

Molecular Dynamics and Mechanical Behavior of Collagen Type I
and its Lysine/Hydroxylysine-derived Crosslinks

Albert Lawrence Kwansa

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in
partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Biomedical Engineering

Joseph W. Freeman, Co-Chair
Raffaella De Vita, Co-Chair
David R. Bevan
Felicia A. Etzkorn
H. Clay Gabler

May 3, 2013
Blacksburg, Virginia

Keywords: Type I Collagen, Crosslink, Molecular Model, Fibril, Microfibril

Copyright © 2013 by Albert L. Kwansa

Molecular Dynamics and Mechanical Behavior of Collagen Type I and its Lysine/Hydroxylysine-derived Crosslinks

Albert L. Kwansa

ABSTRACT

Collagen type I is an extracellular matrix (ECM) protein that affords tensile strength and biological scaffolding to numerous vertebrate and invertebrate tissues. This strength has been attributed to the triple-helical structure of the collagen type I molecules, their organization into fibrils, and the presence of inter-molecular, covalent, enzymatic crosslinks. There are several different types of these crosslinks; their composition is tissue-specific and dependent upon factors such as age and health. Furthermore, these enzymatic crosslinks tend to form specifically at amino/N- and carboxy/C-terminal crosslinking sites. The mechanical behavior of collagen type I has been investigated, via experiment and theory, at the level of the molecule, microfibril, fibril, and fiber. However, the influence of different enzymatic crosslinks and their location (e.g., N- vs. C-site) on the mechanics of collagen type I has not been investigated in the literature.

We employed molecular dynamics to model the mechanical behavior of uncrosslinked and crosslinked ~23-nm-long molecular segments and ~65-nm-long microfibril units of collagen type I. We then used these molecular simulations to construct a model of a single collagen type I fibril by considering the ~65-nm-long microfibril units arranged in series and then in parallel.

When a uniaxial deformation was applied along the long axis of the molecular models, N-crosslinks aligned rapidly at lower strains followed by C-crosslinks more gradually at higher strains, leading to a two-stage crosslink recruitment. Then when comparing the influence of different enzymatic crosslinks, significant differences were observed for the high-strain elastic moduli of our microfibril unit models, namely and in increasing order, uncrosslinked, immature crosslinked (HLKLN and deH-HLNL), mature HHL-crosslinked, and mature PYD-crosslinked. At the fibril level, our low- and high-strain elastic moduli were in good agreement with some literature data, but in over-estimation of several other literature reports. Future work will seek to address simplifications and limitations in our modeling approach. A model such as this, accounting for different enzymatic crosslink types, may allow for the prediction of the mechanics of collagen fibrils and collagenous tissues, in representation of healthy and diseased states.

Dedication

I would like to dedicate this dissertation to God and to my family, in particular, my parents, Dr. Herman Kwansa and Mrs. Vida Kwansa; my brother, Victor Kwansa; our grandfather, Rev. Dr. Albert L. Kwansa; and our family located here in the United States, in Ghana, West Africa, and elsewhere.

Acknowledgements

I would like to thank my committee co-chairs, Dr. Freeman and Dr. De Vita, for their dedication, guidance, encouragement, and patience over the years; and my committee members, Dr. Bevan, Dr. Etzkorn, and Dr. Gabler for all of their time, expertise, and support during my preliminary exam and during meetings that were had. I would like to thank Dr. Freeman for his enthusiasm for research and mentoring, his passion for exposing elementary through high school students to biomedical research, the independence and freedom that he affords his graduate students in the direction of their research, and for the respect, care, and patience that he shows for his students. I would like to thank Dr. De Vita for her dedication to research, teaching, and student outreach, for her attentiveness, her professionalism, and for the excellent example that she sets for her students and for her lab. I am thankful to have had the unique opportunity to work under Dr. Freeman in his Musculoskeletal Tissue Regeneration Lab and then to be co-advised by Dr. De Vita in her Mechanics of Soft Biological Systems Lab. I would also like to thank my fellow graduate students for all of their support as well as the undergraduate and high school students that I have had the opportunity to work with in these labs.

I would like to express gratitude to the Virginia Tech-Wake Forest School of Biomedical Engineering and Sciences (SBES), especially to the department's faculty leadership and its staff leadership (in particular, Amanda Covey, Kathy Cregar, Tess Sentelle, and Pam Stiff) for this opportunity and for their continuous support and kindness. I would also like to thank Dr. Jessica Sparks and SBES for the opportunity to serve as a teaching assistant in the Fall of 2009 for Mammalian Physiology.

Front Matter

I would like to acknowledge the Virginia Tech-Initiative for Maximizing Student Development (VT-IMSD) fellowship program and the Postbaccalaureate Research and Education Program (VT-PREP), their faculty leadership (Dr. David Bevan, Dr. Anne McNabb, Dr. Jill Sible, Dr. Ed Smith, and Dr. Eric Wong), their staff leadership (Mr. Leemar Thorpe, Dr. Anjolie Diaz, and formerly, Dr. Xiaojing Guan), my fellow IMSD scholars, and the PREP scholars. I would like to acknowledge the National Aeronautics and Space Administration (NASA) and the UNCF Special Programs Corporation for their Harriett G. Jenkins Predoctoral Fellowship Project. I am also thankful to have been involved with projects that were funded by Integra LifeSciences and the National Science Foundation.

Finally, I would like to thank friends that I have made at Virginia Tech and in Blacksburg for providing support and encouragement, in particular, Yared Kidane, Jung Ki Hong, the Empson family (Digby, Yvonne, Anthea, and Amelia), and The River Church.

Table of Contents

Dedication..... iii

Acknowledgements..... iii

Table of Contents..... v

List of Equations..... vii

List of Figures..... vii

List of Tables..... viii

List of Abbreviations and Acronyms..... ix

Chapter 1: Introduction..... 1

 1.1. Overview and motivation..... 1

 1.2. Research goal and specific aims..... 3

Chapter 2: Background..... 4

 2.1. Collagen Types..... 4

 2.2. Collagen Type I..... 5

 2.3. Synthesis, Intracellular Post-translational Modifications, and Triple-helical Assembly..... 6

 2.4. Extracellular Post-translational Modifications and Fibril Formation..... 7

 2.4.1. Propeptide removal, fibril nucleation, and early fibril formation..... 7

 2.4.2. Molecular arrangement within microfibrils and fibrils..... 9

 2.4.3. Fibril growth..... 10

 2.5. Enzymatic Collagen Crosslinks..... 10

 2.5.1. Extracellular modification by lysyl oxidases..... 10

 2.5.2. Immature crosslinks: Aldimines and ketoimines..... 11

 2.5.3. Mature crosslinks: Pyridinolines, pyrroles, and histidine-derived crosslinks..... 12

 2.5.4. Crosslinking disorders..... 15

 2.6. Structural Hierarchy and Mechanics of Collagen Type I..... 17

 2.6.1. Structural hierarchy..... 17

 2.6.2. Mechanics: Elastic moduli..... 19

 2.6.3. Mechanics: Deformation mechanisms..... 20

 2.7. Molecular Modeling..... 21

 2.7.1. Molecular mechanics..... 21

 2.7.2. Molecular dynamics (MD)..... 23

 2.7.3. Constant velocity MD pulling techniques..... 24

 2.7.4. Explicit and implicit solvation..... 24

Chapter 3: Molecular Modeling of 23-nm-long Overlap Region..... 27

 3.1. Introduction..... 27

 3.2. Methods..... 28

 3.2.1. Molecular modeling programs and parameters..... 28

 3.2.2. Building of the molecular models..... 29

 3.2.3. Energy minimization, heating, and equilibration..... 30

 3.2.4. Constant velocity pulling..... 31

 3.2.5. Crosslink alignment, crosslink engineering strain, and crosslink strain energy..... 32

 3.2.6. Initial lengths, extensions, and engineering strains..... 33

 3.2.7. Statistical analyses..... 34

 3.2.8. Computing resources..... 35

 3.3. Results and Discussion..... 35

 3.3.1. Higher spring constants improve accuracy in pulling velocity..... 35

 3.3.2. Pulling direction and pulling velocity introduce bias..... 36

 3.3.3. N- and C-crosslinks display two-stage recruitment..... 37

 3.3.4. Extension of N- and C-crosslink recruitment to larger size scales..... 39

Front Matter

Chapter 4: Molecular Modeling of 65-nm-long Microfibril Units.....	41
4.1. Introduction.....	41
4.2. Methods.....	42
4.2.1. Molecular modeling programs and parameters.....	42
4.2.2. Building of the microfibril molecular models.....	43
4.2.3. Energy minimization, heating, and equilibration.....	44
4.2.4. Constant velocity pulling.....	46
4.2.5. Overall strain energy, crosslink strain energy, and crosslink alignment.....	48
4.2.6. Cross-sectional areas.....	49
4.2.7. Initial lengths, extensions, and engineering strains.....	50
4.2.8. Forces, nominal stresses, and elastic moduli.....	51
4.2.9. Statistical analyses.....	53
4.2.10. Computing resources.....	53
4.3. Results and Discussion.....	54
4.3.1. Cross-sectional areas and initial lengths.....	54
4.3.2. Strain energy and stresses: Influence of crosslink state and fixed atoms.....	55
4.3.3. Elastic moduli: Influence of crosslink state and fixed atoms.....	57
4.3.4. Crosslink strain energy and alignment: Influence of crosslink state.....	60
Chapter 5: Multiple-Scale Fibril Model.....	62
5.1. Introduction.....	62
5.1.1. Crosslink compositions of healthy collagenous tissues.....	62
5.1.2. Mechanical properties of single collagen fibrils.....	63
5.2. Methods.....	64
5.2.1. Assumptions and simplifications.....	65
5.2.2. Model formulation.....	66
5.2.3. Qualitative comparison with data from the literature.....	68
5.2.4. Quantitative comparison with data from the literature.....	68
5.3. Results and Discussion.....	69
5.3.1. Qualitative comparison with data from the literature.....	69
5.3.2. Quantitative comparison with data from the literature.....	71
5.4. Future work: Revisiting modeling assumptions and simplifications.....	72
Chapter 6: Summary.....	76
References.....	77
Appendix A: CHARMM19 parameters added for deH-HLNL.....	91
Appendix B: TCL scripts and NAMD configuration file.....	93
B.1. Model building steps for Aims 1 and 2.....	93
B.2. TCL script used to apply molecular kinks within the gap region.....	95
B.3. TCL script used to replace template Pro-Pro-Gly amino acid sequences.....	97
B.4. TCL script used to select fixed atoms and assign pulled atom spring constants.....	98
B.5. NAMD configuration file.....	99
B.6. TCL script used to measure initial lengths.....	101
B.7. TCL script used to obtain energy values for the entire model from the log files.....	102
B.8. TCL script used to calculate energy values for the crosslinks.....	103
B.9. TCL script used to obtain Cartesian coordinates for the crosslink C α atoms.....	104
Appendix C: Mathematica notebook files.....	105
C.1. Mathematica notebook used for the crosslink C α atom Cartesian coordinate data.....	105
C.2. Mathematica notebook used for the crosslink energy data.....	106
C.3. Mathematica notebook used for the energy data for the entire model.....	107
Appendix D: Influence of the pulling velocity on the ~23-nm-long molecular model.....	108
Appendix E: Trajectory snapshots of the ~65-nm-long molecular models.....	109

Appendix F: Influence of the pulling velocity on the ~65-nm-long molecular models..... 117
Appendix G: Influence of the pulling velocity on the moduli of the ~65-nm-long models..... 123
Appendix H: Listing of additional work..... 125
 H.1. Manuscripts in preparation..... 125
 H.2. First-authored publications..... 127
 H.3. Co-authored publications..... 129

List of Equations

Equation 3.1: Total potential energy in the CHARMM19 united-atom force field..... 29
Equation 3.2: RMS difference equation for the calculation of initial lengths..... 34
Equation 3.3: RMS displacement equation for the calculation of changes in length..... 34
Equation 4.1: Total potential energy in the CHARMM22 all-atom force field..... 42
Equation 4.2: RMS difference equation for the calculation of initial lengths..... 50
Equation 4.3: RMS displacement equation for the calculation of changes in length..... 50
Equation 4.4: Strain energy expressed as a 6th order polynomial function..... 51
Equation 4.5: Nominal stress expressed as a 5th order polynomial function..... 52
Equation 4.6: Low-to-high-strain modulus transition point..... 52
Equation 5.1: Fibril force expressed as a summation of identical sub-fibrils in parallel..... 67
Equation 5.2: Fibril force expressed as a summation of different sub-fibrils in parallel..... 68
Equation 5.3: Fibril stress as a function of fibril strain and initial fibril length..... 68

List of Figures

Figure 2.1: Amino acid lengths and sequence similarities for collagen type I α -chains..... 6
Figure 2.2: Chemical structures of amino acids that serve as precursors to enzymatic crosslinks 14
Figure 2.3: Chemical structures of divalent, immature collagen crosslinks..... 14
Figure 2.4: Chemical structures of trivalent, mature collagen crosslinks..... 15
Figure 2.5: A schematic of the hierarchy of collagen type I from the fiber to the molecule..... 18
Figure 3.1: Percent error in the pulling velocity at 100 m/s..... 36
Figure 3.2: Snapshots of the model being pulled in the (a) C- or (b) N-direction..... 36
Figure 3.3: N- and C-crosslink angles (a), engineering strains (b), and strain energy (c)..... 39
Figure 3.4: A schematic of unipolar and bipolar fibrils..... 40
Figure 4.1: Snapshots of our five microfibril unit molecular models after equilibration..... 46
Figure 4.2: A representative schematic of the constant velocity pulling simulation..... 47
Figure 4.3: Cross-sectional areas for the microfibril unit models..... 54
Figure 4.4: Initial cross-sectional areas and initial lengths for the microfibril unit models..... 55
Figure 4.5: Strain energy-extension data for the microfibril unit models at 6.25 m/s..... 56
Figure 4.6: Stress-strain data for the microfibril unit models at 6.25 m/s..... 56
Figure 4.7: Transition point strain values used for determining low- and high-strain moduli..... 57
Figure 4.8: Low- and high-strain elastic moduli for the microfibril unit models at 6.25 m/s..... 59
Figure 4.9: Strain energy in the immature and mature crosslinks at 6.25 m/s..... 61
Figure 5.2: Stress-strain data for our fibril model plotted with data from the literature..... 69
Figure 5.3: Elastic moduli for our fibril model plotted with literature data..... 72
Figure B.1: A schematic of model building steps..... 93
Figure B.2: A schematic of model building steps for Aims 1 and 2..... 94

Figure B.3: A schematic of model building steps for Aim 2.....	94
Figure D.1: Influence of the pulling velocity on the ~23-nm-long model.....	108
Figure E.1: HHL model snapshots (12 fixed C-term C α ; 12 pulled N-term C α).....	109
Figure E.2: HLKLN model snapshots (12 fixed C-term C α ; 12 pulled N-term C α).....	110
Figure E.3: deH-HLNL model snapshots (12 fixed C-term C α ; 12 pulled N-term C α).....	111
Figure E.4: PYD model snapshots (12 fixed C-term C α ; 12 pulled N-term C α).....	112
Figure E.5: Uncrosslinked model snapshots (12 fixed C-term C α ; 12 pulled N-term C α).....	113
Figure E.6: Uncrosslinked model snapshots (3 fixed C-term C α ; 12 pulled N-term C α).....	114
Figure E.7: deH-HLNL model snapshots (3 fixed C-term C α ; 12 pulled N-term C α).....	115
Figure E.8: deH-HLNL model snapshots (12 fixed N-term C α ; 12 pulled C-term C α).....	116
Figure F.1: Influence of the pulling velocity on the HHL-crosslinked model.....	117
Figure F.2: Influence of the pulling velocity on the HLKLN-crosslinked model.....	118
Figure F.3: Influence of the pulling velocity on the deH-HLNL-crosslinked model.....	119
Figure F.4: Influence of the pulling velocity on the PYD-crosslinked model.....	120
Figure F.5: Influence of the pulling velocity on the uncrosslinked model.....	121
Figure F.6: Influence of the pulling velocity on the uncrosslinked model.....	122
Figure G.1: Influence of the pulling velocity on the ~65-nm-long model.....	124

List of Tables

Table 2.1: Examples of vertebrate collagen types and tissues in which they are found.....	4
Table 2.2: Enzymatic collagen crosslinks and vertebrate tissues in which they are found.....	13
Table 2.3: Disorders and conditions affecting collagen crosslinking.....	17
Table 2.4: Hierarchy and mechanics of collagen type I from the molecule to the fiber.....	20
Table 2.5: Verlet integration method employed for MD calculations in NAMD.....	23
Table 3.1: Preparation steps used for our ~23-nm-long molecular models.....	31
Table 4.1: Positions of crosslink precursor amino acids for each crosslink type.....	44
Table 4.2: Preparation steps used for our ~65-nm-long molecular models.....	45
Table 4.3: Pulling velocities employed for the microfibril unit molecular models.....	47
Table 5.1: Experimentally determined elastic moduli for single native collagen type I fibrils....	64
Table 5.2: Theoretically determined elastic moduli for collagen fibrils.....	64
Table A.1: Three new sets of bonds length parameters for the deH-HLNL crosslink.....	91
Table A.2: Four new sets of bond angle parameters for the deH-HLNL crosslink.....	91
Table A.3: Three new sets of bond torsion parameters for the deH-HLNL crosslink.....	92
Table A.4: Two new sets of Lennard-Jones parameters for the deH-HLNL crosslink.....	92
Table B.1: Positions for five molecular segments within a collagen type I unit cell.....	93

List of Abbreviations and Acronyms

Amino acids

Gly	glycine
His	histidine
Hyl	hydroxylysine
Hyp	hydroxyproline
Lys	lysine
Pro	proline

Protein nomenclature

α -chains	alpha chains
C α	alpha carbon atom

Collagen crosslinks

deH-HLNL	dehydro-hydroxylysino-norleucine
deH-LNL	dehydro-lysino-norleucine
HLKNL	hydroxylysino-keto-norleucine
LKNL	lysino-keto-norleucine
PYD	pyridinoline (hydroxylysyl pyridinoline)
DPD	deoxy-pyridinoline (lysyl pyridinoline)
PYL	pyrrole (hydroxylysyl pyrrole)
DPL	deoxy-pyrrole (lysyl pyrrole)
HHL	histidinyl hydroxylysino-norleucine
deH-HHMD	dehydro-histidino-hydroxymerodesmosine

Molecular modeling

CHARMM	Chemistry at HARvard Molecular Mechanics
cv-SMD	constant velocity steered molecular dynamics
GBIS	generalized Born implicit solvent
MD	molecular dynamics
NAMD	NANoscale Molecular Dynamics
RMS	root-mean-square
TCL	tool command language
VMD	Visual Molecular Dynamics

Experimental techniques

AFM	atomic force microscopy
MEMS	microelectromechanical systems
SAXS	small-angle x-ray scattering

1.1. Overview and motivation

Collagens are structural extracellular matrix (ECM) proteins that are found in various vertebrate and invertebrate tissues (Urich, 1994). Collagen proteins are by far the most abundant proteins in mammals, often comprising up to one-third of total body protein (Williams, 1978). Approximately 27 types of collagens have been identified with type I being the most prevalent; type I collagen is found in connective tissues such as tendon, ligament, bone, skin, and the cornea of the eyes (Brinckmann, 2005; K. von der Mark, 2006), where it functions to provide tensile strength and to serve as a biological scaffold for cells and other ECM components (Sweeney et al., 2008).

Collagen type I is a rod-shaped molecule that is approximately 300 nm in length and 1.5 nm in diameter (Birk & Bruckner, 2005). A single collagen type I molecule is comprised of three polypeptide chains (α -chains), each containing just over 1000 amino acids; these three chains are wound together into a triple helix like threads twisted into a rope (Brinckmann, 2005; Orgel et al., 2001). Outside of the cell, these molecules assemble into quarter-staggered assemblies called microfibrils (Orgel, Irving, Miller, & Wess, 2006), which form sub-fibrils and in turn larger structures called fibrils (Cisneros, Hung, Franz, & Muller, 2006; Ezura, Chakravarti, Oldberg, Chervoneva, & Birk, 2000). Thus, collagen type I and other collagens that assemble into fibrils are often termed fibril-forming or fibrillar collagens. In native tissues, these fibrils are bundled together into collagen fibers, with other types of collagens, non-collagenous macromolecules, water, and cells (Weiss & Jayson, 1982).

These fibrillar structures are stabilized by the presence of covalent inter-molecular bonds called crosslinks. There are enzymatic crosslinks that form in early development and throughout maturation, and there are non-enzymatic crosslinks that are more prevalent in old age or in certain disease states (e.g., diabetes) (Avery & Bailey, 2008). The enzymatic crosslinks form initially as immature, divalent crosslinks that connect only two amino acids. In time, these immature crosslinks can react further to form mature, trivalent crosslinks, which can connect three amino acids (Eyre & Wu, 2005). The immature, divalent collagen crosslinks include aldimines and ketoimines, while the mature, trivalent crosslinks include pyridinolines, pyrroles, and a histidine-derived crosslink (Eyre, Koob, & Van Ness, 1984; Eyre & Wu, 2005; Saito &

Chapter 1: Introduction

Marumo, 2010). The prevalence of these different crosslink types has been reported to be tissue-specific and dependent upon factors such as age and health (Eyre, Paz, & Gallop, 1984; Saito, Marumo, Fujii, & Ishioka, 1997).

Regarding health, certain connective tissue disorders have been shown to disrupt the proper formation of these enzymatic crosslinks (e.g., Bruck syndrome type 2, Ehlers-Danlos syndrome (EDS) type 6A, Cutis laxa type 4, and scleroderma). Most of these disorders have been associated with genetic mutations that directly or indirectly affect post-translational modifying enzymes, such as lysyl hydroxylases and lysyl oxidases, that are involved in crosslink formation (Eyre, Paz, & Gallop, 1984; Ishikawa, Kondo, & Miyachi, 1998; van der Slot et al., 2003). Lysyl hydroxylases are enzymes that convert the amino acid lysine to hydroxylysine, which is required for most of the immature and mature collagen crosslinks. Lysyl oxidases are copper-dependent enzymes that convert the side chain amino group of lysines and hydroxylysines into a reactive aldehyde, which is essential for crosslink formation (Eyre & Wu, 2005). Some of the symptoms associated with the aforementioned connective tissue disorders include fragile skin and eyes (EDS type 6A), brittle bones and stiff joints (Bruck syndrome type 2), loose and inelastic skin (Cutis laxa type 4), and overly stiff skin (scleroderma) (Eyre, Paz, & Gallop, 1984; Ishikawa et al., 1998; van der Slot et al., 2003). These symptoms describe alterations to the mechanical behaviors of the affected connective tissues, in which fibril-forming collagens are major load bearing units. Thus, investigations of the mechanical properties of fibril-forming collagens at their different levels of hierarchy (e.g., molecules, microfibrils, sub-fibrils, and fibrils) can contribute to an improved understanding of healthy collagenous tissues and pathologies that result in mechanical dysfunctions of these tissues.

Within the last ~10 years, techniques and technologies have been developed to mechanically characterize and model fibril-forming collagens (e.g., type I collagen) at smaller scales. Such collagens have been characterized and modeled at the molecular level (e.g., optical tweezers (Sun, Luo, Fertala, & An, 2002, 2004), atomic force microscopy (AFM) (Bozec & Horton, 2005), and molecular dynamics (MD) (Lorenzo & Caffarena, 2005)); modeled at the microfibril level (e.g., MD (Gautieri, Vesentini, Redaelli, & Buehler, 2011) and finite-element models (Hambli & Barkaoui, 2012)); and characterized and modeled at the fibril level (e.g., AFM (Hang & Barber, 2011; van der Rijt, 2004; Yang, van der Werf, Dijkstra, Feijen, &

Chapter 1: Introduction

Bennink, 2012), microelectromechanical systems (MEMS) (Eppell, Smith, Kahn, & Ballarini, 2006; Shen, Dodge, Kahn, Ballarini, & Eppell, 2008, 2010), Kelvin-Voigt-based models (Gautieri, Vesentini, Redaelli, & Buehler, 2012), and mesoscopic models (Buehler, 2008)). However, models of crosslinked collagens have, thus far, treated the enzymatic crosslinks generally without distinguishing between different enzymatic crosslink types (e.g., immature and mature crosslinks).

1.2. Research goal and specific aims

The overall goals of this work were to investigate the roles that different enzymatic collagen crosslinks play in the mechanical response of collagen type I at the molecular and microfibril scales, and to model the mechanical behavior of a single collagen fibril while accounting for different enzymatic collagen crosslinks. This work involved a combination of molecular modeling and multiple-scale modeling of a single fibril. In order to accomplish these goals, we had the following specific aims:

- **Specific Aim 1:** To construct a molecular model of the ~23-nm-long overlap region between the ends of two crosslinked collagen type I molecules, and to investigate the molecular dynamics and mechanics of crosslinks located at N- and C-terminal crosslinking sites.
- **Specific Aim 2:** To construct molecular models of ~65-nm-long collagen type I microfibril units, and to investigate the molecular dynamics and mechanics of five different crosslink states (uncrosslinked, immature deH-HLNL and HLKNL crosslinks, and mature HHL and PYD crosslinks).
- **Specific Aim 3:** To develop a multiple-scale model of a single collagen fibril based on the mechanical coupling of the microfibril unit molecular models.

Chapter 2: Background

2.1. Collagen Types

Collagens are a family of structural proteins that are found in both vertebrates and invertebrates; in fact, collagens appear to be present in nearly all eukaryotes except for plants and protozoa (Urich, 1994). Invertebrate collagens are present, for example, in the body walls of hydra and worms, the foot and adductor muscles of sea snails, the spine ligaments of sea urchins, and within the skeletal tissues of sponges (Boot-Handford & Tuckwell, 2003). Vertebrate collagens comprise tissues such as bones, tendons, ligaments, skin, cartilage, and blood vessels, as well as internal organs (Kielty & Grant, 2002; K. von der Mark, 2006). In mammals, collagens are especially abundant, comprising up to one-third of total body protein by weight (Anasuya & Rao, 1970; Harkness, Harkness, & James, 1958; Williams, 1978). Collagen proteins are characterized by their triple-helical domains and by their ability to self-assemble into supra-molecular structures (Exposito, 2004). There are approximately 27 different types of collagen proteins that have been identified (K. von der Mark, 2006); these types have been classified into several sub-families such as fibril-forming collagens (e.g., types I, II, and III), regulatory fibril-forming collagens (e.g., types V and XI), network-forming collagens (e.g., type IV), and fibril-associated collagens with interrupted triple helices (FACITs) (e.g., type XIV) (Exposito, 2004; K. von der Mark, 2006) (see Table 2.1).

Table 2.1: Examples of vertebrate collagen types and tissues in which they are found (Kielty & Grant, 2002; K. von der Mark, 2006).

Type	Sub-family	Found in tissues, such as:	Functions
I	Fibril-forming	Tendons, ligaments, bones, skin, and the cornea of the eyes	Major fibril-forming collagen of non-cartilaginous tissues; associated with tensile strength of connective tissues
II		Cartilage and the vitreous fluid of the eyes	Major fibril-forming collagen of cartilage and related tissues
III		Blood vessels, skin, lung, and liver	Major fibril-forming collagen of non-cartilaginous tissues; often forms heterogeneous fibrils with type I collagen; associated with tissue elasticity
IV	Network-forming	Epithelial and endothelial tissues, such as the basement membranes of the glomeruli & lens capsule of the eyes	Forms thin, mesh-like sheets that serve as interfaces between tissues (e.g., endothelial and epithelial linings of organs and tissues)

Table 2.1 (continued)

Type	Sub-family	Found in tissues, such as:	Functions
V	Regulatory fibril-forming	Bone, tendon, cornea of the eyes, skin, & blood vessels	Minor fibril-forming collagen of non-cartilaginous tissues; associated with fibril-nucleation and fibril formation
VI	Beaded-filament forming	Cartilage, intervertebral discs, skin, placenta, lung, & blood vessels	Forms beaded filaments and branched network structures; associated with stabilizing and reinforcing the ECM
XIV	FACIT	Cartilage, skin, tendon, blood vessels, placenta, lung, & liver	Associated with the control of lateral fibril growth and fibril packing
XI	Regulatory fibril-forming	Cartilage, vitreous fluid of eyes	Minor fibril-forming collagen of cartilaginous tissues; associated with fibril-nucleation and fibril formation

2.2. Collagen Type I

Collagen proteins are composed of three polypeptide chains (also called alpha chains or α -chains) (Brinckmann, 2005). These three α -chains are each folded into left-handed helices, and together they form a right-handed triple helix (Engel & Bächinger, 2005). In the major fibril-forming, vertebrate collagens (i.e., types I, II, and III) as well as the minor fibril-forming collagens (i.e., types V and XI), the α -chains each contain just over 1,000 amino acids; over 95% of these amino acids are a part of an uninterrupted triple-helical domain with the remaining amino acids comprising non-triple-helical extensions called N- and C-terminal telopeptides (The UniProt Consortium, 2012). Fibril-forming, vertebrate collagens are approximately 300 nm in length and 1.5 nm in diameter (Birk & Bruckner, 2005). Interestingly, there is a large variation in the lengths of fibril-forming, invertebrate collagens, which have been reported to range from as short as 14 nm, in the nematocysts (stinging tube-like structures) of hydra (Kurz, 1991), to as long as 2,400 nm, in the cuticle layer of the body walls of worms (Gaill et al., 1991).

Within the triple-helical domain(s) of collagens, the amino acids exhibit a Gly-X-Y repeating sequence, where Gly is an abbreviation for the amino acid glycine. The X and Y positions are commonly occupied by the amino acids proline and hydroxyproline, respectively, but can be occupied by a number of other amino acids. For instance, human Achilles tendon collagens – primarily type I collagen – have been reported to contain 18 different amino acids including 25.4% glycine, 15.2% proline, 12.6% hydroxyproline, 3.29% lysine, 1.50% hydroxylysine, and 0.87% histidine (Eastoe, 1955). The amino acid lengths and sequences of the $\alpha 1$ and $\alpha 2$ chains of collagen type I are fairly well conserved across different mammalian species

Chapter 2: Background

(Figure 2.1). Hydroxyproline has been shown to confer increased thermal stability to the triple-helical domains via hydrogen bonding and inductive effects (Myllyharju, 2005), while lysine, hydroxylysine, and histidine are important for the formation of enzymatic covalent crosslinks (Eyre & Wu, 2005).

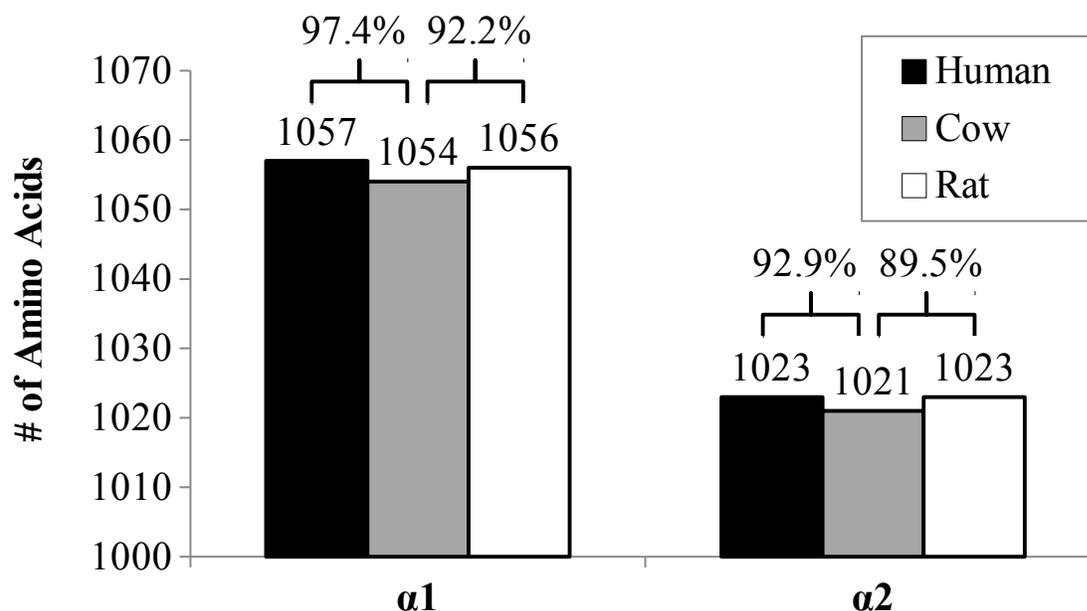


Figure 2.1: Amino acid lengths and sequence similarities for collagen type I α -chains of humans (*Homo sapiens*), cows (*Bos taurus*), and rats (*Rattus norvegicus*). This data is based on the Universal Protein Resource (UniProt) and the following UniProt accession numbers: P02452 and P08123 (humans), P02453 and P02465 (cows), and P02454 and P02466 (rats) for the $\alpha 1$ and $\alpha 2$ chains, respectively (The UniProt Consortium, 2012). The amino acid lengths are expressed as whole numbers and positioned directly above the bars, while the sequence similarities are expressed as percentages and positioned further above with brackets.

2.3. Synthesis, Intracellular Post-translational Modifications, and Triple-helical Assembly

Collagen type I α -chains are encoded by the COL1A1 and COL1A2 genes; in humans, these genes are located on chromosomes 17 and 7, respectively (K. von der Mark, 2006). The α -chains are translated on the rough endoplasmic reticulum, after which they are post-translationally modified in the endoplasmic reticulum (ER) and at the Golgi apparatus (Bateman, Lamande, & Ramshaw, 1996).

Modifications in the ER include the following: Lysines (Lys) and prolines (Pro) are hydroxylated to hydroxylysine (Hyl) and hydroxyproline (Hyp), respectively, by lysyl and prolyl hydroxylases. As mentioned, Lys and Hyl have significant roles in collagen crosslinking, while

Chapter 2: Background

Hyp is important for the thermal stability of the collagen triple helix. Lysyl hydroxylases are intracellular enzymes that act in the post-translational modification of lysines into hydroxylysines. Lysyl hydroxylases are homodimeric proteins with a total molecular weight of approximately 170 kDa (85 kDa per subunit); these enzymes are localized to the luminal (inner) surface of the endoplasmic reticulum (ER) where they act as peripheral membrane proteins (Myllyharju, 2005). Another type of ER post-translational modification is the addition of carbohydrate (sugar) groups onto Hyl residues. Specifically, the intracellular enzyme hydroxylysyl galactosyl-transferase adds a galactose molecule onto specific Hyl residues, and another intracellular enzyme, hydroxylysyl glucosyl-transferase, adds a glucose onto those modified galactosyl Hyl residues (Bateman et al., 1996). The functions of these carbohydrate groups are not yet well understood. Some researchers have proposed that these carbohydrate extensions help to regulate molecular packing and fibril diameter; other researchers have suggested that these carbohydrates may regulate the specificity of crosslink formation and maturation (Yamauchi, 1996).

Within the cell, the alpha chains (two $\alpha 1$ chains and one $\alpha 2$ chain) associate via recognition domains (Khoshnoodi, Cartailier, Alvares, Veis, & Hudson, 2006). These three α -chains fold into a triple-helical molecule called procollagen, under the guidance of associated enzymes and molecular chaperones (Bateman et al., 1996). The resulting procollagen molecule contains a non-triple-helical extended N-terminal propeptide domain (35 kDa), an uninterrupted triple-helical domain (300 kDa), and a globular C-terminal propeptide domain (100 kDa) (Risteli & Risteli, 2006). The C-propeptide domain contains the chain recognition sites responsible for the association of the three chains. Furthermore, these propeptide domains (in particular, the C-propeptide domain) are thought to inhibit the procollagen molecules from assembling into fibrils within cells (Risteli & Risteli, 2006). Once the procollagen molecules are formed, they are packaged into secretory vesicles for release out of the cell via exocytosis (Bateman et al., 1996).

2.4. Extracellular Post-translational Modifications and Fibril Formation

2.4.1. Propeptide removal, fibril nucleation, and early fibril formation

Once released into the extracellular space, enzymes called N- and C-propeptidases (or proteinases) act to cleave the N- and C- terminal propeptides of procollagen, allowing for fibril

Chapter 2: Background

formation to begin. The nucleation or initiation of fibril formation has been attributed to regulatory fibril-forming collagens (e.g., types V and XI) and their association with the main fibril-forming collagens (e.g., types I and III, in the case of ligaments and tendons) (Kadler, Hill, & Canty-Laird, 2008). The initial stages of fibril formation are proposed to occur through the accretion or assembly of collagen molecules, through hydrophobic, electrostatic, and entropic driving forces, into units called microfibrils that are approximately 3.5 to 4 nm in diameter (Christiansen, Huang, & Silver, 2000). It is suggested that these microfibrils serve as building blocks for the formation of early fibrils (Cisneros et al., 2006). *In vivo*, the formation of such early fibrils (also called sub-fibrils, fibril intermediates, or fibril segments) is believed to occur within narrow extracellular fibril-forming compartments that are established by fibroblast cells, after which, the fusion of these compartments is thought to allow for the formation of bundles of fibril segments (Birk, Zycband, Winkelmann, & Trelstad, 1990). Thereafter, it has been proposed that the sub-fibrils can undergo controlled longitudinal and lateral association and fusion, leading to the formation of longer and larger diameter fibrils (Ezura et al., 2000). Collagen sub-fibrils have been reported to be fairly homogenous in diameter (e.g., ~25 nm for rat tail tendons; ~64 nm for mouse digital tendons (Ezura et al., 2000; Zhao, Weinhold, Lee, & Dahners, 2011), while growing/maturing collagen fibrils tend to display a wide distribution of diameters (e.g., tens to hundreds of nm) (Ezura et al., 2000). In addition to the extracellular, fibril-forming compartmentalization provided by cells, extracellular macromolecules such as fibril-associated collagens with interrupted triple helices (e.g., type XIV) and proteoglycans (e.g., decorin, fibromodulin, and lumican) have been shown to influence and guide the growth of collagen fibrils (Ezura et al., 2000).

The kinetics of collagen fibril formation have been studied *in vitro* by monitoring the turbidity of collagen solutions. These studies have revealed a sigmoidal growth curve with an initial lag phase followed by rapid growth, which eventually slows. The lag phase has been attributed to the pre-nucleation of collagen molecules into microfibrils; this lag phase has been shown to be both temperature-dependent and concentration-dependent. Conversely, the rapid growth phase has been attributed to the nucleation and growth of the microfibrils (Farber, Garg, Birk, & Silver, 1986).

2.4.2. Molecular arrangement within microfibrils and fibrils

The molecular packing arrangement of type I collagen molecules into fibrils has been studied extensively using x-ray diffraction techniques. These studies have led to the development of several packing models, most notably, the quasi-hexagonal packing model originally proposed by Hulmes and Miller and the “Smith” 5-bundle model (Hulmes & Miller, 1979; Smith, 1968). The quasi-hexagonal packing model has been further subdivided into sheet-like structures and microfibrillar structures (Wess, Hammersley, Wess, & Miller, 1998).

Fraser *et al.* obtained x-ray diffraction data for rat tail tendon and reported information about a proposed unit cell and lateral positions of the molecules within this unit cell. Here, a unit cell is essentially a very small building block within a structure that tends to be repeated throughout a material, which is highly organized or ordered (Fraser, MacRae, Miller, & Suzuki, 1983). Although the coordinate positions for the five molecular segments were reported in work by Fraser *et al.* (Fraser, MacRae, & Miller, 1987; Fraser *et al.*, 1983), the specific order of the five segments was not yet proposed. However, Wess *et al.* reviewed different collagen molecular packing arrangements and proposed a specific ordering of the five segments for the compressed microfibril based on fitting against experimental x-ray diffraction data (Wess *et al.*, 1998).

More recently, Orgel *et al.* employed x-ray fiber diffraction to experimentally characterize the molecular and microfibrillar structure of collagen type I within rat tail tendons. They reported on the presence of a 1D-staggered microfibril composed of five collagen molecules (Orgel *et al.*, 2006). This microfibril exhibited the following structural characteristics: (1) in the axial/longitudinal (length-wise) direction, the molecules were staggered by a distance 1D (i.e., $1 \cdot D$), where D is approximately 67 nm for rat tail tendon, (2) molecules within each microfibril were inter-digitated with each other and with those of neighboring microfibrils producing a “super twist”, (3) while moving across D-periods, the molecules displayed a wave-like undulation where the molecules were parallel within the overlap region and underwent molecular rearrangement within the gap region to produce the “super twist”, and (4) in the lateral (width-wise) direction, the molecules were arranged in a “quasi-hexagonal” packing array (Orgel *et al.*, 2006).

Chapter 2: Background

2.4.3. Fibril growth

Longitudinal and lateral growth of collagen fibrils have been attributed to the aggregation and fusion of multiple fibril segments (Ezura et al., 2000). However, *in vivo* control mechanisms are in place to prevent certain types of fusion (e.g., to regulate fibril length and diameter) during different stages of tissue development. Collagen-associated biomolecules such as the small leucine-rich proteoglycans (SLRP) decorin and lumican have been shown to inhibit inter-fibrillar interactions and fibrillar fusion. The surface of young and developing collagen fibrils normally has a higher density of these SLRPs concentrated on their sides, but not at their tips, which favors end-to-end fusion in developing fibrils. In contrast, mature fibrils have a reduced surface density of these SLRPs, which allows for additional fusion mechanisms (e.g., side-by-side and side-to-end), which can lead to thicker fibrils as well as branched fibril networks, respectively (Birk, Nurminkaya, & Zycband, 1995).

2.5. Enzymatic Collagen Crosslinks

Enzymatic collagen crosslinks refers to covalent bonds that can form between specific amino acids through the action of enzymes. These crosslinking amino acids include specific lysines (Lys), hydroxylysines (Hyl), and histidines (His). There are several different types of enzymatic collagen crosslinks; some of these crosslinks are immature crosslinks (often called divalent crosslinks since they tend to link only two amino acids), while others are mature crosslinks (often called trivalent crosslinks since most tend to link three amino acids) (Eyre & Wu, 2005). The type of crosslink and the prevalence of each crosslink are dependent upon several factors including the tissue type (e.g., tendon, bone, or skin) (Eyre, Koob, & Van Ness, 1984), the specific tissue and its function (e.g., different tendons) (Fujii, Yamagishi, Nagafuchi, Tsuji, & Kuboki, 1994), the developmental stage or age of the person or animal (e.g., fetal stage, adolescence, or maturity) (Sims & Bailey, 1992), and the health or condition of the tissue in the body (e.g., healthy or compromised by a connective tissue disorder) (Eyre, Paz, & Gallop, 1984).

2.5.1. Extracellular modification by lysyl oxidases

Lysyl oxidases are extracellular copper-dependent enzymes that act selectively on Lys and Hyl residues in the non-helical telopeptide domains, to convert their ϵ -amino groups into aldehyde groups via oxidative deamination (Avery & Bailey, 2008; Kagan & Ryvkin, 2011).

Chapter 2: Background

Lysyl oxidases have high activity on Lys and Hyl residues when their parent collagen molecules have assembled into quarter-staggered molecular aggregates (i.e., microfibrils), in contrast to unassembled collagen molecules. One reason for this preference for assembled collagen molecules is that while acting on a telopeptide Lys or Hyl, a lysyl oxidase enzyme may have a binding site on the helical domain of an adjacent molecule (Eyre, Paz, & Gallop, 1984; Siegel, 1974). After modification, Lys and Hyl aldehydes can react with ϵ -amino groups of other unmodified Lys and Hyl residues, via a condensation reaction, to form crosslinks (Knott & Bailey, 1998). The following sub-sections describe different enzymatic crosslinks that can form; chemical structures for these crosslinks are shown in Figure 2.3 and Figure 2.4.

2.5.2. Immature crosslinks: Aldimines and ketoimines

The immature crosslinks are generally classified as aldimines and ketoimines. Immature crosslinks are usually more abundant in younger tissues, tissues with a higher turn-over of collagens, and in healing tissues (Avery & Bailey, 2008; Yamauchi, 1996).

Aldimines contain a Schiff base functional group ($R-CH=N-R'$, where R and R' denote carbon-based groups) linking the two crosslinking amino acids. Aldimine crosslinks occur between a telopeptide Lys aldehyde (a Lys residue modified by a lysyl oxidase) and either an unmodified helical Hyl or Lys. Thus, there are two types of aldimines, namely, dehydrohydroxylysino-norleucine (abbreviated as deH-HLNL; chemical structure shown in Figure 2.3) and dehydro-lysino-norleucine (abbreviated as deH-LNL; chemical structure shown in Figure 2.3), depending on whether a helical Hyl or Lys is present to react with the telopeptide Lys aldehyde (see Figure 2.3 for the chemical structures of deH-HLNL and deH-NBL) (Eyre & Wu, 2005). Of the enzymatic collagen crosslinks, the aldimines are the most susceptible to being broken or cleaved in the presence of dilute acid and/or with mild heating (Avery & Bailey, 2008). Aldimine crosslinks are found, for example, in skin, tendon, ligament, bone, and cartilage (Saito et al., 1997).

Ketoimines begin as a Schiff base crosslink, but they undergo a chemical rearrangement (Amadori rearrangement) leading to a carbon-nitrogen single bond, instead of a carbon-nitrogen double bond, and a keto group ($R-C(=O)-R'$, where R and R' denote carbon-based groups on either side of a carbonyl group ($C=O$)). This Amadori rearrangement affords greater chemical

Chapter 2: Background

stability since the carbon-nitrogen bond of the crosslink is no longer susceptible to dilute acid and/or mild heating (Simon P. Robins & Brady, 2002). Ketoimine crosslinks occur between a telopeptide Hyl aldehyde (an Hyl modified by a lysyl oxidase) and either an unmodified helical Hyl or Lys. As with aldimines, there are two types of ketoimines, namely, hydroxylysino-keto-norleucine (HLKLN) and lysino-keto-norleucine (LKLN), depending on the presence of either a helical Hyl or Lys (Eyre & Wu, 2005). Ketoimines are found primarily in bones, cartilage, ligaments, tendons, and dentin (Saito et al., 1997; Yamauchi, 1996).

2.5.3. Mature crosslinks: Pyridinolines, pyrroles, and histidine-derived crosslinks

As a tissue matures, the collagens within the tissue tend to experience a reduced turnover. This allows for additional crosslinking reactions to take place resulting in mature crosslinks. The mature crosslinks include the pyridinolines (PYD and DPD), pyrroles (PYL and DPL), and histidine-derived crosslinks (HHL and deH-HHMD). In contrast to immature crosslinks that link only two amino acids, mature crosslinks can link three or more amino acids together; however, trivalent, mature crosslinks linking three amino acids are the most common (Avery & Bailey, 2008).

Pyridinolines are trivalent crosslinks composed of an aromatic, six-membered, nitrogen-containing ring group. Two mechanisms have been proposed for the formation of the pyridinoline crosslinks. Eyre and Oguchi proposed that pyridinolines are derived from the reaction of the ketoimine HLKLN with another ketoimine (HLKLN or LKLN), while Robins and Duncan proposed the reaction of an Hyl aldehyde with a ketoimine (HLKLN or LKLN) (Avery & Bailey, 2008; Eyre & Oguchi, 1980; S. P. Robins & Duncan, 1983). There are two types of pyridinolines that may form: hydroxylysyl pyridinoline (also called pyridinoline or PYD) and lysyl pyridinoline (also called deoxy-pyridinoline or DPD). Pyridinoline crosslinks are found in cartilage, tendon, ligament, bone, and several other tissues (Eyre, Paz, & Gallop, 1984).

Pyrroles are trivalent crosslinks composed of a five-membered, nitrogen-containing ring group. As with pyridinolines, two mechanisms have been proposed for the formation of the pyrroles. Kuypers *et al.* proposed the reaction of a Lys aldehyde with an aldimine (deH-HLNL or deH-LNL), while Hanson and Eyre proposed the reaction of the aldimine deH-HLNL with a ketoimine (HLKLN or LKLN) (Avery & Bailey, 2008; Hanson & Eyre, 1996; Kuypers, Tyler,

Chapter 2: Background

Kurth, Jenkins, & Horgan, 1992). Thus, there are two types of pyrroles that may form: hydroxylysyl pyrrole (also called pyrrole or PYL) and lysyl pyrrole (also called deoxy-pyrrole or DPL). Pyrrole crosslinks are found in mature, mineralized tissues such as bone and some tendons (Eyre & Wu, 2005).

Histidinyl hydroxylysino-norleucine (HHL) is a trivalent crosslink containing histidine. The proposed mechanism for the formation of HHL is the reaction of the aldimine deH-HLNL with a histidine from the triple-helical domain. HHL is found almost exclusively in mature skin and the cornea of the eyes, as well as in some tendons and ligaments (Avery & Bailey, 2008; Yamauchi, Chandler, Tanzawa, & Katz, 1996).

Dehydro-histidino-hydroxymerodesmosine (deH-HHMD) is a tetravalent crosslink. The proposed mechanism for the formation of deH-HHMD involves an initial condensation reaction between two Lys aldehydes, followed by reaction with histidine and Hyl. deH-HHMD is believed to be found in tendons, ligaments, skin, and cornea (Fujii et al., 1994; Yamauchi, 1996). However, the existence of deH-HHMD is still under investigation and debate, since there is a possibility that the detection of deH-HMD may be an artifact of the experimental technique by which it has been characterized (Avery & Bailey, 2008).

Table 2.2: Enzymatic collagen crosslinks and vertebrate tissues in which they are found. The prefixes “Helix” and “Telo” indicate whether the precursor amino acid is from the triple-helical domain or one of the telopeptide domains, respectively. (Eyre & Wu, 2005; Saito et al., 1997; Yamauchi, 1996; S. P. Robins & Duncan, 1983; Kuypers et al., 1992; Yamauchi et al., 1996; Fujii et al., 1994; Amiel & Kleiner, 1988; Marumo, Saito, Yamagishi, & Fujii, 2005; Sims & Bailey, 1992).

Group	Crosslink	Acronym	Precursors	Vertebrate tissues
Aldimines	Dehydro-hydroxylysino-norleucine	deH-HLNL	Telo Lys aldehyde + Helix Hyl	Skin, tendon, ligament, & cartilage
	Dehydro-lysino-norleucine	deH-LNL	Telo Lys aldehyde + Helix Lys	Skin
Ketoimines	Hydroxylysino-keto-norleucine	HLKNL	Telo Hyl aldehyde + Helix Hyl	Bone, cartilage, tendon, & ligament
	Lysino-keto-norleucine	LKNL	Telo Hyl aldehyde + Helix Lys	Bone
Pyridinolines	Hydroxylysyl pyridinoline	PYD	Telo Hyl aldehyde + HLKNL	Mature bone, cartilage, tendon, & ligament
	Lysyl pyridinoline	DPD	Telo Hyl aldehyde + LKNL	Mature bone, dentin, & other calcified tissues

Chapter 2: Background

Pyrroles	Hydroxylysyl pyrrole	PYL	Telo Lys aldehyde + HLKNL	Mature mineralizing tissues & some tendons
	Lysyl pyrrole	DPL	Telo Lys aldehyde + LKNL	Mature bone
Histidine-derived	Histidinyl-hydroxylysino-norleucine	HHL	deH-HLNL + Helix histidine	Mature skin, cornea, & some tendons
	Dehydro-histidino-hydroxymero desmosine	deH-HHMD	Telo Lys aldehyde + Telo Lys aldehyde + Histidine + Hyl	Tendons, ligaments, skin, & cornea

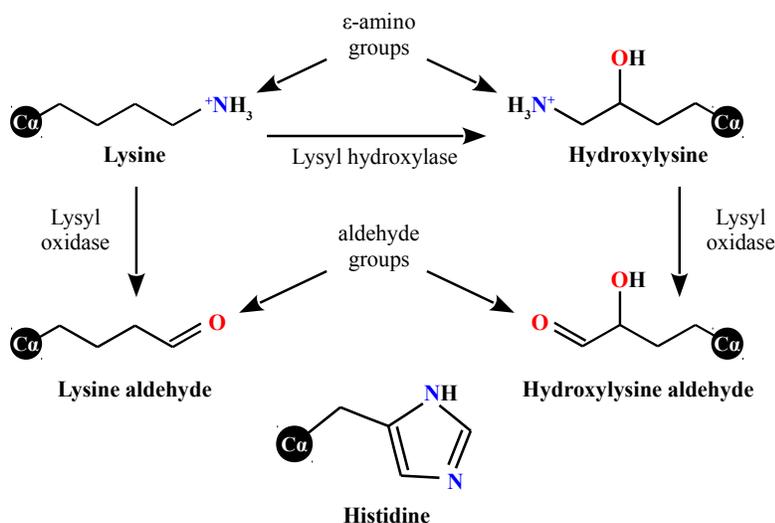


Figure 2.2: Chemical structures of amino acids that serve as precursors to enzymatic crosslinks.

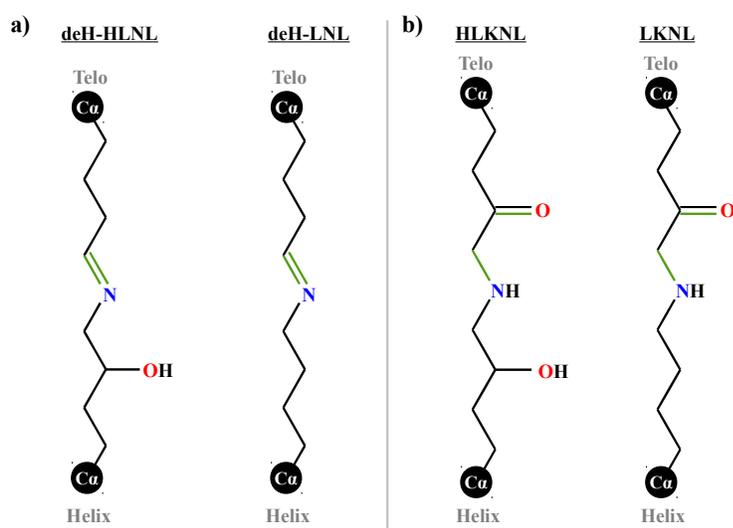


Figure 2.3: Chemical structures of divalent, immature collagen crosslinks. These four crosslinks are grouped as (a) aldimines (deH-HLNL and deH-LNL) and (b) ketoimines (HLKNL and LKNL). “Helix” and “Telo” indicate whether the precursor amino acid is from the triple-helical domain or one of the telopeptide domains, respectively. The color scheme is as follows: alpha carbon atoms (white text within

Chapter 2: Background

black circles), nitrogen atoms (blue text), oxygen atoms (red text), hydrogen atoms (black text), and newly formed covalent crosslink bonds (green lines). These drawings were generated using ACD/ChemSketch based on structures depicted by Bailey *et al.* (Bailey, Paul, & Knott, 1998).

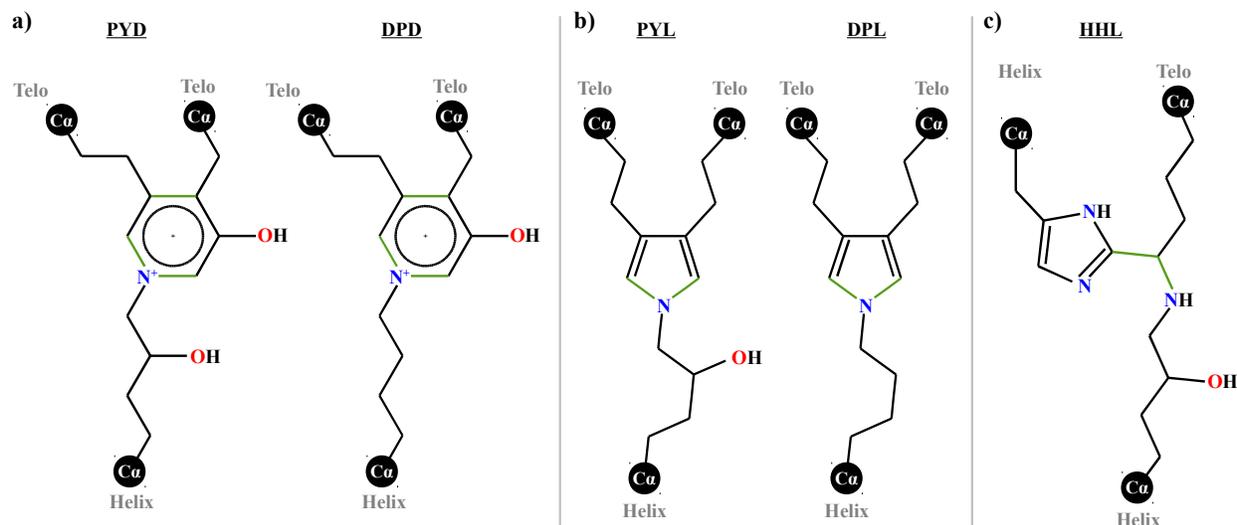


Figure 2.4: Chemical structures of trivalent, mature collagen crosslinks. These five crosslinks are grouped as (a) pyridinolines (PYD and DPD), (b) pyrroles (PYL and DPL), and (c) a histidine-derived crosslink (HHL). “Helix” and “Telo” indicate whether the precursor amino acid is from the triple-helical domain or one of the telopeptide domains, respectively. The color scheme is as follows: alpha carbon atoms (white text within black circles), nitrogen atoms (blue text), oxygen atoms (red text), hydrogen atoms (black text), and newly formed covalent crosslink bonds (green lines). These drawings were generated using ACD/ChemSketch based on structures depicted by Bailey *et al.* (Bailey *et al.*, 1998).

2.5.4. Crosslinking disorders

Connective tissue disorders such as Bruck syndrome type 2, Ehlers-Danlos syndrome type 6A, Cutis laxa type 4, and scleroderma have been linked to improper collagen crosslinking. Most of these disorders have been associated with genetic mutations that directly or indirectly affect post-translational modifying enzymes such as lysyl hydroxylases and lysyl oxidases. Lysyl hydroxylases are enzymes that convert the amino acid lysine to hydroxylysine, which is required for most of the immature and mature collagen crosslinks. Lysyl oxidases are copper-dependent enzymes that convert the side chain amino group of lysines and hydroxylysines into a reactive aldehyde; this reactive aldehyde is essential for crosslink formation (Eyre & Wu, 2005).

Bruck syndrome is a rare congenital disorder that causes brittle bones, joint stiffness, and a lack of joint mobility, leading to a high prevalence of bone fractures. Although the prevalence of Bruck syndrome is not well-known, according to a European report, its prevalence has been estimated at less than 1 in 1 million, and there have been less than 40 published cases (Orphanet

Chapter 2: Background

Report Series, 2010). Mutations to the procollagen-lysine 2-oxyglutarate 5-dioxygenase (PLOD2) gene, also known as the lysyl hydroxylase 2 (LH2) gene, have been linked to Bruck syndrome type 2 (van der Slot et al., 2003). The LH2 gene is located on human chromosome 3, where it codes for a telopeptide lysyl hydroxylase enzyme (van der Slot et al., 2003). Van der Slot *et al.* have implicated two missense mutations in the PLOD2 gene, leading to an underhydroxylation of telopeptide lysines, and thereby, to a reduction of the mature pyridinoline crosslinks (van der Slot et al., 2003). It has been proposed that the telopeptide lysyl hydroxylase affected by Bruck syndrome is bone-specific, since tissues such as ligaments and cartilage are unaffected by this condition (Bank et al., 1999).

Ehlers-Danlos syndrome (EDS) represents a set of disorders that have a prevalence of 1 in 5,000-10,000 people (Bissonnette, Luginbuehl, Marciniak, & Dalens, 2006; H. Chen, 2006). This group of disorders is characterized by skin fragility, joint hypermobility and laxity, and tissue hyperextensibility, amongst other symptoms (H. Chen, 2006). EDS type 6 (also called kyphoscoliotic type EDS) is a sub-type of EDS that displays autosomal recessive inheritance and can cause, in addition, kyphoscoliosis (abnormal lateral and posterior spinal curvatures), severe muscle hypotonia (reduced muscle tone), and fragility of the ocular tissues (Bissonnette et al., 2006). EDS type 6 has been linked with mutations to the PLOD1/LH1 gene on human chromosome 1 that codes for a helical lysyl hydroxylase (H. Chen, 2006). In particular, EDS type 6A exhibits drastically reduced lysyl hydroxylase activity which leads to an underhydroxylation of lysines and a decrease in the ratio of hydroxylysyl to lysyl pyridinoline (PYD:DPD ratio) (H. Chen, 2006).

Cutis laxa encompasses a group of very rare disorders that cause loose, inelastic, and thickened skin, and laxity of other connective tissues (Wynbrandt & Ludman, 2008). Cutis laxa exists in both inherited and acquired forms (Rongioletti, 2003). Cutis laxa type 4, also known as occipital horn syndrome and formerly known as EDS type 9, is an X-linked recessive form of Cutis laxa that can also lead to the formation of calcium deposits on the occipital bone of the skull, diverticula of the bladder, and hernias (Rongioletti, 2003; Wynbrandt & Ludman, 2008). Cutis laxa type 4 has been associated with improper copper transport, due to mutations to the ATP7A gene on the X-chromosome that encodes for the alpha polypeptide of the Cu²⁺-transporting ATPase. These mutations lead to a reduction in the activity of this ATPase leading to

Chapter 2: Background

defective copper uptake, and thereby, a reduction in the activity of the copper-dependent enzyme lysyl oxidase (Wynbrandt & Ludman, 2008).

Morphea, a localized form of systemic scleroderma, is a disorder that causes a localized hardening of skin, usually in the form of skin lesions and plaques. The mechanism of this disease is believed to be due to an accumulation of collagen protein in the skin (Uitto, Bauer, & Eisen, 1979). Additionally, Ishikawa *et al.* have reported that a mature skin crosslink, HHL, was present in elevated levels in the biopsied skin of patients presenting with systemic scleroderma (Ishikawa *et al.*, 1998). Tamura and Ishikawa later showed that HHL was also elevated in localized scleroderma (Morphea) (Tamura & Ishikawa, 2001). Brinckmann *et al.* reported the mature pyridinoline crosslinks to be elevated in skin affected by localized scleroderma (Morphea) (Brinckmann *et al.*, 2001); pyridinoline is not a major mature crosslink in skin (Saito *et al.*, 1997), thus an elevation of pyridinolines in skin is abnormal.

Table 2.3: Disorders and conditions affecting collagen crosslinking.

Condition	Affected enzyme(s)	Crosslink disruption
Bruck syndrome, type 2 (PLOD2/LH2 gene)	↓ telopeptide lysyl hydroxylase	↓ PYD & DPD
EDS, type 6A (PLOD1/LH1 gene)	↓ helical lysyl hydroxylase	↓ PYD:DPD ratio
Cutis laxa, type 4 (ATP7A gene)	↓ copper transport	↓ lysyl oxidase activity
Scleroderma	↑ lysyl oxidase expression	↑ HHL & PYD

2.6. Structural Hierarchy and Mechanics of Collagen Type I

2.6.1. Structural hierarchy

Connective tissues such as tendons and ligaments are rich in collagens (primarily type I collagen), which can often comprise 73-99% of a tendon or ligament by dry weight (Fujii *et al.*, 1994). Within these tissues, collagen type I molecules are assembled into microfibrils, which are further assembled into fibrils. These fibrils are bundled into fibers, which are further grouped into compartments called fascicles. These fibers are observed to have a wavy or crimped morphology when viewed with optical microscopy (the reader is encouraged to see [Plate 3.2](#) from (Moran & Rowley, 2011) for images of these fibrils and fibril bundles). These fibers and fascicle structures also contain fibroblast-like cells and a matrix composed of non-collagenous

Chapter 2: Background

macromolecules such as proteoglycans and elastin, water, ions, and other extracellular matrix components (Weiss & Jayson, 1982) (see Table 2.4 for the size scales of these hierarchical levels).

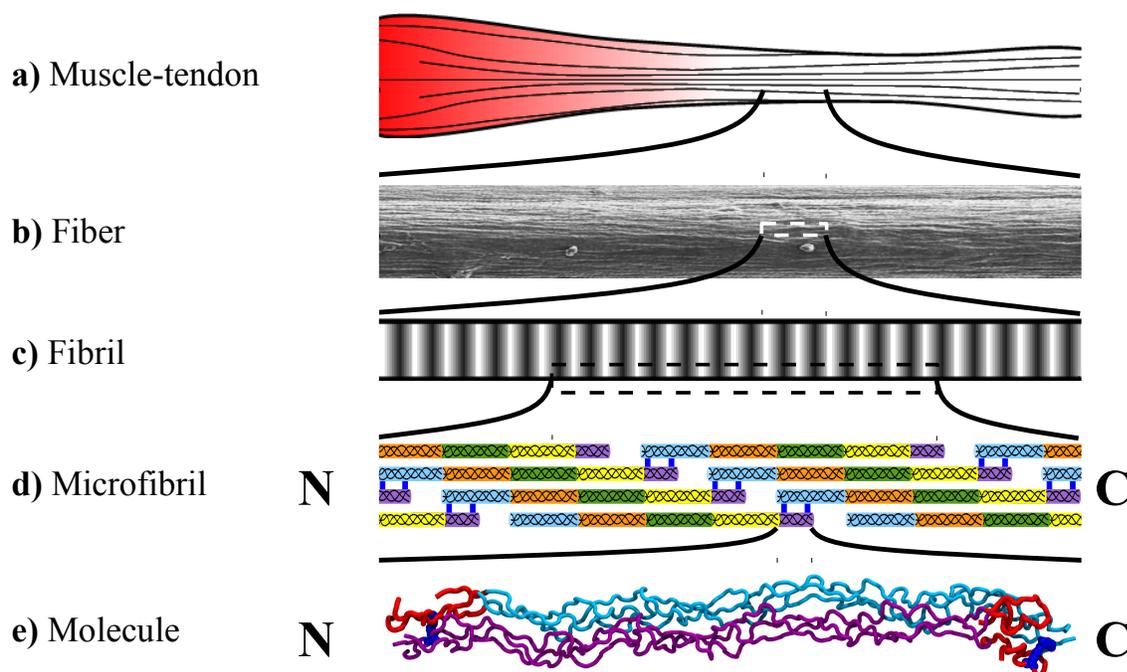


Figure 2.5: A schematic of the hierarchy of collagen type I from the fiber to the molecule. A muscle-tendon unit is shown in (a). Tendons are typically composed of fascicles (not shown), which are comprised of collagen fibers (b). These fibers are composed of collagen fibrils (c), which display a characteristic D-periodic banding pattern (e.g., 64 to 67 nm) (Birk & Bruckner, 2005). These fibrils are further composed of microfibrils (d), each of which contains an assembly of collagen molecules (e.g., five collagen molecules). The molecules are staggered in the axial (length-wise) direction with alternating overlap and gap regions; these overlap and gap regions produce a D-periodic banding pattern that can be visualized experimentally when collagen fibrils are stained with heavy-metal compounds and viewed with transmission electron microscopy. Collagen crosslinks (colored in blue) are most often located within the overlap regions inter-connecting the ends of the collagen molecules. Each molecule is shown, here, using five colors (cyan, orange, green, yellow, and purple) to emphasize the periodicity of their arrangement within collagen microfibrils and fibrils. At the molecular level (e), the crosslinks tend to be specifically located between the triple-helical domain (e.g., cyan and purple) and one of the non-triple-helical telopeptide domains (red). (b) contains an image of a reconstituted rat tail tendon collagen fiber that was taken with an FEI Quanta 600 FEG environmental scanning electron microscope through the Nanoscale Characterization and Fabrication Laboratory, a facility operated by the Institute for Critical Technology and Applied Science. This image was acquired as part of our experimental work (see Appendix H.3). The rendering in (e) is from the molecular modeling study described in Chapter 3 for Aim 1; this molecular rendering was generated using the Visual Molecular Dynamics computer program and the Tachyon ray tracing code (Humphrey, Dalke, & Schulten, 1996; Stone, 1998). The schematics shown in (a), (c), and (d) were all drawn using the LibreOffice and Microsoft Office computer programs.

2.6.2. Mechanics: Elastic moduli

Collagenous tissues, such as tendon, ligament, and skin, display a characteristic stress-strain relationship that usually includes an initial toe-region with a low slope that transitions through an upward curving heel region into a linear region with a higher slope. This linear region continues to a yield region and ultimately to failure. The toe-region has been ascribed to the straightening of wavy or crimped collagen fibers (fibril bundles) and of the collagen molecules within the gap regions of the fibrils (Gupta, 2008). The linear region has been attributed to the stretching of the collagen fibrils (Haut, 2002), while yield and partial failure have been associated with slippage between the fibrils or to the rupture of the fibrils (Curwin, 2005).

The mechanical properties of individual collagen molecules and collagen-like triple-helical molecules have been investigated both experimentally and through theoretical modeling. Experimentally derived elastic moduli at the level of the collagen type I molecule have ranged from 0.35 to 12.2 GPa (Cusack & Miller, 1979; Harley, James, Miller, & White, 1977; Sasaki & Odajima, 1996a; Sun et al., 2002), while theoretical estimates have been made between 4 and 4.8 GPa (Gautieri, Buehler, & Redaelli, 2009; Lorenzo & Caffarena, 2005). At the microfibril level, there have been no experimental investigations yet, but there have been a couple of theoretical predictions. Gautieri *et al.* employed molecular dynamics simulations to predict elastic moduli for hydrated, uncrosslinked microfibrils (0.3 and 1.2 GPa at lower and higher strains, respectively) (Gautieri et al., 2011). Hambli and Barkaoui developed a finite-element model and predicted elastic moduli for hydrated, crosslinked collagen microfibrils (0.1, 0.26, and 0.9 GPa, for unmineralized, mineralized, and further mineralized states, respectively) (Hambli & Barkaoui, 2012). It has been proposed that the reason that there have been no experimental characterizations at the microfibril level is that the microfibrils are highly integrated through the interdigitation of their molecules and through inter-microfibrillar crosslinks, thus making intact microfibrils difficult to isolate (Orgel et al., 2006). At the fibril level, there have been several recent experimental investigations and a couple of theoretical investigations. Experimentally derived moduli have ranged from 0.240 to 2.89 GPa for lower strains (Hang & Barber, 2011; Shen et al., 2008, 2010; Svensson, Hassenkam, Grant, & Magnusson, 2010; van der Rijt, 2004; Yang et al., 2012) and at 12 GPa for higher strains, while theoretical estimates have been made from 2.5 to 4.36 GPa for lower strains (Buehler, 2008; Gautieri et al., 2012) and at 38 GPa for higher strains (Buehler, 2008) (see Table 2.4 for a summary of these low-strain elastic moduli).

Chapter 2: Background

Table 2.4: Hierarchy and mechanics of collagen type I from the molecule to the fiber.

Level	Diameter	Length	Low-Strain Elastic Modulus
Molecule	~ 1.5 nm	~ 300 nm	0.35 to 12.2 GPa (Cusack & Miller, 1979) (Harley et al., 1977) (Sasaki & Odajima, 1996a) (Sun et al., 2002)
Microfibril	~ 3.5 to 4 nm	~ hundreds of nm	0.100 to 1.2 GPa (Gautieri et al., 2011) (Hambli & Barkaoui, 2012)
Fibril	Tens to hundreds of nm	Tens to hundreds of μm	0.240 to 2.89 GPa (Eppell et al., 2006) (Hang & Barber, 2011) (Shen et al., 2008) (Shen et al., 2010) (Svensson et al., 2010) (van der Rijt, 2004) (Yang et al., 2012)
Fiber (fibril bundle)	~ 1 to tens of μm	~ mm to cm	54.3 to 570 MPa (Kato et al., 1989) (Miyazaki & Hayashi, 1999) (Pins & Silver, 1995)

2.6.3. Mechanics: Deformation mechanisms

Nemetschek *et al.* used synchrotron x-ray diffraction to show that 5% macroscopic deformation in a tendon specimen corresponds to 1.5% elongation of the D-period and 0.5% molecular elongation (Nemetschek, Jonak, Nemetschek-Gansler, Riedl, & Rosenbaum, 1978). Fratzl *et al.* conducted small angle x-ray scattering (SAXS) of hydrated rat tail tendons and reported that only 40% of the applied macroscopic strain was transferred to the D-period of the fibrils. This indicated that 60% of the macroscopic strain was a result of other interactions between fibrils (e.g., inter-fibril sliding in a ground substance of proteoglycans). They also present a representative stress-strain diagram, which shows toe- and heel-regions before 5% macroscopic strain followed by a linear region up to approximately 11% strain. The authors also plotted strain in the D-period as a result of macroscopic strain, which showed that prior to the linear elastic region, the D-period remained unstrained. However, within the linear region, the D-period increased in length 40% slower than the macroscopic strain rate. This indicated that the initial non-linear regime was likely associated with mechanisms outside of the fibrils or at the level of the fiber (e.g., removal of fiber or fibrillar crimps) (Fratzl et al., 1998). Sasaki and Odajima employed a small-angle x-ray scattering (SAXS) technique to investigate deformation

Chapter 2: Background

mechanisms of collagen within saline-hydrated bovine Achilles tendon. Their SAXS data led to the proposal of three deformation mechanisms that are most likely to occur within the D-period of collagen fibrils, namely, (1) 85% due to elongation of individual collagen molecules, (2) 12.5% due to increases of the gap region between axially adjacent molecules, and (3) 2.5% due to relative sliding between laterally adjacent molecules. The last mode (relative molecular sliding) was proposed to maintain the tilt angle of the D-period (Sasaki & Odajima, 1996b). Thus, Nemetschek *et al.*, and Sasaki and Odajima have proposed that their respective x-ray diffraction data are supported by a combination of molecular stretching and molecular sliding mechanisms leading to a lengthening of the gap region and a shortening of the overlap region.

2.7. Molecular Modeling

2.7.1. Molecular mechanics

Molecular mechanics is a molecular modeling approach that is based on Newtonian/classical mechanics. In molecular mechanics, a molecule is represented as a collection of hard spheres (atoms). This representation is somewhat analogous to how molecules are represented in commercial molecule building kits that are often used in organic chemistry courses. The representation of molecules in molecular mechanics is a computer representation, but the visualization of a molecular building kit can help to better understand the general concepts of molecular mechanics. In molecular mechanics, covalent bonds between atoms are represented by spring-like connections between the spheres; these spring-like connections store potential energy when compressed or stretched and this potential energy is associated with a force acting to restore the original conformation of the spring-like connection. Non-covalent interactions between atoms are represented by attractions and repulsions between two spheres. Conceptually, this attraction and repulsion can be likened to two magnets that can attract or repel each other. The strength of an attraction and repulsion is dependent upon the distance between the spheres. Thus, there is also potential energy associated with these non-covalent interactions (i.e., an attractive or repulsive energy). These concepts of potential energy, force, and distance/position are underlying principles for this type of molecular modeling.

In order to mathematically model the hard spheres (atoms), their spring-like connections (covalent bonds), and their magnet-like attractions and repulsions (non-covalent interactions), a

Chapter 2: Background

mathematical formulation called a force field is used. A force field is a set of potential energy equations and parameters used to model the connections and interactions between modeled atoms. Although molecular mechanics force fields are mathematical formulations, force field parameters are most often empirical (e.g., based on experimental data or quantum chemistry calculations) (MacKerell et al., 1998). Several force fields have been developed to model biological macromolecules, synthetic polymers, drug-like molecules, solvents, metals, and ions. Furthermore, these force fields have been categorized as classical (e.g., AMBER, CHARMM, and CVFF), second-generation (e.g., CFF and MMFF), polarizable (e.g., variants of some of the previously mentioned force fields), reactive (e.g., ReaxFF) (Xie, 2008), and coarse-grained (e.g., MARTINI) (Marrink, Fuhrmans, Risselada, & Periole, 2009).¹

The Chemistry at HARvard Molecular Mechanics (CHARMM) force fields are a group of classical force fields that have been developed to specifically model macromolecules such as proteins, carbohydrates, nucleic acids, and lipids, as well as small molecules such as water, ions, and metals (e.g., iron). CHARMM has been developed primarily with the TIP3P water model (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983). CHARMM19 is a united-atom (also called polar hydrogen) force field that was developed to model and simulate larger proteins. CHARMM19 is different from all-atom force fields such as CHARMM22 and CHARMM27, in that CHARMM19 excludes non-polar hydrogens (e.g., for the methyl functional group, CH₃, instead of individually modeling the carbon atom and each of the three hydrogen atoms, in CHARMM19, these non-polar hydrogen atoms are grouped with their parent carbon atom, forming a new united atom type called “CH3E”). By grouping all non-polar hydrogens in this manner, the number of total atoms involved in energy calculations is reduced, thereby allowing for the simulation of larger protein structures (Brooks et al., 1983).

In the CHARMM force fields, the potential energy in a molecule is described by covalent terms (bond stretching, angle bending, bond torsion/twisting and out-of-plane bending) and non-covalent terms (Coulombic and Lennard-Jones). The Coulombic term represents electrostatic interactions between atoms (i.e., permanent dipole interactions), while the Lennard-Jones term represents transient or induced dipole interactions between atoms. Each sphere (atom) is

¹ AMBER (Assisted Model Building with Energy Refinement), CHARMM (Chemistry at HARvard Molecular Mechanics), CVFF (Consistent Valence Force Field), CFF (Consistent Force Field), MMFF (Merck Molecular Force Field), and ReaxFF (Reactive Force Field).

Chapter 2: Background

designated by a specific atom type that denotes a specific atomic configuration or bonding state for that atom. For example, each atom (e.g., oxygen) can be represented by several different atom types (e.g., a carbonyl group's oxygen, carboxylic acid's oxygen, hydroxyl group's oxygen, or the oxygen contained in a water molecule). Each atom type is assigned a mass, partial atomic charge, and Lennard-Jones atomic radius, all of which are fixed quantities. Covalent and non-covalent parameters are assigned between two or more atom types for bond stretching, angle bending, bond torsions, out-of-plane bending, Lennard-Jones, and Coulombic/electrostatic interactions. The parameters within the CHARMM force fields have been previously established based on various experimental methods and via quantum mechanics calculations (Brooks et al., 1983).

2.7.2. Molecular dynamics (MD)

Molecular dynamics (MD) is an extension of molecular mechanics that incorporates Newton's equations of motion. In the NANoscale Molecular Dynamics (NAMD) code (Kale, Bhatele, Bohm, & Phillips, 2011), MD is performed using the Verlet integration algorithm (Phillips et al., 2005). MD usually consists of the following stages: (1) the XYZ coordinates of individual atoms are determined for the current timestep, (2) potential energy of each atom is calculated for the current timestep, (3) force vectors for the current timestep are calculated for each atom by taking partial derivatives of the potential energy in the X, Y, and Z directions, (4) acceleration vectors are computed from the force vectors based on the atomic masses ($M^{-1} F_n$), and (5) velocities, positions, and forces are computed for forward timesteps. This sequence of steps is repeated iteratively for each successive timestep.

Table 2.5: Verlet integration method employed for MD calculations in NAMD. Here, r = position, F = force, U = potential energy, v = velocity, M = atomic mass, and Δt = integration timestep. Thus, $M^{-1} * F$ represents force over mass (i.e., acceleration). The subscripts n , $n + 1/2$, $n + 1$ represent quantities for the current timestep, $1/2$ timestep in the future, and 1 timestep in the future, respectively (Phillips et al., 2005).

Start with initial X-Y-Z positions for atoms:	r_n
Compute force vectors:	$F_n = F(r_n) = \nabla U(r_n)$
Compute velocity vectors:	$v_{n+1/2} = v_n + M^{-1} * F_n * \Delta t/2$
Compute new positions:	$r_{n+1} = r_n + v_{n+1/2} * \Delta t$
Compute new force vectors:	$F_{n+1} = F(r_{n+1}) = \nabla U(r_{n+1})$
Compute new velocities:	$v_{n+1} = v_{n+1/2} + M^{-1} * F_{n+1} * \Delta t/2$

2.7.3. Constant velocity MD pulling techniques

Steered molecular dynamics (SMD) is a molecular dynamics technique inspired by atomic force microscopy (AFM). In constant velocity-steered molecular dynamics (cv-SMD), a virtual spring (analogous to an AFM cantilever) is attached to a selected group of atoms. The distal end of this virtual spring (further from the molecule) moves with a constant velocity, while the proximal end of this virtual spring (closer to the molecule) pulls the center-of-mass of the selected atoms in a chosen direction. The cv-SMD pulling simulation can sometimes be conceptualized as a system of two springs in series; the virtual spring and the inherent stiffness associated with the molecule that is being pulled. The potential energy in the virtual spring is given by $U_{cv-SMD} = \frac{1}{2} k_{SMD} [x - x_0 - v_{SMD} * t]^2$, where k_{SMD} is the cv-SMD spring constant (stiffness of the virtual spring), v_{SMD} is the cv-SMD pulling velocity, t is the elapsed time, x is the current position of the virtual spring's proximal end at time t , and x_0 is the initial position of the virtual spring's proximal end. The force needed to pull the selected atoms is given by $F = k_{SMD} [x_0 + v_{SMD} * t - x]$ (Lu, Krammer, Isralewitz, Vogel, & Schulten, 2000). SMD has been used to study the mechanical functions of proteins as well as protein unfolding pathways as reviewed by Isralewitz *et al.* (Isralewitz, Gao, & Schulten, 2001).

Moving constraints is somewhat similar to cv-SMD with a few notable differences. In cv-SMD, the center-of-mass of the designated atoms is attached to a virtual atom through a virtual spring. The virtual atom moves at a constant velocity. In harmonic moving constraints, each designated atom is attached to its own virtual atom through a virtual spring. Each of the virtual atoms moves at a constant velocity along a specified direction. Each timestep, the designated atoms are moved according to $r_{virtual,i}(t) = r_{virtual,i}(0) + v * t$, and the potential energy in each virtual spring is given by $U_i(t) = k_i(r_{virtual,i}(t) - r_{pulled,i}(t))^2$ (Bhandarkar *et al.*, 2011).

2.7.4. Explicit and implicit solvation

The solvent surrounding a protein solute can influence its folding behavior and structure, its conformational dynamics, its stability, and its function (Hassan, Gracia, Vasudevan, & Steinbach, 2005); therefore, simulations involving proteins and other biological molecules are often performed using some form of solvation. Such solvation methods are typically categorized as being explicit or implicit representations of solvent (Mackerell, 2004).

Chapter 2: Background

Explicit solvation involves the actual presence of solvent molecules in the simulation. Explicit solvents generally provide greater accuracy, solvent viscosity effects, and hydrogen bond directionality at the cost of increased computational time and/or resources. Water, in particular, has been represented using three-site (e.g., TIP3P (Jorgensen et al., 1983), SPC, SPC/E (Berendsen, Grigera, & Straatsma, 1987)), four-site (e.g., TIP4P (Jorgensen et al., 1983)), and five-site models (e.g., TIP5P (Mahoney & Jorgensen, 2000)). Validation of such water models against experimental data is often made based on how well the models reproduce certain physical properties of water (e.g., liquid density, diffusion coefficient, and potential energy) (P. Mark & Nilsson, 2001). Since different water models were developed with different force fields, the selection of a water model often depends upon the chosen force field (e.g., TIP3P with the AMBER, OPLS, and CHARMM force fields, TIP4P with the OPLS force fields, and SPC with the GROMOS force field) (Mackerell, 2004) and the chosen molecular modeling software (e.g., TIP3P or TIP4P with NAMD 2.8) (Bhandarkar et al., 2011).

Implicit solvation involves the simulation of solvent effects without the presence of solvent molecules in the simulation. Due to the reduction in the number of atoms in the simulation, implicit solvation usually offers increased computational efficiency and access to increased timescales compared to explicit solvation. However, implicit solvent approaches lead to the absence of solvent viscosity effects that would normally slow down protein dynamics, the inability to account for a non-homogenous solvent environment (e.g., water molecules with sodium and chloride ions), and the absence of directionality in hydrogen bonds (Hassan et al., 2005). Several different implicit solvation approaches have been developed including those based on solutions or approximations to the Poisson/Poisson-Boltzmann equations, solvent-accessible surface area, or the screening/attenuation of electrostatic interactions (Feig & Brooks, 2004).

Generalized Born implicit solvent (GBIS) methods involve approximations to the Poisson-Boltzmann equation. For GBIS methods, each solute atom experiences a de-screening of electrostatic forces depending on the local surroundings of each solute atom. The solute atoms and implicit solvent are assigned different dielectric constants (e.g., consider the case of a lower dielectric constant for the solute atoms and a higher dielectric constant for the implicit solvent). Solute atoms that are surrounded by and in close proximity to other solute atoms experience less

Chapter 2: Background

de-screening due to the low dielectric constant of the solute atoms; in contrast, solute atoms that are surrounded by less solute atoms (i.e., more exposed to the surrounding vacuum) experience more de-screening due to the higher dielectric of the implicit solvent (Bhandarkar et al., 2011; Onufriev, Bashford, & Case, 2000). In the NAMD code (version 2.8 and later), the GBIS approach of Onufriev, Bashford, and Case is implemented (Bhandarkar et al., 2011), namely, $GB^{OBC}(I)$ and $GB^{OBC}(II)$, which differ in their GBISDelta, GBISBeta, and GBISGamma values (Onufriev, Bashford, & Case, 2004). Due to the sizes of our molecular systems (~5,200 and ~38,900 solute atoms for Aims 1 and 2, respectively), we decided to utilize GBIS as implemented in NAMD.

3.1. Introduction

Collagens are extracellular matrix (ECM) proteins that are found in nearly all eukaryotic life except for plants and protozoa (Urich, 1994). In mammals, collagens are especially abundant and can comprise up to one-third of all protein in the body by weight (Williams, 1978). There are approximately 27 different types of collagens that have been identified (K. von der Mark, 2006); type I collagen is the most prevalent and is found in connective tissues such as tendon, ligament, bone, skin, and the cornea of the eyes. Type I collagen functions to provide tensile strength to these connective tissues, and it serves as a structural framework for cells and other ECM components such as fibronectin, proteoglycans, and bone mineral (Sweeney et al., 2008).

Type I collagen is a rod-shaped molecule with a length of ~300 nm and diameter of ~1.5 nm (Birk & Bruckner, 2005). It is composed of three amino acid chains (often called α -chains): two $\alpha 1$ chains (1054 amino acids long) and one $\alpha 2$ chain (1026 amino acids long). The entire molecule is composed of an N-telopeptide domain at the beginning (16 amino acids for $\alpha 1$ and 6 for $\alpha 2$), a triple-helical domain (1014 for $\alpha 1$ and $\alpha 2$), and a C-telopeptide domain at the end (24 for $\alpha 1$ and 6 for $\alpha 2$) (The UniProt Consortium, 2012). The N-telopeptides display an extended coil structure, while the longer $\alpha 1$ C-telopeptides possess a hook/hair-pin structure (Orgel et al., 2006). These amino acid lengths are specific to rat collagen type I and can vary slightly between different species (The UniProt Consortium, 2012).

Fibril-forming collagens (e.g., types I, II, III, V, and XI) are able to assemble into structures called fibrils, which are stabilized by covalent bonds termed crosslinks. Crosslink formation usually involves enzymes termed lysyl oxidases that act on the amino acids lysine (Lys) and hydroxylysine (Hyl). Lysyl oxidases act specifically at the telopeptides, leading to the crucial involvement of the N- and C-telopeptides for these crosslinks. There are several types of crosslinks that can form, depending upon factors such as tissue type, age, and health (Eyre & Wu, 2005; Wynbrandt & Ludman, 2008). In addition, there are certain crosslinks that may form in old age or in certain disease states (e.g., diabetes) outside of the regulation of enzymes and without the involvement of telopeptides (Avery & Bailey, 2008). The focus of this paper is on one of the enzyme-derived crosslinks.

Chapter 3: Specific Aim 1

We investigated the influence of crosslink location (i.e., N vs. C-crosslinking site) using a crosslink named dehydro-hydroxylysino-norleucine (deH-HLNL). We used molecular dynamics to model the stretching and shearing of the ends of two crosslinked collagen type I molecules with deH-HLNL positioned at the N- and C-crosslinking sites; these simulations were designed to model modes of molecular deformation that have been proposed through the x-ray diffraction of bovine Achilles tendons (Sasaki & Odajima, 1996b). In particular, we employed a constant velocity pulling technique called moving constraints, which is similar to constant velocity steered molecular dynamics (cv-SMD). cv-SMD and its related techniques have been used to simulate the mechanical functions of proteins and protein unfolding pathways (Isralewitz et al., 2001), and to predict the Young's modulus of collagen-like molecules (Gautieri et al., 2009; Lorenzo & Caffarena, 2005) in close agreement with experimental data (Sasaki & Odajima, 1996a). The pulling velocity and the stiffness of the virtual springs have been shown to influence such simulations. Faster pulling rates can lead to over-estimated mechanical properties (Gautieri et al., 2009). Overly compliant virtual springs can result in under-estimated mechanical properties, while overly stiff virtual springs can cause increased numerical noise (Lorenzo & Caffarena, 2005). Thus, such simulations require a sufficiently reduced pulling velocity and sufficiently stiff virtual springs.

3.2. Methods

3.2.1. Molecular modeling programs and parameters

The molecular model was constructed using Discovery Studio (Accelrys Software, Inc., 2011) and Visual Molecular Dynamics (VMD) (Humphrey et al., 1996). Simulations were carried out using the NAMD code (Phillips et al., 2005), and the Tachyon ray tracing code was used with VMD to generate molecular renderings (Stone, 1998). The CHARMM19 united-atom force field was used (Brooks et al., 1983) with additional parameters derived for hydroxyproline and hydroxylysine (CHARMm polar-hydrogen (Accelrys Software, Inc., 2011)), and deH-HLNL (CHARMM22 (MacKerell et al., 1998)) (see Appendix A).

Chapter 3: Specific Aim 1

Equation 3.1: Total potential energy in the CHARMM19 united-atom force field expressed as a summation of bond stretching (bonds), angle bending (angles), bond rotation (torsions), out-of-plane bending (impropers), Lennard-Jones, and Coulombic terms. k_b , k_θ , k_φ , and k_ω represent force constants. r , θ , φ , and ω represent variables (bond length, bond angle, torsion angle, and out-of-plane bending angle, respectively). r_0 , θ_0 , and ω_0 represent fixed equilibrium or ideal values (bond length, bond angle, and out-of-plane bending angle, respectively). n is a measure of bond torsion periodicity. A_{ij} and B_{ij} are Lennard-Jones parameters that depend upon E_{min} , R_{min} , and α parameters that are specified in the CHARMM19 force field. Starting from the case of two identical atom types interacting non-covalently, E_{min} is the minimum of their Lennard-Jones energy, R_{min} is $1/2$ of their center-to-center inter-atomic distance at E_{min} , and α is a polarizability parameter. Combining rules are then used to describe interactions between non-identical atom types. q_i and q_j represent fixed partial atomic charges for atoms i and j , while ϵ_0 represents the dielectric constant. S_{vdw} and S_{elec} are switching and shifting functions, respectively. r_{ij} is the center-to-center distance between atoms i and j . r_{on} is the switching distance, and r_{off} is the cutoff distance. $S_{vdw} = 1$ when $r_{ij} \leq r_{on}$, and $S_{vdw} = 0$ when $r_{ij} \geq r_{off}$. $S_{vdw} = (r_{off}^2 - r_{ij}^2)^2 * (r_{off}^2 + 2r_{ij}^2 - 3r_{on}^2) / (r_{off}^2 - r_{on}^2)^3$ when $r_{on} < r_{ij} < r_{off}$. $S_{elec} = (1 - r_{ij}^2 / r_{off}^2)^2$ when $r_{ij} < r_{off}$, and $S_{elec} = 0$ when $r_{ij} \geq r_{off}$ (Brooks et al., 1983; Brunner, Dalke, Gursoy, Humphrey, & Nelson, 2008).

$$E = \sum_{Bonds} k_b (r - r_0)^2 + \sum_{Angles} k_\theta (\theta - \theta_0)^2 + \sum_{Torsions} |k_\varphi| - k_\varphi \cos(n\varphi) + \sum_{Impropers} k_\omega (\omega - \omega_0)^2$$

$$+ \sum_{Lennard-Jones} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right] * S_{vdw} + \sum_{Coulombic} \left[\frac{q_i q_j}{4\pi \epsilon_0 r_{ij}} \right] * S_{elec}$$

3.2.2. Building of the molecular models

The triple-helical structure in our molecular model was based upon a 29-amino acid long collagen-like (proline-proline-glycine)_n crystal structure (1k6f.pdb) (Berisio, Vitagliano, Mazzarella, & Zagari, 2002). Using Discovery Studio, crystallographic water molecules were removed, and the structure was replicated and spliced together, similar to methods described by Chen *et al.* (J. M. Chen, Kung, Fairheller, & Brown, 1991), to form a single collagen-like molecule containing 1054 amino acids per chain (Chapman & Hulmes, 1984; Orgel et al., 2006). This collagen-like molecule was then divided into five molecular segments by breaking the appropriate peptide bonds, after which the cleaved ends were adjusted into positively charged amino groups (-NH₃⁺) or negatively charged carboxylic acid groups (-C(=O)-O⁻) as appropriate. These five segments were then aligned at their N-termini, and they were translated into specific x-y coordinates (see Appendix B.1), in order to represent a quasi-hexagonal compressed microfibril with a length equivalent to an entire collagen type I D-period (Fraser et al., 1983). At this point, the molecular model contained three chains of identical length (1054 amino acids); however, collagen type I contains two identical chains ($\alpha 1$ chains) and a third chain ($\alpha 2$ chain)

Chapter 3: Specific Aim 1

that is different in both length and sequence (e.g., 1026 amino acids for rat collagen type I) (Chapman & Hulmes, 1984; Orgel et al., 2006). Therefore, one of the three chains (selected to be the $\alpha 2$ chain) was trimmed from 1054 to 1026 amino acids leading to an $\alpha 1$ - $\alpha 2$ - $\alpha 1$ collagen-like microfibril unit. Thus, the model contained an N-telopeptide domain at the beginning (16 amino acids for $\alpha 1$ and 6 for $\alpha 2$), a triple-helical domain (1014 for $\alpha 1$ and $\alpha 2$), and a C-telopeptide domain at the end (24 for $\alpha 1$ and 6 for $\alpha 2$). The trimming of the $\alpha 2$ chain was performed from both the N-terminus and the C-terminus; the $\alpha 2$ N-telopeptide was reduced from 16 to 6 amino acids, and the $\alpha 2$ C-telopeptide was reduced from 24 to 6 amino acids. The conformations of the N- and C-telopeptides were then adjusted based upon a rat tail tendon collagen type I x-ray fiber diffraction structure (3hr2.pdb) (Orgel et al., 2006), allowing for favorable alignments between four potential crosslinking sites (Eyre & Wu, 2005). These conformational adjustments were made by changing the positions (i.e., Cartesian coordinates) of the backbone atoms to match those of Orgel *et al.*'s 3hr2.pdb structure; energy minimization was then conducted for 20,000 timesteps to allow the side chain atoms to adjust to the changes to the positions of the backbone atoms. The two main crosslinking segments (see segments 1 and 5 in Figure B.1) were then isolated for further use. Rat (*Rattus norvegicus*) collagen type I amino acid sequences for the 1054-residue-long $\alpha 1$ chains and the 1026-residue-long $\alpha 2$ chain were then used to replace the collagen-like Pro-Pro-Gly sequences (Chapman & Hulmes, 1984; Orgel et al., 2006).

3.2.3. Energy minimization, heating, and equilibration

Energy minimization via the conjugate gradients method was performed until an NAMD gradient tolerance of 0.01 kcal/mol/Å. For heating, temperature reassignment was used to increase the temperature from 0 to 293.15 K over 10 ps. During subsequent equilibration steps, the Langevin dynamics method was used for temperature control at 293.15 K. The generalized Born implicit solvent (GBIS) approach of Onufriev, Bashford, and Case (GB^{OBC} II) was used to indirectly model the effects of solvent and ions (Onufriev et al., 2000). The solvent dielectric, ion concentration, and GBIS alpha cutoff parameters were set to 80.0, 0.150 mol/L, and 14 Å, respectively. The simulation timestep was initially 1 fs during energy minimization, heating, and the initial stages of equilibration; it was later switched to 2 fs during the remaining equilibration time (with rigid bonds and the settle algorithm for covalent bonds involving hydrogens). Non-bonded interactions were treated with a switching distance of 15 Å, a cutoff of 16 Å, and a pair

Chapter 3: Specific Aim 1

list distance of 17.5 Å. The solvent dielectric of 80.0 was selected in representation of water at 293.15 K and 1 atm, and the ion concentration of 0.150 mol/L was selected to match physiological saline solution (i.e., 9% w/v sodium chloride in water). The non-bonded distances (15-16-17.5 Å) and the GBIS alpha cutoff (14 Å) were based upon values suggested in the NAMD 2.8 user's guide for GBIS (Bhandarkar et al., 2011).

Table 3.1: Preparation steps used for our ~23-nm-long molecular models. Here, timesteps refers to the Δt parameter in Table 2.5. The spring constant used for the harmonic atom restraints was $k = 0.5$ kcal/mol/Å².

No.	Stage	Timestep	Duration	Atom constraints	Description
Only the uncrosslinked model					
1	Energy minimization	1 fs	50,000 timesteps	Fix backbone	Conjugate gradients $T = 0$ K
Applied teleopeptide conformations					
2	Energy minimization	1 fs	20,000 timesteps	Fix non-teleopeptide atoms	Conjugate gradients $T = 0$ K
3	a. Duplicated the coordinate file (.pdb) of the uncrosslinked model b. Renamed the new pdb files for the crosslinked model c. Added two identical crosslinks (deH-HLNL) to the new pdb file, and generated corresponding structure files (.psf)				
Both the uncrosslinked and crosslinked models					
4	Energy minimization	1 fs	50,000 timesteps	Restrain all C α	Conjugate gradients $T = 0$ K
5	Heating	1 fs	10,000 timesteps (10 ps)	Restrain all C α	$T = 0$ to 293.15 K
6	Equilibration	1 fs	500,000 timesteps (500 ps)	Restrain all C α	$T = 293.15$ K
7	Equilibration	2 fs	3.5 million timesteps (7 ns)	No restraints	$T = 293.15$ K

3.2.4. Constant velocity pulling

We employed the moving constraints technique implemented in NAMD (Bhandarkar et al., 2011). Each pulled atom was connected through a virtual spring of stiffness, k , to a virtual atom (moving constraint) that was set to move at a constant velocity, v . In our simulations, three N- or C-terminal alpha carbon (C α) atoms of one segment were fixed, while those of the other segment were pulled in the opposite direction (see Figure 3.2). Overall engineering strain of the molecular model was defined as the root-mean-square displacement (RMSD) of the pulled atoms in angstroms divided by the initial length of the molecular model (225.54 Å); this initial length

Chapter 3: Specific Aim 1

was measured as the RMS difference between the positions of the fixed C α atoms and the pulled C α atoms. Three parameters were investigated, namely, the spring constant, the pulling direction, and the pulling velocity.

First, with a velocity of 100 m/s and while pulling in the C-terminal direction, the spring constant parameter was varied from 0.01 to 9,999 kcal/mol/Å² in factors of ~10 (0.01, 0.10, 1.00, 10.0, 100, 1,000, and 9,999). 9,999 was used as an approximation of 10,000, due to the number of characters permitted for this parameter. The observed velocity of the pulled atoms was calculated as the slope of the RMSD of the pulled atoms versus time. Percent error was then calculated as the difference between this observed velocity and the set velocity divided by the set velocity all multiplied by 100%. Numerical noise in the potential energy versus time data was calculated as the error sum of squares (SS_{error}). The potential energy versus time data for each spring constant were fitted with a quadratic polynomial function, and SS_{error} values were calculated using the JMP statistical software as $SS_{error} = \sum (a_i - f_i)^2$ where i increases from 1 to n x-axis data points (i.e., time), a_i represents the actual data value, and f_i represents the fitted value (SAS Institute, Inc., 2010). Second, with a velocity of 100 m/s and a spring constant of 9,999 kcal/mol/Å², the pulling direction was varied in either the C-direction or N-direction. Lastly, with a spring constant of 9,999 kcal/mol/Å² and while pulling in the C-terminal direction, the pulling velocity was varied from 100 down to 1.5625 m/s in factors of two (100, 50, 25, 12.5, 6.25, 3.125, and 1.5625 m/s). A value of 9,999 kcal/mol/Å² was based upon results from the spring constant study; this value led to the least error in the observed velocity for the range of spring constants investigated.

3.2.5. Crosslink alignment, crosslink engineering strain, and crosslink strain energy

Crosslink alignment was calculated based on a line intersecting the two C α atoms of the crosslinked amino acids relative to the pulling direction. Crosslink engineering strain was defined as the change in length of the line intersecting the two C α atoms of the crosslinked amino acids divided by the initial length of this C α -to-C α line. The Cartesian coordinates for these C α atoms were obtained using VMD (see Appendix B.9 for an example TCL script).

Chapter 3: Specific Aim 1

Crosslink orientation angles were calculated as follows, where X_1 , Y_1 , and Z_1 are the coordinates for one C α atom and X_2 , Y_2 , and Z_2 are the coordinates for the second C α atom.

$$\theta_{crosslink} = 90 - \tan^{-1} \left[\frac{|Z_2 - Z_1|}{\sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}} \right]$$

Below is a representative Microsoft Excel formula that was used:

$$=90-DEGREES(ATAN(ABS(Z2-Z1)/SQRT((X2-X1)^2+(Y2-Y1)^2)))$$

Crosslink engineering strains were calculated as follows, where X_1 , Y_1 , and Z_1 are the coordinates for one C α atom and X_2 , Y_2 , and Z_2 are the coordinates for the second C α atom. The variables subscripted with “,0” denote the initial X, Y, and Z coordinates at $t = 0$; those variables not containing this additional subscript denote changing X, Y, and Z coordinates for $t > 0$.

$$\epsilon_{crosslink} = \frac{\sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2 + (Z_2 - Z_1)^2}}{\sqrt{(X_{2,0} - X_{1,0})^2 + (Y_{2,0} - Y_{1,0})^2 + (Z_{2,0} - Z_{1,0})^2}}$$

Below is a representative Microsoft Excel formula that was used:

$$=SQRT((X2-X1)^2+(Y2-Y1)^2+(Z2-Z1)^2)/SQRT((X$2-X$1)^2+(Y$2-Y$1)^2+(Z$2-Z$1)^2)$$

Crosslink strain energy was calculated as the change in total potential energy as calculated by the NAMD code using the same CHARMM19 parameter file that had been used to conduct the simulations (see Equation 3.1 for energy terms contributing to this total potential energy). This total potential energy was calculated for atoms in between the C α atoms of the crosslinks (i.e., excluding other backbone atoms of the crosslinked amino acids) (see Appendix B.8 for an example TCL script).

3.2.6. Initial lengths, extensions, and engineering strains

Initial lengths were determined based on the root-mean-square (RMS) difference between the three fixed C α atoms and the three pulled C α atoms at time zero ($t = 0$). Specifically, the position vector for the i^{th} fixed C α atom was compared to that of the i^{th} pulled C α atom (Theoretical and Computational Biophysics Group, 2012) (see Appendix B.6 for an example TCL script). Extensions (changes in length) were calculated based on the RMS displacement of the three pulled C α atoms; this RMS displacement calculation was similar to the RMS difference calculation, except that the position vector of the i^{th} pulled C α atom at a given time (t) was

Chapter 3: Specific Aim 1

compared to that of the same pulled Ca atom at time zero ($t = 0$). This RMSD displacement data was acquired using VMD's RMSD Trajectory Tool (Humphrey et al., 1996). Engineering strains were then calculated as the extension of each molecular model divided by its initial length.

Equation 3.2: RMS difference equation for the calculation of initial lengths.

$$RMS\ difference = \sqrt{\frac{\sum_{i=1}^3 (\vec{r}_{pulled,i}(0) - \vec{r}_{fixed,i}(0))^2}{3}}$$

where i = atom pair number, n = total number of atom pairs (3 pairs in this case), $\vec{r}_{pulled,i}(0)$ = initial position vector for pulled atom i , and $\vec{r}_{fixed,i}(0)$ = initial position vector for fixed atom i .

Equation 3.3: RMS displacement equation for the calculation of changes in length.

$$RMS\ displacement = \sqrt{\frac{\sum_{i=1}^3 (\vec{r}_i(t) - \vec{r}_i(0))^2}{3}}$$

where i = atom number, n = total number of atoms (3 in this case), $\vec{r}_i(t)$ = current position vector for atom i at time t , and $\vec{r}_i(0)$ = initial position vector for atom i .

3.2.7. Statistical analyses

All statistical analyses were performed using the JMP statistical software (SAS Institute, Inc., 2010). Unless otherwise stated, α (also known as the alpha value, significance level, or type I error rate), was set to 0.05. The Brown-Forsythe and Levene tests were used to initially assess if the variances were homogeneous. If so, a one-way analysis of variance (ANOVA) employing a pooled standard deviation was used, followed by a Tukey-Kramer Honestly Significant Difference (HSD) test, when appropriate. If not, then a Welch's one-way ANOVA and Student's t-tests with Bonferroni correction were used. For the Bonferroni correction, α was set to 0.05 divided by the number of t-tests (e.g., number of t-tests = $g*(g-1)/2$, where g denotes the number of groups).

3.2.8. Computing resources

The molecular dynamics simulations for these ~5,200-atom molecular models for Aim 1 were conducted using two types of computing systems: (1) the Virginia Tech Engineering Science and Mechanics (ESM) Intel 64-bit Linux computing cluster (34 compute nodes, ~20 to 30 GB RAM per node, 8 or 16 processor cores per node, and 2.21 to 2.60 GHz per core) and (2) an HP desktop computer (2 GB RAM and a Core2Duo processor with 3.16 GHz per core).

3.3. Results and Discussion

3.3.1. Higher spring constants improve accuracy in pulling velocity

From our spring constant study, we observed that stiffer virtual springs (higher spring constants) improved the accuracy of the pulling velocity; specifically, the velocity errors were statistically different between spring constants of 0.01, 0.1, 1, 10, and 100, but not between 100, 1,000, and 9,999 kcal/mol/Å² (Welch's ANOVA, $p < 0.0001$; Student's t-tests, $\alpha = 0.0024$) (Figure 3.1). We further analyzed the numerical noise associated with the spring constants; however, no conclusive differences were found via the Student's t-tests (Welch's ANOVA, $p = 0.0085$; Student's t-tests, $\alpha = 0.0024$). Thus, 9,999 kcal/mol/Å² was selected, given the order of magnitude reduction in error between 100, 1,000, and 9,999 kcal/mol/Å² (Figure 3.1). Statistical power values for the velocity error and numerical noise analyses were found to be ~1.00 and 0.695, respectively.

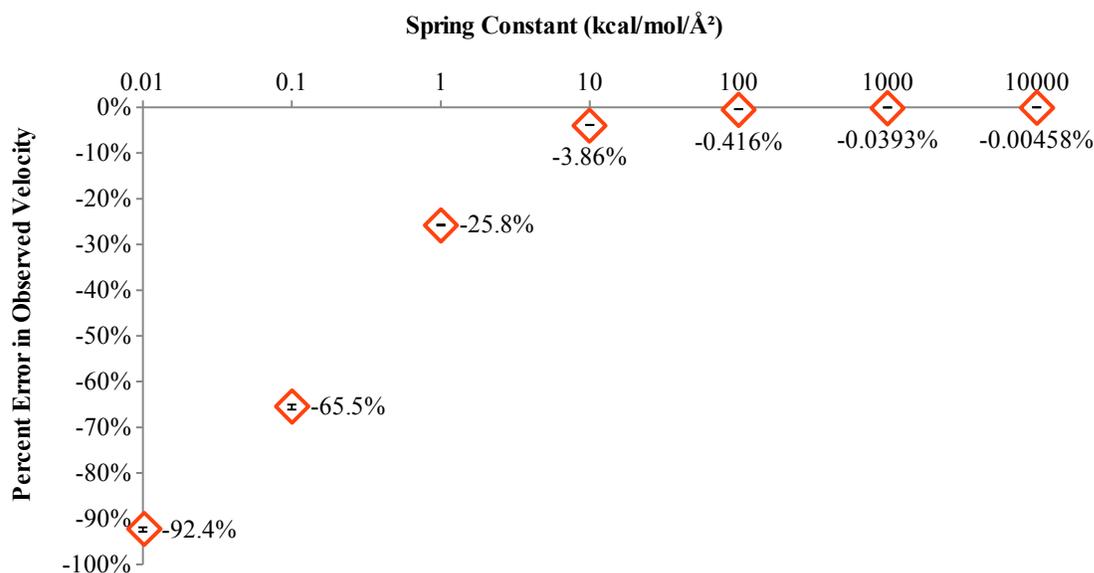


Figure 3.1: Percent error in the pulling velocity at 100 m/s. Each data point represents a mean \pm s.d. ($n = 3$). The x- and y-axes are plotted on log and linear scales, respectively, and the value adjacent to each data point denotes the y-axis value. Negative errors, here, simply indicate that the observed velocity was less than the expected velocity, since these errors were computed as $(\text{observed} - \text{expected})/\text{expected} * 100\%$.

3.3.2. Pulling direction and pulling velocity introduce bias

From our pulling direction and pulling velocity studies, the following results were observed. All strains were calculated as engineering strains, which were defined as the change in length divided by the initial length. Thus, any instance of “strain” used in reference to a change in length (e.g., overall strain or crosslink strain) refers to “engineering strain”. We found that the pulling direction led to a bias in crosslink alignment, crosslink strain, and crosslink strain energy. Specifically, when pulling in the C-direction (Figure 3.2a), the crosslink closer to the pulled atoms (C-crosslink) began to align, strain, and store energy at lower strains relative to the crosslink located further from the pulled atoms (N-crosslink). When pulling in the N-direction (Figure 3.2b), the reverse was observed. This bias was less prominent with slower pulling velocities; it was found that crosslink alignment was the least sensitive to the pulling velocity, followed by crosslink strain, and then by crosslink strain energy. At our slowest pulling velocity of 1.5625 m/s, the bias was essentially non-existent for crosslink strain energy and crosslink strain, but it was still present for crosslink alignment. Thus, data acquired from pulling in the N- or C- directions (Figure 3.3) were averaged to correct for the remaining bias. The influence of the pulling velocity on strain energy and RMS displacement for the ~23-nm-long model, crosslink strain energy, crosslink strain, and crosslink alignment is provided in Appendix D.

Figure 3.2: Snapshots of the model being pulled in the (a) C- or (b) N-direction at 1.5625 m/s. The molecular renderings included in these snapshots were generated using the Visual Molecular Dynamics (VMD) computer program and the Tachyon ray tracing code (Humphrey et al., 1996; Stone, 1998). Triple-helical domains (shown with VMD's “tube” representation, where cyan = upper segment and purple = lower segment), telopeptides (shown with VMD's “tube” representation in red), two crosslinks (shown with VMD's “VDW” representation in blue), three fixed atoms (thick vertical line with diagonal hash lines), three pulled atoms (closed circles), and three virtual atoms (open circles). The “tube” representation shows a line tracing through the backbone atoms of the amino acid chains, while the “VDW” representation depicts each modeled atom as a sphere with its van der Waals radius. “N” and “C” represent the amino/N-terminal and carboxy/C-terminal directions. A schematic of the moving constraints technique is shown in (a) ($k =$ virtual spring stiffness; $v =$ virtual atom velocity). (c) shows close-up views of the N- and C-ends from (a). In each case, three C α atoms were fixed and three C α atoms were pulled; these six atoms can be seen most clearly in (c) at 75% strain. Each of these three pulled C α atoms had its own virtual spring and virtual atom.

3.3.3. N- and C-crosslinks display two-stage recruitment

First, from the averaged data (Figure 3.3), it was observed that the N- and C-crosslinks aligned from being initially more perpendicular to the pulling direction (86.8° and 58.6° , respectively) to being more parallel at 75% overall strain ($9.99 \pm 1.04^\circ$ and $21.44 \pm 2.00^\circ$, respectively, at 1.5625 m/s) (mean \pm s.d.). At this velocity, the change in N- and C-crosslink alignment was fairly linear with increasing overall strain ($R^2 = 0.96$ and 0.86 , respectively); the N-crosslink aligned 1.68x faster than the C-crosslink, and by 75% overall strain, the N-crosslink displayed 2.15-fold greater alignment. Second, N- and C-crosslink strain was initially similar up to ~40% overall strain, after which the N-crosslink reached 1.98x that of the C-crosslink by 75% overall strain. Third, strain energy in the N- and C-crosslinks was similar up to ~60% overall strain, after which the N-crosslink reached 8.46x that of the C-crosslink by 75% overall strain. Neither crosslink began to store appreciable levels of energy until ~60% overall strain. This delayed response in both of the crosslinks is consistent with the delayed crosslink response reported by Uzel and Buehler, who conducted steered molecular dynamics simulations with a C-terminal lysine-lysine crosslink. They further developed a rheological model incorporating an elastic element, a frictional element, and a “delay” element; this delay element introduced a constant force response and was attributed to unfolding of the C-telopeptide (Uzel & Buehler, 2011).

It was also observed that the strain energy responses of the N- and C-crosslinks (Figure 3.3c) lagged behind the N- and C-crosslink strain responses (Figure 3.3b) (i.e., crosslink strain increased before crosslink strain energy). The reason for this observation was most likely due to the method by which crosslink strain was calculated. Specifically, crosslink strain was based upon the change in distance between the $C\alpha$ atoms of the crosslinked amino acids; this measure of crosslink strain incorporates bond strain, angle strain, bond torsion angle strain, as well as non-covalent strains. Thus, it is possible initially for the distance between the $C\alpha$ atoms of the crosslinks to increase without observing an immediate increase in strain energy (e.g., as the bonds within the crosslinks undergo initial rearrangements in response to the applied strains on their parent collagen type I molecules).

Qualitatively, the responses of the N- and C-crosslinks suggest that the C-crosslink responded in a manner that was offset or delayed relative to the N-crosslink. This offset for

Chapter 3: Specific Aim 1

crosslink energy storage was calculated as $10.97 \pm 3.32\%$ overall strain (Figure 3.3c); a similar offset was found for crosslink strain (Figure 3.3b). This means that the N-crosslink responded at lower overall strains and the C-crosslink at higher overall strains – essentially, a two-stage, recruitment-like response. Given a larger collagenous structure, such as a collagen fibril, these results suggest that N-crosslinks may be recruited to aid with initial load bearing (i.e., first responders); then at higher strains, the C-crosslinks can be recruited for additional load bearing (i.e., back-up responders). We propose that the two-stage response predicted here is attributable to the conformations of the N- and C-telopeptides and the position at which the C-crosslink attaches to the C-telopeptide. The C-crosslink is reported to attach after the C-telopeptide hooks around; in our investigation, this positioning allowed the C-telopeptide to undergo a prolonged unfolding (Figure 3.3c), permitting the delayed or offset response in the C-crosslink.

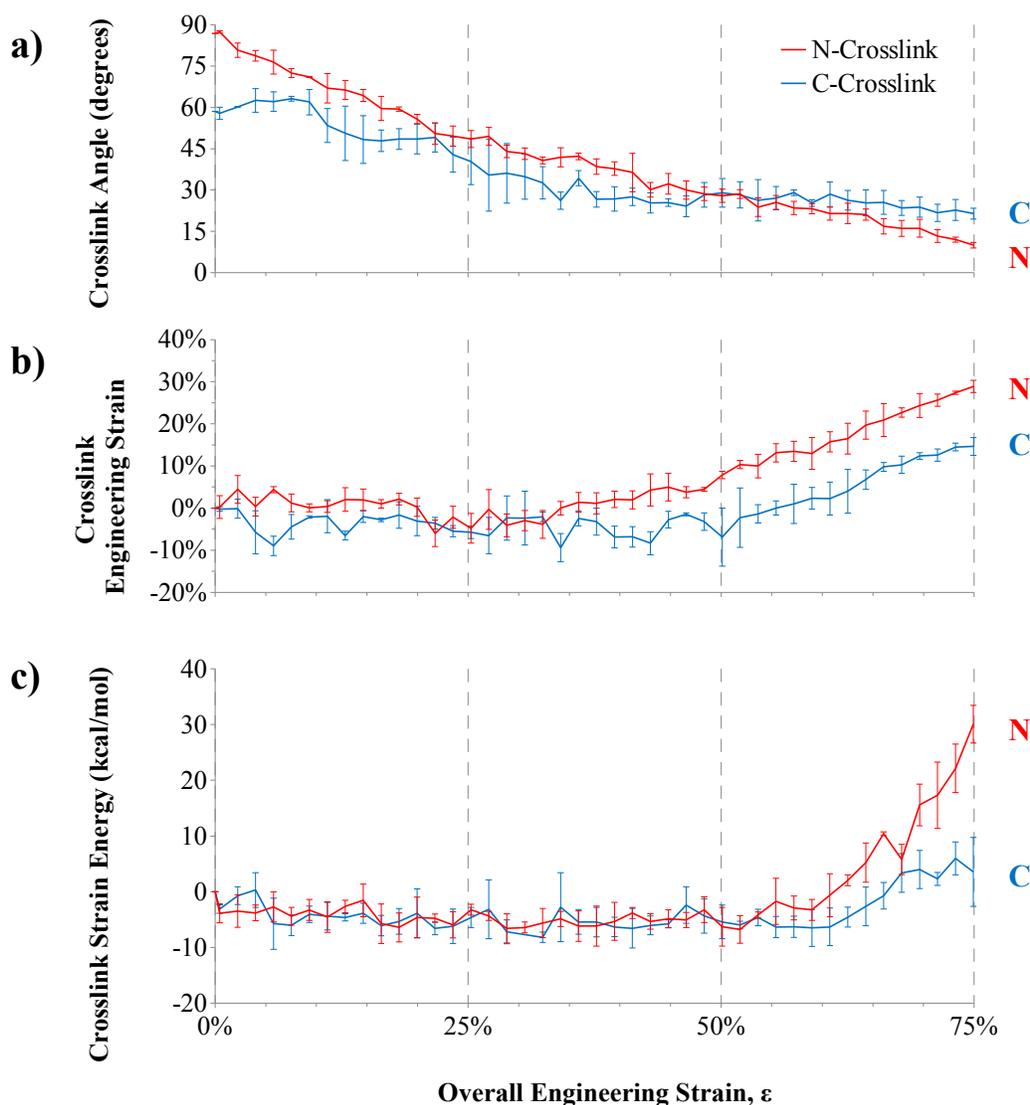


Figure 3.3: N- and C-crosslink angles (a), engineering strains (b), and strain energy (c) at 1.5625 m/s averaged from both pulling directions. Each data point represents mean \pm s.d (n = 3). (a), (b), and (c) all have the same x-axis of overall engineering strain shown in (c). Crosslink angles and crosslink strains were based on the alpha carbon atoms of the crosslinked amino acids, and crosslink angles were calculated relative to the pulling direction.

3.3.4. Extension of N- and C-crosslink recruitment to larger size scales

The extension of this two-stage response predicted here for N- and C-crosslinks to greater length scales, such as collagen fibrils, is dependent upon factors such as the site-specificity of crosslink formation and N/C directionality at the molecular and fibrillar levels. We anticipate that this two-stage response would translate to other immature and mature enzyme-derived crosslinks that form specifically at the telopeptides, but not to non-enzymatic crosslinks that form more haphazardly outside of enzyme and telopeptide involvement. N/C directionality is typically maintained due to a multitude of intracellular and extracellular processes that drive and regulate synthesis and folding of the collagen type I α -chains (Khoshnoodi et al., 2006), the release, modification, and assembly of molecules into fibrils (Kadler, Holmes, Trotter, & Chapman, 1996; Leikin, Rau, & Parsegian, 1995), and the growth of these fibrils within tissues via fibril-to-fibril fusion (Kadler et al., 1996) (see Figure 3.4 for a schematic of two different types of collagen fibrils that may form through such fibril-to-fibril fusion events). This is all to say that the ~23-nm-long region that was modeled here is expected to be a prevalent region of interest within a collagen fibril that often spans hundreds of micrometers in length. Therefore, we propose that the two-stage, recruitment response of N- and C-crosslinks, as predicted here, could be accounted for in the development of larger-scale models for fibrils and ultimately larger structures such as native collagenous tissues, engineered tissues, and collagen-based materials.

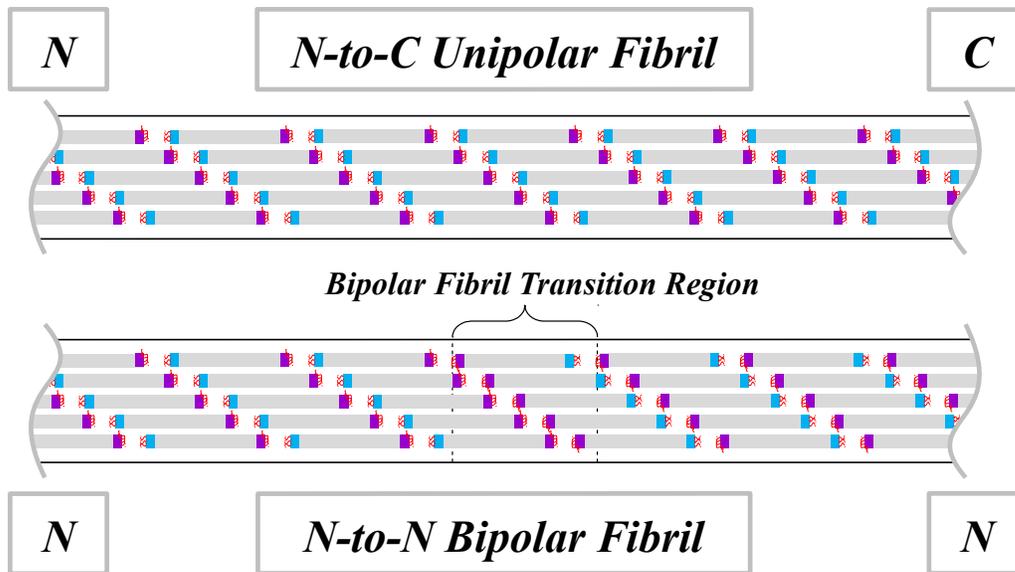


Figure 3.4: A schematic of unipolar and bipolar fibrils with their internal arrangement of collagen type I molecules. “N” and “C” represent the amino/N-terminal and carboxy/C-terminal directions. The wavy lines at the left and right represent truncations of the fibril lengths. The number of molecules depicted in these cross-sections is less than actuality, to simplify the schematic. The color scheme is identical to that described for [Figure 3.2](#). The transition region represents a region of C-C overlap instead of N-C overlap.

4.1. Introduction

The collagen microfibril represents an assembly of fibrillar collagen molecules and a stage of collagen fibrillogenesis. Due to the D-periodicity and semi-crystalline nature of fibrillar collagens, the microfibril may be reduced to a unit cell with a length equivalent to the D-period. This D-period has been reported to be in the range of 64-67 nm depending on the tissue and species (Birk & Bruckner, 2005). Such a unit cell has been characterized experimentally through x-ray diffraction (Fraser et al., 1987, 1983, 1983; Wess et al., 1998).

For collagen proteins, molecular modeling has been useful in the investigation of the molecular and microfibrillar structures of collagens and how these structures relate to the functional properties of collagens. For example, King *et al.* developed a molecular model of a collagen type I microfibril section and were able to show that the charge associated with amino acid side chains, in their collagen microfibril model, correlated well with the experimental banding pattern in positively stained collagen fibrils viewed with transmission electron microscopy (TEM) (G. King, Brown, & Chen, 1996). Brown later employed the microfibril model presented by King *et al.* to identify sites of synthetic collagen crosslinking, with a particular interest in tanning agents used in the manufacture of leather from animal skins (Brown, 2009). Kwansa and Freeman assembled a model of an uncrosslinked collagen type I microfibril unit in representation of dehydrated (*in vacuo*) and hydrated states (explicit TIP3P water and increased inter-molecular spacing) (Kwansa & Freeman, 2010); these microfibril unit models were used to improve results of a collagen fiber model by Freeman and Silver (Freeman & Silver, 2004). Most recently, Gautieri *et al.* constructed molecular models of an uncrosslinked collagen type I microfibril section both in dry and hydrated states. With their model, they investigated mechanical properties in response to tensile loading. Hydration was shown to reduce the Young's modulus (a measure of material stiffness) of their microfibril models from 1.8 to 2.25 GPa for their dehydrated microfibril section to 0.3 to 1.2 GPa for their hydrated microfibril section (Gautieri et al., 2011).

For our work in Aims 1 and 2, we were interested in assembling molecular models of a hydrated collagen type I microfibril, but with the incorporation of different lysine/hydroxylysine-

Chapter 4: Specific Aim 2

derived collagen crosslinks. Aim 1 involved a localized region within the microfibril where crosslinking occurs between two collagen molecules, while Aim 2 involved a section of a microfibril in which five molecules reside.

4.2. Methods

4.2.1. Molecular modeling programs and parameters

Similar to Aim 1, these microfibril unit molecular models were constructed using Discovery Studio (Accelrys Software, Inc., 2011) and Visual Molecular Dynamics (VMD) (Humphrey et al., 1996). Simulations were carried out using the NAMD code (Phillips et al., 2005), and the Tachyon ray tracing code was used with VMD to generate molecular renderings (Stone, 1998). However, for Aim 2, the CHARMM all-atom force field was used (Accelrys Software, Inc., 2011) instead of the CHARMM19 united-atom force field that was used for Aim 1. The CHARMM all-atom force field is based upon the CHARMM22 force field (see Equation 4.1).

Equation 4.1: Total potential energy in the CHARMM22 all-atom force field expressed as a summation of bond stretching (bonds), angle bending (angle), a bond angle correctional term called Urey-Bradley (UB), bond rotation (dihedrals), out-of-plane bending (impropers), and collectively, Lennard-Jones and Coulombic interactions (nonbond). K_b , K_θ , K_{UB} , K_χ , and K_{imp} represent force constants. b , θ , S , χ , and φ represent variables (bond length, UB length, bond angle, torsion angle, and out-of-plane bending angle, respectively). b_0 , θ_0 , S_0 , and φ_0 represent fixed equilibrium values (bond length, bond angle, UB length, and out-of-plane bending angle, respectively). n is a measure of bond torsion periodicity, and δ represents a bond torsion phase angle. Starting from the case of two identical atom types interacting non-covalently, ϵ is the minimum of their Lennard-Jones energy and R_{min} is $1/2$ of their center-to-center inter-atomic distance at ϵ . q_i and q_j represent fixed partial atomic charges for atoms i and j , while ϵ_1 represents the dielectric constant. (MacKerell et al., 1998)

$$\begin{aligned} U(R) = & \sum_{Bonds} K_b (b - b_0)^2 + \sum_{Angle} K_\theta (\theta - \theta_0)^2 + \sum_{UB} K_{UB} (S - S_0)^2 \\ & + \sum_{Dihedrals} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{Impropers} K_{imp} (\varphi - \varphi_0)^2 \\ & + \sum_{Nonbond} \epsilon \left[\frac{R_{min_{ij}}^{12}}{r_{ij}^{12}} - \frac{R_{min_{ij}}^6}{r_{ij}^6} \right] + \frac{q_i q_j}{\epsilon_1 r_{ij}} \end{aligned}$$

4.2.2. Building of the microfibril molecular models

The model building procedure for Aim 2 was similar to that used for Aim 1. The triple-helical structure in our molecular model was based upon a 29-amino acid long collagen-like (proline-proline-glycine)_n crystal structure (1k6f.pdb) (Berisio et al., 2002). Using Discovery Studio, crystallographic water molecules were removed, and the structure was replicated and spliced together, similar to methods described by Chen *et al.* (J. M. Chen et al., 1991), to form a single collagen-like molecule containing 1054 amino acids per chain. This collagen-like molecule was then divided into five molecular segments by breaking the appropriate peptide bonds, after which the cleaved ends were adjusted into positively charged amino groups (-NH₃⁺) or negatively charged carboxylic acid groups (-C(=O)-O⁻) as appropriate. These five segments were then aligned at their N-termini, and they were translated into specific x-y coordinates (see Appendix B.1), in order to represent a quasi-hexagonal compressed microfibril with a length equivalent to an entire collagen type I D-period (Fraser et al., 1983). Then using VMD, molecular kinks were applied to the segments within the gap region based in-part on a concept depicted by Fraser *et al.* (Fraser et al., 1987), in order to establish a microfibrillar twist (Orgel et al., 2001; Wess et al., 1998). Angles were calculated for each of the four longer segments so that the segments could be bent slightly at the start of the gap region and end at the start of the next segment (i.e., if the N-termini of the segments from a second microfibril unit were connected to the C-termini of the first microfibril unit) (see Appendix B.2 for the TCL script that was used to apply these molecular kinks). At this point, the molecular model contained three chains of identical length (1054 amino acids); however, collagen type I contains two identical chains (α 1 chains) and a third chain (α 2 chain) that is different in both length and sequence (e.g., 1021 amino acids for bovine collagen type I) (The UniProt Consortium, 2012). Therefore, one of the three chains (selected to be the α 2 chain) was trimmed from 1054 to 1021 amino acids leading to an α 1- α 2- α 1 collagen-like microfibril unit. Thus, the model contained an N-telopeptide domain at the beginning (16 amino acids for α 1 and 6 for α 2), a triple-helical domain (1014 for α 1 and α 2), and a C-telopeptide domain at the end (24 for α 1 and 1 for α 2). The trimming of the α 2 chain was performed from both the N-terminus and the C-terminus; the α 2 N-telopeptide was reduced from 16 to 6 amino acids, and the α 2 C-telopeptide was reduced from 24 to 1 amino acid. The conformations of the N- and C-telopeptides were then adjusted based upon a rat tail tendon collagen type I x-ray fiber diffraction structure (3hr2.pdb) (Orgel et al., 2006), allowing

Chapter 4: Specific Aim 2

for favorable alignments between four potential crosslinking sites (Eyre & Wu, 2005) (specific sites are shown in Table 4.1). These conformational adjustments were made by changing the positions (i.e., Cartesian coordinates) of the backbone atoms to match those of Orgel *et al.*'s 3hr2.pdb structure; energy minimization was then conducted for 10,000 timesteps to allow the side chain atoms to adjust to the changes to the positions of the backbone atoms. Bovine (*Bos taurus*) collagen type I amino acid sequences for the 1054-residue-long $\alpha 1$ chains and the 1021-residue-long $\alpha 2$ chain were then used to replace the collagen-like Pro-Pro-Gly sequences (Fietzek & Kuhn, 1976; Orgel et al., 2006; The UniProt Consortium, 2012) (see Appendix B.3). For the UniProt Consortium reference cited above, the following UniProt accession numbers were used: [P02453](#) (for the collagen type I $\alpha 1$ chains) and [P02465](#) (for the collagen type I $\alpha 2$ chain).

Table 4.1: Positions of crosslink precursor amino acids for each crosslink type. Numbers represent amino acid positions relative to the most N-terminal amino acid in each domain (i.e., N-telopeptide (N), helical (helix), and C-telopeptide (C) domains). The specific amino acid chain for each crosslink precursor is denoted by A/ $\alpha 1$, B/ $\alpha 2$, or C/ $\alpha 1$. *Note: HHL has not been experimentally detected at the N-terminal position, but it was included for consistency with our modeling approach. It should also be noted that the bovine collagen type I amino acid sequences that we employed allowed for an HHL at the N-terminal position without any sequence modification.

Crosslink	N-crosslink precursors	C-crosslink precursors	Ref.
deH-HLNL	Lys-aldehyde 9 ^N (C/ $\alpha 1$) Hyl 930 ^{helix} (C/ $\alpha 1$)	Hyl 87 ^{helix} (A/ $\alpha 1$) Lys-aldehyde 16 ^C (A/ $\alpha 1$)	(Eyre, Paz, & Gallop, 1984)
HLKNL	Hyl-aldehyde 9 ^N (C/ $\alpha 1$) Hydroxylysine 930 ^{helix} (C/ $\alpha 1$)	Hyl 87 ^{helix} (A/ $\alpha 1$) Hyl-aldehyde 16 ^C (A/ $\alpha 1$)	(Eyre, Paz, & Gallop, 1984)
PYD	Hyl-aldehyde 9 ^N (C/ $\alpha 1$) Hyl 930 ^{helix} (C/ $\alpha 1$) Hyl-aldehyde 9 ^N (B/ $\alpha 2$)	Hyl 87 ^{helix} (A/ $\alpha 1$) Hyl-aldehyde 16 ^C (A/ $\alpha 1$) Hyl-aldehyde 16 ^C (C/ $\alpha 1$)	(Hanson & Eyre, 1996)
*HHL	Lys-aldehyde 9 ^N (C/ $\alpha 1$) Hyl 930 ^{helix} (C/ $\alpha 1$) His 932 ^{helix} (A/ $\alpha 1$)	Hyl 87 ^{helix} (A/ $\alpha 1$) Lys-aldehyde 16 ^C (A/ $\alpha 1$) His 92 ^{helix} (B/ $\alpha 2$)	(Mechanic, Katz, Henmi, Noyes, & Yamauchi, 1987; Yamauchi, London, Guenat, Hashimoto, & Mechanic, 1987)

4.2.3. Energy minimization, heating, and equilibration

The energy minimization, heating, and equilibration steps were nearly identical to those described in section 3.2.3 for Aim 1; however, the durations of certain steps were extended due to the greater size of these microfibril unit molecular models (~38,900 atoms) compared to those molecular models used in Aim 1 (~5,200 atoms). Briefly, energy minimization steps (method:

Chapter 4: Specific Aim 2

conjugate gradients; NAMD gradient tolerance: 0.01 kcal/mol/Å), heating steps (method: temperature reassignment; 0 to 293.15 K over 10 ps), and equilibration steps (method: Langevin dynamics; 293.15 K) were conducted. The generalized Born implicit solvent approach of Onufriev, Bashford, and Case (GB^{OBC} II) was used (Onufriev et al., 2000) (solvent dielectric = 80.0; ion concentration = 0.150 mol/L; GBIS alpha cutoff = 14 Å) (Bhandarkar et al., 2011). Switching distances and cut-off distances were used to reduce the number of calculations required for pair-wise non-bonded interactions (switching distance = 15 Å, cut-off distance = 16 Å; pair list distance = 17.5 Å). Table 4.2 summarizes the details for these steps for Aim 2.

Table 4.2: Preparation steps used for our ~65-nm-long molecular models. Here, timesteps refers to the Δt parameter in Table 2.5. The spring constant used for the harmonic atom restraints was $k = 0.5$ kcal/mol/Å².

No.	Stage	Timestep	Duration	Atom constraints	Description
Only the uncrosslinked model					
Applied telopeptide conformations					
1	Energy minimization	1 fs	10,000 timesteps	Fix non-telopeptide atoms	Conjugate gradients $T = 0$ K
2	Energy minimization	1 fs	500,000 timesteps	Fix backbone	Conjugate gradients $T = 0$ K
3	Energy minimization	1 fs	100,000 timesteps	Restrain all C α	Conjugate gradients $T = 0$ K
4	Heating	1 fs	10,000 timesteps (10 ps)	Restrain all C α	$T = 0$ to 293.15 K
5	Equilibration	1 fs	500,000 timesteps (500 ps)	Restrain all C α	$T = 293.15$ K
6	a. Duplicated the coordinate file (.pdb) of the uncrosslinked model four times b. Renamed the four new pdb files for the four crosslinked models c. Added two identical crosslinks to each of the four new pdb files, and generated corresponding structure files (.psf) (deH-HLNL, HLKLN, HHL, & PYD)				
Only the four crosslinked models					
7	Energy minimization	1 fs	10,000 timesteps	Fix backbone	Conjugate gradients $T = 0$ K
8	Energy minimization	1 fs	100,000 timesteps	Restrain all C α	Conjugate gradients $T = 0$ K
9	Heating	1 fs	10,000 timesteps (10 ps)	Restrain all C α	$T = 0$ to 293.15 K
10	Equilibration	1 fs	500,000 timesteps (500 ps)	Restrain all C α	$T = 293.15$ K
All five models					
11	Equilibration	1 fs	500,000 timesteps (500 ps)	Restrain C α of N- and C-termini	$T = 293.15$ K
12	Equilibration	2 fs	5 million timesteps (10 ns)	Restrain C α of N- and C-termini	$T = 293.15$ K

Chapter 4: Specific Aim 2

The root-mean-square displacement (RMSD) of the backbone atoms was analyzed in several ways in order to determine the appropriate equilibration duration based on the work of Walton and Van Vliet (Walton & Van Vliet, 2006). For example, in one of these approaches, the RMSD vs. time data was fitted with a power function of the form $\text{RMSD}(t) = A \cdot t^B$, for each of the five molecular models. Then the fitted parameters A and B were averaged at each time value. Statistical analysis of these fitted parameters suggested that there was no significant change in these two parameters after 8 nanoseconds.

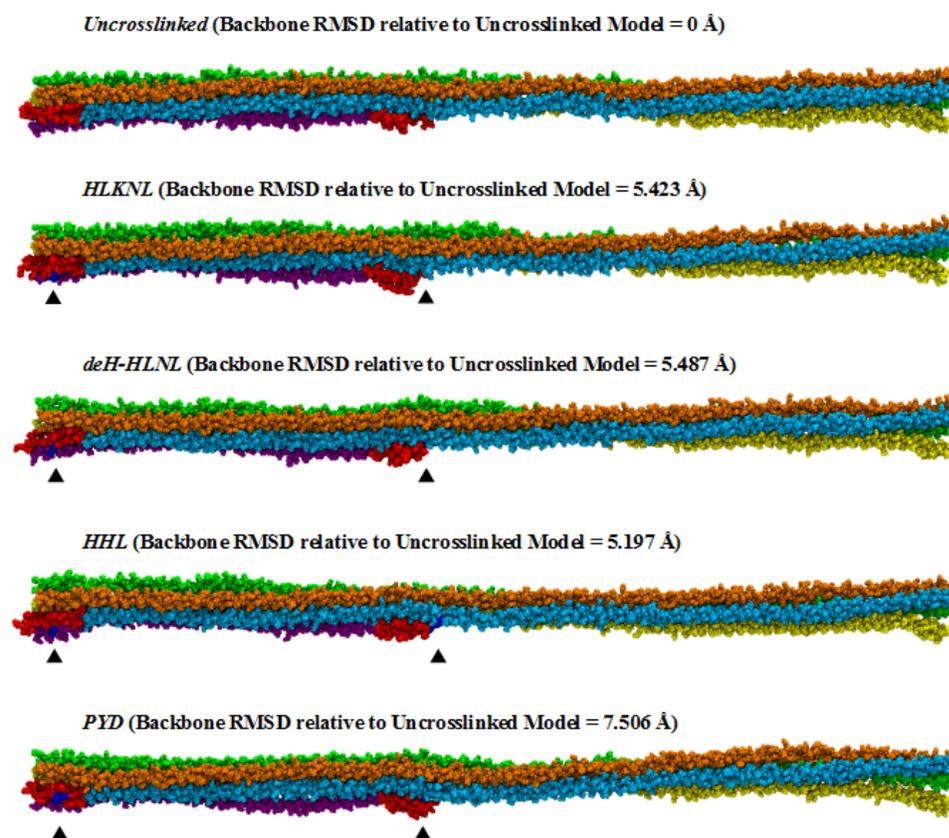


Figure 4.1: Snapshots of our five microfibril unit molecular models after equilibration. The atoms are colored as follows: five molecular segments within each microfibril unit cell (1-cyan, 2-orange, 3-green, 4-yellow, 5-purple), three N-telopeptides and three C-telopeptides (red), and two crosslinks (blue; see arrows). The root-mean-square displacement of the backbone atoms of the four crosslinked models relative to the uncrosslinked model are included.

4.2.4. Constant velocity pulling

In order to conduct constant velocity pulling simulations with the microfibril unit molecular models, C-terminal $C\alpha$ atoms from four segments were fixed (a total of 12 fixed $C\alpha$ atoms) and N-terminal $C\alpha$ atoms from four segments were pulled at a constant velocity using

Chapter 4: Specific Aim 2

moving constraints (a total of 12 pulled C α atoms). Fixed atoms were selected by placing a “1” into the atom’s beta column within the pdb coordinate file, while pulled atoms were selected by placing the moving constraints spring constant into the atom’s occupancy column within the pdb coordinate file. A spring constant of “9999.00” kcal/mol/Å² was used for all pulled atoms, based on work described in section 3.3 (see Appendix B.4 to view the TCL script that was used to select and assign these fixed and pulled atoms). The purpose of the fixed and pulled atoms was to mimic the continuity of each of the collagen molecules. Thus, segment 1 (cyan) was not pulled since the molecule of which it is a part has no continuity into the next microfibril unit to the left. Similarly, segment 5 (purple) was not fixed since the molecule of which it is a part has no continuity into the next microfibril unit to the right. These constant velocity pulling simulations were performed at multiple pulling velocities (see Table 4.3 for a listing of the velocities and Appendix B.5 for a representative NAMD configuration file used for these constant velocity pulling simulations).

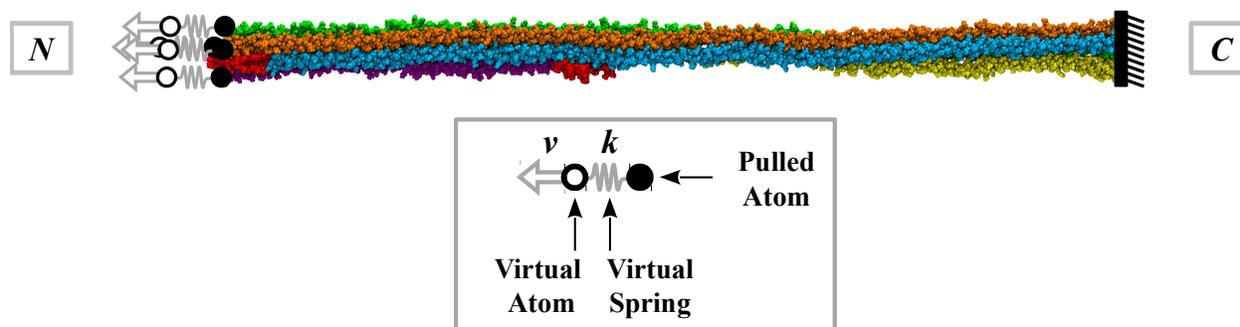


Figure 4.2: A representative schematic of the constant velocity pulling simulation. The color scheme is identical to Figure 3.2. In this figure, 12 C-terminal C α atoms are fixed (thick line with hash lines) and 12 N-terminal C α atoms are being pulled in the N-terminal direction. The boxed inset shows a schematic of the moving constraints technique, where each pulled C α atom is connected through a virtual spring of stiffness, k , to a virtual atom that moves at a constant velocity, v . The number of pulled atom-virtual spring-virtual atom elements with the microfibril unit model has been simplified here for visual clarity; there were actually 12 of these elements – one element for each pulled C α atom.

Table 4.3: Pulling velocities employed for the microfibril unit molecular models.

Velocity (m/s)	Velocity vector (Å/timestep)	Number of timesteps	Displacement (Å)
100.00	0 0 0.00200	234,500	469
50.00	0 0 0.00100	469,000	469
25.00	0 0 0.00050	938,000	469
12.50	0 0 0.00025	1,876,000	469
6.25	0 0 0.000125	3,752,000	469

4.2.5. Overall strain energy, crosslink strain energy, and crosslink alignment

Overall strain energy was calculated as the change in total energy for all atoms in each model relative to its initial unstrained structure (i.e., the structure obtained at the end of the equilibration steps described in section 4.2.3). This total energy was defined as the sum of total potential energy and kinetic energy, with total potential energy being identical to that defined in Equation 4.1. These total energy values were obtained directly from the NAMD log files (see Appendix B.7 for the TCL script that was used to extract these energy values from the NAMD log files). Total potential energy alone could have been used without kinetic energy; it was found that kinetic energy did not change much over the course of these simulations. Thus, the change in total energy was nearly identical to the change in total potential energy. However, for clarity and for consistency with the crosslink strain energy calculations (described next), total potential energy should have been used instead of total energy. This inconsistency will be resolved in our manuscript describing this work.

Crosslink strain energy was calculated as the change in total potential energy as calculated by the NAMD code using the same CHARMM parameter file that had been used to conduct the simulations. This total potential energy was calculated for atoms in between the C α atoms of the crosslinks (i.e., excluding other backbone atoms of the crosslinked amino acid residues) (see Appendix B.8 for the TCL script that was used to calculate these crosslink strain energy values).

Crosslink alignment was calculated based on the two C α atoms (immature crosslinks) or three C α atoms (mature crosslink) from the crosslinked amino acids. The Cartesian coordinates for these C α atoms were obtained using VMD (see Appendix B.9 for the TCL script that was used to obtain these Cartesian coordinates).

Crosslink orientation angles were calculated as follows:

$$\theta = 90 - \tan^{-1} \left[\frac{|Z_2 - Z_1|}{\sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}} \right]$$

Below is a representative Microsoft Excel formula that was used:

=90-DEGREES(ATAN(ABS(Z2-Z1)/SQRT((X2-X1)^2+(Y2-Y1)^2)))

Chapter 4: Specific Aim 2

For the case of the immature crosslinks, deH-HLNL and HLKLN (two C α atoms), this calculation was simply done based on the coordinates of the C α atoms from the two crosslinked amino acids. However, for the case of the mature crosslinks, HHL and PYD (three C α atoms), this calculation was done twice and the average of the two calculated angles was taken. For HHL, one angle was calculated using the C α atom from the telopeptide and one of the C α 's from the triple helix, and a second angle was calculated using the same C α from the telopeptide and the other Ca from the triple helix C α . For PYD, one angle was calculated using the C α atom from the triple helix and one of the C α 's from the telopeptide, and a second angle was calculated using the same C α from the triple helix and the other C α from the telopeptide (see Figure 2.3 for the chemical structures of these immature and mature crosslinks).

4.2.6. Cross-sectional areas

Cross-sectional area measurements were made using a custom code called PDBAN, which was written by the laboratory of Prof. Sergei Sukharev at the University of Maryland (Chiang, Anishkin, & Sukharev, 2004); this custom code was developed using MATLAB® (The MathWorks, Inc., 1984). The “volumetric data creation” option was selected from the PDBAN mainboard window; this tool was used to first create a volumetric data file (.mat). The following settings were used for volumetric data creation: “greed step length or steps #” (1 angstrom; length, A), “smoothing” (no smoothing), “probe radius” (1.4; surface of probe), checked “autoselect matrix size”, selected “VDW radii in the separate file”, “VDW radius file type” (*.txt column), and “VDW radii data type” (data per atom type). The van der Waals (VDW) radii for the atoms of the molecular model were based upon the same CHARMM all-atom force field that was used to carry out the molecular dynamics simulations described in section 4.2. Specifically, the .pdb files and the Lennard-Jones parameter section from Accelrys’ CHARMM parameter file (Accelrys Software, Inc., 2011) were copied into Microsoft Excel, and Excel’s LOOKUP function was used to assign each atom from the .pdb file with a van der Waals radius (the Lennard-Jones R_{min} value). Then the column containing the van der Waals radii was copied into a plain text file for use with the PDBAN code. Subsequently, the “cross-section area calculation by volumetric/PDB/VMRL data” option was selected from the PDBAN mainboard. The following settings were selected for the cross-sectional area calculations: checked “save crosssections’ areas”, “crosssections normal to” (Z-axis; 0 0 1), “sectors” (100), “cros. greed” (length, A; 1), and

“calc. method” (by volumetric density crosssection (*.mat)) (see Figure 4.3 and Figure 4.4 for the resulting cross-sectional areas).

4.2.7. Initial lengths, extensions, and engineering strains

Initial lengths were determined based on the root-mean-square (RMS) difference between the 12 fixed C α atoms and the 12 pulled C α atoms at time zero ($t = 0$). Specifically, the position vector for the i^{th} fixed C α atom was compared to that of the i^{th} pulled C α atom (Theoretical and Computational Biophysics Group, 2012) (see Appendix B.6 for the TCL script that was used and Figure 4.4 for the initial lengths). Extensions (changes in length) were calculated based on the RMS displacement of the 12 pulled C α atoms; this RMS displacement calculation was similar to the RMS difference calculation, except that the position vector of the i^{th} pulled C α atom at a given time (t) was compared to that of the same pulled C α atom at time zero ($t = 0$). This RMSD displacement data was acquired using VMD’s RMSD Trajectory Tool (Humphrey et al., 1996). Engineering strains were then calculated as the extension of each molecular model divided by its initial length.

Equation 4.2: RMS difference equation for the calculation of initial lengths.

$$RMS\ difference = \sqrt{\frac{\sum_{i=1}^{12} (\vec{r}_{pulled,i}(0) - \vec{r}_{fixed,i}(0))^2}{12}}$$

where i = atom pair number, n = total number of atom pairs (12 pairs in this case), $\vec{r}_{pulled,i}(0)$ = initial position vector for pulled atom i , and $\vec{r}_{fixed,i}(0)$ = initial position vector for fixed atom i .

Equation 4.3: RMS displacement equation for the calculation of changes in length.

$$RMS\ displacement = \sqrt{\frac{\sum_{i=1}^{12} (\vec{r}_i(t) - \vec{r}_i(0))^2}{12}}$$

where i = atom number, n = total number of atoms (12 in this case), $\vec{r}_i(t)$ = current position vector for atom i at time t , and $\vec{r}_i(0)$ = initial position vector for atom i .

4.2.8. Forces, nominal stresses, and elastic moduli

Forces and nominal stresses were derived from strain energy vs. extension data (Figure 4.5), which were each fitted with a 6th order polynomial function (Equation 4.4). These strain energy functions were then used to derive nominal stress functions by taking the first derivative of strain energy with respect to extension and dividing by the initial cross-sectional area (Equation 4.5). The 6th order polynomial strain energy functions were constructed such that the following two conditions were satisfied: when extension is zero, both strain energy and nominal stress are zero ($W(0) = 0$ and $\sigma(0) = 0$, respectively). In order to satisfy these two conditions, the coefficients C_0 and C_1 were set to zero. The remaining five coefficients C_2 to C_6 were calculated in Microsoft Excel using Excel's LINEST and INDEX functions (see Equation 4.4).

Due to the low- and high-sloping regions of the nominal stress data (Figure 4.6), two elastic moduli were calculated, namely, a low-strain modulus and a high-strain modulus. The distinction between the low-strain and high-strain moduli was made based on a determination of a transition point, which was defined as the extension value at which the second derivative of the nominal stress function was maximal (i.e., where the third derivative of the nominal stress function was zero) (Equation 4.6).

Equation 4.4: Strain energy expressed as a 6th order polynomial function (boxed equation). Here, strain energy (W) [kcal/mol] is expressed as a function of extension (ΔL) and five fitting parameters (C_2 to C_6). Two of the coefficients (C_0 and C_1) were set to zero to satisfy the following two conditions: when extension is zero, both strain energy and nominal stress are zero ($W(0) = 0$ and $\sigma(0) = 0$, respectively).

Begin with a 6th order polynomial function:

$$W(\Delta L) = C_0 + C_1 * \Delta L + C_2 * \Delta L^2 + C_3 * \Delta L^3 + C_4 * \Delta L^4 + C_5 * \Delta L^5 + C_6 * \Delta L^6$$

Let $C_0 = 0$ and $C_1 = 0$ so that $W(0) = 0$ and $\sigma(0) = 0$, respectively:

$$W(\Delta L) = C_2 * \Delta L^2 + C_3 * \Delta L^3 + C_4 * \Delta L^4 + C_5 * \Delta L^5 + C_6 * \Delta L^6$$

Below is a representative Microsoft Excel formula that was used to find C_2 to C_6 :

$$=INDEX(LINEST(A\$4:A\$74,\$B\$4:\$F\$74,FALSE,TRUE),1,1)$$

The first range references a 1 x 71-cell column with the strain energy values, while the second range represents a 5 x 71-cell matrix with ΔL^n values (ΔL^2 , ΔL^3 , ΔL^4 , ΔL^5 , and ΔL^6 , respectively) where ΔL denotes the extension values.

Chapter 4: Specific Aim 2

Equation 4.5: Nominal stress expressed as a 5th order polynomial function (boxed equation). Here, nominal stress (σ) represents the force (F) divided by the initial cross-sectional area of the microfibril unit (A_0). Nominal stress is expressed as a function of extension (ΔL) and the same five fitting parameters (C_2 to C_6).

$$\sigma(\Delta L) = F * \frac{1}{A_0} = \frac{dW}{d\Delta L} * \frac{1}{A_0} \rightarrow \frac{dW}{d\Delta L} * \frac{1}{A_0} * \left(\frac{CF}{N_A} \right) = \frac{dW}{d\Delta L} * \left(\frac{CF}{A_0 N_A} \right)$$

Given the units of W [kcal/mol], ΔL [Å], and A_0 [Å²], the derived nominal stress was multiplied by a conversion factor (CF) and divided by Avogadro's Number (N_A) [mol⁻¹] to obtain nominal stress in units of GPa.

Derivation of conversion factor (CF):

$$1 \text{ kcal} = 4184 \text{ Joules} = 4184 \text{ N}\cdot\text{m} = 4.184 * 10^{13} \text{ N}\cdot\text{Å}$$

$$1 \text{ GPa} = 10^9 \text{ Pa} = 10^9 \text{ N/m}^2 = 10^{-11} \text{ N/Å}^2$$

$$\frac{dW}{d\Delta L} * \left(\frac{1}{A_0 N_A} \right) \rightarrow \left[\frac{\text{kcal/mol}}{\text{Å}} \right] * \left[\frac{1}{\text{Å}^2 \text{mol}^{-1}} \right] = \left[\frac{\text{kcal}}{\text{Å}^3} \right] = \left[\frac{4.184 * 10^{13} \text{ N}\cdot\text{Å}}{\text{Å}^3} \right]$$

$$= \left[4.184 * 10^{13} \frac{\text{N}}{\text{Å}^2} \right] * \left[\frac{1 \text{ GPa}}{10^{-11} \text{ N/Å}^2} \right] = \left[4.184 * 10^{24} \text{ GPa} \right] = \left[1 \frac{\text{kcal}}{\text{Å}^3} \right]$$

Therefore, $CF = 4.184 * 10^{24} \text{ GPa}/(\text{kcal}/\text{Å}^3)$.

$$\sigma(\Delta L) = (2 C_2 * \Delta L + 3 C_3 * \Delta L^2 + 4 C_4 * \Delta L^3 + 5 C_5 * \Delta L^4 + 6 C_6 * \Delta L^5) * \left(\frac{CF}{A_0 N_A} \right)$$

where engineering strain, $\varepsilon = \Delta L/L_0 \rightarrow \Delta L = \varepsilon * L_0$

A_0 = initial cross-sectional area of the microfibril unit [Å²]

$N_A = 6.02214199(47) \times 10^{23} \text{ mol}^{-1}$ (Fenna, 2002)

Equation 4.6: Low-to-high-strain modulus transition point (boxed equation). The quadratic formula was used to find the transition point associated with the low- and high-strain moduli, namely, the point at which the first derivative of the nominal stress function (elastic modulus) shows a maximum change in its curvature (i.e., where the second and third derivatives of the nominal stress function are maximum and zero, respectively).

First derivative (greatest change in curvature at transition point):

$$\frac{d\sigma}{d\Delta L} = (2 C_2 + 6 C_3 * \Delta L + 12 C_4 * \Delta L^2 + 20 C_5 * \Delta L^3 + 30 C_6 * \Delta L^4) * \left(\frac{CF}{A_0 N_A} \right)$$

Chapter 4: Specific Aim 2

Second derivative (maximal at transition point):

$$\frac{d^2 \sigma}{d \Delta L^2} = (6 C_3 + 24 C_4 * \Delta L + 60 C_5 * \Delta L^2 + 120 C_6 * \Delta L^3) * \left(\frac{CF}{A_0 N_A} \right)$$

Third derivative (zero at transition point):

$$\frac{d^3 \sigma}{d \Delta L^3} = (24 C_4 + 120 C_5 * \Delta L + 360 C_6 * \Delta L^2) * \left(\frac{CF}{A_0 N_A} \right)$$

$$0 = (24 C_4 + 120 C_5 * \Delta L + 360 C_6 * \Delta L^2)$$

$$\Delta L = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-120 C_5 \pm \sqrt{(120 C_5)^2 - 4 * (360 C_6) * (24 C_4)}}{2 * (360 C_6)}$$

4.2.9. Statistical analyses

Statistical analyses were used to compare the elastic moduli for the different microfibril unit models and different simulation cases (e.g., pulling in the N-direction vs. C-direction; fixing 12 Ca atoms vs. fixing only three Ca atoms). As with Aim 1, all statistical analyses were performed using the JMP statistical software (SAS Institute, Inc., 2010). The following is restated from Aim 1 (section 3.2.7). Unless otherwise stated, the significance level, alpha value or α , was set to 0.05. The Brown-Forsythe and Levene tests were used to initially assess if the variances were homogeneous. If so, a one-way analysis of variance (ANOVA) employing a pooled standard deviation was used, followed by a Tukey-Kramer Honestly Significant Difference (HSD) test, when appropriate. If not, then a Welch's one-way ANOVA and Student's t-tests with Bonferroni correction were used. For the Bonferroni correction, the alpha value was set to 0.05 divided by the number of t-tests (e.g., number of t-tests = $g*(g-1)/2$, where g denotes the number of groups).

4.2.10. Computing resources

The molecular dynamics simulations for these ~38,900-atom microfibril unit molecular models were conducted using three types of computing systems: (1) the Virginia Tech Engineering Science and Mechanics (ESM) Intel 64-bit Linux computing cluster (34 compute nodes, ~20 to 30 GB RAM per node, 8 or 16 processor cores per node, and 2.21 to 2.60 GHz per core), (2) the Virginia Tech Advanced Research Computing (ARC) "System X" computing cluster running Mac OS X (1,024 compute nodes, 4 GB RAM per node, 2 processor cores per

node, and 2.3 GHz per core) [www.arc.vt.edu/resources/hpc/systemx.php], and (3) the Virginia Tech ARC “Athena” computing cluster running CentOS Linux 5 (42 compute nodes, 64 GB RAM per node, 32 processor cores per node (quad-socket octa-core nodes), and 2.3 GHz per core) [www.arc.vt.edu/resources/hpc/athena.php].

4.3. Results and Discussion

4.3.1. Cross-sectional areas and initial lengths

Initial cross-sectional areas and initial lengths were determined for each of the five microfibril unit molecular models. It was observed that the cross-sectional areas varied along the lengths of the models, but that the cross-sectional areas were fairly similar amongst the five models (Figure 4.3). This position-dependent variation was primarily due to the gap region and overlap region; that is, in the overlap region, there are five molecular segments, while in the gap region, there are only four molecular segments. Furthermore, one can observe an increase in cross-sectional area around $z = -100 \text{ \AA}$ (Figure 4.3). This increase in cross-sectional area was most likely due to the presence of the C-terminal telopeptides of segment 5; this C-telopeptide domain has a somewhat globular shape due to the hair-pin/hooked conformations of the longer C-telopeptides of the $\alpha 1$ chains.

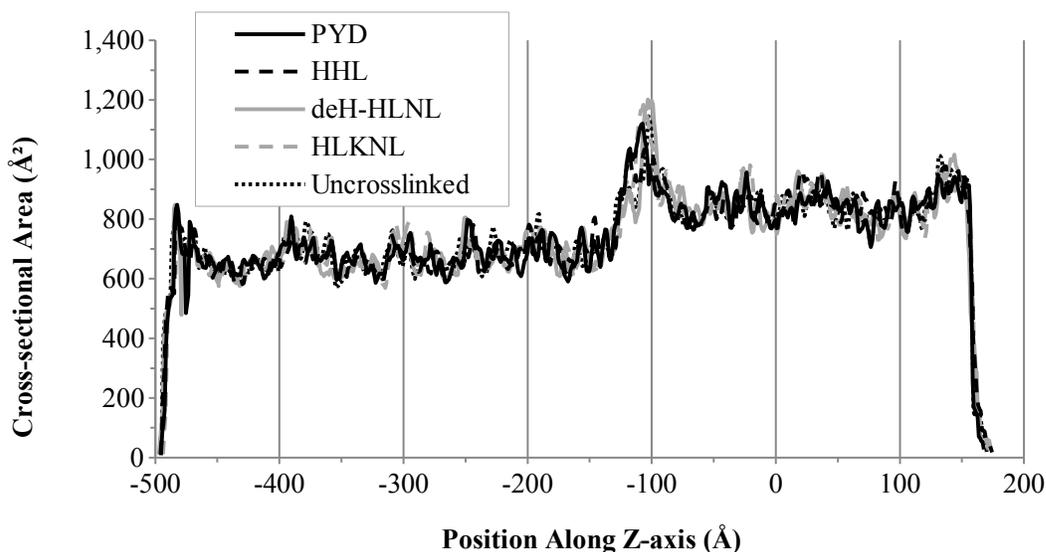


Figure 4.3: Cross-sectional areas for the microfibril unit models. The data are represented as curves so that the data can be seen more easily; the actual data points were acquired at a 1 \AA resolution. The “Position Along Z-axis (\AA)” values on the x-axis are arbitrary in this context. The models could have been translated along their z-axis (i.e., their longitudinal or length-wise axis) leading to different values along the x-axis here. The values here are based on the collagen-like x-ray structure that was used as a template.

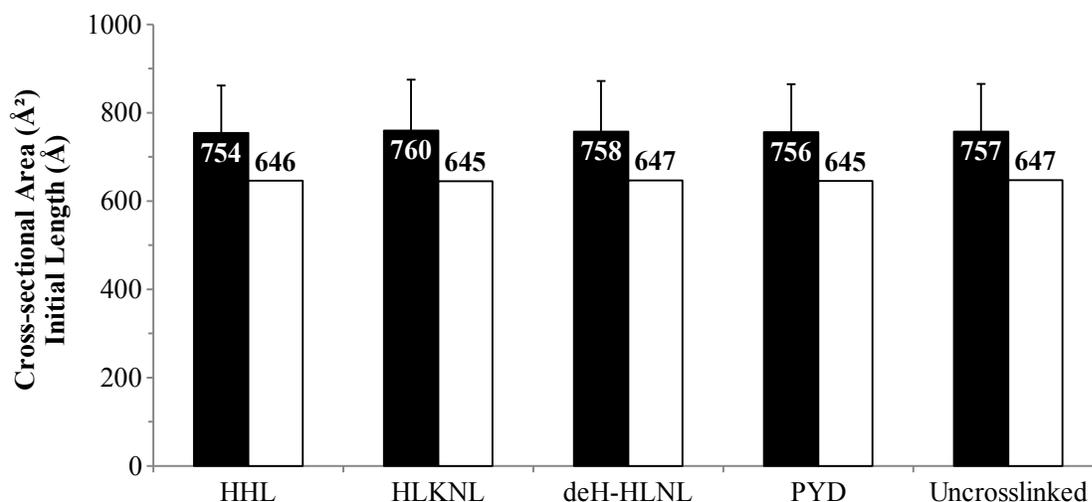


Figure 4.4: Initial cross-sectional areas and initial lengths for the microfibril unit models. The cross-sectional areas (black bars) are plotted as mean + s.d. with mean values, rounded to the nearest integer, shown above the plotted bars ($n = 632$ data points along the lengths of the models). Initial lengths (white bars) are plotted as single values with no s.d. and with data values, rounded to the nearest integer, shown above the plotted bars.

4.3.2. Strain energy and stresses: Influence of crosslink state and fixed atoms

Qualitatively, it was observed that both the crosslink state and the number of fixed $C\alpha$ atoms had an influence over the strain energy-extension data (Figure 4.5) and the nominal stress-strain data (Figure 4.6). All of the strain energy and nominal stress data curves appear to be initially similar at lower extensions and strains; however, the data begin to diverge and become distinguishable at higher extensions and strains. The mature PYD crosslink exhibited the greatest levels of strain energy and nominal stress, followed by the mature HHL crosslink, the immature crosslinks (deH-HLNL and HLKLN), and the uncrosslinked state. Four visually distinct groups can be seen from the strain energy and nominal stress responses, namely, the four crosslinked models with 12 fixed $C\alpha$ atoms, the uncrosslinked model with 12 fixed $C\alpha$ atoms, the deH-HLNL crosslinked model with three fixed $C\alpha$ atoms, and the uncrosslinked model with three fixed $C\alpha$ atoms. The data shown here in Chapter 4 is based on our slowest pulling velocity of 6.25 m/s. Appendix F shows the strain energy-extension data and the nominal stress-strain data for all of the microfibril unit simulations at the five different pulling velocities (100, 50, 25, 12.5, and 6.25 m/s).

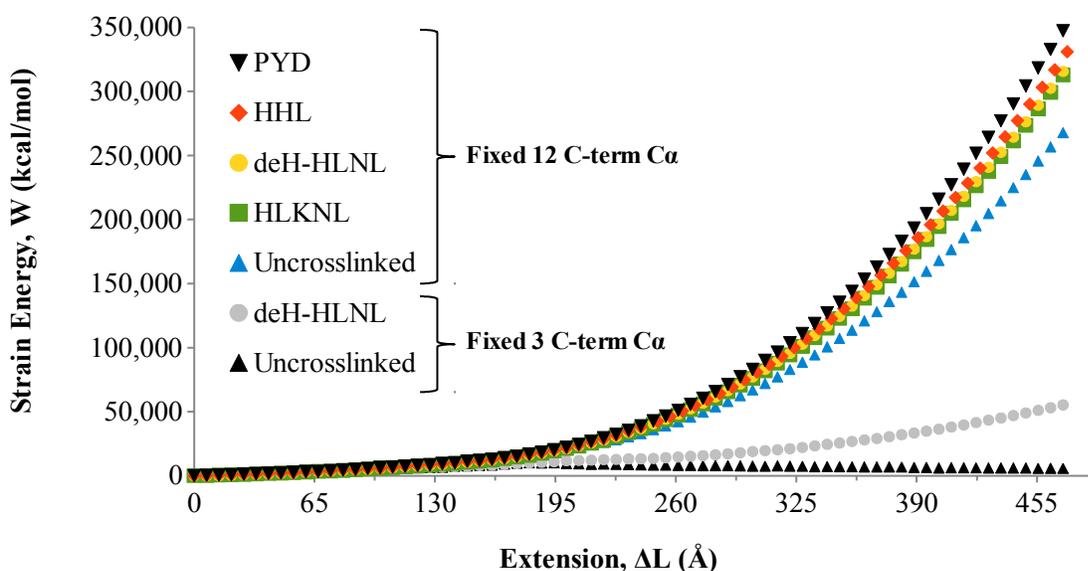


Figure 4.5: Strain energy-extension data for the microfibril unit models at 6.25 m/s. Here, strain energy represents the total energy stored in all ~38,900 atoms in each model as a function of extension; this total energy is the sum of kinetic energy and potential energy (see Equation 4.1). Data is shown for each of the five uniquely prepared models, namely, uncrosslinked and four crosslinked models (PYD, HHL, deH-HLNL, and HLKNL). Two alternative results are shown for the uncrosslinked and deH-HLNL-crosslinked models when the number of fixed C-terminal $C\alpha$ atoms was reduced from 12 to three.

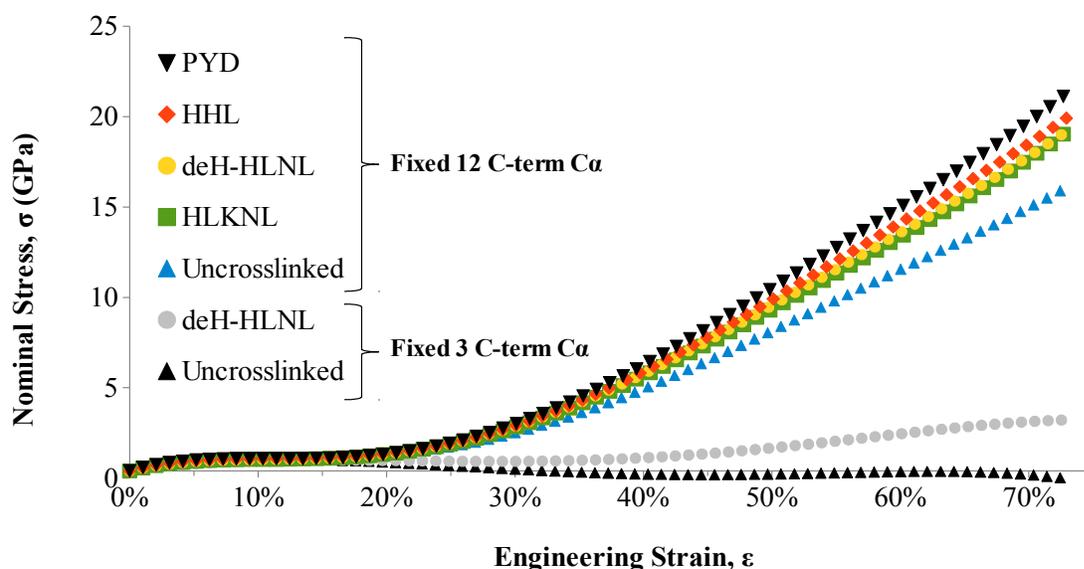


Figure 4.6: Stress-strain data for the microfibril unit models at 6.25 m/s. Here, nominal stress was derived from the total strain energy data (Figure 4.5), and engineering strain was derived from the extension data (Figure 4.5) and the initial lengths (Figure 4.4). Data is shown for each of the five uniquely prepared models, namely, uncrosslinked and four crosslinked models (PYD, HHL, deH-HLNL, and HLKNL). Two alternative results are shown for the uncrosslinked and deH-HLNL-crosslinked models when the number of fixed C-terminal $C\alpha$ atoms was reduced from 12 to three.

4.3.3. Elastic moduli: Influence of crosslink state and fixed atoms

Two elastic moduli were calculated for each of the eight simulation cases ($n = 3$ per simulation). One modulus, termed the low-strain modulus, was found before the transition point, while another modulus, termed the high-strain modulus was found after the transition point. As described in the methods, this transition point was defined as the point at which the second and third derivatives of a nominal stress function were maximal and zero, respectively (see Figure 4.7 for the transition points that were found). This method of objectively computing the transition points worked fairly well for all of the simulations except for the uncrosslinked model with three pulled $C\alpha$ atoms; the low-strain elastic modulus for the uncrosslinked model with three pulled $C\alpha$ atoms came out negative (-1.60 ± 0.12 GPa) (Table 4.4). This negative elastic modulus is due to the combination of the approach used to compute the transition points and the early mechanical failure of this particular simulation of the uncrosslinked microfibril unit (i.e., its nominal stress increases initially, but then decreased towards zero after approximately 10% strain) (Figure 4.6). This decrease in nominal stress can be seen more clearly in Figure F.6 of Appendix F due to its different y-axis scaling.

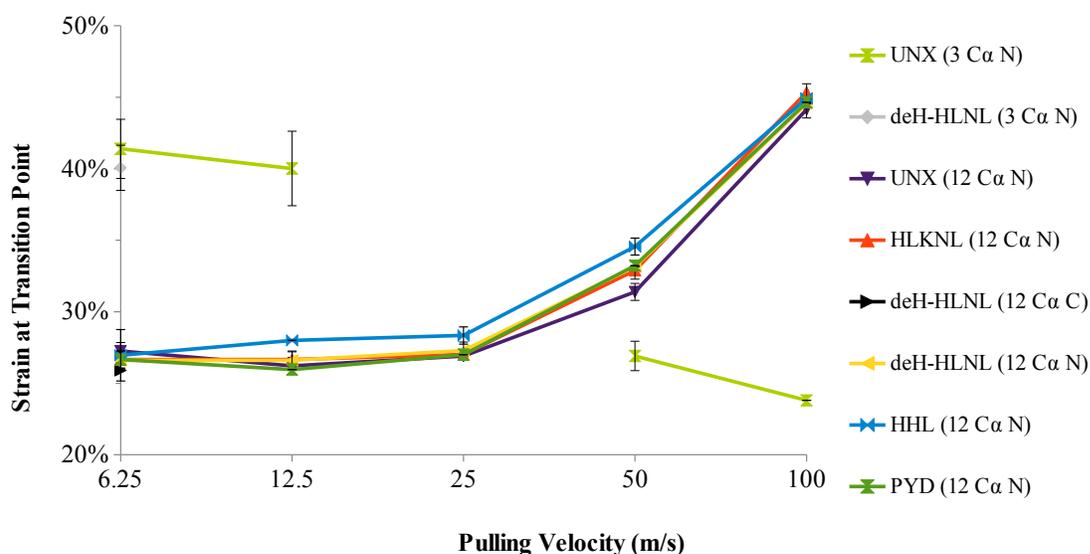


Figure 4.7: Transition point strain values used for determining low- and high-strain moduli. These transition points were determined by solving for the displacement value at which the third derivative of the nominal stress function was zero. Each data point represents a mean \pm s.d. The “3 $C\alpha$ ” or “12 $C\alpha$ ” indicates the number of fixed $C\alpha$ atoms, while the “C” or “N” indicates whether the model was pulled in the C- or N-terminal direction, respectively. Note that the “deH-HLNL (3 $C\alpha$ N)” and “deH-HLNL (12 $C\alpha$ C)” simulations were only run at 6.25 m/s. Also note that a transition point is missing for the “UNX (3 $C\alpha$ N)” simulation at 25 m/s. This was due to the approach used to find the transition point; specifically, imaginary roots were determined for these particular simulations while solving for the transition point.

Chapter 4: Specific Aim 2

The resulting low- and high-strain moduli are tabulated in Table 4.4 and plotted in Figure 4.8. It was found that the low-strain moduli were not significantly affected by the crosslink state (i.e., uncrosslinked, the two immature crosslink types, and the two mature crosslink types); however, when the number of fixed C α atoms was reduced from 12 to three, the low-strain moduli decreased significantly (see “Fixed 12 C α ” vs. “Fixed 3 C α ” for the uncrosslinked and deH-HLNL models in Figure 4.8). The high-strain modulus was significantly affected by both the crosslink state and the number of fixed C α atoms. For the case of 12 fixed C α atoms, the uncrosslinked model had the lowest high-strain modulus, followed by the immature crosslinks (HLKNL and deH-HLNL), the mature HHL crosslink, and then by the mature PYD crosslink. It was also found that the pulling direction had no significant influence over the low- or high-strain moduli (see deH-HLNL *N* vs. deH-HLNL *C* in Figure 4.8). The effects of the five pulling velocities (100, 50, 25, 12.5, and 6.25 m/s) on the low- and high-strain moduli are shown in Appendix G. It was determined that, generally, there was a significant influence from the pulling velocity for 100, 50, and 25, but not for 12.5 and 6.25 m/s, based on Student's t-tests with Bonferroni correction.

Table 4.4: Low- and high-strain elastic moduli for the microfibril unit models at 6.25 m/s.

Crosslink State	Fixed & Pulled Atoms	Low-Strain Modulus (GPa)	High-Strain Modulus (GPa)
Uncrosslinked	Fixed 3 C-term C α	-1.60 \pm 0.12	0.11 \pm 0.05
deH-HLNL	Pulled 12 N-term C α	0.82 \pm 0.10	7.06 \pm 0.42
Uncrosslinked	Fixed 12 C-term C α Pulled 12 N-term C α	4.18 \pm 0.24	31.9 \pm 0.09
HLKNL		4.67 \pm 0.11	37.7 \pm 0.23
deH-HLNL		4.83 \pm 0.30	37.9 \pm 0.31
HHL		4.86 \pm 0.71	39.9 \pm 0.29
PYD		4.63 \pm 0.27	42.4 \pm 0.21
deH-HLNL	Fixed 12 N-term C α Pulled 12 C-term C α	4.80 \pm 0.07	37.5 \pm 0.13

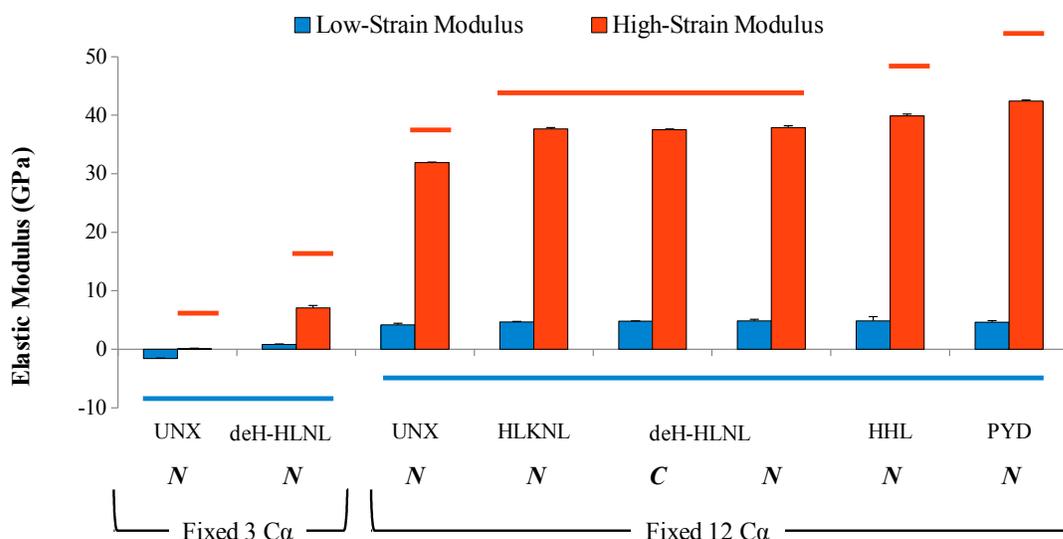


Figure 4.8: Low- and high-strain elastic moduli for the microfibril unit models at 6.25 m/s. The groups are denoted as follows: UNX (uncrosslinked model), deH-HLNL and HLKLN (immature crosslinked models), and HHL and PYD (mature crosslinked models). The “C” and “N” below the model names indicate whether the model was pulled in the C- or N-terminal direction, respectively. “Fixed 3 C α ” and “Fixed 12 C α ” indicate the number of alpha carbon atoms that were fixed at the end opposite from the 12 pulled C α atoms. The plotted values each represent a mean + s.d. (n = 3), and the eight groups are ordered from left to right by increasing high-strain modulus. The thicker horizontal bars above and below the plotted bars, for the high- and low-strain moduli, respectively, denote results from statistical analyses, which was based on a Welch’s one-way ANOVA ($p < 0.0001$ for both low- and high-strain moduli) followed by Student’s t-testing with a Bonferroni-corrected alpha value of 0.0018 (0.05 divided by 28 t-tests for the eight groups).

These results show that the presence of either immature or mature crosslinks led to a significant increase in the high-strain modulus, compared to the uncrosslinked state. Furthermore, the two immature crosslinks provided similar increases in the high-strain modulus, while the two mature crosslinks provided high-strain moduli that were both higher than the immature crosslinks and different from each other (i.e., the mature PYD crosslink provided a greater high-strain modulus than the mature HHL crosslink).

One reason for this observed outcome is that the two immature crosslinks are very similar in their chemical structure, while the two mature crosslinks are fairly different in their structure. Another reason is the manner in which the immature and mature crosslinks were connected between the two crosslinked molecular segments. The two immature crosslink types (deH-HLNL and HLKLN) join the same two amino acid positions, and they both link to a hydroxylysine from the triple-helical domain. Their difference lies in their other half, which comes from a telopeptide

Chapter 4: Specific Aim 2

domain; deH-HLNL links to a telopeptide lysine, while HLKLN links to a telopeptide hydroxylysine (Eyre & Wu, 2005). Additionally, deH-HLNL contains a carbon-nitrogen double bond between the amino acid side chains, while HLKLN contains a carbon-nitrogen single bond (see Figure 2.3); this may explain why deH-HLNL led to slightly higher low- and high-strain moduli, although insignificant, compared to HLKLN. These covalent bonds were modeled as springs; the carbon-nitrogen double bond of deH-HLNL was modeled as a shorter spring (smaller b_0 value) and a stiffer spring (higher K_b value) relative to the carbon-nitrogen single bond of HLKLN, through the use of the CHARMM all-atom force field (Accelrys Software, Inc., 2011). b_0 and K_b refer to the equilibrium length and force constant parameters in the CHARMM/CHARMM22 force fields shown in Equation 4.1.

The two mature crosslink types (HHL and PYD) were also connected to the same two amino acid positions as the immature crosslink types; however, since the mature crosslinks are tri-functional or trivalent, they are able to link a third amino acid side chain. The third amino acid for PYD is a hydroxylysine from a telopeptide domain, while HHL links to a histidine from the triple-helical domain. In PYD, the three amino acid side chains form their crosslink around a six-membered, aromatic ring structure; this aromatic ring contains several partial double bonds, which from a mechanics and modeling stand point are stiffer and shorter than single bonds. In contrast, HHL, while containing the five-membered ring structure from histidine, does not form a ring structure as a part of its actual crosslink; instead, the three amino acids form their crosslink around an sp^3 carbon atom containing three single bonds from each of the three side chains. These structural differences most likely give PYD an increased ability to store strain energy relative to HHL (see Figure 2.3 and Figure 2.4 to view differences in the chemical structures of these immature and mature crosslinks).

4.3.4. Crosslink strain energy and alignment: Influence of crosslink state

Each of our crosslinked microfibril unit molecular models contained two crosslinks: one near the N-terminal end (N-crosslink) and another near the C-terminal end (C-crosslink). It was observed that in all cases, the N-crosslinks exhibited greater levels of strain energy compared to the C-crosslinks. This difference in the N- and C-crosslinks is consistent with that observed for Aim 1. Here in Aim 2, this shows that the two-stage recruitment response observed in Aim 1 for the deH-HLNL crosslink is applicable to other immature and mature crosslinks.

Chapter 4: Specific Aim 2

It was also observed that there was a visible order or rank amongst the four crosslinked models. The mature PYD crosslink displayed the greatest levels of strain energy, followed by the HHL crosslink, and then by the immature crosslinks (deH-HLNL and HLKLN). This order is consistent with that observed for the strain energy and nominal stress data reported in section 4.3.2. The proposed reasoning for this ranked response is, as described in the previous section, due to differences in the chemical structures of the two immature crosslinks, the mature HHL crosslink, and the mature PYD crosslink.

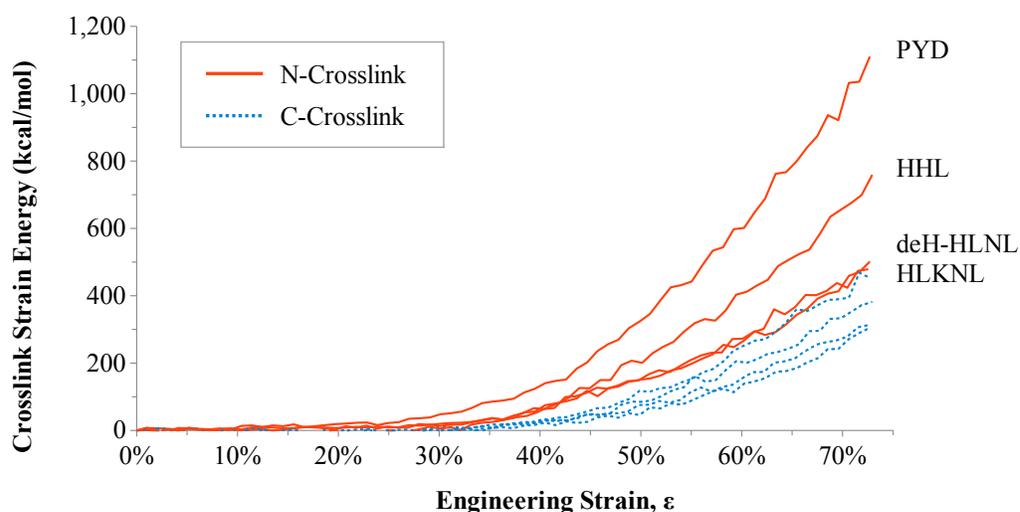


Figure 4.9: Strain energy in the immature and mature crosslinks at 6.25 m/s. Each microfibril unit molecular model contained two identical crosslinks – one located near the N-terminus and one located near the C-terminus. This plot shows how the N- and C-crosslinks in each model stored energy.

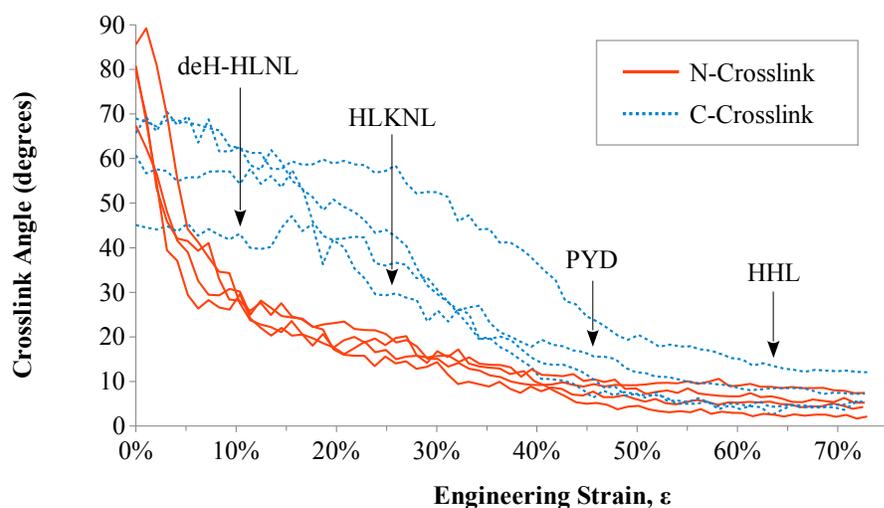


Figure 4.10: Crosslink orientation angles for the immature and mature crosslinks at 6.25 m/s, measured relative to the long axis of the microfibril unit models.

5.1. Introduction

Several experimental and theoretical investigations have been reported in the literature regarding the mechanical properties of collagen type I at the fibril level. We employed experimental and theoretical literature data reported by others to evaluate our fibril model. Such evaluations are important to assess assumptions, simplifications, limitations, and pitfalls in one's modeling approach. In the following two sub-sections, we describe crosslink composition data from the literature that was used while assigning valid crosslink prevalence values and collagen fibril mechanics data to which our model's predictions were compared.

5.1.1. Crosslink compositions of healthy collagenous tissues

There are numerous investigations described in the literature regarding the characterization and quantification of crosslinks in collagenous tissues. Most often, high-performance liquid chromatography (HPLC) has been employed for such work. Eyre *et al.* reported on the development of an HPLC method for the quantification of collagen crosslinks in various human tissues (e.g., articular cartilage, dentin, patellar tendon, ligament, cortical bone, and skin) with fluorescence or ultraviolet absorbance detection (Eyre, Koob, & Van Ness, 1984). Sims and Bailey reported on the age-dependent changes to the collagen crosslinks deH-HLNL, HLKLN, and HHL in bovine skin through the use of HPLC with ninhydrin detection. The immature HLKLN crosslink was reported to decline rapidly during fetal development and the immature deH-HLNL crosslink was shown to increase during fetal development up to approximately 15 months after birth, after which deH-HLNL declined gradually into maturity. In contrast, the mature crosslink HHL was shown to increase during development reaching a plateau into maturity (Sims & Bailey, 1992). Fujii *et al.* have reported on the prevalence of immature and mature crosslinks in different human ligaments and tendons. Age-related changes in collagen crosslinks were also reported for some of these tissues. Immature crosslinks were reduced and radioactively labeled with tritium (^3H) through reaction with NaB^3H_4 followed by HPLC, while the mature pyridinoline crosslinks (PYD and DPD) were quantified by HPLC with fluorescence detection (Fujii *et al.*, 1994). Saito *et al.* employed HPLC with fluorescence detection to quantify the concentration of different immature and mature collagen crosslinks in

human diaphysial femur (bone), articular cartilage, and skin based on tissues from 35-year-old humans; they also reported on the age-dependency of these crosslink concentrations in human diaphysial femur (Saito et al., 1997). By summing the immature and mature crosslink concentration data of Saito *et al.*, we calculated total crosslink concentrations of 1.74 mol/mol for diaphysial femur and 1.87 mol/mol for skin (both representing their 35-year-old human tissue specimens). Here, mol/mol means moles of crosslink per mole of collagen.

5.1.2. Mechanical properties of single collagen fibrils

The mechanical behavior of collagen fibrils have been reported in the peer-reviewed literature based on experimental techniques (e.g., atomic force microscopy (AFM) and microelectromechanical systems (MEMS)) and theoretical approaches (e.g., a viscoelastic spring-dashpot model and a mesoscopic model). van der Rijt also used AFM to obtain force-displacement measurements of single collagen fibrils, in their case, isolated from bovine Achilles tendons. The Young's moduli of their PBS-tested fibrils were reported in the range of 0.240 to 0.790 GPa (van der Rijt, 2004). Shen *et al.* developed a MEMS device for the measurement of tensile mechanical properties of individual collagen fibrils; their collagen type I fibrils were obtained from sea cucumber (*Cucumaria frondosa*) dermis. The average Young's modulus of their hydrated fibrils was reported as 0.470 ± 0.410 GPa (Shen et al., 2010). Table 5.1 provides a summary of these and other experimental reports from the literature.

Buehler developed a two-dimensional (2-D) mesoscopic plane stress model of two full-length microfibrils arranged in series. Each microfibril contained five full-length collagen molecules, and each molecule was in turn comprised of 2,000 2-D beads that were interconnected with springs. This mesoscopic model was a type of coarse-grain model in which several atoms are grouped together to form atom-groups. Potential energy was defined as the summation of extension, bending, and non-covalent terms. Covalent crosslinking was introduced through a numerical scaling of the non-covalent 6-12 Lennard-Jones term; in particular, the Lennard Jones energy parameter was multiplied by a parameter denoted by β , which was defined as being greater than or equal to one ($\beta = 1$ corresponded to 0 moles crosslink/mole collagen; $\beta = 12.5$ to 1 mol/mol; and $\beta = 25$ to 2 mol/mol). Low- and high-strain elastic moduli were reported as 4.36 and 38 GPa, respectively, for $\beta = 25$ (Buehler, 2008). Table 5.2 provides a summary of this and other theoretically based approaches from the literature.

Chapter 5: Specific Aim 3

Table 5.1: Experimentally determined elastic moduli for single native collagen type I fibrils. Acronyms: AFM (atomic force microscopy); ESEM (environmental scanning electron microscopy); HBSS (Hank’s balanced salt solution); MEMS (microelectromechanical systems); MES (2-(N-morpholino)ethanesulfonic acid); PBS (phosphate-buffered saline); SAXS (small-angle x-ray scattering); and UTS (ultimate tensile strength).

Source	Method	Test Conditions	Rate ($\mu\text{m/s}$)	Modulus (GPa)	Strain (%)	n	Ref.
Bovine Achilles tendon	AFM	PBS/MES pH 7.4/5.4	0.2 to 0.91	0.240 to 0.790 (n = 3)	63 \pm 46 <i>at failure</i> (n = 2)	3 or 2	(van der Rijt, 2004)
Sea cucumber dermis	MEMS device	PBS	0.034 \pm 0.001	0.470 \pm 0.410	80 \pm 44 <i>at failure</i>	11	(Shen et al., 2010)
Bovine Achilles tendon	AFM	PBS	0.1 to 10	0.600 \pm 0.200	11 to 15 <i>at failure</i>	11	(Yang et al., 2012)
Deer antler cortical bone	AFM-ESEM	HBSS	Not reported	2.38 \pm 0.37	5.96 \pm 0.62 <i>at UTS</i>	6	(Hang & Barber, 2011)
Human patellar tendon	AFM	PBS 150 mM	9.81 314	2.89 \pm 0.23	4.25 \pm 0.48 <i>cyclic</i>	2	(Svensson et al., 2010)
Sea cucumber dermis	MEMS device	PBS	Not reported	0.400 to 12 <i>fibril B</i>	\sim 40 <i>at failure</i>	1	(Eppell et al., 2006)

Table 5.2: Theoretically determined elastic moduli for collagen fibrils.

Model type	Simulation Conditions	Rate (m/s)	Modulus (GPa)	Ref.
Kelvin-Voigt model	Hydrated	N/A	2.5	(Gautieri et al., 2012)
2-D mesoscopic plane stress model (two full-length microfibrils)	Hydrated (parameters based on MD with TIP3P water); Crosslinked ($\beta = 25$)	0.4	4.36 (lower strains) 38 (higher strains)	(Buehler, 2008)

5.2. Methods

The nominal stress-strain data for each of the microfibril unit molecular models was used as a basis for the mathematical model of a single collagen fibril, which was modeled using principles from linearly viscoelastic mechanical analogs. The following sub-section describes assumptions and simplifications of our approach.

5.2.1. Assumptions and simplifications

1) The N- and C-telopeptides were simplified as being crosslinked with two identical crosslinks, rather than including the cases of combinations of different crosslinks within each microfibril unit. Furthermore, the preferential formation of different crosslinks at the N- or C-crosslink site was not considered. Hanson and Eyre reported that the mature pyridinoline crosslink (PYD) showed no site-specificity for the N- or C- crosslink site (Hanson & Eyre, 1996). In contrast, the mature HHL crosslink has been reported to be concentrated at the C-terminal crosslink site (Yamauchi, 1996). We made the simplification that HHL is equally prevalent at the N-terminal site. Note that the requisite amino acid precursors were already present at appropriate locations to form HHL at the N-terminal site; thus, no sequence modification was necessary.

2) Microfibrils were simplified as being aligned parallel to the long-axis of the fibril. It has been reported that microfibrils actually display tilted orientations such that they are oriented slightly off from the long-axis of the fibril (Orgel et al., 2006). Thus, this simplification may be valid only when considering a pre-loaded fibril, which is typically the case when single fibrils are tested experimentally (van der Rijt, van der Werf, Bennink, Dijkstra, & Feijen, 2006; van der Rijt, 2004).

3) Crosslink composition was assumed to be homogenous along the fibril's length. In whole tissues, localized differences have been reported (e.g., the region of a tendon more proximal to bone can display a different crosslink composition than the more distal region (Knott, Tarlton, & Bailey, 1997)). However, given the difference in length scale between a whole tissue and a collagen fibril, localized differences at the tissue level are not expected to be significant at the single fibril level.

4) A cylindrical shape was assumed for the fibril with a constant diameter throughout its length. Collagen fibrils *in vivo* normally exhibit tapered ends with a smaller diameter than the rest of the fibril, leading to an ellipsoidal rather than cylindrical shape. Goh *et al.*, through mathematical modeling of cylindrical and tapered collagen fibrils, predicted that tapered fibrils display a more uniform stress distribution along their length compared to non-tapered, cylindrical fibrils. It has been proposed that tapered fibrils are less likely to fracture than

cylindrical fibrils, and that per unit volume, tapered fibrils are stronger (Goh, Meakin, Aspden, & Hukins, 2005).

5) Only two inter-molecular crosslinks were accounted for in our approach to link microfibrils in the axial direction of the molecules; inter-molecular crosslinks linking microfibrils in the lateral direction were not incorporated.

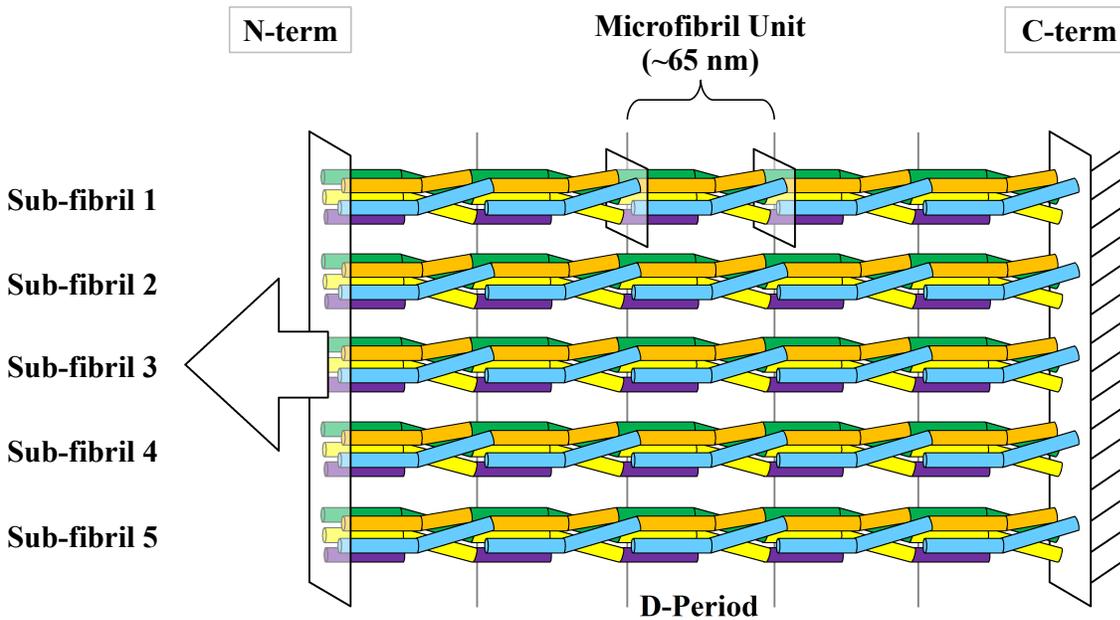


Figure 5.1: A schematic of an axial/longitudinal cross-section of our modeled fibril. The color scheme is identical to that described in Figure 3.2. For visual simplicity, this schematic contains an arbitrarily small number of microfibril units per sub-fibril (i.e., five microfibril units) and an arbitrarily small number of sub-fibrils within the axial cross-section (i.e., five sub-fibrils). Given that fibrils can be tens to hundreds of μm in length and tens to hundreds of nm in diameter and that our microfibril units are $\sim 65\text{-nm}$ in length with cross-sectional areas of $\sim 755 \text{ \AA}^2$, within our modeled fibril, one might expect thousands of microfibril units within each sub-fibril along the length of the fibril and hundreds of sub-fibrils along the diameter of the fibril.

5.2.2. Model formulation

1) Equilibrium equations for parallel and series combinations of microfibril units:

- Series: Let the force and cross-sectional area of a sub-fibril be equal to that of each of its microfibril units ($\mu\text{fibrils}$) arranged in series.

$$F_{sub-fibril} = F_{\mu fibril, 1} = F_{\mu fibril, 2} = \dots = F_{\mu fibril, m} = F_{\mu fibril}$$

$$A_{0, sub-fibril} = A_{0, \mu fibril, 1} = A_{0, \mu fibril, 2} = \dots = A_{0, \mu fibril, m} = A_{0, \mu fibril}$$

Chapter 5: Specific Aim 3

- Parallel: Let the force in a fibril be equal to the summed forces of all of its sub-fibrils arranged in parallel.

$$F_{fibril} = F_{sub-fibril,1} + F_{sub-fibril,2} + \dots + F_{sub-fibril,n}$$

$$F_{fibril} = \sum_{j=1}^n F_{sub-fibril,j} = \left[\frac{A_{0,fibril}}{A_{0,sub-fibril}} \right] * F_{sub-fibril}$$

- Substitute $F_{sub-fibril} = F_{\mu fibril}$ and $A_{0,sub-fibril} = A_{0,\mu fibril}$:

Equation 5.1: Fibril force expressed as a summation of identical sub-fibrils in parallel, where each sub-fibril represents identical microfibril units in series.

$$F_{fibril} = \left[\frac{A_{0,fibril}}{A_{0,\mu fibril}} \right] * F_{\mu fibril}$$

2) Geometry of deformation equations for parallel and series combinations of microfibril units:

- Parallel: Let the displacement and initial length of a fibril be equal to that of each of its sub-fibrils arranged in parallel.

$$\Delta L_{fibril} = \Delta L_{sub-fibril,1} = \Delta L_{sub-fibril,2} = \dots = \Delta L_{sub-fibril,n} = \Delta L_{sub-fibril}$$

$$L_{0,fibril} = L_{0,sub-fibril,1} = L_{0,sub-fibril,2} = \dots = L_{0,sub-fibril,n} = L_{0,sub-fibril}$$

- Series: Let the displacement in a sub-fibril be equal to the summed displacements of all of its microfibril units (μ fibrils) arranged in series.

$$\Delta L_{sub-fibril} = \Delta L_{\mu fibril,1} + \Delta L_{\mu fibril,2} + \dots + \Delta L_{\mu fibril,m}$$

$$\Delta L_{sub-fibril} = \sum_{k=1}^m \Delta L_{\mu fibril,k} \rightarrow \Delta L_{sub-fibril} = \left[\frac{L_{0,sub-fibril}}{L_{0,\mu fibril}} \right] * \Delta L_{\mu fibril}$$

- Substitute $\Delta L_{fibril} = \Delta L_{sub-fibril}$ and $L_{0,fibril} = L_{0,sub-fibril}$:

$$\Delta L_{fibril} = \left[\frac{L_{0,fibril}}{L_{0,\mu fibril}} \right] * \Delta L_{\mu fibril}$$

- Substitute engineering strains (ϵ 's) and initial lengths (L_0 's) for displacements (ΔL 's):

$$\epsilon_{fibril} * L_{0,fibril} = \left[\frac{L_{0,fibril}}{L_{0,\mu fibril}} \right] * (\epsilon_{\mu fibril} * L_{0,\mu fibril}) \rightarrow \epsilon_{fibril} = \epsilon_{\mu fibril}$$

Chapter 5: Specific Aim 3

3) From Equation 5.1, add crosslink prevalence (P_i 's) and express forces as functions of ΔL 's:

Equation 5.2: Fibril force expressed as a summation of different sub-fibrils in parallel, where each sub-fibril represents identical microfibril units in series.

$$F_{fibril}(\Delta L_{fibril}) = \sum_i \left[n_i * P_i * F_{\mu_{fibril},i}(\Delta L_{\mu_{fibril},i}) \right]$$

where $n_i = A_{0,fibril}/A_{0,\mu_{fibril},i}$ and $i = \text{UNX, deH-HLNL, HLKLN, HHL, or PYD}$

4) From Equation 5.2, obtain fibril stress (σ_{fibril}) as a function of fibril strain (ϵ_{fibril}):

- Divide each side by the initial cross-sectional area of a fibril ($A_{0,fibril}$)
- Substitute strains (ϵ 's) and initial lengths (L_0 's) for displacements (ΔL 's)
- Substitute $\sigma = F/A_0$ and $\epsilon_{\mu_{fibril},i} = \epsilon_{fibril}$

Equation 5.3: Fibril stress as a function of fibril strain and initial fibril length. The right-hand side is a summation of microfibril unit stresses as a function of fibril strains and initial microfibril unit lengths. These microfibril unit stresses are weighed via their multiplication with crosslink prevalence parameters.

$$\sigma_{fibril}(\epsilon_{fibril} * L_{0,fibril}) = \sum_i \left[P_i * \sigma_{\mu_{fibril},i}(\epsilon_{fibril} * L_{0,\mu_{fibril},i}) \right]$$

5.2.3. Qualitative comparison with data from the literature

In order to make qualitative comparisons, data obtained from our fibril model was compared to published data describing the mechanical testing of single collagen fibrils. Data from these published works were converted into tabular form using a program called Engauge Digitizer (Mitchell, 2002). First, a snapshot of the figure was exported from its PDF document at a sufficiently high resolution and into a suitable image format (e.g., a 300 DPI .jpg or .png file), followed by an importing of this image file into Engauge Digitizer. Three axis points were then defined (e.g., 0, 0; Xmax, 0; and 0, Ymax), and the data points were manually added using the Curve Point tool. The data points were then exported into a comma-separated plain-text file (.csv) for use in a spreadsheet program for re-plotting.

5.2.4. Quantitative comparison with data from the literature

In order to make quantitative comparisons, elastic moduli were computed from our data for comparison to experimental and theoretical literature data. Low-strain and high-strain moduli were computed as described in section 4.2.8. At this point, means and standard deviations could

not be expressed for our fibril data, because this fibril data were based on the mean nominal stress-strain data from the microfibril unit molecular models.

5.3. Results and Discussion

5.3.1. Qualitative comparison with data from the literature

A re-plotting of literature data with our fibril nominal stress-strain data was conducted (Figure 5.2). Our model data was plotted at total crosslink concentrations of 0 mol/mol ($P_{\text{deH-HLNL}} = P_{\text{HHL}} = P_{\text{HLKNL}} = P_{\text{PYD}} = 0$; $P_{\text{UNX}} = 1$), 1 mol/mol ($P_{\text{deH-HLNL}} = P_{\text{HHL}} = P_{\text{HLKNL}} = P_{\text{PYD}} = 0.125$; $P_{\text{UNX}} = 0.5$), and 2 mol/mol ($P_{\text{deH-HLNL}} = P_{\text{HHL}} = P_{\text{HLKNL}} = P_{\text{PYD}} = 0.25$; $P_{\text{UNX}} = 0$), where mol/mol means moles of crosslink per mole of collagen (see Equation 5.3). These P_i values represent the percentage of a fibril occupied by the uncrosslinked state (UNX), immature crosslinked states (deH-HLNL or HLKNL), or mature crosslinked states (HHL or PYD). These total crosslink concentrations (0, 1, and 2 mol/mol) span from the lower limit (0 mol/mol) to the upper limit (2 mol/mol) for our model (i.e., a maximum of two crosslinks per microfibril unit).

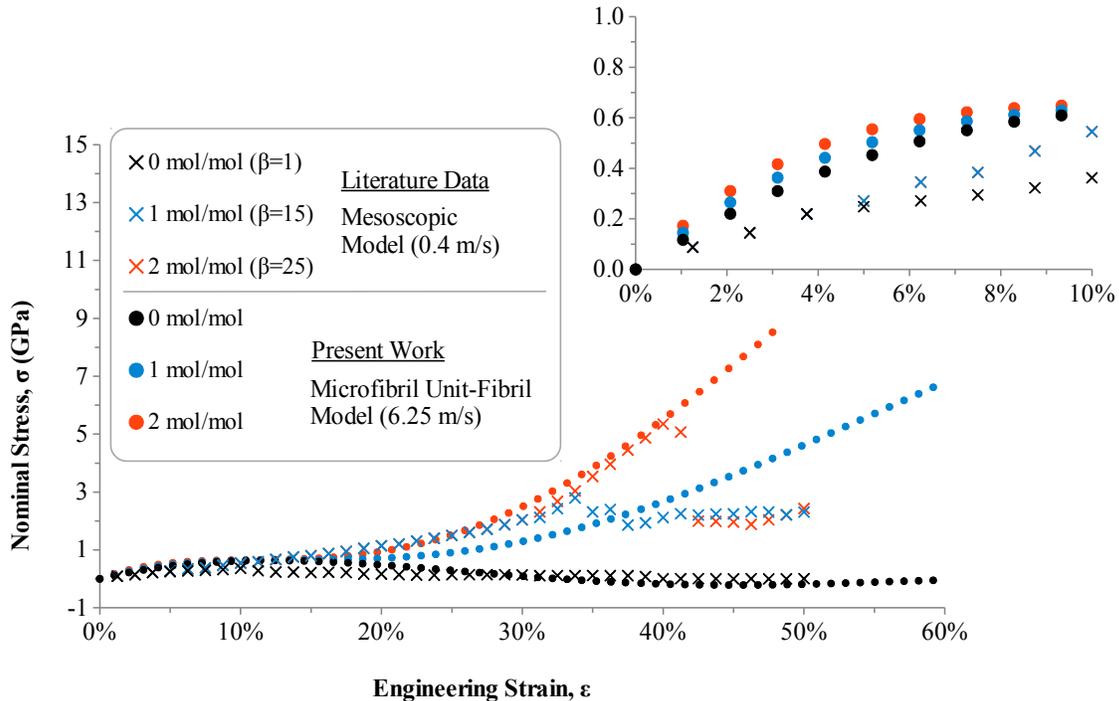


Figure 5.2: Stress-strain data for our fibril model plotted with data from the literature. The main figure is plotted from 0 to 60% strain, while the inset figure is plotted from 0 to 10% strain. The literature data curves are from a 2-D mesoscopic plane stress model run at 0.4 m/s with several different crosslink densities; we have digitized data for three of the author's crosslink densities, namely, 0 moles of crosslink

Chapter 5: Specific Aim 3

per mole of collagen (mol/mol) (black x's), 1 mol/mol (blue x's), and 2 mol/mol (red x's). Corresponding data curves from the present work were plotted with black, blue, and red filled circles, respectively. The digitized literature data curves were included, with permission from Elsevier, from the Journal of the Mechanical Behavior of Biomedical Materials (Vol. 1, No. 1, M. J. Buehler, Nanomechanics of collagen fibrils under varying cross-link densities: Atomistic and continuum studies, pages 59-67, 2008, Fig. 3).

Below approximately 15% strain, our model predicts stress-strain responses that are insensitive to crosslink concentration, as seen by the overlap of our 0, 1, and 2 mol/mol data. However, for higher strains, crosslink concentration leads to marked differences in the stress-strain responses; most notably, our model predicts that increasing crosslink concentrations lead to increases in the slope of the nominal stress-strain curve after ~15% strain. These differences in slope were quantified and are described next in section 5.3.2.

At 0 mol/mol, our model is in fairly good agreement with that of Buehler's 2-D mesoscopic model. However, our model over-estimates at lower strains and under-estimates at higher strains relative to Buehler's model. Our model's over- and under-estimation is most likely due to the use of a 6th order polynomial to fit our strain energy-extension data, which was used to derive this nominal stress-strain data. At 1 mol/mol, our model does not seem to agree with Buehler's model; they agree somewhat under ~15% strain, but afterwards, our model exhibits a decreased elastic modulus. Furthermore, at ~34% strain, Buehler's model is able to account for crosslink failure – a feature that our model does not yet incorporate. At 2 mol/mol, our model shows very good agreement with Buehler's model up to 40% strain, after which Buehler's model exhibits crosslink failure. At 1 and 2 mol/mol, our model also displays a slight over-estimation in nominal stress below 10% strain.

The use of an alternative function could improve our model's predictions at lower strains (e.g., a Fung exponential strain energy function); however, an exponential function would be unable to capture the behavior of the uncrosslinked model, which displays an initial increase in nominal stress followed by a decrease at higher strains. Furthermore, it was observed that the use of an exponential function led to a nominal stress function that diverged notably from the nominal stress-strain data obtained via a numerical derivative approach.

5.3.2. Quantitative comparison with data from the literature

Low- and high-strain moduli were computed for our 1 mol/mol data (2.74 and 20 GPa for low and high strains, respectively) and our 2 mol/mol data (4.49 and 39 GPa for low and high strains, respectively). Standard deviations were not computed because these moduli were based on averaged and combined data from the different microfibril unit molecular models. Our low- and high-strain moduli at 2 mol/mol (4.49 and 39 GPa) are in very good agreement with that reported by Buehler from 2-D mesoscopic modeling (4.36 to 38 GPa (Buehler, 2008)). Our low-strain modulus at 1 mol/mol (2.74 GPa) is also in very good agreement with experimental literature data based on AFM (2.89 GPa (Svensson et al., 2010) and 2.38 GPa (Hang & Barber, 2011)) and theoretical literature data based on a Kelvin-Voigt model (2.5 GPa (Gautieri et al., 2012)); however, our low-strain modulus at 1 mol/mol is also up to an order of magnitude greater than experimental literature data based on MEMS (0.470 ± 0.410 GPa (Shen et al., 2010)) and AFM (0.240 to 0.790 GPa (van der Rijt, 2004) and 0.600 ± 0.200 GPa (Yang et al., 2012)). Our high-strain modulus at 1 mol/mol (20 GPa) is approximately 1.6x greater than that reported from a MEMS-based investigation (12 GPa) (Eppell et al., 2006). To our knowledge, this is the only experimental report that was found to report a high-strain modulus for single collagen fibrils; one reason for this is that there are limitations in the displacement range accessible through AFM-based techniques (van der Rijt, 2004) – one of the two main techniques that has been used to mechanically characterize single collagen fibrils, and while MEMS-based techniques do not necessarily have this limitation, there is a large variation in the failure strains for single collagen fibrils (e.g., $80 \pm 44\%$ with a range from 33 to 183% strain (Shen et al., 2010)) making it difficult to obtain high-strain moduli representative of greater sample sizes. For instance, in the one report of a high-strain modulus mentioned above, the value of 12 GPa was based on one fibril ($n = 1$); two fibrils were tested, but the other fibril failed in the low-strain region.

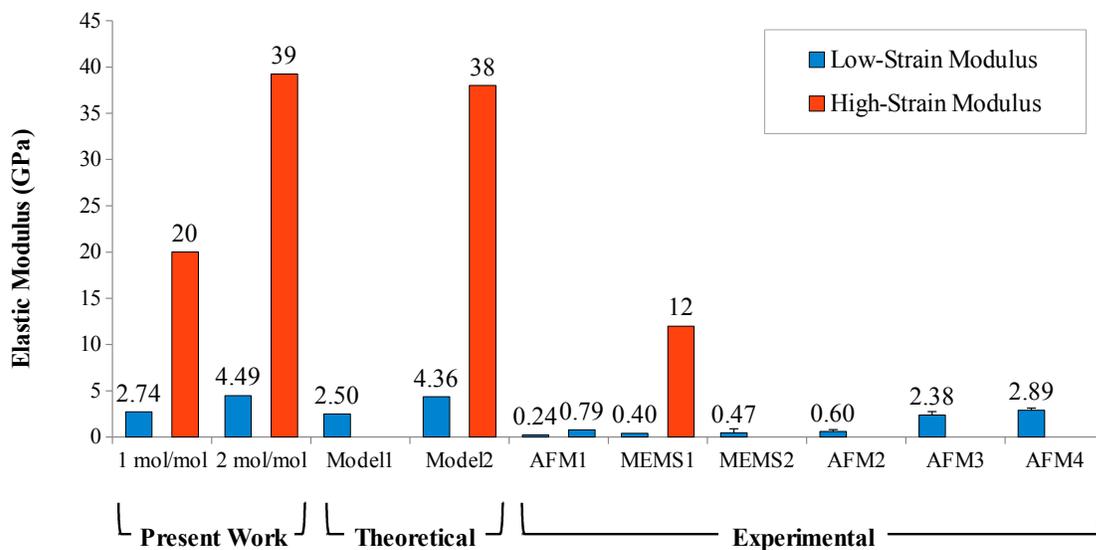


Figure 5.3: Elastic moduli for our fibril model plotted with literature data. Low- and/or high-strain moduli are plotted, as reported, for theoretical and experimental work described in the literature. The theoretical literature data includes Model1 (a Kelvin-Voigt-based model (Gautieri et al., 2012)) and Model2 (a 2-D mesoscopic model (Buehler, 2008)). The experimental literature data includes AFM1 (AFM testing of bovine Achilles tendon collagen fibrils, $n = 1$ (van der Rijt, 2004)), MEMS1 (MEMS testing of a sea cucumber dermis collagen fibril, $n = 1$ (Eppell et al., 2006)), MEMS2 (MEMS testing of sea cucumber dermis collagen fibrils, $n = 11$ (Shen et al., 2010)), AFM2 (AFM testing of bovine Achilles tendon collagen fibrils, $n = 11$ (Yang et al., 2012)), AFM3 (AFM testing of deer antler bone collagen fibrils, $n = 6$ (Hang & Barber, 2011)), and AFM4 (AFM testing of human patellar tendon collagen fibrils, $n = 2$ (Svensson et al., 2010)).

5.4. Future work: Revisiting modeling assumptions and simplifications

There are several underlying assumptions and simplifications that can be addressed in our fibril modeling approach that can improve the model's predictions. It is believed that the most crucial simplification that should be addressed is the stipulation that each sub-fibril represents only one crosslink state (i.e., each sub-fibril was either completely uncrosslinked or crosslinked with a specific crosslink type).

Another simplification was our use of classical molecular mechanics force fields (e.g., CHARMM19 united-atom and CHARMM all-atom) that are not meant to describe covalent bond failure. Force fields such as these are more suitable for low strains where covalent bond lengths and angles do not deviate too far from their equilibrium values. In classical force fields, covalent bonds are modeled as unbreakable springs that can continue to stretch and store strain energy indefinitely.

Together, our simplification of sub-fibrils containing only one crosslink state and covalent bonds that were unable to fail led to crosslinked sub-fibrils that continued to stretch and store energy; with these simplifications, this would theoretically continue indefinitely. Our uncrosslinked sub-fibrils were able to fail through shear and the loss of non-covalent interactions (e.g., hydrogen bonds and salt bridges), but this effect was overshadowed in the presence of completely crosslinked sub-fibrils.

Incorporating covalent crosslink failure: To address this simplification and to at least account for covalent crosslink failure, the following approach could be developed. Additional simulations could be run with our crosslinked models with the N-crosslink removed, since our data suggests that the N-crosslink exhibits greater levels of strain energy and is thus more likely to fail before the C-crosslink. Then one would possess nominal stress-strain data for the uncrosslinked state and for each of the four crosslinked states with two crosslinks (N- and C-crosslinks) and with only one crosslink (C-crosslink). One could then specify a threshold value of crosslink strain after which the N-crosslinks might be expected to fail. Now, when coupling the stress-strain responses from the microfibril units, one could apply these different cases (two crosslinks, one crosslink, or zero crosslinks) for the different crosslinked microfibril units depending on the level of strain. For instance, initially, the crosslinked microfibril units would be represented by the two crosslink case. At some level of strain, these crosslinked microfibril units would switch to the one crosslink case, and at some further level of strain, these microfibril units would switch to the zero crosslink case. The uncrosslinked microfibril units would retain the same zero crosslink case regardless of the level of strain.

Incorporating different crosslink states within each sub-fibril: The next challenge would be allowing sub-fibrils to contain a distribution of different crosslink states (uncrosslinked, immature HLKLN or deH-HLNL, or mature HHL or PYD). Essentially, given a specified crosslink concentration, we would seek to randomize the distribution of these different crosslink states within each sub-fibril. This would allow the sub-fibrils to undergo failure due to the presence of uncrosslinked microfibril units and the switching of the number of crosslinks within the crosslinked microfibril units with higher strains.

Interpolating boundary condition responses (fixed C α atoms): Another simplification to be addressed in connection to the representation of the sub-fibrils is the application of

Chapter 5: Specific Aim 3

boundary conditions for the microfibril units. For our microfibril unit simulations, we used fixed boundary conditions at the C-terminal end (12 fixed C α atoms for the crosslinked microfibril units and three fixed C α atoms for the uncrosslinked microfibril unit). This approach was consistent with the simplification of identical microfibril units within each sub-fibril. However, with a random distribution of different crosslink states within each sub-fibril, additional boundary conditions need to be considered and incorporated.

We pulled one of the crosslinked microfibril units with only three fixed C α atoms; this was meant to represent the situation of having that crosslinked microfibril unit followed by three uncrosslinked microfibril units within the same sub-fibril when moving in the C-terminal axial direction. The case of zero fixed C α atoms would correspond to the case of a microfibril unit being followed by four uncrosslinked microfibril units; zero fixed atoms would better correspond to the case of a completely uncrosslinked sub-fibril than the case of three C α fixed atoms. At the other extreme, the deH-HLNL microfibril unit pulled against 12 fixed C α atoms represents the case when a microfibril unit is followed by four crosslinked microfibril units. In order to account for sub-fibrils containing different crosslink states, the response of the microfibril units could be interpolated between these lower and upper limits (i.e., zero fixed C α atoms and 12 fixed C α atoms, respectively), depending on the types of other microfibril units in the local axial surrounding of each microfibril unit. For instance, if a deH-HLNL-crosslinked microfibril unit was followed by four uncrosslinked microfibril units when going in the axial direction (e.g., the C-terminal direction), then it would be represented by the lower limit (zero fixed C α atoms). However, if this deH-HLNL-crosslinked microfibril unit was followed by four crosslinked microfibril units, then it would be represented by the upper limit (12 fixed C α atoms).

If this deH-HLNL-crosslinked microfibril unit was followed by one crosslinked microfibril unit and three uncrosslinked microfibril units, then the interpolation factor, between the lower and upper limits, would be 1/4. When followed by two crosslinked and two uncrosslinked units, the interpolation factor would be 2/4 (or 1/2). When followed by three crosslinked and one uncrosslinked unit, the interpolation factor would be 3/4. Thus, the local axial surrounding of each microfibril unit (as determined by a randomized distribution of crosslink states based on a desired crosslink concentration) would determine the application of

this interpolation. We also pulled the uncrosslinked microfibril unit with 12 fixed C α atoms. Thus, this interpolation would apply to both the uncrosslinked and crosslinked microfibril units.

Investigating reduced pulling velocities: The pulling velocities employed in our microfibril unit simulations ranged from 100 down to 6.25 m/s, and all of our fibril predictions (nominal stress-strain data and elastic moduli) were based on the simulation data acquired at 6.25 m/s, which may be elevated given that this velocity is several orders of magnitude greater than the experimental velocities to which our predictions were compared (on the order of nm/s to $\mu\text{m/s}$). It has been reported through MD simulations that the mechanical behavior of collagen-like molecules is strain-rate dependent, but that at sufficiently reduced pulling velocities (less than 0.5 to 1 m/s), this behavior is fairly insensitive to the velocity (Gautieri et al., 2009). Given the size of our larger ~ 65 -nm-long microfibril unit models ($\sim 38,900$ atoms), slower velocities (i.e., below 6.25 m/s) have not yet been investigated due to the time and computing resources that would be required. However, this could be explored in future work.

Chapter 6: Summary

In summary, we have discussed our research efforts to better understand the molecular dynamics and mechanics of collagen type I and its lysine/hydroxylysine-derived crosslinks. Chapter 1 presents an introduction. Chapter 2 discusses general background relating to collagen proteins, collagen type I, and enzymatic collagen crosslinks. Chapter 3 describes molecular modeling work involving a ~23-nm-long region (overlap region) where the ends of two adjacent collagen type I molecules are crosslinked, while Chapter 4 presents work involving ~65-nm-long microfibril units in representation of the D-period (overlap and gap region). Lastly, Chapter 5 presents a multiple-scale model to predict the mechanical response of a single collagen fibril based on molecular modeling data and crosslink composition data.

For Aim 1, it was observed that the modeled crosslinks underwent alignment in response to uniaxial deformation along the molecular axis. This alignment was influenced by crosslink location such that N-crosslinks aligned more rapidly at lower strains and C-crosslinks more gradually at higher strains. This led us to propose an N- and C-crosslink recruitment mechanism, wherein N-crosslinks serve as first responders at lower strains and C-crosslinks as back-up responders at higher strains. This recruitment mechanism is believed to stem from the stark differences in the N- and C-telopeptide conformations (i.e., a more extended conformation for the N-telopeptides and a hook/hair-pin conformation for the two longer $\alpha 1$ C-telopeptides).

For Aim 2, it was observed that different crosslink types (e.g., immature HLKLN and deH-HLNL; mature HHL and PYD) led to significant increases in the high-strain elastic moduli of collagen type I, but not the low-strain moduli. This suggests that when modeling crosslinked collagen type I at low strains, it may not be necessary to distinguish between crosslink types; however, accounting for different crosslink types becomes more important at higher strains. The alignment and recruitment behavior of N- and C-crosslinks was also confirmed in Aim 2 with different crosslink types.

For Aim 3, predictions at the fibril level showed very good agreement with a 2-D mesoscopic model, a Kelvin-Voigt model, and a couple of AFM-based experimental reports; however, our predictions were also in large over-estimation of several AFM- and MEMS-based experimental reports. We concluded with a discussion of assumptions and simplifications to be addressed in future work.

References

- Accelrys Software, Inc. (2011). *Discovery Studio Modeling Environment*. San Diego, CA: Accelrys Software, Inc.
- Amiel, D., & Kleiner, J. B. (1988). Biochemistry of tendon and ligament. In M. E. Nimni (Ed.), *Collagen: Biotechnology* (Vol. 3, pp. 223–251). Boca Raton, FL: CRC Press.
- Anasuya, A., & Rao, B. S. (1970). Relationship between body collagen and urinary hydroxyproline excretion in young rats fed on a low-protein or low-calorie diet. *British Journal of Nutrition*, 24(1), 97–107. doi:10.1079/BJN19700013
- Avery, N. C., & Bailey, A. J. (2008). Restraining cross-links responsible for the mechanical properties of collagen fibers: Natural and artificial. In P. Fratzl (Ed.), *Collagen: Structure and Mechanics* (pp. 81–91). New York, NY: Springer US.
- Bailey, A. J., Paul, R. G., & Knott, L. (1998). Mechanisms of maturation and ageing of collagen. *Mechanisms of Ageing and Development*, 106(1-2), 1–56. doi:10.1016/S0047-6374(98)00119-5
- Bank, R. A., Robins, S. P., Wijmenga, C., Breslau-Siderius, L. J., Bardeol, A. F., van der Sluijs, H. A., ... TeKoppele, J. M. (1999). Defective collagen crosslinking in bone, but not in ligament or cartilage, in Bruck syndrome: indications for a bone-specific telopeptide lysyl hydroxylase on chromosome 17. *Proceedings of the National Academy of Sciences of the United States of America*, 96(3), 1054–1058. doi:10.1073/pnas.96.3.1054
- Bateman, J. F., Lamande, S. R., & Ramshaw, J. A. M. (1996). Collagen superfamily. In W. D. Comper (Ed.), *Extracellular Matrix: Molecular Components and Interactions* (Vol. 2, pp. 41–43). The Netherlands: Harwood Academic Publishers.
- Berendsen, H. J. C., Grigera, J. R., & Straatsma, T. P. (1987). The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91(24), 6269–6271. doi:10.1021/J100308a038
- Berisio, R., Vitagliano, L., Mazzarella, L., & Zagari, A. (2002). Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)(10)](3). *Protein Science*, 11(2), 262–270. doi:10.1110/ps.32602
- Bhandarkar, M., Bhatele, A., Bohm, E., Brunner, R., Buelens, F., Chipot, C., ... Zhu, F. (2011). *NAMD User's Guide, Version 2.8*. Retrieved from <http://www.ks.uiuc.edu/Research/namd/2.8/ug/>
- Birk, D. E., & Bruckner, P. (2005). Collagen suprastructures. In J. Brinckmann, H. Notbohm, &

Back Matter: References

- P. K. Muller (Eds.), *Collagen: Primer in Structure, Processing, and Assembly* (Vol. 247, pp. 185–205). New York, NY: Springer Berlin Heidelberg.
- Birk, D. E., Nurminkaya, M. V., & Zycband, E. I. (1995). Collagen fibrillogenesis in situ: fibril segments undergo post-depositional modifications resulting in linear and lateral growth during matrix development. *Developmental Dynamics*, 202(3), 229–243.
doi:10.1002/aja.1002020303
- Birk, D. E., Zycband, E. I., Winkelmann, D. A., & Trelstad, R. L. (1990). Collagen fibrillogenesis in situ. Discontinuous segmental assembly in extracellular compartments. *Annals of the New York Academy of Sciences*, 580, 176–194. doi:10.1111/j.1749-6632.1990.tb17928.x
- Bissonnette, B., Luginbuehl, I., Marciniak, B., & Dalens, B. J. (2006). *Syndromes: Rapid Recognition and Perioperative Implications*. New York, NY: McGraw-Hill Companies.
- Boot-Handford, R. P., & Tuckwell, D. S. (2003). Fibrillar collagen: The key to vertebrate evolution? A tale of molecular incest. *BioEssays*, 25(2), 142–151. doi:10.1002/bies.10230
- Bozec, L., & Horton, M. (2005). Topography and mechanical properties of single molecules of type I collagen using atomic force microscopy. *Biophysical Journal*, 88(6), 4223–4231. doi:10.1529/biophysj.104.055228
- Brinckmann, J. (2005). Collagens at a glance. In J. Brinckmann, H. Notbohm, & P. K. Muller (Eds.), *Collagen: Primer in Structure, Processing, and Assembly* (Vol. 247, pp. 1–6). New York, NY: Springer Berlin Heidelberg.
- Brinckmann, J., Neess, C. M., Gaber, Y., Sobhi, H., Notbohm, H., Hunzelmann, N., ... Scharffetter-Kochanek, K. (2001). Different pattern of collagen cross-links in two sclerotic skin diseases: lipodermatosclerosis and circumscribed scleroderma. *Journal of Investigative Dermatology*, 117(2), 269–273. doi:10.1046/j.0022-202x.2001.01414.x
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2), 187–217.
doi:10.1002/jcc.540040211
- Brown, E. M. (2009). The collagen microfibril as a tool for leather scientists. *International Union of Leather Technologists and Chemists Societies*.
- Brunner, R., Dalke, A., GURSOY, A., HUMPHREY, W., & NELSON, M. (2008). NAMD User Guide,

Back Matter: References

- Version 1.5. Retrieved from <http://www.ks.uiuc.edu/Research/namd/1.5/ug/>
- Buehler, M. J. (2008). Nanomechanics of collagen fibrils under varying cross-link densities: atomistic and continuum studies. *Journal of the Mechanical Behavior of Biomedical Materials*, *1*(1), 59–67. doi:10.1016/j.jmbbm.2007.04.001
- Chapman, J. A., & Hulmes, D. J. S. (1984). Electron microscopy of the collagen fibril. In A. Ruggeri & P. M. Motta (Eds.), *Ultrastructure of the Connective Tissue Matrix* (Vol. 3, pp. 1–33). Boston, MA: Martinus Nijhoff Publishers.
- Chen, H. (2006). Ehlers-Danlos Syndrome. In *Atlas of Genetic Diagnosis and Counseling* (pp. 342–346). Totowa, NJ: Humana Press.
- Chen, J. M., Kung, C. E., Fairheller, S. H., & Brown, E. M. (1991). An energetic evaluation of a “Smith” collagen microfibril model. *Journal of Protein Chemistry*, *10*(5), 535–552. doi:10.1007/BF01025482
- Chiang, C. S., Anishkin, A., & Sukharev, S. (2004). Gating of the large mechanosensitive channel in situ: estimation of the spatial scale of the transition from channel population responses. *Biophysical Journal*, *86*(5), 2846–2861. doi:10.1016/S0006-3495(04)74337-4
- Christiansen, D. L., Huang, E. K., & Silver, F. H. (2000). Assembly of type I collagen: fusion of fibril subunits and the influence of fibril diameter on mechanical properties. *Matrix Biology*, *19*(5), 409–420. doi:10.1016/S0945-053X(00)00089-5
- Cisneros, D. A., Hung, C., Franz, C. M., & Muller, D. J. (2006). Observing growth steps of collagen self-assembly by time-lapse high-resolution atomic force microscopy. *Journal of Structural Biology*, *154*(3), 232–245. doi:10.1016/j.jsb.2006.02.006
- Curwin, S. L. (2005). Rehabilitation after tendon injuries. In N. Maffulli, P. Renström, & W. B. Leadbetter (Eds.), *Tendon Injuries: Basic Science and Clinical Medicine* (pp. 242–266). London, UK: Springer London.
- Cusack, S., & Miller, A. (1979). Determination of the elastic constants of collagen by Brillouin light scattering. *Journal of Molecular Biology*, *135*(1), 39–51. doi:10.1016/0022-2836(79)90339-5
- Eastoe, J. E. (1955). The amino acid composition of mammalian collagen and gelatin. *Biochemical Journal*, *61*(4), 589–600.
- Engel, J., & Bächinger, H. P. (2005). Structure, stability and folding of the collagen triple helix. In J. Brinckmann, H. Notbohm, & P. K. Muller (Eds.), *Collagen: Primer in Structure*,

Back Matter: References

- Processing, and Assembly* (Vol. 247, pp. 7–33). New York, NY: Springer Berlin Heidelberg.
- Eppell, S. J., Smith, B. N., Kahn, H., & Ballarini, R. (2006). Nano measurements with micro-devices: mechanical properties of hydrated collagen fibrils. *Journal of the Royal Society, Interface*, 3(6), 117–121. doi:10.1098/rsif.2005.0100
- Exposito, J.-Y. (2004). Collagens. In J. L. Atwood & J. W. Steed (Eds.), *Encyclopedia of Supramolecular Chemistry* (Vol. 1, pp. 295–301). New York, NY: Marcel Dekker.
- Eyre, D. R., Koob, T. J., & Van Ness, K. P. (1984). Quantitation of hydroxypyridinium crosslinks in collagen by high-performance liquid chromatography. *Analytical Biochemistry*, 137(2), 380–388. doi:10.1016/0003-2697(84)90101-5
- Eyre, D. R., & Oguchi, H. (1980). The hydroxypyridinolinium cross-links of skeletal collagen: their measurement, properties and proposed pathway of formation. *Biochemical and Biophysical Research Communications*, 92(2), 403–410. doi:10.1016/0006-291X(80)90347-2
- Eyre, D. R., Paz, M. A., & Gallop, P. M. (1984). Cross-linking in collagen and elastin. *Annual Review of Biochemistry*, 53, 717–748. doi:10.1146/annurev.bi.53.070184.003441
- Eyre, D. R., & Wu, J.-J. (2005). Collagen cross-links. In J. Brinckmann, H. Notbohm, & P. K. Muller (Eds.), *Collagen: Primer in Structure, Processing, and Assembly* (Vol. 247, pp. 207–229). New York, NY: Springer Berlin Heidelberg.
- Ezura, Y., Chakravarti, S., Oldberg, A., Chervoneva, I., & Birk, D. E. (2000). Differential expression of lumican and fibromodulin regulate collagen fibrillogenesis in developing mouse tendons. *Journal of Cell Biology*, 151(4), 779–788. doi:10.1083/jcb.151.4.779
- Farber, S., Garg, A. K., Birk, D. E., & Silver, F. H. (1986). Collagen fibrillogenesis in vitro: evidence for pre-nucleation and nucleation steps. *International Journal of Biological Macromolecules*, 8(1), 37–42. doi:10.1016/0141-8130(86)90069-3
- Feig, M., & Brooks, C. L. (2004). Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Current Opinion in Structural Biology*, 14(2), 217–224. doi:10.1016/j.sbi.2004.03.009
- Fenna, D. (2002). *Dictionary of Weights, Measures, and Units*. New York, NY: Oxford University Press.
- Fietzek, P. P., & Kuhn, K. (1976). The primary structure of collagen. *International Review of*

Back Matter: References

- Connective Tissue Research*, 7, 1–60.
- Fraser, R. D., MacRae, T. P., & Miller, A. (1987). Molecular packing in type I collagen fibrils. *Journal of Molecular Biology*, 193(1), 115–125. doi:10.1016/0022-2836(87)90631-0
- Fraser, R. D., MacRae, T. P., Miller, A., & Suzuki, E. (1983). Molecular conformation and packing in collagen fibrils. *Journal of Molecular Biology*, 167(2), 497–521. doi:10.1016/S0022-2836(83)80347-7
- Fratzl, P., Misof, K., Zizak, I., Rapp, G., Amenitsch, H., & Bernstorff, S. (1998). Fibrillar structure and mechanical properties of collagen. *Journal of Structural Biology*, 122(1-2), 119–122. doi:10.1006/jsbi.1998.3966
- Freeman, J. W., & Silver, F. H. (2004). Elastic energy storage in unmineralized and mineralized extracellular matrices (ECMs): a comparison between molecular modeling and experimental measurements. *Journal of Theoretical Biology*, 229(3), 371–381. doi:10.1016/j.jtbi.2004.04.018
- Fujii, K., Yamagishi, T., Nagafuchi, T., Tsuji, M., & Kuboki, Y. (1994). Biochemical properties of collagen from ligaments and periarticular tendons of the human knee. *Knee Surgery, Sports Traumatology, Arthroscopy*, 2(4), 229–233. doi:10.1007/BF01845593
- Gaill, F., Wiedemann, H., Mann, K., Kühn, K., Timpl, R., & Engel, J. (1991). Molecular characterization of cuticle and interstitial collagens from worms collected at deep sea hydrothermal vents. *Journal of Molecular Biology*, 221(1), 209–223. doi:10.1016/0022-2836(91)80215-G
- Gautieri, A., Buehler, M. J., & Redaelli, A. (2009). Deformation rate controls elasticity and unfolding pathway of single tropocollagen molecules. *Journal of the Mechanical Behavior of Biomedical Materials*, 2(2), 130–137. doi:10.1016/j.jmbbm.2008.03.001
- Gautieri, A., Vesentini, S., Redaelli, A., & Buehler, M. J. (2011). Hierarchical structure and nanomechanics of collagen microfibrils from the atomistic scale up. *Nano Letters*, 11(2), 757–766. doi:10.1021/nl103943u
- Gautieri, A., Vesentini, S., Redaelli, A., & Buehler, M. J. (2012). Viscoelastic properties of model segments of collagen molecules. *Matrix Biology*, 31(2), 141–149. doi:10.1016/j.matbio.2011.11.005
- Goh, K. L., Meakin, J. R., Aspden, R. M., & Hukins, D. W. (2005). Influence of fibril taper on the function of collagen to reinforce extracellular matrix. *Proceedings of the Royal*

Back Matter: References

- Society B: Biological Sciences*, 272(1575), 1979–1983. doi:10.1098/rspb.2005.3173
- Gupta, H. S. (2008). Nanoscale deformation mechanisms in collagen. In P. Fratzl (Ed.), *Collagen: Structure and Mechanics* (pp. 155–173). New York, NY: Springer US.
- Hambli, R., & Barkaoui, A. (2012). Physically based 3D finite element model of a single mineralized collagen microfibril. *Journal of Theoretical Biology*, 301, 28–41. doi:10.1016/j.jtbi.2012.02.007
- Hang, F., & Barber, A. H. (2011). Nano-mechanical properties of individual mineralized collagen fibrils from bone tissue. *Journal of the Royal Society, Interface*, 8(57), 500–505. doi:10.1098/rsif.2010.0413
- Hanson, D. A., & Eyre, D. R. (1996). Molecular site specificity of pyridinoline and pyrrole cross-links in type I collagen of human bone. *Journal of Biological Chemistry*, 271(43), 26508–26516. doi:10.1074/jbc.271.43.26508
- Harkness, M. L., Harkness, R. D., & James, D. W. (1958). The effect of a protein-free diet on the collagen content of mice. *The Journal of Physiology*, 144(2), 307–313.
- Harley, R., James, D., Miller, A., & White, J. W. (1977). Phonons and the elastic moduli of collagen and muscle. *Nature*, 267(5608), 285–287. doi:10.1038/267285a0
- Hassan, S. A., Gracia, L., Vasudevan, G., & Steinbach, P. J. (2005). Computer simulation of protein-ligand interactions: challenges and applications. In G. U. Nienhaus (Ed.), *Protein-Ligand Interactions: Methods and Applications* (Vol. 305, pp. 451–492). New York, NY: Humana Press.
- Haut, R. C. (2002). Biomechanics of soft tissue. In A. M. Nahum & J. W. Melvin (Eds.), *Accidental Injury: Biomechanics and Prevention* (pp. 228–253). New York, NY: Springer New York.
- Hulmes, D. J., & Miller, A. (1979). Quasi-hexagonal molecular packing in collagen fibrils. *Nature*, 282(5741), 878–880. doi:10.1038/282878a0
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33–38. doi:10.1016/0263-7855(96)00018-5
- Ishikawa, O., Kondo, A., & Miyachi, Y. (1998). Mature type of skin collagen crosslink, histidinohydroxylysinoxorleucine, is significantly increased in the skin of systemic sclerosis patients. *Arthritis and Rheumatism*, 41(2), 376–377. doi:10.1002/1529-0131(199802)41:2<376::AID-ART25>3.0.CO;2-J

Back Matter: References

- Isralewitz, B., Gao, M., & Schulten, K. (2001). Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology*, *11*(2), 224–230. doi:10.1016/S0959-440X(00)00194-9
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, *79*(2), 926–935. doi:10.1063/1.445869
- Kadler, K. E., Hill, A., & Canty-Laird, E. G. (2008). Collagen fibrillogenesis: fibronectin, integrins, and minor collagens as organizers and nucleators. *Current Opinion in Cell Biology*, *20*(5), 495–501. doi:10.1016/j.ceb.2008.06.008
- Kadler, K. E., Holmes, D. F., Trotter, J. A., & Chapman, J. A. (1996). Collagen fibril formation. *Biochemical Journal*, *316* (Pt 1), 1–11.
- Kagan, H. M., & Ryvkin, F. (2011). Lysyl oxidase and lysyl oxidase-like enzymes. In R. P. Mecham (Ed.), *The Extracellular Matrix: An Overview* (pp. 303–306). New York, NY: Springer.
- Kale, L. V., Bhatle, A., Bohm, E. J., & Phillips, J. C. (2011). NANoscale Molecular Dynamics (NAMD). In D. Padua (Ed.), *Encyclopedia of Parallel Computing* (pp. 1249–1254). New York, NY: Springer US.
- Kato, Y. P., Christiansen, D. L., Hahn, R. A., Shieh, S.-J., Goldstein, J. D., & Silver, F. H. (1989). Mechanical properties of collagen fibres: a comparison of reconstituted and rat tail tendon fibres. *Biomaterials*, *10*(1), 38–42. doi:10.1016/0142-9612(89)90007-0
- Khoshnoodi, J., Cartailier, J. P., Alvares, K., Veis, A., & Hudson, B. G. (2006). Molecular recognition in the assembly of collagens: terminal noncollagenous domains are key recognition modules in the formation of triple helical protomers. *Journal of Biological Chemistry*, *281*(50), 38117–38121. doi:10.1074/jbc.R600025200
- Kielty, C. M., & Grant, M. E. (2002). The collagen family: Structure, assembly, and organization in the extracellular matrix. In P. M. Royce & B. Steinmann (Eds.), *Connective Tissue and Its Heritable Disorders: Molecular, Genetic and Medical Aspects* (pp. 183–185). New York, NY: Wiley-Liss.
- King, G., Brown, E. M., & Chen, J. M. (1996). Computer model of a bovine type I collagen microfibril. *Protein Engineering*, *9*(1), 43–49. doi:10.1093/protein/9.1.43
- Knott, L., & Bailey, A. J. (1998). Collagen cross-links in mineralizing tissues: A review of their

Back Matter: References

- chemistry, function, and clinical relevance. *Bone*, 22(3), 181–187. doi:10.1016/S8756-3282(97)00279-2
- Knott, L., Tarlton, J. F., & Bailey, A. J. (1997). Chemistry of collagen cross-linking: biochemical changes in collagen during the partial mineralization of turkey leg tendon. *Biochemical Journal*, 322 (Pt 2), 535–542.
- Kurz, E. M. (1991). Mini-collagens in hydra nematocytes. *Journal of Cell Biology*, 115(4), 1159–1169. doi:10.1083/jcb.115.4.1159
- Kuypers, R., Tyler, M., Kurth, L. B., Jenkins, I. D., & Horgan, D. J. (1992). Identification of the loci of the collagen-associated Ehrlich chromogen in type I collagen confirms its role as a trivalent cross-link. *Biochemical Journal*, 283 (Pt 1), 129–136.
- Kwansa, A. L., & Freeman, J. W. (2010). Elastic energy storage in an unmineralized collagen type I molecular model with explicit solvation and water infiltration. *Journal of Theoretical Biology*, 262(4), 691–697. doi:10.1016/j.jtbi.2009.10.024
- Leikin, S., Rau, D. C., & Parsegian, V. A. (1995). Temperature-favored assembly of collagen is driven by hydrophilic not hydrophobic interactions. *Nature Structural Biology*, 2(3), 205–210. doi:10.1038/nsb0395-205
- Lorenzo, A. C., & Caffarena, E. R. (2005). Elastic properties, Young's modulus determination and structural stability of the tropocollagen molecule: a computational study by steered molecular dynamics. *Journal of Biomechanics*, 38(7), 1527–1533. doi:10.1016/j.jbiomech.2004.07.011
- Lu, H., Krammer, A., Isralewitz, B., Vogel, V., & Schulten, K. (2000). Computer modeling of force-induced titin domain unfolding. In J. Pollack & H. Granzier (Eds.), *Elastic Filaments of the Cell* (pp. 143–162). New York, NY: Kluwer Academic/Plenum Publishers.
- Mackerell, A. D. (2004). Empirical force fields for biological macromolecules: overview and issues. *Journal of Computational Chemistry*, 25(13), 1584–1604. doi:10.1002/jcc.20082
- MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., ... Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*, 102(18), 3586–3616. doi:10.1021/jp973084f
- Mahoney, M. W., & Jorgensen, W. L. (2000). A five-site model for liquid water and the

Back Matter: References

- reproduction of the density anomaly by rigid, nonpolarizable potential functions. *Journal of Chemical Physics*, 112(20), 8910–8922. doi:10.1063/1.481505
- Mark, P., & Nilsson, L. (2001). Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *Journal of Physical Chemistry A*, 105(43), 9954–9960. doi:10.1021/Jp003020w
- Marrink, S. J., Fuhrmans, M., Risselada, H. J., & Periole, X. (2009). The MARTINI force field. In G. A. Voth (Ed.), *Coarse-Graining of Condensed Phase and Biomolecular Systems* (pp. 5–6). Boca Raton, FL: Taylor & Francis Group.
- Marumo, K., Saito, M., Yamagishi, T., & Fujii, K. (2005). The “ligamentization” process in human anterior cruciate ligament reconstruction with autogenous patellar and hamstring tendons: a biochemical study. *American Journal of Sports Medicine*, 33(8), 1166–1173. doi:10.1177/0363546504271973
- Mechanic, G. L., Katz, E. P., Henmi, M., Noyes, C., & Yamauchi, M. (1987). Locus of a histidine-based, stable trifunctional, helix to helix collagen cross-link: stereospecific collagen structure of type I skin fibrils. *Biochemistry*, 26(12), 3500–3509. doi:10.1021/bi00386a038
- Mitchell, M. (2002). *Engauge Digitizer*. Retrieved from <http://digitizer.sourceforge.net/>
- Miyazaki, H., & Hayashi, K. (1999). Tensile tests of collagen fibers obtained from the rabbit patellar tendon. *Biomedical Microdevices*, 2(2), 151–157. doi:10.1023/A:1009953805658
- Moran, D. T., & Rowley, J. C. (2011). Chapter 3: Connective tissue. In *Visual Histology Text-Atlas Book* (p. Plate 3–2). Niwot, CO: VisualHistology.com. Retrieved from http://www.visualhistology.com/products/atlas/VHA_Chpt3_Connective_Tissue.html
- Myllyharju, J. (2005). Intracellular post-translational modifications of collagens. In J. Brinckmann, H. Notbohm, & P. K. Muller (Eds.), *Collagen: Primer in Structure, Processing, and Assembly* (Vol. 247, pp. 115–142). New York, NY: Springer Berlin Heidelberg.
- Nemetschek, T., Jonak, R., Nemetschek-Gansler, H., Riedl, H., & Rosenbaum, G. (1978). On the determination of changes in the large periodic structure of collagen. *Zeitschrift für Naturforschung C*, 33(11-12), 928–936.
- Onufriev, A., Bashford, D., & Case, D. A. (2000). Modification of the generalized Born model suitable for macromolecules. *Journal of Physical Chemistry B*, 104(15), 3712–3720.

Back Matter: References

- doi:10.1021/jp994072s
- Onufriev, A., Bashford, D., & Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins*, 55(2), 383–394. doi:10.1002/prot.20033
- Orgel, J. P. R. O., Irving, T. C., Miller, A., & Wess, T. J. (2006). Microfibrillar structure of type I collagen in situ. *Proceedings of the National Academy of Sciences of the United States of America*, 103(24), 9001–9005. doi:10.1073/pnas.0502718103
- Orgel, J. P. R. O., Miller, A., Irving, T. C., Fischetti, R. F., Hammersley, A. P., & Wess, T. J. (2001). The in situ supermolecular structure of type I collagen. *Structure*, 9(11), 1061–1069. doi:10.1016/S0969-2126(01)00669-4
- Orphanet Report Series. (2010). *Prevalence of rare diseases: Bibliographic data, Rare Diseases collection, Number 1: Listed in alphabetical order of diseases* (pp. 1–29). Retrieved from http://www.orpha.net/orphacom/cahiers/docs/GB/Prevalence_of_rare_diseases_by_alphabetical_list.pdf
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., ... Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16), 1781–1802. doi:10.1002/jcc.20289
- Pins, G. D., & Silver, F. H. (1995). A self-assembled collagen scaffold suitable for use in soft and hard tissue replacement. *Materials Science and Engineering: C*, 3(2), 101–107. doi:10.1016/0928-4931(95)00109-3
- Risteli, J., & Risteli, L. (2006). Products of bone collagen metabolism. In M. J. Seibel, S. P. Robins, & J. P. Bilezikian (Eds.), *Dynamics of Bone and Cartilage Metabolism* (pp. 391–394). Burlington, MA: Elsevier.
- Robins, S. P., & Duncan, A. (1983). Cross-linking of collagen. Location of pyridinoline in bovine articular cartilage at two sites of the molecule. *Biochemical Journal*, 215(1), 175–182.
- Robins, Simon P., & Brady, J. D. (2002). Collagen cross-linking and metabolism. In J. P. Bilezikian, L. G. Raisz, & G. A. Rodan (Eds.), *Principles of Bone Biology* (2nd ed., Vol. 1, pp. 212–214). San Diego, CA: Academic Press.
- Rongioletti, F. (2003). Dermal infiltrates. In F. A. Kerdel & F. Jimenez-Acosta (Eds.), *Dermatology: Just the Facts* (p. 182). New York, NY: McGraw-Hill Companies.
- Saito, M., & Marumo, K. (2010). Collagen cross-links as a determinant of bone quality: a

Back Matter: References

- possible explanation for bone fragility in aging, osteoporosis, and diabetes mellitus. *Osteoporosis International*, 21(2), 195–214. doi:10.1007/s00198-009-1066-z
- Saito, M., Marumo, K., Fujii, K., & Ishioka, N. (1997). Single-column high-performance liquid chromatographic fluorescence detection of immature, mature, and senescent cross-links of collagen. *Analytical Biochemistry*, 253(1), 26–32. doi:10.1006/abio.1997.2350
- SAS Institute, Inc. (2010). *JMP*. Cary, NC: SAS Institute, Inc.
- Sasaki, N., & Odajima, S. (1996a). Stress-strain curve and young's modulus of a collagen molecule as determined by the X-ray diffraction technique. *Journal of Biomechanics*, 29(5), 655–658. doi:10.1016/0021-9290(95)00110-7
- Sasaki, N., & Odajima, S. (1996b). Elongation mechanism of collagen fibrils and force-strain relations of tendon at each level of structural hierarchy. *Journal of Biomechanics*, 29(9), 1131–1136. doi:10.1016/0021-9290(96)00024-3
- Shen, Z. L., Dodge, M. R., Kahn, H., Ballarini, R., & Eppell, S. J. (2008). Stress-strain experiments on individual collagen fibrils. *Biophysical Journal*, 95(8), 3956–3963. doi:10.1529/biophysj.107.124602
- Shen, Z. L., Dodge, M. R., Kahn, H., Ballarini, R., & Eppell, S. J. (2010). In vitro fracture testing of submicron diameter collagen fibril specimens. *Biophysical Journal*, 99(6), 1986–1995. doi:10.1016/j.bpj.2010.07.021
- Siegel, R. C. (1974). Biosynthesis of collagen crosslinks: increased activity of purified lysyl oxidase with reconstituted collagen fibrils. *Proceedings of the National Academy of Sciences of the United States of America*, 71(12), 4826–4830.
- Sims, T. J., & Bailey, A. J. (1992). Quantitative analysis of collagen and elastin cross-links using a single-column system. *Journal of Chromatography*, 582(1-2), 49–55. doi:10.1016/0378-4347(92)80301-6
- Smith, J. W. (1968). Molecular pattern in native collagen. *Nature*, 219(5150), 157–158. doi:10.1038/219157a0
- Stone, J. E. (1998). *An efficient library for parallel ray tracing and animation* (M.S. Thesis). University of Missouri-Rolla, Rolla, MO. Retrieved from <http://hdl.handle.net/10355/22017>
- Sun, Y.-L., Luo, Z.-P., Fertala, A., & An, K.-N. (2002). Direct quantification of the flexibility of type I collagen monomer. *Biochemical and Biophysical Research Communications*,

Back Matter: References

- 295(2), 382–386. doi:10.1016/S0006-291X(02)00685-X
- Sun, Y.-L., Luo, Z.-P., Fertala, A., & An, K.-N. (2004). Stretching type II collagen with optical tweezers. *Journal of Biomechanics*, 37(11), 1665–1669.
doi:10.1016/j.jbiomech.2004.02.028
- Svensson, R. B., Hassenkam, T., Grant, C. A., & Magnusson, S. P. (2010). Tensile properties of human collagen fibrils and fascicles are insensitive to environmental salts. *Biophysical Journal*, 99(12), 4020–4027. doi:10.1016/j.bpj.2010.11.018
- Sweeney, S. M., Orgel, J. P., Fertala, A., McAuliffe, J. D., Turner, K. R., Di Lullo, G. A., ... Antonio, J. D. S. (2008). Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *Journal of Biological Chemistry*, 283(30), 21187–21197. doi:10.1074/jbc.M709319200
- Tamura, M., & Ishikawa, O. (2001). An increase of mature type skin collagen cross-link, histidinohydroxylysinonorleucine, in the sclerotic skin of morphea. *Journal of Dermatological Science*, 25(1), 83–86. doi:10.1016/S0923-1811(00)00112-2
- Tarini, M., Cignoni, P., & Montani, C. (2006). Ambient Occlusion and Edge Cueing for Enhancing Real Time Molecular Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 1237–1244. doi:10.1109/TVCG.2006.115
- The MathWorks, Inc. (1984). *MATLAB*. Natick, MA: The MathWorks, Inc.
- The UniProt Consortium. (2012). Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research*, 40(D1), D71–D75. doi:10.1093/nar/gkr981
- Theoretical and Computational Biophysics Group. (2012). VMD User's Guide, Version 1.9.1. Retrieved from <http://www.ks.uiuc.edu/Research/vmd/current/ug/ug.html>
- Uitto, J., Bauer, E. A., & Eisen, A. Z. (1979). Scleroderma: increased biosynthesis of triple-helical type I and type III procollagens associated with unaltered expression of collagenase by skin fibroblasts in culture. *Journal of Clinical Investigation*, 64(4), 921–930. doi:10.1172/JCI109558
- Urich, K. (1994). Extracellular structural and secretory proteins. In P. J. King (Trans.), *Comparative Animal Biochemistry* (pp. 378–383). New York, NY: Springer-Verlag.
- Uzel, S. G., & Buehler, M. J. (2011). Molecular structure, mechanical behavior and failure mechanism of the C-terminal cross-link domain in type I collagen. *Journal of the Mechanical Behavior of Biomedical Materials*, 4(2), 153–161.

Back Matter: References

- doi:10.1016/j.jmbbm.2010.07.003
- Van der Rijt, J. A. J. (2004). *Micromechanical testing of single collagen type I fibrils* (Ph.D. Dissertation). University of Twente, Enschede, The Netherlands. Retrieved from <http://purl.utwente.nl/publications/41721>
- Van der Rijt, J. A. J., van der Werf, K. O., Bennink, M. L., Dijkstra, P. J., & Feijen, J. (2006). Micromechanical testing of individual collagen fibrils. *Macromolecular Bioscience*, 6(9), 697–702. doi:10.1002/mabi.200600063
- Van der Slot, A. J., Zuurmond, A. M., Bardeol, A. F., Wijmenga, C., Pruijs, H. E., Sillence, D. O., ... Bank, R. A. (2003). Identification of PLOD2 as telopeptide lysyl hydroxylase, an important enzyme in fibrosis. *Journal of Biological Chemistry*, 278(42), 40967–40972. doi:10.1074/jbc.M307380200
- Von der Mark, K. (2006). Structure, biosynthesis and gene regulation of collagens in cartilage and bone. In Markus J. Seibel, S. P. Robins, & J. P. Bilezikian (Eds.), *Dynamics of Bone and Cartilage Metabolism. Principles and Clinical Applications* (2nd ed., pp. 3–40). Burlington, MA: Academic Press/Elsevier.
- Walton, E. B., & Van Vliet, K. J. (2006). Equilibration of experimentally determined protein structures for molecular dynamics simulation. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 74(6 Pt 1), 061901–1 to 061901–8. doi:10.1103/PhysRevE.74.061901
- Weiss, J. B., & Jayson, M. I. V. (1982). *Collagen in Health and Disease*. New York, NY: Longman Group Limited.
- Wess, T. J., Hammersley, A. P., Wess, L., & Miller, A. (1998). A consensus model for molecular packing of type I collagen. *Journal of Structural Biology*, 122(1-2), 92–100. doi:10.1006/jsbi.1998.3991
- Williams, A. P. (1978). The amino acid, collagen and mineral composition of preruminant calves. *Journal of Agricultural Science*, 90(03), 617–624. doi:10.1017/S0021859600056161
- Wynbrandt, J., & Ludman, M. D. (2008). *The Encyclopedia of Genetic Disorders and Birth Defects* (3rd ed.). New York, NY: Facts On File, Inc.
- Xie, X.-Q. (2008). Molecular modeling and in silico drug design. In T. L. Lemke, D. A. Williams, V. F. Roche, & S. W. Zito (Eds.), *Foye's Principles of Medicinal Chemistry* (6th ed., pp. 54–61). Baltimore, MD: Lippincott Williams & Wilkins.

Back Matter: References

- Yamauchi, M. (1996). Collagen: The major matrix molecule in mineralized tissues. In J. J. B. Anderson & S. C. Garner (Eds.), *Calcium and Phosphorus in Health and Disease* (pp. 134–137). Boca Raton, FL: CRC Press.
- Yamauchi, M., Chandler, G. S., Tanzawa, H., & Katz, E. P. (1996). Cross-linking and the molecular packing of corneal collagen. *Biochemical and Biophysical Research Communications*, 219(2), 311–315. doi:10.1006/bbrc.1996.0229
- Yamauchi, M., London, R. E., Guenat, C., Hashimoto, F., & Mechanic, G. L. (1987). Structure and formation of a stable histidine-based trifunctional cross-link in skin collagen. *Journal of Biological Chemistry*, 262(24), 11428–11434.
- Yang, L., van der Werf, K. O., Dijkstra, P. J., Feijen, J., & Bennink, M. L. (2012). Micromechanical analysis of native and cross-linked collagen type I fibrils supports the existence of microfibrils. *Journal of the Mechanical Behavior of Biomedical Materials*, 6, 148–158. doi:10.1016/j.jmbbm.2011.11.008
- Zhao, T. L., Weinhold, P. S., Lee, N. Y., & Dahners, L. E. (2011). Some observations on the subfibrillar structure of collagen fibrils as noted during treatment with NKISK and cathepsin G with mechanical agitation. *Journal of Electron Microscopy*, 60(2), 177–182. doi:10.1093/jmicro/dfr005

Appendix A : CHARMM19 parameters added for deH-HLNL

Table A.1: Three new sets of bonds length parameters for the deH-HLNL crosslink. k_b is the force constant for the modeled covalent bond (k_b units: kcal/mol/Å²) and r_0 is the equilibrium length of the covalent bond (r_0 units: Å) (Brooks et al., 1983).

Atom Types	Parameters	Notes
NS1 CC1B	$k_b = 450.0$ $r_0 = 1.305$	Based on CHARMM19's "CR1E N*" pair $k_b = 450.0$; $r_0 = 1.305$ Compare to CHARMM22's "NS1 CC1B" pair $K_b = 500.000$; $b_0 = 1.276$
CH2E NS1	$k_b = 422.0$ $r_0 = 1.45$	Based on CHARMM19's "CH2E N*" pair, $k_b = 422.0$; $r_0 = 1.45$ Compare to CHARMM22's "CT2 NS1" pair $K_b = 310.000$; $b_0 = 1.440$
CC1B CH2E	$k_b = 250.0$ $r_0 = 1.45$	Based on CHARMM19's "CR1E CH2E" pair $k_b = 250.0$; $r_0 = 1.45$ Compare to CHARMM22's "CC1A CT2" pair $K_b = 365.000$; $b_0 = 1.502$

Table A.2: Four new sets of bond angle parameters for the deH-HLNL crosslink. k_θ is the force constant for the modeled bond angle (k_θ units: kcal/mol/radian²) and θ_0 is the equilibrium angle for the three covalently bonded atoms (θ_0 units: degrees) (Brooks et al., 1983).

Atom Types	Parameters	Notes
CC1B CH2E CH2E	$k_\theta = 70.0$ $\theta_0 = 113.0$	Based on CHARMM19's "C CH2E CH2E" group $k_\theta = 70.0$; $\theta_0 = 113.0$ Compare to CHARMM22's "CC1A CT2 CT2" group $K_\theta = 32.00$; $\theta_0 = 112.20$
CH1E CH2E NS1	$k_\theta = 65.0$ $\theta_0 = 110.5$	Based on CHARMM19's "CH1E CH2E NH3" group $k_\theta = 65.0$; $\theta_0 = 110.5$ Compare to CHARMM22's "CT2 CT2 NS1" group $K_\theta = 67.70$; $\theta_0 = 110.00$ which is identical to CHARMM22's "NH3 CT2 CT2" group $K_\theta = 67.700$; $\theta_0 = 110.0000$
CH2E NS1 CC1B	$k_\theta = 65.0$ $\theta_0 = 110.0$	Based on CHARMM19's "CR1E NR CR1E" group $k_\theta = 65.0$; $\theta_0 = 110.0$ Compare to CHARMM22's "CT2 NS1 CC1B" group $K_\theta = 67.00$; $\theta_0 = 111.00$
NS1 CC1B CH2E	$k_\theta = 90.0$ $\theta_0 = 122.0$	Based on CHARMM19's "C CR1E CH2E" group $k_\theta = 90.0$; $\theta_0 = 122.0$ Compare to CHARMM22's "NS1 CC1B CT3" group $K_\theta = 52.00$; $\theta_0 = 123.00$

Appendix A (continued)

Table A.3: Three new sets of bond torsion parameters for the deH-HLNL crosslink. k_ϕ is the force constant for the modeled bond torsion (k_ϕ units: kcal/mol) and n is a measure of periodicity of the bond torsion (n units: unitless) (Brooks et al., 1983). δ is a phase angle parameter for CHARMM22 (MacKerell et al., 1998).

Atom Types	Parameters	Notes
CH1E CH2E NS1 CC1B	$k_\phi = 0.3$ $n = 3$	Based on CHARMM19's "X CH2E N X" group $k_\phi = 0.3$; $n = 3$ Compare to CHARMM22's "CT2 CT2 NS1 CC1B" group $K_\chi = 0.1000$; $n = 3$; $\delta = 0.00$
CH2E NS1 CC1B CH2E	$k_\phi = 10.0$ $n = 2$	Based on CHARMM19's "X C CR1E X" group $k_\phi = 10.0$; $n = 2$ Compare to CHARMM22's "CT3 NS1 CC1B CT3" group $K_\chi = 12.0000$; $n = 2$; $\delta = 180.00$
NS1 CC1B CH2E CH2E	$k_\phi = 0.0$ $n = 3$	Based on CHARMM19's "X C CH2E X" group $k_\phi = 0.0$; $n = 3$ Compare to CHARMM22's "CC1A CC1A CT2 CT2" group $K_\chi = 0.3000$; $n = 3$; $\delta = 0.00$

Table A.4: Two new sets of Lennard-Jones parameters for the deH-HLNL crosslink. Starting from the case of two identical atom types interacting non-covalently, E_{\min} is the minimum of their Lennard-Jones energy (E_{\min} units: kcal/mol), R_{\min} is $\frac{1}{2}$ of their center-to-center inter-atomic distance at E_{\min} (R_{\min} units: Å), and α is a polarizability parameter (α units: Å³) (Brooks et al., 1983). The subscript 1,4 denotes a modification used for atoms separated by three consecutive covalent bonds. Combining rules are used to describe interactions between non-identical atom types.

Atom Type	Parameters	Notes
NS1	$\alpha = 1.1000$ $E_{\min} = -0.2384$ $R_{\min} = 1.6000$	Based on CHARMM19's "N*" atom type $\alpha = 1.1000$; $E_{\min} = -0.2384$; $R_{\min} = 1.6000$ Compare to CHARMM22's "NS1" atom type $\alpha = 0.000000$; $E_{\min} = -0.200000$; $R_{\min} = 1.850000$ Used CHARMM19's N* rather than CHARMM22's NS1, by analogy of CHARMM22's NS1 to its other nitrogens
CC1B	$\alpha = 1.35$ $E_{\min} = -0.1200$ $R_{\min} = 2.100$ $E_{\min 1,4} = -0.1$ $R_{\min 1,4} = 1.9$	Based on CHARMM19's "CR1E" atom type $\alpha = 1.35$; $E_{\min} = -0.1200$; $R_{\min} = 2.100$; $\alpha_{1,4} = 1.35$; $E_{\min 1,4} = -0.1$; $R_{\min 1,4} = 1.9$ Compare to CHARMM22's "CC1B" atom type $\alpha = 0.000000$; $E_{\min} = -0.0680$; $R_{\min} = 2.0900$

Appendix B : TCL scripts and NAMD configuration file

B.1. Model building steps for Aims 1 and 2

Table B.1: Positions for five molecular segments within a collagen type I unit cell. Here, u and v are fractional coordinates in a non-orthogonal coordinate system based on the dimensions of a collagen type I unit cell reported by Fraser *et al.* ($a = 39.97 \text{ \AA}$, $b = 26.95 \text{ \AA}$, $\gamma = 105.58^\circ$) (Fraser *et al.*, 1983). The specific ordering of the five segments at these positions was based upon Orgel *et al.* (Orgel *et al.*, 2001). The five segments were moved axially along the z-direction followed by lateral x-y movements; x and y coordinates were derived from u and v coordinates (Fraser *et al.*, 1983).

Segment	α	α -chain Length	$x = v*b - u*a*\sin(\gamma - 90)$	$y = -u*a*\sin(180 - \gamma)$	z
1	$\alpha 1$	Pro1 to Pro237	$v = 0.697$	$u = 0.627$	0
	$\alpha 2$	Pro1 to Pro238	$x = 12.1$	$y = -24.1$	
	$\alpha 1$	Pro1 to Gly239			
2	$\alpha 1$	Pro238 to Pro471	$v = 0.619$	$u = 0.262$	+670
	$\alpha 2$	Gly239 to Pro472	$x = 13.9$	$y = -10.1$	
	$\alpha 1$	Pro240 to Gly473			
3	$\alpha 1$	Pro472 to Pro705	$v = 0$	$u = 0$	+1340
	$\alpha 2$	Gly473 to Pro706	$x = 0$	$y = 0$	
	$\alpha 1$	Pro474 to Gly707			
4	$\alpha 1$	Pro706 to Pro939	$v = 0.162$	$u = 0.418$	+2010
	$\alpha 2$	Gly707 to Pro940	$x = -0.121$	$y = -16.1$	
	$\alpha 1$	Pro708 to Gly941			
5	$\alpha 1$	Pro940 to Pro1054	$v = 0.356$	$u = 0.813$	+2680
	$\alpha 2$	Gly941 to Pro1054	$x = 0.866$	$y = -31.3$	
	$\alpha 1$	Pro942 to Pro1054			

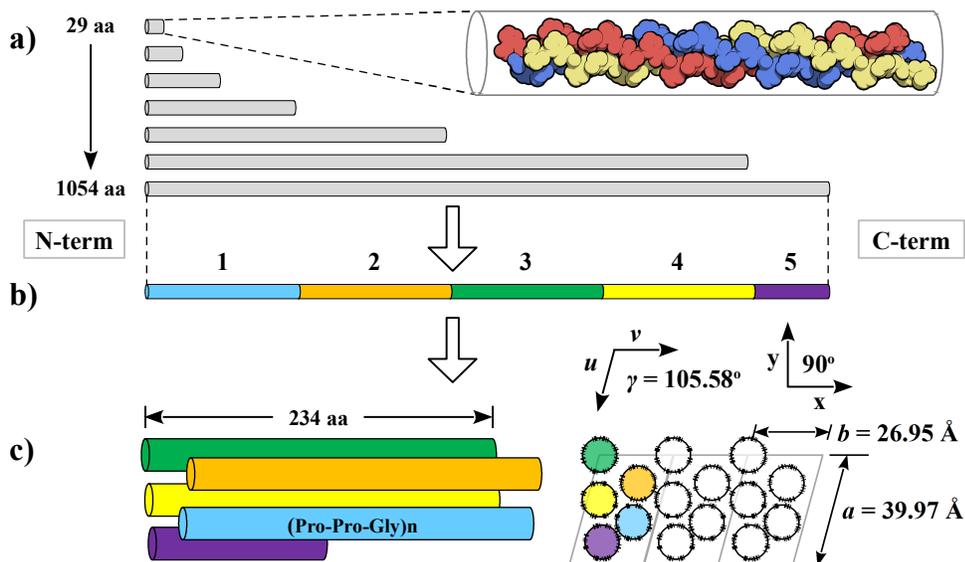


Figure B.1: A schematic of model building steps. (a) lengthening of a short collagen-like x-ray structure (1k6f.pdb) from a length of 29 amino acids (29 aa) to 1054 amino acids (1054 aa), (b) division into five molecular segments, and (c) axial and lateral translations of the five segments to form a microfibril unit with a length of 234 amino acids (234 aa) (Fraser *et al.*, 1983; Orgel *et al.*, 2001). The molecular rendering in (a) is of the 1k6f.pdb x-ray structure that was employed (Berisio *et al.*, 2002); this particular rendering was generated using the computer program QuteMol (Tarini, Cignoni, & Montani, 2006).

B.1 (continued) – Model building steps for Aims 1 and 2

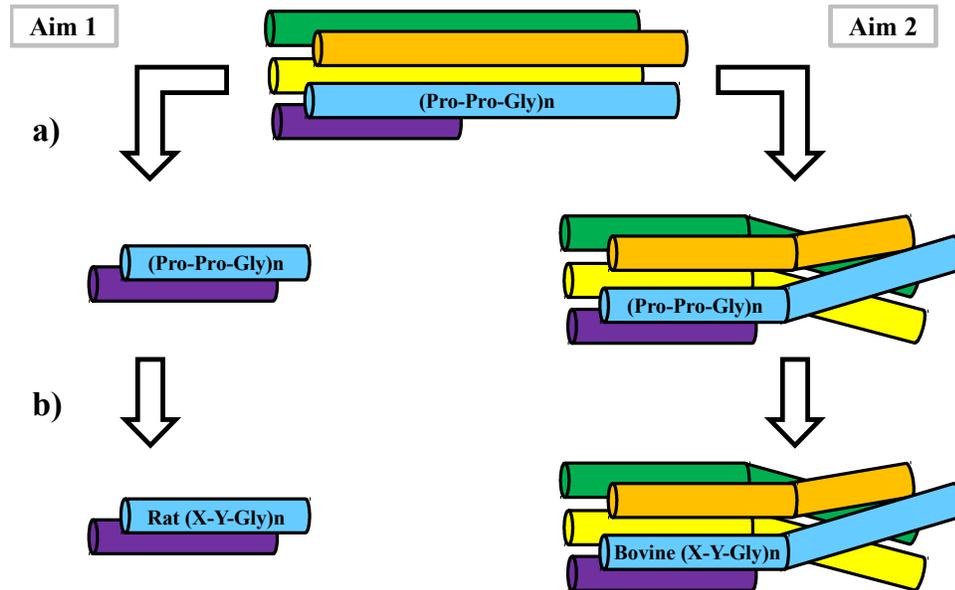


Figure B.2: A schematic of model building steps for Aims 1 and 2. (a) For Aim 1, segment 5 and a portion of segment 1 of comparable length were isolated, while for Aim 2, the segments were kinked/bent within the gap region to form a microfibril twist. (b) The $(\text{Pro-Pro-Gly})_n$ amino acid sequences were replaced with rat and bovine sequences for Aims 1 and 2, respectively, and N- and C-telopeptide conformations were incorporated based on an x-ray fiber diffraction structure for rat tail tendon collagen type I.

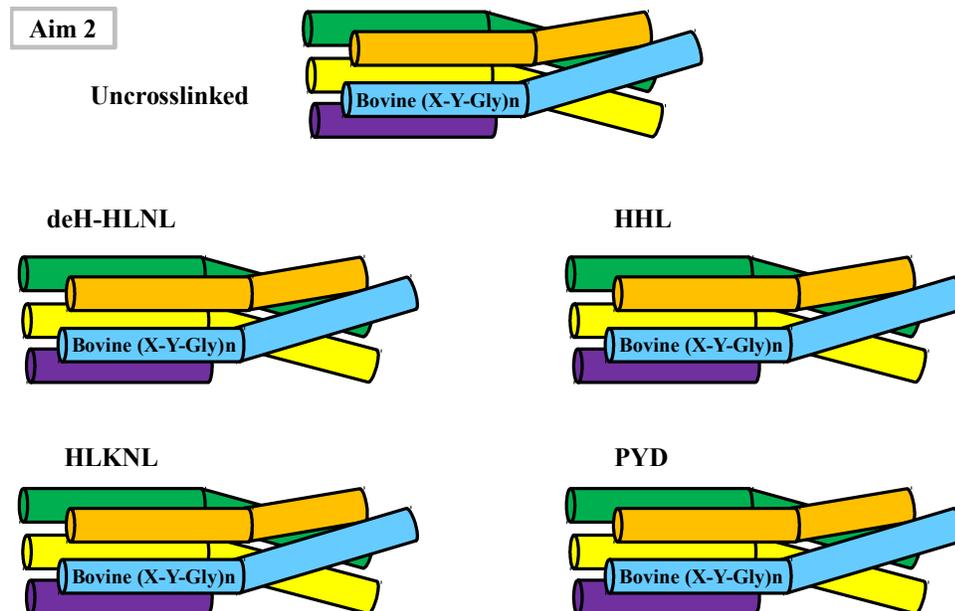


Figure B.3: A schematic of model building steps for Aim 2. The uncrosslinked microfibril unit model was replicated four times and used to derive four differently crosslinked microfibril unit models (deH-HLNL, HLKNL, HHL, and PYD), each containing two identical crosslinks.

B.2. TCL script used to apply molecular kinks within the gap region

```
set Seg1N [atomselect top "(segname P0) and (resid 114) or \  
(segname P5) and (resid 115) or (segname P10) and (resid 116)"]  
set Center1N [measure center $Seg1N]  
set Seg1C [atomselect top "(segname P0) and (resid 237) or \  
(segname P5) and (resid 238) or (segname P10) and (resid 239)"]  
set Center1C [measure center $Seg1C]  
set Seg2C [atomselect top "(segname P1) and (resid 471) or \  
(segname P6) and (resid 472) or (segname P11) and (resid 473)"]  
set Center2C [measure center $Seg2C]  
set G1 [atomselect top "(segname P0) and (resid 114 to 237) or \  
(segname P5) and (resid 115 to 238) or (segname P10) and (resid 116 to 239)"]
```

```
$G1 move [trans angle \  
{24.78777409 -3.468605638 -503.2715607} \  
{24.78777409 -3.468605638 -152.4042053} \  
{26.23713684 10.12509155 -503.2715607} -2.231270452 deg]
```

```
set Seg2C [atomselect top "(segname P1) and (resid 471) or \  
(segname P6) and (resid 472) or (segname P11) and (resid 473)"]  
set Center2C [measure center $Seg2C]  
set Seg2N [atomselect top "(segname P1) and (resid 348) or \  
(segname P6) and (resid 349) or (segname P11) and (resid 350)"]  
set Center2N [measure center $Seg2N]  
set Seg3C [atomselect top "(segname P2) and (resid 705) or \  
(segname P7) and (resid 706) or (segname P12) and (resid 707)"]  
set Center3C [measure center $Seg3C]  
set G2 [atomselect top "(segname P1) and (resid 348 to 471) or \  
(segname P6) and (resid 349 to 472) or (segname P11) and (resid 350 to 473)"]
```

```
$G2 move [trans angle \  
{25.49496269 10.02901983 -501.8819427} \  
{25.49496269 10.02901983 -151.0149536} \  
{11.13636494 20.13290787 -501.8819427} -2.864673456 deg]
```

B.2 (continued) – TCL script used to apply molecular kinks within the gap region

```
set Seg3C [atomselect top "(segname P2) and (resid 705) or \  
(segname P7) and (resid 706) or (segname P12) and (resid 707)"]  
set Center3C [measure center $Seg3C]  
set Seg3N [atomselect top "(segname P2) and (resid 582) or \  
(segname P7) and (resid 583) or (segname P12) and (resid 584)"]  
set Center3N [measure center $Seg3N]  
set Seg4C [atomselect top "(segname P3) and (resid 939) or \  
(segname P8) and (resid 940) or (segname P13) and (resid 941)"]  
set Center4C [measure center $Seg4C]  
set G3 [atomselect top "(segname P2) and (resid 582 to 705) or \  
(segname P7) and (resid 583 to 706) or (segname P12) and (resid 584 to 707)"]
```

```
$G3 move [trans angle \  
{10.51978827 20.63021851 -500.4872589} \  
{10.51978827 20.63021851 -149.6082764} \  
{10.0350008 4.893363953 -500.4872589} -2.569200312 deg]
```

```
set Seg4C [atomselect top "(segname P3) and (resid 939) or \  
(segname P8) and (resid 940) or (segname P13) and (resid 941)"]  
set Center4C [measure center $Seg4C]  
set Seg4N [atomselect top "(segname P3) and (resid 816) or \  
(segname P8) and (resid 817) or (segname P13) and (resid 818)"]  
set Center4N [measure center $Seg4N]  
set Seg5N [atomselect top "(segname P4) and (resid 940) or \  
(segname P9) and (resid 941) or (segname P14) and (resid 942)"]  
set Center5N [measure center $Seg5N]  
set G4 [atomselect top "(segname P3) and (resid 816 to 939) or \  
(segname P8) and (resid 817 to 940) or (segname P13) and (resid 818 to 941)"]
```

```
$G4 move [trans angle \  
{10.11136866 5.725576639 -499.784729} \  
{10.11136866 5.725576639 -148.1298523} \  
{13.02659988 -11.21712017 -499.784729} -2.798842388 deg]
```

Back Matter: Appendices

B.3. TCL script used to replace template Pro-Pro-Gly amino acid sequences

```
set protein [atomselect top protein]
set segments [lsort -unique [$protein get segname]]
foreach S $segments {set sel [atomselect top "segname $S"]}
$sel writpdb ${S}.pdb}

package require psfgen
topology CHARMmTop.inp
pdbalias atom ILE CD1 CD

segment P0 {pdb P0.pdb
mutate 1 GLN; mutate 2 LEU; mutate 3 SER; # ... continued for all amino acids
first NTER; last CTER}
coordpdb P0.pdb P0

# The above block of code was repeated for all 15 segments (P0 to P14)
# and for all amino acids (e.g., a total of 3,129 amino acids for the 15 segments).

guesscoord
writpdb mutated.pdb
writepsf mutated.psf
```

Back Matter: Appendices

B.4. TCL script used to select fixed atoms and assign pulled atom spring constants

```
set xlink_list {HHL HKNL HLNL PYD UNX}
foreach xlink $xlink_list {
set k_pull 9999.00;           # (kcal/mol/angstrom^2)

mol load pdb $xlink.coor psf $xlink.psf

# set PullNterm1 [atomselect top "((segname P0 and resid 1) or \
(segname P5 and resid 8) or (segname P10 and resid 1)) and name CA"]
set PullNterm2 [atomselect top "((segname P1 and resid 238) or \
(segname P6 and resid 239) or (segname P11 and resid 240)) and name CA"]
set PullNterm3 [atomselect top "((segname P2 and resid 472) or \
(segname P7 and resid 473) or (segname P12 and resid 474)) and name CA"]
set PullNterm4 [atomselect top "((segname P3 and resid 706) or \
(segname P8 and resid 707) or (segname P13 and resid 708)) and name CA"]
set PullNterm5 [atomselect top "((segname P4 and resid 940) or \
(segname P9 and resid 941) or (segname P14 and resid 942)) and name CA"]

set FixCterm1 [atomselect top "((segname P0 and resid 237) or \
(segname P5 and resid 238) or (segname P10 and resid 239)) and name CA"]
set FixCterm2 [atomselect top "((segname P1 and resid 471) or \
(segname P6 and resid 472) or (segname P11 and resid 473)) and name CA"]
set FixCterm3 [atomselect top "((segname P2 and resid 705) or \
(segname P7 and resid 706) or (segname P12 and resid 707)) and name CA"]
set FixCterm4 [atomselect top "((segname P3 and resid 939) or \
(segname P8 and resid 940) or (segname P13 and resid 941)) and name CA"]
# set FixCterm5 [atomselect top "((segname P4 and resid 1054) or \
(segname P9 and resid 1027) or (segname P14 and resid 1054)) and name CA"]

set all [atomselect top all]
$all set beta 0
$all set occupancy 0

$FixCterm1 set beta 1
$FixCterm2 set beta 1
$FixCterm3 set beta 1
$FixCterm4 set beta 1

$PullNterm2 set occupancy $k_pull
$PullNterm3 set occupancy $k_pull
$PullNterm4 set occupancy $k_pull
$PullNterm5 set occupancy $k_pull

$all writpdb FixedandPulledCA_ $xlink.pdb
}
```

B.5. NAMD configuration file

```
#####  
## SIMULATION PARAMETERS  
#####  
# Input  
set xlink PYD; set v 006_25ms; set r r1  
set input $xlink  
coordinates $input.coor # coordinate file  
velocities $input.vel # velocities file  
extendedSystem $input.xsc # extended system configuration file  
structure $input.psf # structure file  
parameters CHARMm.inp # parameter file  
paratypecharm on  
  
# Output  
outputname $xlink\_ $v\_ $r  
binaryoutput no # default = yes  
binaryrestart no # default = yes  
  
# Output frequency  
set freq 53600; # multiple of stepspercycle (timesteps)  
dcdFreq $freq  
outputEnergies $freq  
restartfreq $freq  
  
# Dielectric constant and 1-4 scaling parameters  
exclude scaled1-4  
1-4scaling 0.5 # default = 1.0  
dielectric 1.0 # default = 1.0  
  
# Switching, cut-off, and pair list distances  
switching on # default = on  
switchdist 15 # less than or equal to cutoff (angstroms)  
cutoff 16 # cut-off distance (angstroms)  
pairlistdist 17.5 # default = cutoff (angstroms)  
  
# Timestep parameters  
timestep 2 # default = 1 (fs)  
nonbondedFreq 1 # default = 1 (timesteps)  
fullElectFrequency 2 # default = stepspercycle (timesteps)  
stepspercycle 10 # default = 20 (timesteps)  
rigidbonds all # default = none; needed for 2 fs timestep  
useSettle on # default = on
```

B.5 (continued) – NAMD configuration file

```
#####  
## Generalized Born Implicit Solvent (GBIS)  
#####  
GBIS                on                # activate GBIS  
solventDielectric   80                # default: 78.5  
ionConcentration     0.150            # default: 0.2 (M or moles per liter)  
alphaCutoff         14                # default: 15 (angstroms)  
# GBISDelta         1.0                # default = 1.0; used for GB^OBC (II)  
# GBISBeta          0.8                # default = 0.8; used for GB^OBC (II)  
# GBISGamma         4.85              # default = 4.85; used for GB^OBC (II)  
  
#####  
## Langevin dynamics temperature control  
#####  
set temperature     293.15  
langevin            on                # activate langevin dynamics  
langevinDamping     5;                # damping coefficient (5/ps)  
langevinTemp        $temperature;    # temperature (K)  
langevinHydrogen    no;                # do not apply to hydrogen atoms  
  
#####  
## Fixed atoms  
#####  
fixedAtoms          on                # activate fixed atoms  
fixedAtomsFile      FixedandPulledCA_ $xlink.pdb  
fixedAtomsCol       B                # column of fixedAtomsFile  
  
#####  
## Harmonic moving constraints  
#####  
constraints          on                # activate harmonic constraints  
consRef              FixedandPulledCA_ $xlink.pdb  
consKFile            FixedandPulledCA_ $xlink.pdb  
consKCol             O                # column of consKFile with spring constants  
  
movingConstraints    on                # activate moving constraints  
movingConsVel        0 0 0.000125    # velocity vector (angstroms/timestep)  
  
#####  
## Script execution line  
#####  
run                  3752000;        # (timesteps)
```

B.6. TCL script used to measure initial lengths

```
set xlink_list {HHL HKNL HLNL PYD UNX}
foreach xlink $xlink_list {

mol load pdb $xlink.coor psf $xlink.psf

set pulled [atomselect top "\
(((segname P1 and resid 238) or (segname P6 and resid 239) or \
(segname P11 and resid 240)) or ((segname P2 and resid 472) or \
(segname P7 and resid 473) or (segname P12 and resid 474)) or \
((segname P3 and resid 706) or (segname P8 and resid 707) or \
(segname P13 and resid 708)) or ((segname P4 and resid 940) or \
(segname P9 and resid 941) or (segname P14 and resid 942))) and name CA"]

set fixed [atomselect top "\
(((segname P0 and resid 237) or (segname P5 and resid 238) or \
(segname P10 and resid 239)) or ((segname P1 and resid 471) or \
(segname P6 and resid 472) or (segname P11 and resid 473)) or \
((segname P2 and resid 705) or (segname P7 and resid 706) or \
(segname P12 and resid 707)) or ((segname P3 and resid 939) or \
(segname P8 and resid 940) or (segname P13 and resid 941))) and name CA"]

set rmsd_length [measure rmsd $pulled $fixed]
puts "RMSD-based initial length for $xlink model is $rmsd_length angstroms"
}
```

B.7. TCL script used to obtain energy values for the entire model from the log files

This script is based in part on the ft.tcl script that is described in the NAMD tutorials.
<http://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-win-html/node20.html>

```
set xlink_list {HHL HKNL HLNL PYD UNX UNX2}
set v_list {006_25ms 012_50ms 025_00ms 050_00ms 100_00ms}
set r_list {r1 r2 r3}
set sourcedir C:/VMD/LOG; set writedir C:/VMD/LOG

foreach xlink $xlink_list {foreach v $v_list {foreach r $r_list {
set inputfile $xlink\_ $v\_ $r
set outputfile $inputfile

if [catch {open $sourcedir/$inputfile.log r}] {
echo "The file $sourcedir/$ inputfile.log not found."
} else {

# Begin section based on ft.tcl
set file [open $sourcedir/$inputfile.log r]
set output [open $writedir/$outputfile.dat w]
puts $output \
"TS BOND ANGLE DIHED IMPRP \
ELECT VDW BOUNDARY MISC KINETIC \
TOTAL TEMP POTENTIAL TOTAL3 TEMPAVG"

while { [gets $file line] != -1 } {

if {[lindex $line 0] == "ENERGY:"} {
puts $output \
"[lindex $line 1] [lindex $line 2] [lindex $line 3] [lindex $line 4] [lindex $line 5] \
[lindex $line 6] [lindex $line 7] [lindex $line 8] [lindex $line 9] [lindex $line 10] \
[lindex $line 11] [lindex $line 12] [lindex $line 13] [lindex $line 14] [lindex $line 15]"
}}

close $file
close $output
# End section based on ft.tcl
}}}}
```

B.8. TCL script used to calculate energy values for the crosslinks

```
set xlink_list {HHL HKNL HLNL PYD}
set v_list {006_25ms 012_50ms 025_00ms 050_00ms 100_00ms}
set r_list {r1 r2 r3}
set m 0; set sourcedir C:/VMD; set writedir C:/VMD/ENERGY; set namddir C:/NAMD
set procedure {

if {"HHL" == $xlink} {
set sel [atomselect top "((segname P10 and resid 9) or (segname P14 and \
resid 946) or (segname P4 and resid 948)) and not name C N O H"]
namdenergy -all -par CHARMM.inp -sel $sel -ofile \
$writedir/$xlink\_N\_svr.dat -switch 15 -cutoff 16 -exe $namddir/ namd2.exe
set sel [atomselect top "((segname P0 and resid 103) or (segname P4 and \
resid 1046) or (segname P5 and resid 108)) and not name C N O H"]
namdenergy -all -par CHARMM.inp -sel $sel -ofile \
$writedir/$xlink\_C\_svr.dat -switch 15 -cutoff 16 -exe $namddir/namd2.exe

} elseif {"PYD" == $xlink} {
set sel [atomselect top "((segname P10 and resid 9) or (segname P14 and \
resid 946) or (segname P5 and resid 12)) and not name C N O H"]
namdenergy -all -par CHARMM.inp -sel $sel -ofile \
$writedir/$xlink\_N\_svr.dat -switch 15 -cutoff 16 -exe $namddir/namd2.exe
set sel [atomselect top "((segname P0 and resid 103) or (segname P4 and \
resid 1046) or (segname P14 and resid 1046)) and not name C N O H"]
namdenergy -all -par CHARMM.inp -sel $sel -ofile \
$writedir/$xlink\_C\_svr.dat -switch 15 -cutoff 16 -exe $namddir/namd2.exe

} else {
set sel [atomselect top "((segname P10 and resid 9) or (segname P14 and \
resid 946)) and not name C N O H"]
namdenergy -all -par CHARMM.inp -sel $sel -ofile \
$writedir/$xlink\_N\_svr.dat -switch 15 -cutoff 16 -exe $namddir/namd2.exe
set sel [atomselect top "((segname P0 and resid 103) or (segname P4 and \
resid 1046)) and not name C N O H"]
namdenergy -all -par CHARMM.inp -sel $sel -ofile \
$writedir/$xlink\_C\_svr.dat -switch 15 -cutoff 16 -exe $namddir/namd2.exe}}

foreach xlink $xlink_list {
mol load pdb $sourcedir/COOR/$xlink.coor psf $sourcedir/PSF/$xlink.psf
set vr initial; eval $procedure; incr m

foreach v $v_list {foreach r $r_list {
mol load dcd $sourcedir/DCD/$xlink\_v\_r.dcd psf $sourcedir/PSF/$xlink.psf
set vr $v\_r; eval $procedure; incr m}}
mol delete all}
```

B.9. TCL script used to obtain Cartesian coordinates for the crosslink Ca atoms

```
set xlink_list {HHL HKNL HLNL PYD}
set v_list {006_25ms 012_50ms 025_00ms 050_00ms 100_00ms}
set r_list {r1 r2 r3}
set m 0; set sourcedir C:/VMD; set writedir C:/VMD/CRD

set procedure {
    set N1 [atomselect top "(segname P10 and resid 9 and name CA)"]
    animate write crd $writedir/$xlink\_N\_P10_resid9_$vr.crd sel $N1 $m
    set N2 [atomselect top "(segname P14 and resid 946 and name CA)"]
    animate write crd $writedir/$xlink\_N\_P14_resid946_$vr.crd sel $N2 $m
    set C1 [atomselect top "(segname P0 and resid 103 and name CA)"]
    animate write crd $writedir/$xlink\_C\_P0_resid103_$vr.crd sel $C1 $m
    set C2 [atomselect top "(segname P4 and resid 1046 and name CA)"]
    animate write crd $writedir/$xlink\_C\_P4_resid1046_$vr.crd sel $C2 $m

    if {"HHL" == $xlink} {
        set N3 [atomselect top "(segname P4 and resid 948 and name CA)"]
        animate write crd $writedir/$xlink\_N\_P4_resid948_$vr.crd sel $N3 $m
        set C3 [atomselect top "(segname P5 and resid 108 and name CA)"]
        animate write crd $writedir/$xlink\_C\_P5_resid108_$vr.crd sel $C3 $m
    } elseif {"PYD" == $xlink} {
        set N3 [atomselect top "(segname P5 and resid 12 and name CA)"]
        animate write crd $writedir/$xlink\_N\_P5_resid12_$vr.crd sel $N3 $m
        set C3 [atomselect top "(segname P14 and resid 1046 and name CA)"]
        animate write crd $writedir/$xlink\_C\_P14_resid1046_$vr.crd sel $C3 $m}}

foreach xlink $xlink_list {
    mol load pdb $sourcedir/COOR/$xlink.coor psf $sourcedir/PSF/$xlink.psf
    set vr initial; eval $procedure; incr m

    foreach v $v_list {foreach r $r_list {
        mol load dcd $sourcedir/DCD/$xlink\_v\_r.dcd psf $sourcedir/PSF/$xlink.psf
        set vr $v\_r; eval $procedure; incr m}}
    mol delete all}
```

Appendix C : Mathematica notebook files

C.1. Mathematica notebook used for the crosslink Ca atom Cartesian coordinate data

```

TableForm[Flatten[Data1 = Table[
"PYD_C_P0_resid103_" <> ToString[i] <> "_r" <> ToString[j] <> ".crd",
{i, {"006_25ms", "012_50ms", "025_00ms", "050_00ms", "100_00ms"}}, {j, 1, 3}]]];
TableForm[Flatten[Data2 = Table[
"PYD_C_P4_resid1046_" <> ToString[i] <> "_r" <> ToString[j] <> ".crd",
{i, {"006_25ms", "012_50ms", "025_00ms", "050_00ms", "100_00ms"}}, {j, 1, 3}]]];
TableForm[Flatten[Data3 = Table[
"PYD_C_P14_resid1046_" <> ToString[i] <> "_r" <> ToString[j] <> ".crd",
{i, {"006_25ms", "012_50ms", "025_00ms", "050_00ms", "100_00ms"}}, {j, 1, 3}]]];

TableForm[Flatten[Data4 = Table[
"PYD_N_P10_resid9_" <> ToString[i] <> "_r" <> ToString[j] <> ".crd",
{i, {"006_25ms", "012_50ms", "025_00ms", "050_00ms", "100_00ms"}}, {j, 1, 3}]]];
TableForm[Flatten[Data5 = Table[
"PYD_N_P14_resid946_" <> ToString[i] <> "_r" <> ToString[j] <> ".crd",
{i, {"006_25ms", "012_50ms", "025_00ms", "050_00ms", "100_00ms"}}, {j, 1, 3}]]];
TableForm[Flatten[Data6 = Table[
"PYD_N_P5_resid12_" <> ToString[i] <> "_r" <> ToString[j] <> ".crd",
{i, {"006_25ms", "012_50ms", "025_00ms", "050_00ms", "100_00ms"}}, {j, 1, 3}]]];

Clear[DataTable, DataColumn, MergedData]; MergedData = {};
SetDirectory["C:/VMD/CRD/"];
For[j = 1, j < 16, j++, DataTable = Import[Flatten[Data1][[j]], "Table"];
For[k = 1, k < 4, k++, DataColumn = DataTable[[2 ;; 71, k]];
AppendTo[MergedData, DataColumn];]
For[j = 1, j < 16, j++, DataTable = Import[Flatten[Data2][[j]], "Table"];
For[k = 1, k < 4, k++, DataColumn = DataTable[[2 ;; 71, k]];
AppendTo[MergedData, DataColumn];]
For[j = 1, j < 16, j++, DataTable = Import[Flatten[Data3][[j]], "Table"];
For[k = 1, k < 4, k++, DataColumn = DataTable[[2 ;; 71, k]];
AppendTo[MergedData, DataColumn];]
SetDirectory["C:/VMD/CRD/Merged/"];
Export["PYD_C_xlink_crd.dat", Transpose[MergedData]];

```

Note: The last block of lines above was repeated for the N-crosslink by replacing "..._C_xlink_crd.dat" with "..._N_xlink_crd.dat" and *Data 1*, *Data 2*, and *Data 3* with *Data 4*, *Data 5*, and *Data 6*, respectively. All three blocks of lines were then repeated for each of the other crosslinks by replacing PYD with HHL, HKNL, or HLNL and by either omitting steps involving *Data 3* and *Data 6* (HKNL and HHL) or making appropriate modifications to *Data 3* and *Data 6* (HHL).

C.2. Mathematica notebook used for the crosslink energy data

```
TableForm[Flatten[
DataC = Table[ToString[i] <> "_C_" <> ToString[j] <> "_r" <> ToString[k] <> ".dat",
{i, {"HHL", "HKNL", "HLNL", "PYD"}},
{j, {"006_25ms", "012_50ms", "025_00ms", "050_00ms", "100_00ms"}},
{k, 1, 3}]]];
```

```
TableForm[Flatten[
DataN = Table[ToString[i] <> "_N_" <> ToString[j] <> "_r" <> ToString[k] <> ".dat",
{i, {"HHL", "HKNL", "HLNL", "PYD"}},
{j, {"006_25ms", "012_50ms", "025_00ms", "050_00ms", "100_00ms"}},
{k, 1, 3}]]];
```

```
Clear[DataTable, DataColumn, MergedData]; MergedData = {};
SetDirectory["C:/VMD/Energy/"];
For[j = 1, j < 16, j++, DataTable = Import[Flatten[DataC][[j]]];
DataColumn = DataTable[[2 ;; 71, 11]]; AppendTo[MergedData, DataColumn];]
SetDirectory["C:/VMD/Energy/Merged/"];
Export["HHL_C_xlink_energy.dat", Transpose[MergedData]];
```

```
Clear[DataTable, DataColumn, MergedData]; MergedData = {};
SetDirectory["C:/VMD/Energy/"];
For[j = 16, j < 31, j++, DataTable = Import[Flatten[DataC][[j]]];
DataColumn = DataTable[[2 ;; 71, 11]]; AppendTo[MergedData, DataColumn];]
SetDirectory["C:/VMD/Energy/Merged/"];
Export["HKNL_C_xlink_energy.dat", Transpose[MergedData]];
```

```
Clear[DataTable, DataColumn, MergedData]; MergedData = {};
SetDirectory["C:/VMD/Energy/"];
For[j = 31, j < 46, j++, DataTable = Import[Flatten[DataC][[j]]];
DataColumn = DataTable[[2 ;; 71, 11]]; AppendTo[MergedData, DataColumn];]
SetDirectory["C:/VMD/Energy/Merged/"];
Export["HLNL_C_xlink_energy.dat", Transpose[MergedData]];
```

```
Clear[DataTable, DataColumn, MergedData]; MergedData = {};
SetDirectory["C:/VMD/Energy/"];
For[j = 46, j < 61, j++, DataTable = Import[Flatten[DataC][[j]]];
DataColumn = DataTable[[2 ;; 71, 11]]; AppendTo[MergedData, DataColumn];]
SetDirectory["C:/VMD/Energy/Merged/"];
Export["PYD_C_xlink_energy.dat", Transpose[MergedData]];
```

Note: The last four blocks of lines were repeated for the N-crosslink by replacing DataC with DataN and "..._C_xlink_energy.dat" with "..._N_xlink_energy.dat".

C.3. Mathematica notebook used for the energy data for the entire model

```
TableForm[Flatten[DataAll = Table[
ToString[i] <> "_" <> ToString[j] <> "_r" <> ToString[k] <> ".dat",
{i, {"HHL", "HKNL", "HLNL", "PYD", "UNX"}},
{j, {"006_25ms", "012_50ms", "025_00ms", "050_00ms", "100_00ms"}},
{k, 1, 3}]]];
```

```
Clear[DataTable, DataColumn, MergedData]; MergedData = {};
SetDirectory["C:/VMD/Energy/"];
For[j = 1, j < 16, j++, DataTable = Import[Flatten[DataAll][[j]]];
DataColumn = DataTable[[2 ;; 72, 11]]; AppendTo[MergedData, DataColumn];]
SetDirectory["C:/VMD/Energy/Merged/"];
Export["HHL_all_energy.dat", Transpose[MergedData]];
```

```
Clear[DataTable, DataColumn, MergedData]; MergedData = {};
SetDirectory["C:/VMD/Energy/"];
For[j = 16, j < 31, j++, DataTable = Import[Flatten[DataAll][[j]]];
DataColumn = DataTable[[2 ;; 72, 11]]; AppendTo[MergedData, DataColumn];]
SetDirectory["C:/VMD/Energy/Merged/"];
Export["HKNL_all_energy.dat", Transpose[MergedData]];
```

```
Clear[DataTable, DataColumn, MergedData]; MergedData = {};
SetDirectory["C:/VMD/Energy/"];
For[j = 31, j < 46, j++, DataTable = Import[Flatten[DataAll][[j]]];
DataColumn = DataTable[[2 ;; 72, 11]]; AppendTo[MergedData, DataColumn];]
SetDirectory["C:/VMD/Energy/Merged/"];
Export["HLNL_all_energy.dat", Transpose[MergedData]];
```

```
Clear[DataTable, DataColumn, MergedData]; MergedData = {};
SetDirectory["C:/VMD/Energy/"];
For[j = 46, j < 61, j++, DataTable = Import[Flatten[DataAll][[j]]];
DataColumn = DataTable[[2 ;; 72, 11]]; AppendTo[MergedData, DataColumn];]
SetDirectory["C:/VMD/Energy/Merged/"];
Export["PYD_all_energy.dat", Transpose[MergedData]];
```

```
Clear[DataTable, DataColumn, MergedData]; MergedData = {};
SetDirectory["C:/VMD/Energy/"];
For[j = 61, j < 76, j++, DataTable = Import[Flatten[DataAll][[j]]];
DataColumn = DataTable[[2 ;; 72, 11]]; AppendTo[MergedData, DataColumn];]
SetDirectory["C:/VMD/Energy/Merged/"];
Export["UNX_all_energy.dat", Transpose[MergedData]];
```

Note: Data for the UNX2, HLNL2, and HLNL3 simulations were done in the same manner.

Appendix D : Influence of the pulling velocity on the ~23-nm-long molecular model

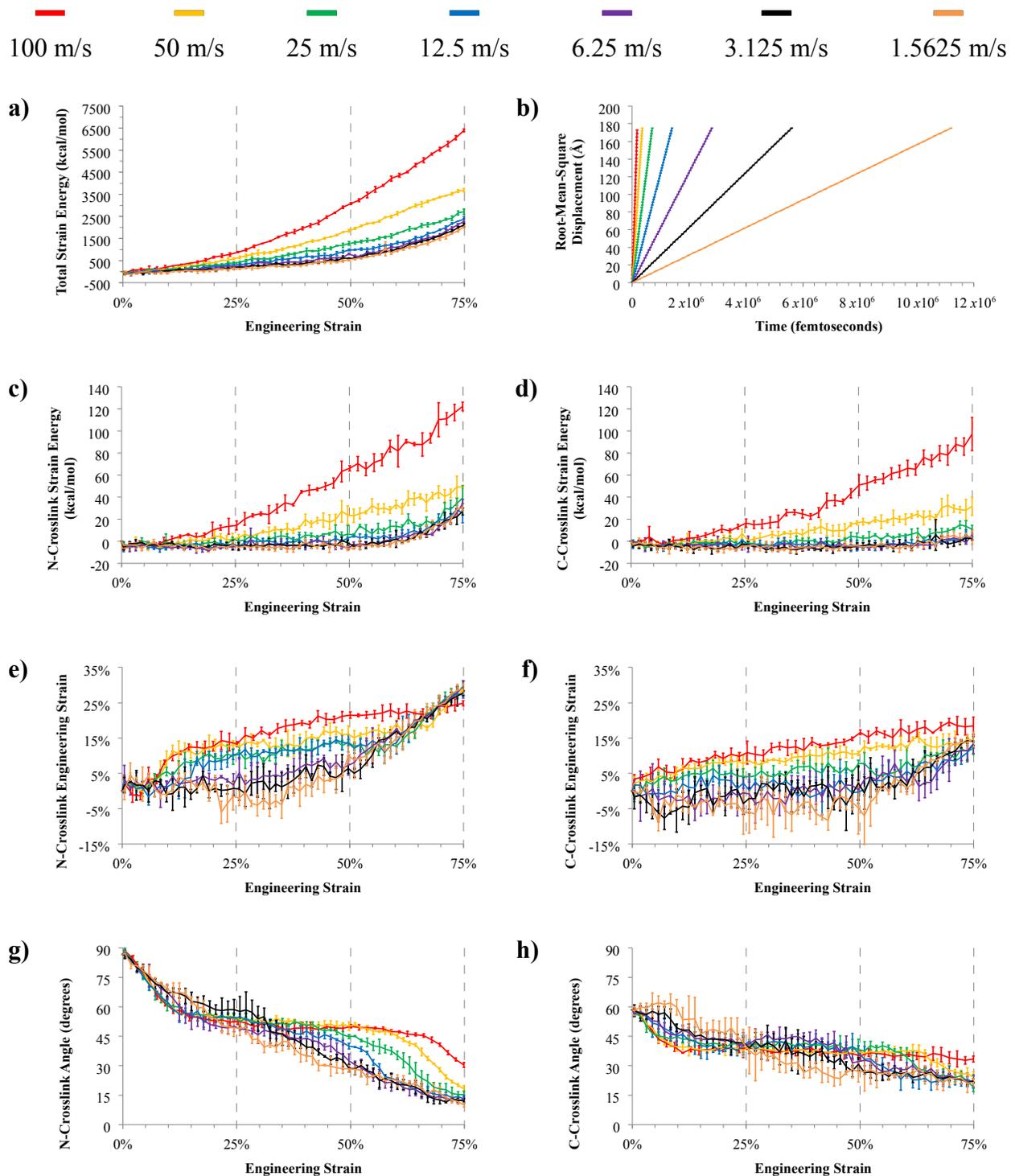


Figure D.1: Influence of the pulling velocity on the ~23-nm-long model. “Engineering Strain” on the x-axes refers to strain of the overall ~23-nm-long model. (a) total strain energy vs. engineering strain, (b) root-mean-square displacement vs. time, (c, d) strain energy in the N- and C-crosslinks, respectively, (e, f) engineering strain in the N- and C-crosslinks, respectively, and (g, h) alignment of the N- and C-crosslinks, respectively, relative to the loading direction. Data points represent means \pm s.d. ($n = 3$).

Appendix E : Trajectory snapshots of the ~65-nm-long molecular models

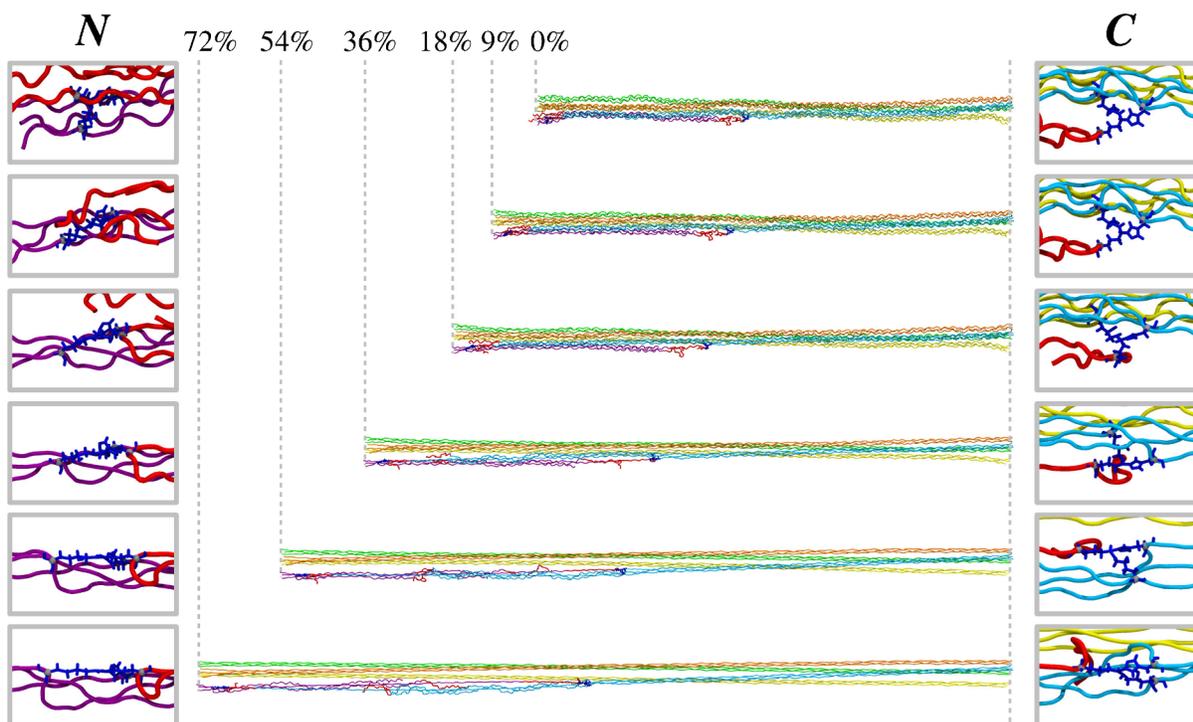


Figure E.1: HHL model snapshots (12 fixed C-term $C\alpha$; 12 pulled N-term $C\alpha$). Snapshots were taken at six different levels of engineering strain, namely, 0, 9, 18, 36, 54, and 72% stain. The color scheme for Figures D.1 through D.8 is identical to that described in Figure 4.1.

Appendix E (continued)

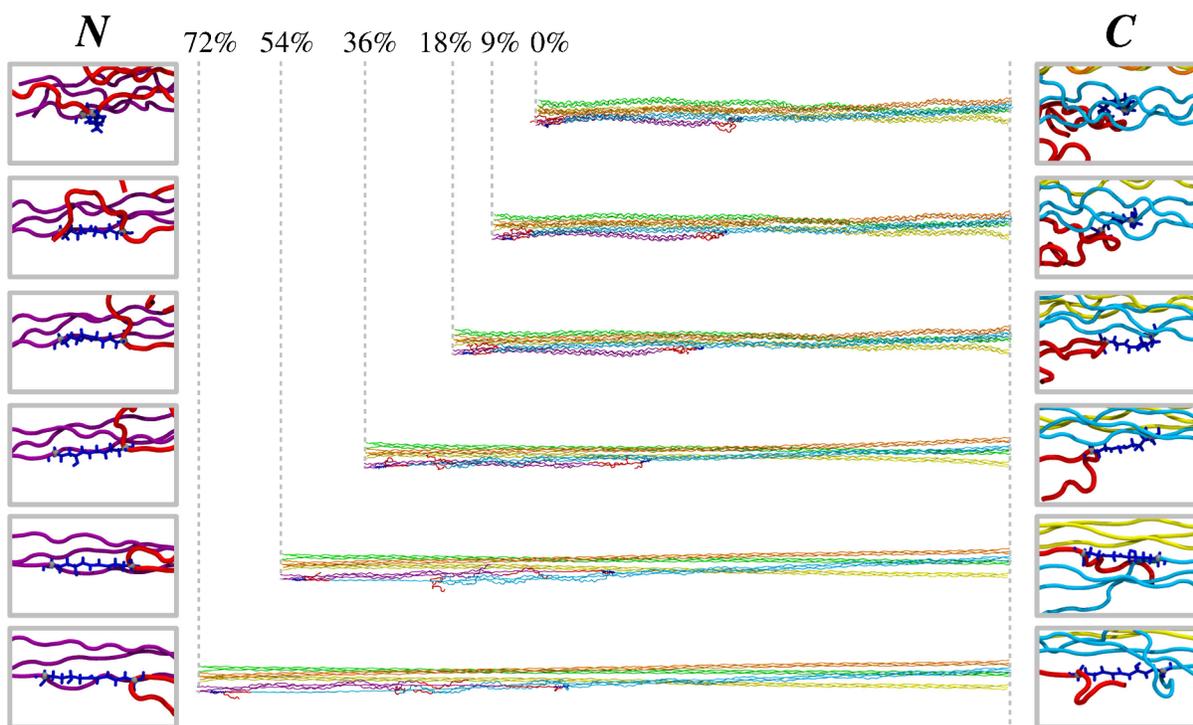


Figure E.2: HLKLN model snapshots (12 fixed C-term $C\alpha$; 12 pulled N-term $C\alpha$). Snapshots were taken at six different levels of engineering strain, namely, 0, 9, 18, 36, 54, and 72% stain.

Appendix E (continued)

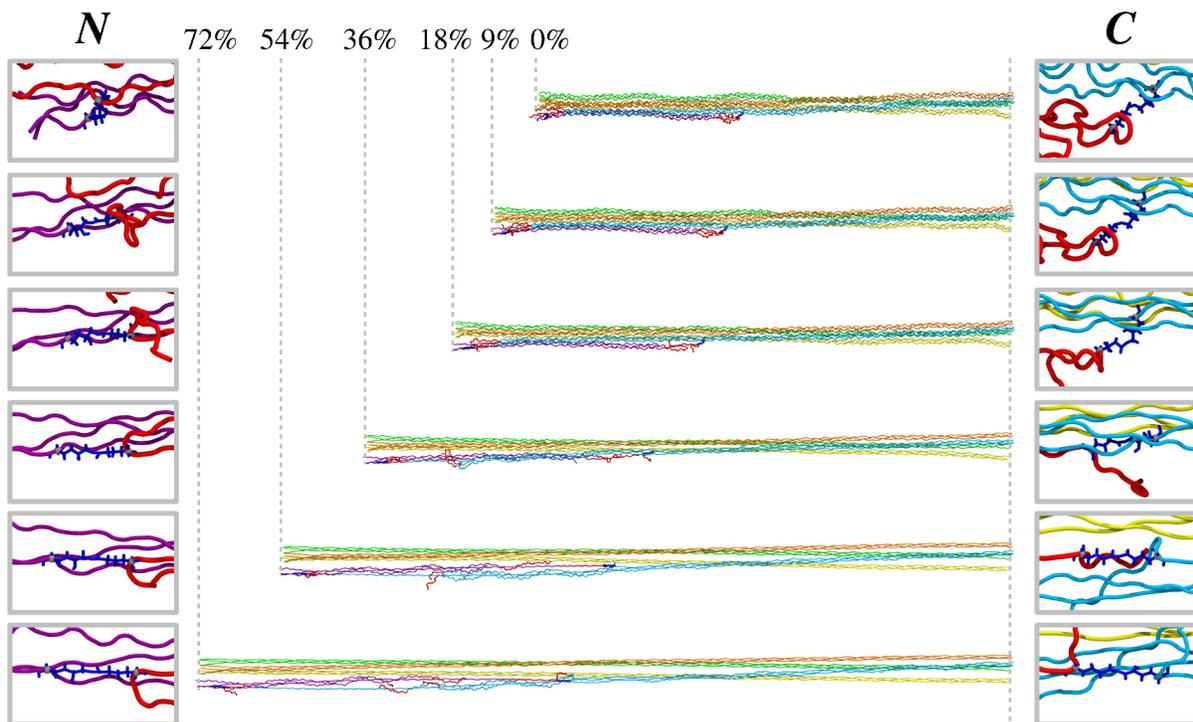


Figure E.3: deH-HLNL model snapshots (12 fixed C-term $\text{C}\alpha$; 12 pulled N-term $\text{C}\alpha$). Snapshots were taken at six different levels of engineering strain, namely, 0, 9, 18, 36, 54, and 72% strain.

Appendix E (continued)

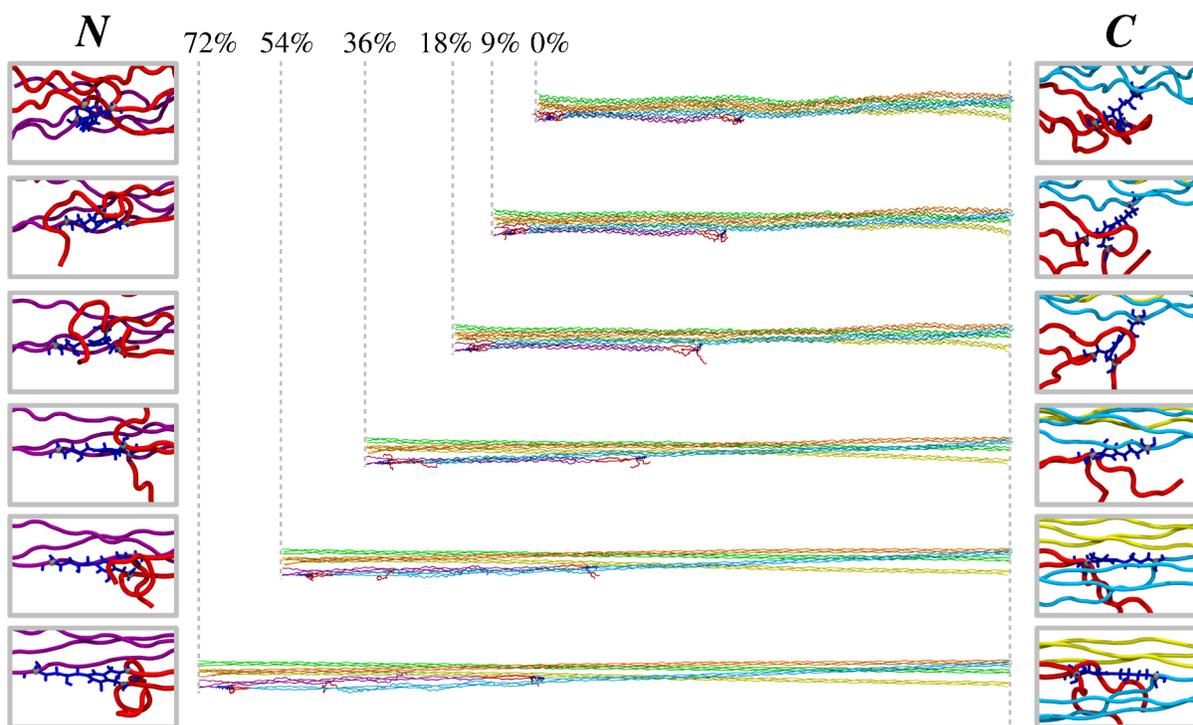


Figure E.4: PYD model snapshots (12 fixed C-term $C\alpha$; 12 pulled N-term $C\alpha$). Snapshots were taken at six different levels of engineering strain, namely, 0, 9, 18, 36, 54, and 72% strain.

Appendix E (continued)

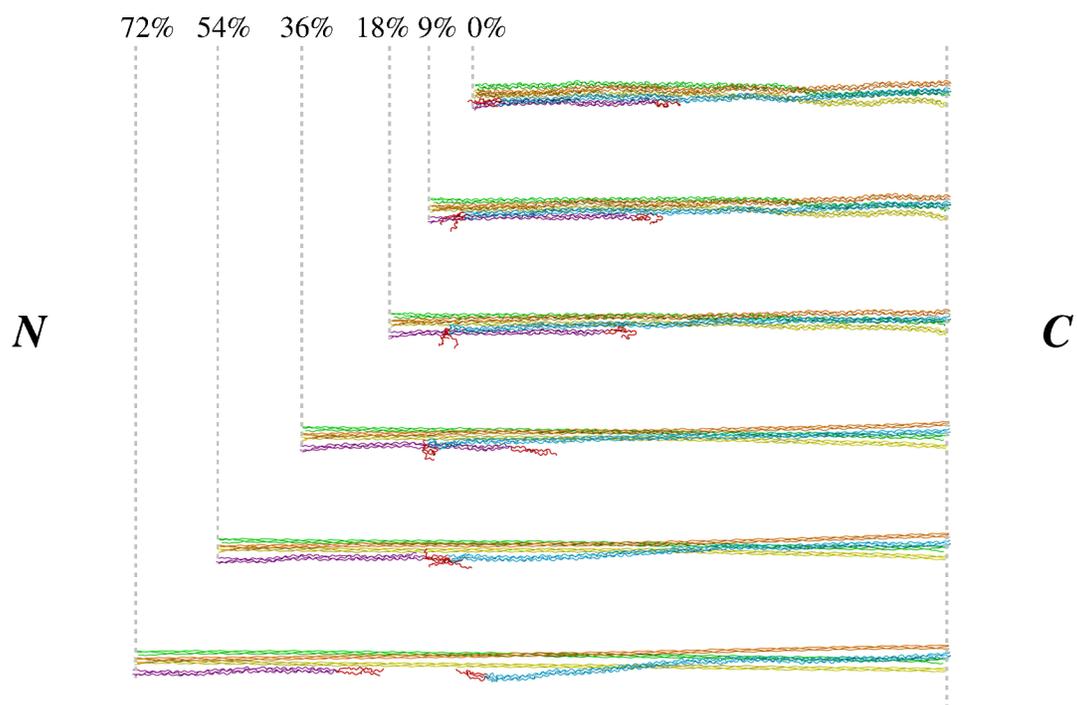


Figure E.5: Uncrosslinked model snapshots (12 fixed C-term $C\alpha$; 12 pulled N-term $C\alpha$). Snapshots were taken at six different levels of engineering strain, namely, 0, 9, 18, 36, 54, and 72% strain.

Appendix E (continued)

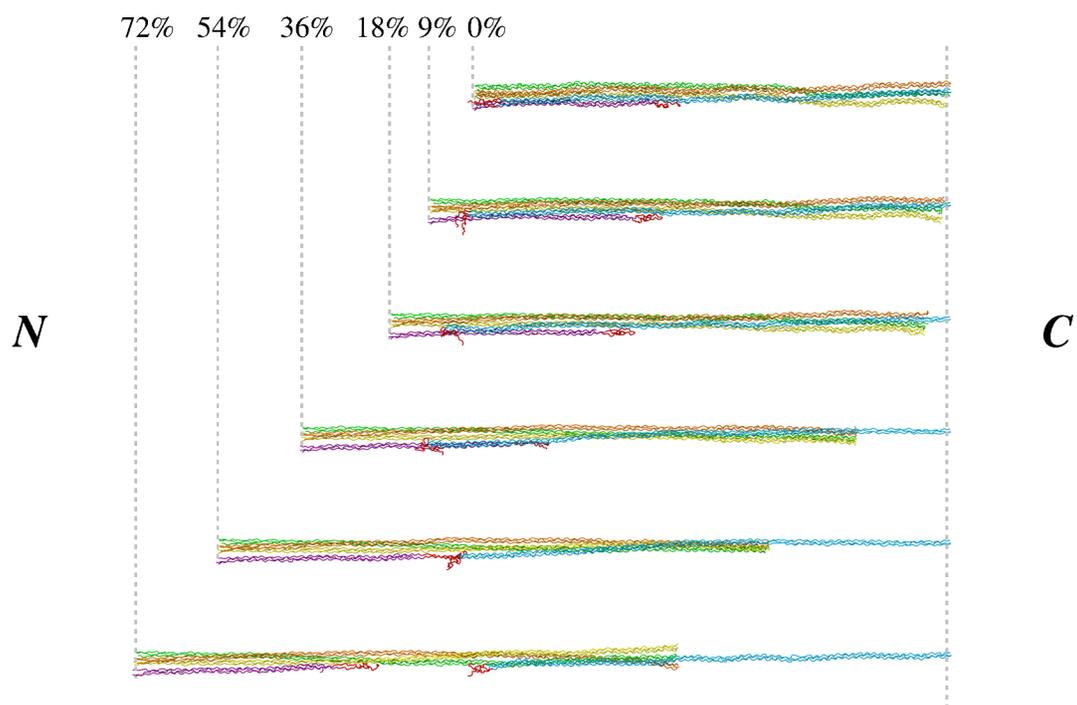


Figure E.6: Uncrosslinked model snapshots (3 fixed C-term $C\alpha$; 12 pulled N-term $C\alpha$). Snapshots were taken at six different levels of engineering strain, namely, 0, 9, 18, 36, 54, and 72% strain.

Appendix E (continued)

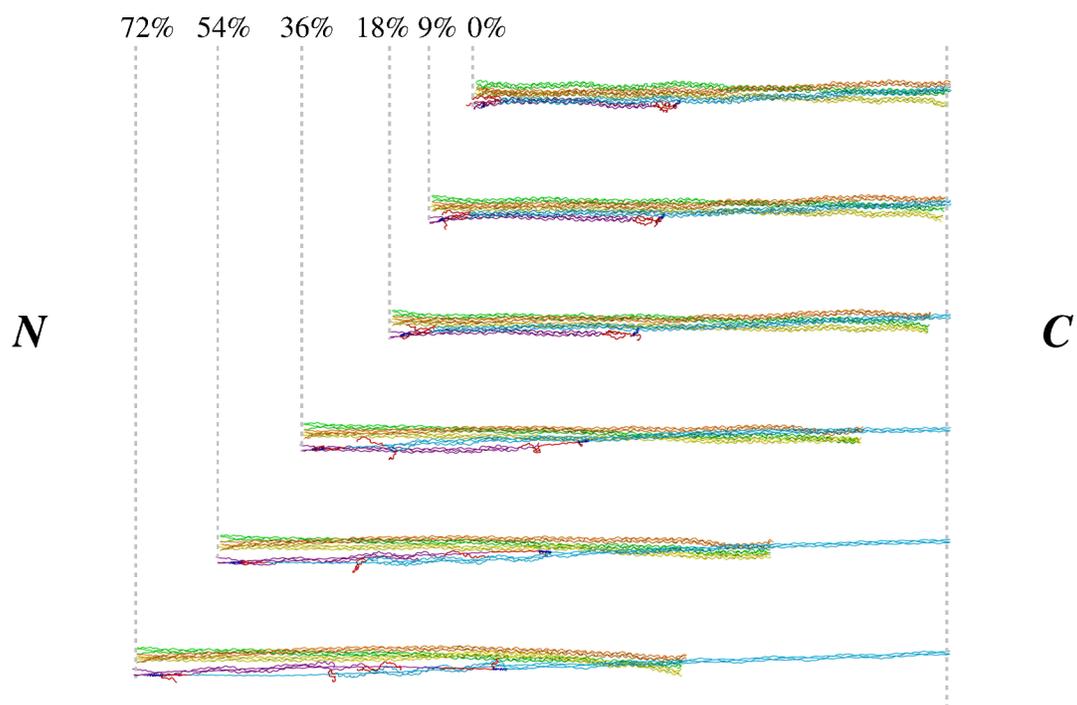


Figure E.7: deH-HLNL model snapshots (3 fixed C-term $C\alpha$; 12 pulled N-term $C\alpha$). Snapshots were taken at six different levels of engineering strain, namely, 0, 9, 18, 36, 54, and 72% strain.

Appendix F : Influence of the pulling velocity on the ~65-nm-long molecular models

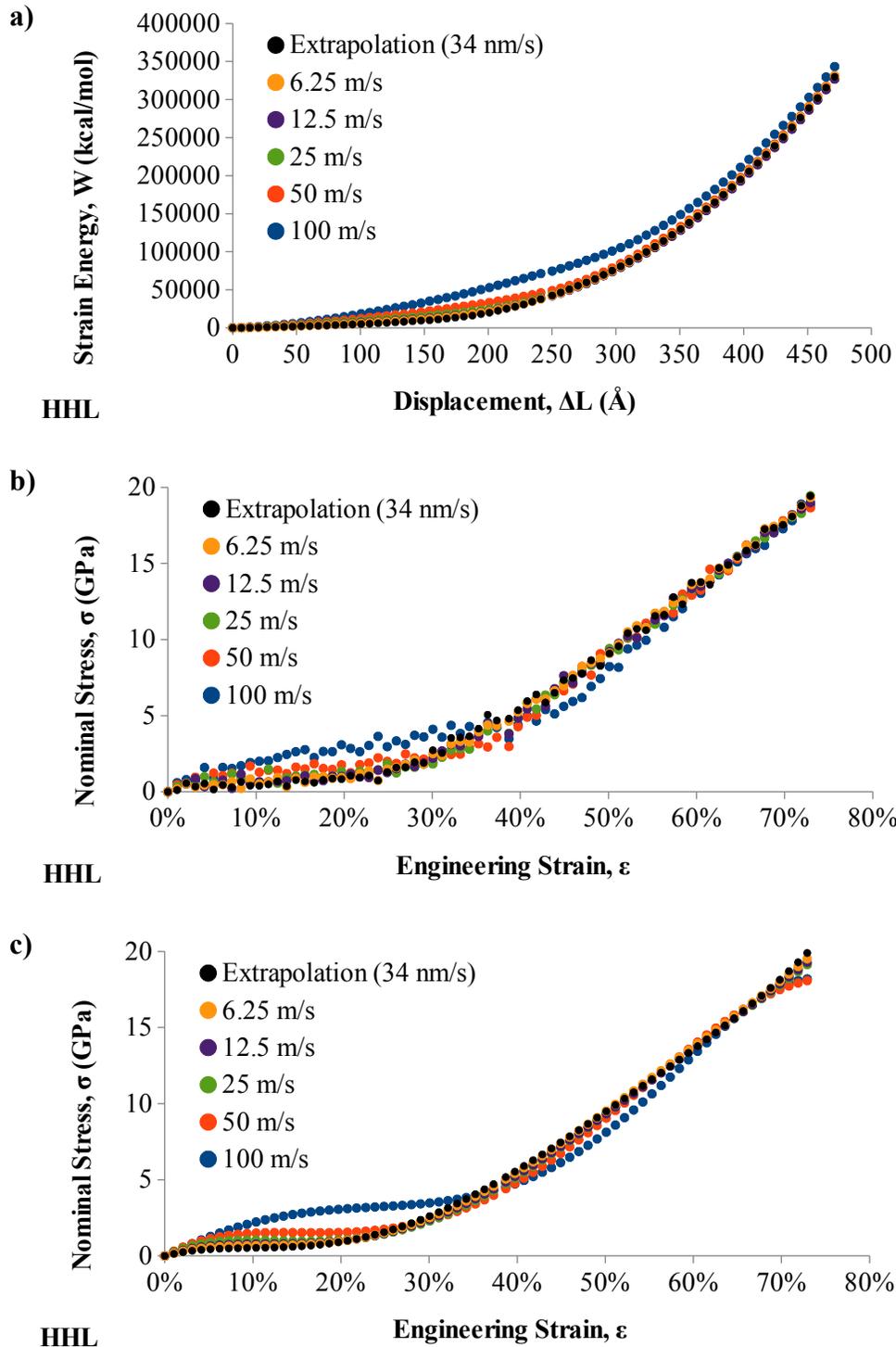


Figure F.1: Influence of the pulling velocity on the HHL-crosslinked model with 12 fixed C-term $C\alpha$ atoms and 12 pulled N-term $C\alpha$ atoms. Shown are two ways that were attempted to convert the strain energy-extension data (a) to nominal stress (b and c). (b) shows results from computing numerical derivatives (i.e., the incremental differences in strain energy and displacement). (c) shows results from fitting the strain energy data to a 6th order polynomial and taking an analytic derivative.

Appendix F (continued)

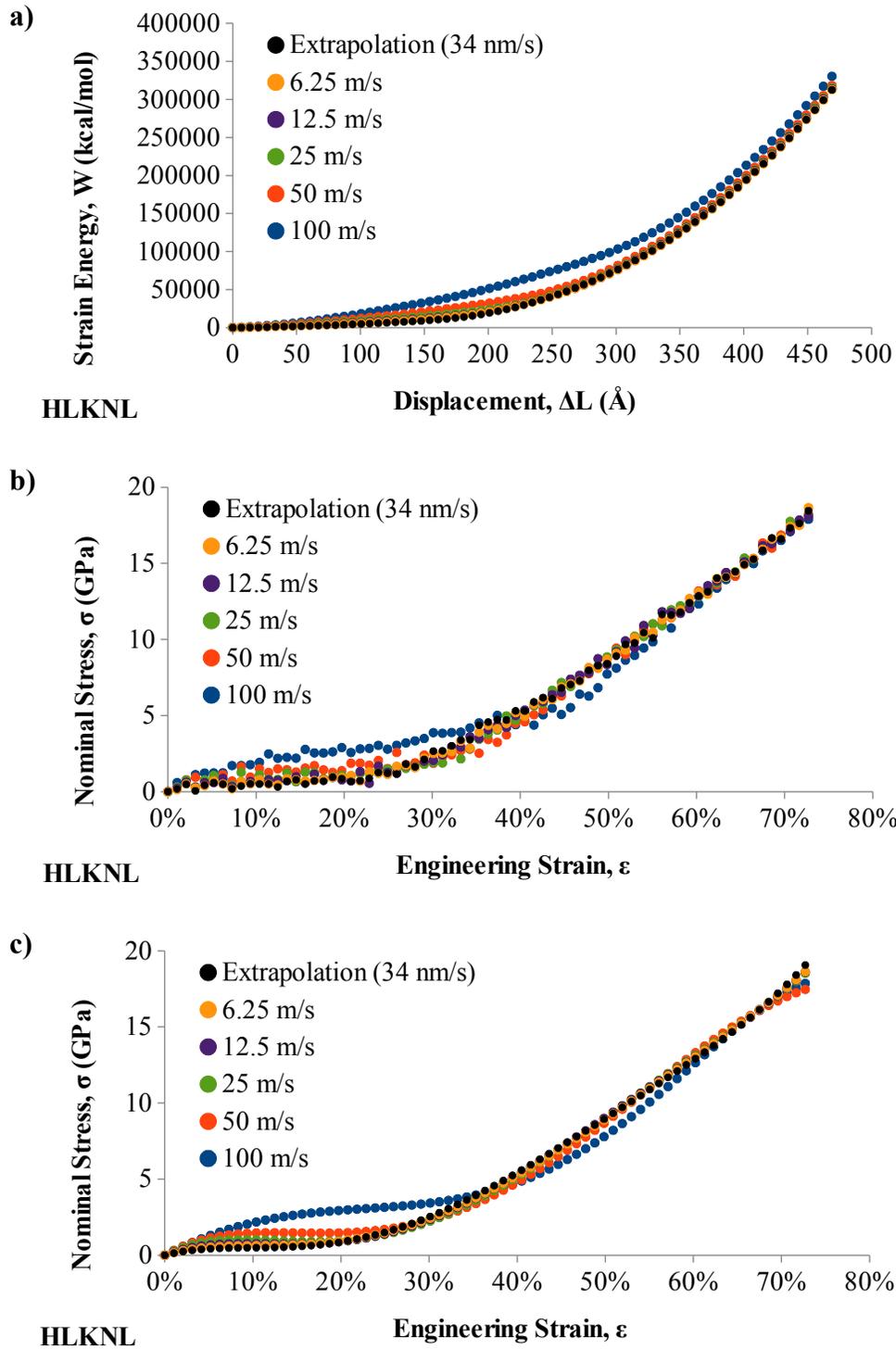


Figure F.2: Influence of the pulling velocity on the HLKLN-crosslinked model with 12 fixed C-term $C\alpha$ atoms and 12 pulled N-term $C\alpha$ atoms. See Figure F.1 for details on plots a, b, and c.

Appendix F (continued)

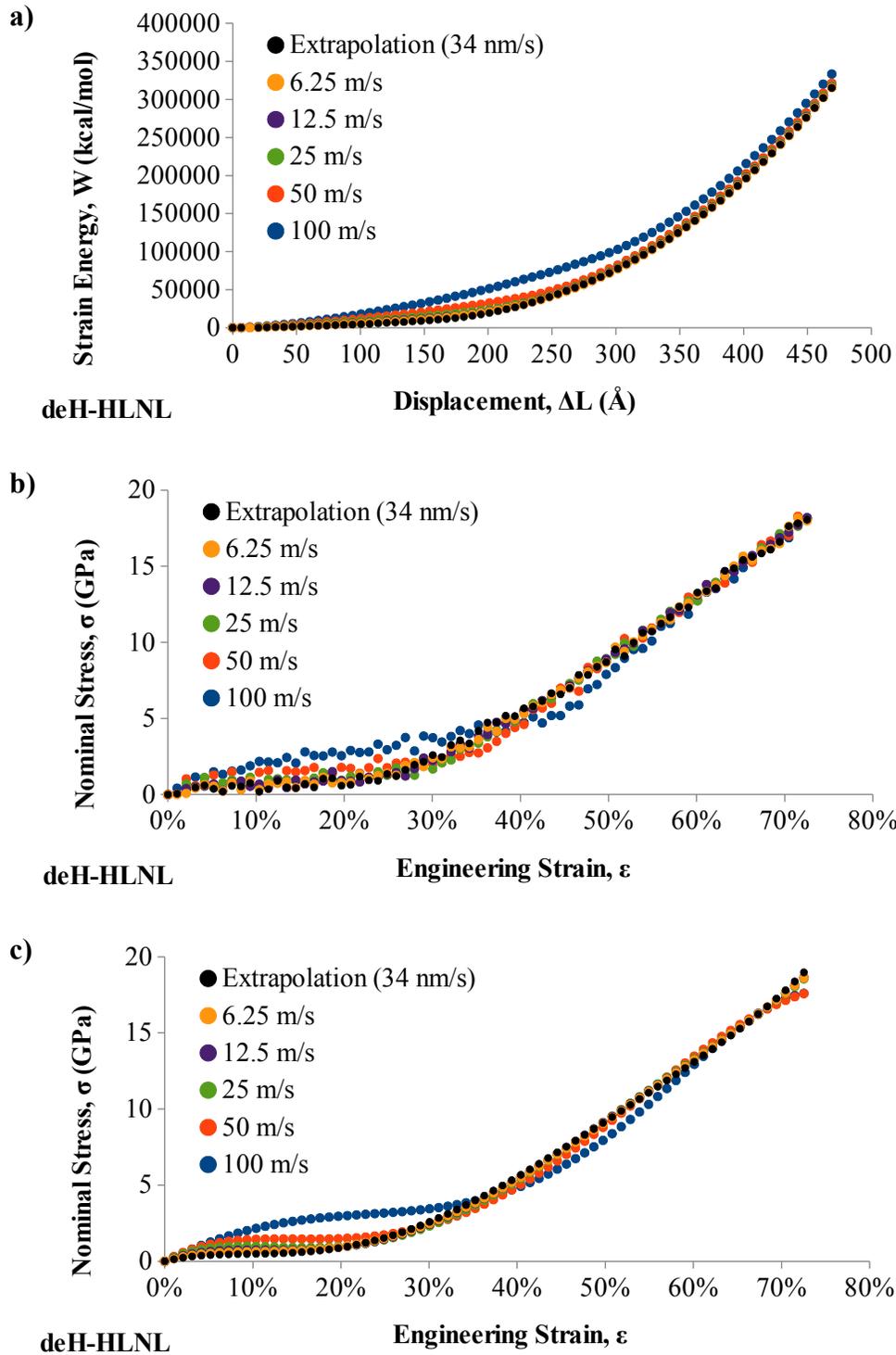


Figure F.3: Influence of the pulling velocity on the deH-HLNL-crosslinked model with 12 fixed C-term $C\alpha$ atoms and 12 pulled N-term $C\alpha$ atoms. See Figure F.1 for details on plots a, b, and c.

Appendix F (continued)

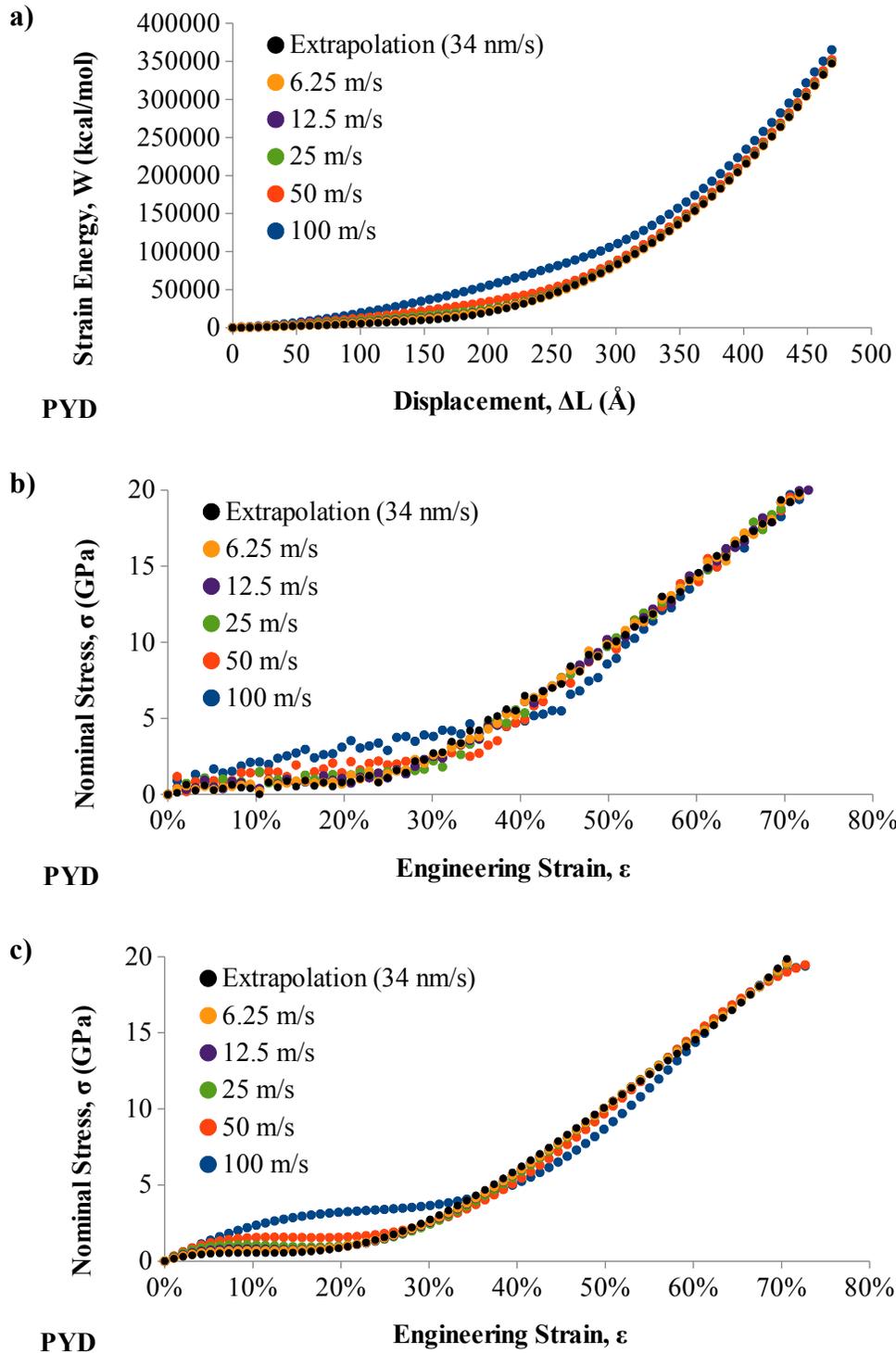


Figure F.4: Influence of the pulling velocity on the PYD-crosslinked model with 12 fixed C-term $C\alpha$ atoms and 12 pulled N-term $C\alpha$ atoms. See Figure F.1 for details on plots a, b, and c.

Appendix F (continued)

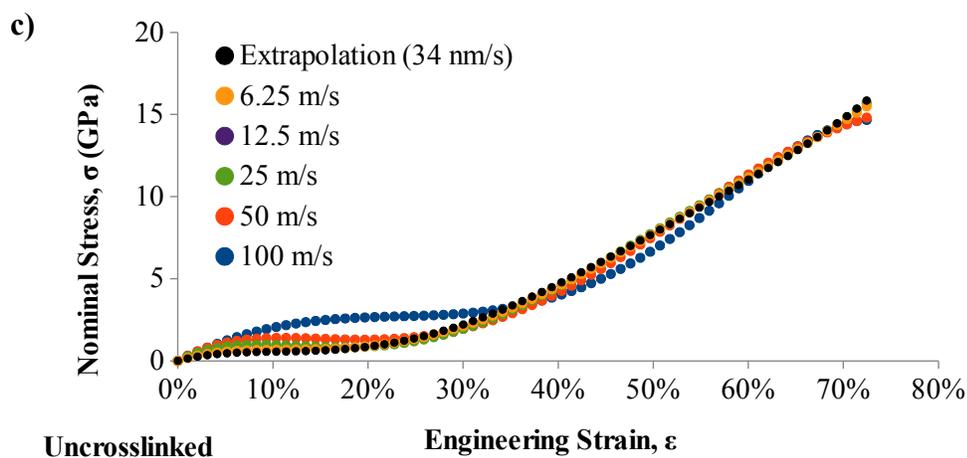
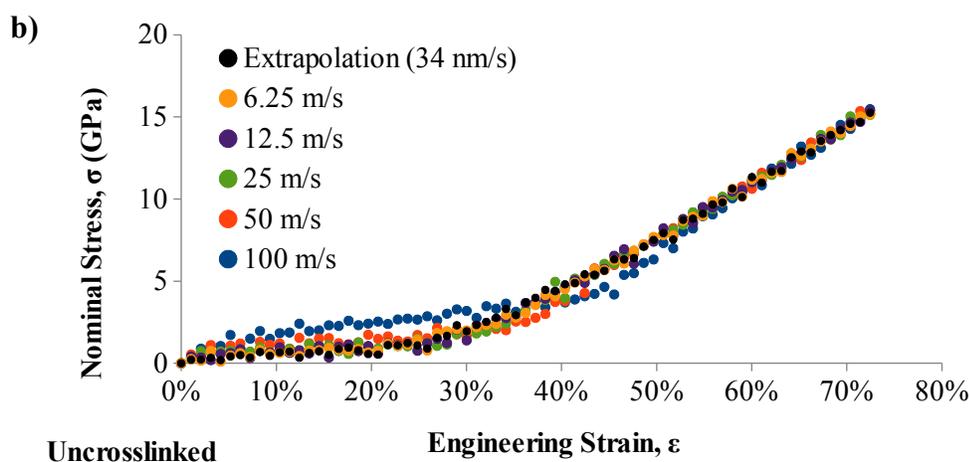
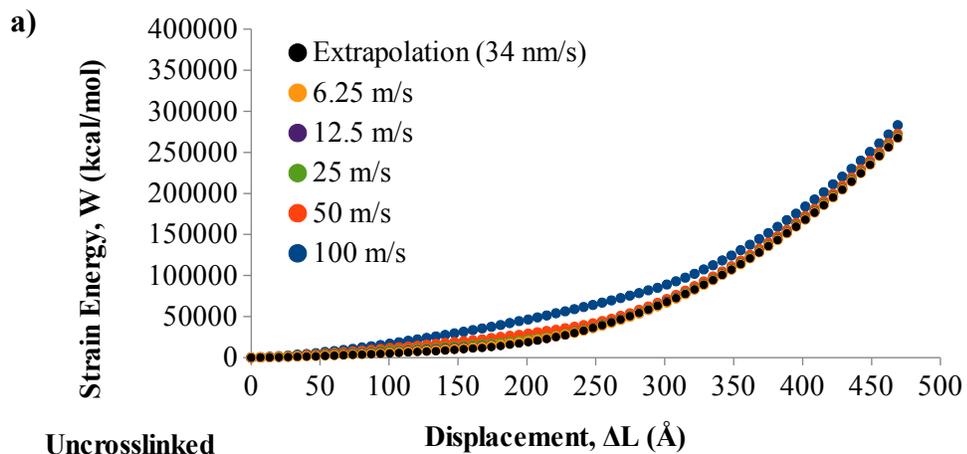


Figure F.5: Influence of the pulling velocity on the uncrosslinked model with 12 fixed C-term $C\alpha$ atoms and 12 pulled N-term $C\alpha$ atoms. See Figure F.1 for details on plots a, b, and c.

Appendix F (continued)

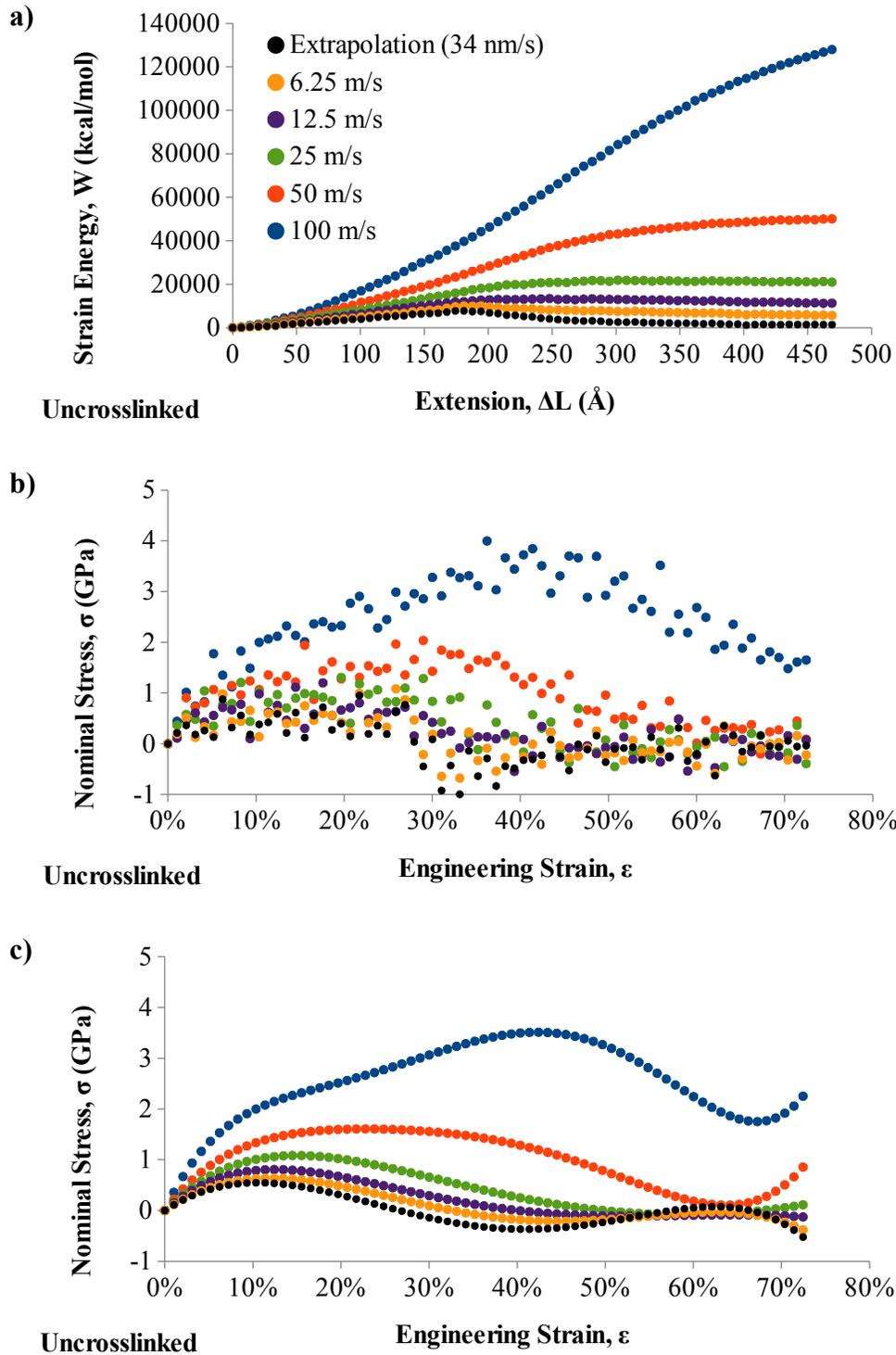
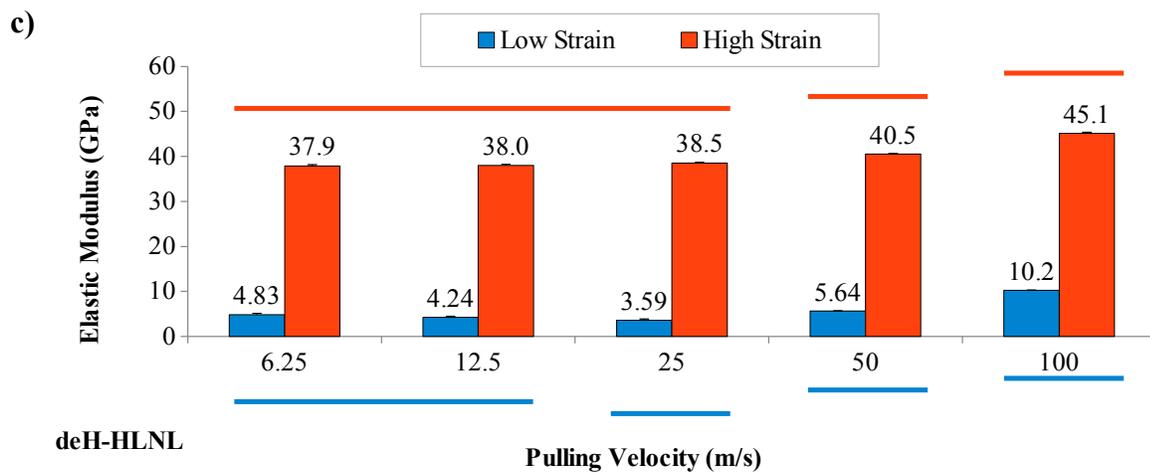
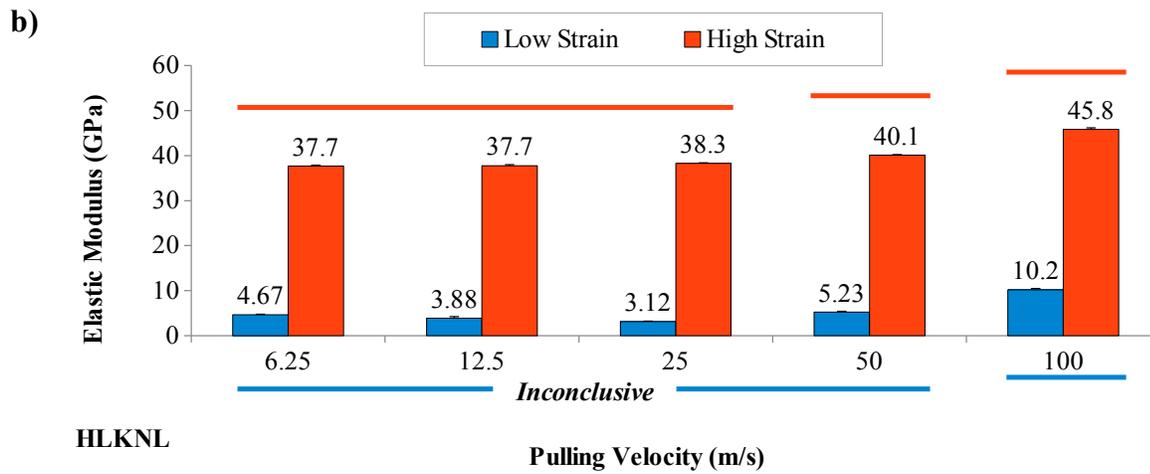
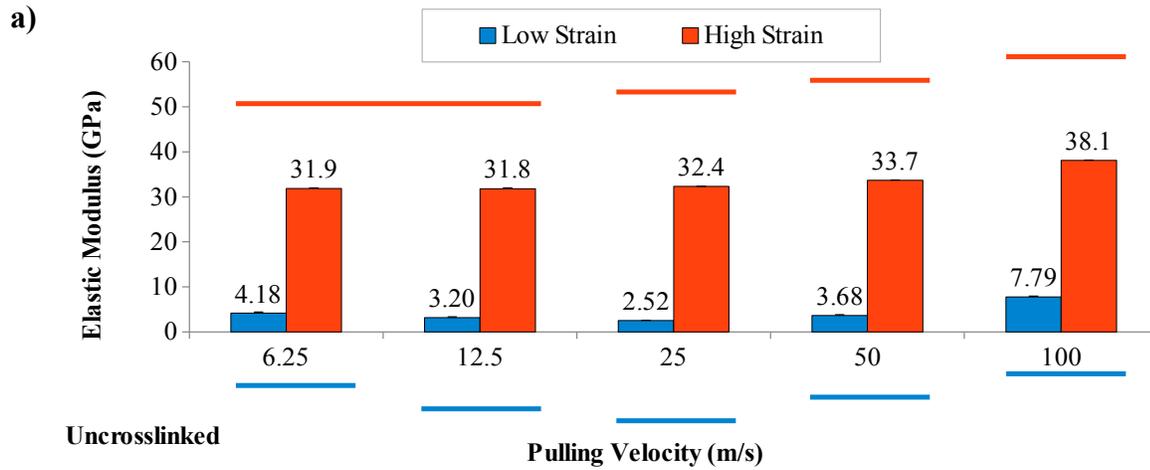


Figure F.6: Influence of the pulling velocity on the uncrosslinked model with 12 fixed C-term $C\alpha$ atoms and 12 pulled N-term $C\alpha$ atoms. See Figure F.1 for details on plots a, b, and c.

Appendix G : Influence of the pulling velocity on the moduli of the ~65-nm-long models



Appendix G (continued)

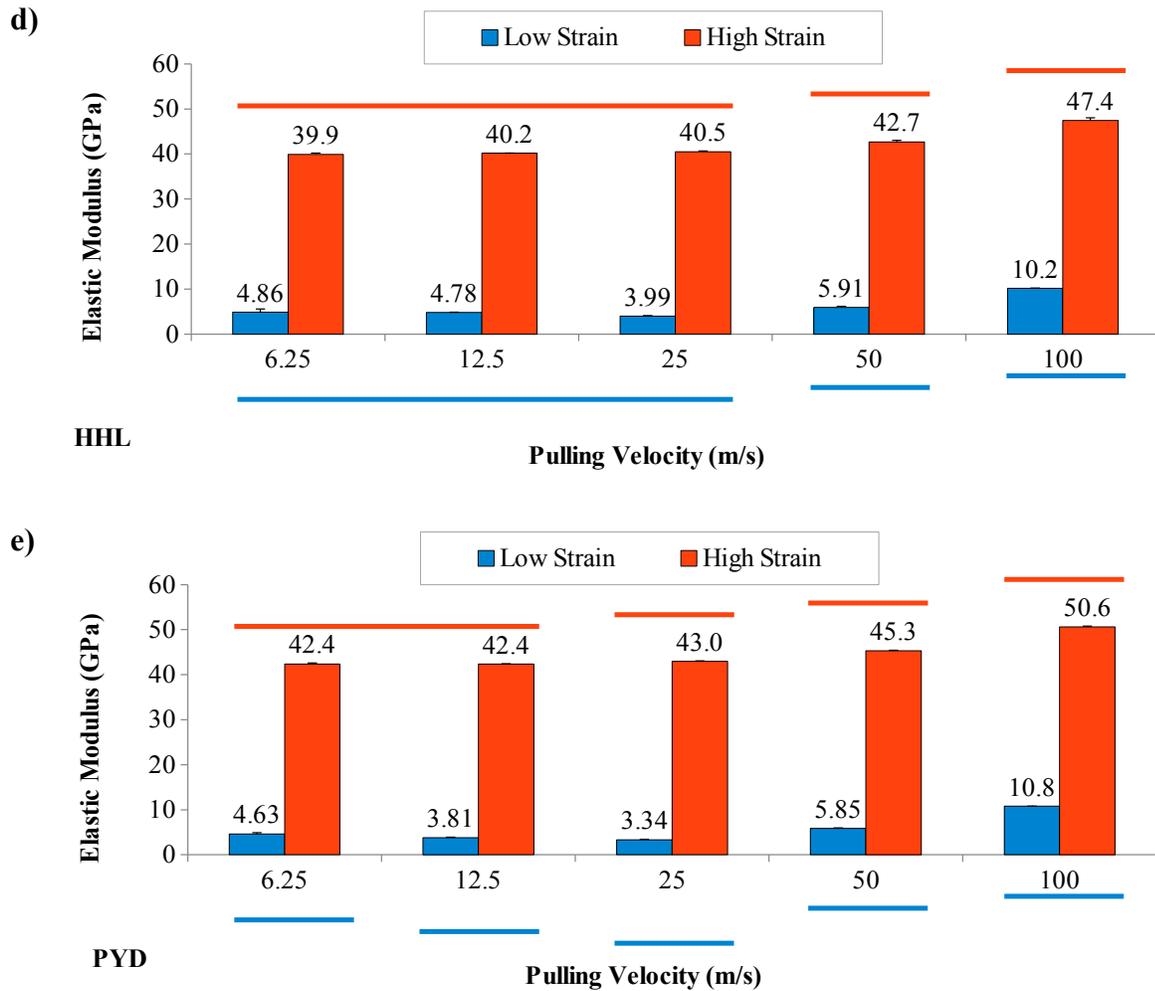


Figure G.1: Influence of the pulling velocity on the ~65-nm-long model. The effect of the pulling velocity on the low- and high-strain elastic moduli for the microfibril unit models are compared using Student's t-tests with Bonferroni correction (i.e., $\alpha = 0.005 = 0.05$ divided by 10 t-tests). The sub-figures include (a) the uncrosslinked model, (b) immature HLKLN-crosslinked model, (c) immature deH-HLNL-crosslinked model, (d) mature HHL-crosslinked model, and (e) mature PYD-crosslinked model. The red (upper) and blue (lower) horizontal lines denote results from the Student's t-tests; horizontal lines connecting the different pulling velocities indicate no statistical significance, while a lack of connection of these lines indicates significance. These lines have different positions to indicate the direction of significance (e.g., 100 m/s resulted in significantly higher low- and high-strain moduli relative to 50 m/s). Data points represent means \pm s.d. ($n = 3$).

Appendix H : Listing of additional work

H.1. Manuscripts in preparation

Kwansa, A.L., De Vita, R., & Freeman, J.W. Mechanical recruitment of N- and C-crosslinks in collagen type I, (in preparation).

Abstract: Collagen type I is an extracellular matrix protein found in connective tissues such as tendon, ligament, bone, skin, and the cornea of the eyes, where it functions mainly to provide tensile strength and to serve as a scaffold for cells and other extracellular matrix components. A single collagen type I molecule is composed of three amino acid chains that form a triple helix for most of the molecule's length; non-triple-helical extensions called N- and C-telopeptides are located at the beginning (amino/N-terminus) and end (carboxy/C-terminus) of the molecule. In two of the three chains, the C-telopeptide has been reported to possess a hair-pin/hook conformation, while the three N-telopeptides display an extended coil structure. These telopeptides are crucial for the formation of enzyme-derived covalent crosslinks that form in collagens near their N- and C-ends. Such crosslinks provide mechanical integrity, stiffness, and strength to collagenous tissues. However, deformation mechanisms of N- and C-crosslinks and functional roles for the N- and C-telopeptide conformations are not yet well known. By performing molecular dynamics simulations, we show that two dehydro-hydroxylysino-norleucine crosslinks positioned at N- and C-crosslinking sites exhibit a two-stage response to the mechanical deformation of their parent molecules. We observed that the N-crosslink served as a first responder to mechanical strain, followed by the C-crosslink at higher strains. The two-stage response predicted here suggests a mechanism of N- and C-crosslink recruitment, which could allow for the non-concurrent loading of N- and C-crosslinks. An improved understanding of this crosslink recruitment mechanism could aid the development of larger-scale predictive models of the mechanical behavior of native collagenous tissues, engineered tissues, and collagen-based materials.

H.1 (continued) – Manuscripts in preparation

Kwansa, A.L., Nelson, A.K., Sampson, A.C., Empson, Y.M., Snowhill, P.B., & Freeman, J.W. Extraction and reconstitution of acid-soluble collagens from enzyme-treated and alkaline-treated (ET/AT) bovine flexor tendon, (in preparation).

Abstract: Collagens are extracellular matrix proteins that are found in nearly all eukaryotic life except for plants and protozoa; collagens are especially abundant in mammals. Type I collagen is the most prevalent of this protein family; in mammals and other vertebrates, it is found in connective tissues such as bone, ligament, skin, and tendon. Collagens have been commercially isolated and purified for use in research, medicine, and the food and cosmetics industries. Most of these collagens have been extracted from bovine tissues such as tendons and skin; however, the yield of collagen type I from bovine tissues is usually quite low, in particular, from mature tissues. The aims of this study were to (1) assess if dilute acidic solutions might further improve the yield of collagen type I from enzyme- and alkaline-treated (ET/AT) mature bovine flexor tendon with and without sonication, (2) fabricate reconstituted collagen films, and (3) characterize the mechanical properties of these films. It was observed that (1) conventional stirring of the ET/AT tendon in dilute hydrochloric acid yielded insufficient quantities of collagen protein (0.095% yield w/w), (2) sonication provided a marked improvement to this yield ($2.10 \pm 1.37\%$ yield w/w), (3a) in vitro storage of reconstituted rat tail tendon collagen films increased their total slope (a quantity related to Young's modulus) and elastic fraction (a measure of elastic energy storage), and (3b) rat tail tendon collagen films exhibited a significantly higher ultimate tensile strength, higher Young's modulus, lower swelling ratio, and greater thickness than the ET/AT bovine flexor tendon, suggesting that the collagen reconstitution was more effective for rat tail tendon collagen compared to the ET/AT mature bovine flexor tendon collagen isolated with the aid of sonication.

H.2. First-authored publications

Kwansa, A.L., Empson, Y.M., Ekwueme, E.C., Walters, V.I., Freeman, J.W., & Laurencin, C.T. (2010). Novel matrix based anterior cruciate ligament (ACL) regeneration. *Soft Matter*, 6, 5016-5025. <http://dx.doi.org/10.1039/c0sm00182a>.

Abstract: Among the five ligaments in the knee the anterior cruciate ligament (ACL) is among the most important for stability and also the most commonly injured. Due to a lack of vascularization, the ACL has poor healing potential, therefore moderate to severe damage warrants medical intervention. Ligaments are complex, highly organized tissues; they are longitudinally arranged with a great deal of order that begins at the molecular level and carries through to the tissue level. The components of the ligament and their location and orientation heavily influence the tissue's mechanical behavior. ACL replacements have faced a variety of limitations that prevented their extensive use, including implant fatigue or fraying of the device. In the face of these problems, investigators have begun to examine a variety of matrix based techniques to create options for ACL repair, replacement, and regeneration. This article will discuss ACL structure and mechanics, past replacement options and their limitations, and recent tissue engineered options for ACL repair. These techniques employ a wide variety of designs, materials, and methods to heal damaged ACL tissue or regenerate lost tissue in order to regain full ACL strength and mechanics.

H.2 (continued) – First-authored publications

Kwansa, A.L., & Freeman, J.W. (2010). Elastic energy storage in an unmineralized collagen type I molecular model with explicit solvation and water infiltration. *Journal of Theoretical Biology*, 262(4), 691-697. <http://dx.doi.org/10.1016/j.jtbi.2009.10.024>.

Abstract: Collagen type I is a structural protein that provides tensile strength to tendons and ligaments. Type I collagen molecules form collagen fibers, which are viscoelastic and can therefore store energy elastically via molecular elongation and dissipate viscous energy through molecular rearrangement and fibrillar slippage. The ability to store elastic energy is important for the resiliency of tendons and ligaments, which must be able to deform and revert to their initial lengths with changes in load. In an earlier paper by one of the present authors, molecular modeling was used to investigate the role of mineralization upon elastic energy storage in collagen type I. Their collagen model showed a similar trend to their experimental data but with an over-estimation of elastic energy storage. Their simulations were conducted in vacuum and employed a distance-dependent dielectric function. In this study, we performed a re-evaluation of Freeman and Silver's model data incorporating the effects of explicit solvation and water infiltration, in order to determine whether the model data could be improved with a more accurate representation of the solvent and osmotic effects. We observed an average decrease in the model's elastic energy storage of $45.1\% \pm 6.9\%$ in closer proximity to Freeman and Silver's experimental data. This suggests that although the distance-dependent dielectric implicit solvation approach was favored for its increased speed and decreased computational requirements, an explicit representation of water may be necessary to more accurately model solvent interactions in this particular system. In this paper, we discuss the collagen model described by Freeman and Silver, the present model building approach, the application of the present model to that of Freeman and Silver, and additional assumptions and limitations.

H.3. Co-authored publications

Walters, V.I., **Kwansa, A.L.**, & Freeman, J.W. (2012). Design and analysis of braid-twist collagen scaffolds. *Connective Tissue Research*, 53(3), 255-266.

<http://dx.doi.org/10.3109/03008207.2011.634532>.

Freeman, J., Walters, V.I., & **Kwansa, A.L.** Ligaments, biomaterials, and tissue-engineering opportunities. (2012). In J.O. Hollinger, *An Introduction to Biomaterials*, 2nd Edition. Boca Raton, FL: CRC Press. Ch. 32, p. 565-582.

<http://www.crcpress.com/product/isbn/9781439812563>.

Sopakayang, R., De Vita, R., **Kwansa, A.L.**, & Freeman, J.W. (2012). Elastic and viscoelastic properties of a type I collagen fiber. *Journal of Theoretical Biology*, 293, 197-205.

<http://dx.doi.org/10.1016/j.jtbi.2011.10.018>.

Freeman, J.W., & **Kwansa, A.L.** (2008). Recent advancements in ligament tissue engineering: The use of various techniques and materials for ACL repair. *Recent Patents on Biomedical Engineering*, 1(1), 18-23. <http://eurekalect.com/95052/article/recent-advancements-ligament-tissue-engineering-use-various-techniques-and-materials>.