SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Individual Study Semester – MBB 481-5 Honours Thesis

Department of Molecular Biology and Biochemistry

Faculty of Science

Spring 2018

Exploring the Molecular Properties of Collagen Type IV with Atomic Force Microscopy

by

Alaa Al-Shaer

13-04-18

11:30 am at SSB 6178

Senior Supervisor: Nancy R. Forde

Committee members: Lisa Craig and Jenifer Thewalt

Abstract

Collagen type IV is a network-forming collagen that provides support and anchorage to cells. Its basic structural unit is a 410 nm long and 1.5 nm in diameter triple helix, with natural discontinuities in the triple-helical defining Gly-X-Y sequence. The C-terminal globular domain (NC1) in a collagen IV molecule plays an important role in forming networks, and has recently been reported to be structurally triggered by chloride ions to form hexamers outside the cell. How this hexamer assembles *in vitro* remains unknown. Here, I aim to use atomic force microscopy (AFM) to investigate the molecular basis of collagen type IV network assembly by studying the effects of different solvent conditions on the stability of the NC1 domain. Studying the dissociation of this hexametric domain can shed light onto how it assembles in solution and under what ionic conditions. The flexibility of the collagen type IV molecule is also investigated by performing statistical analysis of AFM-imaged chains and estimating persistence length, a mechanical property that quantifies the flexibility of a polymer. Here, I investigate the effects of triple helix interruptions on the flexibility of the molecule, by comparing collagen type IV to other fibrillar collagens that are continuously triple-helical. In addition, I determine a position-dependent flexibility profile of the molecule showcasing the effects of over-lapping interruptions, from a $\alpha1(IV)]_2$–$\alpha2(IV)$ mouse collagen type IV, on the persistence length.

Introduction

Collagen is the most abundant protein in the animal kingdom. It is a major structural component of our extracellular matrix and connective tissue, comprising one-third of the total protein in the human body [1]. Collagen contributes to the mechanical stability, organization, and shape of a wide variety of tissues; where different types of collagen are found in different types of tissues. There are twenty-eight distinct collagen types that have been reported in the human body, with five classes of higher-order structures identified [1]. The most prevalent class is fibrillar collagen which has a unique hierarchical structure. Briefly, collagen molecules self-assemble into fibrillar-nanostructures which later form fibers that become the predominant structural protein in tissues. Fibrillar collagens include collagen type I which is the most abundant type of collagen that makes up over 90% of the total collagen in the human body [1]. Collagen types IV, VI, VIII, and X are network-forming collagens that align in lateral associations to form open networks rather than fibrils [2]. Molecular defects in collagen are linked to many human diseases such as osteogenesis imperfecta, epidermolysis bullosa, and Alport syndrome [3, 4]. However, mechanical descriptions of the different types of collagens as biopolymers still remains unresolved. Having a profound knowledge of their mechanical properties may improve our understanding of how molecular defects correlate with altered higher-order structures of collagen.

Collagen type IV, henceforth collagen IV, is predominantly found in basement membranes (BMs) where numerous other extracellular matrix proteins work together to provide tensile strength surrounding epithelial, endothelial, Schwann, and muscle cells. How these various components assemble into an organized network with regulatory functions remains unclear. Collagen IV scaffolds tether diverse macromolecules, including proteoglycans and growth factors,

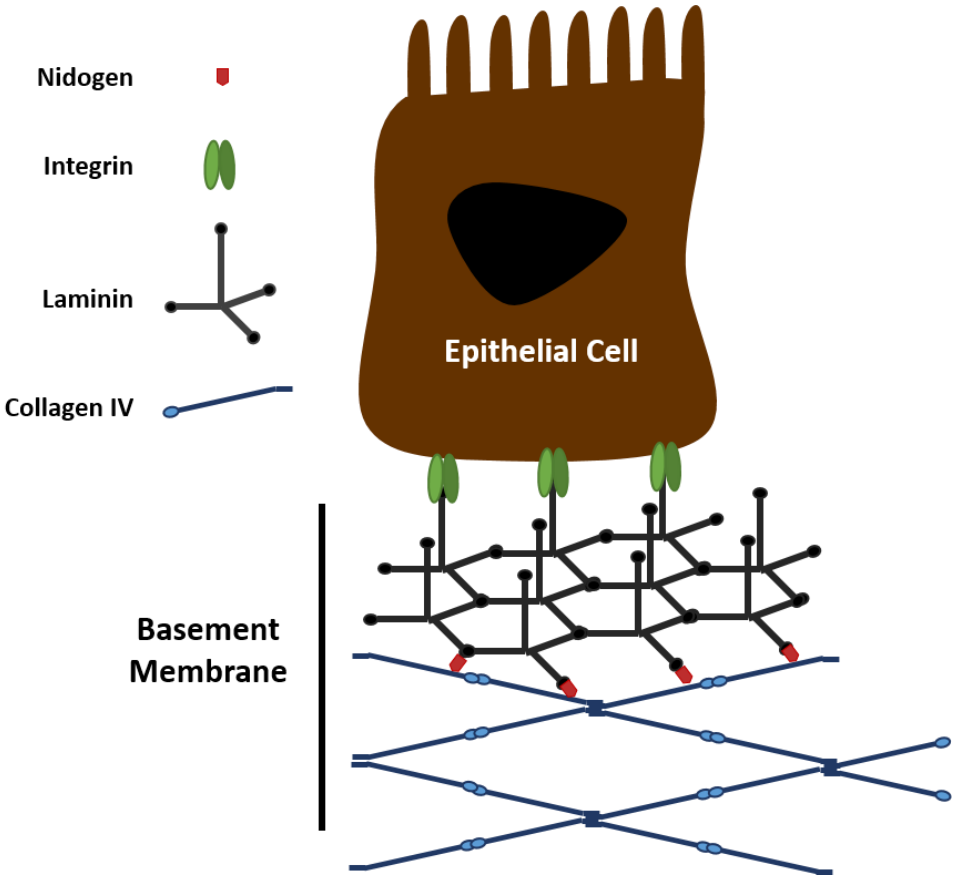

Fig. 1. **Cell-matrix interactions in basement membranes.** Collagen IV scaffolds are bridged to laminin networks through nidogen, a basal lamina glycoprotein. Laminin networks are anchored to the cells through integrins, transmembrane receptors that facilitate cell-matrix adhesion.

and serve as a ligand for cell-surface receptors including laminin (see Fig. 1) [5]. Collagen IV networks form a complex permeable barrier that functions as a molecular filter, and are involved in cell-matrix interactions which provide support and anchorage for cells and tissues [2].

The assembly of collagen IV protomers, the basic structural units, starts in the endoplasmic reticulum compartment inside the cell. Three left-handed α-chains consisting of about 1700 amino acids are synthesized independently and coil around each other to form a right-handed triple helix, forming the principal building block for supramolecular collagen IV networks in BMs. Proper

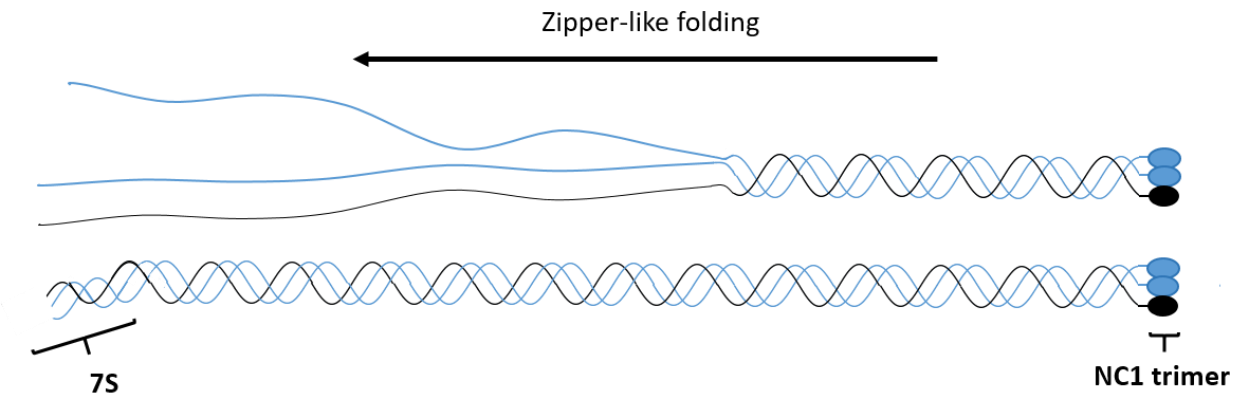

Fig. 2. **Structure of collagen IV protomer.** Three collagen IV α-chains fold forming a heterotrimer. Each protomer has a 7S domain at the N-terminus ~ 30 nm long, a long triple helical segment that runs for ~ 360 nm, and a noncollagenous NC1 trimer at the C-terminus.

alignment and association of the α-chains is an important prerequisite for correct folding of the collagen protomer. The C-terminal globular domain in an α-chain serves as the interaction region where the folding mechanism is initiated [6]. Collagen folding is described as proceeding from the C- to the N- terminus in a zipper-like fashion [7], as shown in Fig. 2. Six different collagen IV α-chains have been found with only three sets of heterotrimers identified in our tissues: $[\alpha1(IV)]_2$–α2(IV), α3(IV)–α4(IV)–α5(IV), and $[\alpha5(IV)]_2$–α6(IV) [4]. The central triple helical segment in a collagen IV protomer is 360 nm long with more than 20 interruptions in the triple-helix-defining Gly–X–Y amino acid sequence of each α-chain (where X and Y are often proline and

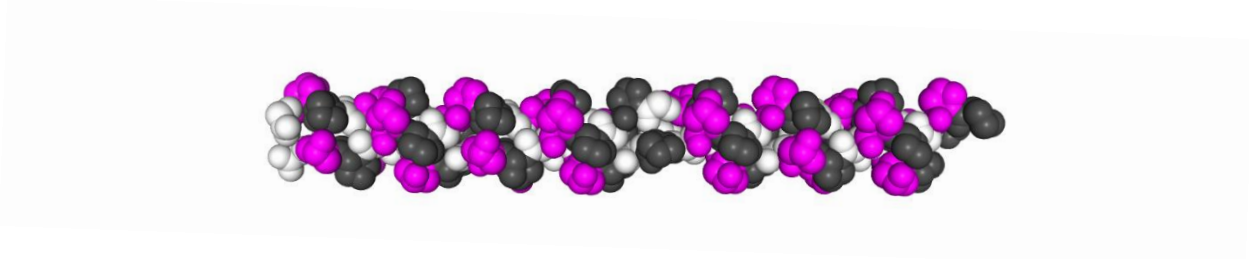

Fig. 3. **Segment of collagen triple helix.** From x-ray crystallographic studies of a collagen model peptide with sequence $(Gly\text{-}Hyp\text{-}Pro)_5$-Gly-Pro-$(Gly\text{-}Hyp\text{-}Pro)_4$ that contains a central Gly-Pro-Gly (white-black-white) interruption in the consensus collagen sequence. From 1EI8 file of Protein Data Bank.

hydroxyproline) (see Fig. 3). These interruptions lead to a delay in triple helix folding [8], and have been reported to contribute to the overall flexibility of the structure [9]. Glycine residues fit inside the triple helix, facing the interior, and are required for compact structure of the helix. The side chains of the X and Y amino acids are exposed to the exterior of the chain, where imino (proline)-rich regions help to form a tighter triple helix [1]. The three α-chains are hydrogen bonded to each other and are cross-linked through disulfide bonds [6]. In addition, a collagen IV protomer has two distinct end-domains that play an important role in forming higher-order structures. Collagen IV's assembly into larger networks is triggered outside of the cell, however, much remains to be learned about the molecular basis of this process.

Collagen IV networks are known to be formed via two distinct interactions, as shown in Fig. 4. The cysteine- and lysine-rich N-terminal 7S domain, which is approximately 30 nm long, forms tetramers by aligning four collagen IV protomers. The tetrameric 7S domain is joined together through interchain disulfide bridges and lysyl cross links [10]. It is also heavily glycosylated making it resistant to collagenase activity, and easily purifiable [11]. This posttranslational modification plays an important role in the assembly and stabilization of collagen IV networks [12]. At the other end, the C-terminal noncollagenous 1 (NC1) domain connects two

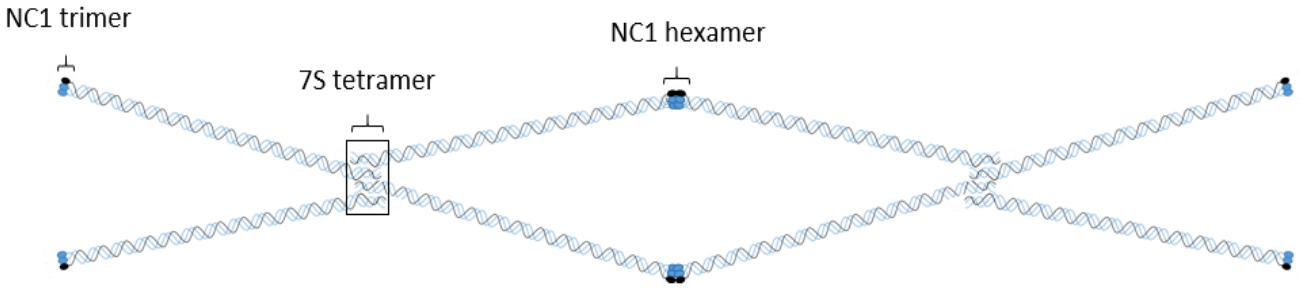

Fig. 4. **Collagen IV network assembly interactions.** Two protomers self-associate at their NC1 domains, forming an NC1 hexamer, and four protomers are cross-linked at the 7S domain.

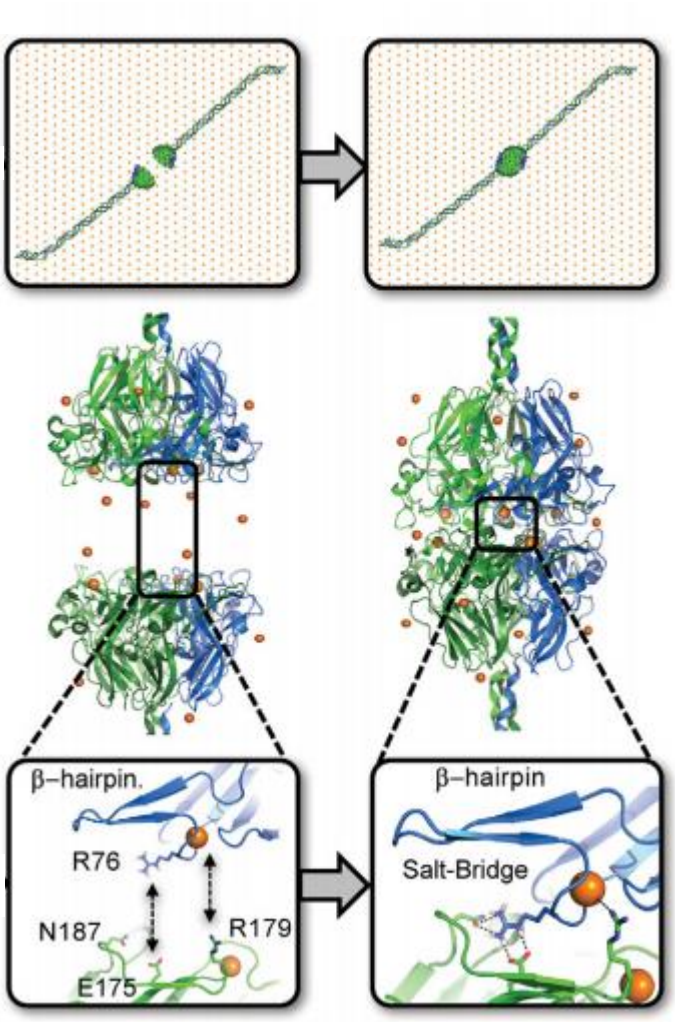

Fig. 5. **The molecular switch triggered by chloride ions inducing hexamer formation of NC1 domains.** Chloride ions disrupt the intramolecular salt bridges by electrostatic screening. The R76 orients itself to an opposing NC1 trimer, forming intermolecular salt bridges with E175 from an opposing NC1 trimer. From [13].

collagen IV protomers by the end-on association of protomers in the extracellular space. Inside the cell, the NC1 domain of each α-chain in a collagen IV protomer laterally associates with others to form NC1 trimers, which is the onset of alignment promoting triple helix folding [6]. The trimeric NC1 domains dimerize extracellularly and form hexamers as shown in Fig. 4. The molecular basis behind what triggers the formation of collagen IV networks outside of the cell has been studied for

over 30 years, with recent literature reporting a surprising mechanism that controls collagen IV assembly [13]. This recent work shows that chloride ions trigger the assembly of collagen IV networks by triggering a molecular switch in the NC1 domains and inducing hexamer formation [13]. Chloride disrupts intramolecular salt bridges, yielding a conformational change that allows the formation of intermolecular interactions with an opposing NC1 trimer (Fig. 5 [13]). This finding provides insight into the mechanisms of collagen IV assembly in the BMs, where chloride ions exist at a higher concentration extracellularly than intracellularly. Specifically, it has been shown that chloride ions trigger network assembly via the NC1 domains, where the chloride ions are not required for trimerization but rather for hexamerization, associating two protomers together. The NC1 trimer is formed in the endoplasmic reticulum compartment, where chloride concentration is very low. On the outside of the cell, protomers are exposed to high chloride concentration which induces stable head-to-head hexamer assembly.

In BMs, covalently cross-linked sulfilimine (S=N) bonds are formed between adjoining collagen IV protomers via the NC1 hexamers [14]. This post-translational modification is essential for the stability of collagen IV networks and also tissue development [15]. The sulfilimine bonds are covalent cross-links that occur between a methionine sulfur and hydroxylysine nitrogen in the trimer-trimer interface of NC1 hexamers, as shown in Fig. 6 [14]. A BM enzyme called peroxidasin is responsible for catalyzing the formation of the sulfilimine bonds using a halogenation cycle, where bromide is a required cofactor. Bromide, converted to hypobromous acid, forms a bromosulfonium-ion intermediate that energetically selects for sulfilimine formation. The presence of these cross-links protects the NC1 hexamer from dissociation in a chloride-free environment. There is a maximum of six cross-links in an NC1 hexamer, where two cross-links join opposing α-chains of two protomers. One can anticipate, upon denaturation of a cross-linked

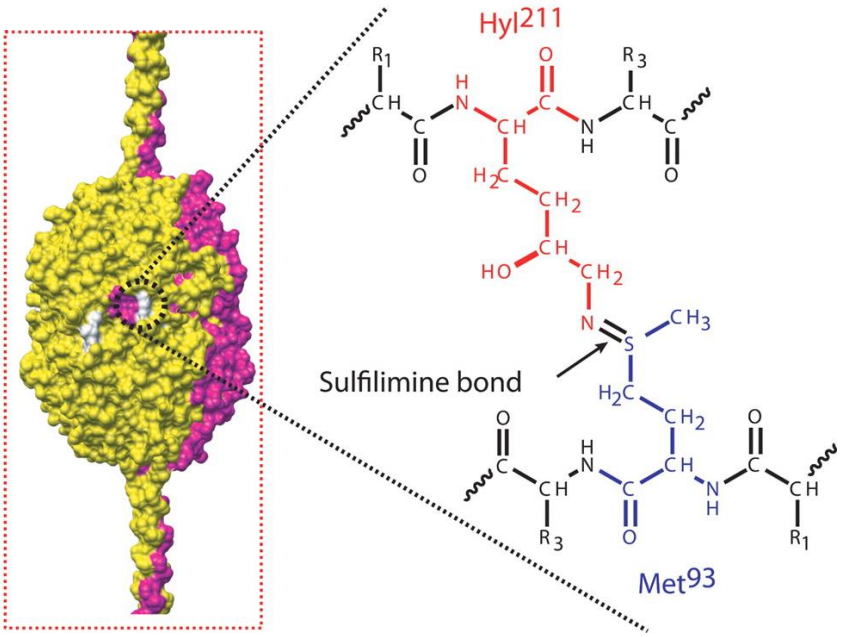

Fig. 6. **Chemical structure of sulfilimine cross-links.** The sulfilimine bond connects the side chains of Met^{93} and Hyl^{211}, as shown for the α1(IV)-α1(IV) NC1 dimer in yellow and the α2(IV)-α2(IV) NC1 dimer in magenta. Adapted from [14].

hexamer, dissociation into cross-linked dimeric subunits (D) and un-cross-linked monomeric subunits (M), as shown in Fig. 7 [15]. The extent of cross-linking varies in different tissues in animals. For example, NC1 hexamers from bovine lens capsule basement membrane (LBM) are ~ 10 % cross-linked, while bovine placenta basement membrane (PBM) is ~ 50 % cross-linked [21].

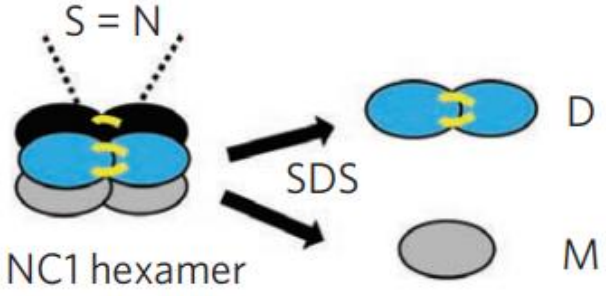

Fig. 7. **A collagen IV NC1 hexamer with sulfilimine cross-links.** Upon addition of SDS, the hexamer dissociates into dimeric subunits (D) and un-cross-linked monomeric subunits (M). From [15].

Sulfilimine bond formation endows mechanical strength to collagen IV scaffolds, and in the case of humans, confers immune privilege to the collagen IV antigen of Goodpasture autoimmune disease [17].

A mechanical description of a collagen molecule as a biopolymer remains a challenge, with a significant amount of disagreement reported in the literature and with only a few studies conducted at the single-molecule level. Atomic Force Microscopy (AFM) is a common experimental method used to study biopolymers because of its high-resolution imaging ability [18]. It uses a fine probe to image the nanoscale topological features of a surface, yielding a 3D profile of the surface. The AFM uses a laser-beam detection system to record the deflections of the cantilever, as shown in Fig. 8. AFM imaging can be used to look at ideal chain conformations in 2D and characterize structural properties. The flexibility of a polymer can be described by the persistence length which is defined as the length over which tangent vector correlations are lost. To calculate the persistence length of collagen, the worm-like chain (WLC) model is used. This is a good model that well-describes stiff polymers of which flexibility is due to thermal fluctuations about a straight-line [19]. The WLC model has been used to characterize the flexibility of various collagen types. A recent study on the persistence length of different types of fibrillar collagen molecules reports that in a given solution condition the persistence length is not exceedingly variable [20]. In the case of a network-forming collagen such as collagen type IV, I hypothesize that the persistence length will be significantly different than that of fibrillar collagens. I attribute this to the triple-helical defining Gly-X-Y sequence in collagen type IV being interrupted in various positions, contributing to more flexible regions along its contour and its overall flexibility as a molecule.

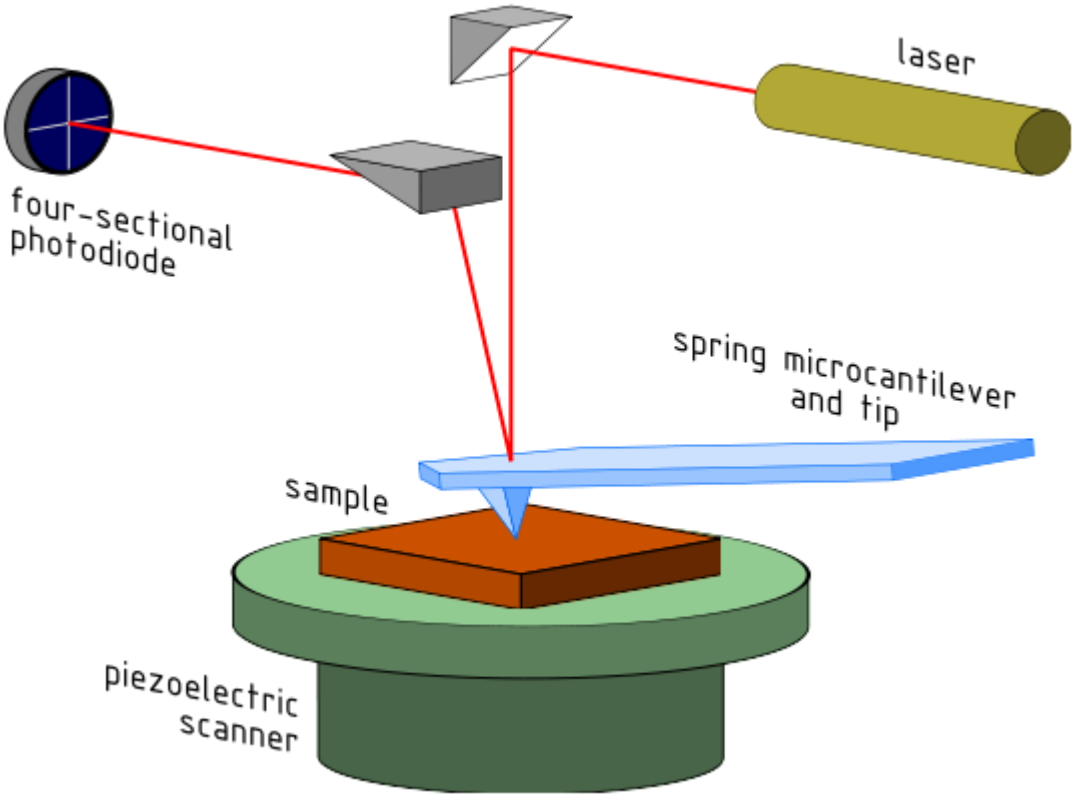

Fig. 8. **Schematic of AFM setup.** In tapping mode, the mechanical probe's (cantilever and tip) oscillation amplitude changes in response to surface topography. The cantilever deflects when a bump is encountered on the surface, and these deflections are monitored by a laser-beam detection system. (Wielgoszewski, 2011)

Although average persistence lengths provide us with valuable insight into the mechanical properties of different collagen types, having a position-dependent persistence length of collagen along its contour would significantly add to our understanding of this protein. Average persistence lengths treat collagen as a homogeneous polymer, which is not a reasonable assumption considering the variability of the amino acid composition along its sequence. To analyse collagen as an inhomogeneous polymer has its challenges, especially when it comes to fibrillar collagens where the N- and C-terminal ends of the molecules are not easily distinguished. By contrast, for collagen type IV, the C-terminal end has a globular NC1 domain that makes its ends easily

distinguishable with AFM. The presence of this globular domain allows us to align the chains and perform statistical analysis in a position-dependent manner. Furthermore, mapping the persistence length of collagen type IV and will surely provide insight into the effects of these triple-helical interruptions on the molecular flexibility.

In this thesis, I use AFM to explore the molecular basis of collagen type IV network assembly with a focus on the stability of NC1 domains. I study the effects of different solvent conditions on the stability of the NC1 domains from two sources: bovine lens capsule basement membrane and bovine placenta basement membrane. These NC1 domains have different structural characteristics and behave differently under certain solution conditions. In addition, I explore the mechanical properties of collagen type IV by extracting a position-dependent flexibility profile of the molecule. The flexibility of the collagen type IV molecule is investigated by performing statistical analysis of AFM-imaged chains. Here, I show the effects of the triple helix interruptions on the molecular flexibility, and I compare the average persistence length of collagen type IV to fibrillar collagens. These studies on the structural and mechanical properties of collagen contribute to our understanding of how changes at the molecular-level can presumably affect higher-order structure assembly of collagen.

Materials and Methods

Collagen Type IV and Isolated NC1 Sources

All samples have been kindly provided by S. Budko and B. G. Hudson (Vanderbilt University). Engelbreth-Holm-Swarn (EHS) tumor heterotrimeric [α1(IV)]$_2$–α2(IV)] collagen type IV was grown subcutaneously in mice and purified at a final concentration of 113 μg/mL in 0.5 M acetic acid and stored in 4 ºC (from Klaus Kuhn). NC1 hexamers from bovine lens capsule basement membrane (LBM) and bovine placenta basement membrane (PBM) were isolated by treating the collagen IV matrix with bacterial collagenase enzyme [21]. Native LBM NC1 hexamers are ~10 % cross-linked, whereas native PBM NC1 hexamers are ~50 % cross-linked. The LBM NC1 hexamers used in this study were re-assembled from un-cross-linked dimers, and the PBM NC1 hexamers were re-assembled from heavily cross-linked dimers. Both NC1 samples and 7S domains were at a final concentration of 100 μg/ml, in tris-buffered saline (TBS) [50 mM Tris-Cl, 150 mM NaCl; pH 7.6] and sodium azide for storage. These domains come from [α1(IV)]$_2$–α2(IV)] heterotrimers, and are also stored in 4 ºC.

SDS-PAGE gel electrophoresis

A 12% resolving gel was casted, with a 5% stacking gel. The samples were prepared for SDS-PAGE by adding 4X protein loading dye, and boiling for 10 min at 80ºC. The gel was loaded with ~ 2 μg of the protein samples, and 10 μL of the Fermentas PageRuler™ Unstained Protein Ladder. The gel was run at 220V for 1.5 hours, until the dye-front was 0.5 cm above the bottom edge of the gel. The resolved gel was stained with Coomasie Blue R-250 for 30 min, and unstained with destaining solution (4% methanol, 1 % acetic acid) for 30 min. The gel was transferred into water and left overnight before it was imaged.

Atomic Force Microscopy Imaging

Imaging was done with an Asylum Research MFP-3D atomic force microscope using AC tapping mode in air. AFM tips with a 160 kHz resonance frequency and 5 N/m force constant (MikroMasch, HQ: NSC14/AL BS) (325 kHz, 40 N/m from MikroMasch, HQ: NSC15/AL BS for collagen type I images) were used. To acquire images, the AFM uses a laser-beam detection system to record the deflection of the laser-beam from the back of the cantilever, which is caused by the forces produced between the mechanical probe and the surface. In this mode, the AFM tip is oscillating up and down touching the surface at short times in discrete locations (pixels).

Sample Preparation for AFM

The sample preparation is done on mica (Highest Grade V1 AFM Mica Discs, 10 mm, Ted Pella), an atomically-flat surface, and is relatively straightforward. The protein of interest is diluted into a desired solution and 50 μL is deposited onto freshly cleaved mica. The excess unbound proteins are washed with either a salt solution or ultra-pure water. The mica is then dried under filtered air. All proteins were imaged under dry conditions, and the solution conditions of the samples refer to the conditions in which they were deposited onto mica.

Chain Tracing and Analysis

The tracing algorithm program used to study the mechanical properties of collagen chains is a custom-built MATLAB code named SmarTrace, which was developed by Naghmeh Rezaei (former PhD student, [22]). The analysis starts off by tracing a chain with a few user-defined points. The SmarTrace program then successfully detects the backbone of the chain even in noisy environments, with sub-pixel resolution. The traced chains are sampled and later used to perform position-independent and position-dependent statistical analysis.

Results

I. Exploratory Mission on End-Domain Interactions

Atomic force microscopy is used to explore the full-length collagen type IV protomer, as well as isolated end-domains. The full-length collagen IV is from *Mus musculus* (mouse), while the isolated 7S and NC1 domains are from *Bos taurus* (bovine). The samples are imaged on mica – a negatively-charged flat and smooth surface. The molecules adhere to the surface via electrostatic interactions. Collagen is positively charged in acidic pH, and thus sticks to mica without the need to do any surface chemistry manipulation.

An AFM image of $\alpha1(IV)]_2$–$\alpha2(IV)$ collagen IV protomers is shown in Fig. 9, where individual molecules with a triple helix chain and a distinguishable globular NC1 domain can be seen. NC1 hexamers of collagen IV dimers and NC1 trimers of an individual protomer are easily identified, as denoted by the blue and black arrows respectively. This sample was prepared in the presence of chloride ions (100 mM KCl and 1 mM HCl), which was enough to induce partial collagen IV oligomerization through NC1 domains. The contour lengths of these protomers is close to the expected 400 nm, which includes the triple helical region and the 7S domain.

The 7S domain in collagen type IV is a short N-terminal triple-helical domain that is approximately 150 amino acid residues. It is responsible for forming the tetrameric complexes in the collagen IV network. The first 15 residues of the 7S domain from α1 (bovine) form a non-triple-helical region that contains four cysteine residues, and these cysteine residues are also conserved in α1 (mouse). This first region appears to be functionally important in forming the disulfide cross-links in the 7S tetramer after the alignment of the four triple helices which are oriented in an anti-parallel fashion. The amino acid sequence of the 7S domain and collagenous

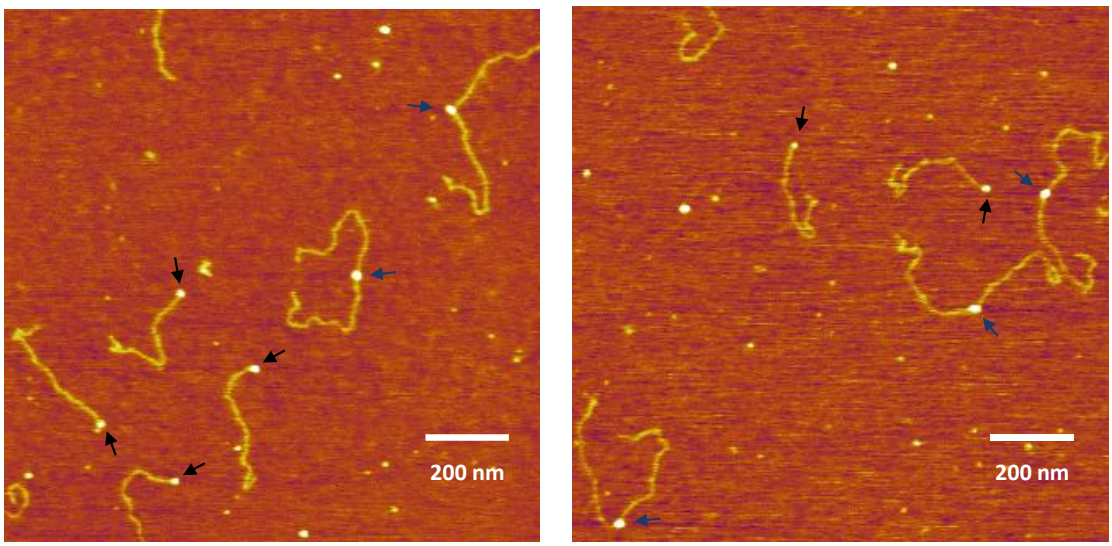

Fig. 9. **AFM image of collagen IV protomers.** The long triple helical region is clearly visible, with different configurations shown on the mica surface. The C-terminal globular NC1 domain is easily distinguishable, as it appears as a "blob" at the end of the triple helix. Black arrows point to NC1 trimers of individual collagen IV protomers. Blue arrows point to NC1 hexamers of collagen IV dimers. Sample deposited from 100 mM KCl and 1 mM HCl and 1 mM Acetic Acid (pH= 3.2). The stock solution (110 µg/mL of collagen IV in 5 M Acetic Acid) was diluted straight into 100 mM KCl and 1 mM HCl, and the final concentration of collagen IV is 0.2 µg/mL and was incubated on the mica surface for 20s followed by a water wash.

domain of α2 (bovine) has not yet been determined. However, the 7S domain amino acid sequence of α2 from mouse is available, and appears to also have a non-triple-helical region in the beginning of the sequence, that runs for 32 residues which is double that of α1 (bovine and mouse). The 7S domain of both α1 and α2 also appear to end with an interrupted region, 12 residues in α1 (bovine and mouse) and 14 residues in α2 (mouse). This interrupted region could be contributing to the flexibility of the collagen protomer at the N-terminus end, as seen in Fig. 9 where the ends appear to have hooks. In addition, the flexibility of this region causes the triple helices to avert from one another and arrange in a tetragonal array [23], favouring the formation of the supramolecular network [24].

AFM images of the isolated 7S domains from bovine placenta basement membrane (PBM) and lens capsule basement membrane (LBM) are shown in Fig. 10. What we can expect to see for these domains is either a spider or a stick and in this case we see the latter. Isolated 7S domains can appear to look like a spider due to the four triple helices projecting outward from the heavily cross-linked tetramer. I assume that for this particular sample, the 7S domains appear to look like sticks because they were heavily digested with collagenase, leaving little to no triple helical residues. However, in some images, we can see fainter strands near the ends of the sticks, which I believe could be the triple helices projecting outwards from the tetramer. This is particularity

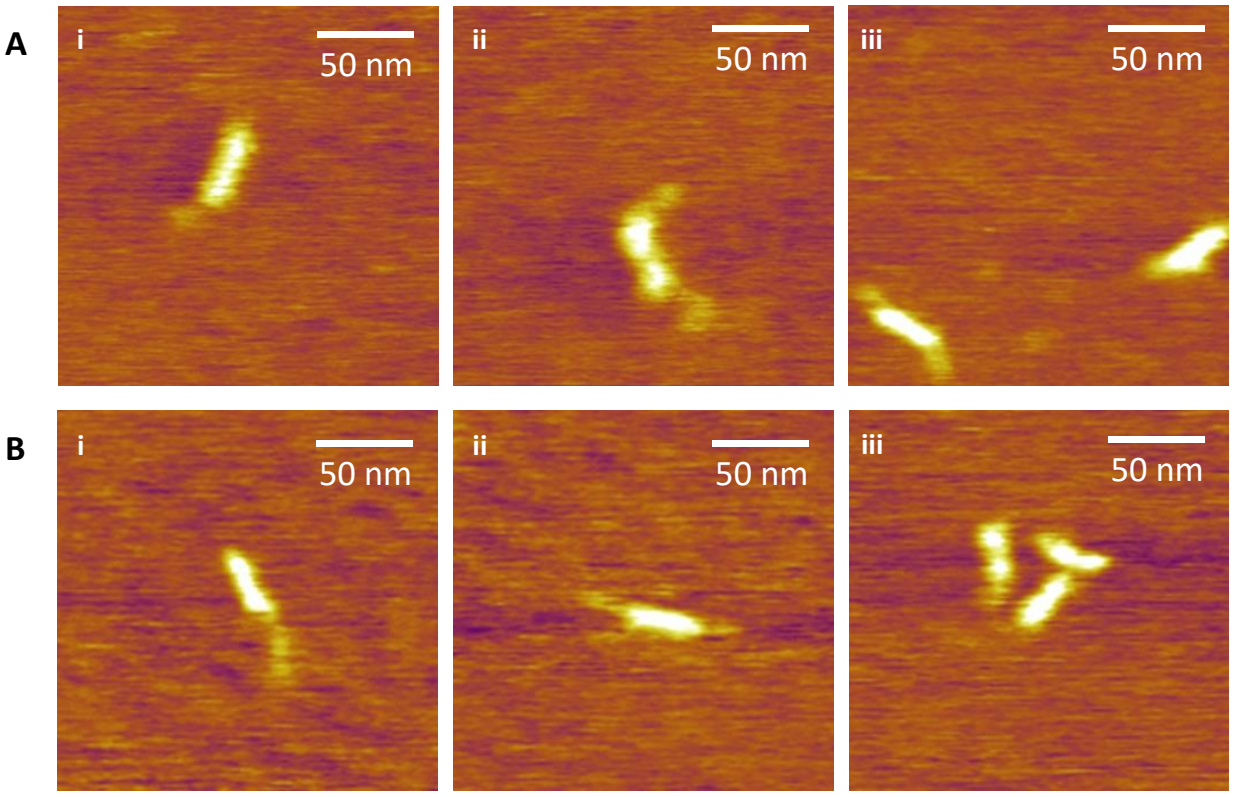

Fig. 10. **PBM and LBM isolated 7S domain. A**. 0.5 µg/mL of 7S domain from PBM in TBS. **B**. 0.5 µg/mL of 7S domain from LBM in TBS. The 7S domains appear to look like sticks, and close to the expected size 40 nm. The brighter area corresponds to the tetramer where four triple helices are cross-linked and held together. The triple helices projecting from the tetramer can roughly be seen in some cases.

apparent in panel A (ii) where we can see three strands coming out of the tetramer. Furthermore, comparing the structures from PBM and LBM, they appear to look very similar.

The NC1 domains from various basement membranes differ significantly by composition and degree of cross-linking [21]. The NC1 domains analyzed in this study come from $[\alpha1(IV)]_2$–α2(IV) heterotrimers that have been purified using a collagenase digestion, leaving only a few triple-helical residues, insufficient to stabilize and induce trimer formation [25]. The PBM NC1 hexamers have been re-assembled from heavily cross-linked NC1 dimers, and the LBM NC1 hexamers have been re-assembled from completely un-cross-linked monomers. As a matter of fact, the native PBM NC1 hexamers are 50% cross-linked, while the LBM NC1 hexamers are actually 10% cross-linked [21]. A schematic of anticipated PBM and LBM dissociation patterns is shown

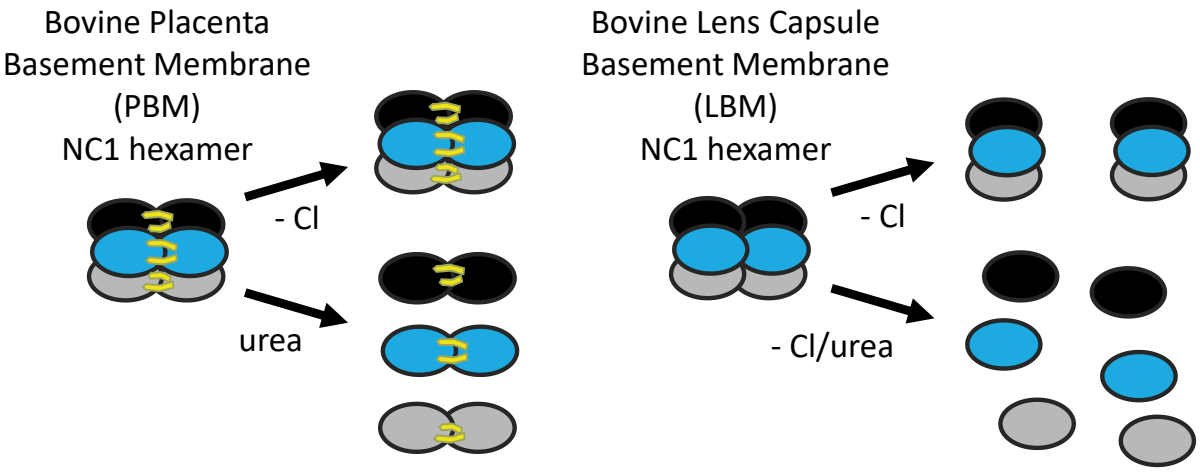

Fig. 11. **Anticipated PBM and LBM NC1 hexamer dissociation patterns. A**. Upon washing with urea, the PBM NC1 hexamer is expected to dissociate into three NC1 dimers. Upon removal of Cl-, the hexamer is expected to remain intact due to cross-links. **B**. Upon removal of chloride, the LBM NC1 hexamer is expected to dissociate into two NC1 trimers or completely dissociate to six NC1monomers. Upon washing with urea, the hexamer is expected to dissociate into six monomeric subunits. This figure was made by me and inspired from Fig. 7 [15].

in Fig. 11. Having some background information on these samples composition, I layout potential dissociation patterns of these hexamers in the presence and in the absence of chloride, as well as upon denaturation with urea. The cross-linking in a hexamer occurs between α1-α1 (in blue) and α2-α2 (in black), with a maximum of two cross-links between the NC1 domains from two protomers, and a maximum of six cross-links in a hexamer. To quantify the cross-linking in these samples, I ran an SDS-PAGE gel, shown in Fig. 12. We can see that for the PBM NC1 hexamers, there are some monomeric NC1 domains (~ 20 kDa) along with heavily cross-linked NC1 dimers (~ 40 kDa), stabilized by one or two sulfilimine cross-links. So, we can expect the PBM NC1 hexamer to dissociate into monomeric subunits because it is not a completely cross-linked

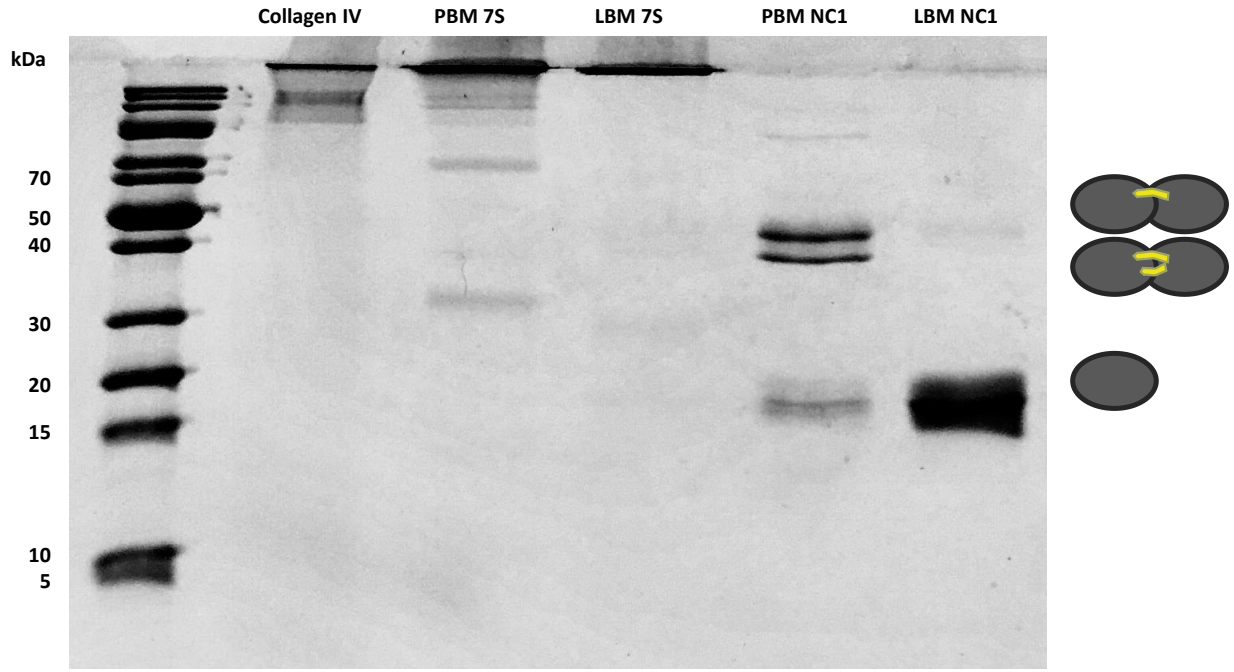

Fig. 12. **SDS-PAGE gel of collagen IV and end-domains from PBM and LBM.** To resolve collagen IV and 7S domains, I would need to cast a 7.5% gel. The expected sizes of the α1 and α2 chains from the full-length collagen type IV are roughly ~ 160 kDa, but cannot be seen in this gel. All PBM and LBM lanes were loaded with 75 μg/mL. NC1 domains from PBM are composed of heavily cross-linked dimers with a few monomers. NC1 domains from LBM are composed of mostly monomers, with little to no cross-linked domains (faint band). Dimers are stabilized by one or two sulfilimine cross-links.

hexamer. On the other hand, the LBM NC1 hexamer is almost completely monomeric, however, there is a faint band (~ 40 kDa) that could suggest there is a small fraction of cross-linked dimers.

I decided to use AFM imaging in different conditions to test the anticipated dissociation patterns shown in Fig. 11. AFM images of NC1 domains from PBM and LBM are shown in Fig. 13A and Fig.13B respectively. The PBM NC1 hexamers deposited from TBS (chloride-containing buffer) followed by washing the surface with water – removing the chloride – did not induce dissociation of the cross-linked hexamers. This was an expected result because the cross-links hold the hexamer together in the absence of chloride. In addition, there was little events of dissociation

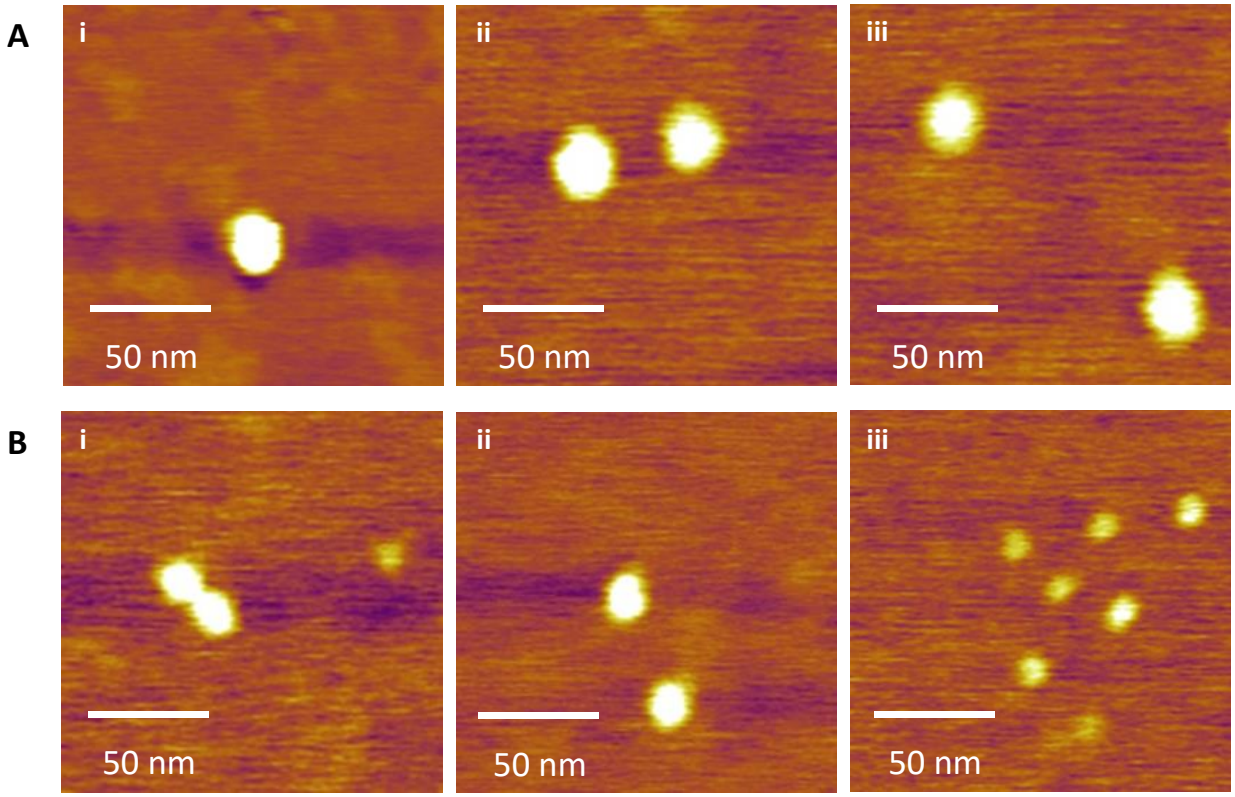

Fig. 13. **PBM and LBM NC1 hexamer dissociation upon removal of chloride.** Samples were prepared by depositing 0.3 µg/mL PBM/LBM in TBS on mica for 5s, followed by washing with water to remove unbound proteins and salts, including chloride. **A**. The PBM NC1 hexamers remains intact upon removal of chloride. **B**. The LBM NC1 hexamers dissociate into a mixture of two NC1 trimers (i) (ii) and six NC1 monomers (iii) upon removal of chloride.

of the PBM NC1 hexamer. This can be attributed to the fact that there are un-cross-linked NC1 domains in this sample, as shown in the gel. In contrast, almost all of the LBM NC1 hexamers dissociated on the mica surface upon removal of chloride. These hexamers are assembled from un-cross-linked monomers. We can see that the LBM hexamer dissociates into two trimers (i and ii) or six monomers (iii). The dissociation of this hexamer was expected since the chloride is needed to stabilize the trimer-timer interface which holds the hexamer together. We can expect that upon removal of chloride, the hexamer dissociates into two trimers. However, the NC1 trimer is not covalently linked and has been shown to be unstable in the absence of a sufficient triple helical domain [13]. So, it is rather surprising that there appears to be NC1 trimers on the surface. This could be the result of interactions between the NC1 domains and the charged mica surface. To address this issue, I decided to look at LBM NC1 hexamers in TNA (chloride-free buffer) [25mM TrisAcetate, 150mM NaAcetate; pH 7.4] and see if trimers are stabilized in the absence of chloride.

LBM NC1 hexamers were left overnight in TNA to induce complete dissociation of the hexamers prior to plating on mica. The results are shown in Fig. 14, where there are many scattered NC1 monomers, as well as some hexamers (black arrows) and trimers (blue arrows). It appears that NC1 trimers can rather be stabilized in solution, contradictory to what has been discussed in the literature [13]. We can see an intact hexamer (iii) and a fragmented hexamer (i) that appears to be at the onset of dissociation. The NC1 trimer is shown in (ii), where there are also three monomers adjacent to it. Here, I have shown that trimers and hexamers can be stable in solution prior to plating. Therefore, the trimers shown in Fig. 13B may not be the result of surface-effects but rather represent trimers that are independently stable in solution.

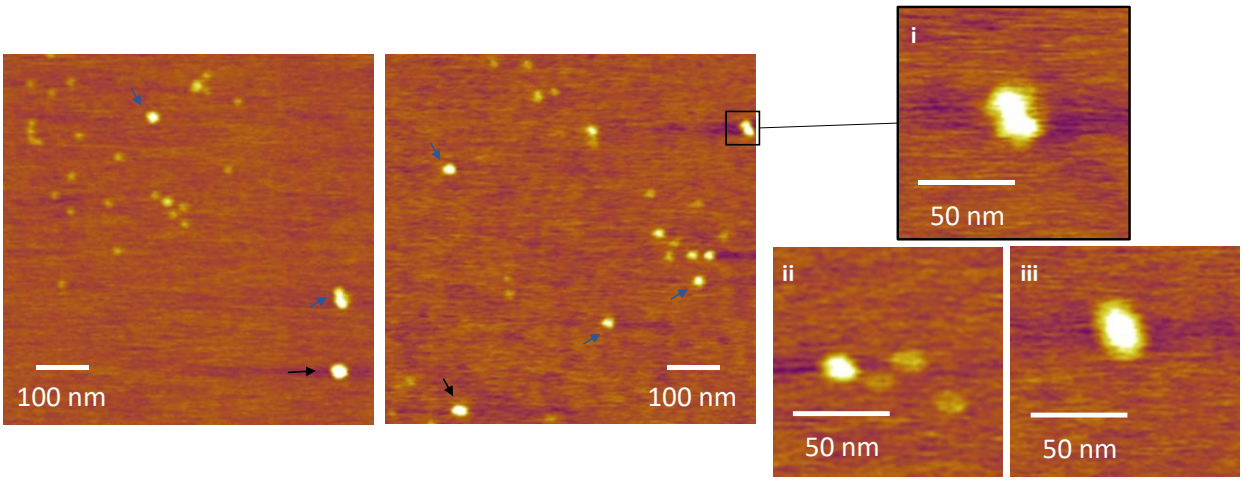

Fig. 14. **LBM NC1 hexamer dissociation in TNA.** Samples were prepared the night before plating. 0.5 µg/mL LBM in TNA was deposited onto mica for 5-10s, followed by washing with water. The LBM NC1 hexamers are mostly dissociated into monomers. There are a few trimers (ii) and hexamers (iii) as denoted by the blue and black arrows respectively.

Next, I decided to look at the effects of urea on the PBM and LBM NC1 hexamer, where the expected dissociation patterns are discussed in Fig. 10. The results are shown in Fig 15, where PBM and LBM NC1 hexamers were deposited onto mica in TBS, followed by 1 M urea incubation on the mica surface. Indeed, the urea induced partial (i) and complete (ii) dissociation of the PBM NC1 hexamer into three dimers, as shown in Fig. 14A. The 1 M urea incubation for one minute on the mica surface was not enough to induce total dissociation of all hexamers, as seen in (iii) and many other cases not shown here. Lastly, the LBM NC1 hexamers were mostly completely dissociated upon urea treatment, as shown in Fig. 15B. There was little sign of intact hexamers, as denoted by the arrow in (i). These results show that the urea induces dissociation of the NC1 hexamers without denaturing the protein domains completely. The domains still have an intact shape, and this can be attributed to the fact that the NC1 domains themselves are held together compactly by intramolecular disulfide cross-links, contributing to the structures rigidity.

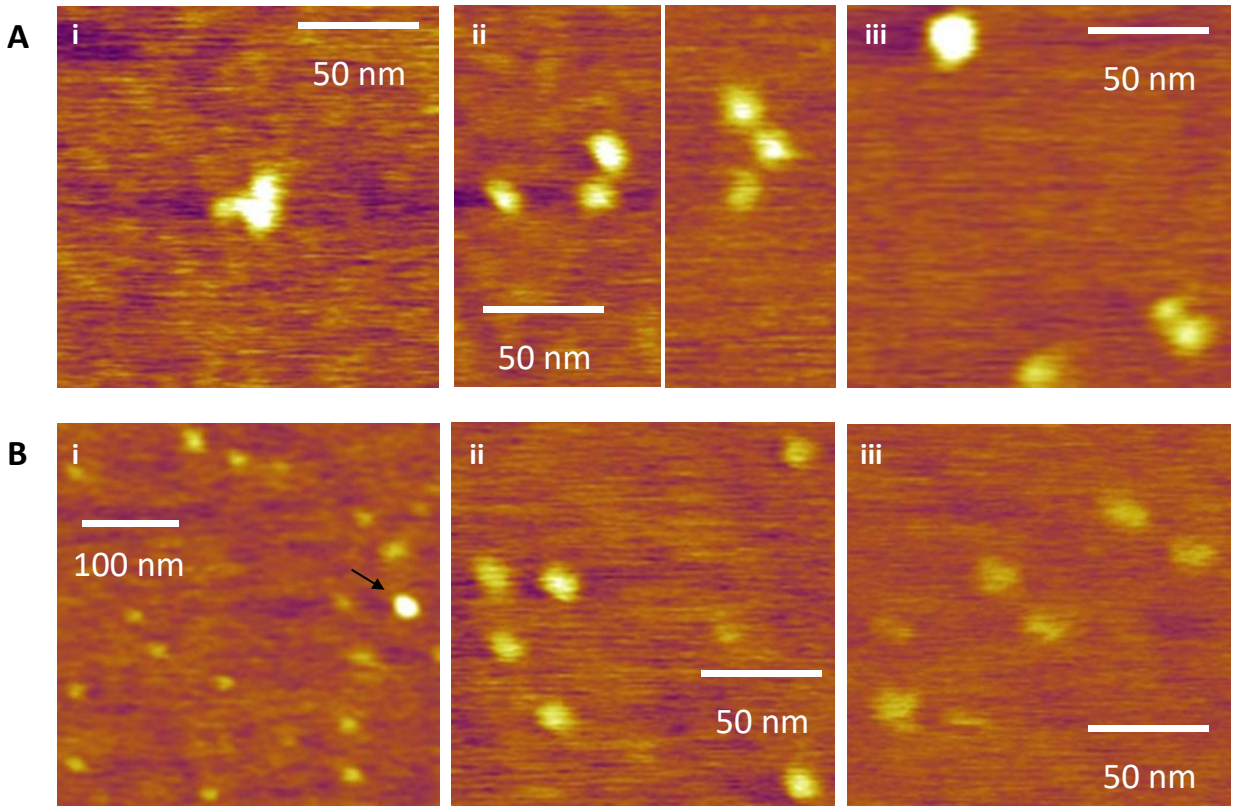

Fig. 15. **PBM and LBM NC1 hexamer dissociation upon urea treatment.** Samples were prepared by depositing 0.3 µg/mL PBM/LBM in TBS on mica for 5s, followed by 1 min 1 M urea incubation, and washing with water. **A**. The PBM NC1 hexamers dissociate into three dimers as expected, shown in (i), (ii) and (iii). There are also intact hexamers (iii) that have not dissociated. **B**. The LBM NC1 hexamer dissociates into monomers mostly, and there are only a very few intact hexamers, as denoted by the arrow in (i).

II. Position-Dependent Flexibility Profile of Collagen Type IV

To study the mechanical properties of collagen, I used a tracing algorithm developed by Naghmeh Rezaei [26]. In this algorithm, the backbone of a collagen molecule is traced from an AFM images to obtain coordinates along the chain. A few user-defined inputs along the chain are sufficient enough for the backbone contour identification, as shown in Fig. 16. The chain tracing algorithm provides conformational analysis of the chains traced, allowing for the implementation of polymer physics tools to determine structural properties of the molecule. The worm-like chain (WLC) model is used in this work to fit the data, collected from SmarTrace, and estimate a

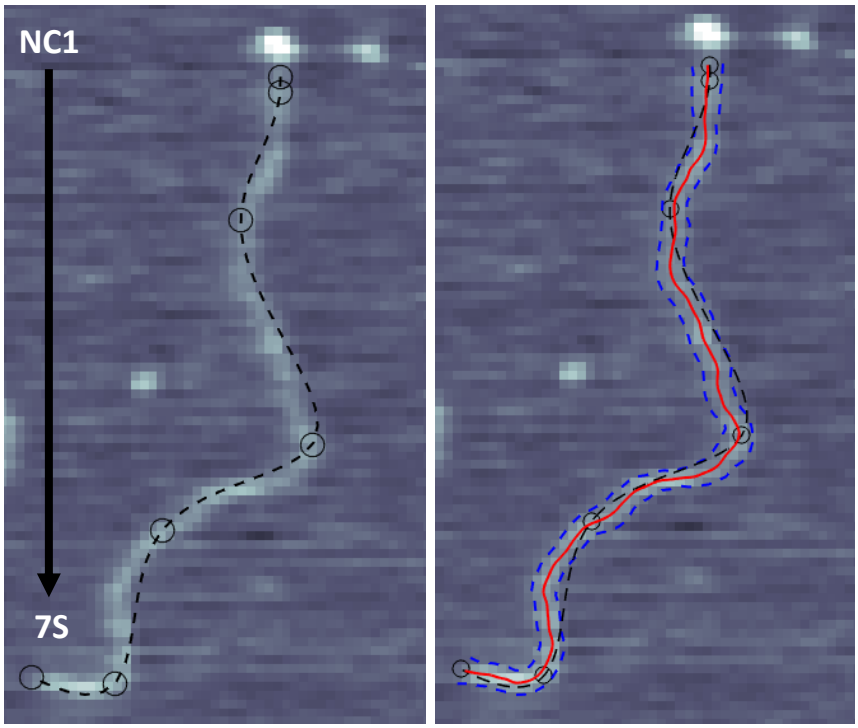

Fig. 16. **Tracing collagen chains using SmarTrace chain tracing program.** A few user-defined pointed along the chain (black cirlces) are sufficient enough for the program to identify the centerline backbone of the chain (in red). This chain was traced from the NC1 domain to the 7S domain of a collagen type IV molecule.

persistence length. The WLC model is suited for describing semi-flexible polymers with successive segments whose orientation is correlated. This model assumes that a polymer adopts a smoothly curved conformational ensemble at room temperature, and at T=0 K exhibits only a rigid rod conformation. The two WLC model equations implemented in this work to determine the persistence length are the mean-squared end-to-end distance $\langle R^2(s)\rangle$ and the tangent vector correlation function$\langle \cos\theta(s)\rangle$, described in Fig. 17 and shown below:

$$\langle R^2(s)\rangle = 4sp\left[1-\frac{2p}{s}\left(1-e^{-\frac{s}{2p}}\right)\right] \quad \textbf{(1)}$$

$$\langle \vec{t}_1 \,.\, \vec{t}_2\rangle = \langle \cos\theta(s)\rangle = e^{-\frac{s}{2p}} \quad \textbf{(2)}$$

Equations (1) and (2) were used to fit the collagen type IV statistical data collected from SmarTrace, as shown in Fig.18 and Fig 19. These results show that collagen IV is well-described

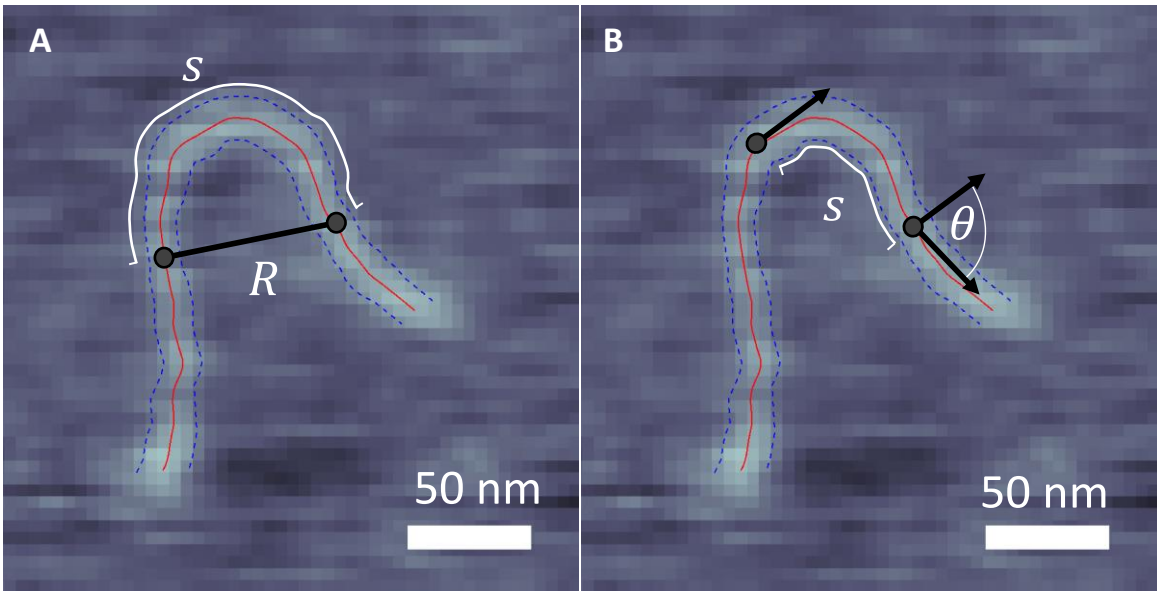

Fig. 17. **Schematic showing how the mean-squared end-to-end distance and tangent correlation function are calculated. A.** Relates the distance along the contour, segment *s*, to the linear distance between the start and the end point, denoted as *R*. **B**. Takes the dot product of two tangent vectors that are a segment *s* apart. Figure courtesy Aaron Lyons.

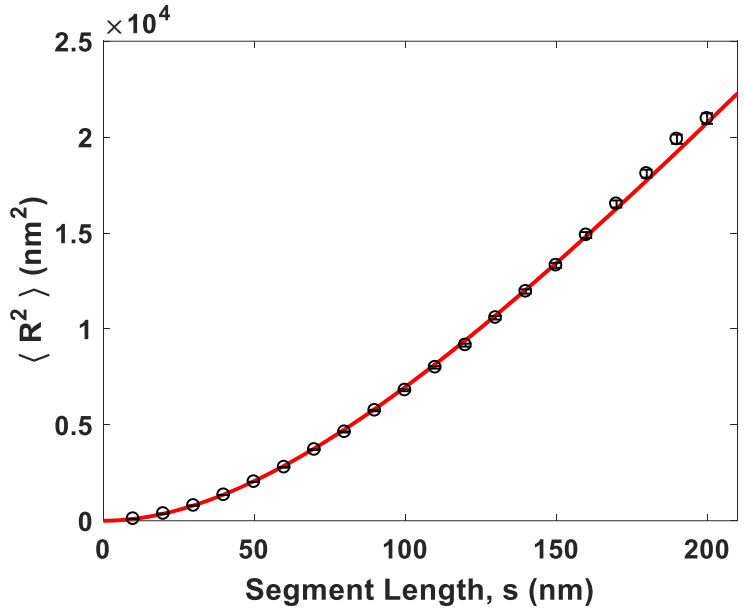

Fig. 18. **Mean-squared end-to-end distance as a function of segment length of collagen type IV.** Collagen IV is well-described by the worm-like chain model in 100 mM KCl, 1 mM HCl and 1 mM Acetic Acid. The curve fit (in red) using equation **(1)** yields a persistence length of p = 43.0 ± 0.6 nm.

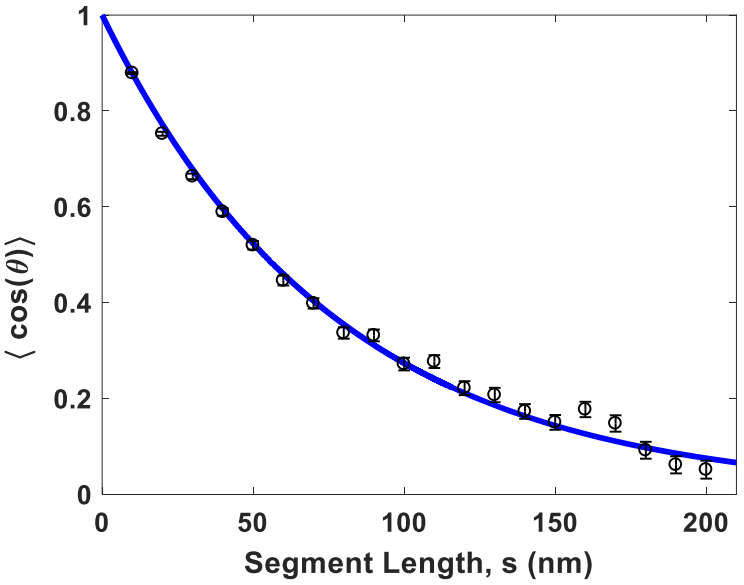

Fig. 19. **Tangent correlation function as a function of segment length of collagen type IV.** Collagen IV is well-described by the worm-like chain model in 100 mM KCl ,1 mM HCl and 1 mM Acetic Acid. The curve fit (in blue) using equation **(2)** yields a persistence length of p = 39.3 ± 1.4 nm.

by the WLC model, yielding persistence lengths of p = 43.0 ± 0.6 nm and p = 39.3 ± 1.4 nm. Comparing these persistence length values to the contour length of collagen IV, which is 400 nm long without the NC1 domain, shows us that it is a flexible polymer.

Different collagen types are associated with the heterogeneity of various tissues mechanical properties. The extent of amino acid composition heterogeneity can differ across many collagen types, where network-forming collagens have a more profound heterogeneity along their contour than fibrillar collagens. This is due to interrupted regions in the Gly-X-Y amino acid sequence that forms the triple helix of network-forming collagens. These interruptions have been reported to destabilize the triple helix and increase the flexibility of the overall structure [9]. Collagen type I

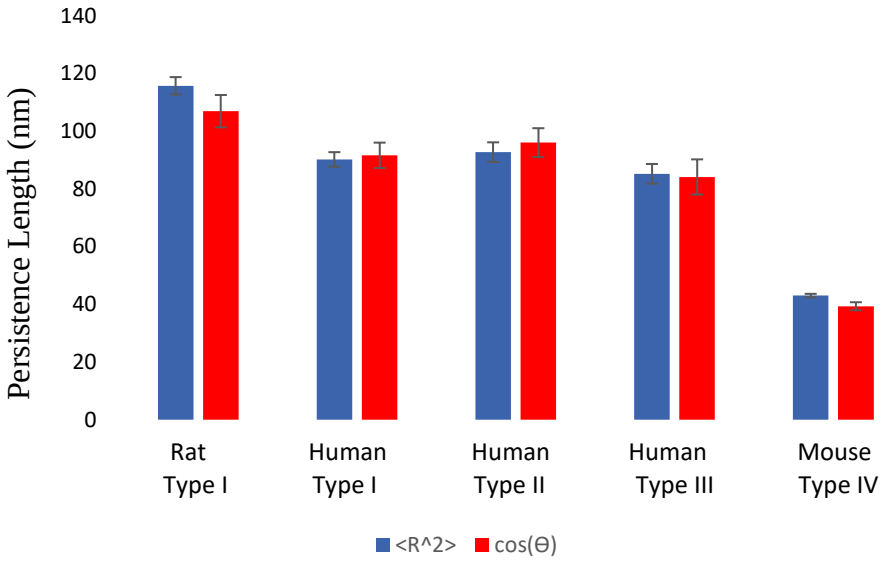

Fig. 20. **Comparing the persistence length of collagen type IV to different fibrillar collagens.** Collagen types I, II, and II from rat and human were deposited from 100 mM KCl and I mM HCl, while collagen type IV was deposited from 100 mM KCl, 1 mM HCl, and 1 mM Acetic Acid. The persistence lengths from the $<R^2>$ fits (in blue) are 115.7 ± 3, 90.2 ± 2.5, 92.7 ± 3.4, 85.2 ±3.4, and 43.0 ± 0.6 nm. The persistence lengths from the $<\cos(\theta)>$ fits (in red) are 106.9 ± 5.6, 91.6 ± 4.4, 96.0 ± 5, 84.1 ± 6.1, 39.3 ± 1.4 nm.

is 96 % triple helical (with non-helical regions confined to its ends) while collagen type IV is 80% triple helical. This difference in triple helix composition is evident in the persistence length of the collagen type IV molecule when compared to fibrillar collagens, as shown in Fig. 20. The persistence lengths of the different fibrillar collagens are very similar, within the range of approximately 80-120 nm [26], while collagen type IV is less than half of that. This striking difference in flexibility can also be seen when comparing AFM images of collagen type IV to collagen type I, shown in Fig. 21A and Fig. 21B respectively. We can see that the flexibility of the collagen IV molecule varies along its contour, and seems to be more flexible near the 7S domain, evident by the "hook" curvature.

I decided to map out where the triple helix interruptions lie in the $\alpha1(IV)]_2$–$\alpha2(IV)$ collagen IV protomer, shown in Fig. 22. It was striking to find out that most of the interruptions are located in the N-terminal half of the chain, where we can also see an increase in the flexibility of the chains in the AFM images. This finding sheds light onto the heterogeneity of the collagen molecule, which for the case of collagen type IV appears to be less triple helical and more flexible towards the N-terminal end. Thus far, I have treated collagen as a homogeneous polymer to estimate persistence lengths, which is clearly not the case. Next, I decided to investigate the position-dependent flexibility of collagen type IV by assigning persistence lengths in segments along the contour.

Collagen is an inhomogeneous polymer because it possesses a heterogeneous amino acid composition along its contour. Experimentally characterizing collagen as an inhomogeneous polymer has challenges, especially for fibrillar collagens because they do not have a marker that allows for the distinction of the N- from the C-terminus (see Fig. 21B). Collagen type IV has a C-terminal globular domain (NC1) that allows for this identification with AFM. The NC1 domain

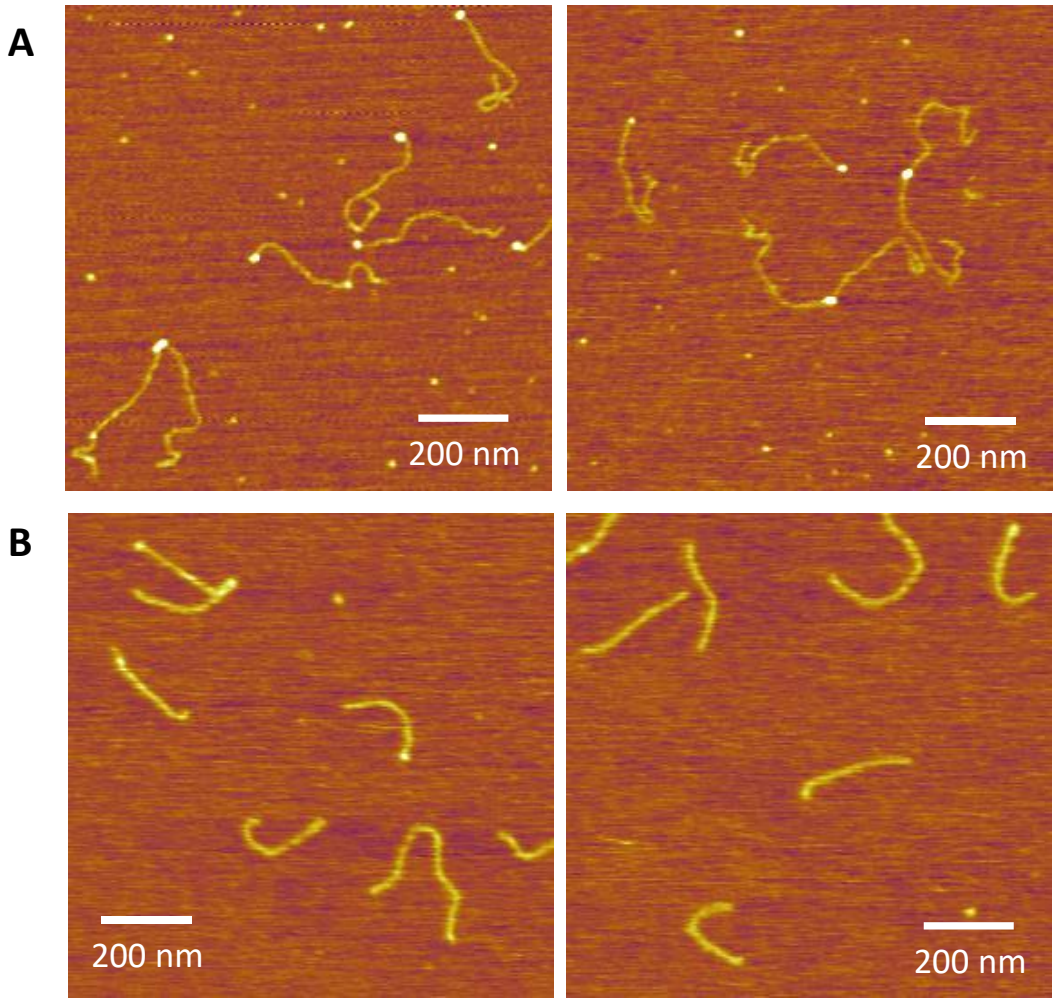

Fig. 21. **AFM images comparing collagen type IV to collagen type I. A.** Collagen type IV deposited from 100 mM KCl, 1mM HCl and 1 mM Acetic Acid. **B**. Collagen type I (rat) deposited from 100 mM KCl and 1 mM HCl.

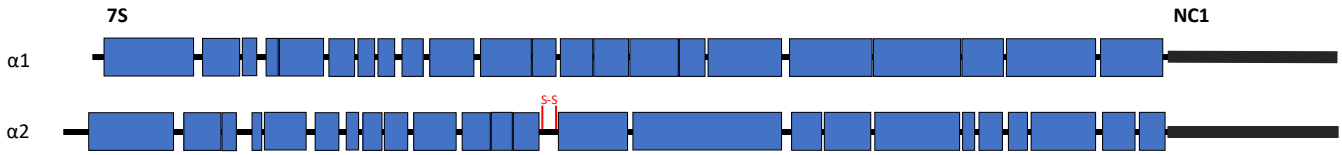

Fig. 22. **Collagen type IV triple helix interruptions map.** This presentation is to scale with the amino acid sequence length (uniprot: P02463, P08122). The blue boxes are triple helical regions, which I defined as (GXY)nG with n ≥ 4. These can be seen more frequently towards the C-terminal end (NC1). The longest interruption, 26 amino acids long in α2, is flanked by two cysteines (in red).

can be easily distinguished and thus the chains can be traced from one end to another and can be statistically aligned for the implementation of position-dependent tools. For this part of my thesis, I used an inhomogeneous chain mapping algorithm developed by a fellow undergraduate student, Aaron Lyons. The details of this program can be found in his undergraduate thesis [27].

To start off, we wanted to test the robustness of this programs in capturing interrupted regions using simulated chains. The methodology behind these simulations relies on sampling from probability distributions of bending angles using Monte Carlo methods [27]. The simulated chains (see Fig. 23A) used in this study have globular domains at one end, and a chain with a persistence length of 85 nm interrupted by segments with a significantly lower persistence length of 5 nm. The 85 nm persistence length was chosen because it resembles what we have seen with fibrillar collagens that are continuously triple helical. The segments of lower persistence length are spaced approximately 80 nm apart, and run for 1 nm, 2 nm, 3 nm and 4 nm along the chain. The analysis program uses segments of 30 nm that move in increments of 1 nm across the chain to calculate position-dependent persistence lengths. The persistence length map of the simulated chains is shown in Fig. 23B, where we can clearly see four minima that correspond to the interrupted regions – shown on the number line at the top of the plot. We can also see that the minima follow the expected trend, where the persistence length drops for longer interrupted segments. Note that the algorithm over-estimates the persistence length in regions with no interruptions, which should retrieve a persistence length of 85 nm. Nevertheless, the expected value is roughly within the 95% confidence intervals of this estimate, at least within the regions of greater rigidity. These results show that the program is capable of capturing interrupted regions with higher flexibilities, and so I decided to apply it to collagen type IV.

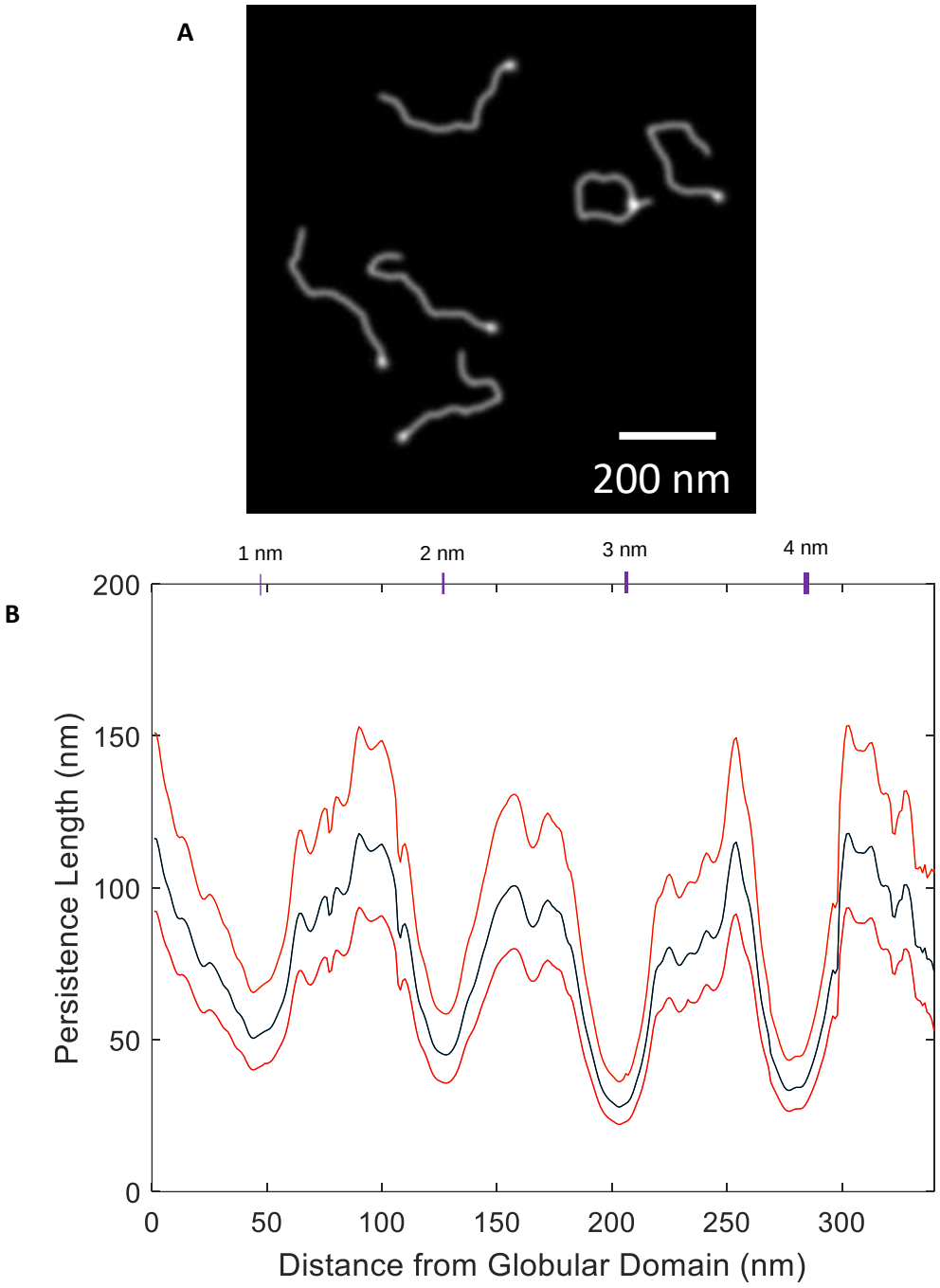

Fig. 23. **Monte Carlo Simulations of chains with interrupted regions. A.** Simulated chains that are 400 nm long, with a 10 nm globular domain. The chains have four interrupted regions – significantly lower persistence length. This image was kindly provided by Aaron Lyons. **B**. Position-dependent persistence length map of the simulated chains starting from the globular end. The red curves represent 95% confidence intervals. The positions of the interrupted regions are shown on the number line at the top of the plot.

The aligned amino acid sequence of collagen type IV (mouse) from α1 and α2 is shown in Fig. 24A, while the position-dependent flexibility profile is shown in Fig. 24B. Details on the alignment of the sequences can be found in the appendix. There are many interrupted regions in α1 and α2, shown in red and blue respectively, towards the 7S domain. The flexibility profile shows a clear trend where the persistence length decays towards the 7S domain, which was an expected result due to more interruptions near that end. There are also some over lapping interrupted regions, shown in purple, which are clearly evident in the flexibility profile. The over-

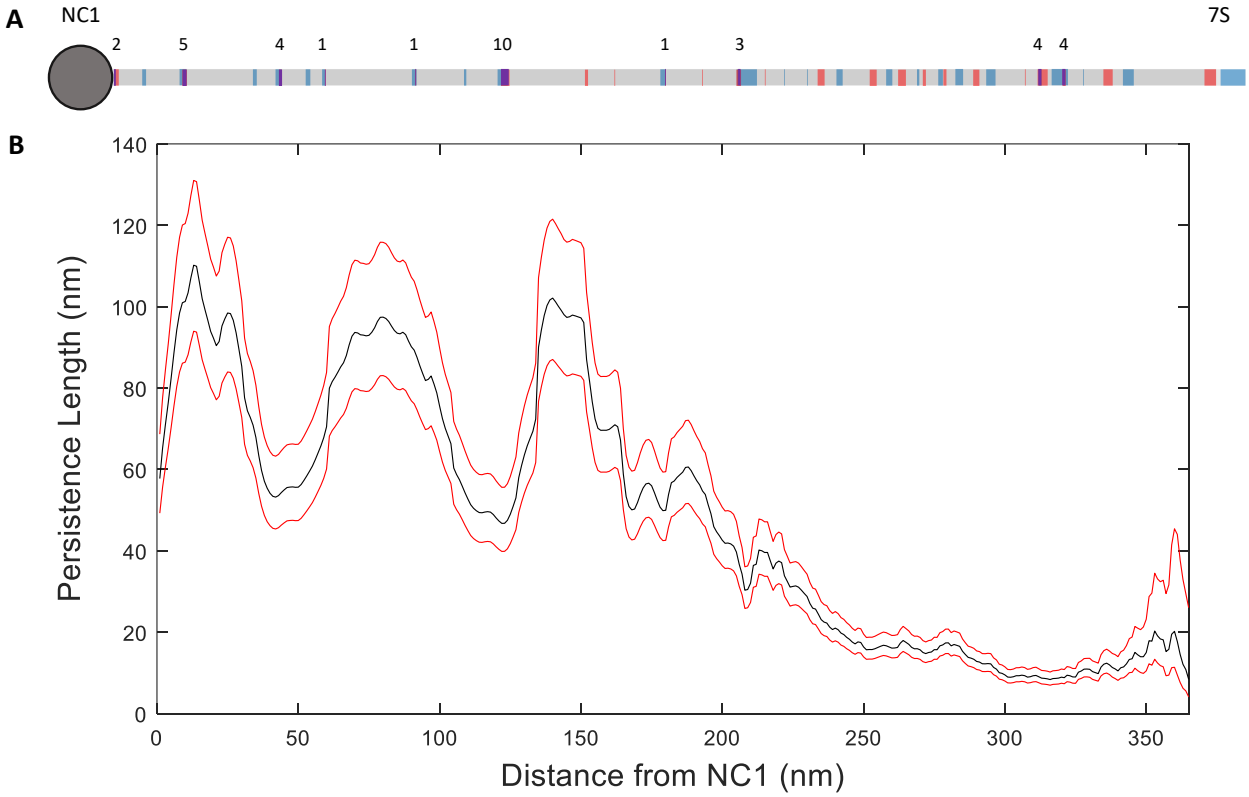

Fig. 24. **Position-dependent flexibility profile of collagen type IV. A.** Alignment of the α1 and α2 amino acid sequence from mouse collagen type IV, showcasing triple helix interruptions. Interruptions from α1 are shown in red, while interruptions from α2 are shown in blue. Over-lapping interruptions are in purple, with the numbers representing the amino acid length of these regions. **B**. Position-dependent persistence length map of collagen type IV starting from the NC1 domain. The red curves represent 95% confidence intervals.

lapped regions that run for 4 and 10 amino acids, near the NC1 domain, are very well aligned with the minima in persistence length. Furthermore, earlier we saw that collagen type IV has a lower persistence length than fibrillar collagens, which have no triple helix interruptions. The fibrillar collagens' persistence length spanned from 80-120 nm for types I, II, and III. It is striking that the persistence length of the triple helical segments in collagen type IV, the three maxima in the plot, share the same persistence length as fibrillar collagens. These results verify the effect of the triple helix interruptions on the flexibility of the collagen molecule, where the interruptions lower the persistence length, contributing to the overall flexibility of the structure.

Discussion and Future Directions

In this thesis, I explored the molecular basis of collagen type IV assembly by studying the stability of NC1 domains. Understanding how collagen type IV is assembled into networks is the first step in understanding the organization of BMs into scaffolds with regulatory functions. It remains unknown how the NC1 hexamer assembles in solution, and these results can provide insight into NC1 hexamer stability and shed light on the importance of cross-links. Through analyzing dissociation patterns of an un-cross-linked hexamer in the absence of chloride, I hypothesize that the hexamer first falls apart into two trimers before readily dissociating into monomers. Upon plating, I expect that the NC1 hexamers in solution fall onto mica as hexamers and dissociate upon water washing. We saw that the LBM NC1 hexamer dissociated into trimers and monomers upon removal of chloride. I believe that this result could be due to the drying the mica before the domains were able to completely dissociate into monomers. The AFM takes a snapshot of all states and revealed some trimers on mica for the LBM NC1 hexamer sample, while gel-filtration of this sample, a ~ 20 min process, shows no trimers upon chloride removal [13]. With these results, it has been established that NC1 trimers in solution are unstable in the absence of a sufficient triple helical domain. However, the NC1 domains do form trimers inside the cell to assemble the collagen IV protomer. This association inside the cell is described as a transient step, and so I believe that the trimers seen with AFM are intermediate steps that have been captured on the mica and dried. An experiment that may address this is to try a longer water wash and see if that induces complete dissociation, which I have not yet tried unfortunately. Furthermore, the results of the LBM NC1 hexamers imaged following overnight incubation in TNA also reveal trimers on the surface. Here, I expect that the NC1 hexamers are dissociating in solution. However, the NC1 domains can also come together transiently and form trimers, which can also be captured

on the mica upon plating and drying. One experiment that I hoped would have work out this term, and which I spent a lot of time on, is washing the mica with salt-containing solutions, such as TBS. Washing the mica with salt solutions and drying caused salt crystals to form on the surface, making it hard to resolve the proteins. This technique of TBS washing would have allowed me assess the fraction of NC1 hexamers intact in the diluted samples I used for AFM imaging, which are at a low concentration which could lead to dissociation of the hexamers even in the presence of chloride. In principal, some of the domains seen on mica could have come from dissociation in solution rather than dissociation induced by water washing. To keep track of this, I worked with higher concentration conditions for the NC1 domains, and decreased the incubation time on the mica so that the surface would not get overcrowded with proteins.

AFM imaging is a useful method that can provide insight into the structure and dissociation of proteins. However, AFM is not common method used to estimate sizes of proteins because of tip convolution – an inherent feature of AFM causing the images to be convolutions of the shape of the probe, and the shape of the sample. The imaged proteins present over-estimates of their actual sizes, because the tip tapping on the surface is larger than the protein domains on the surface. In order to distinguish between NC1 hexamers, trimers, and monomers in my experiments, I compared the sizes of the 'blobs' in the images.

Having a well-developed technique to study NC1 domain assembly and dissociation can allow for using them as recognition modules to understand how specific mutations can cause diseases through interfering with protomer assembly and stabilization. It can also allow for studying defects in collagen type IV that impair the organization of BMs for both hereditary and acquired diseases, such as Alport syndrome and Goodpasture syndrome, respectively [28]. These diseases have been found to be associated with eliciting antibodies that bind to different epitopes

on the α3(IV)–α4(IV)–α5(IV) NC1 hexamer. The NC1 domain of the α3 chain is the target for autoantibodies in Goodpasture syndrome leading to symptoms of bleeding in the lungs, and renal failure. The epitope for the Goodpasture autoantibodies reside in the trimer-trimer interface of a α3(IV)–α4(IV)–α5(IV) NC1 hexamer, making it inaccessible for antibody binding unless exposed by dissociation of the hexamer. For BMs that have been subjected to the post-translational modification that forms the sulfilimine cross-links in collagen IV networks, this confers an immune privilege against Goodpasture syndrome because the epitope remains inaccessible, buried inside the hexamer. In contrast, Alport syndrome is a genetic disease that has mutations in the α3(IV)–α4(IV)–α5(IV) collagen protomer that causes defective assembly of the collagen type IV network and leads to hearing loss, eye abnormalities, and also progressive renal failure. The immune system of the Alport patient recognizes the α3(IV)–α4(IV)–α5(IV) network as foreign, which elicits alloantibodies against the NC1 domains in response to blood transfusion or transplantation. Here, the epitopes for the Alport alloantibodies reside on the hexamer surface, making it accessible for antibody binding without the need for hexamer dissociation. With my studies in developing a well-established technique with AFM to study NC1 domain stability, one can study how these antibodies bind to different complexes of NC1 domains. These studies can help us understand the molecular identity of autoantigens related to diseases that affect collagen type IV.

In addition to studying the end-domains of collagen type IV, I also studied the molecular properties of the collagenous domain using polymer physics analysis. The results in this section shed light onto the heterogeneity of collagen molecules and the effects of triple helix interruptions on the flexibility of the molecule – a feature that is incompletely understood and lacking a well-defined role. Defining the structural consequences of interruptions in a sequence may advance our

understanding of its biological role, where studies have shown that the interruptions can promote supramolecular association. It has also been suggested that the interruptions play a role in molecular flexibility, collagen degradation, and ligand binding [29]. Moreover, there are contradictory reports in the literature on the effects of interruptions on the stability of the triple helix. Some believe that the interruptions within the collagenous domain play a role in destabilizing the triple helix, and a recent publication reports stabilization of the triple helix, where they have found that collagen IV has similar, and even higher melting temperatures than fibrillar-forming collagens with continuous triple-helical collagenous domain [24].

We saw that the interruptions in α1 and α2 of collagen type IV are not conserved, where some interruptions are found in one chain but not in the other, and some are over lapping, which I have shown to have a significant effect on the persistence length. Some of these interruptions have been suggested to form loops, to maximize triple helical lengths. There is one particular interruption in α2 that is 26 amino acids long and has two flanking cysteines (in red shown in Fig. 22). This region is thought to loop and form a disulfide cross-link [30]. Furthermore, we saw that the incidence of these interruptions is higher in the N-terminal end near the 7S domains, which also correlated with the flexibility of the chains seen in the AFM images. We saw that collagen type IV is a more flexible polymer than fibrillar collagens, showcasing the effects of triple-helical interruptions. I believe that the differences seen in their flexibilities is also illustrated in the type of higher-order structures they assemble into, where fibrillar collagens form fibers which are more rigid structures than the networks formed by collagen type IV.

In addition, using a method that allows for yielding position-dependent flexibility profiles, I have mapped out persistence length as a function of distance along the contour of collagen type IV. Flexibility profiles of collagen type IV have been previously published in the literature [31,

32] by using electron micrographs. Their calculations were done by taking pictures and tracing chains manually using a magnetic pen on a data tablet and later using a computer program to calculate radii of curvature at successive positions along the contour. This method has many limitations and one can expect large errors associated with their calculations. The methods used in this work are an advancement because they use a chain tracing algorithm rather than a user-input spline, and prove to be a valuable system for extracting flexibility profiles of polymers.

Acknowledgement

I would like to express my deep gratitude to my senior supervisor, Nancy Forde, who has guided me throughout my research with patience and encouragement. I would also like to thank the members of the Forde lab for their continued support and help throughout this term. I sincerely acknowledge Aaron Lyons for contributing his time in teaching me how to use the programs executed in this work. I am particularly grateful for the assistance provided by our collaborators, Sergey Budko and Billy G. Hudson, and for the samples they sent us. A special thanks to ChangMin Kim who helped with using the AFM instrument. Finally, this journey would not have been possible without the continued support and love from my parents. Thank you for being there for me throughout the toughest of times and for pushing me beyond my limits in becoming the best version of myself.

References

[1] M. D. Shoulders and R. T. Raines. Collagen Structure and Stability. *Annu. Rev. Biochem.*, 78: 929-958, 2009.

[2] C. Knupp and J. M. Squire. Molecular Packing in Network-Forming Collagens. *Adv. Protein Chem.*, 70: 375-403, 2005.

[3] S. Ricard-Blum and F. Ruggiero The collagen superfamily: from the extracellular matrix to the cell membrane. *Pathologie Biologie.*, 53: 430-442, 2005.

[4] B.G. Hudson, S.T. Reeders, and K. Tryggvason. Type IV collagen: structure, gene organization, and role in human diseases. Molecular basis of Goodpasture and Alport syndromes and diffuse leiomyomatosis. *J. Biol. Chem.*, 268: 26033-26036, 1993.

[5] P. D. Yurchenco. Basement Membranes: Cell Scaffoldings and Signaling Platforms. *Cold Spring Harb. Perspect. Biol*., 3(2), 2011.

[6] R. Dölz, J. Engel, and K. Kühn. Folding of collagen IV. *Eur. J. Biochem*., 178: 357-366, 1988.

[7] J. Engel, and D. J. Prockop. The Zipper-Like Folding of Collagen Triple Helices and the Effects of Mutations that Disrupt the Zipper. *Annu. Rev. Biophys. Biophys. Chem.*, 20: 137-52, 1991.

[8] E. S. Hwang and B. Brodsky. Folding Delay and Structural Perturbations Caused by Type IV Collagen Natural Interruptions and Nearby Gly Missense Mutations. *J. Biol. Chem.*, 287: 4368 - 4375, 2012.

[9] P. Vandenberg, A. Kern, A. Ries, L. Luckenbill-Edds, K. Mann, K. Kühn. Characterization of a type IV collagen major cell binding site with affinity to the alpha 1 beta 1 and the alpha 2 beta 1 integrins. J. Cell Biol., 113: 1475-1483, 1991.

K u H N ’ Folding of collagen IV

[10] H. K. Kleinman, M. C. Kibbey, H. W. Schnaper, M. A. Hadley, M. Dym, D. S. Grant. Role of basement membrane in differentiation. In R. Timpl, & D. Rohrbach (Eds.), *Molecular and Cellular Aspects of Basement Membranes*, New York: Academic Press, 309-326, 1993.

[11] J. Khoshnoodi, V. Pedchenko, B. G. Hudson. Mammalian Collagen IV. *Microsc. Res. Tech*., 71: 357370, 2008.

[12] B. R. Nayak, R. G. Spiro. Localization and structure of the asparagine-linked oligosaccharides of type IV collagen from glomerular basement membrane and lens capsule. *J. Biol. Chem*., 266: 13978-13987, 1991.

[13] C. F. Cummings, V. Pedchenko, K. L. Brown, S. Colon, M. Rafi, C. Jones-Paris, E. Pokydeshava, M. Liu, J. C. Pastor-Pareja, C. Stothers, I. A. Ero-Tolliver, A. S. McCall, R. Vanacore, G. Bhave, S. Santoro, T. S. Blackwell, R. Zent, A. Pozzi, and B. G. Hudson. Extracellular chloride signals collagen IV network assembly during basement membrane formation. *J. Cell Biol.*, 213: 479-494, 2016.

[14] R. Vanacore, A. J. Ham, M. Voehler, C. R. Sanders, T. P. Conrads, T. D. Veenstra, K. B. Sharpless, P. E. Dawson, B. G. Hudson. A sulfilimine bond identified in collagen IV, *Science*, 325: 1230-1234, 2009.

[15] G. Bhave, C. F. Cummings, R. Vanacore, C. Kumagai-Cresse, I. A. Ero-Tolliver, M. Rafi, J. Kang, V. Pedchenko, L. I. Fessler, J. H. Fessler, and B. G. Hudson. Peroxidasin forms sulfilimine chemical bonds using hypohalous acids in tissue genesis. *Nat. Chem. Bio*., 8: 784-790, 2012.

[16] E. C. Tsilibary and A. S. Charonis. The Role of the Main Noncollagenous Domain (NC1) in Type IV Collagen Self-Assembly. *J. Cell Biol*., 103: 2467-2473, 1986.

[17] R. M. Vanacore, A. L. Ham, J. Cartailler, M. Sundaramoorthy, P. Todd, V. Pedchenko, Y. Sado, D. Borza, and B. G. Hudson. A Role for Collagen IV Cross-links in Conferring Immune Privilege to the Goodpasture Autoantigen. *J. Biol Chem*., 283: 22737-22748, 2008.

[18] D. J. Müller1 and Y. F. Dufrêne. Atomic force microscopy: a nanoscopic window on the cell surface. *Trends in Cell Biology*, 21: 461–469, 2011.

[19] M. Rubinstein and R. H. Colby. Polymer physics. Oxford University Press, 2003.

[20] N. Rezaei, A. Lyons, N. R. Forde. Conformational flexibility of collagen molecules: composition and environmental effects. *Submitted*

[21] S. P. Boudko, N. Danylevych, B. G. Hudson, V. K. Pedchenko. Basement membrane collagen IV: Isolation of functional domains. *Methods Cell Biol*, 143: 171-185, 2018.

[22] N. Rezaei. Mechanical Studies of Single Collagen Molecules Using Imaging and Force Spectroscopy. Simon Fraser University, 2016.

[23] K. Kuhn, H. Wiedemann, R. Timpl, J. Risteli, H. Dieringer, T. Voss and R. W. Glanville.

Macromolecular Structure of Basement Membrane Collagens: Identification of 7 S collagen as a crosslinking domain of type IV collagen. *FEBS Letters*, 125: 123-128, 1981.

[24] D. Brazel, I. Oberbaumer, H. Dieringer, W. Babel, R. W. Glanville, Rainer Deutzmann and K. Kühn. Completion of the amino acid sequence of the a1 chain of human basement membrane collagen (type IV) reveals 21 non-triplet interruptions located within the collagenous domain. *Eur. J. Biochem.* 168: 529-536, 1987.

[25] B. R. Nayak, R. G. Spiro. Localization and structure of the asparagine-linked oligosaccharides of type IV collagen from glomerular basement membrane and lens capsule. *J. Biol. Chem*., 266: 13978-13987, 1991.

[26] N. Rezaei. Mechanical Studies of Single Collagen Molecules Using Imaging and Force Spectroscopy, 2016. PhD (Simon Fraser University, Burnaby, Canada).

[27] A. Lyons. Collagen Mechanics and the Study of Inhomogeneous Biopolymers, 2017. Undergraduate Honours Thesis (Simon Fraser University, Burnaby, Canada).

[28] B. G. Hudson. The Molecular Basis of Goodpasture and Alport Syndromes: Beacons for the Discovery of the Collagen IV Family. JASN, 15: 2514-2527, 2004.

[29] E. S. Hwang, G. Thiagarajan, A. S. Parmar, and B. Brodsky. Interruptions in the collagen repeating tripeptide pattern can promote supramolecular association. *Protein Science*. 19: 1053-1064, 2010.

[30] D. Brazel, R. Pollner, I. Oberbaumer and K. Kühn. Human basement membrane collagen (type IV) The amino acid sequence of the a2(IV) chain and its comparison with the al(1V) chain reveals deletions in the al(1V) chain. *Eur. J. Biochem*. 172: 35-42, 1988.

[31] H. Hofmann, T. Voss, K. Kühn, and J. Engel. Localization of flexible sites in thread-like molecules from electron micrographs. Comparison of interstitial, basement membrane and intima collagens. *J. Mol. Biol*. 172: 325-43, 1984.

[32] G. P. Lunstrum, H. Bachinger, L. I. Fessler, K. G. Duncane, R. E. Nelson, and J. H. Fessler. Drosophila Basement Membrane Procollagen IV I. Protein Characterization and Distribution. *J. Biol. Chem.* 263: 18318-18327, 1988.

Appendix

> Sequence Alignment of α1(IV)]$_2$–α2(IV) Collagen (mouse)

Red : α1(IV)
Blue: α2(IV)]
Underlined : 7S domain
Highlighted: triple helix interruptions
Highlighted: α1(IV) NC1 domain
Highlighted: α2(IV) NC1 domain

```
------------------------------------KGDCGGSGCGKCDCHGVKGQKGERGLPGLQGV
LLAQSVLGGVKKLDVPCGGRDCSGGCQCYPEKGARGQPGAVGPQGYNGPPGLQGFPGLQGRKGDKGERGV

IGFPGMQGPEGPHGPPGQKGDAGEPGLPGTKGTRGPPGAAGYPGNPGLPGIPGQDGPPGPPGIPGCNGTK
PGPTGPKGDVGARGVSGFPGADGIPGHPGQGGPRGRPGYDGCNGTRGDAGPQGPSGSGGFPGLPGPQGPK

GERGPLGPPGLPGFSGNPGPPGLPGMKGDPGEILGHVPGTLLKGERGFPGIPGMPGSPGLPGLQGPVGPP
GQKGEPYALSKEDRDKYRGEPGEPGLVGYQGPPGRPGPIGQMGPMGAPGRPGPPGPPGPKGQPGNRGLGF

GFTGPPGPPGPPGPPGEKGQMGSSFQGPKGDKGEQGVSGPPGVPGQAQVKEKGDFAPTGEKGQKGEPGFP
YGQKGEKGDIGQPGPNGIPSDITLVGPTTSTIHPDLYKGEKGDEGEQGIPGVISKGEEGIMGFPGIRGFP

GVPGYGEKGEPGKQGPRGKPGKDGEKGERGSPGIPGDSGYPGLPGRQGPQGEKGEAGLPGPPGTVIGTMP
GLDGEKGVVGQKGSRGLDGFQGPSGPRGPKGERGEQGPPGPSVYSPHPSLAKGARGDPGFQGAHGEPGSR

LGEKGDRGYPGAPGLRGEPGPKGFPGTPGQPGPPGFPTPGQAGAPGFPGERGEKGDQGFPGVSLPGPSGR
GEPGEPGTAGPPGPSVGDEDSMRGLPGEMGPKGFSGEPGSPARYLGPPGADGRPGPQGVPGPAGPPGPDG

DGAPGPPGPPGPPGQPGHTNGIVECQPGPPGDQGPPGTPGQPGLTGEVGQKGQKGESCLACDTEGLRGPP
FLFGLKGSEGRVGYPGPSGFPGTRGQKGWKGEAGDCQCGQVIGGLPGLPGPKGFPGVNGELGKKGDQGDP

GPQGPPGEIGFPGQPGAKGDRGLPGRDGLEGLPGPQGSPGLIGQPGAKGEPGEIFFDMRLKGDKGDPGFP
GLHGIPGFPGFKGAPGVAGAPGPKGIKGDSRTITTKGERGQPGIPGVHGMKGDDGVPGRDGLDGFPGLPG

GQPGMPGRAGTPGRDGHPGLPGPKGSPGSIGLKGERGPPGGVGFPGSRGDIGPPGPPGVGPIGPVGEKGQ
PPGDGIKGPPGDAGLPGVPGTKGFPGDIGPPGQGLPGPKGERGFPGDAGLPGPPGFPGPPGPPGTPGQRD

AGFPGGPGSPGLPGPKGEAGKVVPLPGPPGAAGLPGSPGFPGPQGDRGFPGTPGRPGIPGEKGAVGQPGI
CDTGVKRPIGGGQQVVVQPGCIEGPTGSPGQPGPPGPTGAKGVRGMPGFPGASGEQGLKGFPGDPGREGF

GFPGLPGPKGVDGLPGEIGRPGSPGRPGFNGLPGNPGPQGQKGEPGIGLPGLKGQPGLPGIPGTPGEKGS
PGPPGFMGPRGSKGTTGLPGPDGPPGPIGLPGPAGPPGDRGIPGEVLGAQPGTRGDAGLPGQPGLKGLPG

IGGPGVPGEQGLTGPPGLQGIRGDPGPPGVQGPAGPPGVPGIGPPGAMGPPGGQGPPGSSGPPGIKGEKG
ETGAPGFRGSQGMPGMPGLKGQPGFPGPSGQPGQSGPPGQHGFPGTPGREGPLGQPGSPGLGGLPGDRGE
```

FPGFPGLDMPGPKGDKGSQGLPGLTGQSGLPGLPGQQGTPGVPGFPGSKGEMGVMGTPGQPGSPGPAGTP
PGDPGVPGPVGMKGLSGDRGDAGMSGERGHPGSPGFKGMAGMPGIPGQKGDRGSPGMDGFQGMLGLKGRQ

GLPGEKGDHGLPGSSGPRGDPGFKGDKGDVGLPGMPGSMEHVDMGSMKGQKGDQGEKGQIGPTGDKGSRG
GFPGTKGEAGFFGVPGLKGLPGEPGVKGNRGDRGPPGPPPLILPGMKDIKGEKGDEGPMGLKGYLGLKGI

DPGTPGVPGKDGQAGHPGQPGPKGDPGLSGTPGSPGLPGPKGSVGGMGLPGSPGEKGVPGIPGSQGVPGS
QGMPGVPGVSGFPGLPGRPGFIKGVKGDIGVPGTPGLPGFPGVSGPPGITGFPGFTGSRGEKGTPGVAGV

PGEKGAKGEKGQSGLPGIGIPGRPGDKGDQGLAGFPGSPGEKGEKGSAGTPGMPGSPGPRGSPGNIGHPG
FGETGPTGDFGDIGDTVDLPGSPGLKGERGITGIPGLKGFFGEKGAAGDIGFPGITGMAGAQGSPGLKGQ

SPGLPGEKGDKGLPGLDGVPGVKGEAGLPGTPGPTGPAGQKGEPGSDGIPGSAGEKGEQGVPGRGFPGFP
TGFPGLTGLQGPQGEPGRIGIPGDKGDFGWPGVPGLPGFPGIRGISGLHGLPGTKGFPGSPGVDAHGDPG

GSKGDKGSKGEVGFPGLAGSPGIPGVKGEQGFMGPPGPQGQPGLPGTPGHPVEGPKGDRGPQGQPGLPGH
FPGPTGDRGDRGEANTLPGPVGVPGQKGERGTPGERGPAGSPGLQGFPGISPPSNISGSPGDVGAPGIFG

PGPMGPPGFPGINGPKGDKGNQGWPGAPGVPGPKGDPGFQGMPGIGGSPGITGSKGDMGLPGVPGFQGQK
LQGYQGPPGPPGPNALPGIKGDEGSSGAAGFPGQKGWVGDPGPQGQPGVLGLPGEKGPKGEQGFMGNTGP

GLPGLQGVKGDQGDQGVPGPKGLQGPPGPPGPYDVIKGEPGLPGPEGPPGLKGLQGPPGPKGQQGVTGSV
SGAVGDRGPKGPKGDQGFPGAPGSMGSPGIPGIPQKIAVQPGTLGPQGRRGLPGALGEIGPQGPPGDPGF

GLPGPPGVPGFDGAPGQKGETGPFGPPGPRGFPGPPGPDGLPGSMGPPGTPSVDHGFLVTRHSQTTDDPL
RGAPGKAGPQGRGGVSAVPGFRGDQGPMGHQGPVGQEGEPGRPGSPGLPGMPGRSVSIGYLLVKHSQTDQ

CPPGTKILYHGYSLLYVQGNERAHGQDLGTAGSCLRKFSTMPFLFCNINNVCNFASRNDYSYWLSTPEPM
EPMCPVGMNKLWSGYSLLYFEGQEKAHNQDLGLAGSCLARFSTMPFLYCNPGDVCYYASRNDKSYWLSTT

PMSMAPISGDNIRPFISRCAVCEAPAMVMAVHSQTIQIPQCPNGWSSLWIGYSFVMHTSAGAEGSGQALA
APLPMMPVAEEEIKPYISRCSVCEAPAVAIAVHSQDTSIPHCPAGWRSLWIGYSFLMHTAAGDEGGGQSL

SPGSCLEEFRSAPFIECHGRGTCNYYANAYSFWLATIERSEMFKKPTPSTLKAGELRTHVSRCQVCMRRT
VSPGSCLEDFRATPFIECNGGRGTCHYFANKYSFWLTTIPEQNFQSTPSADTLKAGLIRTHISRCQVCMK

--
NL