, P. Ghosh, and T.K.F. Taylor, "A comparative analysis of the differential spatial and temporal distributions of the large (aggrecan, versican) and small (decorin, biglycan, fibromodulin) proteoglycans of the intervertebral disc," *Journal of Anatomy*, vol. 198, 2001, pp. 3-15.
- [19] J.A. Buckwalter, K.C. Smith, L.E. Kazarien, L.C. Rosenberg, and R. Ungar, "Articular cartilage and intervertebral disc proteoglycans differ in structure: an electron microscopic study," vol. 7, 1989, pp. 146-151.
- [20] J.L. DiFabio, R.H. Pearce, B. Caterson, and H. Hughes, "The heterogeneity of the non-aggregating proteoglycans of the human intervertebral disc," *Biochemical Journal*, vol. 244, 1987, pp. 27-33.
- [21] L.S. Lohmander and J.H. Kimura, "Biosynthesis of cartilage proteoglycan," *Articular Cartilage Biochemistry*, 1986, pp. 93-110.
- [22] P.J. Roughley, L.I. Melching, T.F. Heathfield, R.H. Pearce, and J.S. Mort, "The structure and degradation of aggrecan in human intervertebral disc.," *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society*, vol. 15 Suppl 3, Aug. 2006, pp. S326-32.
- [23] L. Ng, A.J. Grodzinsky, P. Patwari, J. Sandy, A. Plaas, and C. Ortiz, "Individual cartilage aggrecan macromolecules and their constituent glycosaminoglycans visualized via atomic force microscopy," *Journal of Structural Biology*, 2003.
- [24] L.J. Ng, "Probing nanomechanics of aggrecan and the aggrecan-rich pericellular matrix of chondrocytes in cartilage," Massachusetts Institute of Technology, 2005.

- [25] H.Y. Lee, "Ultrastructure and nanomechanical properties of aggrecan from native cartilage and engineered tissue," Massachusetts Institute of Technology, 2010.
- [26] J.A. Buckwalter, A. Pedrini-Mille, V. Pedrini, and C. Tudisco, "Proteoglycans of human infant intervertebral disc. Electron microscopic and biochemical studies," *The Journal of Bone and Joint Surgery*, vol. 67, 1985, pp. 284-294.
- [27] C.Q. Zhao, L.M. Wang, L.S. Jiang, and L.Y. Dai, "The cell biology of intervertebral disc aging and degeneration," *Ageing research reviews*, vol. 6, 2007, pp. 247-261.
- [28] S.A. Fellini, J.H. Kimura, and V.C. Hascall, "Polydispersity of proteoglycans synthesized by chondrocytes from the swarm rat chondrosarcoma," 1981, pp. 7883-7889.
- [29] P.A.S. Mourão, "Distribution of chondroitin 4-sulfate and chondroitin 6-sulfate in human articular and growth cartilage," *Arthritis & Rheumatism*, vol. 31, 1988, pp. 1028-1033.
- [30] R.M. Lauder, T.N. Huckerby, G.M. Brown, M.T. Bayliss, and I.A. Nieduszynski, "Age-related changes in the sulphation of the chondroitin sulphate linkage region from human articular cartilage aggrecan.," *Biochemical Journal*, vol. 358, 2001, p. 523.
- [31] D. Dean, L. Han, A.J. Grodzinsky, and C. Ortiz, "Compressive nanomechanics of opposing aggrecan macromolecules," *Journal of Biomechanics*, 2006.
- [32] D. Dean, L. Han, C. Ortiz, and A.J. Grodzinsky, "Nanoscale conformation and compressibility of cartilage aggrecan using microcontact printing and atomic force microscopy," *Macromolecules*, vol. 38, 2005, pp. 4047-4049.
- [33] L. Han, D. Dean, L.A. Daher, A.J. Grodzinsky, and C. Ortiz, "Cartilage Aggrecan Can Undergo Self-Adhesion," *Biophysical Journal*, vol. 95, 2008, pp. 4862-4870.
- [34] A. Papagiannopoulos, T.A. Waigh, T. Hardingham, and M. Heinrichs, "Solution structure and dynamics of cartilage aggrecan," *Biomacromolecules*, vol. 7, 2006, pp. 2162-2172.
- [35] M. Bathe, A.J. Grodzinsky, B. Tidor, and G.C. Rutledge, "Optimal linearized Poisson--Boltzmann theory applied to the simulation of flexible polyelectrolytes in solution," *The Journal of chemical physics*, vol. 121, 2004.
- [36] M. Bathe, G.C. Rutledge, A.J. Grodzinsky, and B. Tidor, "Osmotic pressure of aqueous chondroitin sulfate solution: a molecular modeling investigation," *Biophysical Journal*, vol. 89, 2005, pp. 2357-2371.
- [37] M. Bathe, G.C. Rutledge, A.J. Grodzinsky, and B. Tidor, "A coarse-grained molecular model for glycosaminoglycans: application to chondroitin, chondroitin sulfate, and hyaluronic acid," *Biophysical Journal*, vol. 88, 2005, pp. 3870-3887.

- [38] M.D. Buschmann and A.J. Grodzinsky, "A molecular model of proteoglycan-associated electrostatic forces in cartilage mechanics," *J. Biomech. Eng.*, vol. 117, 1995, pp. 179-192.
- [39] D. Dean, J. Seog, C. Ortiz, and A.J. Grodzinsky, "Molecular-level theoretical model for electrostatic interactions within polyelectrolyte brushes: applications to charged glycosaminoglycans," *Langmuir*, vol. 19, 2003, pp. 5526-5539.
- [40] J. Seog, D. Dean, A.H.K. Plaas, S. Wong-Palms, A.J. Grodzinsky, and C. Ortiz, "Direct measurement of glycosaminoglycan intermolecular interactions via high-resolution force spectroscopy," *Macromolecules*, vol. 35, 2002, pp. 5601-5615.
- [41] A. Subbotin, M. Saariaho, R. Stepanyan, O. Ikkala, and G. ten Brinke, "Cylindrical Brushes of Comb Copolymer Molecules Containing Rigid Side Chains," *Macromolecules*, vol. 33, Aug. 2000, pp. 6168-6173.
- [42] A. Subbotin, M. Saariaho, O. Ikkala, and G. ten Brinke, "Elasticity of Comb Copolymer Cylindrical Brushes," *Macromolecules*, vol. 33, May. 2000, pp. 3447-3452.
- [43] L. Feuz, F.A.M. Leermakers, M. Textor, and O. Borisov, "Bending rigidity and induced persistence length of molecular bottle brushes: A self-consistent-field theory," *Macromolecules*, vol. 38, 2005, pp. 8891-8901.
- [44] R.J. Nap and I. Szleifer, "Structure and Interactions of Aggrecans: Statistical Thermodynamic Approach," *Biophysical Journal*, vol. 95, 2008, pp. 4570-4583.
- [45] J.A. Buckwalter, P.J. Roughley, and L.C. Rosenberg, "Age-Related changes in cartilage proteoglycans: Quantitative electron microscopic studies," *Microscopy research and technique*, vol. 28, 1994, pp. 398-408.
- [46] S.O. Nielsen, C.F. Lopez, G. Srinivas, and M.L. Klein, "Coarse grain models and the computer simulation of soft materials," *Journal of Physics: Condensed Matter*, vol. 16, 2004, p. R481-R512.
- [47] K.N. Kirschner, A.B. Yongye, S.M. Tschampel, J. Gonza, C.R. Daniels, B.L. Foley, and R.J. Woods, "GLYCAM06 : A Generalizable Biomolecular Force Field .," *Journal of Computational Chemistry*, 2007.
- [48] others, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *The Journal of Physical Chemistry B*, vol. 102, 1998, pp. 3586-3616.
- [49] J.R. Fried, "Computational Parameters," *Macromolecules*.
- [50] A.D. Mackerell, "Empirical force fields for biological macromolecules: overview and issues.," *Journal of computational chemistry*, vol. 25, 2004, pp. 1584-604.

- [51] others, "LAMMPS-large-scale atomic/molecular massively parallel simulator," *Sandia National Laboratories*, 2007.
- [52] W. Humphrey, A. Dalke, and K. Schulten, "{VMD} -- {V}isual {M}olecular {D}ynamics," *Journal of Molecular Graphics*, vol. 14, 1996, pp. 33-38.
- [53] M.-yi Shen My and K.F. Freed, "Long time dynamics of Met-enkephalin: comparison of explicit and implicit solvent models.," *Biophysical journal*, vol. 82, Apr. 2002, pp. 1791-808.
- [54] S. Plimpton, "Fast Parallel Algorithms for Short-Range Molecular Dynamics," *Sciences-New York*, vol. 117, 1995, pp. 1-19.
- [55] D. Cremer and J.A. Pople, "General definition of ring puckering coordinates," *Journal of the American Chemical Society*, vol. 97, 1975, pp. 1354-1358.
- [56] A. Almond, A. Brass, and J.K. Sheehan, "Deducing polymeric structure from aqueous molecular dynamics simulations of oligosaccharides: predictions from simulations of hyaluronan tetrasaccharides compared with hydrodynamic and X-ray fibre diffraction data.," *Journal of molecular biology*, vol. 284, Dec. 1998, pp. 1425-37.
- [57] B.M. Sattelle, J. Shakeri, I.S. Roberts, and A. Almond, "A 3D-structural model of unsulfated chondroitin from high-field NMR: 4-sulfation has little effect on backbone conformation.," *Carbohydrate research*, vol. 345, 2010, pp. 291-302.
- [58] B.M. Sattelle and A. Almond, "Less is More When Simulating Unsulfated Glycosaminoglycan 3D-Structure : Comparison SCC-DFTB-D / TIP3P Predictions With Experiment," *Journal of Computational Chemistry*, 2010.
- [59] G. Cilpa, M.T. Hyvönen, A. Koivuniemi, and M. Riekkola, "Atomistic Insight into Chondroitin-6-Sulfate Glycosaminoglycan Chain Through Quantum Mechanics," *Journal of Computational Chemistry*, 2010.
- [60] H. Sugeta and T. Miyazawa, "General method for calculating helical parameters of polymer chains from bond lengths, bond angles, and internal-rotation angles," *Biopolymers*, vol. 5, 1967, pp. 673-679.
- [61] D. Heinegard and I. Axelsson, "Distribution of keratan sulfate in cartilage proteoglycans," *The Journal of Biological Chemistry*, vol. 252, 1977, pp. 1971-1979.
- [62] B. Caterson, J.E. Christner, and J.R. Baker, "Identification of a monoclonal antibody that specifically recognizes corneal and skeletal keratan sulfate. Monoclonal antibodies to cartilage proteoglycan.," *The Journal of biological chemistry*, vol. 258, Jul. 1983, pp. 8848-54.

- [63] S. Roberts, B. Caterson, H. Evans, and S.M. Eisenstein, "Proteoglycan components of the intervertebral disc and cartilage endplate: an immunolocalization study of animal and human tissues.," *The Histochemical journal*, vol. 26, May. 1994, pp. 402-11.
- [64] J.M. Sorrell, F. Mahmoodian, and B. Caterson, "of chondroitin sulfates and keratan sulfate in embryonic chick bone marrow," *Cell and Tissue Research*, vol. 252, 1988, pp. 523-531.
- [65] D. Reith, H. Meyer, and F. Mu, "Mapping Atomistic to Coarse-Grained Polymer Models Using Automatic Simplex Optimization To Fit Structural Properties," *Society*, 2001, pp. 2335-2345.
- [66] P.G. de Gennes, *Scaling concepts in polymer physics*, Cornell Univ Pr, 1979.
- [67] K.J. Doege, M. Sasaki, T. Kimura, and Y. Yamada, "Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. Human-specific repeats, and additional alternatively spliced forms.," *The Journal of biological chemistry*, vol. 266, Jan. 1991, pp. 894-902.
- [68] L. Monticelli, S.K. Kandasamy, X. Periole, R.G. Larson, D.P. Tieleman, and S.-J. Marrink, "The MARTINI Coarse-Grained Force Field: Extension to Proteins," *Journal of Chemical Theory and Computation*, vol. 4, May. 2008, pp. 819-834.
- [69] S. Park and K. Schulten, "Calculating potentials of mean force from steered molecular dynamics simulations.," *The Journal of chemical physics*, vol. 120, Apr. 2004, pp. 5946-61.
- [70] C. Jarzynski, "Nonequilibrium Equality for Free Energy Differences," *Physical Review Letters*, vol. 78, Apr. 1997, pp. 2690-2693.
- [71] M.J. Buehler, "Atomistic and continuum modeling of mechanical properties of collagen: Elasticity, fracture, and self-assembly," 2006, pp. 1947-1961.
- [72] H.Y. Lee, P. Kopesky, L. Daher, A. Mosquera, D. Frisbie, J. Kisiday, A. Grodzinsky, and C. Ortiz, "Morphology of Aggrecan Produced By Adult Equine Mesenchymal Stem Cells and Chondrocytes in Self-Assembling Peptide Hydrogels," *submitted to the Orthopaedic Research Society Meeting, San Francisco, CA*, 2008.