

# Mechanical Studies of Single Collagen Molecules Using Imaging and Force Spectroscopy

by

Naghmeh Rezaei

B.Sc., University of Tehran

Dissertation Submitted in Partial Fulfillment of the  
Requirements for the Degree of  
Doctor of Philosophy

in the  
Department of Physics  
Faculty of Science

© Naghmeh Rezaei 2016  
SIMON FRASER UNIVERSITY  
Fall 2016

All rights reserved.

However, in accordance with the *Copyright Act of Canada*, this work may be reproduced without authorization under the conditions for "Fair Dealing." Therefore, limited reproduction of this work for the purposes of private study, research, education, satire, parody, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

# Approval

**Name:** Naghmeh Rezaei  
**Degree:** Doctor of Philosophy (Physics)  
**Title:** *Mechanical Studies of Single Collagen Molecules Using Imaging and Force Spectroscopy*  
**Examining Committee:** **Chair:** Malcolm Kennett  
Associate Professor

**Nancy Forde**  
Senior Supervisor  
Associate Professor

---

**John Bechhoefer**  
Supervisor  
Professor

---

**Eldon Emberly**  
Supervisor  
Associate Professor

---

**David Sivak**  
Internal Examiner  
Assistant Professor

---

**Laurent Kreplak**  
External Examiner  
Professor  
Department of Physics and Atmospheric  
Science  
Dalhousie University

---

**Date Defended:** December 7, 2016

---

# Abstract

Collagen is a key component of the extracellular matrix and is the most abundant protein in vertebrates. Collagen is found in almost every connective tissue of the body including skin, bone, tendon, cartilage, arteries and cornea, where it plays a crucial role in providing structural support. Collagen molecules self-assemble to form hierarchical structures, from single molecules to fibrils to fibers and tissues. Structural and mechanical changes at the molecular level may affect self-assembly of the molecules and the resulting tissue. Despite its significance, the mechanics of collagen and its flexibility at the molecular level remain contentious, and collagen has been variously described as a flexible polymer to a semi-rigid rod.

In this thesis, I present my work developing and utilizing experimental and analytical tools to study the mechanical properties of molecular collagen. I carefully designed and controlled a wide variety of experimental conditions, such as different collagen types and sources, solution pH and salt concentrations, and analysed the results in search of potential reasons for inconsistency in reported results of collagen flexibility at the basic molecular level.

Atomic force microscopy (AFM) imaging is used to study effect of environmental factors such as ionic strength and pH on molecular conformations and flexibility of single collagen molecules. In addition, molecular conformations of different types of collagen from different sources are compared using AFM imaging. I measure persistence length of collagen molecules, a measure of flexibility, arising due to the conformational sampling of collagen. My results link the bending energy of collagen molecules to how tightly the helix is wound.

In order to analyse AFM images of collagen, I developed an image and statistical analysis algorithm, SmarTrace, optimized for my images of collagen. The program was validated using images of DNA with known persistence length, then applied to collagen molecules.

Analysis of different types of collagen in two different solutions and type I collagen in solutions of different ionic strength and pH show that collagen's flexibility depends strongly on ionic strength and pH. In addition, it shows that different types of collagen show similar average conformational characteristics in a given solution environment.

In addition, mechanical properties and force-response of single collagen and procollagen molecules are studied using optical tweezers. I discuss the challenges of stretching single collagen proteins,

whose length is much less than the size of the microspheres used as manipulation handles, and show how instrumental design and biochemistry can be used to overcome these challenges.

The result of this work is an improved understanding of the sensitivity of molecular flexibility, stability and response of collagen to environmental factors. This can shed light on identifying underlying mechanisms of collagen-related diseases as well as designing and producing improved engineered biomaterials with tunable properties.

**Keywords:** Collagen; DNA; Structural protein; Molecular conformation; Biophysics; Atomic Force Microscopy; Optical Tweezers; Single molecule; Force spectroscopy; Imaging

To mom and dad

# Acknowledgements

I have been extremely fortunate to be surrounded by many caring, loving and inspiring people in my life. Countless individuals have helped me throughout this wonderful journey of pursuing my education, and I am thankful to every one of them.

First, I would like to express my deepest gratitude to my senior supervisor Dr. Nancy Forde. She provided me with continuous support and guided my project with her insight and patience. I have greatly benefited from her invaluable advice, immense knowledge, enthusiasm for science and extraordinary style of thinking over these years.

Thanks to all members of my thesis committee, Dr. John Bechhoefer and Dr. Eldon Emberly for their thought-provoking questions and insightful comments during this work. I am grateful to my thesis examining committee Dr. David Sivak and Dr. Laurant Kreplek for discussions and providing constructive feedback on my research.

My sincere appreciation goes to all past and present members of Forde lab, for valuable discussions and being incredible teammates. Special thanks to Andrew Wieczorek for sharing his invaluable "decades of experience" in biology and biochemistry and for his sense of humor and always being supportive. I sincerely acknowledge Aaron Lyons for his help with AFM imaging experiments, Marjan Shayegan for being a great colleague and a caring and supportive friend.

There were many brilliant faculty, researchers and fellow graduate students with whom I have had the opportunity to collaborate. Dr. Patricia Mooney provided the AFM facility and Chang Min Kim helped with using the instrument. Our collaborators Dr. Fred Keeley and Dr. Megan Miao at University of Toronto provided DNA-elastin samples. Dr. Leah Keshet at University of British Columbia taught me biology from a mathematical point of view. Support from faculty and staff at Simon Fraser University and Department of Physics played a meaningful role in my education.

My wonderful family and friends all around the world were part of my joyful memories. I shared laughter, new experiences and adventures with them and they were always there for me.

My heartfelt feelings and greatest thanks goes to my lovely mom and dad for their unconditional love and never-ending support, for being inspiring, believing in me and giving me the strength to pursue my passions in life. Thanks to my only brother Saba, for his invaluable support and humor, and for inspiring me to believe in myself to follow my dreams. My dear Pooya has been an incredible friend and loving partner. I am thankful for his support and for being understanding and a constant source of encouragement all the time.

# Table of Contents

<b>Approval</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Dedication</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vi</b>
<b>Table of Contents</b>	<b>viii</b>
<b>List of Tables</b>	<b>x</b>
<b>List of Figures</b>	<b>xi</b>
<b>Abbreviations</b>	<b>xiv</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Atomic Force Microscopy: Image Analysis of Single Polymers</b>	<b>8</b>
2.1 AFM setup . . . . .	8
2.2 AFM Image Analysis Program . . . . .	10
2.3 Principles of Data Analysis for Traced AFM Images of Single Polymers . . . . .	21
2.4 Validation of the Image Analysis Method Using AFM Images of DNA . . . . .	22
2.5 Summary and Discussion . . . . .	24
<b>3 Atomic Force Microscopy: Imaging Single Collagen Molecules</b>	<b>27</b>
3.1 Sample Preparation for AFM Imaging of Collagen Molecules . . . . .	27
3.2 AFM Imaging of Different Types of Collagen . . . . .	28
3.2.1 Results . . . . .	28
3.2.2 Non-WLC behaviour of chains . . . . .	38
3.2.3 Summary and Discussion . . . . .	41
3.3 Effect of Ionic Strength and pH on Collagen Mechanics . . . . .	45
3.3.1 Results . . . . .	45
3.3.2 Summary and Discussion . . . . .	48

<b>4</b>	<b>Optical Tweezers Stretching Experiments</b>	<b>54</b>
4.1	Introduction . . . . .	54
4.2	Optical Tweezers . . . . .	58
4.2.1	Principles of Optical Trapping . . . . .	58
4.2.2	Instrument . . . . .	59
4.3	Measurements and Calibration . . . . .	62
4.3.1	Position Measurement . . . . .	63
4.3.2	Calibration . . . . .	65
4.4	Worm-Like Chain (WLC) Model . . . . .	70
4.5	Stretching 11.7 kbp (4 $\mu$ m) <i>dsDNA</i> Molecules . . . . .	72
4.6	Stretching Single Collagen Molecules . . . . .	74
<b>5</b>	<b>Studying Mechanical Properties of Procollagen Molecules Using Optical Tweezers</b>	<b>75</b>
5.1	Materials and Methods . . . . .	75
5.1.1	Microsphere and Procollagen Preparation . . . . .	75
5.1.2	Specifically Tethering Procollagen Between Microspheres . . . . .	76
5.1.3	Force Measurements and Analysis . . . . .	76
5.2	Results and Discussions . . . . .	77
<b>6</b>	<b>Stretching Type III Collagen with Optical Tweezers</b>	<b>79</b>
6.1	Materials and Methods . . . . .	79
6.1.1	Covalent Binding of Collagen Molecules to Microspheres . . . . .	79
6.2	Results and Discussions . . . . .	81
<b>7</b>	<b>Summary and Future Directions</b>	<b>84</b>
	<b>Bibliography</b>	<b>89</b>

# List of Tables

Table 1.1	Examples of connective tissue diseases arising from mutations in collagen.	2
Table 1.2	Persistence lengths ( $l_p$ ) of collagen in the literature. . . . .	3
Table 3.1	Different types of collagen molecules studied in Section 3.2. . . . .	29
Table 3.2	Persistence lengths from $\langle R^2 \rangle$ , $\langle \cos \theta \rangle$ and bending energy analysis for different types of collagen. Reported values represent the average $l_p$ over the range of $150 < s_{\max} < 250$ nm. The reported values beneath for $\langle R^2 \rangle$ and $\langle \cos \theta \rangle$ analyses represent the minimum and maximum values found over this $s_{\max}$ fitting range. The quoted error for the bending energy analysis (Equation 2.9) represents the 95% confidence interval. . . . .	37
Table 3.3	$l_p$ results obtained from standard WLC and the curved model fit. . . . .	43
Table 3.4	List of solution conditions and ionic strengths used in this work. . . . .	45
Table 3.5	Debye screening length of solutions with different ionic strength. . . . .	50

# List of Figures

Figure 1.1	Schematic of collagen structure hierarchy. . . . .	3
Figure 1.2	Collagen triple helix. . . . .	4
Figure 1.3	Force-extension curve of a <i>dsDNA</i> manipulated in the OT instrument. . . . .	6
Figure 2.1	Schematic of the AFM setup. . . . .	9
Figure 2.2	The repulsive and attractive force regimes in AFM. . . . .	10
Figure 2.3	Examples of tracing noisy images of collagen chains using Reference [62]. . . . .	11
Figure 2.4	Examples of tracing noisy images of chains using Reference [63]. . . . .	12
Figure 2.5	Examples of chain mistracing by other algorithms and more accurate tracing using SmarTrace. . . . .	13
Figure 2.6	Examples of traced backbone of chains imaged by SmarTrace program. . . . .	14
Figure 2.7	Tracing noise-free image of DNA. . . . .	15
Figure 2.8	Intensity examples of plot of a subsection of a chain in the search window. . . . .	17
Figure 2.9	Intensity plot, and 3D plot of the quartic polynomial filter. . . . .	18
Figure 2.10	Examples of collagen AFM images with overlapping and unresolved ends. . . . .	20
Figure 2.11	Example of DNA strands analyzed with SmarTrace. . . . .	23
Figure 2.12	Distribution of sampled segment lengths for 24 DNA molecules. . . . .	23
Figure 2.13	$\langle R^2 \rangle$ vs. length along the DNA molecules and WLC fit. . . . .	24
Figure 2.14	$\langle \cos \theta \rangle$ vs. length along the DNA molecules and WLC fit. . . . .	24
Figure 2.15	Determination of bending energy at different separations. . . . .	25
Figure 2.16	Extracting persistence length from the angular distribution fit parameter $B$ . . . . .	26
Figure 3.1	Representative AFM images of collagen molecules. . . . .	29
Figure 3.2	Results of analysis of 49 chains of human collagen type II from cartilage. . . . .	30
Figure 3.3	Results of analysis of 125 chains of collagen type I from rat tail tendon. . . . .	30
Figure 3.4	Results of analysis of 95 chains of recombinant human collagen type I. . . . .	31
Figure 3.5	Results of analysis of 78 chains of recombinant human collagen type III. . . . .	31
Figure 3.6	Determination of bending energy at different contour separations. . . . .	32
Figure 3.7	Extracting persistence length of human collagen type II from fit parameter $B$ . . . . .	33
Figure 3.8	Schematic of calculation of curvature of a chain. . . . .	34
Figure 3.9	Curvature plots at 10 nm contour separation of collagen. . . . .	35

Figure 3.10	Persistence length values for collagen in 1 mM HCl and 100 mM KCl. . .	36
Figure 3.11	Persistence length values for collagen in 1 mM HCl and 100 mM KCl. . .	36
Figure 3.12	AFM images of collagen molecules deposited from 20 mM HOAc. . . . .	37
Figure 3.13	$l_p$ results for different types of collagen in HOAc. . . . .	38
Figure 3.14	Examples of non-WLC behaviour of collagen chains. . . . .	39
Figure 3.15	Persistence length results of the WLC fit to $\langle R^2 \rangle$ and $\langle \cos \theta \rangle$ data as a function of $s_{\max}$ . . . . .	40
Figure 3.16	Example of the fit from curved and standard WLC models to $\langle \cos \theta \rangle$ . . .	41
Figure 3.17	Results of curved model fit for different types of collagen deposited from 20 mM HOAc. . . . .	42
Figure 3.18	$l_p$ and $l_p^*$ results for different types of collagen deposited from 20 mM HOAc.	43
Figure 3.19	AFM images of collagen I molecules deposited from different solutions. .	46
Figure 3.20	Experimental plots and WLC fits for collagen I in 10 mM KCl at neutral pH.	47
Figure 3.21	Dependence of $l_p$ on ionic strength at neutral and acidic pH. . . . .	48
Figure 3.22	Dependence of collagen radius of curvature $R_0$ on ionic strength of the solution. . . . .	49
Figure 3.23	Dependence of apparent persistence length on Debye length. . . . .	50
Figure 3.24	Electrostatic interactions in collagen in different ionic strengths. . . . .	51
Figure 3.25	Estimate of the pH-dependent charge of triple helical collagen. . . . .	52
Figure 3.26	Electrostatic interactions in collagen at low pH and different ionic strengths.	53
Figure 4.1	Comparison of single-molecule force spectroscopy techniques. . . . .	55
Figure 4.2	Schematic of an atomic force microscopy stretching experiment. . . . .	56
Figure 4.3	Schematic of magnetic tweezers setup. . . . .	57
Figure 4.4	Stretching experiment in optical tweezers. . . . .	57
Figure 4.5	A trapped bead in a parabolic potential well from the trap. . . . .	58
Figure 4.6	Ray optics schematic of the gradient force for a dielectric microsphere. .	60
Figure 4.7	Schematic of optical tweezers setup. . . . .	61
Figure 4.8	Photo of the single beam optical tweezers. . . . .	61
Figure 4.9	Schematic of the microfluidic chamber. . . . .	62
Figure 4.10	Schematic of flow system. . . . .	63
Figure 4.11	Image beads stuck on the surface of a coverslip in a sealed chamber. . . .	64
Figure 4.12	Stuck bead displacements as a function of time. . . . .	66
Figure 4.13	Probability density and Gaussian fits of displacements of a stuck bead. .	67
Figure 4.14	Pixel calibration of CCD camera. . . . .	68
Figure 4.15	Calculation of the volts to nanometers conversion factor. . . . .	69
Figure 4.16	Histogram of trap stiffness values. . . . .	70
Figure 4.17	Power spectral density of a trapped bead. . . . .	71
Figure 4.18	Force-extension curve of extension and relaxation of a <i>dsDNA</i> . . . . .	73

Figure 4.19	Stretching curve of a <i>ds</i> DNA with a WLC fit of entropic elasticity. . . . .	74
Figure 5.1	Western blot to monitor binding of procollagen II to anti-N beads. . . . .	76
Figure 5.2	Optical tweezers stretching curve of type II procollagen and the WLC fit.	77
Figure 5.3	Dependence of $l_p$ on the maximum force used in fitting the WLC model.	78
Figure 6.1	Force-extension curve of type III collagen. . . . .	80
Figure 6.2	Force-extension curve of a single type III collagen . . . . .	81
Figure 6.3	Example force-extension curves for collagen type III. . . . .	83
Figure 7.1	Schematic of a collagen-DNA hybrid system . . . . .	88

# Abbreviations

AFM – atomic force microscopy  
BSA – bovine serum albumin  
CCD – charge-coupled device  
CD – circular dichroism  
dsDNA – double-stranded DNA  
EDTA – ethylenediaminetetraacetic acid  
EM – electron microscopy  
FFB – fibril forming buffer  
Gly – glycine  
HCl – hydrochloric acid  
HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid  
HOAc – acetic acid  
HRP – horseradish peroxidase  
MMP – matrix metalloproteinase  
MW – molecular weight  
KCl – potassium chloride  
PBS – phosphate buffered saline  
PD – photodiode  
PSD – power spectral density  
SPDP – succinimidyl 3-(2-pyridyldithio)propionate)  
WLC – worm-like chain

# Chapter 1

## Introduction

Collagens are the most abundant proteins in mammals, responsible for providing structural support and mechanical stability to cells and tissues [1]. The collagen family has 28 known types, which are found in different parts of the body [1] and exist in almost every tissue with a mechanical role [2]. Most collagens form supramolecular assemblies and hierarchical structures: single collagen molecules assemble into highly organized fibrillar nanostructures, which later form more complex structures and composite materials. These structures are the foundations of different tissues with a variety of mechanical properties and biological functions [3].

Different types of collagen are found in different tissues [1]. Three fibrillar collagens, type I, II and III, are the focus of this work. Each plays a crucial role in structural support within the body. Type I collagen, the most abundant type, is the main constituent of skin, blood vessels and tendon, and is the organic part of bone. Type II collagen is the major protein constituent of cartilage. Collagen type III is mainly found alongside collagen type I.

Defects in collagen structure, production, or stability arising from genetic modifications can lead to a broad range of serious and sometimes life-threatening connective tissue diseases [4]. Chemical modifications at the molecular level can significantly affect mechanical properties of resulting structures. Some of these modifications are relevant to conditions such as diabetes and aging [5–7]. A wide range of connective tissue diseases exhibiting altered mechanical characteristics arises from point mutations in collagen (Table 1.1) [5,8]. For example, glycine substitution has been linked to several connective tissue diseases such as osteogenesis imperfecta and Ehler-Danlos syndrome [9]. Changes in the composition of the building blocks of collagen structures, *i.e.* single molecules of collagen, can lead to changes both in molecular stiffness (as found by simulations for glycine substitutions [10]) and in properties of the resulting tissues.

Mechanical properties of collagen have been the subject of numerous studies at the fibrillar, fiber and tissue levels [3]. However, mechanical properties at the molecular level are still unresolved. There are only a few studies directly examining structural conformations, stability and mechanical response at the single molecule level. There is a significant disagreement in the literature over both force-dependent stability and inherent flexibility of collagen molecules. Estimates

Table 1.1: Examples of connective tissue diseases arising from mutations in collagen (types I, II and III) and associated mechanical phenotypes. [5, 8]

Collagen Type	Related Disease	Target Organ	Description
I	Osteogenesis imperfecta Ehlers-Danlos syndrome Marfan syndrome Caffey disease	bone skin, joint arteries bone	brittle bones hyperflexibility of skin and joints vascular defects bone thickening
II	growth defects Legg-Perthes disease Stickler syndrome type I	bone and cartilage hip joint, eye, ear	dwarfism, bone abnormalities hip necrosis myopia, deafness
III	Ehlers-Danlos syndrome type IV Arterial aneurysms	arteries, intestine aorta	fragile blood vessels dilation of aorta

of the flexibility of collagen molecules, *i.e.*, persistence lengths, vary over an order of magnitude (Table 1.2) [11–20]. When compared to its 300 nm contour length, these values ranging from 12–170 nm describe collagen as a flexible polymer to a semi-rigid rod.

A close examination of the varying persistence length values in the literature suggests that a wide variety of experimental conditions (different collagen types and sources, pH and solvent salt concentrations) have been used. Carefully designed and controlled experiments and analysis in this work aims to address the inconsistency in the collagen flexibility at the basic molecular level. My focus on studying molecular stability and response of collagen to environmental factors can shed light on identifying underlying mechanisms of collagen-related diseases, as well as designing and producing improved engineered biomaterials with tunable properties.

**Collagen Structure:** Procollagen is the precursor form of collagen formed initially inside the cell. It consists of a 300-nm-long triple helix flanked by two globular domains, N- and C-propeptides. These bulky domains inhibit assembly of collagen into higher order structures such as fibrils. Upon secretion to the extracellular matrix, specific enzymes cleave the propeptide regions and collagen molecules assemble in a staggered fashion to form fibrils, which later develop into fibers and tissues [22] (Figure 1.1). Post-translational modifications enable crosslinking of neighbouring collagen molecules in fibrils, resulting in stable structures.

Each collagen molecule is made up of three chains, so-called  $\alpha$ -chains (Figure 1.2) [23, 24]. Each  $\alpha$ -chain is a left-handed polyproline II-type helix and consists of about 1020 amino acids. These three chains wind into a  $\approx$ 300 nm long and  $\approx$ 1.5 nm diameter right-handed triple helix with a short sequence ( $\sim$  20 residues) of non-helical structure (telopeptides) at each end [1, 25]. The triple helical region of collagen consists of a repeated sequence of three amino acids with glycine repeating at every third amino acid (Gly-X-Y). Glycine, the smallest amino acid, fits inside the triple helix and is required for compact structure of the helix. While glycine residues face the interior of the triple helix, the side chains of the X and Y amino acids are exposed to the exterior of the chain. The X and Y residues in collagen are often proline and hydroxyproline, creating "elbows" in each  $\alpha$ -chain [26]. X and Y residues can also be other amino acids such as lysine and

Table 1.2: Persistence lengths ( $l_p$ ) of collagen in the literature.

Technique	Collagen Source Type	$l_p$ (nm)	Solution
Optical Tweezers Stretching [11]	Recombinant Procollagen II	11	25 mM KCl, 1 mM EDTA, 0.2% Tween 20, 0.1% casein, 25 mM Hepes, pH 7.4
AFM Imaging [18]	Short model peptide	12	H <sub>2</sub> O
Molecular Dynamics Simulation [21]	Collagen type I	13	H <sub>2</sub> O, 10 mM NaCl
Optical Tweezers Stretching [19]	Recombinant Procollagen I	15	25 mM KCl, 1 mM EDTA, 0.2% Tween 20, 0.1% casein, 25 mM Hepes, pH 7.4
Molecular Dynamics Simulation [12]	Short model peptide	16	H <sub>2</sub> O, Neutral pH
Molecular Dynamics Simulation [13]	Model Peptide	51	H <sub>2</sub> O, Physiological pH
Electron Microscopy [20]	Skin of dermosparactic calves I	60	N/A
Viscometry [15]	Rat skin	130	60 mM potassium acetate, 0.001 M sodium bisulfate, pH 4.8
AFM Imaging [18]	Short model peptide	135	Fibril formation buffer
AFM Imaging [18]	Short model peptide	165	PBS
Dynamic Light Scattering [14]	Bovine dermal I	165	0.012 N HCl
Rheometry [17]	Skin of Lathyrtic rat I	170	150 mM NaCl, 50 mM Tris, 20 mM EDTA, 1mM N-ethyl maleimide, 5mM Benzamidine HCl, 1mM Phenyl methyl sulfonyl fluoride, 1µg/ml Pepstatin, 1µg/ml Leupeptin, pH 7.4

hydroxylysine. Exposure of ionizable residues to the exterior of the helix and their electrostatic interactions can play a role in stability of the triple helix [27, 28].

High-resolution X-ray diffraction studies of synthetic collagen-like peptides show that helical conformation of collagen molecule have subtle variations along its length, depending on its amino acid sequence [29]. It has been found that imino (proline)-rich regions form a tightly wound helix with 7/2 helical symmetry with an axial repeat of 20 Å [24, 29], while imino-poor regions adopt a 10/3 helical symmetry with 28.6 Å axial repeat leading to a less tightly wound (longer pitch) triple helical conformation [29, 30].

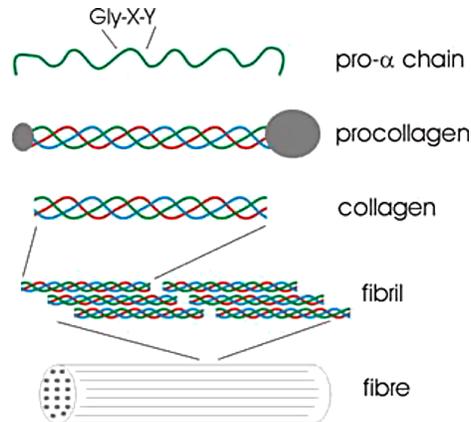


Figure 1.1: Schematic of collagen structure hierarchy (not to scale). Figure courtesy N. Forde.

It has been shown that hydrogen bonds between chains can stabilize the triple helix [29, 31]. Having hydroxyproline in the Y position in Gly–X–Y repeats also plays an important role in stabilization of the triple helix [1].

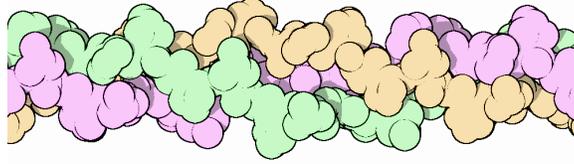


Figure 1.2: Part of collagen triple helix from x-ray crystallographic studies of a collagen-like peptide. Figure from 1BKV file of Protein Data Bank [32].

The three main fibrillar collagen types examined in this study are discussed below.

**Collagen Type I:** The most abundant collagen is type I collagen, a fibrillar protein found in bone, skin and other connective tissues. Normally collagen I is a heterotrimeric collagen with two  $\alpha 1(I)$  and one  $\alpha 2(I)$  chains encoded by *COL1A1* and *COL1A2* genes [5]. Collagen type I mutations are associated with a range of bone, skin and joint disorders (Table 1.1). A homotrimeric isoform of collagen I also exists, which consists of three  $\alpha 1$  strands. The homotrimeric collagen I is normally found in fetal tissues and it is also a pathogenic form associated with cancer and fibrosis [33–35]. The properties of homo- and heterotrimer collagen at the molecular level are compared below. Fibrils of homotrimeric collagen have been found to have greater thermal stability (or approximately the same, depending on solvent conditions) compared to the wildtype heterotrimeric form [36].

**Collagen Type II:** Collagen type II is the second-most abundant collagen and the main protein component of the articular cartilage. Collagen type II provides tensile strength to tissues and is responsible for guiding skeletal development [37]. Type II collagen is a fibrillar homotrimer encoded by the gene *COL2A1* [5]. Mutations and defects in collagen type II can disrupt development of bones and connective tissues, and are also associated with blindness [5].

**Collagen Type III:** A fibrillar homotrimeric protein, collagen type III consists of three  $\alpha 1(III)$  chains found in the extracellular matrix, skin and other extensible organs such as aorta, lungs and intestine, where it frequently accompanies type I collagen [37]. Mutations in its gene *COL3A1* can lead to fragile tissues and ruptured blood vessels [5]. In healthy fibrillogenesis of type I collagen in the cardiovascular system and other organs, type III collagen is an essential protein that regulates the diameter of type I collagen fibrils [38].

**Homo vs Hetero isoforms:** As mentioned, collagen type I is usually a heterotrimer, while a homotrimeric isoform also exists, which has altered mechanical properties of the resulting fibrillar structures. Previous studies of proteolytic cleavage of collagen type I indirectly compared the

stability of the naturally occurring heterotrimeric form with pathogenic homotrimers [39, 40]. These results indicate that the homotrimeric isoform has a higher local triple helix stability near the matrix metalloproteinase (MMP) cleavage site. It has been also found that homotrimeric collagen type I has a higher thermal stability compared to wildtype collagen I [39].

On the other hand, a simulation study compared mechanical properties of type I heterotrimer and homotrimer molecules of collagen [41]. In this study, which is based on a fully atomistic model of a small section of collagen triple helix, the mechanical response of both isoforms to an applied force was studied. The results showed that the hetero isoform exhibits less bending flexibility than the homo form. The authors concluded that kinks formed in the homotrimer collagen are likely to be responsible for reduced mechanical stability of the resulting tissues [41].

In all of the studies mentioned above, the normal heterotrimer isoform is compared with the pathogenic homotrimer isoform. The results of these studies contradict each other on the helix stability. However, very little is known about the mechanical properties of different types of collagen in their healthy occurring homo or hetero states. In this work, the overall stability and configurations of collagen types I, II and III are directly compared using atomic force microscopy imaging.

**Conformation and stability:** As discussed, most of the experimental studies investigated stability of collagen through its thermal denaturation or susceptibility to proteolytic cleavage [39–41]. In this work, I investigate the conformational stability of collagen by measuring persistence length using AFM imaging. My hypothesis is that the bending energy of collagen molecules is related to how tightly the helix is wound.

**Environmental factors:** Studying how environmental factors affect collagen conformation and functionality can help to understand underlying mechanisms of structural transitions and collagen mechanics. It has been shown that electrostatic interactions can affect the conformation and stability of the collagen triple helix [28, 29, 42, 43]. One study found, using circular dichroism and differential scanning calorimetry, that an increase in the ionic strength of solution increases collagen’s triple helical thermal stability [28]. Another study found, using atomic force microscopy imaging that collagen becomes less flexible as a result of increasing salt concentration [18]. In this work, we directly quantify collagen flexibility in solutions with different pH and ionic strengths, using images of individual collagen molecules obtained with AFM. The results show that ionic strength and pH of the solution strongly affect the flexibility of collagen.

**Force-Induced Structural Transitions.** Collagen’s mechanical response plays a crucial role in load-bearing tissues. Therefore, force-dependent stability of collagen molecules is of biological relevance. Collagen’s triple helical flexibility and molecular response to the force can reveal mechanical properties at the fundamental level of collagen hierarchical structures.

Previous studies used optical tweezers to probe elastic response of isolated types I and II procollagen molecules [11, 44]. The observed nonlinear force-extension behaviour of procollagen at low forces ( $< 10\text{pN}$ ) was fit to the worm-like chain (WLC) model of polymer elasticity, which is the same model that has been applied successfully to describe the force-extension relationship of double-stranded DNA at low forces [45, 46]. The WLC model assumes that force is required to overcome entropic elasticity of the system, without any structural deformations such as change in bond length or bond angles. Sun *et al.* found persistence length values of  $\approx 15\text{ nm}$  for procollagen molecules for forces up to  $10\text{ pN}$  [11, 44]. Computer simulations have also investigated structure and mechanics at the molecular level using collagen model peptides, finding a similarly low value of persistence length [12, 21]. These values describe collagen as a flexible molecule and are significantly shorter than persistence length values obtained by most other measurement techniques (Table 1.2).

How does the structure of collagen's triple helix change as a result of force? There are three competing hypotheses: collagen mechanics is solely governed by entropic elasticity as suggested by the earlier optical tweezers studies at low force [11, 44]; cooperative melting/unwinding occurs over a narrow force range, as with DNA [47, 48]; or helix denaturation occurs continuously during collagen stretching, as has been suggested by MD simulations [49]. Some experimental studies of proteolytic cleavage of single molecules of collagen suggest that there is a force-mediated local unwinding of the triple helix [50, 51], while others suggest that force tightens the triple helix [52]. By performing optical tweezers experiments, we explore mechanical properties of collagen under force to unravel the flexibility of single proteins as well as force-induced structural changes.

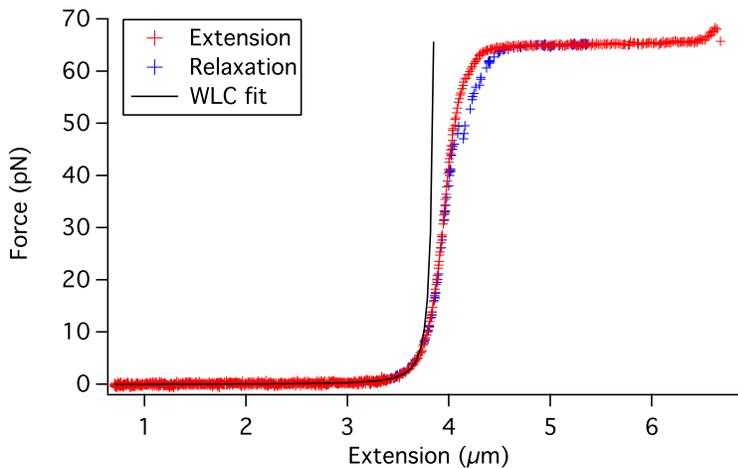


Figure 1.3: Force-extension curve of a 11.7 kbp double-stranded DNA manipulated in our OT instrument. Stretch and relaxation data are shown as points. The black line represents a worm-like chain (WLC) fit to low force data ( $< 10\text{pN}$ ) giving a persistence length of  $50\text{ nm}$  [45, 46].

As an example of this approach, Figure 1.3 shows the force-extension curve of a  $4\text{ }\mu\text{m}$ -long dsDNA molecule stretched in our optical tweezers instrument. These data illustrate that the WLC model is valid only at relatively low forces, consistent with findings of others [45, 46]. At higher

forces there is a disagreement between the WLC model and experimental data points, indicating that DNA mechanics is no longer only governed by entropic elasticity and that deformations occur. The DNA double helix undergoes an extreme structural transition at  $\approx 65$  pN of force [47].

**Overview:** In this thesis, two different experimental techniques are used to characterize mechanical properties of collagen: studying the flexibility of collagen by imaging individual collagen molecules using atomic force microscopy (AFM) (Chapter 3); and studying the response of single collagen molecules to applied force using optical tweezers (OT) (Chapters 5 and 6). For direct study of collagen molecular conformations on surfaces, an algorithm was developed for tracking individual molecules and for performing statistical analysis of AFM-imaged chains (Chapter 2). Conformations of different types of collagen from different sources are directly compared under the identical ionic strength and pH (Section 3.2). AFM images of type I collagen from rat-tail tendon are studied under different salt concentrations and pH (Section 3.3). The second half of the thesis focuses on collagen's response to force applied using optical tweezers. Chapter 4 gives an overview of the technique, while chapters 5 and 6 represent published work on our collagen production system and optical tweezers studies of collagen. The thesis concludes with a summary and proposed directions for future experiments.

## Chapter 2

# Atomic Force Microscopy: Image Analysis of Single Polymers

The Atomic Force Microscope (AFM) was invented in 1986 by IBM scientists, Gerd Binnig and his colleagues, to investigate surfaces of insulators at the atomic scale [53]. Since then, it has been widely used in different fields such as material science, physics, biology and chemistry [54, 55]. An AFM is a tool for high-resolution imaging of surfaces at the atomic level and characterizing properties of materials [56]. In addition to imaging, AFMs are also used for stretching molecules, microindentation and surface manipulation studies [57, 58]. An AFM can be used to probe dried or wet samples. An AFM has advantages over traditional microscopes. Unlike conventional microscopes, the resolution is not limited by the diffraction limit of visible light; it is rather defined by the size of the probe (also referred to as the tip) used to investigate the surface. It also enables three-dimensional measurements of surface features.

Because they can image at high resolution, AFMs have been widely used to study single biological molecules such as DNA and proteins. Mechanical properties of single polymers can be deduced from AFM images [59]. Sample preparation for AFM imaging is relatively straightforward. Molecules can be imaged in physiological conditions without the need to chemically modify the protein. This makes the AFM a powerful analytical tool to study proteins such as collagen and extract their configurations and mechanical properties in different pH and ionic strengths. This chapter describes the AFM instrument, analysis method along with its verification, and foundations of the statistical analysis used for collagen studies.

### 2.1 AFM setup

A typical AFM comprises a sharp tip or probe attached to a weak cantilever spring that deflects when it touches a sample substrate. The deflection of the cantilever is measured by a laser beam which reflects from the back of the cantilever and is directed to a split photodiode detecting displacements with sub-Ångström resolution [60]. The substrate and the tip are moved relative to

each other with a piezoelectric stage enabling scanning of the surface. A feedback loop regulates the force applied to the surface by the tip via the height of the tip relative to sample. The amount of force against the substrate is given by the product of the spring constant of the cantilever with the displacement of the cantilever (obtained from the deflection of the laser beam from the back of the cantilever). A schematic of the AFM setup is shown in Figure 2.1.

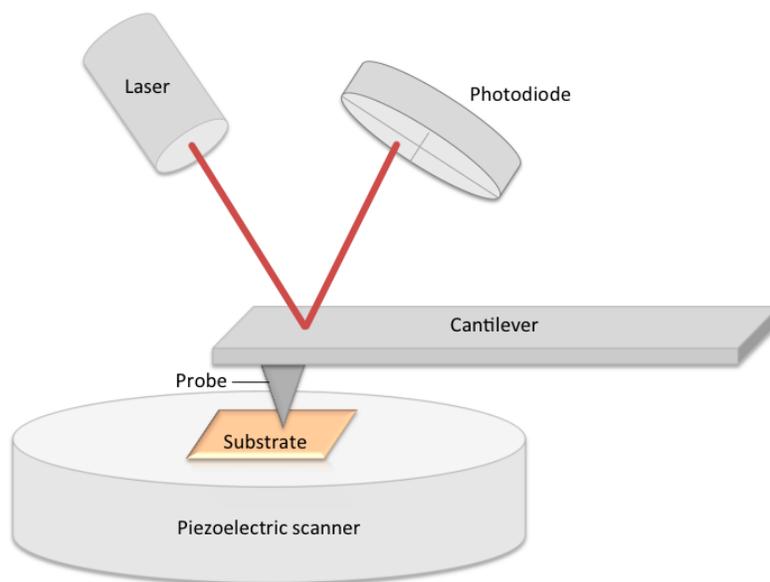


Figure 2.1: Schematic of the AFM setup.

AFMs operate in two fundamental modes: contact mode and tapping mode (Figure 2.2). In contact mode, the tip and the surface are always in contact (*i.e.* the force on the probe is in the repulsive region). The tip drags on the surface, experiencing a finite adhesive force. In this mode, the feedback circuit keeps the force constant during the scan. The lateral force exerted on the sample during the scan can be quite high, resulting in damaging the sample or displacing relatively loosely bound molecules along the surface. In tapping mode, the tip oscillates up and down at the resonant frequency of the probe-cantilever system. The tip comes into contact with the surface for a short time at discrete locations (pixels). Tapping mode is used throughout the AFM studies in this work. The advantage of tapping mode over contact mode is that it avoids the lateral force issue and reduces the possibility of deformation or displacement of molecules on the substrate while scanning.

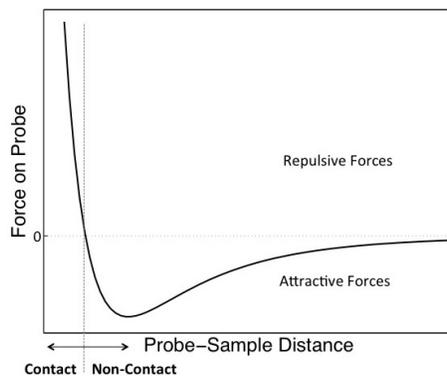


Figure 2.2: The repulsive and attractive force regimes as a function of sample-probe distance in an AFM instrument (Adapted from Reference [61]).

## 2.2 AFM Image Analysis Program

### Tracing Single Molecules

In this work, the goal of AFM imaging of single chains is to analyze their configurations and extract their mechanical properties. The first step in analyzing the images is to trace each molecule's backbone, to later perform statistical analysis.

For this purpose, first the centreline of each molecule needs to be traced. This can be done in a way ranging from a fully manual to a fully automatic approach depending on the images. A number of available chain tracing algorithms, ranging from fully automatic to semi-automatic, were tested on collagen chains imaged in this work (see Chapter 3 for experimental details) [62–64]. These programs either required extensive user intervention for each separate chain or did not successfully trace our collagen images and lacked precision. The fully manual tracing strongly depends on human knowledge and the results might change from one operator to the next. For proper statistical analysis of the molecules, it is necessary to have data from a large number of chains. When it comes to gathering this much data, manual tracing would be extremely time-consuming. As an example, the code used in Reference [62], searches for the brightest point along each chain, which can result in erroneous detection of the backbone of the chains with inhomogeneous brightness or adjacent parts of the chain depending on the size of search-lines (Figure 2.3). It also requires the user to introduce several points on the backbone of the chain. If the points are not located on the chain or fewer points are used, it can lead to detecting contours that clearly do not follow correctly the original imaged chains (Figure 2.3). It is a time-consuming process to introduce "enough" initial points located on the backbone of the chain.

The fully automatic tracing in reference [63], on the other hand, does not have these issues. It is fast and not prone to bias from the user. With automatic tracing, however, there is little control over the process. I tested the algorithm in Reference [63]. Automated tracing of images also needs low-noise images. If the quality of the image is low or parts of the chain in the image have different brightness or are faded compared to other parts, the detection algorithm may not

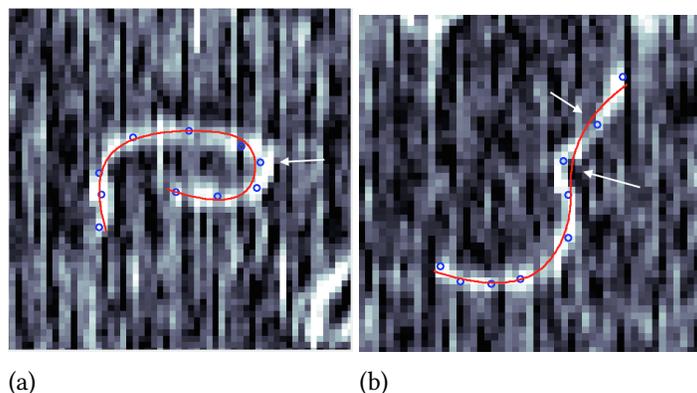


Figure 2.3: Examples of tracing noisy images of collagen chains using the Easyworm algorithm from Reference [62]. Blue dots represent initial user-selected points. Arrows indicate regions in which the identified contour (red line) clearly does not overlap with the (bright) backbone of the polymer from the original image.

correctly detect the chain (Figure 2.4). Although it is a fully automated tracing program, user intervention is needed to excluded erroneously detected traces. For example, bright background (noise) lines can be detected as chains, which need to be removed by user. Although not impossible, it is a labour-intensive process when hundreds of them are detected. Hence, a good approach is to use human supervision to eliminate erroneously detected chains. Alternatively, by adjusting algorithm parameters to exclude more noise lines, many chains can be left undetected, which is also not desirable.

An intermediate method would combine manual and automatic tracing. Therefore, a custom-built MATLAB code, **SmarTrace**, was developed to analyze AFM images of single molecules (Figures 2.5–2.7). This program requires minimal user intervention, to introduce only a few points near each chain (this step is very fast), and noisy images or chains with uneven brightness can be detected. The algorithm is described in this chapter.

The performance of this newly developed algorithm was validated on a series of test images of DNA with known persistence length. This new method is not very sensitive to noise and can successfully detect the chain backbone in noisy images. If parts of the chain are slightly faded, the code successfully traces the whole chain without cutting the rest of the chain (Figure 2.6). The SmarTrace program can successfully detect the backbone of the chain images even in a noisy environment with few user-defined points that show the whereabouts of the chain. The results do not depend on where the user selects the points, and the initially selected points do not need to be located exactly on the chain. These features, along with an efficient computation algorithm, make the code very fast and easy to use for tracing and analyzing images of single polymers, as obtained for example by AFM or EM (electron microscopy). The program traces the chain backbone with sub-pixel resolution using a pattern-matching algorithm.

Here is an overview of how SmarTrace works:

1. The program starts by asking the user for an initial manual trace of a chain.

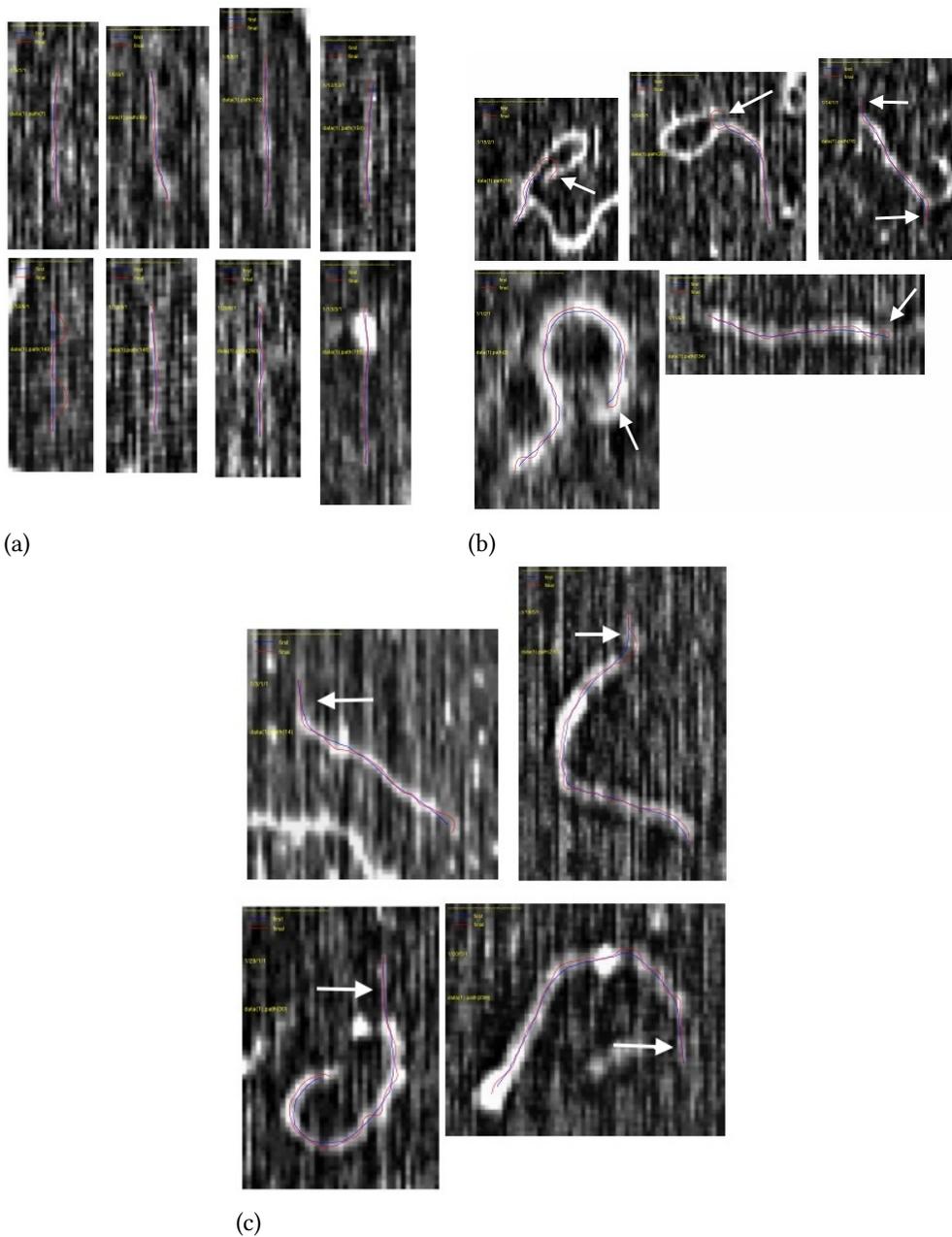


Figure 2.4: Examples of tracing noisy images of chains using Reference [63]. The red line represents the final traced backbone of the chain. This automated chain-tracing program requires user intervention at the end to exclude erroneous traces. For noisy images, there are issues (arrows) with (a) picking up many noise lines as molecules, (b) detecting adjacent parts of the chains, or highly flexible ends, and (c) detection of the ends of the chains. The algorithm leaves many chains undetected.

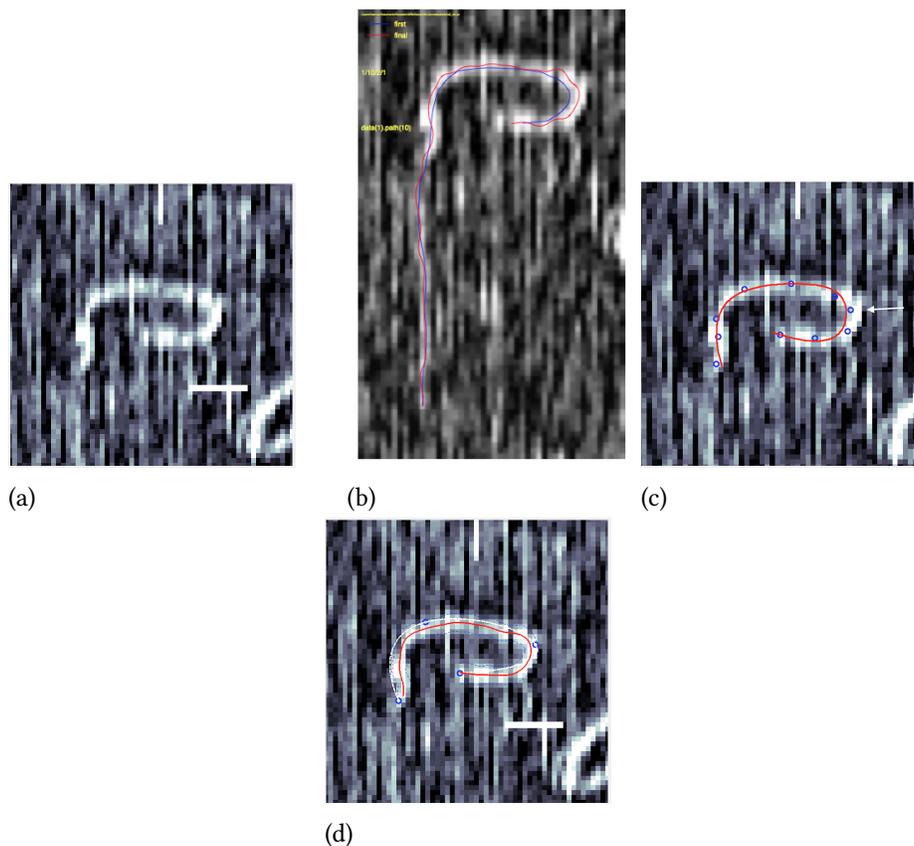


Figure 2.5: Examples of chain mistracing by other algorithms and more accurate tracing using SmarTrace. (a) A noisy image of collagen. (b) Issues with detection of ends of the chain by the fully automatic tracing algorithm from reference [63]. The red line is the traced backbone of the chain. Traced  $L = 491$  nm. The algorithm also leaves many chains undetected. It is very sensitive to noise. (c) An example of inaccurate tracing of the backbone of the chain from the method of [62]. Points selected by user (open circles). Traced  $L = 269$  nm. (d) Successful tracing of a chain (red line) using the newly developed algorithm, SmarTrace. The dashed white line is the initial B-spline fit to the selected points (open circles), and the dashed blue lines represent the calculated width of the chain. Traced  $L = 273$  nm. Scale bar is 50 nm.

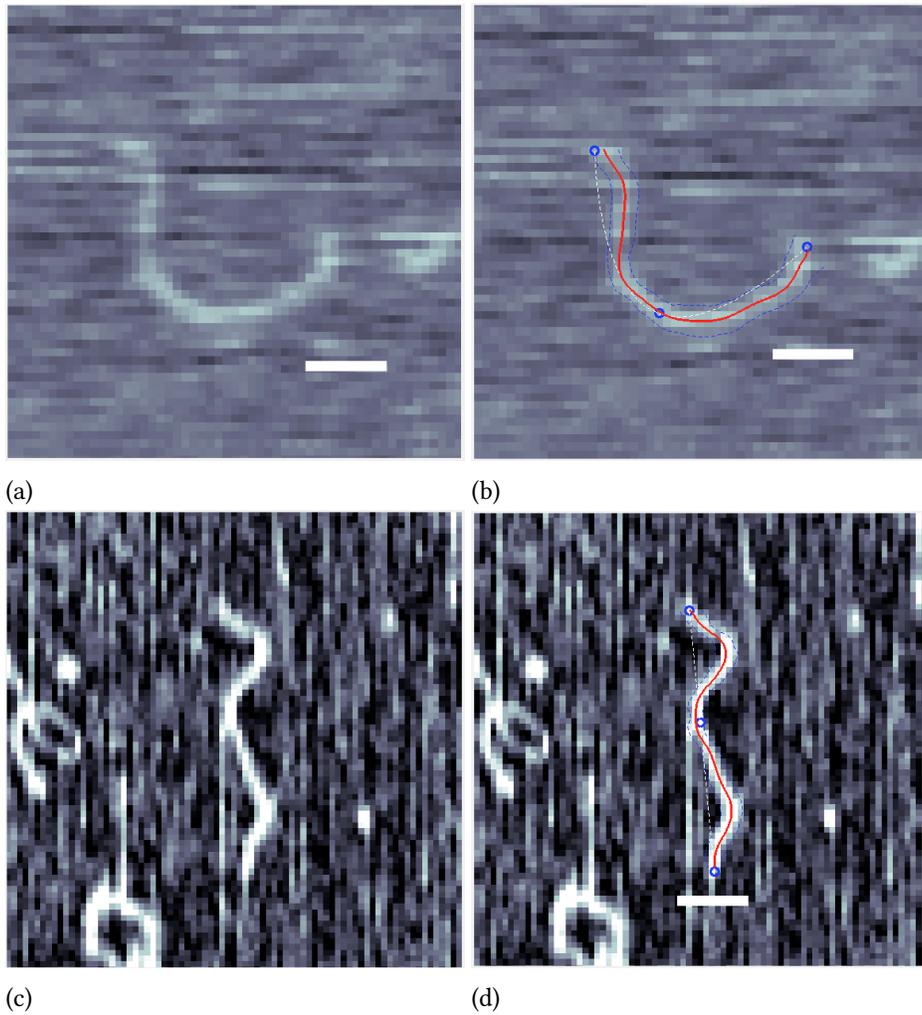


Figure 2.6: (a, c) Examples of images of collagen on the surface of mica. (b, d) Traced backbone of chains images by SmarTrace program. Only three points selected by the user (open circles); the initial B-spline fit to the already selected points is also shown (dashed lines). This spline is the starting point for the algorithm to detect the chain backbone. The final traced path is shown in red, and the blue lines represent the detected width of the chain. Noisier images of collagen are chosen to show the capabilities of the SmarTrace algorithm in tracing the chains even in a noisy image. (b) Traced  $L = 260$  nm. (d) Traced  $L = 237$  nm. Scale bar is 50 nm.

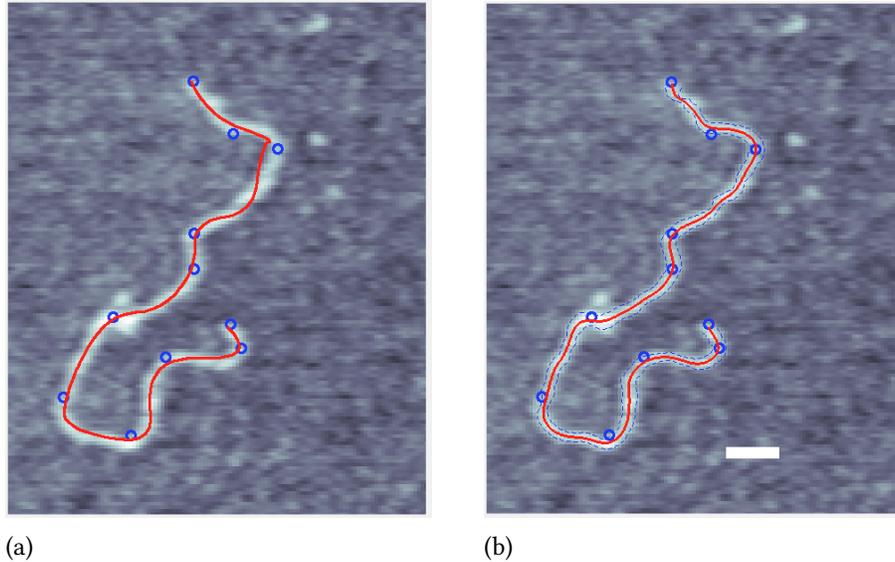


Figure 2.7: Examples of tracing noise-free images of a DNA chain (provided in the Easyworm package [63]). Same user-defined points are used to trace the chain using (a) Easyworm (traced contour length = 910 nm), and (b) SmarTrace (traced contour length = 909 nm). More user-defined points are required for Easyworm to trace the backbone of the chain more precisely. The scale bar is 50 nm.

2. It then fits an initial spline to these defined points.
3. The program removes the background and enhances the image.
4. To automatically detect the best path describing the chain, a rotating window is slid along the tangential direction of the initial spline.
5. Looking only at the window around a point on the chain, a pattern is matched against the chain.
6. A matching score is calculated for each possible width and location of this pattern. The goal is to find the best width and location of a center point on the path at which the pattern best describes the chain.
7. The traced points are then fine-tuned and the location and width of the chain are extracted.

These steps are described in detail below.

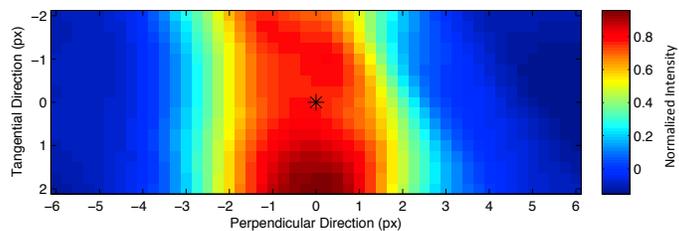
### SmarTrace Implementation Details

1. **Initial Manual Trace:** In a visual user interface (adapted from the EasyWorm package [62]), the user selects points ( $P_{\text{user}}$ ) on a chain roughly tracing the center points of the chain. Simultaneously, an initial spline curve is fit to these selected points and shown back to the user. This spline can be used to guide the user as to where to click such that the tentative

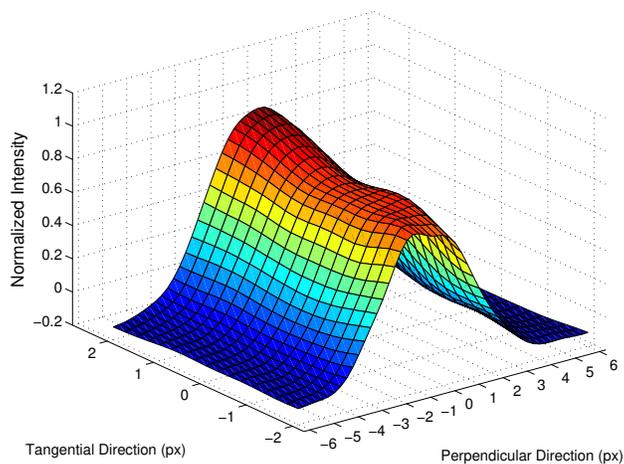
spline curve almost passes through the chain centre line (Figure 2.6). The user interface allows users to rapidly trace the whole chain by giving them options to zoom in on the chain, delete and re-select the points interactively.

2. **B-spline Fit to User Points:** After the user selects the chain, an initial spline curve is fit (cubic spline interpolation) to the X and Y coordinates of these user-defined points (Figure 2.6). B-spline curves, piecewise polynomial interpolations, are widely used in computer vision and computer graphics to represent curves and objects [65]. The B-spline fit is then used to extract equally distanced points along the curve ( $P_{\text{spline}}$ ). Each point is equally separated by a distance equal to one pixel along the spline.
3. **Image Enhancement:** To improve image quality, a background removal step is performed on the grayscale version of the image. Background removal is done using top-hat filtering, a standard method used in computer vision [65], with a disk-shaped smoothing filter. The second step after background removal is noise removal. A median filter reduces speckle-type noise but keeps the image sharp [65].
4. **Search Window:** For each point along the B-spline, tangential and perpendicular directions are defined based on the B-spline curve. A step-wise two-dimensional grid is created along these directions by extracting pixel intensity values at each grid point. Step sizes are selected to be fractions of one pixel and a cubic interpolation method is used to estimate sub-pixel values within this new, tangentially aligned grid. This high-resolution grid of intensity values should contain a small and almost vertically aligned segment of the chain (Figure 2.8a).
5. **Pattern Matching:** In this step, a search algorithm extracts the centre position of the chain for each point ( $P_{\text{spline}}$ ) along the chain. A template is matched against the chain's image in the search area. The goal is to find the width and location where the template best matches the image of the chain. A template that resembles the intensity pattern of a cross-section of the chain is used. For different widths of the template, the cross-correlation of the template and the chain in the search area are calculated.

In this work, the template used for the search is a basic fourth-degree (quartic) polynomial of form  $y = (\frac{w}{2})^4 - (\frac{w}{2} - x)^4$  with varying widths  $w$  (Figures 2.8b and 2.9). The minimum  $w$  is 1 pixel, and the maximum  $w$  equals to the width of the search grid in the perpendicular direction. In implementing this polynomial, only the positive values are kept; the rest are set to zero. The area under the template curve is kept constant. Two borders with negative values are added to the polynomial curve, to boost the score when the edges of both template and the chain match. The quartic 1D template is then repeated along the chain tangential direction to form the final filter (Figure 2.9). This template will now best match when exactly placed on the centre of a completely vertical chain. Additionally, the finite length of the filter in the tangential direction makes it less sensitive to noise in the perpendicular direction.

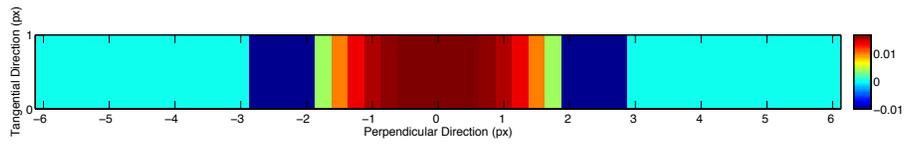


(a)

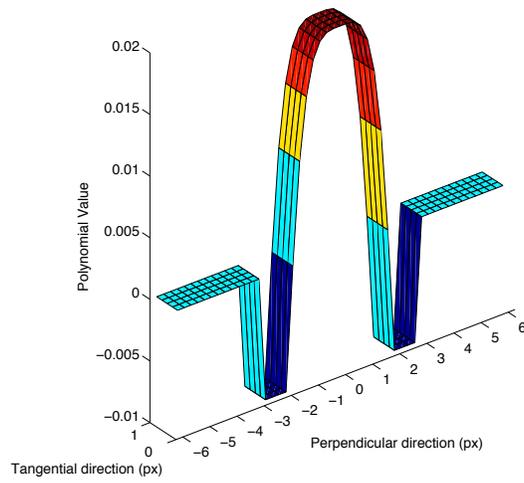


(b)

Figure 2.8: (a) Intensity (height) map of a subsection of a chain in the search window. The star represents a point of the B-spline ( $P_{\text{spline}}$ ) around which the search grid is formed. The size of the search grid can be adjusted based on the image of the molecule. Here, it is set to be 13 by 5 pixels. (b) 3-dimensional intensity plot of this subsection of the chain in the search area.



(a)



(b)

Figure 2.9: (a) Intensity plot, and (b) 3-dimensional plot of the quartic polynomial filter with  $w = 4$  pixels.

6. **Scoring:** Using these generated filters at different widths of  $w$ , a pattern-matching algorithm gives an initial score to each point in the search grid. The scores are determined by calculating the cross-correlation between the intensity values in the search grid centred at that point and each template of width  $w$ . This score represents how well the centre and width of the chain matches the template of the same width in the search grid. The result of this step is a cross-correlation value for the template of each different width centred at each point in the search grid.

The cross-correlation scores are used to determine the best centre line position and width of the contour. Rather than directly pick out the maximum score, a few additional steps are followed.

First, the cross-correlation scores for each 2D search window are collapsed into a 1D window perpendicular to the original spline. The aggregate score for each point is calculated by a weighted Gaussian average of scores in the vertical (tangential) direction. This average smooths out the noise while most heavily weighting the region in the centre of the search window near the original  $P_{\text{spline}}$  point. This allows small disconnects in the chain to still be given a high matching score, if most points directly below and above have high values. This scoring process is repeated for all the points in  $P_{\text{spline}}$  along the chain. The results are then fine-tuned for stability as follows.

7. **Fine-tuning the Scores:** Based on the scores, the chain can be defined in terms of its width and centre at each point along its length. The most straightforward way to do this would be to pick the highest score for each tangent location. Instead, one final step involves fine-tuning to achieve a more stable result and to reduce the effect of image noise. If there is noise close to a part of the chain or where the chain crosses itself, it can lead to an abrupt change in the detected width of the chain. By adding two more stability fine-tuning steps, the results become more stable and less sensitive to the noise. We achieve stability by re-scoring the pattern-matching scores with two additional measures: width stability and direction stability.

**Width stability:** In order to achieve width stability, cross-correlation scores are re-weighted based on the difference between their width and an average, smoothed value of the width in that region of the chain. To do this, first the highest scoring width for each location along the chain is chosen. These width values are smoothed with a moving average, providing a smoothly varying chain width. Following this step, the original scores for each different width of the template are reweighted, based on  $\Delta w = w_l - w_{l,\text{smooth}}$ , such that a smaller  $\Delta w$  has a larger Gaussian weight, and vice versa. Therefore, scores that deviate too much from the smoothed width are assigned a lower score. If there is a sudden change in the chain image as a result of noise, this criterion provides a penalty, and the width results are stabilized.

**Direction stability:** For this step, the scores are fine-tuned in the dimension representing centre values. For the next point to describe the chain, the centres are re-weighted based on their distance from the current chain direction. For each point  $l$  along the chain, the current direction of the chain is linearly projected from the previous tangent:  $P_{\text{spline}}(l - 1) - P_{\text{spline}}(l - 2)$ . The perpendicular distance from the current direction for each point  $l$  is calculated and given a Gaussian weight based on the distance from the chain direction: the higher the distance, the lower the score. The score function is chosen carefully to penalize sharp bends along the chain, thus preventing the calculated centreline from unnecessary oscillations.

**Final results:** After the scores are finalized, the re-weighted maximum scores are chosen to extract the width and centre location of the chain at each length. The centre-line locations are separated by approximately 1 pixel along the chain. A B-spline is fit to these centre-line locations resulting in a piecewise smooth polynomial that represents the polymer chain (Figure 2.6). This spline is used in subsequent chain conformation analysis.

### Sampling Chain Segments

After the chain coordinates are obtained, the traced chains are used for statistical analysis of their flexibility. To do so, first, different contour length segments of the chains are extracted. The chain-sampling approach used in this work modifies the method introduced in [63] as follows: Each molecule is randomly divided into segments of lengths drawn from a set of input values (e.g. 10 nm, 20 nm, 30 nm, ...). No chain segment is used twice. The maximum segment length does not need to be the same as the chain contour length. This sampling method allows the use of partially traced chains, which is particularly helpful when ends of a molecule are not clear or chains intersect. This is seen, for example, in collagen images in low pH and low salt (Figure 2.10). As rare, longer segments are less useful for mathematical fitting and persistence length calculations, choosing a relatively short maximum segment length provides more data for statistical analysis.

After the segments are sampled, the segment sizes are shuffled to avoid accumulating shorter lengths towards one end of the molecule (which was the case in the original sampling method in [63]). This process is repeated 50 times, to sample different parts of each chain. Quoted error bars are calculated based on the standard error of the mean for the data with that length of segment.

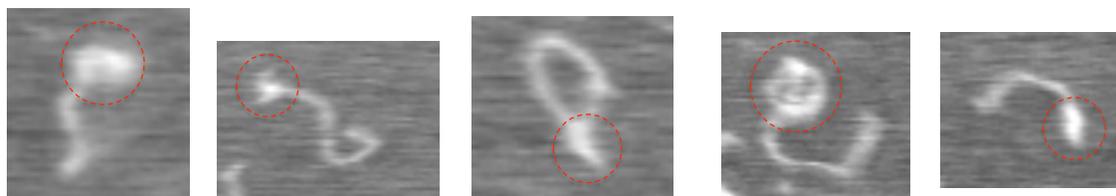


Figure 2.10: Examples of collagen AFM images with overlapping and unresolved ends (red circles).

## 2.3 Principles of Data Analysis for Traced AFM Images of Single Polymers

We are interested in the mechanical properties of the imaged polymers. Persistence length, the length over which the correlation of chain tangent vectors drops by a factor of  $e^{-1}$ , characterizes the chain's flexibility. After the coordinates of chains are traced, and segments of different contour lengths are sampled for analysis, several complementary techniques are applied to calculate the persistence length of the traced polymers. One standard technique is to determine the mean squared end-to-end distance as a function of length along the contour. A second is to determine the mean cosine of the angle between tangent vectors as a function of separation along the contour. Also, from the probability distribution of angles at each given separation along the contour, the form of the bending potential can be determined. For each of these approaches, mathematical models are fit to the resulting data, to estimate the persistence length.

Once deposited on the surface, the molecules can equilibrate in two dimensions before sticking to the surface (2D), can be kinetically trapped (3D) or can otherwise be affected by the surface. In 2D equilibrium, the molecules can rearrange themselves in the vicinity of the surface and then bind to the mica, while in the 3D limit of kinetic trapping, the configuration of molecules in the solution is projected onto the two-dimensional surface of the mica [66]. Each of these scenarios results in different formulations of the statistical parameters of the chain. Here, I focus on the 2D equilibration conditions. The following paragraphs summarize the analytical expressions used in this work to extract persistence lengths from traced chain contours.

The probability of finding a system at energy  $E$  is given by the Boltzmann distribution.

$$P(E) \propto e^{-\frac{E}{k_B T}}. \quad (2.1)$$

According to elasticity theory, the energy of a semi-flexible polymer for small bend angles on a two-dimensional surface is given by [67]

$$E(\theta, s) = \frac{k_B T l_p}{2s} \theta^2, \quad (2.2)$$

where  $k_B = 1.38 \times 10^{-23} \text{ J.K}^{-1}$  is the Boltzmann constant,  $T = 293 \text{ K}$  is the absolute temperature of the laboratory room,  $l_p$  is the persistence length and  $\theta$  is the bending angle between two segments of the polymer separated by a distance along contour  $s$ . The probability density combining 2.1 and 2.2 gives for bending angle  $\theta$

$$P(\theta; s)_{2D} = \sqrt{\frac{l_p}{2\pi s}} e^{-\frac{l_p}{2s} \theta^2}. \quad (2.3)$$

Since the bending energy is a quadratic function of bending angle, the odd moments of the angle distribution are zero. The second moment is

$$\langle \theta^2 \rangle_{2D} = \frac{s}{l_p}. \quad (2.4)$$

The mean square end-to-end distance between segments of a semi-flexible polymer separated by a distance  $s$  equilibrated in 2D is given by [59]

$$\langle R^2 \rangle_{2D} = 4l_p s \left( 1 - \frac{2l_p}{s} \left( 1 - e^{-\frac{s}{2l_p}} \right) \right), \quad (2.5)$$

and the average cosine of the angle between the segments at a distance  $s$  in 2D is given by

$$\langle \cos \theta \rangle_{2D} = e^{-\frac{s}{2l_p}}. \quad (2.6)$$

The assumption of a harmonic chain can be tested experimentally: the bending energy can be obtained from the angular probability distribution

$$\frac{E(\theta; s)_{2D}}{k_B T} = \text{constant} - \ln(P(\theta; s)). \quad (2.7)$$

If the chain energy is harmonic in bending angle, the energy can be written in the form of

$$E(\theta, s) = A + B(s)\theta^2. \quad (2.8)$$

Comparison with equation (2.2) shows that  $B$  is related to persistence length as

$$B = \frac{k_B T l_p}{2s}. \quad (2.9)$$

## 2.4 Validation of the Image Analysis Method Using AFM Images of DNA

The chain tracing and analysis method are validated by analyzing a series of images of DNA with known persistence length. In this section, analysis results for DNA molecules are presented as a validation for the analysis method described above.

Images of DNA molecules with 1  $\mu\text{m}$  contour length were obtained under standard conditions corresponding to 2D equilibrated chains and a persistence length of 52 nm [62]. These were provided with the software package Easyworm [62]. A total of 24 molecules was used in the analysis. An example of DNA chains traced with the SmarTrace algorithm is shown in Figure 2.11.

After the chains were traced, each chain was randomly chopped into segment lengths of multiples of 10 nm, with a minimum length of 10 nm and a maximum of 300 nm. This step was repeated 50 times, and the resulting segments were pooled for analysis by their segment length. The result-

ing distribution of segment lengths is shown in Figure 2.12.

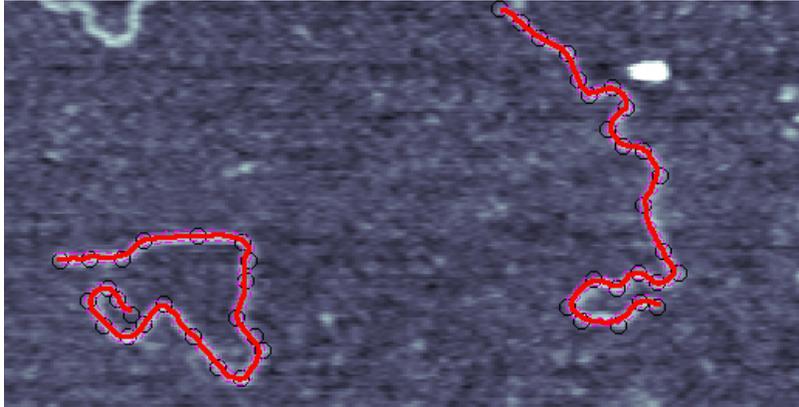


Figure 2.11: Example of DNA strands analyzed with SmarTrace (Section 2.2). AFM image of DNA image is provided in the Easyworm package [62].

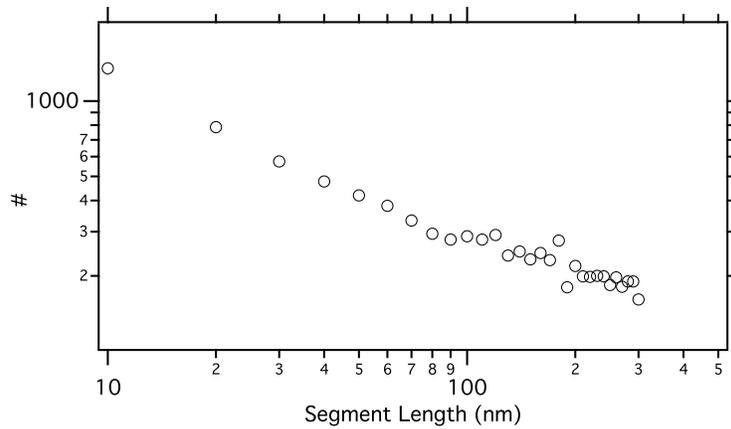


Figure 2.12: Distribution of sampled segment lengths for 24 DNA molecules of 1  $\mu\text{m}$  contour length, after 50 resamplings of the parent chains.

The experimental results for mean square end-to-end distance,  $\langle R^2 \rangle$ , as a function of length are shown in Figure 2.13 along with a 2D worm-like chain fit (Equation 2.5). Figure 2.14 shows the experimental data for tangent-tangent correlation,  $\langle \cos \theta \rangle$ , vs length and the 2D WLC fit (Equation 2.6). A persistence length of 52 nm is obtained from each of these analyses, which is expected for DNA [46, 62].

Analysis of the bending angle distribution leads to the form of the bending energy (Equation 2.7). All distributions seem relatively well described by harmonic potentials (Figure 2.15). The angular distribution at 10 nm deviates slightly from a normal distribution, which could be due to surface-induced curvature and/or local properties or defects on mica [68, 69]. Because of this, 10 nm separations were not included in the following analysis. The quadratic fit results from Figure

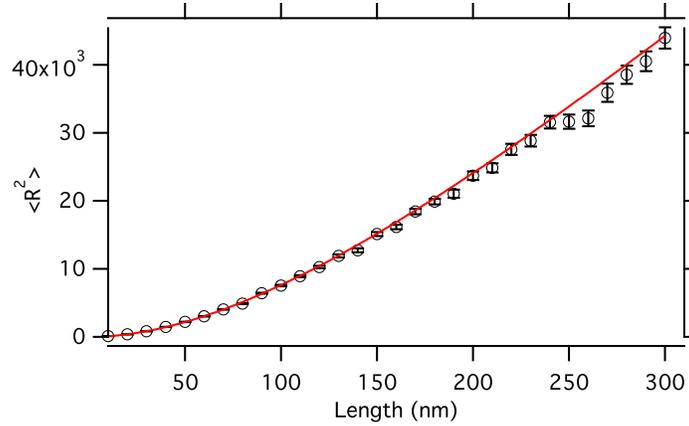


Figure 2.13: Mean square end-to-end distance vs. length along the DNA molecules and 2D WLC fit. Symbols represent the mean for each segment length while error bars show the standard error of the mean. The persistence length from the fit is  $52 \pm 1$  nm (reduced chi-squared  $\chi_{red}^2 = 5.5$ ).

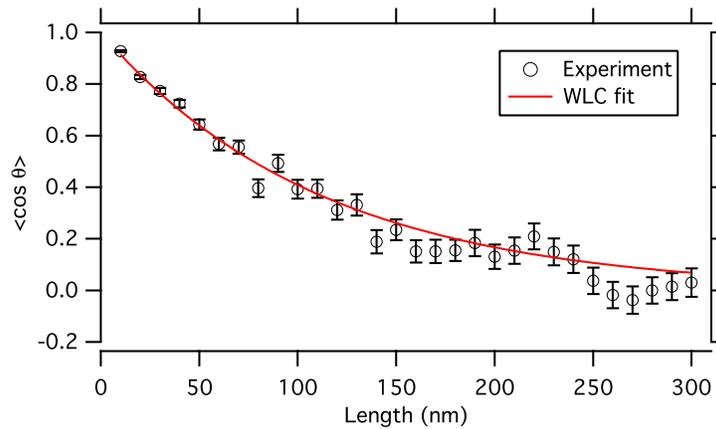


Figure 2.14: Average tangent-tangent correlation vs. length along the DNA molecules and a 2D WLC fit. The persistence length from the fit is  $52 \pm 3$  nm (reduced chi-squared  $\chi_{red}^2 = 2.2$ ).

2.15 are used to extract the persistence length using Equation 2.2. Figure 2.16 shows the compiled results of this analysis to obtain the persistence length, giving 53 nm.

The results of all of these analyses give the expected value of  $l_p \approx 52$  nm for DNA. In the next chapter, the SmarTrace algorithm is used in analysis of collagen images obtained with AFM.

## 2.5 Summary and Discussion

From AFM images of single molecules, their mechanical properties can be extracted. In this chapter, the analysis tool developed to extract mechanical properties from obtained AFM images of single molecules of collagen was described. The AFM images are first traced using a program I developed in MATLAB, SmarTrace. It is optimized for AFM images of single molecules in both noise-free and noisy environments. SmarTrace performs high-precision detection of centrelines

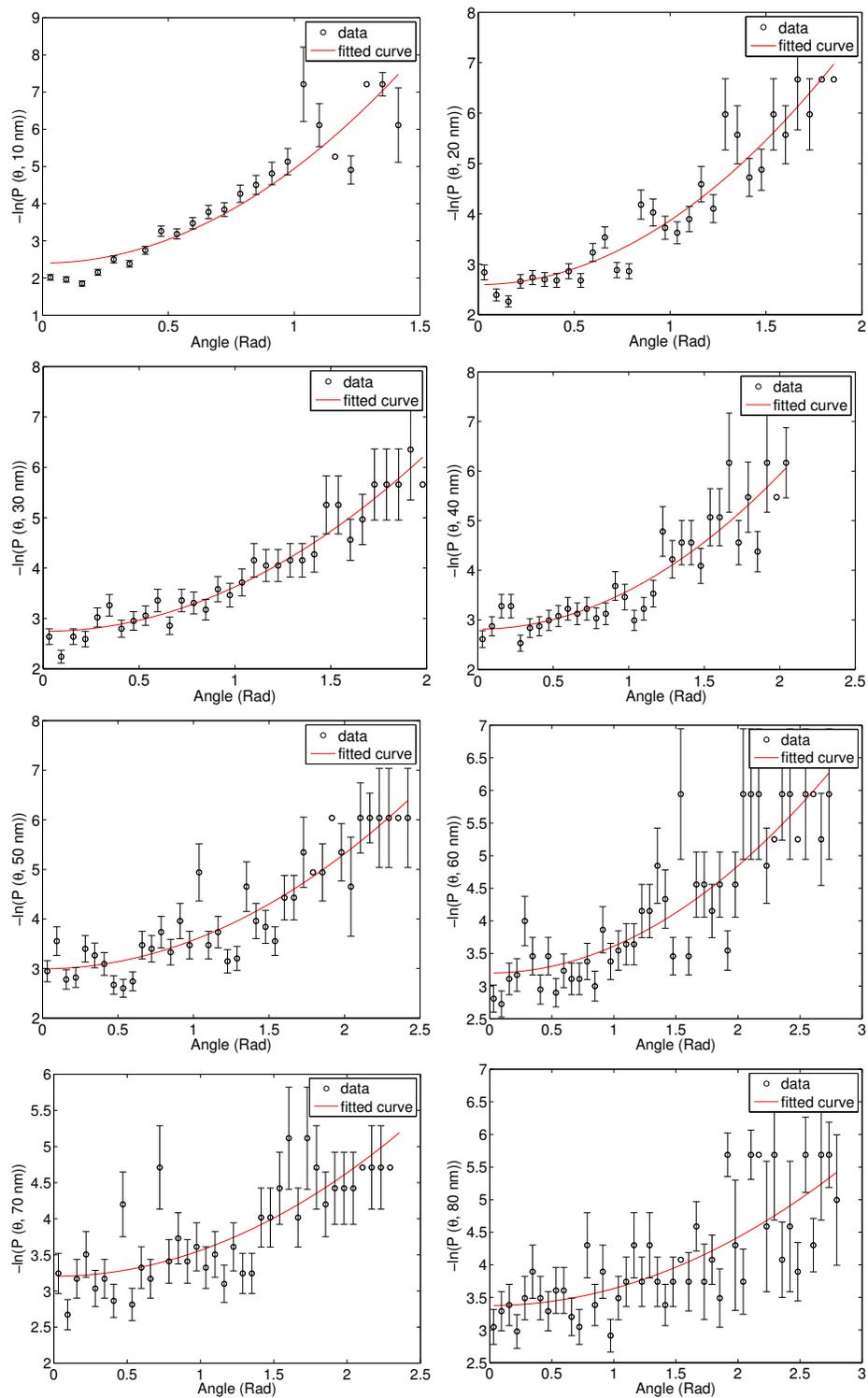


Figure 2.15: Determination of bending energy at separations of 10, 20, 30, 40, 50, 60, 70 and 80 nm along the contour, with fits to harmonic potentials. The error bars represent  $\sqrt{N}$ , where  $N$  is number of counts in each angle bin.

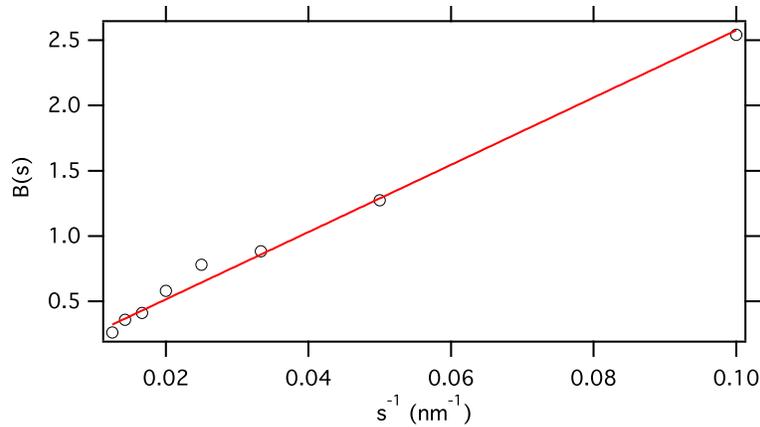


Figure 2.16: Extracting persistence length from the angular distribution fit parameter  $B$  (Equation 2.9). The persistence length from the linear fit is  $53 \pm 2$  nm.

of molecules without being sensitive to image noise. It starts with the user selecting a few points along the chain to be traced, including start and end points. The algorithm then automatically detects the centreline of the chain. The detection is based on matching a pattern template to the chain, in order to find the best width and location for the pattern that describes the intensity profile of the chain. Then a series of fine-tuning steps are performed to stabilize the results of the location of the backbone of the chain.

After the centreline of each molecule is detected, statistical parameters that describe mechanical properties of the chain such as mean square end-to-end distance,  $\langle R^2 \rangle$ , and mean tangent-tangent correlation of the angles,  $\langle \cos \theta \rangle$ , are calculated from which the persistence length,  $l_p$ , of the molecules is extracted. The WLC model is used to extract persistence length from these curves.

AFM images of DNA molecules were traced and analyzed using the SmarTrace program to validate the technique. The persistence length of DNA was independently calculated using  $\langle R^2 \rangle$ ,  $\langle \cos \theta \rangle$  and probabilities of the bending energy data. These analyses all give persistence lengths of about 50 nm, which is expected for DNA [46, 62].

## Chapter 3

# Atomic Force Microscopy: Imaging Single Collagen Molecules

In this chapter, atomic force microscopy imaging is used to study the conformation and flexibility of collagen molecules in solutions with different pH and ionic strength, as well as different types of collagen. Tracking and analysis of the images are performed using the SmarTrace analysis program described in Section 2.2, and the analysis follows that outlined in Section 2.3. In the first part of this chapter, the conformation and flexibility of different types of collagen under the same ionic strength and pH are directly compared. The different types of collagen samples investigated are listed in Table 3.1. In the second half of this chapter, effects of ionic strength and pH of solution on collagen flexibility are studied. Collagen type I from rat-tail tendon is used in this part. From analysis of the images of the molecules, the persistence length of collagen in each condition is calculated.

### 3.1 Sample Preparation for AFM Imaging of Collagen Molecules

In order to image molecules with AFM, the samples need to be immobilized on a flat smooth substrate such as mica. Mica is widely used for imaging biological molecules. A freshly cleaved mica is negatively charged. Thus, the molecules stick to the surface mainly due to electrostatic attractions between opposing charges on the molecule and the surface [70]. In case of DNA, its backbone is negatively charged and either a counter ion such as  $\text{Ni}^{2+}$  or  $\text{Mg}^{2+}$  is needed in the DNA solution or the mica surface needs to be treated to become positively charged [71]. Collagen (isoelectric point 7.7 [28]) on the other hand is positively charged in the neutral and acidic pH solutions used in this study; therefore, it adheres to the mica without the need for counter ions or mica surface manipulation.

Collagen stock solutions are stored at a concentration of between 2 and 5 mg/ml in 20 mM acetic acid. This concentration is too high for AFM imaging, and control is also desired over the solvent. Thus, before deposition, this solvent is exchanged with the desired solvent using Millipore

Amicon Ultra-0.5 spin filters (NMWL 50 kDa, UFC505096). This solution is then diluted down to 1  $\mu\text{g}/\text{ml}$  prior to AFM imaging. 50  $\mu\text{l}$  of the desired collagen solution is deposited onto freshly cleaved mica (Highest Grade V1 AFM Mica Discs, 10mm, Ted Pella) for 20 seconds. The deposition time and concentration of collagen sample should be such that the surface-adhered molecules do not overlap. After deposition, the excess unbound molecules are rinsed away with ultra pure water, and the mica is dried with a weak flow of filtered nitrogen gas. It is important to note that all collagen molecules were imaged in these dry conditions. Thus, solution conditions quoted refer to the condition under which collagen was deposited onto mica.

The AFM used in this work is the Asylum Research MFP-3D SPM (courtesy Patricia Mooney), working in tapping mode in air. The probes used are Mikromasch HQ:NSC15/AL BS with nominal spring constant of 40 nN/nm and nominal resonance frequency of 325 kHz. The probes have a typical tip radius of 8 nm and 12–18  $\mu\text{m}$  tip height.

## 3.2 AFM Imaging of Different Types of Collagen

In order to characterize differences between different types of collagen, in this section, different types of molecular collagen are imaged using AFM. The sample preparation follows the method introduced in Section 3.1. The types of collagen used in this study are pepsin-digested type I collagen from rat tail tendon (Cultrex Invitrogen), type I recombinant human collagen<sup>1</sup> (Fibrogen), type II collagen from human cartilage (EMD Millipore Corporation) and type III recombinant human collagen<sup>2</sup> (Fibrogen) (Table 3.1).

### 3.2.1 Results

Examples of AFM images of different types of collagen deposited onto mica from a 1 mM HCl and 100 mM KCl solution are shown in Figure 3.1. The images of molecules in this solvent condition show that the chains have fairly extended conformations, and their contour lengths appear to be close to the expected 300 nm. Analysis of the flexibility of different types of collagen in this solution of pH 3 is discussed in the following. Images of human collagens type I, II and III in this solvent condition were recorded and traced by undergraduate Aaron Lyons, who has graciously provided them for this study.

Following the analysis procedure described in the previous chapter, mean-square end-to-end distance,  $\langle R^2 \rangle$ , and the mean cosine of the angular difference  $\langle \cos \theta \rangle$  of tangent vectors at different contour separations are calculated. Figure 3.2 shows plots of these data for type II collagen from human cartilage with a WLC model fit (Equations 2.5 and 2.6). For these analyses, 250 nm was used as the maximum segment length for chain segmentation. The WLC model is fit to a range of segment lengths from 10 nm to a maximum,  $s_{\text{max}}$ , which is chosen to be 200 nm for this solvent

---

<sup>1</sup>expressed in yeast

<sup>2</sup>expressed in yeast

Genetic Source	Collagen Type	Biological Source	Trimeric Form
Rat	I	Rat tail tendon	Hetero
Human	I	Yeast	Hetero
Human	II	Human cartilage	Homo
Human	III	Yeast	Homo

Table 3.1: Different types of collagen molecules studied in Section 3.2.

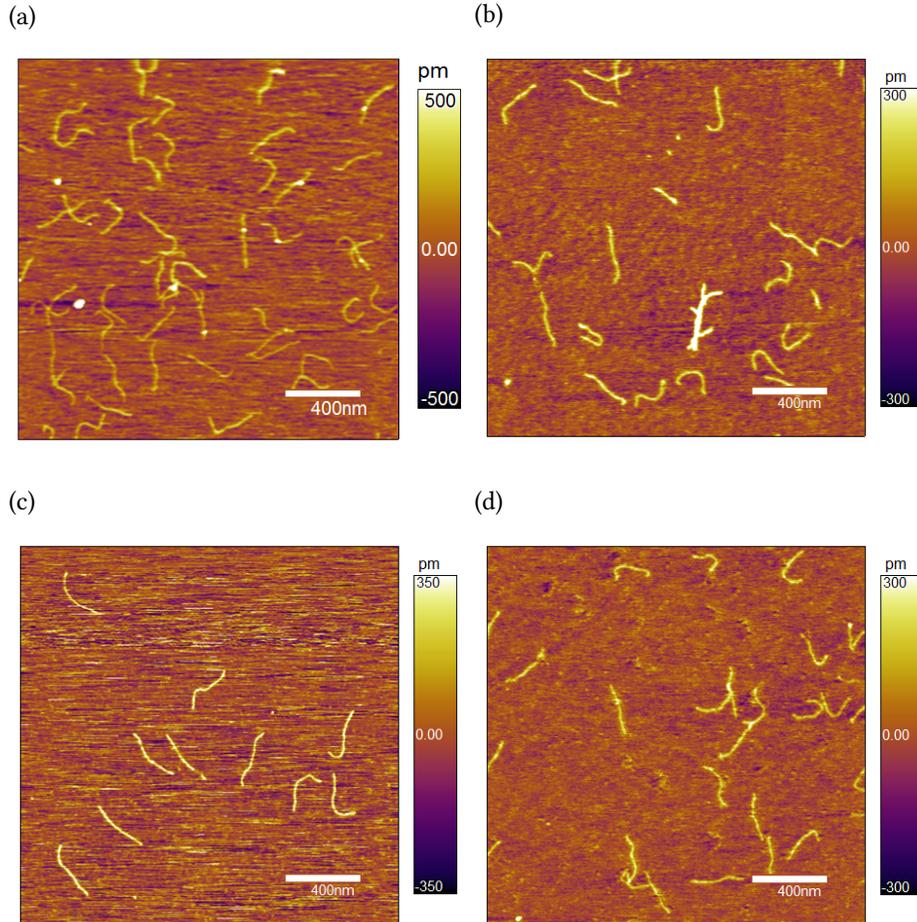


Figure 3.1: Representative AFM images of collagen molecules deposited from 1 mM HCl and 100 mM KCl solution. (a) Type I collagen from rat tail tendon, (b) recombinant human collagen type I, (c) human collagen type II from cartilage, (d) recombinant human collagen type III. The scale bar on the right of the each image indicates relative height.

condition. The results of these analyses give  $l_p = 89 \pm 5$  nm and  $l_p = 96 \pm 5$  nm for  $\langle R^2 \rangle$  and  $\langle \cos \theta \rangle$  respectively. Figures 3.3, 3.4 and 3.5 show the experimental data and the corresponding WLC fits for  $\langle R^2 \rangle$  and  $\langle \cos \theta \rangle$  for rat tail type I, recombinant type I and recombinant type III collagen respectively.

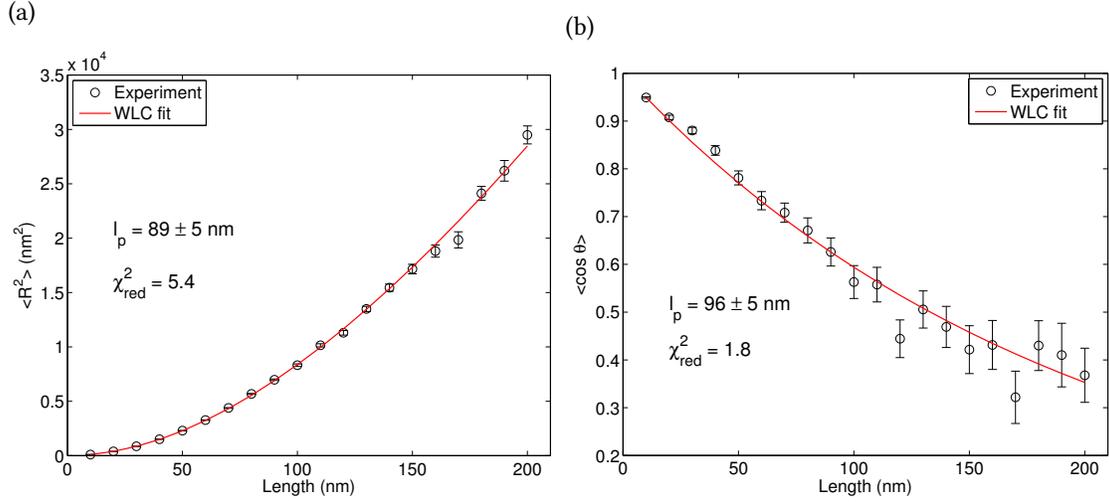


Figure 3.2: Results of analysis of 49 chains of human collagen type II from cartilage in 1 mM HCl and 100 mM KCl. Experimental data and WLC fits for (a)  $\langle R^2 \rangle$  and (b)  $\langle \cos \theta \rangle$  as a function of different contour separations. The error bars on the data points represent standard errors of the means. The resulting persistence lengths for  $\langle R^2 \rangle$  and  $\langle \cos \theta \rangle$  are respectively  $89 \pm 5$  nm ( $\chi^2_{red} = 5$ ) and  $96 \pm 5$  nm ( $\chi^2_{red} = 2$ ). Errors on persistence lengths represent 95% confidence intervals.

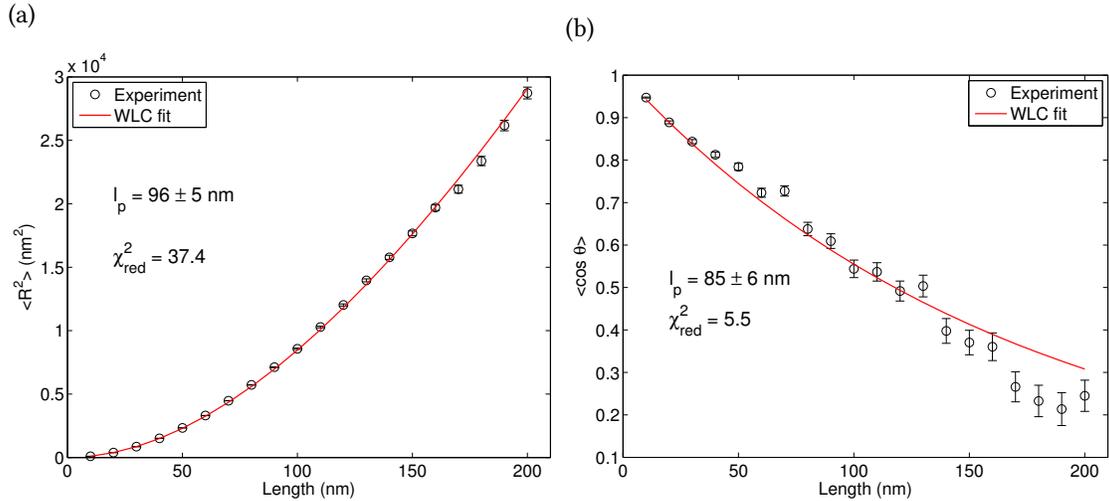


Figure 3.3: Results of analysis of 125 chains of collagen type I from rat tail tendon in 1 mM HCl and 100 mM KCl. Experimental data and WLC fits for (a)  $\langle R^2 \rangle$  and (b)  $\langle \cos \theta \rangle$  as a function of different contour separations. The red line represents the WLC model fit to the experimental data from (a) Equation 2.5 with resulting  $l_p = 96 \pm 5$  nm ( $\chi^2_{red} = 37$ ), and (b) Equation 2.6 with resulting  $l_p = 85 \pm 6$  nm ( $\chi^2_{red} = 15$ ). The error bars in the figures represent standard errors of the means. The error bars in (a) are smaller than the figure symbols. Errors on persistence lengths represent 95% confidence intervals.

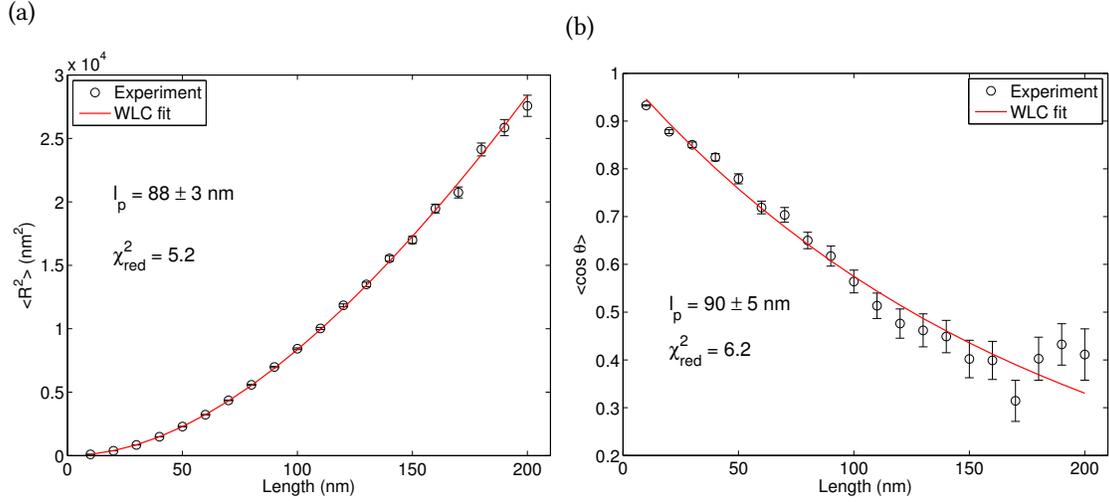


Figure 3.4: Results of analysis of 95 chains of recombinant human collagen type I in 1 mM HCl and 100 mM KCl. Experimental data and WLC fits for (a)  $\langle R^2 \rangle$  and (b)  $\langle \cos \theta \rangle$  as a function of different contour separations. The error bars on the data points represent standard errors of the means. The resulting persistence lengths for  $\langle R^2 \rangle$  and  $\langle \cos \theta \rangle$  are respectively  $90 \pm 5$  nm ( $\chi^2_{red} = 6.2$ ) and  $88 \pm 3$  nm ( $\chi^2_{red} = 5.2$ ). The errors on  $l_p$  represent 95% confidence intervals of the fits.

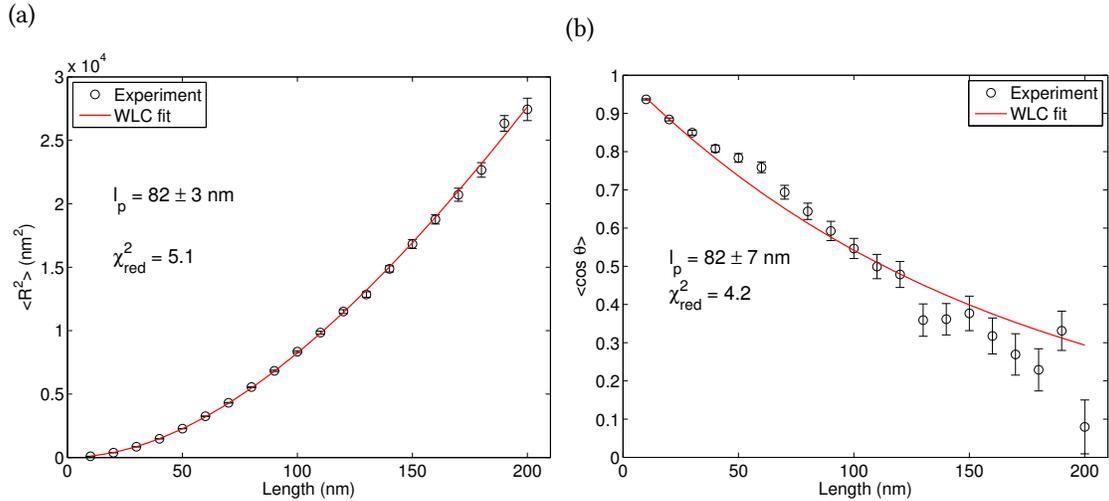


Figure 3.5: Results of analysis of 78 chains of recombinant human collagen type III. Experimental data and WLC fits for (a)  $\langle R^2 \rangle$  and (b)  $\langle \cos \theta \rangle$  as a function of different contour separations. The error bars on the data points represent standard errors of the means. The resulting persistence lengths for  $\langle R^2 \rangle$  and  $\langle \cos \theta \rangle$  are respectively  $82 \pm 3$  nm ( $\chi^2_{red} = 5.1$ ) and  $82 \pm 7$  nm ( $\chi^2_{red} = 4.2$ ). Errors on persistence lengths represent 95% confidence intervals.

Figure 3.6 shows analysis of the bending angle distributions for collagen type II from human cartilage. The resulting  $l_p = 92$  nm is very close to the values obtained from  $\langle R^2 \rangle$  and  $\langle \cos \theta \rangle$  data analysis.

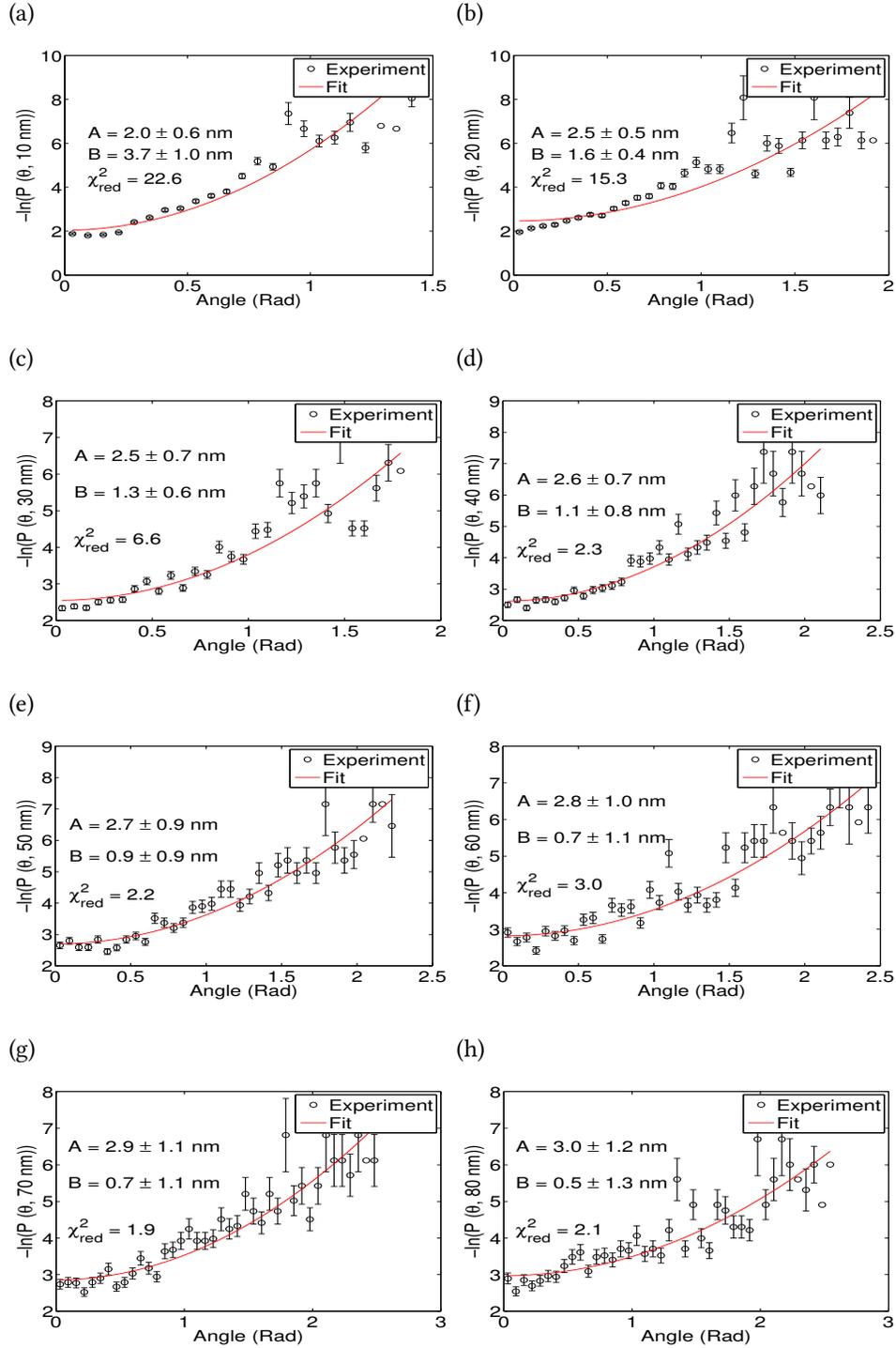


Figure 3.6: (a-h) Determination of bending energy at different contour separations (Equation 2.8) ranging from  $s = 10$ – $80$  nm, for collagen type II from human cartilage in 1 mM HCl and 100 mM KCl. The resulting  $B$  parameters (Equation 2.9) were used to extract persistence length Figure 3.7). A total of 49 molecules were analyzed.

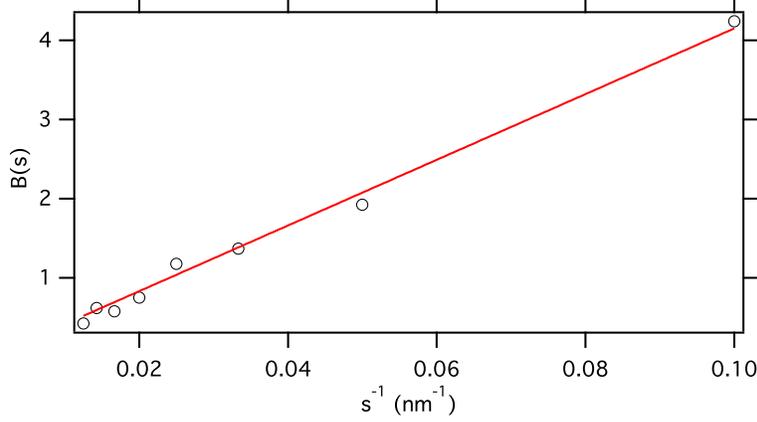


Figure 3.7: Extracting persistence length of human collagen type II from the angular distribution fit parameter  $B$  (Equation 2.9). A linear fit to  $B$  values obtained from Figure 3.6 results in a persistence length of  $l_p = 92 \pm 3$  nm.

Gaussian behaviour of the molecules requires additionally that the probability of finding curvature  $\kappa$  over a segment length  $ds$ , which is governed by bending modulus, is normally distributed [69]:

$$P(\kappa; ds) \propto \exp\left(-\kappa^2 \frac{ds l_p}{2}\right). \quad (3.1)$$

A plot of the logarithm of the probabilities as a function of curvature will give a linear slope if Gaussian chains exist. To assess whether this is true for collagen, the curvature was determined from the traced chains. Curvature is the magnitude of the rate of change of the tangent vector over arc length  $ds$ :

$$\kappa = \left| \frac{d\hat{t}}{ds} \right|, \quad (3.2)$$

where  $\hat{t}$  is the unit tangent vector. For calculation of curvature, tangent vectors were defined by connecting points separated by  $ds = 5$  nm along collagen's contour. The curvature at a separation of  $\Delta s = 10$  nm was calculated using tangent vectors whose ends were separated by 10 nm (Figure 3.8). Thus:

$$\kappa = \left| \frac{\vec{T}' - \vec{T}}{\Delta s} \right|. \quad (3.3)$$

Probability plots on a logarithmic scale as a function of curvature squared for collagen deposited from 1 mM HCl + 100 mM KCl show linear behaviour (Figure 3.9), which indicates that collagen behaves as a Gaussian chain in this solution condition. In addition, kurtosis, the fourth moment of the distribution, was calculated. Kurtosis measures peakedness or flatness of a distribution; for a normal distribution its value is 3. Values for kurtosis of all collagen types in these solvent conditions are close to 3 (Figure 3.9).

In fitting  $\langle R^2 \rangle$  and  $\langle \cos \theta \rangle$ , the data in each segment length bin are weighted by their respective standard error of the mean (*i.e.*,  $\sigma/\sqrt{N}$ , where sigma is the standard deviation of values and  $N$

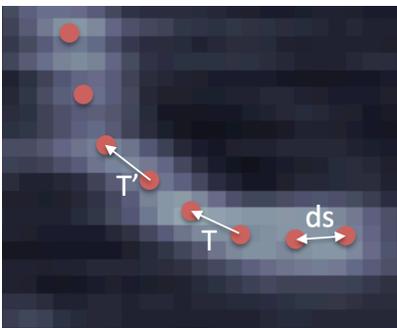


Figure 3.8: Schematic of calculation of curvature of a chain. The red points represent points along the backbone of collagen spaced by  $ds = 5$  nm. The tangent vectors ( $\vec{T}'$  and  $\vec{T}$ ) connect successive points. The curvature is calculated over an arc length  $\Delta s = 10$  nm.

is the number of data points in each bin). Therefore, smaller segment lengths contribute more strongly to the fit results, because of the smaller standard deviation and the larger number of data points in each bin. The weighting of the data allows extracting reduced chi squared  $\chi_{red}^2$  to quantitatively assess goodness of fits. Although this has been done, the  $R^2$  and  $\cos \theta$  values in each segment length bin are not normally distributed. Thus, this statistical error analysis should not be interpreted too strongly. Another systematic error can arise because it is seen that the fit results can be affected by the maximum chain segment length  $s_{max}$  used in the fitting. To present results of the analysis, two different approaches are used: 1) Each presented persistence length value is the average of values obtained over a range of different fit  $s_{max}$ . Error bars represent the standard deviation of  $l_p$  obtained at different values of  $s_{max}$ , separated by 10 nm increments and ranging from 150 nm to 250 nm. The error bars in this case represent the minimum and maximum values of  $l_p$  obtained over the range of fit  $s_{max}$ . 2) Each presented  $l_p$  value is the persistence length when  $s_{max}$  is 200 nm. Error bars represent uncertainty of the fit (95% confidence interval) at  $s_{max} = 200$  nm. The results from each approach are shown side by side for  $\langle R^2 \rangle$  of different collagen types (Figure 3.10). For this solution condition, there is a negligible variation in the  $l_p$  values over different  $s_{max}$  and the average value of  $l_p$  is close to the value at  $s_{max} = 200$  nm. Unless otherwise stated, the reported persistence lengths in the rest of this chapter are based on approach 1: the  $l_p$  value is the average of values obtained over a range of different fit  $s_{max}$ , and error bars represent the maximum and minimum values of  $l_p$  over this range.

Figure 3.11 compares the persistence lengths obtained for different types of collagen in 1 mM HCl and 100 mM KCl from fitting the WLC model to  $\langle R^2 \rangle$  or  $\langle \cos \theta \rangle$  data. The results from the two analyses are in agreement to within the uncertainties. For easy comparison measures, reported  $l_p$  values are from  $\langle R^2 \rangle$  analysis in the rest of this chapter, unless otherwise mentioned.

Persistence lengths resulting from WLC fits for different types of collagen deposited from a 1 mM HCl and 100 mM KCl solution are summarized in Table 3.2. Persistence lengths of different types of collagen are similar in this solvent. The results indicate the unexpected finding that these different collagen types from different genetic and biological sources have comparable overall

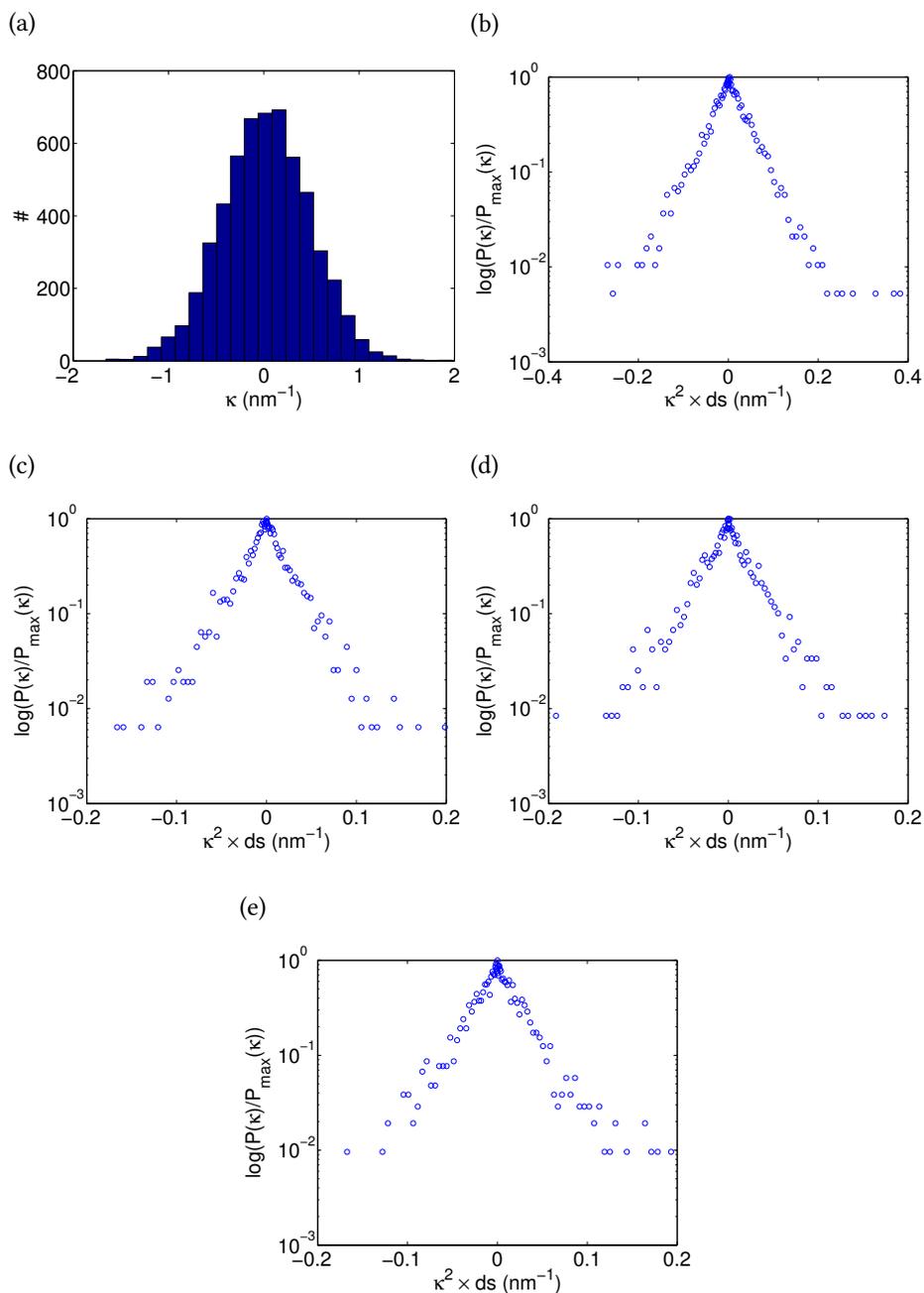


Figure 3.9: Curvature plots at 10 nm contour separation of collagen deposited from 1 mM HCl + 100 mM KCl solution. (a) Histogram of curvature values for type I collagen from rat tail. Probability plots on a logarithmic scale as a function of curvature squared for (b) type I collagen from rat tail tendon (kurtosis = 3.13), (c) recombinant human type I collagen (kurtosis = 3.17), (d) type II collagen from human cartilage (kurtosis = 3.35), and (e) recombinant human type III collagen (kurtosis = 3.15). Plotted this way, these data show linear behaviour.

flexibility within uncertainty of the analysis. The persistence lengths of different types of collagen in this solvent condition fall in the mid-range of values reported in the literature.

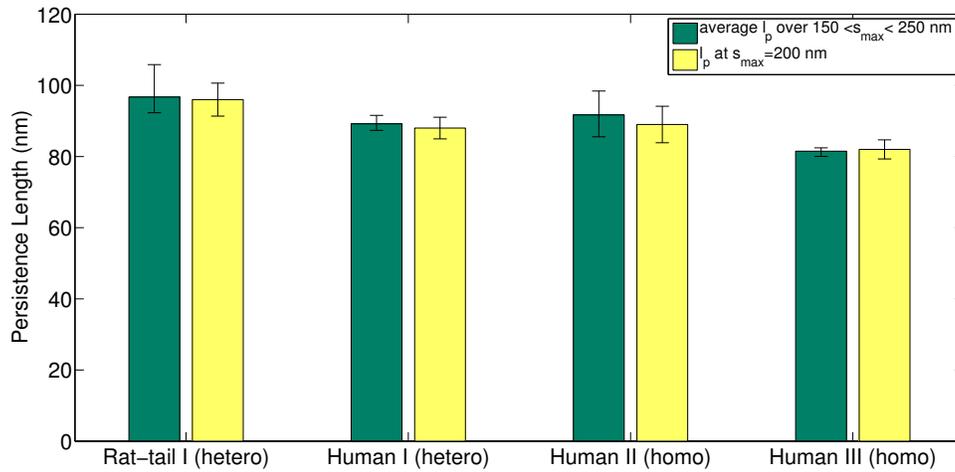


Figure 3.10: Persistence length values for collagen in 1 mM HCl and 100 mM KCl obtained from a WLC fit for  $\langle R^2 \rangle$  data when determined  $l_p$  values are averaged over a range of  $s_{max}$  between 150 nm and 250 nm (green) or are determined at 200 nm  $s_{max}$  (yellow). The error bars on the green bars represent the lowest and highest values over this range. The error bars on the yellow bars represent the 95% confidence interval of the fit at  $s_{max} = 200$  nm.

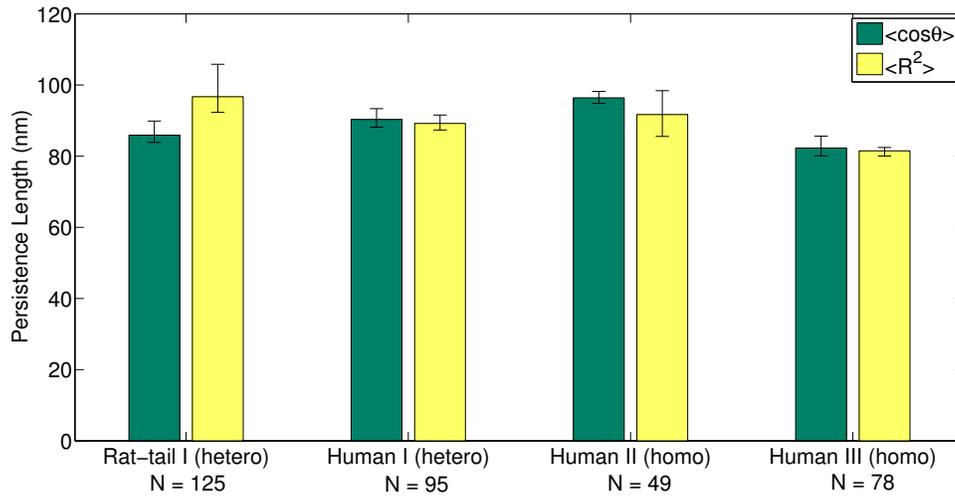


Figure 3.11: Persistence length values for collagen in 1 mM HCl and 100 mM KCl are obtained from WLC fit for  $\langle \cos \theta \rangle$  (green) and  $\langle R^2 \rangle$  (yellow) data when  $l_p$  values are averaged over a range of  $s_{max}$  between 150 nm and 250 nm. The error bars represent the lowest and highest values of  $l_p$  over the range of  $150 < s_{max} < 250$  nm.  $N$  is the total number of molecules used in the analysis.

AFM images of different collagen types deposited from 20 mM acetic acid solution were also studied (Figures 3.12 and 3.13). These molecules exhibited significantly increased flexibility compared to collagen in 1 mM HCl and 100 mM KCl. In acetic acid, however, the experimental data deviate from WLC behaviour.  $\langle R^2 \rangle$  and  $\langle \cos \theta \rangle$  data are not completely described by the WLC

Table 3.2: Persistence lengths from  $\langle R^2 \rangle$ ,  $\langle \cos \theta \rangle$  and bending energy analysis for different types of collagen. Reported values represent the average  $l_p$  over the range of  $150 < s_{\max} < 250$  nm. The reported values beneath for  $\langle R^2 \rangle$  and  $\langle \cos \theta \rangle$  analyses represent the minimum and maximum values found over this  $s_{\max}$  fitting range. The quoted error for the bending energy analysis (Equation 2.9) represents the 95% confidence interval.

Type of Collagen	$l_p$ (nm) $\langle R^2 \rangle$	$l_p$ (nm), $\langle \cos \theta \rangle$	$l_p$ (nm), B(s)*
Rat tail I	97 (92, 106)	86 (84, 90)	$73 \pm 6$
Recombinant human I	89 (87, 92)	90 (88, 93)	$64 \pm 9$
Human cartilage II	92 (86, 98)	96 (95, 98)	$89 \pm 5$
Recombinant human III	82 (80, 82)	82 (80, 86)	$71 \pm 7$

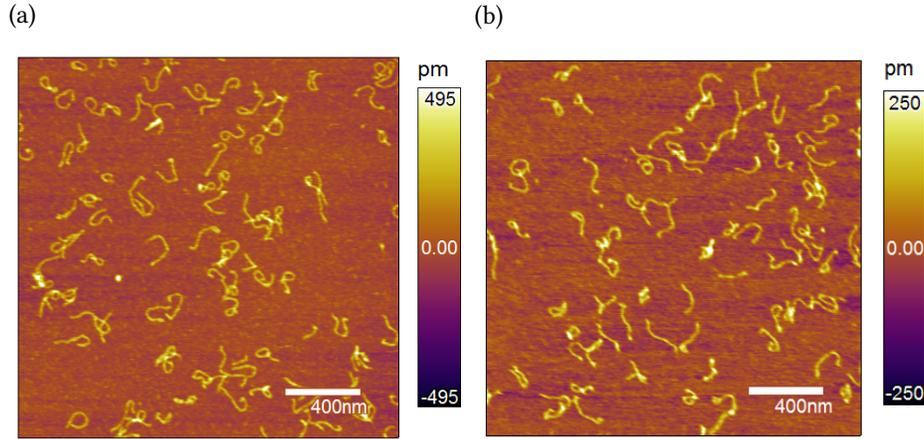


Figure 3.12: Examples of AFM images of collagen molecules deposited from 20 mM acetic acid solution (pH 3.2). (a) Recombinant human collagen type I, (b) recombinant human collagen type III. The scale bar on the right of the each image indicates relative height.

model. The deviation between the model and the  $\langle R^2 \rangle$  data increases at larger separations, with data at smaller segment lengths having a better agreement with the model (Figure 3.14). The values of  $l_p$  obtained from the WLC fit depend on the value of  $s_{\max}$  and decrease (up to 40%) as  $s_{\max}$  increases from 100 to 200 nm (Figure 3.15).

Although the WLC model does not appear to describe my observations in all solvent conditions, it is used to obtain an apparent persistence length as a comparative measure to assess flexibility of different types of collagen. Based on these results, different types of collagen, from different sources, exhibit similar persistence lengths.

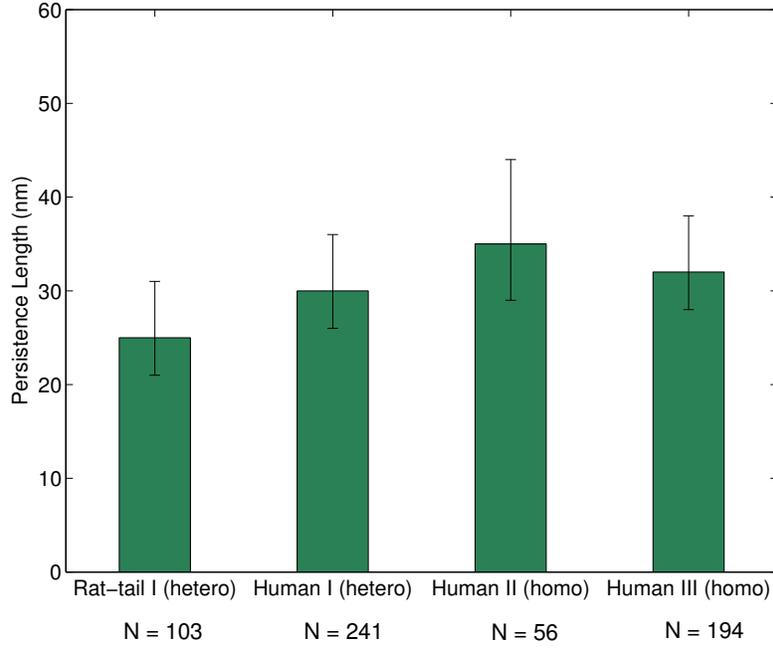


Figure 3.13: Persistence lengths for different types of collagen deposited from 20 mM acetic acid solution (pH 3.2). Quoted values are the average  $l_p$  over the range of  $100 < s_{\max} < 200$  nm. Error bars represent the maximum and minimum  $l_p$  obtained from fitting over this range.  $N$  represents the total number of molecules used in the analysis.

### 3.2.2 Non-WLC behaviour of chains

As mentioned, the behaviour of collagen in some solutions shows a deviation from the WLC model (Figure 3.14). Apparent sinusoidal modulation of the exponentially decaying  $\langle \cos \theta \rangle$  data observed in some solution conditions (Figure 3.14b) led me to explore briefly other models. For polymers with inherent or induced curvature, when there is no twist or torsion, in three dimensions the tangent-tangent correlation data can be described by the following equation [72]:

$$\langle \cos \theta \rangle = \exp\left(-\frac{s}{l_p^*}\right) \cos\left(\frac{s}{R_0}\right). \quad (3.4)$$

Here  $R_0$  is the inherent / induced radius of curvature and  $l_p^*$  is the effective persistence length. The effective persistence length here is defined as "the average chain length that the chain curls up with a radius  $R_0$  before it forgets to curve and wanders off";  $l_p^*$  is not directly comparable with  $l_p$  in the standard WLC model [73]. For Gaussian chains, *i.e.* zero intrinsic curvature or equivalently very large  $R_0$ , Equation 3.4 converges to the standard WLC model in 3D.

Several assumptions, such as torsion and twist in the chain, need to be taken into account to convert this model into a 2D model which leads to a form similar to Equation 3.5 [72]. For simplicity it is assumed these conditions are satisfied, resulting in the 2D equivalent of Equation

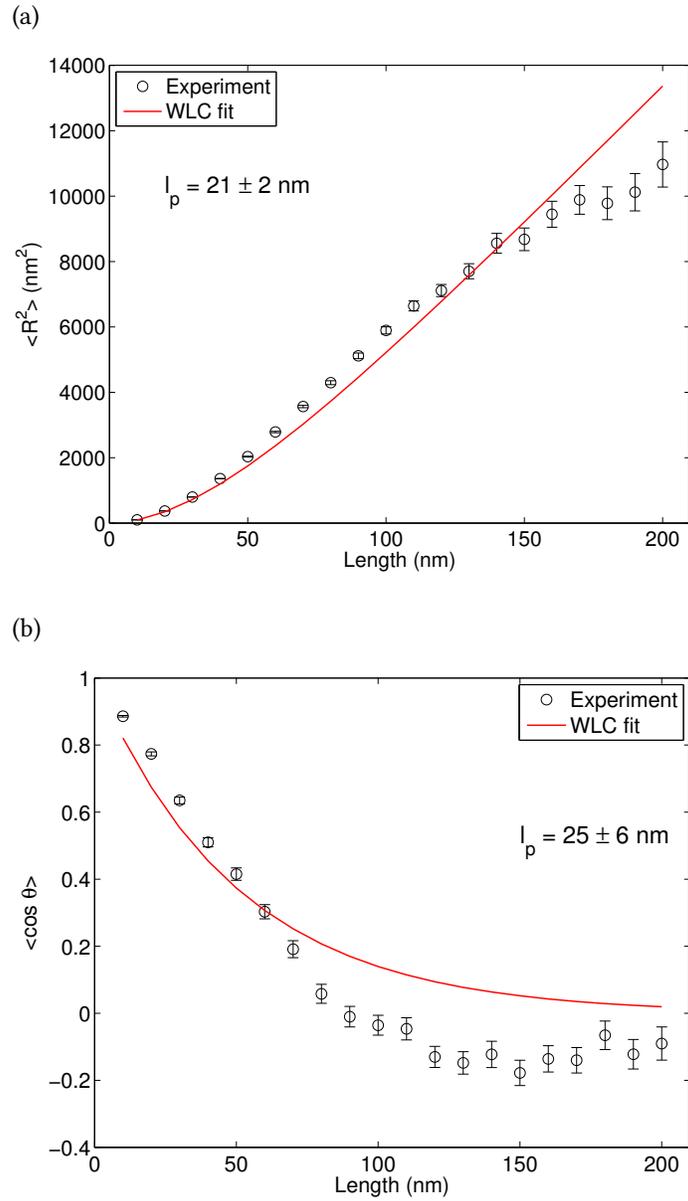


Figure 3.14: (a)  $\langle R^2 \rangle$  and (b)  $\langle \cos \theta \rangle$  data of type I rat tail collagen deposited from 20 mM acetic acid (pH 3.2) exhibit non-WLC behaviour. In this solution condition, the  $\langle \cos \theta \rangle$  tangent correlation data does not decay exponentially as expected for a WLC chain, but rather appears to exhibit periodic behaviour.

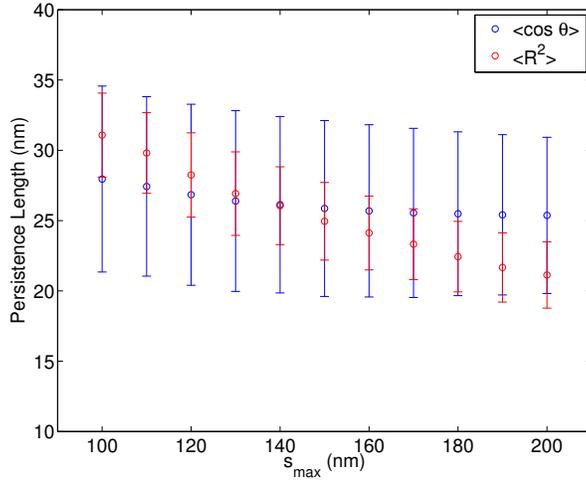


Figure 3.15: Persistence lengths resulting from a WLC fit to  $\langle R^2 \rangle$  and  $\langle \cos \theta \rangle$  data as a function of  $s_{\max}$  for rat tail collagen type I deposited on mica from 20 mM acetic acid. The values obtained from  $\langle R^2 \rangle$  analysis depend strongly on  $s_{\max}$ , while for  $\langle \cos \theta \rangle$  there is not a strong dependence.

3.5 of

$$\langle \cos \theta \rangle = \exp\left(-\frac{s}{2l_p^*}\right) \cos\left(\frac{s}{R_0}\right). \quad (3.5)$$

This model accounts for oscillatory decay of the correlation function, and gives better agreement with experimental  $\langle \cos \theta \rangle$  data that deviate from the standard WLC model (Figure 3.16). The  $\chi_{red}^2$  value is dramatically improved with the curved model. The resulting  $l_p^*$  value of 46 nm does not vary with  $s_{\max}$  (100 to 250 nm). By contrast, the standard WLC model fit to the  $\langle \cos \theta \rangle$  data results in a value for  $l_p$  that varies between 21 nm and 32 nm for  $s_{\max}$  ranging from 100 to 200 nm.

The standard WLC assumes  $l_p$  arises from thermal bending, while the curved model takes into account the inherent / induced curvature, in addition to the thermal bending. This model does not distinguish whether the curvature is intrinsic or surface-induced. Future investigations can shed light on the origin of curvature by imaging collagen molecules on different surfaces such as graphite or surface-treated mica.

Length-dependent non-WLC behaviour was previously observed for DNA molecules [63, 68]. Possible causes of the observed deviation of collagen from WLC behaviour are intrinsic curvature of molecules, surface-induced curvature or other surface-molecule interactions. Surface-induced curvature was previously observed for chiral amyloid fibrils deposited on mica [69]. Another study found surface-induced deformations of DNA as a result of interactions with the charged surface of mica [63]. Future experiments could address the origins and strengths of surface-molecule interactions by comparing conformations of collagen on different surfaces such as graphite and surface-treated-mica [59]. My preliminary experiments with mica surface treatments did not reveal qualitative differences, but a more extensive study is required. In addition, the possible intrinsic curvature of collagen could be studied by developing a model of twist-bend coupling in

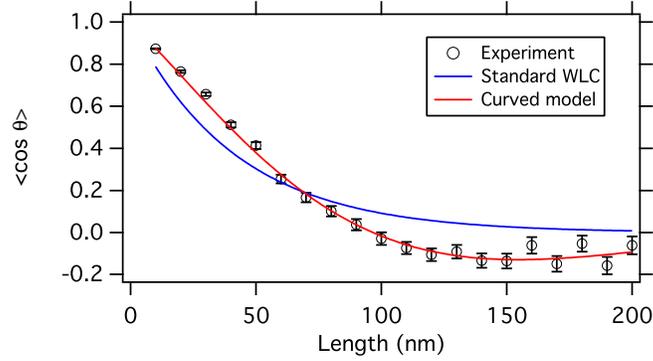


Figure 3.16:  $\langle \cos \theta \rangle$  data of of type I collagen from rat tail in 20 mM acetic acid are better described by the curved WLC model (equation 3.5; red) than by the standard WLC model (equation 2.6; blue). The standard WLC gives a persistence length of  $l_p = 32$  nm ( $\chi_{red}^2 = 47$ ), while the curved model gives an effective persistence length  $l_p^* = 46$  nm and radius of curvature of  $R_0 = 60$  nm ( $\chi_{red}^2 = 1.7$ ). Using the curved model dramatically improves  $\chi_{red}^2$ . The plotted error bars represent standard errors of the means.

the helix for collagen. It is likely that collagen exhibits a sequence-dependent flexibility, due to the heterogeneity of its sequence. For all of these reasons, one end of collagen could be labeled in order to assess the directionality of the polymer chain. Then questions of curvature and its sequence-dependence could be investigated.

By comparing persistence length values from the standard WLC and the curved model, we see higher values for the curved model. However, the effective persistence length cannot be used to extract persistence length value and to directly compare the chain flexibility from the WLC model. This model is used in the next section to extract the radius of curvature for collagen data that do not behave as a WLC.

The results of fitting the curved model (Equation 3.5) to  $\langle \cos \theta \rangle$  of different types of collagen in 20 mM acetic acid (Figure 3.17) are presented in Figure 3.18 and Table 3.3. Interestingly, values of  $R_0$  in this solvent condition are close to the spacing of periodic banding patterns in collagen fibrils [3]; however this is not the case in all solvent conditions (e.g. Figure 3.22) so is likely a coincidence. Interestingly, when analysed with the curved model, tissue-derived type II collagen appears to possess a significantly higher effective persistence length than the other collagens studied in this solvent condition. This behaviour contrasts that of found from analysis using the standard WLC model and is more in line with the expected stiffening of collagen resulting from crosslinking that occurs in tissues.

### 3.2.3 Summary and Discussion

Different types of collagen are found in different parts of the body in tissues with different mechanical properties. AFM imaging of different types of collagen at the molecular level can give insights into mechanical properties of the molecules and provides a direct comparison of flexibilities of the molecules.

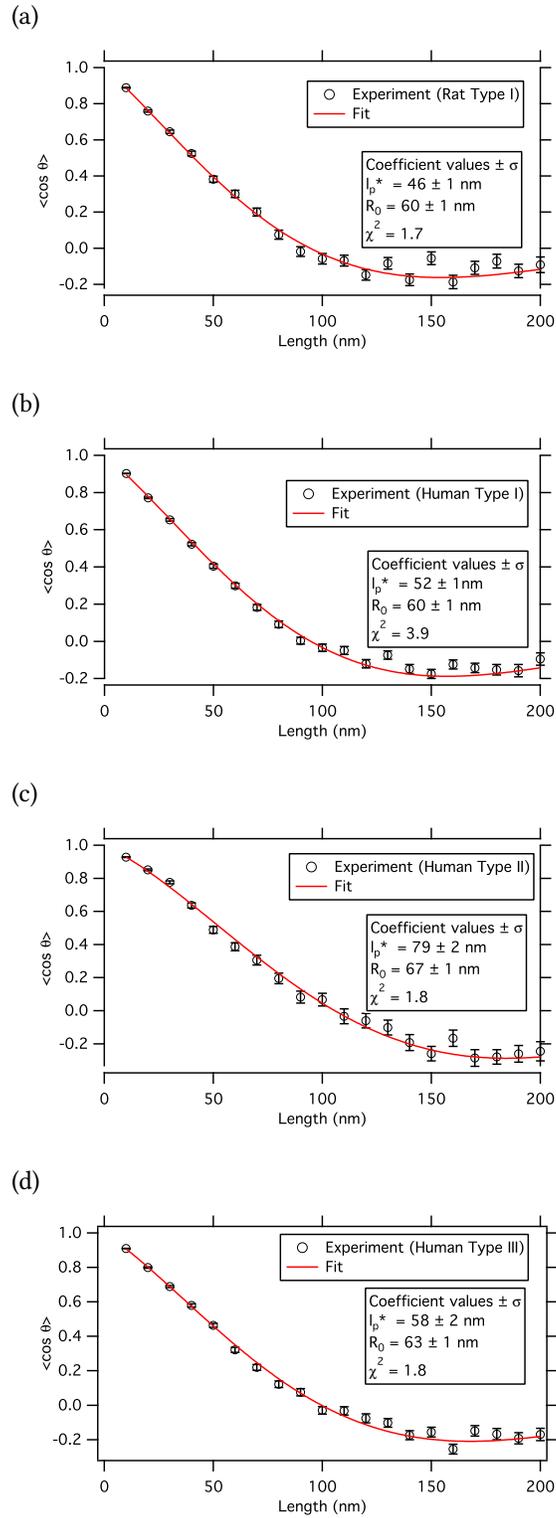


Figure 3.17: Results of fitting the curved model to  $\langle \cos \theta \rangle$  data for different types of collagen deposited from 20 mM acetic acid.

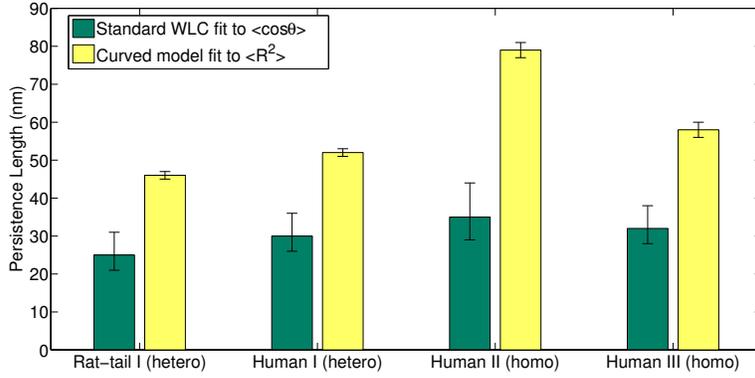


Figure 3.18:  $l_p$  and  $l_p^*$  results for different types of collagen deposited from 20 mM acetic acid solution (pH 3.2) obtained from the standard WLC fit to  $\langle R^2 \rangle$  (green), and the curved model (Equation 3.5) fit to  $\langle \cos \theta \rangle$  (yellow).

Table 3.3: Persistence lengths obtained for different collagen types deposited from 20 mM acetic acid. Shown are results from the standard WLC fit to  $\langle R^2 \rangle$  data and from the curved model fit to  $\langle \cos \theta \rangle$ . The last column represents the radius of curvature derived from the curved model. Errors represent 95% confidence intervals.

Type of Collagen	$l_p$ (nm), Standard WLC fit, $\langle R^2 \rangle$	$l_p$ (nm), Curved model fit, $\langle \cos \theta \rangle$	$R_0$ (nm), Curved model fit, $\langle \cos \theta \rangle$
Rat tail I	27 ± 2	46 ± 1	60 ± 1
Recombinant human I	32 ± 1	52 ± 1	60 ± 1
Human cartilage II	38 ± 2	79 ± 2	67 ± 1
Recombinant human III	38 ± 1	58 ± 2	63 ± 1

The persistence lengths of different types of collagen obtained in this work indicate that in 1 mM HCl and 100 mM KCl, types I, II and III collagens from different sources all have comparable flexibility, with  $l_p$  values between 80 and 100 nm. The values of persistence length obtained for collagen molecules from this solution fall in the middle of the range of the persistence length values found in the literature (Table 1.2). The apparent increase in flexibility of collagen molecules in the low ionic strength acetic acid solution, 20 mM acetic acid (Figure 3.13) correlates with persistence lengths towards the lower end reported for collagen in the literature [11, 12, 18, 19]. Although the persistence length appears to vary with solution conditions, here too I find similar values for  $l_p$  among different collagens. However, the data deviate from the standard WLC model. Using the curved model (Equation 3.5) to fit  $\langle \cos \theta \rangle$ , I find better agreement with the data. The inherent/induced curvatures found for different collagens using equation 3.5 are similar (Table 3.3). Fits to this model suggest that homotrimeric collagen types II and III exhibit slightly higher persistence lengths than the heterotrimeric collagens. A change in pH or ionic strength can affect the stability of homo and heterotrimers differently [74], and in this case, in the acetic acid solution, collagen type II and III homotrimers show slightly higher persistence lengths.

In previous studies, optical tweezers stretching experiments in a complex ionic solution at neutral pH (Table 1.2) found a similar flexibility for two types of procollagen, I (heterotrimer) and II (homotrimer) [11, 44]. These results agree with our finding of comparable overall flexibility between different types of collagen (in 1 mM HCl and 100 mM KCl).

One key assumption in our analysis is that collagen is a homogeneous polymer. In our analysis all segments of a given length are treated equivalently and independent of their position or orientation in the chain. This approach reveals the overall conformation and flexibility of the molecules but cannot determine sequence-dependent flexibility. Collagen in fact has sequence heterogeneity, which may or may not lead to sequence-dependent flexibility. Type I collagen from human and rat have a high sequence identity ( $\sim 90\%$ ), while the similarity is less when comparing types I, II and III [75].

Earlier studies of proteolytic cleavage of collagen type I molecules indirectly compared mechanical properties of molecules of the conventional heterotrimeric form with pathogenic homotrimers. It was found that molecules of the wildtype heterotrimeric collagen are more susceptible to proteolytic cleavage than the homotrimeric isoforms [39, 40]. On the contrary, a simulation study, based on mechanical response of a small section of collagen triple helix to an applied force, found that the hetero isoform has lower bending flexibility than the homo form [41]. The discrepancy between these results may arise from the fact that collagen has different local stabilities along its length. The enzymatic experimental studies observed an increased local stability of the triple helix and increased resistance to enzymatic cleavage by matrix metalloproteinase (MMP) at the cleavage site for the homotrimer [40]. The simulation work studied a short collagen-like peptide [41], and found that homotrimeric isoform has shorter persistence length and forms kinks. The effects of kinks on persistence length of the chain are likely averaged out when the flexibility of the entire chain is considered.

Collagen molecules from different biological sources undergo different extents of post translational modifications [76–78]. Post-translational modifications such as hydroxylation of proline depend on the type of tissue [79]. Age of the tissue also affects the extent of post-translational modifications [80]. These modifications can change the local stability of the molecules [1, 78, 81]. It has been shown that some post-translational modifications affect the stability of collagen. For example, hydroxylation of proline to hydroxyproline increases the thermal stability of collagen [1, 82].

In spite of these potential differences among the source collagens studied here, the current results show that collagen molecules of different types from different sources have comparable overall flexibility. These results indicate that overall flexibility of collagen is more strongly affected by solution conditions (*e.g.* ionic strength) and that variations in the amino acid compositions and modifications play a much more limited role, if any, in overall flexibility. It is possible that differences would be noted by studying local and sequence-dependent flexibility of collagen. This approach could help to elucidate underlying differences arising due to variations in the type and modifications of collagen. Future studies implementing AFM imaging to study specifically

<b>Solution</b>	<b>KCl Concentration (mM)</b>	<b>pH</b>	<b>Ionic Strength (mM)</b>	<b>N<sup>**</sup></b>
H <sub>2</sub> O	0	~ Neutral	0.0001*	59
KCl	0.1	~ Neutral	0.1	26
KCl	1	~ Neutral	1	50
KCl	10	~ Neutral	10	70
KCl	100	~ Neutral	100	77
1 mM HCl,KCl	0	3	1	86
1 mM HCl,KCl	10	3	11	138
1 mM HCl,KCl	100	3	101	125

Table 3.4: List of solution conditions and ionic strengths from which type I collagen molecules from rat tail were imaged to study the effects of salt and pH on molecular conformations. \* The ionic strength is calculated for pure water at pH 7:  $\text{H}_2\text{O} \rightleftharpoons \text{H}^+ + \text{OH}^-$  where  $[\text{H}^+] = [\text{OH}^-]$ . Distilled water is somewhat acidic due to dissolved  $\text{CO}_2$ , but this is disregarded here..

\*\* Number of molecules in the analysis.

end-labeled collagen molecules would enable examining of sequence-dependent structural heterogeneity and flexibility.

### 3.3 Effect of Ionic Strength and pH on Collagen Mechanics

In this section, the flexibility of collagen molecules is quantified under a range of solvent conditions, to assess how the persistence length of the molecules changes as pH and ionic strength are varied. Type I collagen from rat tail tendon is prepared in solutions of different salt concentration and pH according to the sample preparation method introduced in Section 3.1. The different solution conditions in which collagen is solvated before deposition on mica are summarized in Table 3.4.

#### 3.3.1 Results

Figure 3.19 shows example images of collagen deposited from different solution conditions. It is apparent that collagen molecules adopt more extended conformations at higher ionic strengths. The pH of the solution also affects the collagen flexibility: molecules in solution conditions of similar ionic strength have a more extended conformation at neutral pH compared to an acidic pH.

Figure 3.20 shows plots of the mean-square end-to-end distance,  $\langle R^2 \rangle$ , and mean cosine of angular difference of tangent vectors,  $\langle \cos \theta \rangle$ , as a function of the measured separation along the contour, for collagen type I molecules deposited from 10 mM KCl at neutral pH. The experimental results are described by the standard WLC model introduced in Equations 2.5 and 2.6. The data show reasonable agreement with the this model.

As observed for collagen in 20 mM acetic acid (Section 3.2.2), collagen does not behave as a 2D WLC in all solution conditions. Here, for the purpose of comparing flexibility of collagen, we

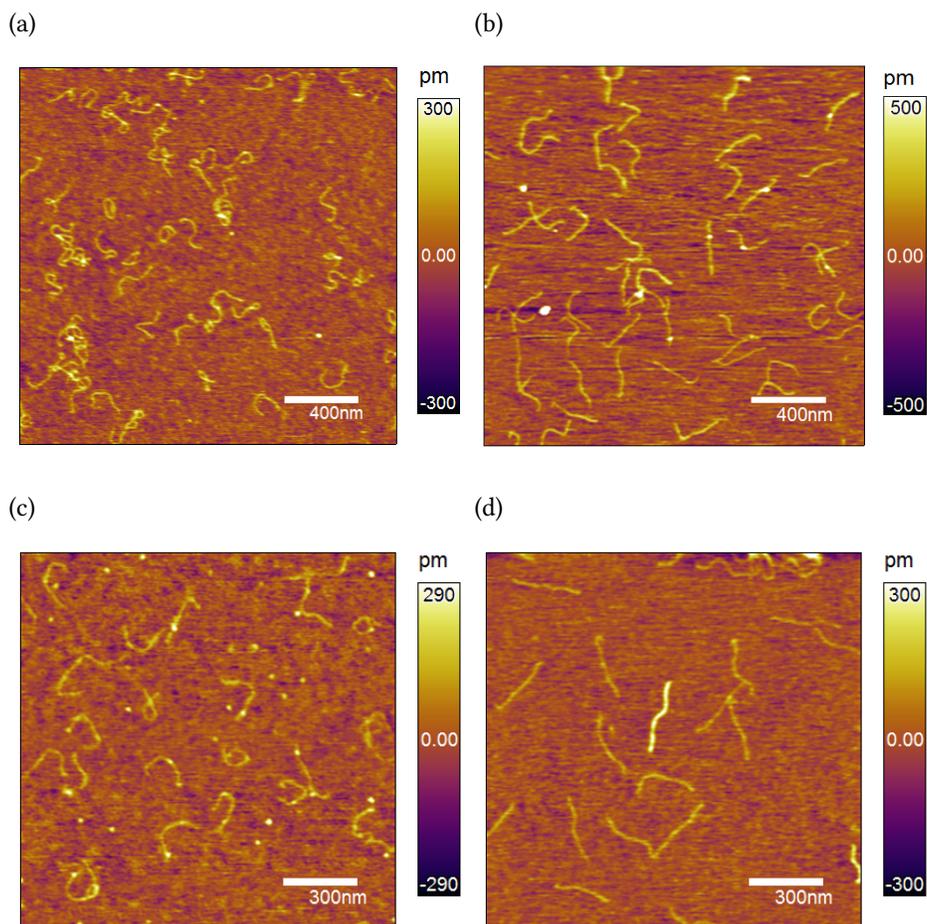


Figure 3.19: Representative AFM images of type I collagen molecules deposited from acidic pH solutions of (a) 1 mM HCl and (b) 1 mM HCl + 100 mM KCl. AFM images of collagen molecules deposited from solutions with similar ionic strength but a neutral pH with (c) 1 mM KCl and (d) 100 mM KCl show less flexible conformations compared to the acidic pH solutions. A higher ionic strength greatly decreases flexibility in both pH conditions. The scale bar on the right of the each image indicates relative height.

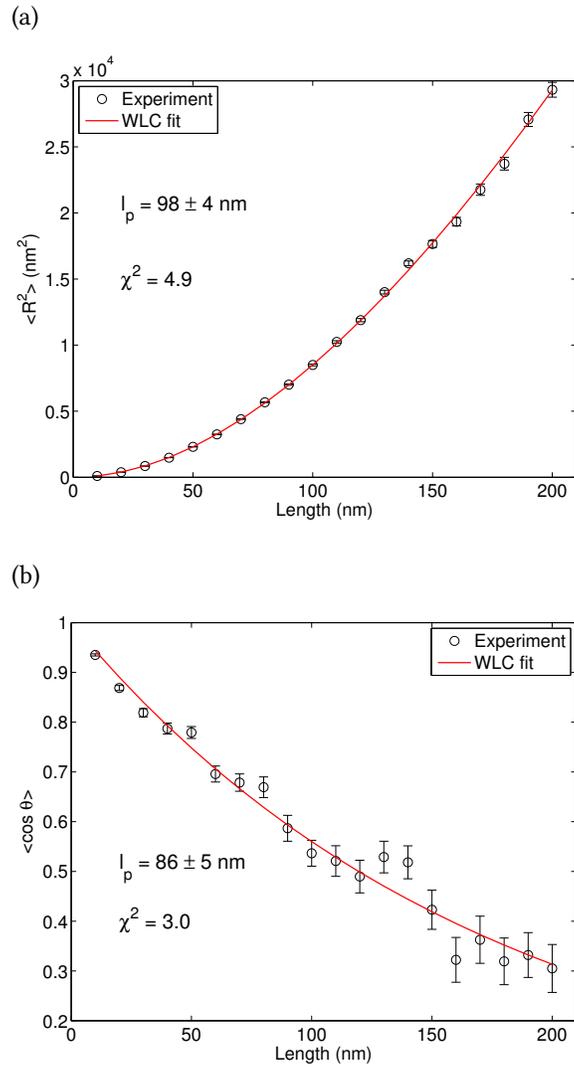


Figure 3.20: Experimental plots and corresponding WLC fits of (a)  $\langle R^2 \rangle$  and (b)  $\langle \cos \theta \rangle$  data for type I collagen deposited from 10 mM KCl at neutral pH. Error bars represent standard errors of the means.  $l_p$  values obtained from these fits with  $s_{\max} = 250 \text{ nm}$  are respectively  $98 \pm 4 \text{ nm}$  ( $\chi_{red}^2 = 4.9$ ) and  $86 \pm 5 \text{ nm}$  ( $\chi_{red}^2 = 3$ ). The errors on  $l_p$  represent 95% confidence intervals of the fits.

use the output persistence length from  $\langle R^2 \rangle$  analysis, as these data show better agreement with the WLC model. We use this only in a comparative sense, to investigate the contributions of ionic strength and pH to the apparent flexibility of collagen.

Figure 3.21 shows how the apparent persistence length of collagen varies with ionic strength at acidic and neutral pH. A strong dependence of apparent flexibility is observed on ionic strength at neutral pH. The dependence of flexibility on ionic strength is also clear, but weaker, in acidic pH.

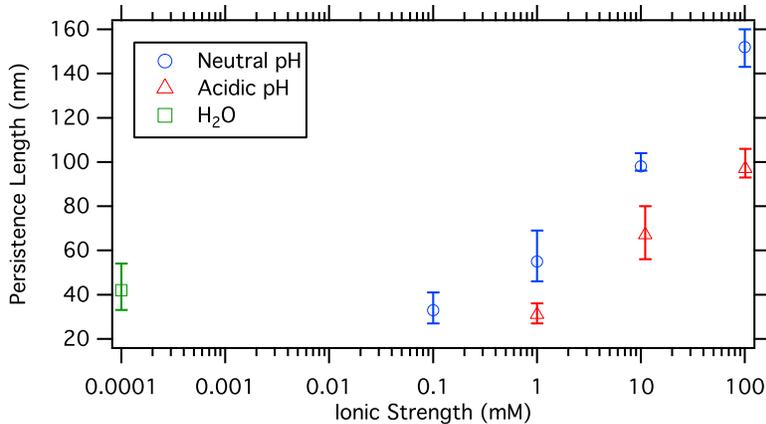


Figure 3.21: Dependence of persistence length on ionic strength at neutral and acidic pH. The shown values are the average  $l_p$  over the range of  $100 < s_{\max} < 200$  nm\*. Error bars represent the maximum and minimum  $l_p$  obtained from fitting over this range.

\*  $150 < s_{\max} < 250$  nm for 10 mM KCl and 100 mM KCl at neutral pH, and 100 mM KCl at acidic pH.

Considering instead the curved WLC model, I find that the value for  $R_0$ , *i.e.* inherent/induced curvature of the chain (Equation 3.5), increases with ionic strength, as shown in Figure 3.22. As  $R_0$  increases, the chains converge to standard WLC behaviour. No values are derived for higher ionic strengths since  $R_0$  becomes very large: in these cases, the data are well described by the standard WLC model (Equation 2.6). These experiments do not distinguish intrinsic molecular curvature from surface interactions.

### 3.3.2 Summary and Discussion

Ionizable amino acid residues are found in abundance in collagen's  $\alpha$ -chains [83]; hence, it is expected that electrostatic interactions contribute to stability of the triple helix. It has been found that electrostatic interactions play a key role in thermal stability of the triple helix [27, 28, 42]. In order to unravel the effect of solution conditions on mechanical properties of the chains, conformations of type I collagen were studied in different solutions using AFM imaging. These measurements were performed following deposition from solution and subsequent drying. Solution conditions of different pH and ionic strengths used in this study can be found in Table 3.4. The ionic strengths of the collagen solution were incrementally increased in both neutral and acidic pH.

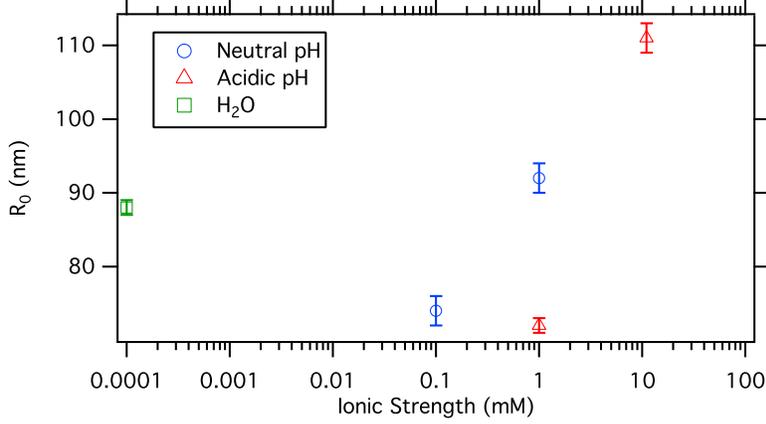


Figure 3.22: Dependence of collagen radius of curvature  $R_0$ , on ionic strength of the solution from which rat tail collagen type I is deposited on mica. Error bars represent 95% confidence of the fit.

In both neutral and acidic pH, at higher ionic strengths molecule configurations were straighter, exhibiting less bending flexibility. Therefore, the apparent persistence length increases with ionic strength. An increase in persistence length with ionic strength is in agreement with a previous study [18], which found larger values for persistence length in high salt buffers at neutral pH compared with a low ionic strength acidic solution. This study employed a simpler analysis strategy, in which only the ends of the molecules (the square of end-to-end separations) were used to determine the persistence length of collagen.

By separating the effects of pH and ionic strength, I have shown that the persistence length depends independently on both of these parameters. The dependence of apparent persistence length on ionic strength at neutral pH is stronger than that at acidic pH. Possible mechanisms are proposed later in this section. My results show that collagen's conformations, as imaged on mica, depend strongly on ionic strength as well as on pH of the solution.

Collagen's increasing persistence length with ionic strength contrasts with the expected behaviour of a flexible polyelectrolyte. For example, DNA becomes more flexible (decreased persistence length) at higher monovalent salt concentrations [84]. One proposed mechanism for the observed behaviour of DNA is that at higher ionic strengths, interaction of the solution's ions with the charged phosphates of DNA screen the electrostatic repulsion between neighbouring regions of the negatively charged backbone, which results in a lower energy penalty for bending [84].

The range of electrostatic interactions for a polyelectrolyte in solution is governed by a characteristic length, the Debye screening length. This is the length over which the electric potential decreases by  $1/e$ :

$$\lambda_D = \sqrt{\frac{\epsilon\epsilon_0 k_B T}{2e^2 N_A I}}. \quad (3.6)$$

Here,  $\epsilon$  is the dielectric constant,  $\epsilon_0$  the vacuum permittivity,  $N_A$  Avogadro's number,  $e$  the electron charge and  $I$  the ionic strength of the solution. Table 3.5 summarizes Debye lengths for ionic

Table 3.5: Debye screening length of solutions with different ionic strength.

Solution	Ionic Strength (mM)	Debye Length (nm)
H <sub>2</sub> O	0.0001	961
0.1 mM KCl	0.1	30.40
1 mM KCl	1	9.61
10 mM KCl	10	3.04
100 mM KCl	100	0.96
1 mM HCl	1	9.61
1 mM HCl + 1mM KCl	2	6.80
1 mM HCl + 10 mM KCl	11	2.90
1 mM HCl + 100 mM KCl	101	0.96

strengths of the solutions studied here. Figure 3.23 plots the dependence of persistence length of collagen on Debye length.

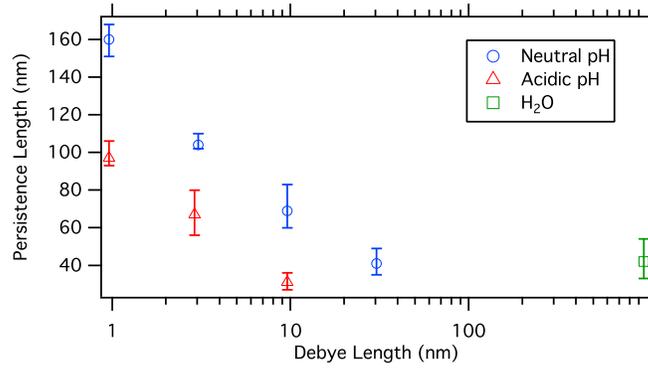


Figure 3.23: Dependence of apparent persistence length on Debye length at neutral and acidic pH. The shown  $l_p$  values are based on fit results of the WLC model to  $\langle R^2 \rangle$  data for  $s_{\max} = 200$  nm. Error bars represent the maximum and minimum resulting  $l_p$  when  $s_{\max}$  changes between 150 nm and 250 nm.

As seen from Equation 3.6, the Debye screening length decreases with the ionic strength of the solution. Therefore, the extent of electrostatic interactions around a charged polymer is affected by the ionic strength of the solution. For collagen at neutral pH, negatively and positively charged residues are distributed throughout the slightly positively charged triple helix. At low ionic strengths, charged residues along the chain can interact over relatively long distances because of the longer Debye length (Equation 3.6). This implies that each charged residue may interact with multiple other charged residues. As the ionic strength is increased, long-range electrostatic interactions are screened and local "salt bridges" are formed between charged units at close proximity (Figure 3.24) [28, 85, 86]. The term "salt-bridge" is used to describe stabilizing electrostatic interactions formed between oppositely charged side chains (*e.g.*, arginine or lysine and aspartic acid or glutamic acid) within 0.4 nm of each other [85, 87]. There are less favourable yet stabilizing interactions beyond this range; however, the diffuse electrostatic interactions between long-range ion pairs are destabilizing [87], because interactions with other charged residues in a

variety of directions weaken the net electrostatic force. Note that as the ionic strength increases, the electric potential from the charged particle drops as  $V(r) \propto \frac{1}{r}e^{-r/\lambda_D}$ , hence, electrostatic interactions decrease for charges at all separations. However, the strength of interactions between neighbouring charges can still remain strong (relative to thermal energy), while longer-range interactions become of negligible strength. In my experiments, at the highest ionic strength, the Debye length of  $\approx 1$  nm is still greater than the length of a salt bridge, meaning that the salt-bridges will be intact [85], but other long-range interactions are screened out. The salt bridge mechanism was proposed to explain the increased thermal stability and helicity of collagen's triple helix with increased KCl concentration found in a previous study [28]. This salt-bridge mechanism can contribute to the dependence of  $l_p$  on ionic strength found here at neutral pH.

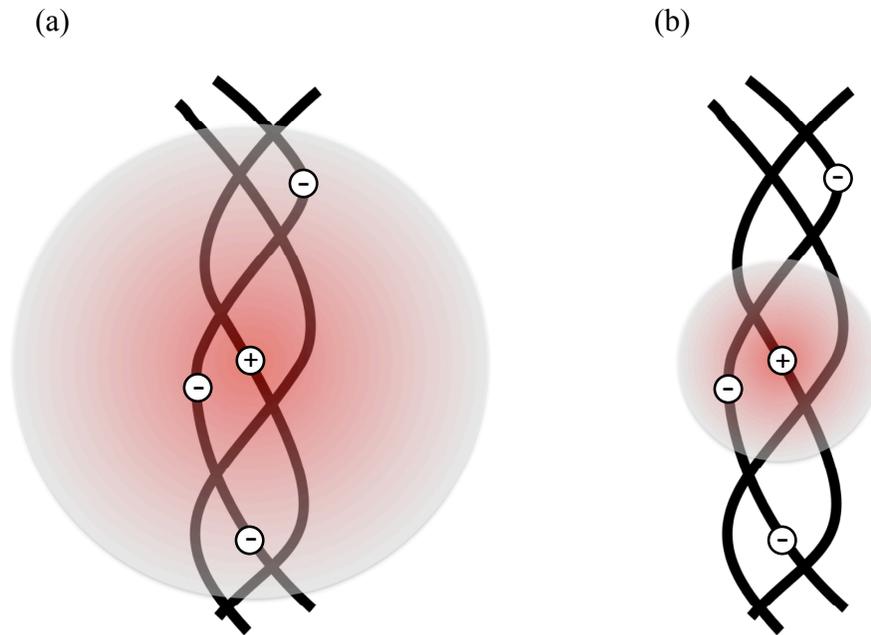


Figure 3.24: Schematic of electrostatic interactions in collagen in solutions of different ionic strengths. The shaded circles represent the Debye length from the central positive charge, beyond which electrostatic interactions are decreased significantly (not to scale). (a) At low ionic strength, charged residues on the chain feel long-range electrostatic interactions that detract from a local stabilizing attraction. (b) As the ionic strength is increased, long-range electrostatic forces are screened and nearby oppositely charged residues can form local, stabilizing "salt bridges". Adapted from Reference [28].

In acidic pH, collagen's persistence length also increases with ionic strength but not as strongly as at neutral pH. At  $\text{pH} \sim 3$ , collagen's triple helix exhibits a high positive charge (Figure 3.25), with very few, if any, negatively charged residues. Thus, salt-bridge formation cannot explain

the increase in persistence length with ionic strength. A decreased thermal stability of the triple helix was observed for collagen at acidic pH [88–90]. At acidic pH, collagen is positively charged, and repulsion between strands could lead to an underwound triple helix (Figure 3.26). Selective ion-binding by positively charged residues on the collagen triple helix [91, 92] may account for the increased stability of collagen with increasing ionic strength at low pH. Addition of salt and selective binding to positively charged sites could lead to formation of a more tightly wound helix, which would be expected to exhibit a higher persistence length (Figure 3.26).

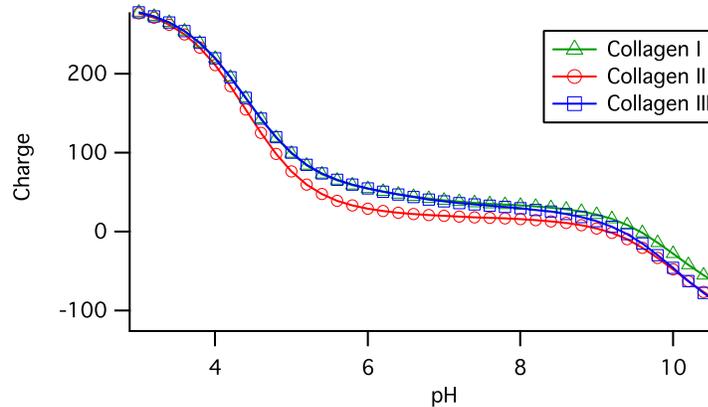


Figure 3.25: Estimate of the pH-dependent charge of triple helical collagen. The charge is calculated using a protein calculator [93] and assumes that all residues have pKa values equal to their isolated values. Post-translational modifications are not taken into account. UniProt Knowledgebase accession numbers: human CO1A1: P02452; human CO1A2: P08123; human CO2A1: P02458; human CO3A1: P02461.

The pH of the medium influences a range of collagen-based processes from fibrillogenesis [94] to bone resorption [95]. It is well known that collagen self-assembly to form fibrils is accelerated at neutral compared with acidic pH [94, 96]. It has been proposed that the extended conformation of collagen molecules at higher pH facilitates alignment of the molecules for self-assembly and exposes the interacting sites for fibril formation [18]. My study on the effect of pH on collagen molecules indicates collagen is more flexible at acidic pH than at neutral pH. Perhaps the loosely wound triple helix and weakening of collagen’s structure in acidic conditions accelerates its degradation, in turn facilitating the demineralization of bones [95].

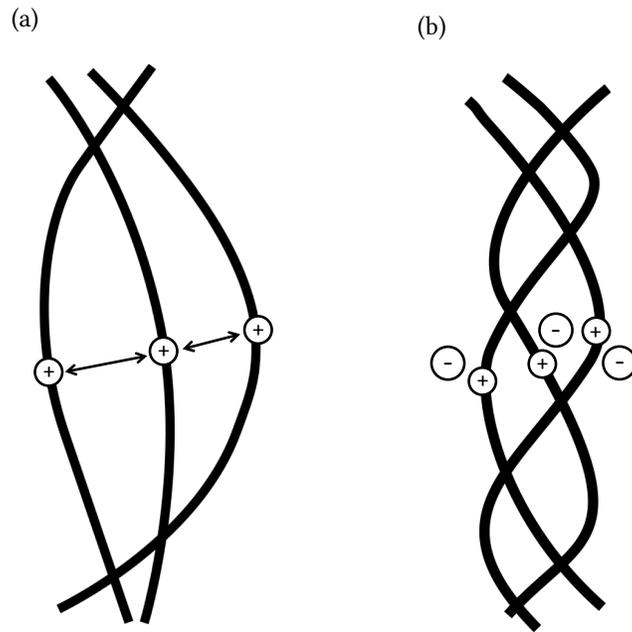


Figure 3.26: Schematic of a region of collagen's triple helix at acidic pH. (a) At low ionic strengths, electrostatic repulsion between charged residues may cause local unwinding of the helix. (b) At higher ionic strengths, it is likely that the positive charges along the helix interact with negative ions present in solution. This would screen repulsive electrostatic interactions between the  $\alpha$ -chains, permitting a tighter helix.

## Chapter 4

# Optical Tweezers Stretching Experiments

### 4.1 Introduction

In 1970, Arthur Ashkin introduced the concept of optical trapping with highly focused laser light [97]. Later, Optical Tweezers (OT) were invented by Ashkin and his colleagues, who trapped microscopic dielectric objects at the centre of a highly focused laser beam by a transfer of momentum of light [98]. This technique has found a broad range of applications in biology and physics [99]. It has been widely used by scientists to study biological systems ranging from cells and filamentous networks to single molecules and molecular motors. Optical tweezers apply biologically relevant forces to study mechanical properties of biological molecules and their response to tension. Because of their ability to manipulate and study individual molecules, optical tweezers are powerful tools in molecular biology, biochemistry and biophysics. In the current and following chapters, the use of a home-built OT setup is described to stretch single collagen molecules in order to study their mechanical response to applied force [100–102].

One of the advantages of single-molecule force spectroscopy is that it is not associated with population averaging in ensemble measurements. Population averaging can obscure rare or transient properties of the system. Among the different techniques for single-molecule force spectroscopy, optical tweezers, atomic force microscopy and magnetic tweezers are the most commonly used [103]. Each of these techniques has its own applications and provides different force ranges, and temporal and spatial resolutions. The ranges and limitations of these techniques are summarized by Neuman and Nagy (Table 4.1 [103]) and are outlined briefly below.

#### Atomic Force Microscopy

Besides mapping surface properties at Ångström resolution, atomic force microscopes (AFM) can also be used to manipulate single molecules [104, 105]. In this method, the setup is similar to that shown in Figure 2.1. However, when stretching a single molecule, it is attached between the tip

	<b>Optical tweezers</b>	<b>Magnetic (electromagnetic) tweezers</b>	<b>AFM</b>
Spatial resolution (nm)	0.1–2	5–10 (2–10)	0.5–1
Temporal resolution (s)	$10^{-4}$	$10^{-1}$ – $10^{-2}$ ( $10^{-4}$ )	$10^{-3}$
Stiffness (pN nm <sup>-1</sup> )	0.005–1	$10^{-3}$ – $10^{-6}$ ( $10^{-4}$ )	$10$ – $10^5$
Force range (pN)	0.1–100	$10^{-3}$ – $10^2$ (0.01– $10^4$ )	$10$ – $10^4$
Displacement range (nm)	0.1– $10^5$	5 – $10^4$ (5– $10^5$ )	0.5– $10^4$
Probe size (μm)	0.25–5	0.5–5	100–250
Typical applications	3D manipulation Tethered assay Interaction assay	Tethered assay DNA topology (3D manipulation)	High-force pulling and interaction assays
Features	Low-noise and low-drift dumbbell geometry	Force clamp Bead rotation Specific interactions	High-resolution imaging
Limitations	Photodamage Sample heating Nonspecific	No manipulation (Force hysteresis)	Large high-stiffness probe Large minimal force Nonspecific

Figure 4.1: Comparison of single-molecule force spectroscopy techniques. With permission from Reference [103].

and the surface, and the surface is displaced vertically relative to the tip in order to stretch and relax the tethered molecule (Figure 4.2). The tip is mounted on a cantilever, which bends when the molecule is being manipulated. The end-to-end separation of the tethered molecule can be monitored by measuring the deflection of the laser beam reflecting from the back of the cantilever. The applied force, which can range between 10 pN and 10 nN, depends on cantilever stiffness.

The most straightforward attachment of molecules to the surface (usually mica or glass) is through non-specific adsorption, which lacks precision on where the molecule is bound on the surface. Molecules can be specifically bound to the surface covalently [106] or using antibodies [107]. Another portion of the molecule becomes attached to the AFM tip when it touches the surface and subsequently, the tethered molecule can be manipulated. From the resulting force-extension curve, the molecule's response to applied force can be studied.

AFM has limitations due to the relatively high stiffness of its cantilever, leading to a high loading rate and large minimal force (> 10 pN) which is beyond the range of some biological forces such as the entropic elasticity of DNA [104]. Another drawback of AFM stretching experiments is the non-specific interactions commonly used to adhere the molecule between the probe and the surface, which results in experiments stretching an uncontrolled portion of the molecule of interest.

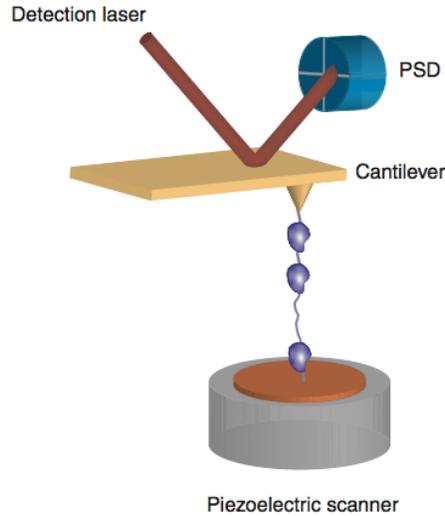


Figure 4.2: Schematic of an atomic force microscopy stretching experiment. The molecule (purple) represents a polyprotein attached to the substrate surface (copper) from one end and to the AFM tip from the other end. (Figure from Reference [103] with permission).

### Magnetic Tweezers

Magnetic tweezers are based on a pair of magnets held above a sample substrate in an inverted microscope [108]. A microscopic magnetic particle experiences a force proportional to the gradient of the square of the external magnetic field. With one end of the molecule attached to the sample substrate surface and the other end attached to the magnetic bead, the tethered molecule can be manipulated (Figure 4.3). Magnetic tweezers allow controlled rotation of the bead and can be used not only to stretch and relax the molecule [51, 108] but also for twisting experiments [108, 109]. The image of the magnetic bead is recorded by a camera, which limits the spatial and temporal resolution of this technique compared to optical tweezers and atomic force microscopy methods.

### Optical Tweezers

Optical tweezers provide a trap for a microsphere at centre of highly focused light. This technique provides high temporal and spatial resolution of particle position detection and enables scientists to study molecules in the low force regime by slowly stretching and relaxing the molecule (Figure 4.4). Details of the technique are provided in Section 4.2. The low force range is of special importance when it comes to studying the mechanical properties of collagen. The response of the molecule to low applied forces reveals information on the entropic regime and flexibility of the molecule. In the latter half of this thesis, mechanical properties of type II human procollagen and type III collagen are studied using optical tweezers.

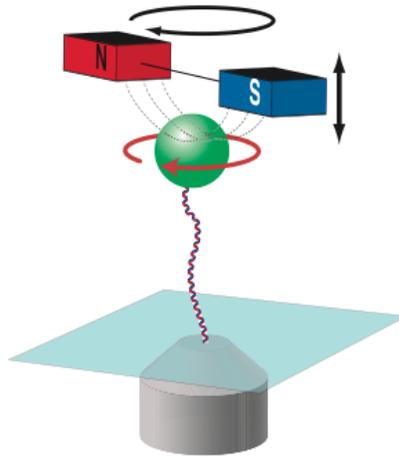


Figure 4.3: Schematic of magnetic tweezers setup. The figure shows a super-paramagnetic bead (green) pulled by the magnetic field gradient produced by the magnets (dashed lines). The tethered molecule represents a DNA molecule attached to the bead and at its other end to the surface of the sample chamber. The magnets can move up and down and also rotate. (Figure from Reference [103] with permission).

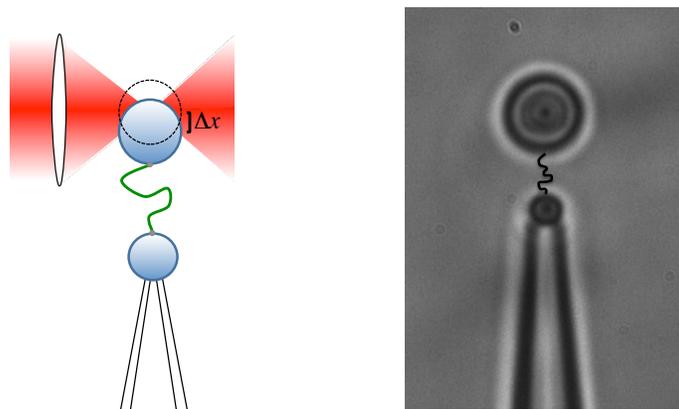


Figure 4.4: Schematic (left) and experimental (right) figures of a stretching experiment using optical tweezers. The tethered molecule represents a single molecule of DNA (green) tethered between two microspheres, one trapped in the optical trap and other immobilized on tip of a micropipette.  $\Delta x$  represents the displacement of the trapped bead from equilibrium. Red represents the focused laser light.

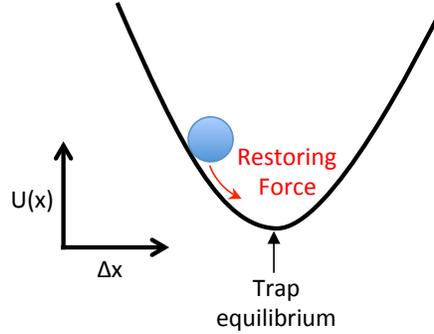


Figure 4.5: Schematic of a trapped bead in a parabolic potential well from the trap. For small displacements from the trap equilibrium, the trapped object experiences a Hookean restoring force towards the equilibrium point.

## 4.2 Optical Tweezers

### 4.2.1 Principles of Optical Trapping

Optical tweezers trap a micron-sized dielectric object placed near a highly focused laser beam [97]. As a result of the interaction between the object and the light, a momentum exchange confines the object within the focus of the beam. The principles of optical trapping and its applications in studying biological systems are thoroughly reviewed by Neuman and Block [110]. Their approach forms the basis of this section.

In an optical tweezers setup, the laser light is highly focused by a high numerical aperture (NA) objective lens to a diffraction-limited spot. A dielectric particle located near the focus interacts with the light, changing its momentum, which results in an optical force confining the bead in the trap. Depending on the size of the particle (here, a spherical dielectric object of radius  $R$ ) relative to the wavelength of the light ( $\lambda$ ), there are two regimes for which the trapping force can be exactly calculated: 1) Mie scattering regime for  $R \gg \lambda$ , and 2) Rayleigh scattering regime for  $R \ll \lambda$ .

Conventionally, the trapping force has been decomposed into two components: 1) scattering force, which is along the direction of propagation of light (optical axis), and 2) gradient force, in both axial and lateral directions. The scattering force arises from the scattering of photons of the incident light. The net scattering force for an isotropic scatterer is in the forward direction (like a "fire hose" of photons). A steep intensity gradient of the light along the optical axis is required for stable trapping in three dimensions and to overcome the scattering force. This steep intensity gradient is created by sharply focusing a laser light with a gradient in the transverse direction (such as a Gaussian profile laser beam) using a high numerical aperture lens. This gives rise to the second force component: the gradient force. The balance between the gradient and scattering forces creates a restoring force towards the focus of the laser. For small displacements from the equilibrium position, i.e. the centre of the trap, the trap acts as a Hookean spring (Equation 4.3), with stiffness determined by intensity of the light (Figure 4.5).

In the Mie regime, ray optics can be used to compute the optical forces. The incident light is refracted by the microsphere, which leads to change in the momentum of light ( $\Delta p_{\text{net}}$ ). The bead's momentum change, from Newton's third law, is equal but in the opposite direction of the change in light momentum, resulting in an optical restoring force on the bead ( $F_{\text{bead}}$ ) (Figure 4.6).

In the Rayleigh regime, the particle can be treated as a point dipole interacting with the radiation. The gradient force is given by [110]

$$F_{\text{gradient}} = \frac{2\pi\alpha}{cn_m^2} \nabla I_0, \quad (4.1)$$

where  $I_0$  is the intensity of the incident light,  $c$  is the speed of light,  $n_m$  is the refractive index of the medium,  $\alpha = n_m^2 R^3 \left(\frac{m^2-1}{m^2+2}\right)$  is the polarizability of the microsphere,  $n_p$  is the refractive index of the particle, and  $m = \frac{n_p}{n_m}$  is the relative refractive index between the particle and the medium. The scattering force is given by [110]

$$F_{\text{scattering}} = \frac{\sigma n_m}{c} I_0, \quad (4.2)$$

where  $\sigma = \frac{128\pi^5 R^6}{3\lambda^4} \left(\frac{m^2-1}{m^2+2}\right)^2$  is the scattering cross-section of the microsphere.

In our studies as well as most biophysical studies, the size of trapped microspheres is comparable to the wavelength of the trapping laser (*i.e.*  $\frac{\lambda}{R} \approx 1$ ). A rigorous treatment of the electromagnetic theory is required to quantify the optical trapping in the intermediate regime [111], which is outside the scope of this thesis. Here, both the ray optics and the Rayleigh approximations provide a basis to understand the principles of the optical trapping, and provide the same relevant result, *i.e.*, that the trap can be treated as harmonic for small displacements of the particle from equilibrium.

## 4.2.2 Instrument

In this section, the OT setup used in this work is briefly described (Figure 4.7 and 4.8). A high numerical aperture objective lens (Olympus UPlanApo/IR, NA 1.2, 60 X water-immersion) is used to focus an 835 nm, 200 mW diode laser Gaussian beam (assembled by Melles Griot) into a microfluidic flow chamber. This wavelength is not strongly absorbed by water, thus minimizing heating in the sample [112]. A manually pulled glass micropipette is inserted in a sample chamber (Figure 4.9). The chamber is mounted on a piezo-electric stage (Mad City Labs, Nano H-50), allowing it to be moved relative to the optical trap with nanometer-scale precision. The manipulation is in a plane perpendicular to the optical axis. Using a second, identical objective lens, the laser light is re-collimated and directed onto a duolateral photodiode (UDT Sensors, DL-10) that images the back-focal plane of the second objective. The photodiode detects deflections of the light arising due to the trapped object's displacement from the trap centre in directions perpendicular to the optical axis. These deflections are recorded at a high bandwidth of up to 100 kHz. In addition, images are recorded at 60 Hz using a CCD camera (Flea, Point Grey Research), from which the

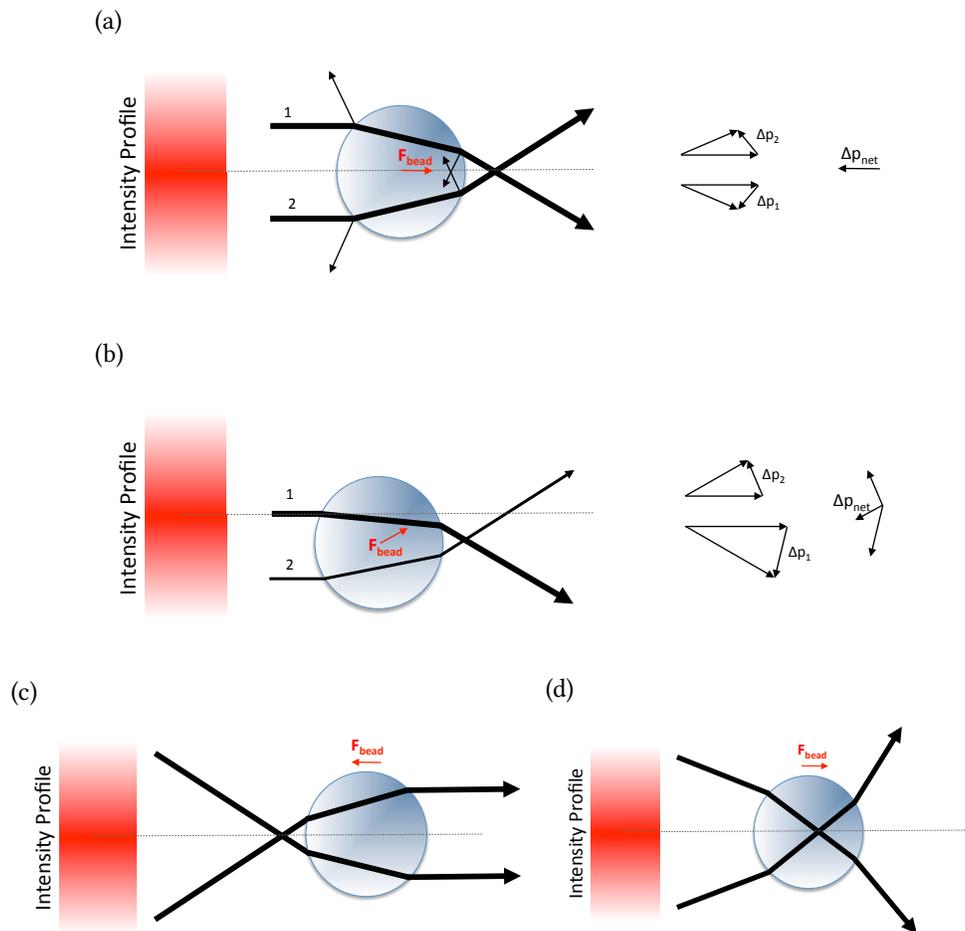


Figure 4.6: Ray optics schematic of the gradient force for a dielectric microsphere. As the incident laser beam passes through the bead, its momentum vector changes as a result of refraction ( $\Delta p_{\text{net}}$ ) creating an optical force on the bead in the opposite direction ( $F_{\text{bead}}$ ). (a) Refraction of a collimated light by a bead when the bead's centre is on the optical axis (dashed line). The lateral forces cancel each other out resulting in a net force only along the optical axis. (b) For the centre of the bead laterally displaced from the optical axis of collimated light, the intensity gradient in the transverse direction creates a restoring force toward the centre of the beam. (c) For a bead downstream of a focus, the net force pulls the bead towards the focus (upstream). (d) For a bead downstream of the focused laser beam, the net force pushes the bead upstream.

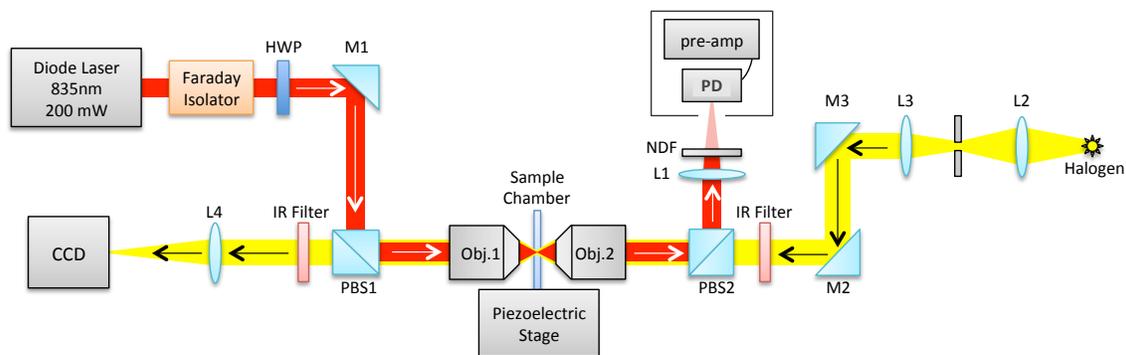


Figure 4.7: Schematic of the single beam optical tweezers setup used for single molecule force spectroscopy of collagen. List of abbreviations: HWP: half wave plate, M: mirror, PBS: polarized beam splitter, Obj: objective lens, L: lens, NDF: neutral density filter, PD: photodiode, pre-amp: pre-amplifier, IR: infra-red, CCD: camera.

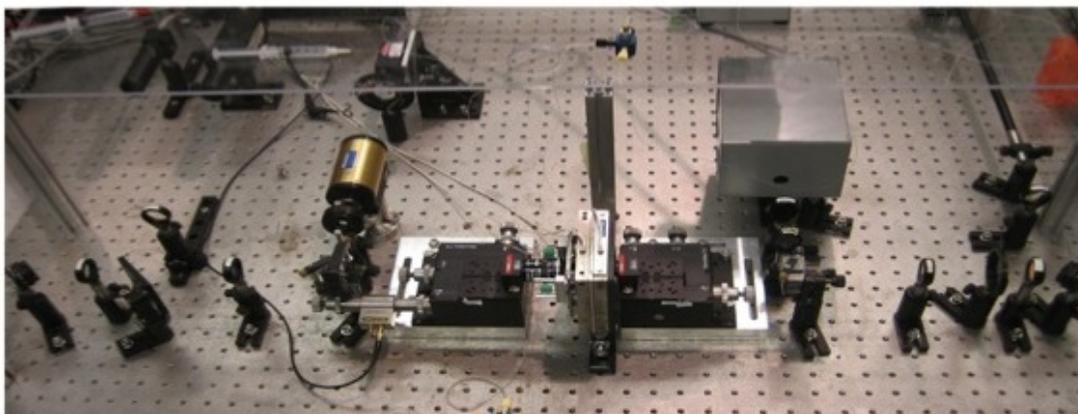


Figure 4.8: Photo of the single beam optical tweezers used to stretch single molecules. The layout is approximately that in Figure 4.7

positions of the optically trapped and pipette-immobilized beads can be determined (Figures 4.7 and 4.8).

### Microfluidic Flow System

Solutions of microspheres suspended in aqueous buffer are kept in syringes connected to polyethylene tubes (PE-10, Intramedic). These tubes connect the samples to the chamber where the trapping and manipulation occurs (Figure 4.9). The chamber consists of two layers of Nescofilm (Karlan, N-1040), a tube for pipette insertion is placed between the two layers of Nescofilm, which in turn are sandwiched between two coverslips. One of the coverslips has precisely drilled small holes, which are connected to tubes that form a watertight seal allowing fluids to be flowed in and out of the chamber.

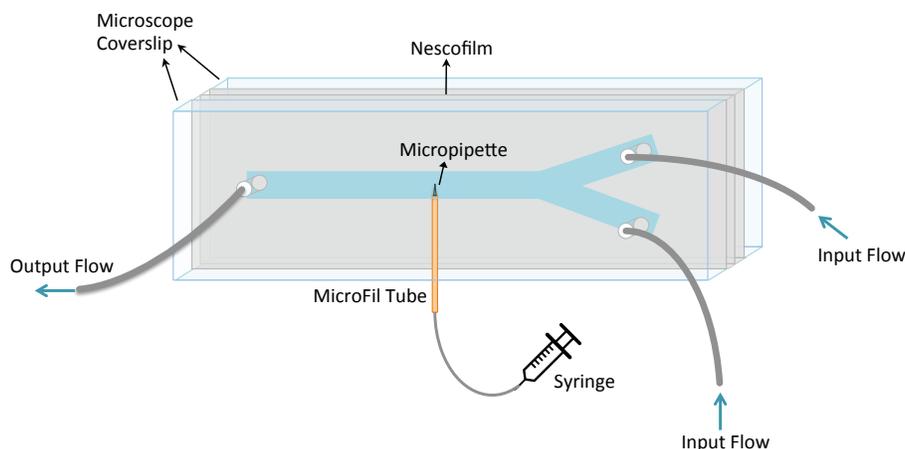


Figure 4.9: Schematic of the microfluidic chamber used for optical trapping for single molecule manipulation experiments. The Y-shaped channel cut in the Nescofilm is where the beads flow. The Nescofilms are then sandwiched between two microscope coverslips. The micropipette positioned inside a MicroFil tube (World Precision) is secured between the Nescofilms. The syringe connected to the micropipette via a polyethylene tube provides suction to immobilize the beads during the experiment. The pipette tip has an approximate diameter of 1  $\mu\text{m}$ , smaller than the beads used in these experiments. The input and output flow are allowed through small holes drilled in the top coverslip and are connected to polyethylene tubes through watertight seals.

Because of the dimensions of the chamber (75 mm length by 25 mm width by 0.17 mm thick) and the flow rates, the flow inside the chamber is laminar. There are two channels entering the chamber, one containing beads and the other, buffer (no beads). The buffer channel is used to regulate the flow boundary between beads and buffer, and also to flush the chamber to avoid bead interference during the experiment. The flow speed is controlled by adjusting the height of external containers containing the samples and the buffer with respect to the outlet waste container (Figure 4.10).

In a pulling experiment (Figure 4.7), two different sizes are used in order to distinguish the microspheres: the smaller bead is held on the tip of the micropipette and the larger bead held in the trap. The small beads are flowed until a bead is trapped and then sucked on the micropipette by suction. Then the bead sample channel is switched to deliver the larger microspheres with the molecules attached. Next, the sample channel is closed and the buffer only flows continuously until the excess beads are washed from the chamber. This avoids unwanted interference during the experiment. Flow is stopped by closing valves before measurements begin.

### 4.3 Measurements and Calibration

The mechanical response of a protein can be probed with optical tweezers by applying picoNewton-scale forces to a microsphere tethered to the end of the protein, thereby manipulating its extension. The resulting force-extension curve provides information on the protein's mechanical properties

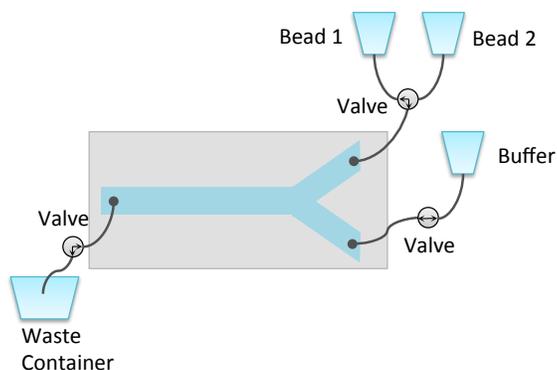


Figure 4.10: Schematic of flow system. The relative height of the sample and the buffer containers with respect to the waste container controls speed of the flow.

such as elasticity and extensibility. Force experienced by the molecule is linearly proportional to displacement of the trapped bead from the trap centre with the proportionality constant equal to the trap stiffness. The extension of the molecule can be calculated from the position of the microspheres with respect to each other. Therefore, calibration of the trap for trap stiffness and also measurement of the bead positions are essential to obtain the force-extension curve from the experiment.

### 4.3.1 Position Measurement

The positions of the two microspheres are measured using two independent methods. In the first method, a CCD camera captures the images during the experiment for later analysis. Secondly, the deflection of the laser light from the trapped bead is detected with a photodiode (PD). In our instrument, the forward-scattered light is used for the position measurement of the trapped bead. The piezoelectric stage voltage is used to measure the position of the immobilized bead on the micropipette. Each of these methods has its own advantages and disadvantages are discussed in the following sections.

#### Low Bandwidth Position Measurement

The CCD camera records images of  $640 \times 480$  pixels, 8-bit grey scale at 60 frames per second. The camera sampling rate is fast enough to capture the changes in the position of the microspheres during the experiment but is not fast enough to track all of their Brownian motion. The saved images of the beads are used for position analysis. Two different algorithms have been tested for this: Convolution and Edge finder.

**Convolution-Based Image Analysis** A home-written LabVIEW program is used to measure displacement of a bead using a spatial correlation algorithm [113]. In this analysis method, a template image of the bead of interest is selected along with a search area around it. Then for

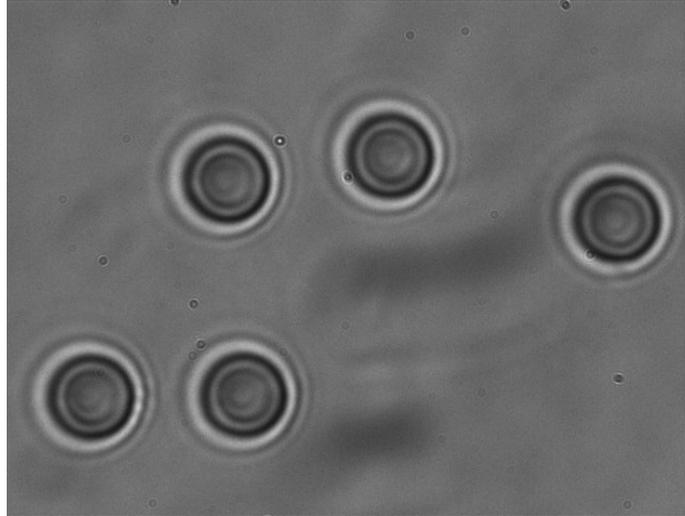


Figure 4.11: Image of 2.89  $\mu\text{m}$  diameter polystyrene beads were stuck on the surface of a coverslip in a sealed chamber. The positions of the beads were analyzed using the two image analysis approaches and from the uncertainty of the positions, the accuracy of each algorithm was derived.

each image frame a correlation matrix of the template image in register with each position within the search area is calculated. A  $7 \times 7$  pixel region around the maximum of the correlation matrix is fitted to a two dimensional parabolic function that gives a sub-pixel measurement of the position of the particle in the current image frame. This method gives the relative position of each microsphere with respect to the template image, usually chosen to be the first frame.

**Edge-Finder Algorithm** In this method, images saved from the CCD camera are analyzed using a LabVIEW function, "IMAQ Find Circular Edge". It identifies the edges of the microsphere along radial search lines based on the intensity gradient. The intensity gradient parameters are specified by the user. A circle is then fit to the edge points and the centre of the circle represents the bead position.

**Comparison of Image Analysis Methods** Each of the two methods introduced above has its own advantages and disadvantages. The precision depends on the size of the microspheres, illumination of the field of view and alignment of the setup. As an example the precision of each algorithm was calculated for 2.89  $\mu\text{m}$  diameter beads. The images of five microspheres attached to the surface of the coverslip in a sealed chamber were saved for 4 minutes at 60 Hz (Figure 4.11).

The displacements of the beads were analyzed using both image analysis algorithms. The drift of the system was subtracted from the displacements by assuming the separations between the beads remain constant at all times (Figure 4.12). The spread in the extracted displacements of each particle with respect to its drift-corrected mean is assigned to be the uncertainty of the algorithm (Figure 4.13). Two times the standard deviation of the displacements was considered for the algorithm precision. For the convolution-based algorithm, it is  $1.4 \pm 0.1$  nm. For the circle

edge-finder algorithm, the precision is  $5.0 \pm 0.1$  nm. These values are specific to this size of bead with particular illumination condition.

The precision of the circle edge-finder algorithm strongly depends on how circular the bead appears. When the image is obscured by surrounding objects such as the pipette or overlap of the edges of beads in close proximity, the resolution decreases significantly. However, the portions of the beads that overlap during the stretching experiments can be excluded by defining only a slice of the bead images for analysis.

The resolution of the convolution algorithm is higher if the appearance and illumination of the images don't change in different frames. For example, the convolution algorithm can produce a biased location of the trapped bead when the image of the bead on the pipette enters the search area. Moreover, at higher forces the trapped bead displaces also along the optical axis. Moving to a different focal plane makes the trapped bead appear different compared to its appearance at lower forces and the convolution algorithm precision drops. Therefore, this algorithm cannot be used when analysis at higher forces is desired.

### **High Bandwidth Position Measurement**

As mentioned earlier, the PD and piezoelectric stage can sample the positions of the microspheres at much higher bandwidth than the CCD camera rate. As the trapped bead moves in the trap during a stretching experiment, the trapping laser light is deflected. The amount of deflection depends on the lateral displacement of the bead from its equilibrium position in the trap. The focused laser light is then re-collimated by a second objective lens; the change in the position of the trapped bead shifts the centre of the re-collimated light which is then detected by the PD.

A piezoelectric stage positions the flow chamber with sub-nanometer precision. Therefore, the relative position of the trapped bead and the bead on the micro-pipette can be controlled enabling stretching and relaxing of the tethered molecule. The piezoelectric stage data are saved to give the positions of the micropipette-immobilized bead. These data are used to calculate the relative position of the two beads. Using stage data gives the precise position of the chamber, however, it does not account for drift of the micropipette during the experiment.

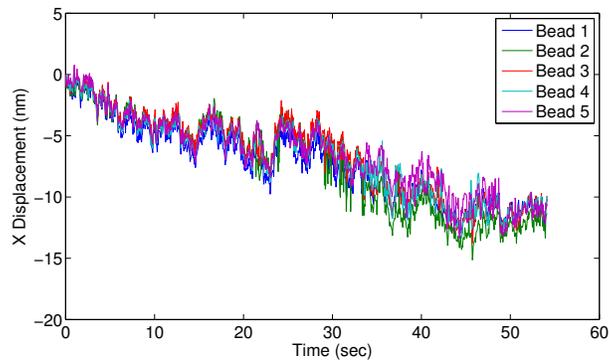
The PD and stage position measurements are recorded at either 1 kHz (stretching experiment) or 100 kHz (trap calibration).

### **4.3.2 Calibration**

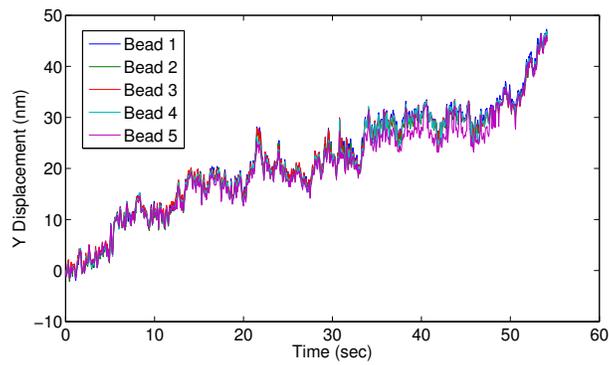
#### **CCD Camera**

The CCD camera captures a magnified image of the microspheres. The bead displacements are determined in pixels and need to be converted into nanometers. For pixel calibration, a microsphere immobilized on the tip of a micropipette is stepped through the field of view in horizontal (X) and vertical (Y) direction by moving the flow chamber with the piezoelectric stage. The piezoelectric stage is calibrated by the manufacturer with sub-nanometer precision, thus stepping the stage is

(a)



(b)



(c)

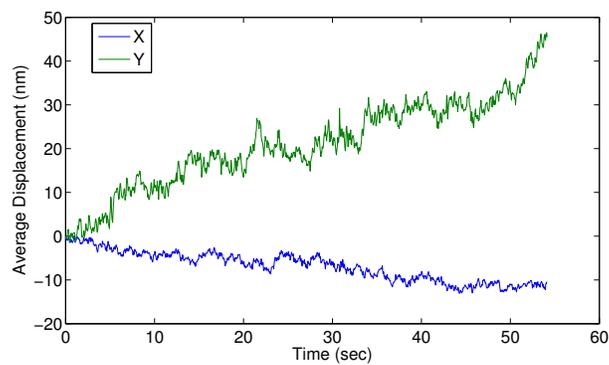


Figure 4.12: Stuck bead displacements as a function of time as calculated using convolution-based algorithm. (a) X direction and (b) Y direction. (c) The average displacements of the beads as a function of time.

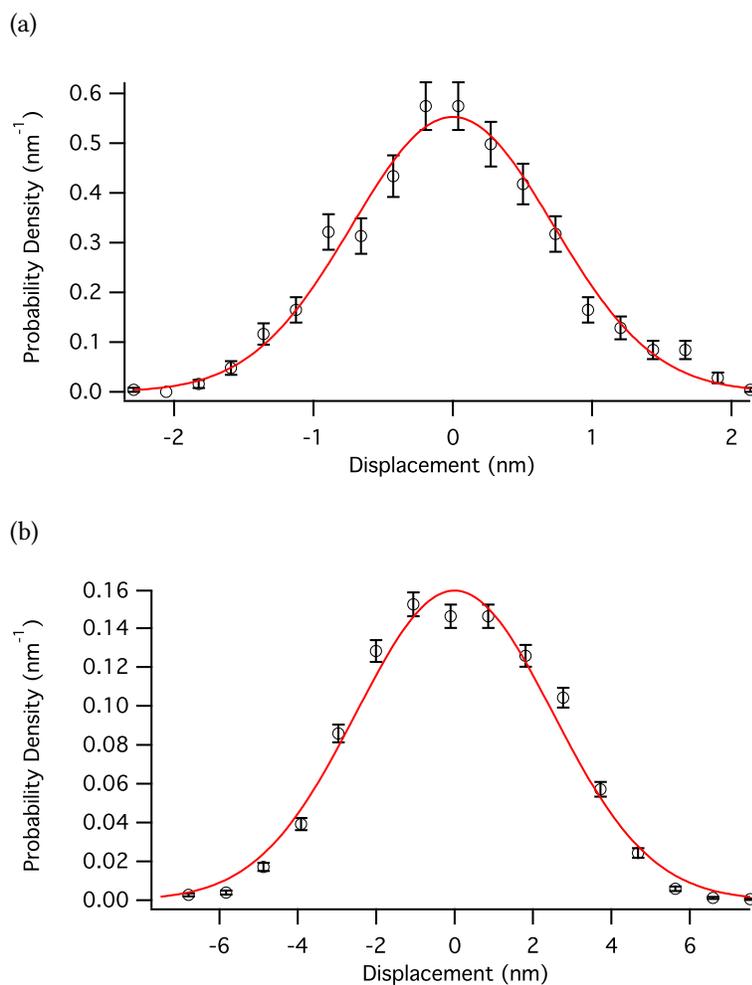


Figure 4.13: Representative probability density (black circle) and Gaussian fit (red) of drift-corrected displacements of one of the stuck beads calculated with (a) convolution-based algorithm and (b) circle edge-finder algorithm. Twice the standard deviation of the positions was considered for the algorithm precision. The position uncertainty for convolution-based algorithm is  $1.4 \pm 0.1$  nm, while it is  $5.0 \pm 0.1$  nm for the circle edge-finder algorithm. The error bars represent normalized values of  $\sqrt{N}$ , where  $N$  is number of counts in each bin.

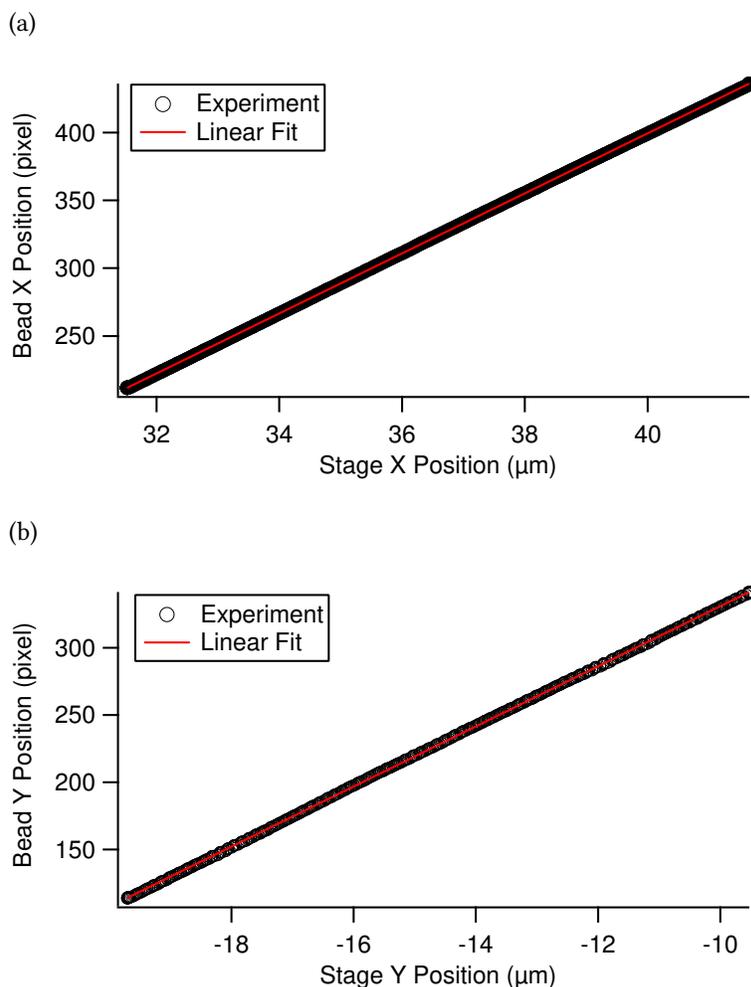


Figure 4.14: Pixel calibration of CCD camera. The calibration factors are (a)  $22.122 \pm 0.004$  pixel/ $\mu\text{m}$  in x direction and (b)  $22.334 \pm 0.008$  pixel/ $\mu\text{m}$  in y direction.

used as the displacement standard. I have independently confirmed the calibration using a microscope ruler. The position coordinates of the bead are analyzed using the image analysis methods described earlier. The resulting bead locations are plotted against stage readings and from a linear fit the nanometers per pixel conversion factor is calculated (Figure 4.14). This calibration is independent of bead size, but would change for different alignment of the imaging optics.

### Position Sensitive Detector

The PD data are saved in volts and can be converted to units of length using the already calibrated camera data. For this calibration the trapped bead is displaced through extension or relaxation of the tethered molecule. Because the trapped bead position is simultaneously recorded with both camera and PD, the PD data can be plotted against the camera positions. A linear fit gives the conversion factor from Volts to nanometers (Figure 4.15). Given that there is variation in bead

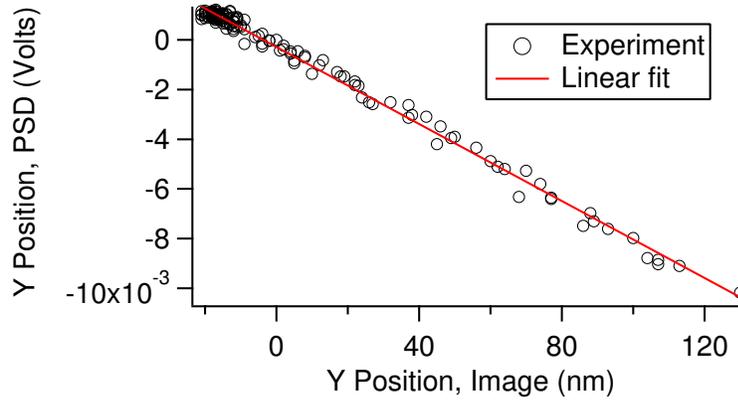


Figure 4.15: From the slope of a linear fit to PD data, the conversion factor for volts to nanometers is calculated. The PD conversion factor depends on the bead size and needs to be calibrated separately for each trapped bead. For this particular curve the conversion factor is  $-7.75 \times 10^{-5}$  volts/nm.

sizes ( $\approx 5\%$  according to the manufacturer) and the conversion value depends on the bead size, the PD response must be calibrated independently for each bead.

### Trap Stiffness

The optical trap can be described as a Hookean spring for small displacements of the bead from its equilibrium position. A trapped microsphere displaced  $\Delta x$  from equilibrium experiences a restoring force  $F$  given by

$$F = -\kappa \Delta x, \quad (4.3)$$

where  $\kappa$  is the trap stiffness. In experiments, it is determined by power spectral analysis using the PSD. The Brownian motion of the trapped bead is measured. Its power spectral density  $P(f)$  obtained from the Fourier transform squared can be described by a Lorentzian function [114]:

$$P(f) = \frac{D}{2\pi^2(f_c^2 + f^2)}. \quad (4.4)$$

Here  $D$  is the bead's diffusion constant and  $f$  is the frequency. From the fitting parameter of corner frequency  $f_c = \kappa/2\pi\gamma$ , the trap stiffness  $\kappa$  is calculated.  $\gamma$  is the drag coefficient of the trapped bead given by

$$\gamma = 6\pi\eta R, \quad (4.5)$$

where  $\eta$  is the solvent viscosity and  $R$  is the nominal bead radius.  $D$  is taken to be  $D = \frac{k_B T}{\gamma}$  in the fitting. The corner frequency of the power spectrum depends on the size of the trapped bead, laser power and optical alignment.

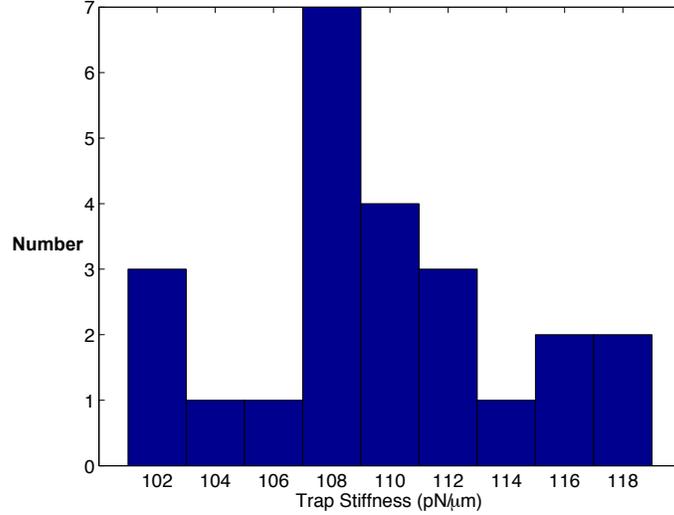


Figure 4.16: Histogram of trap stiffnesses determined for 24 polystyrene beads with nominal 2.1  $\mu\text{m}$  diameter, and trapped by an identical laser power.

In optical tweezers experiments, bead positions are sampled at  $N$  discrete times, with  $\Delta t = 1/f_{\text{sample}}$ , which results in an aliased power spectrum. For discrete sampling, the aliased power spectral density is [115],

$$P_f = \frac{D(1 - c^2)\Delta t}{2\pi f_c (1 + c^2 - 2c \cos(2\pi f/N))}, \quad (4.6)$$

where  $c = \exp(-2\pi f_c \Delta t)$ .

For measurement of fluctuations of the trapped bead, the Nyquist frequency (half of the sampling frequency) should be well above the corner frequency. For 2.1  $\mu\text{m}$  diameter beads, typical corner frequencies range between 810 and 960 Hz, corresponding to trap stiffnesses of 100 and 120  $\text{pN}/\mu\text{m}$  (Figure 4.16). In these trap calibration experiments, the position signal was sampled at  $f_{\text{sample}} = 100$  kHz,  $\Delta t = 10^{-5}$  seconds, for 100 seconds. The data was then divided into 100 independent sets with  $N = 10^5$  data points in each set. Power spectral density analysis was performed on each set and then averaged over the 100 sets (Figure 4.17).

#### 4.4 Worm-Like Chain (WLC) Model

The force-dependent stretching response of a polymer reveals valuable information about its mechanical properties. When there is no external force, the end-to-end separation of a long polymer ( $L/l_p \gg 1$ ) is significantly shorter than its contour length, because of thermal fluctuations of the chain. Larger end-to-end separations decrease the accessible spectrum of fluctuations, creating an unfavourable state in which entropy has decreased. Thus, the molecule tends to remain at shorter end-to-end separations, and an applied force is required to increase its extension. Depending on  $L/l_p$ , this behaviour, called entropic elasticity, the low-force response of long polymers to an ap-

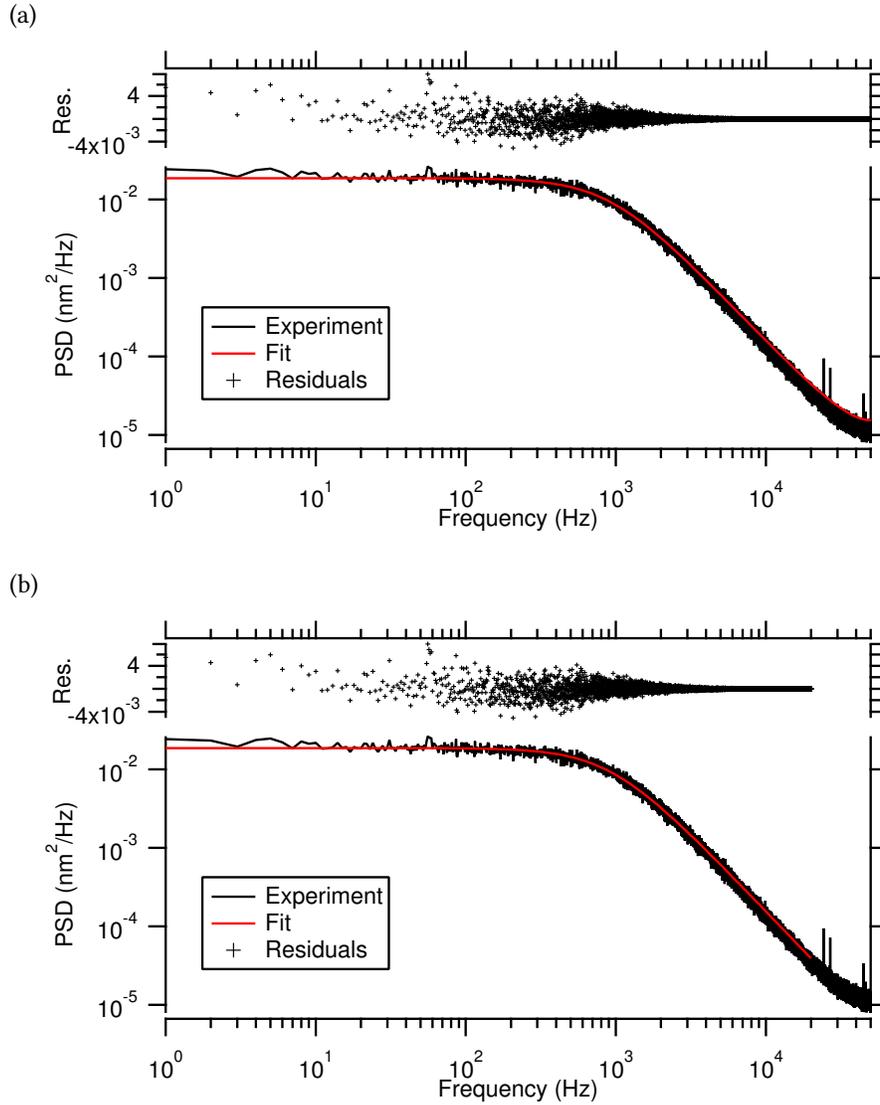


Figure 4.17: Power spectral density (PSD) of a 2.1  $\mu\text{m}$  trapped polystyrene bead with the fit (bottom) and the residual plot (top). (a) Aliased Lorentzian fit (Equation 4.6), and (b) Lorentzian fit (Equation 4.4). Corner frequency is (a) 913 Hz, and (b) 915 Hz, resulting in similar values of 114  $\text{pN}/\mu\text{m}$  for the trap stiffness from both fits.

plied force (usually up to  $\approx 10$  pN) and has been found to represent the response of a variety of biopolymers including double-stranded DNA [45, 46] and perhaps collagen [11, 44].

There are several models describing the entropic elasticity of polymers, of which the worm-like chain (WLC) model has been applied to interpreting collagen's force-extension behaviour [11, 44]. Its predictions are outlined below.

The entropic force can be estimated from the entropic Hamiltonian of an inextensible semi-flexible polymer. When the polymer is stretched to an end-to-end extension  $x$  with external force  $F$ , the Hamiltonian is given by [45]

$$H = \frac{k_B T}{2} \int_0^L l_p \left( \frac{\partial^2 \vec{r}(s)}{\partial s^2} \right)^2 ds - xF. \quad (4.7)$$

As defined before, the persistence length,  $l_p$ , is the length along the polymer over which the tangent vector  $\hat{t}$  correlations drop to  $e^{-1}$ :

$$\langle \hat{t}(s) \cdot \hat{t}(0) \rangle = \langle \cos \theta(s) \rangle = \exp \left( \frac{-|s|}{l_p} \right). \quad (4.8)$$

The difference between this equation and equation 2.6 is arises from the 3D conformational space here.

An interpolation formula called the inextensible worm-like chain approximates the response of the polymer to the applied force [45]:

$$F(x) = \frac{k_B T}{l_p} \left[ \frac{1}{4(1-x/L)^2} - \frac{1}{4} + \frac{x}{L} \right]. \quad (4.9)$$

At higher forces ( $>10$  pN for *dsDNA*), the elastic response comprises both to entropic and enthalpic contributions. The chain extensibility can be taken into account by introducing an enthalpic term to the Hamiltonian in Equation 4.7. At high forces, where the polymer is significantly extended, the chain behaviour can be approximated by the following [116]:

$$x = L \left( 1 - \frac{1}{2} \sqrt{\frac{k_B T}{F l_p} + \frac{F}{K}} \right). \quad (4.10)$$

Here,  $K$  is the stretch modulus of the polymer.

## 4.5 Stretching 11.7 kbp (4 $\mu$ m) *dsDNA* Molecules

Because DNA and its mechanical properties have been widely studied, DNA molecules are used to calibrate our measurements and stretching experiments. Double stranded DNA molecules of length 11.7 kbp ( $\approx 4$   $\mu$ m) were tethered between two polystyrene microspheres of 1.2  $\mu$ m and 2.1  $\mu$ m in diameter [117]. DNA was attached to the larger optically trapped bead via an antibody-

antigen (anti digoxigenin – digoxigenin) interaction. The other end of the DNA was labeled with biotin, which enables binding to a streptavidin-coated smaller bead, immobilized on the micropipette (Figure 4.4). By moving the sample chamber via the piezoelectric stage, the tethered molecule was stretched and subsequently relaxed. Note that here, the contour length of the DNA is long compared to the sizes of the two beads.

An example force-extension curve of DNA is shown in Figure 4.18. The curve can be divided into three regions [104]:

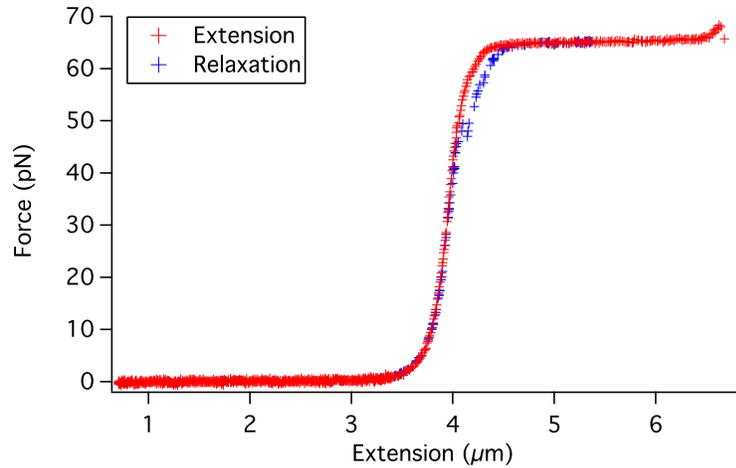


Figure 4.18: Force-extension curve of extension and relaxation of a *dsDNA* measured in our OT instrument.

1. Entropic elasticity regime: Thermal fluctuations lead to chain bending, which shortens the end-to-end separation of the molecule below its contour length. A force is needed to increase extension and decrease the chain's configurational entropy. The inextensible Worm-Like Chain model (Equation 4.9) successfully describes the entropic elasticity of *dsDNA* at forces up to 10 pN [104].
2. Intrinsic elasticity regime: At forces above 10 pN, the end-to-end distance becomes larger than its theoretical B-form <sup>1</sup> contour length [47]. This arises due to the elasticity of the chain. In this regime, the molecule behaves like a stretchable chain and its contour length increases linearly with applied force. Response of the molecule to force is no longer purely entropic and the inclusion of a stretch modulus in the model is required to describe the behaviour of the chain [116]. The extensible WLC (Equation 4.10) fits the response of *dsDNA* to mechanical forces below around 35 pN [118, 119].
3. Overstretching transition: At a force of  $\approx 65$  pN, torsionally relaxed DNA undergoes structural changes and its contour length increases by 70% [47, 118]. While molecular details of

---

<sup>1</sup> $\approx 0.34$  nm per bp

this overstretching transition are complex, sequence-dependent unstacking of DNA and/or melting of basepairs is expected [48, 118].

The inextensible WLC fit in the entropic elasticity region is shown in Figure 4.19, giving a persistence length of 50 nm, in agreement with the established value for *dsDNA* in the literature [46, 119].

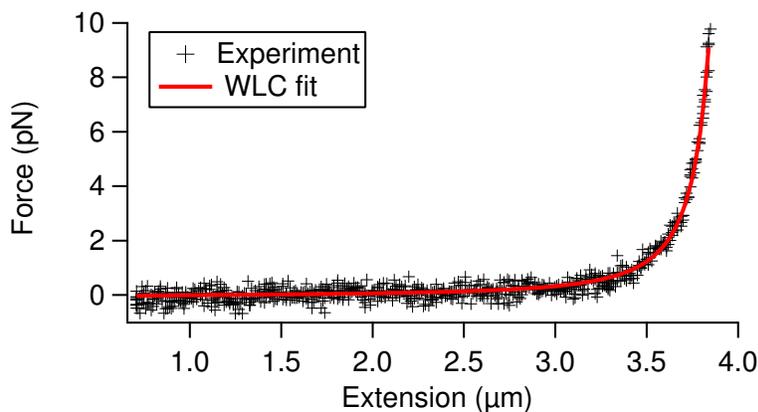


Figure 4.19: Stretching curve of a 11.7 kbp *dsDNA* with a WLC fit of entropic elasticity gives a persistence length of 50 nm to the low force data ( $< 10\text{pN}$ ) [45, 46].

## 4.6 Stretching Single Collagen Molecules

Unlike DNA, collagen's mechanical properties are understudied and unresolved (e.g. Table 1.2). Biological roles of collagen are closely related to its mechanical properties and how it responds to force. These properties are known at larger scales, but remain relatively unknown at the molecular level. Knowledge of collagen behaviour at the microscopic scale can reveal the underlying mechanisms for behaviour of collagen tissues and networks. It can lead to better understanding of collagen-related diseases and better design of artificial tissues. Similar to DNA, collagen has a helical structure. By performing optical tweezers experiments, we seek evidence of a force-induced structural transition, akin to the *dsDNA* overstretch transition [47].

The following chapters present my published results of attempts to measure the flexibility of collagen via optical tweezers.

## Chapter 5

# Studying Mechanical Properties of Procollagen Molecules Using Optical Tweezers

In this chapter, I investigate stretching molecules of type II procollagen using optical tweezers. (See Figure 1.1 for the distinction between procollagen and collagen.) This work has been published in [100, 101]. This study utilized an expression system for human type II procollagen, which was developed in the Forde lab and has been described in [101].

### 5.1 Materials and Methods

The experiments were conducted using the single-beam optical tweezers instrument, described in the previous chapter.

#### 5.1.1 Microsphere and Procollagen Preparation

Microspheres were purchased from commercial sources and modified as necessary. I conjugated streptavidin (Molecular Probes) to COOH-terminated 1.27  $\mu\text{m}$  diameter polystyrene microspheres (Spherotech) as described [117]. Antibodies specific to the N-propeptide of human type II collagen, "anti-N" (003-02, Abcam), were bound to protein G beads (PGP-20-5, 2.1  $\mu\text{m}$  diameter polystyrene, Spherotech) via affinity interactions. To stretch single molecules of type II procollagen, cysteines present only in its globular propeptide ends were reduced using beta-mercaptoethanol (Bioshop), and were covalently linked to biotin using maleimide-biotin (EZ-Link Maleimide-PEG2-Biotin, Thermo Scientific). Successful biotinylation was verified by Western blotting with streptavidin.

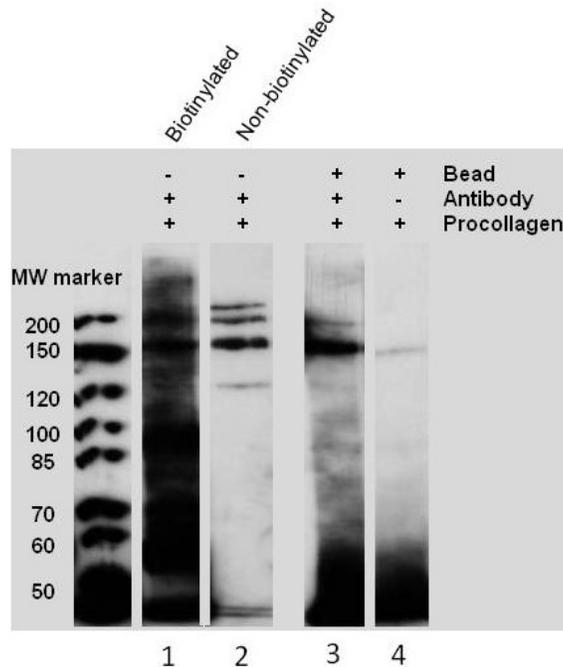


Figure 5.1: Binding of type II procollagen to anti-N labeled beads is monitored using a 6% polyacrylamide gel and Western blotting using streptavidin-HRP. The high-MW bands in lane #3 indicate binding of procollagen to antibody-coated beads, with minimal contributions from nonspecific binding (lane #4, no antibody present on protein G beads). Beads were blocked by BSA prior to incubating with procollagen to reduce nonspecific interactions.

### 5.1.2 Specifically Tethering Procollagen Between Microspheres

The optical tweezers stretching experiment shown schematically in the inset to Figure 5.2 requires the specific binding of procollagen by its ends to microspheres, such that the attachment points in the protein and contour length between them are well defined. To achieve this, I first used anti-N beads to specifically bind the N-propeptide of type II procollagen. After trapping one of these spheres in the optical trap, I then used a streptavidin-coated bead held on the micropipette to link to the biotinylated C-propeptide of procollagen. I verified that the anti-N antibodies can bind to biotinylated type II procollagen by immunoprecipitation using anti-N beads. Figure 5.1 shows that the anti-N beads capture and bind a much greater population of biotinylated procollagen than do control beads lacking the antibody specific to procollagen. These results demonstrate successful binding of type II procollagen specifically to microspheres. The small amount of nonspecific binding observed with unlabeled beads suggests that only a small fraction of the tethers we stretch in the optical tweezers experiments may arise from nonspecifically bound procollagen.

### 5.1.3 Force Measurements and Analysis

Collagen stretching experiments were performed to obtain force-extension curves. Tethered proteins were stretched in PBS buffer pH 7.4. Laser deflection was used to determine the position of

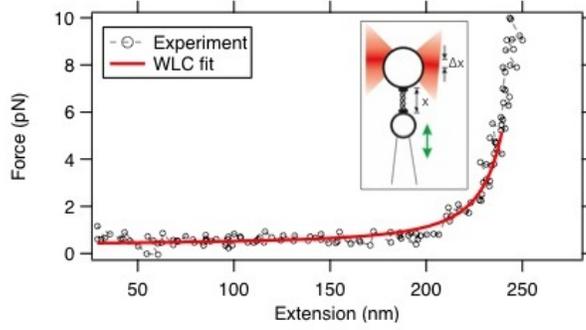


Figure 5.2: Optical tweezers stretching curve of type II procollagen described at low force by entropic elasticity. The Worm-Like Chain (WLC) model (red; equation (5.1)) is fit to an example force-extension curve (black dots), giving a persistence length of 32 nm for a molecule of 300 nm contour length, when a maximum force of 5 pN is used for the fit. (The fit is plotted to 5 pN.) Inset: a schematic showing procollagen stretching in the optical tweezers and illustrating the extension  $x$  and bead offset from trap  $\Delta x$ , from which force is determined (not to scale).

the trapped bead, while the pipette bead's relative position was given by a read-out of the position of the piezo stage (Section 4.3.1). These 1 kHz bandwidth data were low-pass filtered using a boxcar average to produce data of approximately 60 Hz bandwidth. Forces on the trapped microsphere were determined using the trap stiffness and bead offset from equilibrium (Section 4.3.2). Molecular extensions were determined from the separation between microspheres, and included an additional distance offset term when fitting the data to account for potential off-axis tethering. The force-extension curves are fitted using the inextensible worm-like-chain model (Equation 4.10) from which collagen's persistence length is extracted. Because the positions of the beads are known only relatively and the exact binding point on the pipette bead is unknown, a length offset parameter,  $x_0$ , is added to equation (4.10):

$$F(x) = \frac{k_B T}{l_p} \left[ \frac{1}{4 \left(1 - \frac{x-x_0}{L}\right)^2} - \frac{1}{4} + \frac{x-x_0}{L} \right]. \quad (5.1)$$

## 5.2 Results and Discussions

Single procollagen molecules were tethered between two microspheres in the optical tweezers instrument, as described above. Figure 5.2 shows a sample force-extension curve of this protein. The low-force part of this curve ( $F < 5$  pN) is well described as an inextensible worm-like chain, as shown by the good agreement between the experimental data and the model (Equation 5.1).

Preliminary findings suggest that the persistence lengths found from many of the fits are longer than those obtained in previous single-molecule optical tweezers experiments on type II procollagen [11], and are in closer agreement with larger values found using other techniques [15, 16, 20]. I find that the values of persistence length obtained from the fits depend strongly on the force range used for the fit (Figure 5.3). The results are filtered to analyse only curves known to contain a sin-

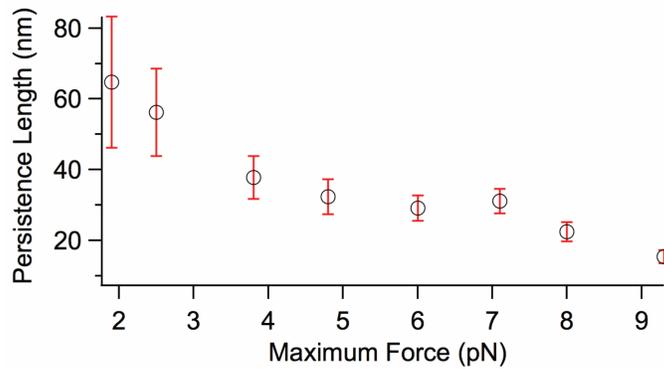


Figure 5.3: The persistence length from fitting the WLC model to the data in Figure 5.2 decreases as the maximum force used in the fitting increases. The error bars show the uncertainty of the fitting parameter.

gle molecule tether between particles (as demonstrated by a single tether rupture event). I found additionally that relaxation curves are easier to analyse than extension curves, as the latter appear to contain multiple tethers acquired from having approached the two particles to close separations. Among several procollagen molecules stretched, five of them were single tethers which contained low-force data and were analyzed. The dependence of persistence length on the maximum force used in the WLC fit was observed for all of them.

The binding of collagen to beads through antibodies is not strong enough to reach high force regimes and tethers break before reaching higher forces. In addition, the presence of propeptides in the procollagen molecules could contribute to the flexibility of the molecules being stretched in the optical tweezers. These globular regions at the ends of the triple helix provide chemically accessible "handles" for linking to beads, yet their contributions to the mechanical response measured for this relatively short polymer (with respect to the micron-sized beads) may not be negligible. The next chapter is designed to address these concerns, by stretching collagen (with its globular propeptides removed) rather than procollagen, covalently bound to microspheres, and comparing the results with previous measurements on procollagen.

## Chapter 6

# Stretching Type III Collagen with Optical Tweezers

In this chapter, I investigate mechanical properties of type III collagen molecules by stretching them with optical tweezers. This work has been published in part in [102]. Results of elasticity of collagen type III beyond the low-force entropic regime are discussed.

### 6.1 Materials and Methods

#### 6.1.1 Covalent Binding of Collagen Molecules to Microspheres

Human type III collagen (Fibrogen, San Francisco) was treated to introduce biotin moieties on its N-terminus and to utilize the cysteines present in the C-terminus to link covalently to derivatized beads. Briefly, collagen was treated with PLP (Sigma) to introduce N-terminal aldehydes, which were subsequently reacted with biotin hydrazide (Sigma) [51]. Western blots using streptavidin alkaline phosphatase confirmed the presence of covalently linked biotin. Cysteines at the C-terminus were reduced to produce sulfhydryl groups, which were then reacted with SPDP (Pierce). This allowed subsequent covalent coupling to amino-terminated polystyrene spheres (2.89  $\mu\text{m}$ , Spherotech).

Single molecules of type III collagen were stretched in PBS buffer pH 7.4. The proteins were tethered between two polystyrene microspheres, linked covalently to the 2.89  $\mu\text{m}$  diameter-sphere held in the optical trap and via N-terminal biotins to a 1.03  $\mu\text{m}$  streptavidin-coated sphere (carboxy-terminated polystyrene, Spherotech, labeled as described [117]) held on a micropipette by suction (Figure 6.1, inset). In the experiments described here, the pipette bead was moved continuously at a rate of 75 nm/sec to stretch collagen. The trap stiffness was obtained from fitting of a power spectrum of the same trapped bead (before tethering to collagen) to a Lorentzian function [114], and was 62 pN/ $\mu\text{m}$  for the measurements reported here. Forces were obtained from the displacement of the trapped bead from its rest position, while extensions were obtained from the relative separation between the two microspheres. Curves were analyzed only if they could be verified to

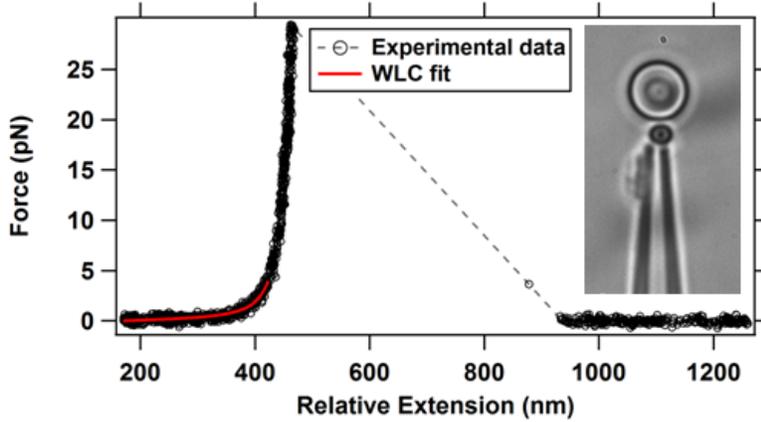


Figure 6.1: Force-extension curve of type III collagen with the WLC fit (red line) over the range used for fitting to the data ( $F \approx 4$  pN). Fit parameters are described in the text. The longest values of relative extension ( $\approx 1000$  nm) correspond to the bead-bead separation after the tether breaks (presumably due to rupture of the biotin-streptavidin bond). The inset shows an image obtained during this stretching experiment, to illustrate the geometry in which the collagen molecule is stretched between an optically trapped bead (top) and a bead held on a movable micropipette (bottom).

be from a single tether, as indicated by a single rupture event bringing the force on the trapped bead to zero.

For the data presented here, the displacement of the trapped bead was obtained from the deflection of the trapping laser beam on the position-sensitive photodiode (DL-10, OSI), which was read at 1 kHz and calibrated based on coincident lower bandwidth video imaging (Flea, Point Grey). The position of the pipette bead was found from the output of the piezo stage sensor (Nano H50, Mad City Labs), also read at 1 kHz. To use decorrelated measurements in our analysis, we used every fourth data point (effective sampling frequency 250 Hz, below the effective corner frequency of our collagen-bead system at all forces), and determined the mean and standard error of the mean of four consecutive data points for use in data fitting. This total sampling time (16 msec) per data point ensures that the pipette bead moves less than  $1/3$  of the expected standard deviation of trapped bead motion within the force range used for analysis.

These position readings reveal the change in each bead's position, but not the absolute distance between them. To determine the extension, these readings were offset using the mean bead radii and the distance between bead centers found by image analysis following tether rupture. Because of uncertainty in the extension (due to bead polydispersity) and zero force (force on the trapped bead following tether rupture), we included in the fitting algorithm an extension offset,  $x_0$  (Equation 5.1) which we constrained to lie within 10% of the sum of bead radii, and a force offset,  $F_0$ , which we constrained to lie within the standard deviation of the force measurements following tether rupture.

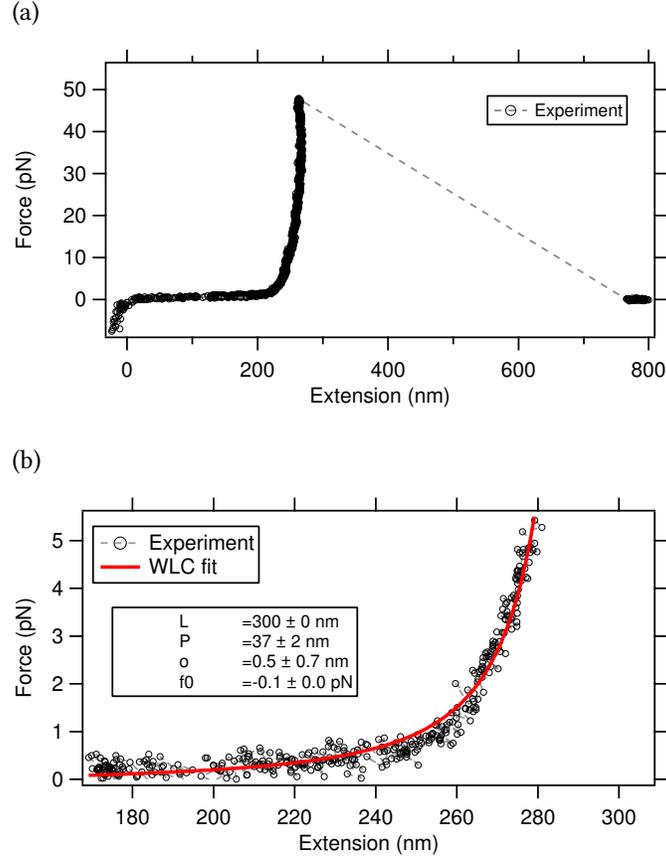


Figure 6.2: (a) Force-extension curve of a single type III collagen, in which the dashed line indicates a single rupture to zero force. (b) Corresponding WLC fit to the low-force region of the curve. For this fit,  $F_{\max} = 5$  pN and the contour length was held constant at 300 nm (expected  $L$  for native collagen). The resulting persistence length is  $37 \pm 2$  nm for this curve. Errors represent 95% confidence intervals for the fit parameters.

## 6.2 Results and Discussions

Examples of force-extension curves for type III collagen are presented in this section. A rupture to zero force was required to indicate that a single molecule was responsible for the observed force-extension behaviour.

Figure 6.1 presents a force-extension curve for a type III collagen molecule fit with the inextensible worm-like chain model (Equation 5.1). Here, the fitting procedure constrained the contour length to the expected 300 nm; this resulted in an offset extension  $x_0 = 169 \pm 0.7$  nm and  $F_0 = -0.01 \pm 0.01$  pN, both within the ranges of allowed extension and force offsets as described above. Errors represent 95% confidence intervals in the fit parameters. This fit gave a persistence length of  $l_p = 11.3 \pm 0.3$  nm ( $\chi_{\text{red}}^2 = 5.2$ ,  $N_{\text{points}} = 596$ ), within the range of previous estimates of the persistence lengths of procollagens type I and II using optical tweezers [11, 44]. Constraining the contour length instead to 400 nm, a length that is possible if the triple helix were completely dena-

tured by these low forces [50, 51], resulted in a fit with a slightly longer persistence length of  $l_p = 16.3 \pm 0.5$  nm. Leaving the contour length as a free fit parameter resulted in a fit with significantly large uncertainties on the resulting length parameters ( $L = 536 \pm 140$  nm;  $x_0 = -60 \pm 140$  nm;  $l_p = 26 \pm 12$  nm).

There is a strong correlation between length fitting parameters, which indicates the importance of determining independently at least one of these length parameters (e.g. the absolute, rather than relative, extension) in order to obtain the others through fitting. The development of measurement and analysis algorithms for stretching this relatively short molecule in optical tweezers will assist in determining statistically meaningful assessments of competing descriptions of its flexibility.

Examples of force-extension curves recorded and analysed for type III collagen are shown in Figures 6.2 and 6.3. The covalent binding strategy allows repeating stretching-relaxation cycles several times. However, at close proximity of the two beads, there is a high chance for formation of multiple tethers, which break at higher forces and eventually lead to a single tether (as signified later by a single rupture event to zero force). About 15 single-molecule curves were obtained and the resulting persistence length values range between 50 nm to 70 nm depending on the fit conditions (*i.e.* fitting parameters and geometric offset). As seen in the previous chapter for procollagen II under force (Figure 5.3), here too I find that the values found for  $l_p$  results decrease with maximum force used in the fit ( $F_{\max}$ ) (data not shown).

According to the AFM studies presented in Chapter 3 of this thesis, collagen's persistence length increases at higher ionic strengths. Although these optical tweezers studies were performed in high ionic strength solutions, the values of persistence length are significantly shorter than those found using AFM. This may be a result of the short length of the collagen tethers compared to the sizes of the microspheres used in optical tweezers. It has been shown that force-extension measurements using optical tweezers to stretch molecules of DNA with  $L < 100l_p$  yield values of persistence lengths that decrease with contour length, by as much as a factor of 2 for  $L \approx 2l_p$  [120, 121]. Thus, the analysis of collagen's force-extension response could yield a low effective persistence length in part due to geometric constraints of this type of experiment. Future directions that could address this issue are proposed in Chapter 7.

The covalent binding strategy implemented in this chapter allows investigation of collagen's force response at higher forces, in order to study collagen beyond the entropic elasticity regime. Unlike the characteristic overstretch transition observed at high force for DNA (Figure 4.18), the experiments here on type III collagen showed no evidence of a cooperative force-induced helix melting up to at least  $F \approx 60$  pN (Figure 6.3). Higher forces than this were applied and similarly showed no evidence of a cooperative structural transition; however, the higher forces could not be exactly calibrated and quantified due to trap nonlinearities and a displacement of the trapped bead along the optical axis.

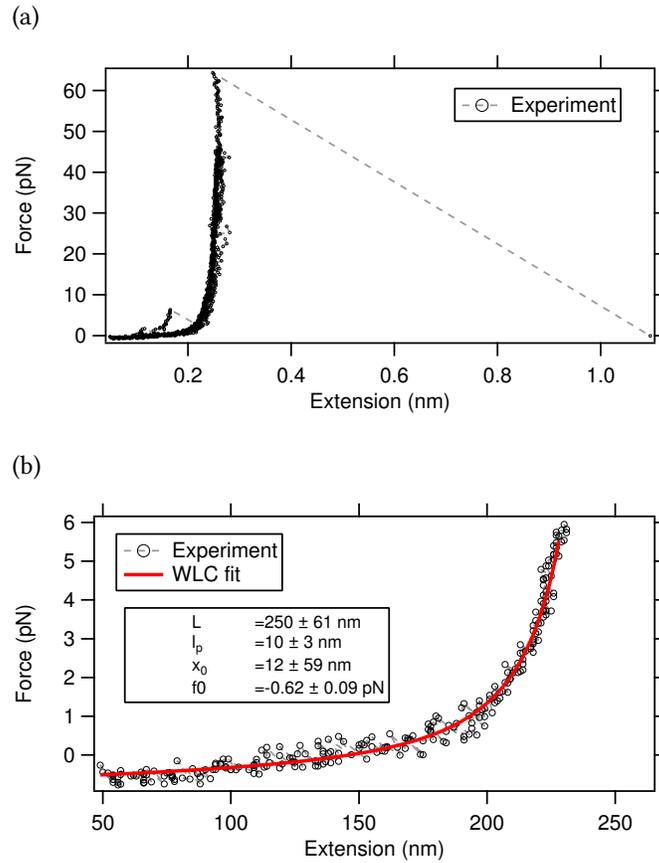


Figure 6.3: (a) Stretch-relaxation curves of type III collagen molecules. At short distances, multiple tethers form, which break at higher forces (dashed lines represent the ruptures) until a single tether remains. A rupture to zero force indicates that it is a single tether. In this example, multiple tethers break at around 8 pN and form a single tether, which breaks to a zero force at around 65 pN after several stretching-relaxation cycles. (b) Corresponding WLC fit to the low-force region of the curve for the tether that precedes the single rupture to zero force. For this fit,  $F_{\max} \approx 6$  pN and the contour length was constrained between 250 and 300 nm (expected  $L$  for native collagen).  $L = 250 \pm 61$  nm,  $l_p = 10 \pm 3$  nm.

## Chapter 7

# Summary and Future Directions

In this thesis, the development and implementation of experimental and analytical tools for nanomechanical studies of collagen molecules have been described.

Molecular conformations of single collagen molecules were studied using Atomic Force Microscopy (AFM) imaging. Persistence lengths of different types of collagen molecules were measured. The effect of salt and pH on molecular conformations of collagen molecules were also studied through AFM imaging. To analyze their conformations, a new analysis routine, SmarTrace, was developed (Chapter 2).

Comparison of AFM images and resulting persistence lengths of the collagen types studied here (type I from rat-tail tendon, recombinant human type I expressed in yeast, human type II from cartilage, recombinant human type III expressed in yeast) shows that they have comparable overall flexibility. Two different solution conditions from which collagen molecules were deposited on mica were presented in this work: 1 mM HCl and 100 mM KCl at pH 3, and 20 mM acetic acid at pH 3.5. The resulting  $l_p$  values for the first solution range between 70 and 90 nm, which fall towards the mid-range of  $l_p$  values found in the literature (Table 1.2). The persistence length values for the latter solution are between 25-35 nm, which is on the shorter range. Previous optical tweezers experiments reported a similar value of persistence length for collagens type I and II [11, 44], which is in line with my results of similar flexibility for type I and II collagens.

Based on my results, the global molecular conformations are mainly affected by the solution from which the molecules are imaged, rather than type of collagen molecules or their biological source.

A fundamental assumption in our analysis is that collagen is a homogeneous polymer, a common assumption when analyzing AFM images of molecules [18, 59, 63]. In this approach all segments of the chains are treated identically and the direction of the molecules are not discriminated. Therefore, neither internal sequence nor the orientation of the chains is taken into account. This approach reveals the global conformation of the chains. In order to study sequence heterogeneity of the chains and its effect on local flexibilities, the ends of the chains need to be distinguished.

In future experiments, imaging of end-labeled chains could be performed to study the sequence-dependent flexibility of collagen.

Among the types of collagen studied here, type I is a heterotrimer and types II and III are homotrimers. To our knowledge, this is the first direct comparison of AFM images of hetero and homo trimeric collagen in their native states. Previous studies compared mechanical properties of healthy hetero and pathogenic homo forms of collagen type I and found a change in local stability of the triple helix [39,40,122]. It has been shown the two forms have different local flexibility [39,40,122], which correlates with altered mechanical properties of the resulting tissue [41]. Local kinks in the homotrimeric collagen are likely to be responsible for the altered mechanical properties of the resulting fibrils [122]. Effect of kinks on the flexibility of collagen can be locally studied in future experiments. Here however, local kinks are averaged out in determining the average persistence length of collagen.

Apart from sequence dependent flexibility, the homo and hetero trimers might have different stabilities in different solutions. A study on engineered collagen peptides found that decreasing pH or increasing salt concentrations can promote homotrimer stability compared to the heterotrimer's [74]. The persistence length results of different types of collagen in 20 mM acetic acid of (pH 3.5) show a slightly higher persistence length for tissue-derived type II collagen, while persistence lengths are more similar among collagen types in 1 mM HCl and 100 mM KCl. A future direction would be measurement of the thermal stability of homo and hetero collagen types (*e.g.* type I heterotrimer and type II homotrimer, in solutions of the same ionic strength but different pH).

I found the flexibility of collagen on mica to be significantly affected by the ionic strength and pH of the solution from which it is deposited on mica. My results show that increasing ionic strength rigidifies collagen's triple helix. In neutral pH, the persistence length of collagen increases from  $\sim 20$  nm to  $\sim 150$  nm as a result of an increase in the solution ionic strength from 0.1 mM to 100 mM. Moreover, for similar ionic strengths, collagen is more flexible in acidic pH. These results contradict the expected behaviour of a flexible polyelectrolyte such as DNA, for which an increase in ionic strength causes an increased flexibility [84]. Decreased intramolecular electrostatic repulsion is the proposed mechanism for the observed DNA mechanics in the presence of increasing salt [84]. By contrast, at neutral pH collagen's increased mechanical stability at higher ionic strengths can be explained by formation of salt bridges between neighbouring residues [28,86,87]. The formation of salt bridges has a stabilizing effect on collagen's triple helix [28], which stiffens the chain resulting in longer persistence lengths. This happens because the increase in ionic strength of solution decreases the Debye length, thereby localizing electrostatic interactions (Equation 3.6).

My results are in agreement with another study in which a higher persistence length was observed for collagen in buffers with higher ionic strengths such as fibril forming buffer (FFB) and phosphate buffered saline (PBS) when compared to collagen in water [18]. My result of increased persistence length of collagen in the presence of salt correlates well with the salt-promoted triple-helical stability of collagen, found using differential scanning calorimetry and circular dichroism spectroscopy [28].

In acidic pH, an increase in ionic strength more weakly increases collagen's persistence length. Being positively charged at the acidic pH, collagen's increase in rigidity at higher ionic strength cannot be attributed to salt bridge formation. One possible mechanism is screening of electrostatic repulsion between positively charged residues on neighbouring  $\alpha$ -chains, thereby stabilizing the triple helix by overcoming local unwinding in the molecule [123].

My results support the hypothesis that conformational stability of collagen, *i.e.* persistence length, is related to tightness of the helix. Increase in the ionic strength has been found to increase thermal stability of collagen [28], which corresponds to increase in persistence length in my experiments. In neutral pH, the results can be explained by formation of salt bridges, which lead to a tightly wound helix. Examination of the distribution of charged amino acid residues along individual  $\alpha$ -chains of collagen shows that they are distributed throughout the helix and supports the results of change in helical stability as a result of change in ionic strength.

AFM imaging of end-labeled collagen chains in future experiments would make it possible to study changes in local flexibilities (attributed to changes in local unwinding) as a result of sequence and its dependence on pH and ionic strength. In these experiments, chain heterogeneity and local flexibility can be determined through studying specific parts of the molecule along its contour.

Collagen is a key load-bearing unit in the body. Its response to applied force can help to elucidate the mechanical properties of this structural protein. We applied force to single collagen molecules in my single-beam optical tweezers instrument (Chapters 4-6) to measure collagen's flexibility and answer a long-standing question of whether collagen molecules undergo structural transitions as a result of an applied force. Before studying the force-induced behaviour of collagen, biochemical assays were developed and applied to attach single molecules of collagen to microspheres that are used as handles to manipulate collagen in the optical tweezers [100–102]. Major efforts have been done to develop techniques and verify the effectiveness of the procedure of binding collagen molecules to microspheres for optical tweezers studies.

Entropic elasticity of procollagen type II was studied by optical tweezers stretching. Binding of molecules to microspheres was done through antibody-antigen interactions. The persistence length results are on the low end of the values and describe collagen as a flexible chain, which are in agreement with results from a previous study by Sun *et al.* [11]. However, in my experiments, the measured persistence length exhibits a force dependence, decreasing as the force range for fitting increases (even for forces less than 10 pN) [101]. This force-dependent apparent persistence length might arise due to microunfolded collagen due to stretching and structural transitions of the chain under force [50].

In these experiments, the binding between the collagen and the antibody was not strong enough to reproducibly reach high forces. Covalent attachment between type III collagen molecules and optically trapped microspheres was used to achieve higher forces [51]. My stretching experiments of collagen type III in optical tweezers showed no evidence of structural transitions at higher forces. No cooperative unwinding such as seen for *dsDNA* was observed up to  $\sim 80$  pN of force (data not shown). The persistence length obtained from fitting the WLC model of entropic

elasticity to low force data [102] results in  $l_p$  values of 11 nm, in agreement with previous optical tweezers studies [11, 44].

Due to the short length of collagen molecules, extracted values of persistence length may not be good characteristic of collagen's flexibility. For one, the results of  $l_p$  from stretching experiments depend strongly on the length offset parameter,  $x_0$  (Sections 6.1.1 and 6.2). The short length of the molecule makes  $l_p$  very sensitive to this offset length parameter, which is not precisely known due to uncertainty in the size of the trapped bead, attachment point on the pipette bead and pulling surface. More fundamentally, it has been shown that for short lengths of tethered molecules, persistence length is underestimated [120, 121]. Another challenge arising due to short length of tethered molecules is formation of multiple tethers. When the two beads are brought in close proximity to tether collagen, there is a high chance that more than one tether forms, because the length of the molecule is short and the beads need to be brought very close. Although diluting the collagen samples to decrease the number of molecules on the bead can decrease the effect, there is still a chance that more tethers are formed. Too low concentrations decreases the experimental throughput.

To address these concerns, a suggested future direction for these optical tweezers studies would be to use DNA handles serially linked to the collagen molecule as spacers, thus overcoming challenges introduced by the short length of collagen. Figure 7.1 shows a schematic setup proposed for manipulating collagen-DNA hybrid systems in optical tweezers. In this setup the DNA spacer would be bound to a polystyrene microsphere via digoxigenin-antidigoxigenin and to the N-terminus of collagen via biotin-streptavidin interaction (Sections 4.5 and 6.1.1). Collagen type III is covalently attached to a derivatized bead through SPDP (succinimidyl 3-(2-pyridyldithio)propionate) bonds formed between sulfide group (-SH) in cysteine residues in its C-terminus and amine (-NH<sub>2</sub>) coated beads (Section 6.1.1). The DNA handle works as a spacer to overcome challenges inherent to short-length tethers such as extension offset effects [102], formation of multiple tethers and underestimation of the persistence length [120, 121]. The force-extension response of DNA molecules is well-characterized and allows characterization of mechanical properties of the collagen stretched as part of the DNA-collagen hybrid system. Moreover, one can make sure it is a single tether if it undergoes the characteristic over-stretching transition at 65 pN of force [47] (Figure 4.18).

In this proposed configuration for the hybrid system one DNA spacer is chosen with attachment strategies already tested in our instrument. Other configurations, such as two DNA handles with a collagen in the middle could also be used. A benefit of this approach is that it minimizes interactions of collagen with the bead surface. Using one DNA handle on the other hand, decreases the number of linkages and their mechanical contributions.

In this interdisciplinary project, the nanomechanical properties of collagen molecules were investigated at the molecular scale, using different analytical and experimental tools. Collagen is an important load-bearing structural protein found in almost every tissue in the body. Its conformations and force response are of direct relevance to its biological functions. By using AFM imaging, molecular conformations of different types of collagen molecules and the effects of environmental

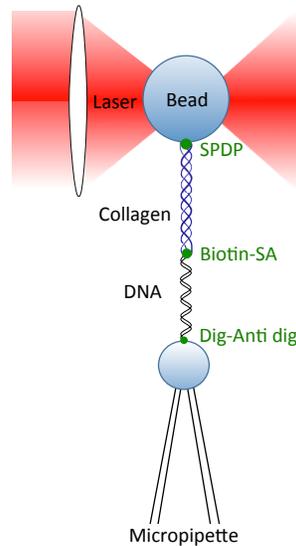


Figure 7.1: Schematic of a collagen-DNA hybrid system (not to scale). The DNA handle is linked to a bead to be held on a micropipette via digoxigenin-anti digoxigenin (Dig-Anti dig) interactions. The other end of the DNA attaches to one end of collagen through a biotin-streptavidin (Biotin-SA) interaction. The collagen molecule is covalently attached to a bead to be optically trapped through SPDP linking an SH group from a C-terminal cysteine and an amine group of a derivatized bead.

factors such as salt and pH were studied. As part of characterizing the nanomechanical properties of collagen, the force response of collagen was studied using optical tweezers. This research on nanomechanical characterization of collagen molecules helps to resolve long-standing questions about the flexibility of this crucial structural protein. Mechanical properties of collagen molecules can affect higher order structures made of this protein [124]. This fundamental knowledge is required to shed light on properties of the extracellular matrix, cancer cell migration and collagen-related diseases, and has potential benefits for drug delivery research as well as tissue engineering. I hope that the results presented in this work will lead to a better understanding of the underlying mechanics of collagen molecules, which could feed into future studies of the extracellular matrix and resulting tissues.

# Bibliography

- [1] MD Shoulders and RT Raines. Collagen structure and stability. *Annual Review of Biochemistry*, 78:929, 2009.
- [2] AV Persikov and B Brodsky. Unstable molecules form stable tissues. *Proceedings of the National Academy of Sciences*, 99:1101–1103, 2002.
- [3] P Fratzl. *Collagen: structure and mechanics, an introduction*. Springer, 2008.
- [4] JF Bateman, RP Boot-Handford, and SR Lamandé. Genetic diseases of connective tissues: cellular and extracellular effects of ECM mutations. *Nature Reviews Genetics*, 10:173–183, 2009.
- [5] J Myllyharju and KI Kivirikko. Collagens and collagen-related diseases. *Annals of Medicine*, 33:7–21, 2001.
- [6] Allen J Bailey, R Gordon Paul, and L Knott. Mechanisms of maturation and ageing of collagen. *Mechanisms of Ageing and Development*, 106:1–56, 1998.
- [7] A Gosain and L A DiPietro. Aging and wound healing. *World Journal of Surgery*, 28:321–326, 2004.
- [8] H Kuivaniemi, G Tromp, and DJ Prockop. Mutations in collagen genes: causes of rare and some common diseases in humans. *The FASEB Journal*, 5:2052–2060, 1991.
- [9] H Kuivaniemi, G Tromp, and DJ Prockop. Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels. *Human Mutation*, 9(4):300–315, 1997.
- [10] A Gautieri, S Vesentini, A Redaelli, and MJ Buehler. Single molecule effects of osteogenesis imperfecta mutations in tropocollagen protein domains. *Protein Science*, 18:161–168, 2009.
- [11] YL Sun, ZP Luo, A Fertala, and KN An. Stretching type II collagen with optical tweezers. *Journal of Biomechanics*, 37:1665–1669, 2004.
- [12] MJ Buehler and SY Wong. Entropic elasticity controls nanomechanics of single tropocollagen molecules. *Biophysical Journal*, 93:37–43, 2007.
- [13] S Vesentini, A Redaelli, and A Gautieri. Nanomechanics of collagen microfibrils. *Muscles, Ligaments and Tendons Journal*, 3:23, 2013.
- [14] K Claire and R Pecora. Translational and rotational dynamics of collagen in dilute solution. *The Journal of Physical Chemistry B*, 101:746–753, 1997.

- [15] H Utiyama, K Sakato, K Ikehara, T Setsuiye, and M Kurata. Flexibility of tropocollagen from sedimentation and viscosity. *Biopolymers*, 12:53–64, 1973.
- [16] T Saito, N Iso, H Mizuno, N Onda, H Yamato, and H Odashima. Semi-flexibility of collagens in solution. *Biopolymers*, 21:715–728, 1982.
- [17] EJ Amis, CJ Carriere, JD Ferry, and A Veis. Effect of pH on collagen flexibility determined from dilute solution viscoelastic measurements. *International Journal of Biological Macromolecules*, 7:130–134, 1985.
- [18] HH Lovelady, S Shashidhara, and WG Matthews. Solvent specific persistence length of molecular type I collagen. *Biopolymers*, 101:329–335, 2014.
- [19] YL Sun, ZP Luo, and KN An. Stretching short biopolymers using optical tweezers. *Biochemical and Biophysical Research Communications*, 286:826–830, 2001.
- [20] 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. *Journal of Molecular Biology*, 172:325–343, 1984.
- [21] S Varma, JPRO Orgel, and JD Schieber. Nanomechanics of type I collagen. *Biophysical Journal*, 111:50–56, 2016.
- [22] EG Canty and KE Kadler. Procollagen trafficking, processing and fibrillogenesis. *Journal of Cell Science*, 118:1341–1353, 2005.
- [23] A Rich and F Crick. The structure of collagen, 1955.
- [24] J Bella, M Eaton, B Brodsky, and HM Berman. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science*, 266:75–81, 1994.
- [25] JP Orgel, TJ Wess, and A Miller. The in situ conformation and axial location of the intermolecular cross-linked non-helical telopeptides of type I collagen. *Structure*, 8:137–142, 2000.
- [26] DJ Prockop and KI Kivirikko. Collagens: molecular biology, diseases, and potentials for therapy. *Annual Review of Biochemistry*, 64:403–434, 1995.
- [27] AV Persikov, JAM Ramshaw, A Kirkpatrick, and B Brodsky. Electrostatic interactions involving lysine make major contributions to collagen triple-helix stability. *Biochemistry*, 44:1414–1422, 2005.
- [28] U Freudenberg, SH Behrens, PB Welzel, M Müller, M Grimmer, K Salchert, T Taeger, K Schmidt, W Pompe, and C Werner. Electrostatic interactions modulate the conformation of collagen I. *Biophysical Journal*, 92:2108–2119, 2007.
- [29] RZ Kramer, J Bella, P Mayville, B Brodsky, and HM Berman. Sequence dependent conformational variations of collagen triple-helical structure. *Nature Structural Biology*, 6:454–457, 1999.
- [30] K Okuyama, C Hongo, G Wu, K Mizuno, K Noguchi, S Ebisuzaki, Y Tanaka, N Nishino, and HP Bächinger. High-resolution structures of collagen-like peptides [(Pro-Pro-Gly) 4-Xaa-Yaa-Gly-(Pro-Pro-Gly) 4]: Implications for triple-helix hydration and Hyp (X) puckering. *Biopolymers*, 91(5):361–372, 2009.

- [31] OP Boryskina, TV Bolbukh, MA Semenov, AI Gasan, and VYa Maleev. Energies of peptide-peptide and peptide-water hydrogen bonds in collagen: Evidences from infrared spectroscopy, quartz piezogravimetry and differential scanning calorimetry. *Journal of Molecular Structure*, 827:1–10, 2007.
- [32] <http://pdb101.rcsb.org/motm/4>.
- [33] E Makareeva, S Han, JC Vera, D L Sackett, K Holmbeck, CL Phillips, R Visse, H Nagase, and S Leikin. Carcinomas contain a matrix metalloproteinase-resistant isoform of type I collagen exerting selective support to invasion. *Cancer Research*, 70:4366–4374, 2010.
- [34] S Minafra, C Luparello, F Rallo, and I Pucci-Minafra. Collagen biosynthesis by a breast carcinoma cell strain and biopsy fragments of the primary tumour. *Cell Biology International Reports*, 12:895–905, 1988.
- [35] M Rojkind, MA Giambrone, and L Biempica. Collagen types in normal and cirrhotic liver. *Gastroenterology*, 76:710–719, 1979.
- [36] CA Miles, T J Sims, NP Camacho, and AJ Bailey. The role of the  $\alpha 2$  chain in the stabilization of the collagen type I heterotrimer: a study of the type I homotrimer in OIM mouse tissues. *Journal of Molecular Biology*, 321:797–805, 2002.
- [37] ME Nimni. *Collagen: Biochemistry, biomechanics, biotechnology*. CRC Press Inc., Boca Raton, FL, 1988.
- [38] X Liu, H Wu, M Byrne, S Krane, and R Jaenisch. Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. *Proceedings of the National Academy of Sciences*, 94:1852–1856, 1997.
- [39] NV Kuznetsova, DJ McBride, and S Leikin. Changes in thermal stability and microunfolded pattern of collagen helix resulting from the loss of  $\alpha 2$  (I) chain in osteogenesis imperfecta murine. *Journal of Molecular Biology*, 331:191–200, 2003.
- [40] S Han, E Makareeva, NV Kuznetsova, AM DeRidder, MB Sutter, W Losert, CL Phillips, R Visse, H Nagase, and S Leikin. Molecular mechanism of type I collagen homotrimer resistance to mammalian collagenases. *Journal of Biological Chemistry*, 285:22276–22281, 2010.
- [41] SW Chang, SJ Shefelbine, and M J Buehler. Structural and mechanical differences between collagen homo- and heterotrimers: relevance for the molecular origin of brittle bone disease. *Biophysical Journal*, 102:640–648, 2012.
- [42] MG Venugopal, JAM Ramshaw, E Braswell, D Zhu, and B Brodsky. Electrostatic interactions in collagen-like triple-helical peptides. *Biochemistry*, 33:7948–7956, 1994.
- [43] EP Katz and CW David. Energetics of intrachain salt-linkage formation in collagen. *Biopolymers*, 29:791–798, 1990.
- [44] YL Sun, ZP Luo, A Fertala, and KN An. Direct quantification of the flexibility of type I collagen monomer. *Biochemical and Biophysical Research Communications*, 295:382–386, 2002.

- [45] JF Marko and ED Siggia. Stretching DNA. *Macromolecules*, 28:8759–8770, 1995.
- [46] C Bustamante, JF Marko, ED Siggia, and S Smith. Entropic Elasticity of  $\lambda$ -Phage DNA. *Science*, 265:1599–1600, 1994.
- [47] SB Smith, Y Cui, and C Bustamante. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, 271:795–799, 1996.
- [48] J van Mameren, P Gross, G Farge, P Hooijman, M Modesti, M Falkenberg, GJL Wuite, and EJG Peterman. Unraveling the structure of DNA during overstretching by using multi-color, single-molecule fluorescence imaging. *Proceedings of the National Academy of Sciences*, 106:18231–18236, 2009.
- [49] A Gautieri, MJ Buehler, and A Redaelli. Deformation rate controls elasticity and unfolding pathway of single tropocollagen molecules. *Journal of the Mechanical Behavior of Biomedical Materials*, 2:130–137, 2009.
- [50] AS Adhikari, J Chai, and AR Dunn. Mechanical load induces a 100-fold increase in the rate of collagen proteolysis by MMP-1. *Journal of the American Chemical Society*, 133:1686–1689, 2011.
- [51] AS Adhikari, E Glassey, and AR Dunn. Conformational dynamics accompanying the proteolytic degradation of trimeric collagen I by collagenases. *Journal of the American Chemical Society*, 134:13259–13265, 2012.
- [52] RJ Camp, M Liles, J Beale, N Saeidi, BP Flynn, E Moore, SK Murthy, and JW Ruberti. Molecular mechanochemistry: low force switch slows enzymatic cleavage of human type I collagen monomer. *Journal of the American Chemical Society*, 133:4073–4078, 2011.
- [53] G Binnig, CF Quate, and C Gerber. Atomic force microscope. *Physical Review Letters*, 56:930, 1986.
- [54] A Engel, Y Lyubchenko, and D Müller. Atomic force microscopy: a powerful tool to observe biomolecules at work. *Trends in Cell Biology*, 9:77–80, 1999.
- [55] Y Sugimoto, P Pou, M Abe, P Jelinek, R Pérez, S Morita, and O Custance. Chemical identification of individual surface atoms by atomic force microscopy. *Nature*, 446:64–67, 2007.
- [56] N Jalili and K Laxminarayana. A review of atomic force microscopy imaging systems: application to molecular metrology and biological sciences. *Mechatronics*, 14:907–945, 2004.
- [57] L Wilson, PT Matsudaira, BP Jena, and JKH Horber. *Atomic force microscopy in cell biology*, volume 68. Academic Press, 2002.
- [58] TE Fisher, PE Marszalek, and JM Fernandez. Stretching single molecules into novel conformations using the atomic force microscope. *Nature Structural & Molecular Biology*, 7:719–724, 2000.
- [59] C Rivetti, M Guthold, and C Bustamante. Scanning force microscopy of DNA deposited onto mica: Equilibration versus kinetic trapping studied by statistical polymer chain analysis. *Journal of Molecular Biology*, 264:919–932, 1996.

- [60] JL Hutter and J Bechhoefer. Calibration of atomic-force microscope tips. *Review of Scientific Instruments*, 64:1868–1873, 1993.
- [61] D Qu  net, EK Dimitriadis, and Y Dalal. *Atomic force microscopy of chromatin*. INTECH Open Access Publisher, 2012.
- [62] G Lamour, JB Kirkegaard, H Li, TPJ Knowles, and J Gsponer. Easyworm: an open-source software tool to determine the mechanical properties of worm-like chains. *Source Code for Biology and Medicine*, 9:1, 2014.
- [63] FGA Faas, B Rieger, LJ van Vliet, and DI Cherny. DNA deformations near charged surfaces: electron and atomic force microscopy views. *Biophysical Journal*, 97:1148–1157, 2009.
- [64] I Usov and R Mezzenga. FiberApp: an open-source software for tracking and analyzing polymers, filaments, biomacromolecules, and fibrous objects. *Macromolecules*, 48:1269–1280, 2015.
- [65] M Sonka, V Hlavac, and R Boyle. *Image processing, analysis, and machine vision*. Cengage Learning, 2014.
- [66] C Bustamante, C Rivetti, and DJ Keller. Scanning force microscopy under aqueous solutions. *Current Opinion in Structural Biology*, 7(5):709–716, 1997.
- [67] LD Landau and EM Lifshitz. Statistical Physics, part I. *Course of Theoretical Physics*, 5:468, 1980.
- [68] PA Wiggins, T van der Heijden, F Moreno-Herrero, A Spakowitz, R Phillips, J Widom, C Dekker, and PC Nelson. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature Nanotechnology*, 1:137–141, 2006.
- [69] S Jordens, EE Riley, I Usov, L Isa, PD Olmsted, and R Mezzenga. Adsorption at liquid interfaces induces amyloid fibril bending and ring formation. *ACS Nano*, 8:11071–11079, 2014.
- [70] K El Kirat, I Burton, V Dupres, and YF Dufrene. Sample preparation procedures for biological atomic force microscopy. *Journal of Microscopy*, 218:199–207, 2005.
- [71] D Pastr  , O Pi  trement, S Fusil, F Landousy, J Jeusset, MO David, L Hamon, E Le Cam, and A Zozime. Adsorption of DNA to mica mediated by divalent counterions: a theoretical and experimental study. *Biophysical Journal*, 85:2507–2518, 2003.
- [72] Surya K Ghosh, K Singh, and A Sain. Effect of intrinsic curvature on semiflexible polymers. *Physical Review E*, 80:051904, 2009.
- [73] A Craig and EM Terentjev. Auxiliary field theory of polymers with intrinsic curvature. *Macromolecules*, 39:4557–4565, 2006.
- [74] AS Parmar, M Joshi, PL Nosker, NF Hasan, and V Nanda. Control of collagen stability and heterotrimer specificity through repulsive electrostatic interactions. *Biomolecules*, 3:986–996, 2013.
- [75] L Kukkola, R Hieta, K I Kivirikko, and J Myllyharju. Identification and characterization of a third human, rat, and mouse collagen prolyl 4-hydroxylase isoenzyme. *Journal of Biological Chemistry*, 278:47685–47693, 2003.

- [76] A Lamberg, T Helaakoski, J Myllyharju, S Peltonen, H Notbohm, T Pihlajaniemi, and KI Kivirikko. Characterization of human type III collagen expressed in a baculovirus system production of a protein with a stable triple helix requires coexpression with the two types of recombinant prolyl 4-hydroxylase subunit. *Journal of Biological Chemistry*, 271:11988–11995, 1996.
- [77] J Myllyharju, M Nokelainen, A Vuorela, and KI Kivirikko. Expression of recombinant human type I-III collagens in the yeast *Pichia pastoris*. *Biochemical Society Transactions*, 28:353–357, 2000.
- [78] M Yamauchi and M Sricholpech. Lysine post-translational modifications of collagen. *Essays in Biochemistry*, 52:113–133, 2012.
- [79] E Pokidysheva, KD Zientek, Y Ishikawa, K Mizuno, JA Vranka, NT Montgomery, DR Keene, T Kawaguchi, K Okuyama, and HP Bächinger. Posttranslational modifications in type I collagen from different tissues extracted from wild type and prolyl 3-hydroxylase 1 null mice. *Journal of Biological Chemistry*, 288:24742–24752, 2013.
- [80] MT Bayliss, FPJG Lafeber, A Maroudas, and JM Tekoppele. Ageing and zonal variation in post-translational modification of collagen in normal human articular cartilage: the age-related increase in non-enzymatic glycation affects biomechanical properties of cartilage. *Biochemical Journal*, 330:345–351, 1998.
- [81] WA Cabral, I Perdivara, MA Weis, M Terajima, AR Blissett, W Chang, JE Perosky, EN Makareeva, EL Mertz, and S Leikin. Abnormal type I collagen post-translational modification and crosslinking in a cyclophilin B KO mouse model of recessive osteogenesis imperfecta. *PLoS Genet*, 10:e1004465, 2014.
- [82] CL Jenkins, LE Bretscher, IA Guzei, and RT Raines. Effect of 3-hydroxyproline residues on collagen stability. *Journal of the American Chemical Society*, 125:6422–6427, 2003.
- [83] VC Chan, JAM Ramshaw, A Kirkpatrick, K Beck, and B Brodsky. Positional preferences of ionizable residues in Gly-X-Y triplets of the collagen triple-helix. *Journal of Biological Chemistry*, 272:31441–31446, 1997.
- [84] CG Baumann, SB Smith, VA Bloomfield, and C Bustamante. Ionic effects on the elasticity of single DNA molecules. *Proceedings of the National Academy of Sciences*, 94:6185–6190, 1997.
- [85] S Kumar and R Nussinov. Close-range electrostatic interactions in proteins. *ChemBioChem*, 3:604–617, 2002.
- [86] T Gurry, PS Nerenberg, and CM Stultz. The contribution of interchain salt bridges to triple-helical stability in collagen. *Biophysical Journal*, 98:2634–2643, 2010.
- [87] S Kumar and R Nussinov. Relationship between ion pair geometries and electrostatic strengths in proteins. *Biophysical Journal*, 83:1595–1612, 2002.
- [88] T Hayashi and Y Nagai. Effect of pH on the stability of collagen molecule in solution. *The Journal of Biochemistry*, 73:999–1006, 1973.
- [89] DE Mechling and HP Bächinger. The collagen-like peptide (GER) 15GPCCG forms pH-dependent covalently linked triple helical trimers. *Journal of Biological Chemistry*, 275:14532–14536, 2000.

- [90] L Chen, S Cai, J Lim, SS Lee, and SG Lee. Elucidating pH-Dependent Collagen Triple Helix Formation through Interstrand Hydroxyproline-Glutamic Acid Interactions. *ChemBioChem*, 16:407–410, 2015.
- [91] A Weinstock, PC King, and RE Wuthier. The ion-binding characteristics of reconstituted collagen. *Biochemical Journal*, 102:983–988, 1967.
- [92] AE Russell. Effect of pH on thermal stability of collagen in the dispersed and aggregated states. *Biochemical Journal*, 139:277–280, 1974.
- [93] Protein calculator v3.4. <http://protcalc.sourceforge.net/cgi-bin/protcalc>, Date accessed: September 2016.
- [94] Y Li, A Asadi, MR Monroe, and EP Douglas. pH effects on collagen fibrillogenesis in vitro: Electrostatic interactions and phosphate binding. *Materials Science and Engineering: C*, 29:1643–1649, 2009.
- [95] TR Arnett. Extracellular pH regulates bone cell function. *The Journal of Nutrition*, 138:415S–418S, 2008.
- [96] F Jiang, H Hörber, J Howard, and DJ Müller. Assembly of collagen into microribbons: effects of pH and electrolytes. *Journal of Structural Biology*, 148:268–278, 2004.
- [97] A Ashkin. Acceleration and trapping of particles by radiation pressure. *Physical Review Letters*, 24:156, 1970.
- [98] A Ashkin, JM Dziedzic, JE Bjorkholm, and S Chu. Observation of a single-beam gradient force optical trap for dielectric particles. *Optics Letters*, 11:288–290, 1986.
- [99] JR Moffitt, YR Chemla, SB Smith, and C Bustamante. Recent advances in optical tweezers. *Biochemistry*, 77:205, 2008.
- [100] N Rezaei, BPB Downing, A Wieczorek, CKY Chan, RL Welch, and NR Forde. Using optical tweezers to study mechanical properties of collagen. In *Photonics North 2011*, pages 80070K–80070K. International Society for Optics and Photonics, 2011.
- [101] A Wieczorek, N Rezaei, CK Chan, C Xu, P Panwar, D Brömme, and NR Forde. Development and characterization of a eukaryotic expression system for human type II procollagen. *BMC Biotechnology*, 15:1, 2015.
- [102] M Shayegan, N Rezaei, NH Lam, T Altindal, A Wieczorek, and NR Forde. Probing multi-scale mechanics of collagen with optical tweezers. In *SPIE NanoScience+ Engineering*, pages 88101P–88101P. International Society for Optics and Photonics, 2013.
- [103] KC Neuman and A Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature Methods*, 5:491–505, 2008.
- [104] C Bustamante, SB Smith, J Liphardt, and D Smith. Single-molecule studies of DNA mechanics. *Current Opinion in Structural Biology*, 10:279–285, 2000.
- [105] C Baldock, AF Oberhauser, L Ma, D Lammie, V Siegler, SM Mithieux, Y Tu, JYH Chow, F Suleman, and M Malfois. Shape of tropoelastin, the highly extensible protein that controls human tissue elasticity. *Proceedings of the National Academy of Sciences*, 108:4322–4327, 2011.

- [106] ASM Kamruzzahan, A Ebner, L Wildling, F Kienberger, CK Riener, CD Hahn, PD Pollheimer, P Winklehner, M Hölzl, and B Lackner. Antibody linking to atomic force microscope tips via disulfide bond formation. *Bioconjugate Chemistry*, 17:1473–1481, 2006.
- [107] P Hinterdorfer and YF Dufrêne. Detection and localization of single molecular recognition events using atomic force microscopy. *Nature Methods*, 3:347–355, 2006.
- [108] T Strick, JF Allemand, V Croquette, and D Bensimon. Twisting and stretching single DNA molecules. *Progress in Biophysics and Molecular Biology*, 74:115–140, 2000.
- [109] TR Strick, V Croquette, and D Bensimon. Single-molecule analysis of DNA uncoiling by a type II topoisomerase. *Nature*, 404:901–904, 2000.
- [110] KC Neuman and SM Block. Optical trapping. *Review of Scientific Instruments*, 75:2787–2809, 2004.
- [111] AAR Neves, A Fontes, LY Pozzo, AA De Thomaz, E Chillce, E Rodriguez, LC Barbosa, and CL Cesar. Electromagnetic forces for an arbitrary optical trapping of a spherical dielectric. *Optics Express*, 14:13101–13106, 2006.
- [112] KC Neuman, EH Chadd, GF Liou, K Bergman, and SM Block. Characterization of photo-damage to Escherichia coli in optical traps. *Biophysical Journal*, 77:2856–2863, 1999.
- [113] A van der Horst and NR Forde. Calibration of dynamic holographic optical tweezers for force measurements on biomaterials. *Optics Express*, 16:20987–21003, 2008.
- [114] K Berg-Sørensen and H Flyvbjerg. Power spectrum analysis for optical tweezers. *Review of Scientific Instruments*, 75:594–612, 2004.
- [115] Simon F Nørrelykke and Henrik Flyvbjerg. Harmonic oscillator in heat bath: Exact simulation of time-lapse-recorded data and exact analytical benchmark statistics. *Physical Review E*, 83:041103, 2011.
- [116] T Odijk. Stiff chains and filaments under tension. *Macromolecules*, 28:7016–7018, 1995.
- [117] A Farré, A van der Horst, GA Blab, BPB Downing, and NR Forde. Stretching single DNA molecules to demonstrate high-force capabilities of holographic optical tweezers. *Journal of Biophotonics*, 3:224–233, 2010.
- [118] P Gross, N Laurens, LB Oddershede, U Bockelmann, EJG Peterman, and GJL Wuite. Quantifying how DNA stretches, melts and changes twist under tension. *Nature Physics*, 7:731–736, 2011.
- [119] MD Wang, H Yin, R Landick, J Gelles, and SM Block. Stretching DNA with optical tweezers. *Biophysical Journal*, 72:1335, 1997.
- [120] YF Chen, DP Wilson, K Raghunathan, and JC Meiners. Entropic boundary effects on the elasticity of short DNA molecules. *Physical Review E*, 80:020903, 2009.
- [121] Y Seol, J Li, PC Nelson, TT Perkins, and MD Betterton. Elasticity of short DNA molecules: Theory and experiment for contour lengths of 0.6–7  $\mu\text{m}$ . *Biophysical Journal*, 93:4360–4373, 2007.

- [122] SW Chang, BP Flynn, JW Ruberti, and MJ Buehler. Molecular mechanism of force induced stabilization of collagen against enzymatic breakdown. *Biomaterials*, 33:3852–3859, 2012.
- [123] K Ravikumar, J Humphrey, and W Hwang. Spontaneous unwinding of a labile domain in a collagen triple helix. *Journal of Mechanics of Materials and Structures*, 2:999–1010, 2007.
- [124] A Gautieri, S Vesentini, A Redaelli, and MJ Buehler. Hierarchical structure and nanomechanics of collagen microfibrils from the atomistic scale up. *Nano Letters*, 11:757–766, 2011.