2-2017

Design of Collagen-Mimetic Peptides

Parminder Jeet Kaur

The Graduate Center, City University of New York

Recommended Citation

https://academicworks.cuny.edu/gc_etds/1930

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Materials Chemistry Commons, and the Other Chemistry Commons
Design of Collagen-mimetic peptides

by

Parminder Jeet Kaur

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2017
Design of Collagen- mimetic peptides

by

Parminder Jeet Kaur

This manuscript has been read and accepted for the Graduate Faculty in Chemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date

Yujia Xu
Chair of Examining Committee

Date

Brian Gibney
Executive Officer

Supervisory Committee:

Hiroshi Matsui

Charles M. Drain

Vikas Nanda

Dorthe Eisele

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

Design of Collagen-mimetic peptides

by

Parminder Jeet Kaur

Advisor: Yujia Xu

Collagen is the major component of the extracellular matrix and is involved in a wide range of cellular functions during tissue development and action. At the core of the diverse functions of this versatile protein is its remarkable ability to form supramolecular structures through self-assembly and/or through interactions with other biomolecules. One typical example of such supramolecular structure is the fibrils of collagen types I, II and III of the connective tissues. These fibrils have staggered arrangement of the laterally associated collagen triple helices – the structural unit of collagen – to form long, smooth fibrils with a characteristic 67 nm axial structural feature known as the D-periodicity.

The fibrillogenesis of collagen in tissues is a complex process involving other macromolecules; the fibril formation itself, however, is a self-assembly process proceeding from the self-association of the triple helix. While many of the structural details of both the triple helix and the collagen fibrils have been elucidated, the molecular recognition mechanisms of the self-assembly process remain poorly understood. Efforts of producing collagen-like, self-assembled fibrils through protein design have not been successful.

In this work, we describe the self-assembly of a collagen-like periodic mini-fibril from a recombinant collagen triple helix. The triple helix, designated Col108, is expressed in E. coli using an artificial gene and consists of a 378–residue triple helix domain organized into three pseudo-repeating sequence units. The Col108 peptide forms a stable triple helix with a melting temperature of 41°C. Upon increases of pH and temperature, The Col108 peptide self-assembles in solution
into smooth mini-fibrils with the cross-striated banding pattern typical of fibrillar collagens. The banding pattern is characterized by an axially repeating feature of ~ 35 nm as observed by TEM and AFM. Both the negatively stained and the positively stained TEM patterns of the Col108 fibrils are consistent with a staggered arrangement of triple helices with a staggering value of 123 residues, a value closely connected to the size of one repeating sequence unit.

A mechanism is proposed for the fibril formation of Col108 with axial periodicity, which is stabilized by the optimized interactions between the triple helices in a 1-unit staggered arrangement, pointing towards the similar underlying molecular mechanisms i.e. inherent sequence periodicity at play in guiding the self-assembly of collagen. Col108 establishes itself as a system comparable to native collagen to establish a connection in the self-assembly mechanisms as it is significantly larger than other collagen-like peptides which are only 30-45 residues long.
In Dedication to my parents, Sardar Davinder Jeet Singh & Sardarni Nirmal Kaur, whose sacrifices have made every step possible
Acknowledgements

My sincere thanks to several excellent colleagues and mentors in the Department of Chemistry at The Hunter College and The Graduate Center, CUNY: to Dr. Hiroshi Matsui, Dr. Charles M. Drain and Dr. Dorthe Eisele for helping me develop contextual knowledge in nanotechnology and protein scaffolding; to Gamage Aruggoda, Sam Wong, Dr. Hanying Bai, and Dr. Yoshiaki Maeda for technical guidance around protein chemistry; Dr. Zengyan Wei for sharing his technical knowledge of TEM; Dr. Kristina Fabijanic and Dr. Christopher Farley for assistance with AFM experiments; and to Fangfang Chen, Dr. Rebecca Strawn and Dr. Yujia Xu for support in literature reviews and scientific writing. I am thankful to undergraduate research students Gabriel Ordas and Pascal Maguin for assisting me in purifying Col108. I am also grateful to Dr. Vikas Nanda of Rutgers University for sharing his valuable perspectives on critical aspects of my research and Dr. Stephen Doty for assistance with TEM imaging. Most importantly, my deepest thanks are reserved for my doctoral advisor, Dr. Yujia Xu. Yujia steered my academic development and research efforts across seven years, which I will always regard as one of the most challenging times in my life. In essence, Yujia taught me how to think - a gift that is far beyond expectations, and seldom received.
INDEX

1. Introduction

1.1. Collagen, from triple helix to fibrils ................................................................. 1
1.2. Biosynthesis of Collagen fibers ........................................................................... 5
1.3. In vitro fibrillogenesis of type I collagen ......................................................... 8
1.4. Peptides as model system for collagen .............................................................. 11

2. Fibril formation studies using commercial rat tail type I collagen

2.1. Introduction ........................................................................................................ 24
2.2. Methods ............................................................................................................. 26
2.3. Results .............................................................................................................. 27

2.3.1. Effect of temperature and salt concentration on the turbidity curve of fibril formation

2.3.2. Effect of temperature and salt concentration on fibril morphology

2.4. Conclusion ........................................................................................................ 37

3. Col108- a designed collagen like fibril forming peptide

3.1. Introduction to the sequence design of Col108 ................................................. 38
3.2. Construct of the recombinant plasmid for Col108 ............................................ 41

4. Optimization of the expression and the purification of Col108

4.1. Introduction ........................................................................................................ 44
4.2. Methods ............................................................................................................. 46

4.2.1. Expression and cell lysis

4.2.2. Purification of His-tagged fusion protein using His-affinity column

4.2.3. Thrombin cleavage
4.2.4. Separation of His-tagged thioredoxin from Col108 using high performance liquid chromatography

4.3. Results........................................................................................................................................................................55

4.3.1. The effects of expression strain and the wash/extract buffer on the purification of His-tagged fusion form of Col108

4.3.2. Optimized thrombin cleavage

4.3.3. High performance liquid chromatography as final purification step

4.4. Conclusion........................................................................................................................................................................82

5. The folding and the thermal stability of Col108 peptide

5.1. Introduction....................................................................................................................................................................83

5.2. Methods ..........................................................................................................................................................................83

5.2.1. Preparing Col108 solutions

5.2.2. Conformation studies

5.2.3. Thermal stability studies

5.3. Results..........................................................................................................................................................................84

5.3.1. Conformation of peptide Col108

5.3.2. Thermal stability of peptide Col108

5.4. Conclusion....................................................................................................................................................................87

6. Fibril formation studies using Col108

6.1. Introduction...................................................................................................................................................................89

6.2. Methods .........................................................................................................................................................................90

6.2.1. Fibril formation of Col108

6.2.2. TEM characterization of the Col108 mini-fibrils
6.2.3. AFM characterization of the Col108 mini-fibrils

6.3. Results.................................................................................................................................92

6.3.1. The mini-fibrils of Col108 with ~35nm periodicity

6.3.2. The $d$-periodicty of the mini-fibrils of Col108 characterized by AFM


6.4.1. The interaction curves

6.4.2. The unit staggered model of Col108 mini-fibrils

6.5. Conclusion.............................................................................................................................................106

6.6. Discussion.....................................................................................................................................................109

Supplementary Information.......................................................................................................................................115

Bibliography.........................................................................................................................................................116
List of Figures

1 Chapter 1: Introduction

1-1 Electron microscopy image of collagen fibrils .................................................................4
1-2 Major steps in biosynthesis of fibril forming collagen ......................................................7
1-3 Kinetics of in vitro self-assembly of type I collagen fibrils .............................................11
1-4 Higher-order assembly of unmodified collagen-like peptides ......................................16
1-5 Cysteine knots in controlled supramolecular assembly of heterotrimeric helices of collagen-like peptides .........................................................................................................18
1-6 Sticky-end assembly strategy relying on hydrophobic interactions from terminally placed aromatic residues in collagen-like peptides .........................................................20
1-7 Assembly strategy of collagen-like peptides relying on electrostatic interactions ....22

2 Chapter 2: Fibril formation studies using commercial rat tail type I collagen

2-1 Turbidity Curves showing an increase in optical density with time as collagen fibrils form ......................................................................................................................................29
2-2 Transmission electron micrographs of self-assembled collagen fibrils using high salt (135mM NaCl) buffer and high incubation temperature (37°C) ........................................33
2-3 Transmission electron micrographs of self-assembled collagen fibrils using high salt (135mM NaCl) buffer and low incubation temperature (26°C) ........................................34
2-4 Transmission electron micrographs of self-assembled collagen fibrils using low salt (67.5mM NaCl) buffer and high incubation temperature (37°C) .....................................35
2-5 Transmission electron micrographs of self-assembled collagen fibrils using low salt (67.5mM NaCl) buffer and low incubation temperature (26°C) .................................36
3 Chapter 3: Col108- a designed collagen-like fibril-forming peptide

3-1 The amino acid sequence architecture of peptide Col108 ...........................................40
3-2 The gene construct of peptide Col108 .........................................................................43

4 Chapter 4: Optimization of the expression and the purification of Col108

4-1 Amino acid sequence of fusion form of Col108..........................................................45

4-2 Time Program Graphs

4-2.1 Time program graph used for analytical column representing % gradient acetonitrile..................................................................................................................52

4-2.2 Time Program graph used for preparative column representing % gradient acetonitrile ..................................................................................................................53

4-2.3 Modified Time Program graph used for preparative column representing % gradient acetonitrile ..................................................................................................54

4-3 SDS-PAGE

4-3.1 Expression of His-tagged Col108 in JM109...............................................................57

4-3.2 Effects of wash/extract buffer on the purification of His-tagged Col108 expressed in JM109 ..............................................................................................................60

4-3.3 Continued from 4-3.2 .................................................................................................61

4-3.4 Effects of wash/extract buffer on the purification of His-tagged Col108 expressed in BL21 ..............................................................................................................64

4-3.5 Continued from 4-3.4 .................................................................................................65

4-4 A Schematic representation of thrombin cleavage reaction to separate His-tagged thioredoxin from fusion Col108 .............................................................................70

4-5 SDS-PAGE
Chapter 5: Characterization of Col108 by circular dichroism (CD) spectroscopy

5-1 Circular dichroism spectra of Col108 ................................................................. 86

5-2 Temperature melt curve of Col108 ................................................................. 88

Chapter 6: Fibril formation studies using Col108 peptide

6-1 The electron micrographs of Col108 mini fibrils formed after 6 hours incubation at 37°C ................................................................. 94
6-2 The electron micrographs of Col108 mini fibrils formed after 24 hours incubation at
37°C .................................................................................................................95

6-3 The electron micrographs of Col108 mini-fibrils formed at 26°C ......................96

6-4 AFM analysis of the Col108 mini-fibrils ..................................................................98

6-5 The interaction curves for optimal interactions .................................................102

6-6 The model on the Col108 mini-fibrils with 35nm periodicity ..............................105

6-7 Comparison of self-assembly of Col108 and native collagen highlighting the similarity
in assembly mechanism and differences in morphological features ......................108
CHAPTER 1
INTRODUCTION

1.1 Collagen, from triple helix to fibrils

Collagen is the most abundant protein found in the connective tissue\textsuperscript{11-12}. It is most prevalent as the fibrous component of skin, bone, tendon, cartilage and teeth in vertebrates. So far at least 28 different types of collagen have been reported with 46 different polypeptide chains\textsuperscript{12}. Collagen is involved in a wide variety of developmental processes and tissue homeostasis. This versatile protein is known for its diversity in cellular functions and has a remarkable ability to form supramolecular structural networks like microfilament or fibrils\textsuperscript{1}. Formation of fibrils, i.e. fibrillogenesis, is a trait of the family of fibrillar collagen including collagen types I, II, III, V, and XI; they form fibrils with characteristic banding patterns\textsuperscript{8, 13}. Among fibrillar collagens, type I collagen establishes itself as the major fibril component of bones, tendons, ligaments, skin and other tissues and accounts for \textasciitilde 90\% of the total collagen content of the body. Type II collagen is mostly the structural component of cartilage and the vitreous humor of the eye. Type III collagen is predominantly in blood vessel walls, skin, and internal organs and coexists in a fibrillary form with type I collagen. Types V and XI are quantitatively less abundant collagens and found in association with types I and II collagen fibrils\textsuperscript{14}.

Basement membrane collagens, also known as network-forming collagens, are found in a particular extracellular matrix which supports epithelial cells, muscle fibers, and peripheral nerves; types IV, VIII, XV, and XVIII belong to this family. Types IX, XII, XIV, XVI, XIX are fibril-associated collagens with interrupted triple helices, i.e. FACIT’s\textsuperscript{9, 12, 15}. MACIT’s are membrane-associated collagens with interrupted triple helices, also known as transmembrane collagens, and...
Type XXV belongs to this family. Type XV and XVIII have multiple triple helical domains and interruptions, therefore known as MULTIPLEXIN’s 12.

Irrespective of the kind of collagen, one characteristic of this protein is that they all have the same basic structural element: the triple helix11-12. A collagen triple helix is composed of three parallel polypeptide chains in the left-handed polyproline II (PPII) conformation, which further supercoil in a right-handed fashion around a common axis. Every third amino acid residue in the polypeptide chain must be a Glycine (Gly). This results in a repeating pattern of the sequence of (Gly-X-Y), where X and Y could be any amino acid but is often imino acids proline (Pro) and hydroxyproline (Hyp), respectively, at X and Y positions. Glycine residues with their small side chains are packed at the center of the triple helix, whereas the side chains of X- and Y- residues are exposed to the solvents and are responsible for direct molecular interactions between different triple helices during fibril formation16-17. The –NH group of the Gly-, amino acid residue, forms an inter-chain hydrogen bond (H-Bond) with –C=O of the X-position residue on the neighboring polypeptide chain. These inter-chain or intra-helical H-bonds (one H-bond per one Gly-X-Y-triplet) are oriented perpendicularly to the axis of the triple helix and are known as the dominant stabilizing factor of the triple helix18-20.

While all collagens share the same conformation of the triple helix, various types of collagen can be differentiated by the interactions between helical and non-helical regions leading to the assembly of different supramolecular structures. Type I collagen is very abundant in tendon-rich tissue and due to this abundance was the first of the collagens to be characterized structurally.

A triple helix of collagen type I is a heterotrimer formed from two identical α1(I) polypeptide chains, and one α2(I) chain having a different amino acid sequence. Each polypeptide chain has 1,050 amino acid residues in the non-interrupted (Gly-X-Y) repeating sequence. Each
triple helix molecule of collagen (also known as the tropocollagen) has a width of 1.5nm, a length of 300nm and a molecular weight of ~285kDa. The triple helices further interact laterally to form fibrils of 50-200nm in diameter and more than 10µm long. In these fibrils, adjacent triple helices are all in parallel to each other but displaced from one another in a stagger of 67 nm also known as the D-period\textsuperscript{4}. There are in total 4.4D periods in one triple helix of native collagen, and the D-staggered arrangement resulted in the 67nm axial repeat pattern of dark and light bands as observed in electron micrographs if the sample is stained negatively: the 0.4D overlap zone (27nm) forms the light band, and the 0.6D gap region (40nm) appears as a Dark band in observed electron micrographs (Figure 1-1). Structural studies have related this 67nm D-periodicity to defined sections of the triple helix comprised of 234 amino acid residues (per single chain)\textsuperscript{21-23}. 
Figure 1-1 Electron microscopy image of Collagen fibrils. A. Schematic drawing of axial packing arrangement of triple helices of collagen in a fibril, B. Negatively stained collagen fibril, C. Positively stained collagen fibril. Image adapted from Kadler et al.¹
1.2 Biosynthesis of collagen fiber

Collagen self-assembly in vivo is a multistep, complex but well-organized process. From collagen synthesis to the folding of the triple helix and fibril assembly, many intracellular and extracellular steps are involved (Figure 1-2). Collagen synthesis begins at the membrane of the rough endoplasmic reticulum (RER). Ribosomes synthesize the individual chains. These procollagen chains at the surface of the RER are in a soluble precursor form known as pro-α chains. These pro-α chains contain the signature (Gly-X-Y)_n sequence and globular N- and C- propeptides which flank the (Gly-X-Y)_n sequence, respectively, on the N-and C-termini.

The C-propeptides control the intracellular assembly of pro-α chains into triple helices i.e. procollagen molecules, which can be either homotrimeric or heterotrimeric. The C-propeptides play a crucial role in nucleation and folding of the pro-α chains into homo- or heterotrimeric triple helical molecules. Further, C-propeptides confer the chain selectivity intracellularly during assembly of different types of collagen and are also known to confer solubility to the procollagen molecules to prevent fibril assembly during intracellular steps.

Before the procollagen molecule folds into a triple helix, the peptide chains undergo post-translational modifications. These modifications require the enzymes: Prolyl hydroxylases, glycosyltransferases, isomerases, proteinases, and oxidases. Hydroxylation of almost all of proline residues to 4R-hydroxyproline (Hyp) in the Y-position and a few lysine residues in the Y positions to hydroxylysine occurs. All the hydroxylation takes place in RER. The Hyp in the Y residue position provides a stabilizing effect to the triple helical structure. The conversion of proline into hydroxyproline in the Y position in a Gly-X-Y set leads to the formation of Gly-X-Hyp triplets at the C-terminus, which in turn acts as a nucleation site for triple helix
folding/formation. From the nucleation site, the triple helix folding propagates towards N-terminus\textsuperscript{32}.

Additionally, some of the prolines in the X- and Y- residue positions are hydroxylated to 3-Hydroxyproline (3-Hyp) instead of 4R-Hyp. What specific roles these residues play in collagen structure and formation remains unclear, but in collagen, the locations of 3-Hyp residues are highly conserved. The hydroxylation of certain Lys in Y-positions to hydroxylsine (Hyl) residues also takes place. The hydroxylation of Lys is significant since the hydroxyl group of Hyl -is a site for attachment of carbohydrates and for inter-helical cross-linking\textsuperscript{13, 33}; which contribute to the stabilizing the collagen. Peptidyl proline cis-trans isomerase and protein disulfide isomerase (PDI) mediates the formation and rearrangement of inter- and intrachain disulfides in N-and C-propeptides\textsuperscript{33}.

Once a procollagen triple helix molecule is formed, it is secreted into the extracellular matrix via transport vesicles. In the extracellular matrix, this precursor form of collagen undergoes further modifications\textsuperscript{1, 8, 34-35}. Appropriate proteinases cleave respective propeptides, leading to the formation of tropocollagen which contains triple helix molecules with only (Gly-X-Y-)\textsubscript{n} sequence repeats. Finally, lysyl oxidase leads to oxidative deamination of specific lysine and hydroxylsine residues to allysine and hydroxyallysine to make them available for cross-linking and glycosylation\textsuperscript{36-37}.

The processed tropocollagen molecules self-assemble to form fibrils in the extracellular matrix. As the tissues mature, covalent cross-links between neighboring triple helices are formed between the α-amino groups of lysine and hydroxylsine residues and the aldehyde group of allysine and hydroxyallysine. The cross-links stabilize the fibrils and contribute to the tensile strength and mechanical stability of connective tissue\textsuperscript{38}. 
**Figure 1-2 Major steps in biosynthesis of fibril forming collagen.** A schematic drawing of the journey of collagen from its biosynthesis in the endoplasmic reticulum into the extracellular matrix via the golgi complex. Intracellular steps involve the synthesis of individual collagen pro-α chains by ribosomes, which are injected into the lumen of endoplasmic reticulum (ER). Before folding of these biosynthesized chains in the ER, the chains undergo many post translational modifications over ~10% of all residues i.e. hydroxylation of proline and lysine residues, glycosylation of some hydroxylysine residues, and protein disulfide isomerase(PDI) mediated rearrangements of disulfides in propeptides. Collagen folding into a triple helix takes place inside the cell starting at C- terminus. After triple helix formation is completed, the procollagen molecules are secreted into the extracellular matrix, where propeptides from the N-and C- termini are cleaved and the collagen molecules self-associate to form fibrils. Covalent cross-links are formed between different collagen triple helices to stabilize the fibrils. Image adapted from Myllyharju et al.8
1.3 *In Vitro* fibrillogenesis of type I collagen

Although collagen fibril formation *in vivo* is mediated by many enzymes present in the cell, it has been shown that the triple helices, as the building block of the complicated structure of the extracellular matrix, contain much of the required information for fibril assembly. Thus, fibril formation of purified collagen molecules can be studied *in vitro* in defined conditions. Studies of the *in vitro* fibrillogenesis have provided much insight into fibril formation mechanisms. *In vitro* studies have shown collagen fibril formation to be an entropy-driven self-assembly/polymerization process, which is achieved by increasing the temperature and pH of cold, acid-dissolved collagen extracts from tissues to physiological conditions. Reconstituted collagen fibrils formed *in vitro* exhibit all the features observed for fibrils formed *in vivo*, but are less uniform in diameter. This suggests that the regulating factor of fibril size and length is lost during the extraction of collagen, or that the mechanism of regulation is different during the generation of fibrils from pure solutions.

The mechanistic growth of a fibril during *in vitro* fibrillogenesis has been shown to be proceeded by nucleation of monomers and resulting growth to form fibrils of constant length and diameter. *In vitro* fibril formation is often monitored by turbidity measurements at 313nm.

*In vitro* fibril formation starts with a lag period, which is attributed to fibril nucleation. This phase is considered the formation of precursors of the fibril. These precursors grow linearly until a critical length is reached. This lag period has important roles in determining the final fibril size. The value of the pH of the solution, the ionic strength and ion types present in the buffer, and the temperature are a few parameters known to affect the duration of the lag phase. Light scattering studies indicated that the initial aggregates show a linear association of collagen molecules along the fibril axis instead of lateral organization. The second phase of the turbidity
time curve is known as the growth phase where the microfibrils formed during the lag phase grow linearly and laterally. The increased size of the fibrils is the reason for observed solution turbidity (Figure 1-3). The last phase is the plateau phase in which fibrils with an axial repeat pattern, i.e. D-period, are readily observable.

Electron microscopy and light scattering techniques have further confirmed in vitro fibrillogenesis to be a multi-step process in which linear aggregate formation is followed by lateral growth. Quasi-elastic light scattering studies of collagen solutions have indicated that in early stages of in vitro fibrillogenesis, 4.4D staggered interactions are already present. Also, electron microscopy studies have supported the presence of 4.4D staggered dimers during early stages of in vitro assembly of collagen.

It is not completely understood what properties of the triple helix lead to the D-periodicity of the self-assembled fibrils. Hulmes and colleagues have shown by using a computer-aided sequence alignment approach that by a mutual staggering of multiple units of 234 residues, or multiples of a D-period, hydrophobic and charged interactions between neighboring triple helices were maximized. Therefore, it has been proposed that maximized interaction is the driving force which leads to spontaneous self-assembly of collagen triple helices into fibrils with the precise structural specificity of a 67nm D-period. However, due to the complexity of fibrillogenesis, the critical role of known factors in fibril formation are difficult to isolate and evaluate. The possible contribution of other factors to the features of the supra-molecular fibrous structure of native collagen have also been explored: the role of N- and C-telopeptides in fibrillogenesis, which are considered essential for anchoring ends of neighboring collagen molecules, the role of aromatic residues that contribute to CH—Π interactions; and the role of hydroxyproline, which adds additional stability due to stereoelectronic effects and hydration network. The fundamental
biological process is still poorly understood, and much work is needed to advance the field of collagen research. Unfortunately, our current understanding of the molecular recognition mechanisms involved in self-assembly of collagen is insufficient and has limited the efforts of engineering macroscopic ECM scaffolds by protein design.
**Figure 1-3 Kinetics of *in vitro* self-assembly of type I collagen fibrils.** (i) Turbidity increases significantly as collagen self-assembly into fibrils takes place. Experiments were conducted for collagen solution 0.5mg/mL, pH 6.9 and temperature 21°C. Inset schematics show that in vitro self-assembly of collagen starts as triple helical collagen molecules and ends with well-ordered collagen fibrils that form gel. A. an initiation step where soluble collagen aggregates formation takes place and these aggregates/ precursors grow linearly until a critical length is reached. B. Growth of the fibrils takes place in this step leading to linear and lateral growth of fibrils. Solution turbidity increases during this step. C. A plateau is the last step when native banded fibrils are obtained. (ii) TEM image of a self-assembled collagen fibril showing dark-light D-banding pattern. Image adapted from Shayegan et al.6
1.4 Peptides as model systems for collagen

The understanding of roles/contributions of the factors mentioned above towards molecular recognition mechanisms of this peculiar assembly of collagen has been explored using synthetic triple helical peptides by different researchers\textsuperscript{10, 49}.

Collagen-like peptides have played a very significant role in our current understanding of collagen. They were essential for elucidating the structure of the triple-helical conformation of collagen. These polypeptides also informed our understanding of how sequence drives the structure and function of native collagen.

A conceptual outline of the structure of the collagen triple helix was developed from structural studies of polyproline\textsuperscript{50} and polyglycine\textsuperscript{51}. These expositions provided insight into the fundamental stabilizing interactions. Once the structure of collagen triple helix was resolved, it set the stage for synthesis and characterization of collagen-like oligopeptides and polypeptides\textsuperscript{52-56}. Deriving the relationship between conformation, amino acid sequence, and stability of the triple helix was the aim of such studies. These studies highlighted the significance of residues in X-and Y-positions of the Gly-X-Y sequence and the role of the solvent in stabilizing the triple helix. X-ray crystallography was used to explore the high-resolution structures of collagen-like peptides. (PPG)\textsubscript{10} was the first 30-mer peptide whose crystal structure was solved by Okuyama et al.\textsuperscript{57}. Hydration network\textsuperscript{58-59}, hydrogen bonding\textsuperscript{60-61} and helical parameters\textsuperscript{57} were some of the many features that were addressed by high-resolution structures of collagen-like peptides. The sequence dependency of the triple helical parameters was decoded by studying structures of peptides with different sequences. Variability in the super-helix twist / helical symmetry in the imino acid rich- and poor- peptides was highly significant, indicating a high probability of having subtle variations.
in the helical conformations\textsuperscript{62-64}. This difference in helical symmetry can be used for fibril assembly or functionalizing collagen with other molecules.

Further advances from sequence-structure findings included the structural consequences of pathological mutations. Host-guest peptides of sequence (GOP)\textsubscript{3}-G-X-Y-(GOP)\textsubscript{4} were used to relate the amino acid sequence of the collagen triple helix to its thermal stability\textsuperscript{65}. Triple-helical stabilization propensities of all 20 residues in the X-position of the G-X-O guest triplet, and for 20 residues in the Y-position of G-P-Y triplets were evaluated. Also, interactions between X-and-Y- residues in a G-X-Y guest triplet along with interactions between neighboring triplets G-X\textsubscript{1}-Y\textsubscript{1}-G-X\textsubscript{2}-Y\textsubscript{2} were similarly measured. The results culminated in a collagen stability calculator. The relative stability predictions from the algorithm and the observed experimental values of melting temperature differ by only 3-4°C. Evaluating the effects of single point mutations in the X- and Y- positions is a common application of this algorithm. Consequently, this algorithm has been very useful in protein design.

Collagen-like peptides have played a significant role in providing conceptual information of the effects of Gly mutations in collagen. It is well known that abnormal degradation or cross-linking of collagen can lead to many common diseases like arthritis and diabetes. Also, specific genetic mutations in collagen can lead to rare genetic disorders. One example of a collagen-related genetic disease is Osteogenesis Imperfecta (OI), which has a clinical manifestation of bone fragility resulting from a mutation in type I collagen\textsuperscript{66}. In recent years, it was found that at least 682 Gly missense mutations in \( \alpha_1 \) and/or \( \alpha_2 \) chains in the Type I heterotrimer cause bone fragility\textsuperscript{67}. The severity of the disease is dependent on factors like location of the mutation with respect to the C-terminus; type of residue that is replacing glycine; chain in which the mutation
occurs etc. The effect that these mutations have on the folding of the helix has also been studied using host-guest triple helical peptides.

While collagen-like designer peptides have been very successful in providing a detailed story of the collagen triple helix at a molecular level, much emphasis has not been placed on understanding the higher order assembly of such collagen peptides. The abundance of collagen in the extracellular matrix, its structural integrity and its role in cellular growth and tissue development, makes it an ideal candidate for biomedical applications such as drug delivery, tissue regeneration, and bone repair. However, the risk of immunogenic responses as well as the transfer of toxic agents like prions, pose challenges in the commercialization of natural collagen. Another drawback of natural collagen is that its structure cannot be modified in a sequence-specific manner. Also, natural collagen scaffolds when used as transplants in the extracellular matrix cannot withstand the collagenase-induced degradation. Some of these challenges are addressed through nano- and macro-assembly strategies, which rely on short peptides that are easily synthesized and can self-assemble into higher order structures upon physiological conditions.

Brodsky et al. have initially reported a higher order aggregated structure of a 30 residue long (POG)10 unmodified collagen peptide. This research focuses on elucidating the similarities between natural collagen self-assembly and the self-association of this small peptide. This study underlines the critical link between peptide and collagen self-association - the role of hydroxyproline in mediating the lateral interactions between the building blocks of self-assembly, i.e., triple-helical molecules. The aggregates of (POG)10 show no morphological similarity to collagen fibrils but provide the grounds for the studies focusing on the development of synthetic collagen fibers (Figure 1-4A). Techniques like chemical cross-linking and native chemical ligation have also been used to generate polymerized collagen peptides with poly-(POG) sequence.
(Figure 1-4B). Both of these methods produced nanofiber-like structures, which were longer than the aggregates of (POG)$_{10}$. 
Figure 1-4 Higher order Assembly of unmodified collagen-like peptides. A. TEM image of aggregates formed after self-assembly of (Pro-Hyp-Gly)$_{10}$ peptide B. Native chemical ligation mechanism combining a C-terminal thioester and N-terminal cysteine facilitating polymerization of Cys-Hyp-Gly-(Pro-Hyp-Gly)$_{9}$ peptide and fibers formed from polymerized peptide. Image A adapted from Kar et al. Image B adapted from Paramonov et al. and Przybyla et al.
Additional modified collagen peptides include functional incorporations at one or both ends. One popular modification is cysteine knots, which are used to cross-link three collagen peptide sequences in a triple helix, covalently. The cysteine knots increase the thermal stability of the triple helix and induce the generation of higher order collagen peptide assemblies using heterotrimeric helices. Koide et al. have shown that using disulfide linkages between three (POG) based 24-mer peptides results in staggered overhangs of 12 and 13 residues, allowing for self-complementary supramolecular assembly of these peptides into 0.6 and 14µm particles (Figure 1-5A). Koide et al. have further modified this prototype peptide to generate staggered overhangs of 15- and 18 residues. These peptides, unlike the prototype, have managed to produce hydrogels. However, a comparison between this hydrogel and native collagen gel has never been studied. Similarly, Raines et al. have used disulfide linkages to tether collagen peptides, resulting in higher order assemblies (Figure 1-5B).

In 2009, Kiick et al. had reported a hydrophilic, non-repeating and hydroxyproline-lacking collagen mimetic peptide that includes a C-terminal Cys-knot and forms covalently crosslinked homotrimers. Strikingly, the C-terminal Cys-knot was not critical for the assembly. The synergistic effect of hydrophobic clusters of Gly-Pro-Pro sequences at both termini and the presence of charged residues facilitating electrostatic interactions were sufficient to guide the higher order assembly of a 30-mer collagen peptide (CLP-Cys).
Figure 1-5 Cysteine knots in controlled supramolecular assembly of heterotrimeric helices of collagen-like peptides. A. Assembly strategy of self-complementary collagen like peptides designed by Koide et al.; B. Assembly strategy of staggered collagen sequence designed by Raines et al. Image adapted from Przybyla et al.
Maryanoff et al., Raines et al. and Brodsky et al. used a sticky end approach relying on noncovalent hydrophobic interactions to promote the self-assembly of collagen peptides (Figure 1-6). These groups used terminally placed aromatic residues on designed collagen peptide sequences. Maryanoff et al. and Raines et al. engineered the peptide assembly by an end-to-end stacking of triple helices, which was caused by ordered, stacking hydrophobic interactions between the terminally placed aromatic residues\(^5\), \(^80\text{81}\). Brodsky et al. also demonstrated the collagen peptide assembly using terminally positioned aromatic residues\(^47\). In such assemblies, along with \(\pi\text{-}\pi\) stacking interactions, the CH- \(\pi\) interactions between the terminally positioned aromatic residue at the end and imino acids within the adjacent triple helix occur. Both of these self-assembling strategies relied on hydrophobic interactions producing self-assembled supramolecular structures of collagen peptides.
Figure 1-6 Sticky-end assembly strategy relying on hydrophobic interactions from terminally placed aromatic residues in collagen-like peptides. A. Sequence design of a collagen-like peptide with hydrophobic residues at both termini; a pentafluorophenylalanine residue at N-terminus and a Phe residue at C-terminus of (Pro-Hyp-Gly)$_{10}$ peptide. B. TEM image of assembled collagen fibers. Image adapted from Przybyla et al.
Electrostatic interactions have also been used in generating collagen mimetic microfibers. Chaikof et al. have created a homotrimeric peptide with three different Gly-X-Y domains flanked with a positively charged amino acid (Arg) or a negatively charged amino acid (Glu). This collagen peptide domain has a poly-(POG) sequence as a hydrophobic core, which increases the thermal stability of this peptide design (Figure 1-7). The electrostatic interactions of terminally positioned, charged residues serve a two-fold purpose: 1) to facilitate the linear oligomerization within the fibrils; and 2) to promote the staggered orientation between neighboring triple helices. However, one critical step in achieving this observed self-association is thermal annealing- heating the peptide solution to 75°C for 40 minutes and followed by cooling at room temperature.
Figure 1-7 Assembly strategy of collagen-like peptides relying on electrostatic interactions. A. Sequence design of (Pro-Arg-Gly)₄-(Pro-Hyp-Gly)₄-(Glu-Hyp-Gly)₄ collagen peptide B. Model showing charged residues Arg and Glu facilitate linear assembly of triple helix through electrostatic-interactions and D-periodic collagen peptide fiber. Image adapted from Przybyla et al.¹⁰
Additionally, Chmielewski et al. reported the directional assembly of collagen-based peptides into large and well-ordered structures with tunable shapes promoted by metal-ligand interactions\textsuperscript{82-83}.

Over time, our understanding of the structure of collagen, especially its biological diversity, has promoted its potential as a novel biomaterial. From regenerative medicine to stem cell differentiation, the biomedical importance of collagen is apparent. Site-specific modifications in collagen-like peptides allow researchers to use biological and physical properties of native collagen and functionalize modified biological sequences to create useful biomedical scaffolds. However, mimicking natural collagen self-assembly in the actual sense is still a long journey. The size of these functional collagen peptides does not match the real length of native collagen, and therefore, does not provide complete insight into the molecular mechanism of the self-assembly of native collagen. While this does not preclude the development of functional scaffolds, an improved understanding of native collagen assembly would contribute substantially to the basic research underlying the design and synthesis of mimetic peptides.
CHAPTER 2

Fibril formation studies using Commercial rat tail type I collagen

2.1 Introduction

In vitro assembly of collagen under physiological conditions results in fibrils formed with the characteristics of those observed in vivo. The similarities found between native fibrils and reconstituted fibrils from purified type I collagen were the driving force for studies focused on understanding collagen self-assembly. The effects of factors such as pH between 5.0-8.5, ionic strength between 0.1 and 0.8, temperatures between 15-37°C, the identity of ion species in solution, the presence of surfactants, etc. on the self-assembly process have been characterized. Collagen fibrillogenesis is an endothermic, thermodynamically favorable and entropy driven process. Theoretical models implicate hydrogen bonding and hydrophobic and electrostatic interactions as the stabilizing factors of self-assembly.

It is well known that temperature affects the fibrillogenesis rate as well as the morphology of the fibrils. Self-assembly of collagen i.e. transition of collagen from a liquid phase to a solid phase resulting in the particular D-periodic feature of fibril morphology was explained as a physical process of diffusion-limited aggregation (DLA). Using this model, Parkinson et al. showed that the fibril formed using their model resembled the fibrils formed at a high temperature of 37°C. One of the major assumptions made in this example was that a total of 4.4D periods exist per collagen molecule; all the staggers (0D, 1D, 2D, 3D or 4D) were considered to have an equal stabilizing effect on the fibril structure. During warming, collagen monomers diffuse slowly onto the formed micro-fibril or nucleation aggregate. Spontaneous, thermodynamically favorable and diffusion-limited aggregation at warming temperatures can lead to the formation of well-organized...
and long fibrils. Regarding warming the neutralized collagen solutions, it was observed that the collagen solutions at either 26°C or 37°C led to the formation of gels and fibrils with the distinct D-banding pattern. It has been reported that increases in temperature favor the release of structured water from the hydrophilic residues to bulk solution. Also, increases in temperature facilitate alignment of protein side chains to the most favorable attractive configuration.

In vitro fibrillogenesis of collagen has been extensively studied using electron microscopy with the aim of developing insights into the complex molecular mechanisms involved in natural collagen fibrillogenesis. The variation in the diameter of fibrils formed in vitro is a common observation. Controlling the dimensions of fibrils formed during these experiments is difficult. Under high salt conditions, the fibrils formed at both temperature conditions have shown smaller diameters compared to fibrils formed at low salt concentration. This observation is in agreement with a study reported by J.R.Harris et al.

Ion species in the buffer also have a strong influence on fibrillogenesis. At neutral pH and low salt conditions, it has been shown that the dibasic phosphate (HPO$_4^{2-}$) binding sites per collagen molecule increase relative to high salt conditions. Phosphate anions are either directly bound to the collagen or present as free anions in the interstitial water. Binding of dibasic phosphate to the collagen fibrils is known to facilitate the longitudinal growth of collagen fibrils as phosphate binding has been shown to affect the electrostatic interactions significantly. Also, dibasic phosphate anion binding has a weak promoting effect by reducing the critical concentration required for collagen fibrillogenesis.

The purpose of our study on fibril formation using rat tail type I collagen was to find the optimal experimental conditions to guide the study of the fibril formation using our designed collagen-like peptide. We wanted to identify the optimal conditions regarding salt concentration
and incubation temperature for fibril formation. The method for fibrillogenesis was adapted from the assembly procedure used by Piez et al.\textsuperscript{39} The fibril formation of the type I collagen isolated from the rat tail tendon was studied in two different salt concentrations: high salt (135mM NaCl) and low salt (67.5mM NaCl) (after mixing, see section 2.2 for details) and at two different temperatures: 26°C and 37°C.

2.2 Methods

The buffer conditions used in this experiment were 5mM acetic acid (HAc, pH 4) for dissolving collagen; double strength neutralization buffer (60mM TES, 60mM Na\textsubscript{2}HPO\textsubscript{4}, pH 7.4) with high salt (270mM NaCl) and low salt (135mM NaCl) for fibril formation experiment. The final concentration of buffer after neutralizing the acidic solution of collagen is 2.5mM HAc, 30mM TES, 30mM Na\textsubscript{2}HPO\textsubscript{4} and high salt (135mM NaCl) and low salt (67.5mM NaCl) both at pH 7.4.

The rat tail collagen (Sigma-Aldrich C7661) was dissolved in 5mM Acetic acid (HAc) to the concentration of ~1mg/mL (estimated by weight) and equilibrated overnight at 4°C to ensure the complete dissolution and folding of the protein. To start fibrillogenesis, this acidic solution of collagen was mixed in 1:1 equal volume ratio with a double strength (2X) neutralization buffer. The final concentration of salt after mixing was 135mM (the high salt condition) or 67.5 mM (the low salt condition) NaCl, pH 7.4. The buffers were pre-chilled on ice, and mixing was completed under cold conditions. The sample was transferred to a water bath set at 37°C. The process of fibril formation was monitored by measuring the turbidity at different time intervals. Samples to be used to study the formation of fibrils under transmission electron microscope were prepared after a plateau was observed in the turbidity curve. The turbidity change was monitored by recording
optical density of the sample at 313nm using a DU800 spectrophotometer. The raw data was plotted to obtain a turbidity curve.

The Transmission Electron Microscope (TEM) imaging samples were prepared on 400 mesh formvar-carbon coated grids and imaged under a JEOL2100 TEM. 3µL of the sample were placed on the grid and allowed to sit for for 100 seconds. The grids were washed using deionized water for 5 seconds. The grid samples were stained with 3µL of 1 % Phosphotungstic acid for 100 seconds, and the grids were rinsed again with deionized water for 5 seconds. The samples were left to air dry before imaging.

The same protocol was used for fibril formation at 26°C except for the water bath temperature, which was set at 26°C.

2.3 Results

2.3.1 Effect of temperature and salt concentration on the turbidity curve of fibril formation

The fibril formation was monitored by the optical density at 313 nm, which has been shown to be comparative to the amount of product that is assembled. The turbidity curves obtained in the experiment using high salt (135mM NaCl) and low salt (67.5mM NaCl) neutralization buffer are shown in Figure 2-1A. A typical sigmoidal curve with a lag phase, growth phase, and a plateau region was observed. No noticeable increase in turbidity during the lag phase was observed. Turbidity increased very quickly during the growth phase and the plateau, remaining constant; confirming the nucleation-polymerization process.

At high temperature, i.e. 37°C, the plateau was reached in about 45 minutes in both low and high salt conditions. However, the lag phase under a high salt condition was observed to be shorter than the lag phase in low salt condition. The growth phase, where the linear and lateral
growth of fibrils takes place, appeared to start earlier in high salt conditions as compared to low salt conditions. The plateau in high salt conditions was also reached earlier, compared to low salt conditions.
Figure 2-1 Turbidity curves showing an increase in optical density with time as collagen fibrils form. The concentration of the starting acidic solution of collagen was 1.0mg/mL. The starting sample was brought to physiological conditions by mixing in 1:1 volume ratio with double strength high salt (Red circles) or low salt (Blue circles) neutralization buffer (pH 7.4). The final concentration of the sample after neutralization is ~0.5mg/mL. A. The sample incubated at 37°C. B. The sample incubated at 26°C.
The turbidity curve observed at 26°C shown in Figure 2-1B was comparable to that observed at 37°C. It clearly showed a sigmoidal curve with a lag phase, a growth phase, and a plateau region. At low temperature, the total time for fibrillogenesis to reach plateau was observed to be almost 5 times longer than at high temperature. The total time for fibril formation at low temperature using high salt concentration in neutralization buffer is ~200 minutes, and for the sample neutralized with low salt concentration was ~250 minutes. The effect of changing the salt concentration (on turbidity) was not prominent at low temperature.

From Figure 2-1, it is clear that under high and low salt conditions, the fibrillogenesis rate was influenced by temperature. At high temperature, collagen fibrillogenesis proceeded at a faster pace compared to the same process conducted at low temperature. Also, the total time of the lag phase before the onset of growth phase was observed to be greater for the sample neutralized with low salt buffer as compared to high salt buffer. Electrostatic interactions have long been considered to play a role during self-association of the triple helix. However, these results show that the effects of salt on fibrillogenesis are still not fully understood\textsuperscript{96-97}. The time to reach a plateau of fibril formation completion at both high and low salt conditions, as seen from the turbidity curve, was comparable.

### 2.3.2 Effect of temperature and salt concentration on fibril morphology

The morphology of the fibrils formed under different conditions -low salt solutions and high salt solutions incubated at 37°C or 26°C was examined using TEM. For samples incubated at 37°C or 26°C, the D-period of the fibrils was clearly distinguishable after staining the sample with either 1% Phosphotungstic acid or 1% Sodium Phosphotungstate. The majority of the fibrils appeared smooth, unbranched, and very long. It was difficult to follow the length of a fibril from end to end. However, the diameter of formed fibrils was able to be characterized.
For high salt containing samples incubated at high temperature (37°C), the diameter of the formed fibrils was observed to be 29-58nm (Figure 2-2). And, at low temperature (26°C), the fibrils formed were 30-85nm in diameter (Figure 2-3). In Figure 2-2(i) the fibrils were stained negatively with 1% sodium phosphotungstate staining reagent. The fibril surface appeared smooth and clear D-banding pattern on the fibril was observed. Salt/staining reagent deposits were clearly seen on the formed fibrils. In Figure 2-2(ii), it can be seen that samples stained with phosphotungstic acid were stained positively. These fibrils did not appear smooth, and a lot of backgrounds (probably coming from salt/staining reagent) besides the fibrils was observed on the grid. In Figure 2-3, different areas of the grid are showing the presence of D-banded fibrils. Some non-banded clusters of fibrils were seen in the background.

For low salt-containing samples incubated at high temperature, the fibrils formed were of 50-127 nm in diameter (Figure 2-4) and for samples incubated at low temperature the diameter of the fibrils observed was 40-92nm in diameter (Figure 2-5). In Figure 2-4(i), for samples stained with 1% sodium phosphotungstate, the fibrils appeared to have a clear D-banding pattern irrespective of a lot of backgrounds. For the samples stained with 1% phosphotungstic acid (shown in Figure 2-4(ii)), the fibrils appeared to have a very clear and defined D-banding pattern. But some salt/staining reagent deposit was seen on the fibrils as well. For samples incubated at a low temperature in low salt conditions (Figure 2-5), the fibrils appeared to have a very clear D-banding pattern, and the fibrils appeared very smooth and with no extra salt/stain deposit.

Comparing different salt conditions and different incubation temperatures, it appeared that a low salt condition seems to work better to give well-defined D-bands on the fibrils. Also, the salt/staining reagent deposit was not so significant in the low salt condition. Incubation temperature variation did not bring any significant changes to the fibril morphology at a low salt
condition. Samples showed both positive and negative staining patterns using the same heavy metal staining reagent. These estimates were based on 7-16 measurements of a single grid prepared for each condition.
Figure 2-2 Transmission electron micrographs of self-assembled collagen fibrils in high salt (135mM NaCl) buffer and at high incubation temperature (37°C). Samples for Electron microscopy were taken after a plateau phase was reached during fibrillogenesis as observed using turbidity curve studies and stained with (i) 1% sodium phosphotungstate and (ii) 1% phosphotungstic acid. The banding pattern observed in the fibrils is similar to that of fibril formed in vivo. (i) (A) and (B) show different magnifications and different areas of the grid. A 67nm D-period on the fibrils is clearly seen in the negatively stained fibrils. Fibrils appear smooth and long. (ii) (A) and (B) show different magnifications and of the same area. Samples stained positively. Polymorphic fibrils were seen.
Figure 2-3 Transmission electron micrographs of self-assembled collagen fibrils in high salt (135mm NaCl) buffer and at low incubation temperature (26°C). Samples for electron microscopy were taken after plateau is reached in turbidity curve studies (see above) and stained with 1% phosphotungstic acid. A negative staining pattern was observed in the fibrils. The fibrils formed were long and uniform in diameter. Also, some non-banded fibrils were observed in the background. (A), (B) and (C) show different areas of the grid showing fibrils.
Figure 2-4 Transmission electron micrographs of self-assembled collagen fibrils in low salt (67.5 mm NaCl) buffer and high incubation temperature (37°C). Samples for electron microscopy were taken after plateau is reached in turbidity curve studies (see above) and stained with (i) 1% sodium phosphotungstate and (ii) 1% phosphotungstic acid. The banding pattern observed in the fibrils is similar to that of fibrils formed in vivo. (i) And (ii), (A) and (B) show different magnifications and different fibril area. Both show the presence of a 67nm D-period on the fibrils. Fibrils appear smooth and long.
Figure 2-5 Transmission electron micrographs of self-assembled collagen fibrils in low salt (67.5mm NaCl) buffer and low incubation temperature (26°C). Samples for electron microscopy were taken after plateau is reached in turbidity curve studies (see above) and stained with 1% sodium phosphotungstate. The banding pattern observed in the fibrils is similar to that of fibril formed in vivo. (A), (B) and (C) show different areas of the grid and different fibril areas. The presence of a 67nm D-period on the fibrils can be seen clearly. Fibrils appear smooth, long and some non-banded fibrils in the background are also seen.
2.4 Conclusion

The fibril formation experiments using commercial rat tail collagen were conducted with two different neutralization buffer compositions and at two different temperatures. The fibril formation appears to be faster at high temperature. The fibril formation at high salt buffer conditions also appear to be faster than at low salt, but the differences disappeared at low-temperature incubation. No significance difference was observed in the morphology of the fibrils, but the D-periods of the fibrils in low salt buffer were better defined compared to those incubated in high salt buffer. At low salt conditions, there were also less salt deposits and less overall background on the TEM grids. Based on these observations, we decide to use the low salt neutralization buffer to study the fibril formation process of the designed peptides. Both 37°C and 26°C appear to be adequate for fibril study. The fibril formation for Col108 would be studied at both high (37°C) and low (26°C) temperature.
CHAPTER 3

Col108 - a designed collagen-like fibril-forming peptide

3.1 Introduction to Sequence design of Col108

Col108 is a designed collagen-like peptide that contains a total of 417 amino acid residues. Out of these 417 residues, 378 amino acids residues are in a non-interrupted Gly-X-Y repeating sequence – the Gly-X-Y domain. The triple helical domain of Col108 has an inbuilt sequence periodicity, i.e. it contains three segments of pseudo-identical amino acids sequences. Each segment, herein called a sequence unit, contains 108 residues, referred to as the ‘Col domain’ (hence the peptide is named Col108), which is a composite of residue sequence selected from the alpha I chain of type I collagen; they are residues from positions 242 through 256, 296 through 322, 434 to 478 and 515 to 535 of the collagenous domain. These residues were selected based on their relatively high propensity for triple helix stabilization according to the work of Brodsky et al.\textsuperscript{65}.

Each sequence unit also contains a (Gly-Pro-Pro)\textsubscript{4} sequence the N-terminal to the Col domains. The purpose of incorporating the (Gly-Pro-Pro)\textsubscript{4} was to increase the stability of the peptide\textsuperscript{65}. An additional (Gly-Pro-Pro)\textsubscript{4} was inserted at the C-terminal of the entire Gly-X-Y domain which mimics the C-terminal imino acid rich region of fibrillar collagen thought to play a role in facilitating the triple helix formation with the correct chain register\textsuperscript{101-102}. Also to ensure the formation of the stable triple helix, a sequence of (-Gly-Pro-Cys-Cys) was incorporated at both the N- and C-termini. The cysteine residues were expected to form inter-chain disulfide bonds, known as the Cys-knots to covalently link three Col108 monomer chains in a triple helix\textsuperscript{75, 103}. A foldon domain of 27 residues from bacteriophage T4 fibritin was incorporated at the C-terminus.
to serve as a nucleation site for the triple helix folding. The foldon domain forms a stable, compact three-stranded beta barrel. The complete amino acid sequence of Col108 is shown in Figure 3-1.

In summary, the Col108 peptide is comprised of three pseudo-identical sequence units. Each unit has a total of 120 amino acid residues containing a (GPP)₄ and a 108 residue ‘Col domain. The sequence has one additional set of (GPP)₄ at the C-terminus, and the central repeating unit contains an additional pair of the Gly-X-Y triplet (GSR and GTP) introduced by restriction enzyme sites used during the ligation of the gene.
Figure 3-1. The amino acid sequence architecture of peptide Col108; (i) Four different residue sequence selections from alpha-1 chain of type I collagen indicated by residue position and identity of the corresponding 36 Gly-X-Y triplets in Col domain. (ii) Col108 sequence design corresponding to gene construct; three different amino acids sequence sets encoded by corresponding codon sequences which are ligated using restriction endonuclease enzyme site; the three repeating units of Col108 are marked with double headed arrows. The GS, SR and GT in boxes are the residues coming from the restriction enzyme sites. (iii) The amino acid sequence of Col108 shown in detail.
3.2 Construct of Recombinant plasmid for Col108

The expression plasmid was constructed by ligating three separately synthesized DNA sequences. All sequences were acquired from GenScript Corp. The three sequences were designed for the convenience of the gene manipulation and were not in the exact correspondence to the three sequence units in Figure 3-1. The three genes were flanked by restriction enzyme cutting sites used for gene ligation. In order to avoid any interruptions in the Gly-X-Y repeating sequence pattern of the gene product, additional bases were inserted in conjunction with the restriction enzyme sites, and in combination, these sequence inserts would translate to a Gly-X-Y amino acid triplet.

The first gene encodes for a Cys-knot (-GPCC), a (GPP)$_4$ repeat, the first Col domain, and a second set of (GPP)$_4$ repeats. It also includes a BamH1 restriction site at the 5’-end (___GGATCC) and a (GGCTCTAGA) nucleotide sequence at the 3’-end encoding the amino acid sequence Gly-Ser-Arg; corresponding to an Xba1 restriction enzyme site.

The second gene encoded for the second Col-domain and was flanked by an Xba1 site at the 5’-end and the sequence GGTACCCCG at the 3’ end; corresponding to a Kpn1 site, and the final base sequence is ultimately translated into the amino acid triplet Gly-Thr-Pro.

The third gene supplied a Kpn1 restriction site at the 5’-end (codon sequence GGTACCCCG), the third set of (GPP)$_4$, the third Col-domain, the C-terminal (GPP)$_4$, a sequence for a Cys-knot, the foldon domain, and an EcoR1 site at the 3’-end.

The three genes (or Gene 1, Gene 2 and Gene 3) were treated with a corresponding restriction enzyme and ligated to form the final COL108 gene having BamH1 and EcoR1 restriction sites on the 5’-and 3’-ends, respectively. The resulting gene construct is shown in
Figure 3-2. The gene insert was cloned into a modified pET32a(+) plasmid between the BamH1 and EcoR1 sites. In this modified plasmid, the initial Thioredoxin and His-tag positions in the original pET32a(+) were reversed, and the 3’-end His-tag was removed. The sequence of Col108 gene was confirmed by DNA sequencing using primers from 5’-end (T7-promotor primer) and 3’-end (T7 terminator primer) of the gene.
Figure 3-2. The gene construct of peptide Col108. The three genes and corresponding to the amino acid sequence are shown in red, blue and green boxes. Respective restriction enzyme sites flanking each gene at 5’ and 3’ are also shown in the boxes confirming the ligation mechanism used in synthesizing gene of Col108. Stop codon EcoRI is shown at 3’ end of the gene.
CHAPTER 4

Optimization of the Expression and the Purification of Col108

4.1 Introduction

The peptide Col108 was expressed in *E. coli* under the control of the *Lac* operon. The direct gene product of Col108 is a fusion protein that includes a His-tagged thioredoxin at the N-terminus and a thrombin cleavage site for the removal of the tag (Figure 4-1). The purification of Col108 involves three steps: (i) separation of the His-tagged fusion protein from the cell lysate; (ii) enzymatic cleaving of the His-tagged thioredoxin from Col108 peptide using thrombin; and (iii) the final separation of the Col108 peptide from the thrombin and the His-tagged thioredoxin by High performance liquid chromatography (HPLC). The purified Col108 is then, stored as lyophilized powder. The buffers, the cell line for expression and other conditions were optimized to ensure rapid and efficient purification.
Figure 4-1 Amino acid sequence of fusion form of Col108. Fusion Col108 contains a polyhistidine tag at N-terminus (shown in brown colored text; followed by thioredoxin (shown in black colored text) and thrombin cleavage site (shown in pink colored text) linking His-tagged thioredoxin to Col108 sequence. A foldon domain is present at the C-terminal of sequence (shown in bold black colored text). The sequence of Col108 includes two Cys-Knots at both N-and C-terminal (Aqua colored text); three repeating units i.e. ‘Col domain’ of 108 residues (shown in red, green and blue colored text).
4.2 Methods

4.2.1 Expression and cell lysis

A plasmid containing recombinant Col108 gene was transformed into bacterial expression strains. The expression includes the following steps. In the first inoculation, the 10mL culture containing antibiotic ampicillin (10μL of 50mg/mL stock solution) (Amp+) was incubated overnight in the shaker set at 200 rpm at 37°C. The next day, the 10mL culture was inoculated into 1L ampicillin-containing LB media (1 mL of 50mg/mL stock solution, AMP+ LB media). The growth was monitored by determining the optical density of the culture at 600nm using DU800 spectrophotometer. Once the optical density reached a value between 0.4-0.6, 100μL of 1M isopropyl-β-D-thiogalactoside (to a final concentration of 0.1mM IPTG) was added to induce the expression of Col108, and the culture was grown in the shaker at 300rpm, at 25°C for overnight (~16-18 hours)

In the next step, the cells of a 1-liter liquid culture were centrifuged at 2000rpm for 20 minutes at 4°C (Beckman Coulter centrifuge with a rotor the JA-12 conical rotor, serial #03U1169). The supernatant media was removed, and the collected cell pellet was transferred into a 50mL disposable centrifuge tube. The cells were suspended in 20mL of chilled phosphate wash/extract buffer (50mM Na₂HPO₄, 150mM NaCl buffer, pH 8.0); 1mL of 20mg/mL of freshly prepared lysozyme solution was added to dissolve the cell wall. Sonication was used to lyse the cell (Vibracell, three 60 second pulses, machine output 3, duty cycle 30%, microtip limit 4). The cell lysate was kept on the ice whenever possible. To prevent the proteolysis of Col108, 0.5mL of freshly made protease inhibitor solution (0.2M PhenylMethylSulfonyl Fluoride (PMSF) in dry ethanol) was added after completion of sonication to a final concentration of 0.1mM of PMSF.
The cell lysate was centrifuged at 5000 rpm for 30 minutes at 4°C using a JA-12 conical rotor, serial #03U1169 to separate the fusion Col108 containing supernatant from lysed cell debris.

4.2.2 Purification of His-tagged fusion protein using His-affinity column

The polyhistidine tag, consisting of six His- residues (His)_6 on the N-terminus of the protein, enables the purification of the recombinant peptide Col108, using widely employed, immobilized metal-affinity chromatography (IMAC) with commercially available Ni^{2+}-affinity resin (Qiagen, Cat No./Id: 30310). The resin has nickel-nitrilotriacetic acid (Ni^{2+}-NTA) matrix coupled to supporting matrix of Agarose. The transition metal ion (Ni^{2+}) forms a coordinate bond with electron donor groups available on histidine residues’ imidazole rings. The presence of polyhistidine sequences, i.e. consecutive histidine residues, ensures efficient retention of tagged protein on the column of Ni^{2+}-NTA resin. The purification was performed under native conditions following protocols suggested by Qiagen\textsuperscript{106}. The wash/extract buffer is a pH 8 phosphate buffer with no imidazole (50mM Na_2HPO_4, 150mM NaCl).

The resin was prepared by transferring 2mL of well suspended Ni^{2+}-NTA metal affinity resin slurry into a 50mL disposable centrifuge tube and spun down in the centrifuge (Eppendorf centrifuge 5430R, F-35-6-30, CE 01167) at 2000 rpm for 5 minutes. The supernatant was removed, and the resin was resuspended in 10 mL of wash/extract buffer. The re-suspended resin was centrifuged again at 2000 rpm at 4°C and supernatant was removed. This step was repeated one more time to ensure the saturation of the resin with the wash/extract buffer.

The resin was then mixed with the cell lysate containing the fusion Col108, sealed with parafilm and kept on the shaker in at 4°C for 3 hours for the His-tag to bind to the resin. After binding. The suspension was transferred into a clean gravity flow column which was rinsed with
10 mL of Wash/extract buffer. The resin was left to settle in the column, and the supernatant was allowed to flow through and was collected in a 50 mL disposable centrifuge tube and labeled as a non-binding (NB).

The resin bed in the column was then washed with 20 mL of wash/extract buffer containing 10 mM imidazole and collected as wash 1 (W1). The resin bed was again washed with 20 mL of Wash/extract buffer containing 50 mM imidazole and collected as wash 2 (W2).

The fusion Col108 was finally eluted in fractions of 1 mL with 10-12 mL of wash/extract buffer containing 300 mM Imidazole. The collected samples and fractions were analyzed using SDS-PAGE on a 12% acrylamide gel. The molecular weight of protein was compared to high range and low range protein molecular weight markers. The bands in high range molecular weight marker (Sigma-Aldrich, Cat# S8320) corresponds to 200 kDa, 116 kDa, 97 kDa, 66 kDa, 55 kDa, 45 kDa, 36 kDa (top to bottom). The bands in low range molecular weight marker (Sigma-Aldrich Cat# M3913) correspond to 66 kDa, 45 kDa, 29 kDa, 24 kDa, 20 kDa, 14.2 kDa (top to bottom).

The fractions showing a significant amount of protein were pooled and dialyzed against a phosphate wash/extract buffer to remove imidazole. The dialysis was carried out in the refrigerator for one day, with one change after 4-6 hour. The ratio of the protein sample (in mL) and dialysis buffer was 1:100.

The Tris wash/extract buffer (pH 7) was also used in place of the phosphate wash/extract buffer. The Tris buffer has 50 mM Tris, and pH of Tris buffer was adjusted to be 7, the concentration of NaCl was kept same as that of the Phosphate wash/extract buffer i.e. 150 mM.
The Tris buffer was prepared with different concentrations of imidazole: 0mM, 10mM, 50mM, and 300mM.

4.2.3 Thrombin cleavage.

The powder form of human Thrombin enzyme was purchased from Thermo Fisher (cat number); each vial contains 120.4 NIH units of the enzyme, and each NIH unit equals to 0.324µg of the enzyme (120.4 NIH=39µg). The enzyme was dissolved in 100µL of 50mM Tris, 150mM NaCl (pH 7) buffer. Based on the instruction manual the enzyme is effective in a 1:500 enzyme to protein ratio, i.e. one vial of human thrombin enzyme powder can cleave ~19mg of protein.

The samples of fusion-Col108 after Ni²⁺-affinity column were estimated to be ~ 2.6mg/mL concentration. The concentration of protein in the sample was estimated by measuring absorbance at 280nm using Beckman Coulter DU800 spectrophotometer with extinction coefficients of 1mg/mL=0.232AU. The extinction coefficient was calculated by using ExPaSy ProtParam tool. To optimize the thrombin enzyme to fusion protein ratio, as well as the optimal temperature and incubation time conditions for removal of His-tagged thioredoxin from Col108, in five Eppendorf tubes each containing 20 µL of peptide sample and 4µL, 1µL, 0.1 µL, 0.01 µL and 0.001 µL of enzyme solution were added, respectively, and incubated for one (18-20 hours) and two overnights (36-40 hours) at 4°C or at room temperature. The amount of the fusion protein in each of the sample solutions was ~52µg; the amount of thrombin are, respectively, 1.56µg (4µL), 0.39µg (1µL), 39ng (0.1µL), 3.9ng (0.01µL) and 0.39ng (0.001µL). The respective enzyme to protein ratio for each sample are calculated to be 1:33, 1:133, 1:1,333, 1: 13,333 and 1:133,333.

In another trial, samples were prepared with 1µL (0.39µg), 2µL (0.78µg), or 3µL (1.17µg) of the enzyme added to 10µL of protein sample (26µg) and incubated for one overnight (~18-20
hours) at 4°C. The ratio of thrombin to the fusion protein of these samples are, respectively: 1:67, 1:33 and 1:22. The SDS-PAGE analysis was used to study the results of thrombin cleavage.

4.2.4 Separation of His-tagged thioredoxin from Col108 using HPLC

The Col108, human thrombin enzyme, and the His-tagged thioredoxin-containing mixture solution were run on high-performance liquid chromatography (HPLC - Beckmann Coulter System Gold 126 solvent module, detector 168, SC100 fraction collector running 32 Karat 7.0 software) with reverse phase C8 analytical column (Vydac, Cat# 208TP52) first.

For analytical analysis, 20µL of fusion Col108 solution; human thrombin enzyme solution, and Col108 solution after thrombin cleavage were run on the analytical column separately with a flow rate of 0.2mL/minute. The analytical program measured the acetonitrile gradient, which increased linearly from 5% to 95% in 90 minutes starting from 5th minute onward till 90th minute with 1% increase per minute (Figure 4-2.1).

For the preparative column, the Col108 sample was concentrated in a centriprep tube with a molecular weight cutoff membrane of 3kDa (Amicon Ultra centrifugal filter, Cat No.UFC900324) in a fixed angle rotor at 4000X g for 30 minutes at 4°C to reduce the sample volume to 2-5mL and run on reverse phase high-performance liquid chromatography (RP-HPLC). The concentrated Col108 mixture containing human thrombin enzyme and His-tagged thioredoxin was run on reverse phase C8 column (Vydac Cat # 208TP1010) with a time program graph where the initial acetonitrile gradient changes from 5% to 20% (15% increase) in 20 minutes, followed by 1% acetonitrile gradient increase per 2 minutes, changing the % acetonitrile gradient from 20% to 75% in a total time of 110 minutes with a flow rate of 2ml/minute (Figure 4-2.2). To decrease the total run time on the column from 110 minutes, and increase the elution time gap between
Col108 and Human thrombin enzyme, the time program graph was changed. In the modified program, initial acetonitrile gradient change from 5% to 30% is done in 15 minutes followed by the gradient change from 30% to 45% in 30 minutes (1% per two minutes) and from 45% to 70% in 25 minutes (1% per minute) as shown in Figure 4-2.3.

In a preparative column run as shown by time program graph in Figure 4-2.2 and 4-2.3, 40 elution fractions with each fraction volume to be 1.5mL were collected after the 21st and the 22nd minute of the run, respectively. The presence of protein in collected fractions was confirmed by running the collected fraction samples on a 12% SDS-PAGE.
Figure 4-2.1. Time program graph used for analytical column representing % gradient acetonitrile. A linear gradient showing acetonitrile % change from 5% to 95% in 90 minutes (1% increase per minute).
Figure 4-2.2. Time program graph used for preparative column representing % gradient acetonitrile. Initial change of acetonitrile gradient from 5% to 20% in 20 minutes (1% increase per minute) followed by increase in acetonitrile gradient from 20% to 75% in 110 minutes (1% increase in acetonitrile gradient per 2 minutes).
Figure 4-2.3. Modified time program graph used for preparative column representing change in % gradient acetonitrile. Initial change of acetonitrile gradient from 5% to 30% in 15 minutes (~1.67 % increase in acetonitrile gradient per minute) followed by increase in acetonitrile gradient from 30% to 45% in 30 minutes (0.5% increase in acetonitrile gradient per minute), followed by increase in acetonitrile gradient from 45% to 70% in 25 minutes (1% increase in acetonitrile gradient per minute).
4.3 Results: The optimization of expression and purification.

4.3.1 The effects of the expression strain and the wash/extract buffer on the purification of the His-tagged, fusion form Col108.

In order to choose the optimum condition for the expression and purification of the His-tagged fusion form Col108, the Col108 was expressed using the two commonly used bacterial expression strains of pET32a(+) plasmid: the *E. coli* strain JM109 and the *E. coli* strain BL21(DE3); and the purification by Ni$^{2+}$-affinity resin were carried out using two different kinds of wash/extract buffer: Tris buffer (pH 7) and phosphate buffer (pH 8). The results are shown in Figures 4-3.1 to 4-3.5.

The results in Fig 4-3.1 show the typical purification outcome. The protein was expressed in JM109 cell. The fusion protein can be purified in either the monomer form (the predominant form, the red box, lanes 9-14) or the trimer form (the orange box, lanes 10-11). The His-tagged Col108 can be seen only in the supernatant (lane 2) but not in the cell pellet (lane 3). The streaks in both lanes are often associated with the precipitation of protein and/or the concentration of the protein was too high. It was concluded that even the His-tagged fusion Col108 is a large protein, it is expressed in soluble form. The fusion form of Col108 was first purified using Ni$^{2+}$-affinity resin. In this particular trial, some protein is present in non-bound (lane 5) but wash 1 and wash 2 (lane 6 & 7) did not show any traces of fusion Col108, indicating the binding with Ni$^{2+}$-NTA resin was effective. Fusion Col108 was eluted with 300mM Imidazole in phosphate wash/extract buffer. It was clear that eluted fractions contain the monomer (lanes 9-14) and the trimer form (lanes 10-12) of the protein. Several strong bands were observed in the low molecular weight range (the black box, lane 10-13). The identification of these bands was not made, but it was assumed that these bands were non-specifically bound proteins. Non-specific binding was a common drawback of
Ni²⁺-NTA resin. Alternatively, these bands could be fragments of the His-tagged Col108 due to incomplete translation during expression or due to protease degradation during purification.
Figure 4-3.1. Expression of His-tagged Col108 in JM109. Lanes 1 and 7 are, respectively, the high range and the low range protein markers. Lanes 2 and 3 are the samples of the supernatant and the cell pellet, respectively; lane 4 is non-bound, lanes 5 and 6 are wash 1 and wash 2, respectively; lanes 8-14 the elution 1-7 (collected with a volume of 1 mL each). Blue, purple, yellow and red boxes highlight the band of the monomer form of Col108 (molecular weight ~52 kDa); orange box highlight the trimer form (molecular weight ~156 kDa). The bands in the black box are unidentified species showing the same affinity to the Ni$^{2+}$-resin as that of His-tagged Col108.
Once certain that His-tagged Col108 can be expressed in a bacterial expression system, efforts were directed towards improving the purification conditions by comparing different wash/extract buffers and different *E Coli* expressions strains.

After expressed in JM109, the His-tagged Col108 were purified using two different buffers, and the results are as shown in Figures 4-3.2 and 4-3.3. In Figure 4-3.2, no His-tagged Col108, monomer or trimer form, was seen in the cell pellet using Tris (pH 7) wash/extract buffer (lane 2) pointing towards the solubility of fusion Col108 and a complete breakdown of the *E.coli* cells. The cell pellet using phosphate (pH 8) buffer showed the presence of trimer form of the protein (orange box, lane 3); the monomer of fusion Col108 was not clear due to smearing (red box, lane 3). The presence of trimer in cell pellet sample could be due to the incomplete breakdown of cells during sonication step of purification. Further, viewing lanes 4 and 5, no band corresponding to molecular weight of fusion Col108 monomer or trimer was seen in the non-binding samples, whether using Tris or Phosphate, indicating efficient binding of fusion Col108 present in a cell lysate to Ni$^{2+}$-NTA resin. In lane 5, many non-specific protein molecular weight bands were seen. Focusing more specifically on the area on lane 5 where fusion Col108 band for monomer was anticipated, it was unclear if monomer was present or not since there was a cluster of bands found in the specified area (the black box, lane 5). The presence of fusion Col108 monomer (the purple box, lanes 6 and 7) was seen in the first wash using Tris (pH 7) or phosphate (pH 8) wash/extract buffers containing 10mM imidazole Trace amount of protein were also seen in the second wash (yellow box, lane 8 & 9). The elutions using Tris or Phosphate buffers were shown in lanes 11 – 14. The presence of fusion Col108 monomer in elutions 1 and 2 in phosphate (pH 8) buffer are shown in the green boxes (lanes 12 and 14). Some high molecular weight bands were also observed in lane 12, 13 and 14 indicating the presence of dimers (pink box). Similarly, elutions collected in Tris wash/extract
buffer (pH 7) containing 300 mM imidazole ((lanes 11 and 13) showed the presence of fusion Col108 monomer (white boxes).

Similarly, in **Figure 4-3.3**, elutions collected using Tris wash/extract buffer (pH 7) containing 300 mM imidazole (lanes 2, 4, 6, 8, 10, 12 & 14). The presence of Col108 monomer was shown in white boxes. And elutions collected using phosphate buffer with 300mM imidazole showed the presence of Col108 monomer (green boxes, lanes 3, 5, 7, 9, 11 & 13). Some bands in high molecular weight range were seen (shown in the pink box), indicating the presence of either dimer of fusion Col108. Also, some low molecular weight bands around 14kDa area were seen (the black box).

Comparing the elution collections using Tris or phosphate wash/extract buffers, it was observed that bands representing fusion Col108 monomer in tris buffer were much thicker. This difference is especially obvious in Fig 4-2B. We, therefore, decided to change the wash/extract buffers to be used during the purification protocol to Tris (pH 7).
**Figure 4-3.2** Effects of wash/extract buffer on the purification of His-tagged Col108 expressed in JM109.

The His-tagged proteins were purified using Tris (pH 7) buffer (Lanes 2, 4, 6, 8, 11 & 13) and phosphate (pH 8) wash/extract buffer (Lanes 3, 5, 7, 9, 12 & 14). Lanes 2 and 3 are cell pellet, Lanes 4 and 5 are the non-binding, lanes 6 and 7 are wash 1, lanes 8 and 9 are the wash 2, lanes 11 & 12 are the elution 1, lanes 13 & 14 and the elution 2. Lanes 1 and 10 are, respectively the high range and the low range protein marker. The orange, yellow purple, white and green boxes highlight the monomer form of His-tagged Col108, the magenta box identifies the higher order of aggregates of His-tagged Col108.
Figure 4-3.3 (continuing from Fig. 4-3.2) Elution 3-9 are in lanes 2 & 3 – lanes 14 &15; where samples in lanes 2, 4, 6, 8, 10, 12 and 14 are eluted using Tris buffer containing 300 mM imidazole, and samples in lanes 3, 5, 7, 9, 11, 13 and 15 are eluted using phosphate buffer (pH 8) containing 300 mM imidazole. The white and green boxes highlight the monomer form of His-tagged Col108, the magenta box highlight the oligomers, and the black box the unidentified species.
Also, after expression of fusion form of Col108 in BL21(DE3), purification steps were optimized for Tris and phosphate buffers. The results are shown in Figure 4-3.4 and 4-3.5. In Figure 4-3.4, no presence of fusion Col108 monomer or trimer in cell pellet for tris buffer was seen (lane 2), fusion form of Col108 was present only in cell lysate (monomer: the dark red and trimer: the pink box, lane 4). The non-bound sample appeared to contain a very faint band at the same molecular weight as fusion Col108 monomer (light blue box, lane 6). The first wash sample showed the presence of non-specific proteins bound to the column to be washed out with 10mM imidazole (lane 8). But, it was unclear if 10mM imidazole eluted any fusion Col108 monomer from the column in the first wash. A faint band was seen in the fusion form col108 monomer molecular weight range (the blue box, lane 8). The second wash of column sample pointed clearly towards no loss or elution of bound fusion form Col108 in the second wash (the orange box, lane 10). The presence of fusion form of Col108 monomer was seen in first eluted fraction (the tan box, lane 12). Lane 3 showed cell pellet sample in phosphate wash/extract buffer (pH 8). The presence of fusion form of Col108 monomer indicated towards improper lysis of cells during sonication (the yellow box). The sample for supernatant/cell lysate indicated the presence of fusion Col108 monomer and trimer (the green and the light green box respectively, lane 5). The non-bound sample collected after settling the resin in matrix column in lane 7 showed the clear presence of fusion Col108 monomer (the white box) and trimer (the purple box). The first wash sample showed the presence of non-specific proteins bound to the column to be washed out with 10mM imidazole (lane 9). A faint band in the fusion col108 monomer molecular weight range was seen (the blue box). The second wash sample showed no presence of fusion Col08 (the orange box, lane 11). SDS sample of first elution fraction collected in phosphate buffer showed the very faint band at an expected molecular weight range of fusion Col108 monomer (the tan box, lane 13).
In Figure 4-3.5, the presence of fusion form of Col108 monomer (the white boxes, Lanes 1, 3, 5, 7, 9 & 11) and fusion form of the Col108 trimer (the red boxes, lanes 3, 5 & 7) in Tris purification was very clear. Lanes 2, 4, 6, 8, 10 and 12 showed the presence of fusion Col108 monomer (the green boxes) and lanes 4 and 6 showed the presence of a faint band corresponding to fusion Col108 trimer (the yellow boxes) in phosphate purification. Also, few bands were observed between fusion Col108 monomer and trimer, which could be due to either degradation of Col108 fusion trimer or formation of dimers of Col108 from fusion monomers. It could also be possible that some non-specific, high molecular weight proteins with the same affinity for Ni\(^{2+}\)-NTA resin as fusion Col108 got eluted with 300mM imidazole (the pink box). Some low molecular weight bands were seen (the black box). The identity of these bands was not clear. These could be either some non-specific proteins or degraded fusion Col108. Based on the comparative analysis, both bandwidth and intensity indicated the presence of more protein in purification with Tris wash/extract buffer compared to phosphate. Also, binding to Ni\(^{2+}\)-NTA resin was more effective when we used Tris buffer as we did not lose any Col108 fusion monomer and trimer in non-bound cell lysate and washes. So we decided to switch our buffer during purification from phosphate wash extract buffer (pH 8) to Tris wash/extract buffer (pH 7).
Figure 4-3.4 Effects of wash/extract buffer on the purification of His-tagged Col108 expressed in BL21. The His-tagged proteins were purified using Tris (pH 7) buffer (Lanes 2, 4, 6, 8, 10 & 12) and phosphate (pH 8) wash/extract buffer (Lanes 3, 5, 7, 9, 11 & 13). Lanes 2 and 3 are cell pellet, Lanes 4 and 5 are the non-binding, lanes 6 and 7 are wash 1, lanes 8 and 9 are the wash 2, lanes 11 &12 are the elution 1, lanes 13 &14 and the elution 2. Lanes 1 and 14 are, respectively the low range and the high range protein marker. The red, yellow, dark red, dark green, light blue, white, blue, orange and tan boxes highlight the monomer form of His-tagged Col108, the pink, light green and purple boxes identify the Fusion form of Col108 trimer.
Figure 4-3.5 (continuing from Fig. 4-3.4) Elution 2-7 are in lanes 1 & 2 – lanes 11 &12; where samples in lanes 1, 3, 5, 7, 9 & 11 are eluted using Tris buffer containing 300 mM imidazole, and samples in lanes 2, 4, 6, 8, 10 & 12 are eluted using phosphate buffer (pH 8) containing 300 mM imidazole. Lanes 13 and 14 are respectively high and low range protein markers. The white and green boxes highlight the monomer form of His-tagged Col108; the red and yellow boxes highlight the trimer form of His-tagged Col108; the magenta box highlights the oligomers; and the black box highlights the unidentified species.
4.3.2 Optimized Thrombin cleavage

Thrombin cleavage was used to remove His-tagged thioredoxin from fusion Col108 (Figure 4-4). Several trials were conducted to find the optimum thrombin enzyme to fusion protein to ensure complete removal of His-tagged thioredoxin.

The respective enzyme to fusion protein ratio of 1:33, 1:133, 1:1,333, 1: 13,333 and 1:133,333 were used, and the incubation time were between one overnight (18-20 hours) and two overnight (36-40 hours). In Figure 4-5.1, lanes 1 and 2 showed efficient thrombin cleavage with an enzyme to protein ratio of 1:33 as indicated by the presence of a band corresponding to Col108 monomer (the tan box). The Col108 trimer was present (the blue box, lane 1). Col108 monomer and trimer bands were easily identified by comparing to, respectively, the bands of fusion Col108 monomer (lanes 11 and 12, light blue box) and trimer (shown in aqua box) and also comparing to the protein molecular weight marker (lane 13) The sample in lane 1 was incubated at 4°C for one overnight (~18-20 hours), in lane 2 at room temperature for one overnight. The bands in lane 2 appeared to shift slightly upward, but still showed the complete removal of His-tagged thioredoxin. The band for the Col108 is clearly visible, and the band for the fusion protein disappeared. There was no Col108 trimer in Lane 2. The intensity of the bands was lighter in lane 2 compared to lane 1, which could be caused by the degradation/ precipitation of Col108 due to incubation at room temperature.

Samples with an enzyme to protein ratio of 1:133 after one overnight incubation at 4°C or room temperature were observed (lane 3 and 4 respectively). The sample incubated at 4°C showed the presence of Col108 monomer and trimer (the red and the dark red box), but the sample incubated at room temperature showed the only presence of Col108 monomer (dark red box). The
intensity of Col108 monomer band in lane 4 was lesser compared to the respective band in lane 3 (the dark red box).

Lanes 5 and 6 showed the samples with an enzyme to protein ratio of 1:1333 after overnight incubation at 4°C or at room temperature, respectively. At 4°C, the incomplete thrombin cleavage due to the presence of Col108 fusion monomer (the black box). A band corresponding to the Col108 monomer (the orange box) and a very faint band corresponding to the trimer (the yellow box) was observed. Lane 6 showed completion of thrombin cleavage at room temperature as no fusion Col108 was observed on the SDS-PAGE; the Col108 monomer band was clearly seen (in the orange box).

The fusion form of Col108 monomer was observable in samples with the enzyme to protein ratio of 1:13333 (lanes 7 and 8, the dark green box) or 1:1333333 (lanes 9 and 10 in dark pink box). The trimer band of the fusion form of Col108 were also visible in lane 7 and 8, 9 and 10 (light green box and light pink box respectively). It is clear that one overnight incubation at these low concentrations of the enzyme was ineffective at 4°C or room temperature. These findings were expected since the recommended range of enzyme to protein ratio for efficient thrombin cleavage was 1:500, much higher than those used in lanes 7-10.

We also examined the digestion after prolonged incubation – two overnights at room temperature or 4°C, all together for about 36-40 hours. SDS-PAGE samples were prepared, and the results were shown in Figure 4-5.2. The findings were similar to those after 18hr incubation. The enzyme digestion was complete at ratios 1:33 (lanes 4-5) and 1:133 (lanes 6-7) after 48 hours incubation at either room temperature (lanes 4 and 6) or 4°C (lanes 5 and 7). The monomer band of the Col108 monomer (the orange box in lanes 4-5 and the dark red box in lanes 6 and 7) and trimer band (the blue box in lanes 4-5 and red box in lane 7) are clearly visible. The sample with
an enzyme to protein ratio of 1:1333 incubated at 4°C that the quantity of enzyme in protein solution was not effective for efficient thrombin cleavage (lane 9). We observed Col108 fusion monomer (shown in the black box). While no fusion protein was observed in samples with the same enzyme to protein ratio but incubated at room temperature (lane 8), the bands were so faint it is difficult to draw the conclusion. In fact, the bands of samples incubated at room temperature were all significantly lighter in intensity comparing to those at 4°C, even the samples without any addition of Thrombin (lane 2), indicating the potential loss of protein to degradation at this elevated temperature.

Once again Samples with an enzyme to protein ratio of 1: 13333 (lanes 10-11) and 1:133333 (lanes 12-13) confirmed that the quantity of enzyme in these samples is so small that even a longer incubation time was not leading to the removal of His-tagged thioredoxin from the fusion protein. Only fusion form of Col108 monomers and trimers were observed (the green and the light green boxes in lanes 10-11, the pink and the light pink boxes in lanes 12-13).

To further define the most effective and also economical level of enzymatic cleavage, we studied the effects of enzyme: protein ratio of 1:67, 1:33 and 1:22 incubated for one overnight at 4°C. While 1:133 show complete thrombin cleavage in Figure 4-5.1 and 4-5.2, but occasionally we did observe a faint band of the fusion protein. Therefore, we decided not to carry out the reaction with enzyme: protein ratio of 1:133 or lesser enzyme. The results are shown in Figure 4-5.3. A complete digestion in samples with an enzyme to protein ratio of 1:67, 1:33 and 1: 22 was clearly seen in lanes 3, 4 and 5, respectively. The Col108 monomer was shown in light blue, yellow and pink box respectively. The His-tagged thioredoxin was also clearly seen in lanes 3-5 (the black box). A band can be identified as that of the thrombin enzyme (Lane 1). This Thrombin band was
visible in lane 5 (bright red box) and in other samples having a relatively low concentration of the enzyme in Figures 4-5.1 (Lanes 1-3) and 4-5.2 (Lanes 4-7).
Figure 4.4. A schematic representation of thrombin cleavage reaction to separate His-Tagged thioredoxin from fusion Col108.
Figure 4-5.1 Optimization of thrombin cleavage after overnight (18 hours) incubation. The enzyme:protein ratio is 1:33 for lanes 1 and 2, 1:133 for lanes 3 and 4, 1:1333 for lanes 5 and 6, and 1:133,333 for lanes 9 and 10. Lanes 11 and 12 are samples of the fusion protein prior to the addition of thrombin. Samples of lanes 1, 3, 5, 7, 9 and 11 were incubated at 4°C during the digestion, and those in lanes 2, 4, 6, 8, 10 and 12 were kept at room temperature (28°C). Lane 13 is high range protein marker.
Figure 4-5.2 optimization of the thrombin cleavage after two overnights incubation (≈ 40 hours). Lanes 2 and 3 are samples of the fusion protein prior to the addition of thrombin. The enzyme:protein ratio are 1:33 for lanes 4 and 5, 1:133 for lanes 6 and 7, 1:1,333 for lanes 8 and 9, 1:13,333 for lanes 10 and 11 and 1:133,333 for lanes 12 and 13. The samples in lanes 2, 4, 6, 8, 10 and 12 were incubated at room temperature (≈ 28°C), and those in lanes 3, 5, 7, 9, 11 and 13 were incubated at 4°C. Lane 1 is the low range protein marker, and lane 15 is the high range protein marker. Lane 14 is another sample prior to the addition of thrombin incubated at 4°C for 18 hours.
Figure 4-5.3. Optimization for enzyme:protein (E:P) ratio for thrombin cleavage using lyophilized human thrombin enzyme; lane 1 showing human thrombin enzyme; lane 2 showing fusion Col108; lane 3 showing E:P ratio of 1:67; lane 4 showing E:P ratio of 1:33; lane 5 showing E:P ratio of 1:22; lane 6 showing low range protein molecular weight marker (6,500-66,000 Da). Samples in lanes 3-5 were incubated at 4°C for one overnight (~18 hours).
Even though the recommended enzyme/protein ratio is 1:500, after analyzing different thrombin cleavage set ups to find optimal conditions and ensure the complete removal of His-tagged thioredoxin from Col108, we decided to use an enzyme: protein ratio of 1:66. We also determined that one overnight incubation (~18-20 hours) at 4°C is the optimal condition for thrombin cleavage.

4.3.3 High-performance liquid chromatography as final purification step

After thrombin cleavage, the next and final step was a complete separation of Col108 from His-tagged thioredoxin and human thrombin enzyme. We first did the analytical run separately on fusion Col108, human thrombin enzyme solution, and Col108. The analytical program was then scaled up for the preparative column.

In the analytical run spectra shown for Fusion Col108 in Figure 4-6.1(i) showing absorbance spectrum at wavelength 214nm and 280nm; and (ii) showing absorbance spectrum at wavelength 280nm, we observed that fusion Col108 (His-tagged thioredoxin containing Col108) elutes at 62.5% acetonitrile gradient (~62nd minute - shown in the orange box).

For human thrombin enzyme run on the same analytical column using the same time program graph (Figure 4-2A), we observed in Figure 4-6.2(i) showing absorbance spectrum at wavelength 214nm and 280nm; and (ii) showing absorbance spectrum at wavelength 280nm that human thrombin enzyme elutes at 53% acetonitrile (~53rd minute - shown in light blue box).

Using the same time program graph for the Col108 sample after thrombin cleavage, i.e. sample containing Col108, human thrombin enzyme and His-tagged thioredoxin: the absorbance spectrum observed at wavelength 214nm and 280 nm as shown in Figure 4-6.3(i); and spectrum observed only at 280nm as shown in Figure 4-6.3(ii) showed that Col108 eluted at 46%
acetonitrile (~46th minute - shown in pink box in Figure 4-6.3(i) & (ii)), human thrombin enzyme eluted at 53% acetonitrile (~53rd minute - shown in light blue box in Figure 4-6.3(i) & (ii)) and thioredoxin eluted at 62.5% acetonitrile gradient (~62nd minute), respectively (shown in orange box in Figure 4-6.3(i) & (ii)).

For the preparative column after finding the acetonitrile % gradient for elution of Col108, human thrombin enzyme and His-tagged thioredoxin, we scaled up the program graph as shown in Figure 4-2.2 and 4-2.3, and observed the HPLC spectra showing Col108 eluting at 38% acetonitrile (~56th minute - shown in pink box), Human thrombin enzyme eluting at 44% acetonitrile (~70th minute - shown in light blue box) and thioredoxin eluting at 54% acetonitrile (~88th minute - shown in orange box) (Figure 4-7). But in the modified program, the Col 108 elutes at ~32nd minute (38 % Acetonitrile), thrombin elutes at 45th minute (44% acetonitrile), and thioredoxin elutes at the 61st minute (54% acetonitrile) (Figure 4-8).
Figure 4-6.1 (i) HPLC spectra of Fusion Col108 at 214nm and 280nm; (ii) HPLC spectrum of Fusion Col108 at 280nm; Fusion Col108 elutes at 60% to 65% acetonitrile (~ from 60 minutes to 65 minutes). Peak corresponding to fusion Col108 on the spectrum is shown in an orange box.
Figure 4-6.2 (i) HPLC spectra of Thrombin enzyme at 214nm and 280nm; (ii) HPLC spectrum of Thrombin enzyme at 280nm; Human thrombin enzyme on the spectrum elutes at 53% acetonitrile gradient (~53 minutes). Peak corresponding to human thrombin enzyme is shown in a light blue box.
Figure 4-6.3 (i) HPLC spectra of Col108 after thrombin cleavage at 214nm and 280nm; (ii) HPLC spectrum of Col108 after thrombin cleavage at 280nm: Col108 elutes at 46% acetonitrile gradient (~46 minute); human thrombin enzyme elutes at 53% acetonitrile gradient (~53 minutes) and thioredoxin elutes at 63.5% acetonitrile gradient (~63.5 minute). Peak corresponding to Col108 protein on the spectrum is shown in a pink box. A peak corresponding to human thrombin enzyme on the spectrum is shown in a light blue box. A peak corresponding to His-tagged thioredoxin is shown in an orange box.
Figure 4-7. HPLC spectra for Col108 after thrombin cleavage at 280nm. Col108 elutes at 38% acetonitrile (56th minute); Human Thrombin enzyme elutes at 44% (70th minute) and His-tagged thioredoxin elutes at 54% acetonitrile (88th minute). A peak corresponding to Col108 protein on the spectrum is shown in a pink box. A peak corresponding to human thrombin enzyme on the spectrum is shown in a light blue box. A peak corresponding to His-tagged thioredoxin is shown in an orange box.
Figure 4-8. HPLC spectra for Col108 after thrombin cleavage at 280nm. Col108 elutes at 38% acetonitrile (32nd minute); Human Thrombin enzyme elutes at 44% (45th minute) and His-tagged thioredoxin elutes at 54% acetonitrile (61st minute). Peak corresponding to Col108 protein on the spectrum is shown in a pink box. A peak corresponding to human thrombin enzyme on the spectrum is shown in a light blue box. A peak corresponding to His-tagged thioredoxin is shown in an orange box.
Figure 4-9. A 12% SDS PAGE confirming the presence of Col108 peptide in the collected HPLC fractions as shown in the black box. No low molecular weight bands were seen. This indicated towards the purity of Col108.
4.3 Conclusion

From these studies, we learned that the bacterial strain has little effects on the expression. Both JM109 and BL21 were used for the subsequent experiments. On the other hand, the Tris (pH 7) clearly show less non-specific binding to the resin, less loss of protein during binding; therefore, it became our choice of buffer. The optimal conditions of thrombin cleavage were found to be one overnight incubation at 4°C with thrombin enzyme to fusion protein ratio of 1:66. One crucial step before thrombin cleavage was to dialyze the fusion protein sample to remove imidazole from the solution. The thrombin was usually added after the first change of dialysis buffer (the change usually take place after ~ 4 hr dialysis). The HPLC appeared to be an effective method to separate Col108 from thrombin and His-tagged thioredoxin. The SDS-PAGE confirmed that final Col108 sample was pure without contamination of His-Thioredoxin or thrombin (Figure 4-9).
CHAPTER 5

The folding and the thermal stability of Col108 peptide

5.1 Introduction

The Col108 peptide is unique in the sense that it is significantly longer compared to synthetic collagen-like peptides studied so far. This peptide has a total of 417 amino acid residues, out of which 378 are in a non-interrupted collagen-like Gly-X-Y sequence. Additionally, it has a built-in sequence periodicity arising from three pseudo-identical sequence units placed in tandem; each unit contains ~123 amino acid residues (Figure 3-1). After successful purification of Col108, the next goal was to ensure that the peptide folded into a triple helix and is stable under the experimental conditions for further studies.

5.2 Methods

5.2.1 Preparing Col108 solutions

Lyophilized Col108 powder was dissolved in a 5mM acetic acid (HAc, pH 4) solution to a concentration of ~1mg/mL. The concentration was determined by measuring absorbance at a 280nm wavelength with an extinction coefficient of 0.232AU (1mg/mL). The extinction coefficient was calculated by the ExPaSy Protparam tool. The samples were equilibrated in the refrigerator for ~10 days to ensure triple helix formation. The conformation was checked by the wavelength scan using circular dichroism (CD) spectroscopy. Samples for the wavelength scan were prepared by diluting to a concentration of 0.2mg/mL. The wavelength scans were also performed on Col108 fibrils in fibril-forming buffer (2.5mM HAc, 30mM TES, 30mM Na₂HPO₄ and 67.5mM NaCl, pH 7.4). The concentration of the samples were adjusted to 0.2mg/mL from 0.5mg/mL. The samples used for thermal stability studies were prepared at a concentration of
1mg/mL or 0.5mg/mL in 5mM HAc and fibril-forming buffer, respectively. The samples were further equilibrated for ~2-4 days after each dilution before any CD data were taken. Once folded, the conformation was stable. The spectra of wavelength scans are highly reproducible even after extended storage at 0.2 mg/mL in the refrigerator.

5.2.2 Conformation Studies

The triple helix conformation of Col108 was determined using a circular dichroism spectrometer (AVIV Biomedical, Model 202-01). The wavelength scans were conducted at 4°C between wavelength 190 and 300nm, with the 1nm interval. Hellma quartz CD cuvettes (1mm pathlength) were used. CD wavelength data was normalized to a mean residue ellipticity (MRE), the molecular weight of 38.156 kDa and the number of residues (417) per single chain of Col108 were used.

5.2.3 Thermal studies

The thermal unfolding was studied to determine the stability of Col108 in different buffers. The temperature melt experiment was conducted by heating the sample from 4°C to 65°C with an equilibration time of two minutes at each temperature; the estimated average heating rate is 0.3°C/minute, and the unfolding of Col108 was monitored at 225nm. The data was normalized to the fraction of folded peptides, as described in Xu et al.108

5.3 Results

5.3.1 Conformation of peptide Col108

Col108 adopts a triple helical conformation at 4°C in 5mM HAc (pH 4) as shown by the characteristic CD spectra with a positive peak at 225nm and a low, negative peak at 197nm (Figure
Spectra was also observed in pH 7.4 buffer. The ratio of positive to negative peak intensity, known as the ‘Rpn’ of Col108 is ~0.067. This Rpn is lower compared to the Rpn value observed for (POG)$_n^{109-110}$ but is in good agreement with that of the peptide (GPP)$_{10}$, (GPP)$_{10}$-foldon, and other synthetic triple helical peptides with high (Gly-Pro-Pro) content$^{111-113}$. 
Figure 5-1 Circular dichroism spectra of Col108 at 0.2mg/mL and 4°C in 5mM acetic acid (pH 4; open blue circles) and in fibril forming buffer (pH 7.4; open red triangles).
5.3.2 Thermal stability of peptide Col108

Col108 triple helix is quite stable with a melting temperature of ~41°C in pH 4 buffer and pH 7.4 buffer (Figure 5-2).

5.4 Conclusion

Despite lacking hydroxyproline at Y-position, Col108 folds into a stable triple helix with a $T_m$ value of 41°C. The higher than usual content of charged residues in the Col domain, the presence of Cys-knots at both N- and C-termini, the regular inserts of (GPP)$_4$ and the foldon domain could be the sequence design features of Col108 that are contributing to its thermal stability. The Col108 exists as triple helices in pH 4 buffer, but form mini-fibrils in the fibril-forming buffer as will be described in the next chapter. However, the thermal stability is the same under both conditions.
Figure 5-2 **Temperature melt curve of Col108** at 1.2mg/mL and 4°C in 5mM acetic acid (pH 4; open blue circles) and in fibril forming buffer (pH 7.4; open red triangles).
CHAPTER 6

The self-assembled Col108 mini-fibrils*

*This chapter is adapted from manuscript Kaur et al.115

6.1 Introduction

Collagen is studied for its versatility and involvement in a wide range of cellular functions during tissue development. This versatility of collagen made it: a successful scaffold to be studied extensively for biomedical applications like tissue engineering116-117 and in vitro substrate for studying phenomena like cell differentiation118, adhesion119, and migration120. Also, collagen has been extensively studied to understand the role of mutations in structure and/or synthesis of this protein leading to diseases such as Osteogenesis imperfecta and Ehlers-Danlos syndrome121-123. To understand the structure, functions, and interactions involved in the self-assembly of this extensively abundant protein in the body, self-assembling collagen-mimetic peptides that model specific properties of collagen fibrils have been sought after by researchers.

Many of the reported self-assembled collagen mimetic peptides have often relied on the introduction of chemical moieties at the ends of the triple helices. The self-assembled fibrils observed in these cases lacked the specific structural features of native collagen fibrils5, 7, 9, 47, 123-125. The one exception is the D-periodic microfibers formed by the self-assembly of a 36-residues synthetic collagen-mimetic peptide reported by Rele et al.3. The self-assembled microfibers revealed the banding pattern of 18nm which is larger than what is the anticipated end-to-end distance of 10-12nm predicted for a 36-residue long triple helix. Also, the microfibers were formed only after thermal annealing, i.e. unfolding of triple helices. So the underlying mechanisms of the
observed self-assembly were assumed to be significantly different from the staggered lateral packing of the triple helices as observed in native collagen fibrils.

In brief, it can be stated that even though functional collagen-mimetic peptides have been developed in the past, such studies rely only on short peptides which are usually 30-35 residue long and also, many of these bottom-up assemblies employ techniques like chemical cross-linking, native chemical ligation and thermal annealing. Therefore, a major limitation of such systems is that it is very difficult to ascribe a direct association between designed and native when it comes to defining the molecular mechanisms involved.

Therefore, our goal was to develop a collagen-like peptide that does not involve the external chemical moieties but includes the features which are inherent to the native collagen. We specifically focused on the sequence periodicity and engaged it in our design to accomplish a two-fold goal: to develop a scaffold which not only mimics the native collagen fibril under the defined optimal condition of self-assembly but also gives insight into the underlying molecular mechanisms of self-assembly. The significance of this system is that it can be used as a potential functional scaffold in tissue engineering and regenerative medicine.

After triple helical conformation of Col18 peptide had been confirmed by CD spectroscopy, our next goal was to study if Col108 self-assembles into a higher-order assembly, which bears any resemblance to native collagen fibrils.

6.2 Methods

6.2.1 Fibril formation of Col108

The fibrillogenesis experiment was set up using 1mg/mL of Col108 in 5mM HAc (pH 4) kept at 4°C. The sample was then mixed in 1:1 volume ratio with double strength neutralization
buffer (2X buffer: 30mM TES, 30mM Na₂HPO₄, 67.5mM NaCl; pH 7.4). The final concentration of Col108 was 0.5mg/mL in the fibril-forming buffer, which has a composition of 2.5mM HAc, 30mM TES, 30mM Na₂HPO₄ and 67.5mM NaCl composition (pH 7.4). The neutralized Col108 solution was incubated at 37°C or 26°C. The pH was determined in a separate test trial using fibril forming buffer, formed by mixing an equal volume of pH 4 HAc and pH 7.4 2X neutralization buffer.

6.2.2 TEM characterization of Col108 mini-fibrils

TEM grids of Col108 fibrils were prepared at 6 hours and 24 hours after fibrillogenesis initiation at 26°C or 37°C using 400 mesh Formvar-carbon coated copper grids. 3µL of incubated sample was allowed to sit on the grid for 100 seconds. The grids were rinsed with deionized water for 5 seconds and stained with 3µL of 1% sodium phosphotungstate for 100 seconds. The grids were rinsed again with deionized water for 5 seconds. The grids were air-dried before being examined under a Jeol 2100 electron microscope (Jeol Inc.). The collected images were analyzed using ImageJ software.¹²⁶

The axial periodicity was determined by manually measuring the negatively stained images using a ruler on the original picture printout of the micrograph. Approximately, 40 such measurements from 4 separate TEM images (two from a sample after 6 hours of incubation at 37°C and two from a sample after 24 hours of incubation at 37°C) were taken. The periodicity of the mini-fibrils formed at 26°C was determined based on 11 such measurements taken from two separate TEM images after 6 hours of incubation.

6.2.3 AFM characterization of Col108 mini-fibrils
For tapping mode AFM imaging, samples were prepared on a freshly cleaved mica substrate. A 10µL volume of Col108 sample after overnight incubation was placed directly on the substrate. The sample was rinsed gently with deionized water after drying. AFM imaging of the substrate was carried out using NSC15/AlBS probes from Mikromasch in the tapping mode (spmtips, nominal radius ~8nm, force constant 40N/m and resonance frequency 325 kHz). All the imaging was carried out using a TM-AFM and scanner (Agilent 5500). Random locations of the sample were imaged. The image analysis and measurements were performed using WSxM 5.0 software. The axial periodicity was measured as the distances between the peaks of the phase retrace plots based on ~20 measurements.

6.3 Results

Upon incubation of Col108 in the fibril-forming buffer at 26°, or 37°C, the self-assembled aggregate were examined using TEM and AFM. The fibrils were examined after 6 hours or 24-hours incubation at 26°C or 37°C. Analytical ultracentrifugation study has indicated the self-associated aggregates are formed in solutions.

6.3.1 The mini-fibrils of Col108 with ~35nm periodicity

Under TEM, the assembled aggregates of Col108 were observed to be fibrillary. The fibrils appear to have variation in diameter and length. Most fibrils after 6 hours of incubation are 400-600nm long and have a diameter of about ~50nm (Figure 6.1-3). Fibrils formed after overnight incubation are longer (800nm -1µm) and have a diameter mostly in the range of 75nm with some having a width of 100nm (Figure 6.1 to 6-3). The exact length of some fibrils is difficult to determine because the ends of individual long mini-fibrils were difficult to discern.
The most striking observation is the banding pattern on the fibrils under all the conditions examined, irrespective of variations in length and diameter. Under negative staining fibrils formed under different conditions, all showed a repeating banding pattern of a dark and light band repeats of ~35nm; very similar to the $D$-period of 67nm of fibrillar collagen comprised of one dark and light band. The dark band in Col108 fibril is 25nm and light band is 10nm observed through the entire length of the fibril (Figure 6-1(i), 6-2(i), 6-3(i)B). The average value of axial periodicity of a mini-fibril formed at 37°C was determined to be 34.6nm. The mini-fibrils formed after incubation at 26°C have a similar average periodicity of 34.2nm. This ~35nm axial periodicity was named the “$d$-period” because of its comparable yet different features from the $D$-period of type I collagen fibril. The comparable banding features of collagen fibrils (ref) and Col108 fibrils led us to conclude that Col108 fibrils have a similar alternating gap and overlap regions as collagen, except the periodicity, is ~35nm.

The Col108 fibrils stained positively show different banding pattern under the same assembly for different TEM grids. (Figures 6-1(ii), 6-2(ii), 6-3(i)A and (ii)). The pattern can be characterized by three pairs of alternating thick and thin dark bands, intercalated with light bands along the length of one triple helix of Col108, which is ~120nm long. Unlike the negative staining, in a positive staining pattern, the dark bands represent the direct binding of heavy metal ions to the charged residues on the surface of fibrils. The banding pattern observed on the surface of Col108 fibrils reflects the finer distribution of charged residues on the fibrils, and reflect an ordered arrangement of triple helices. Any random arrangement of triple helices with misaligned charged zones would result in an undefined distribution of charges regions on the surface of the mini-fibril, and will, therefore, lead to an undefined non-specific staining appearance. Our observations, on the contrary, were indicating a defined and specific positive staining pattern on mini-fibrils of
Col108 pointing towards a specific arrangement of triple helices of Col108. In this arrangement, there are three pairs of thick-thin dark bands every ~100- nm, which is the approximate length of the triple helical domain of Col108 triple helix.
Figure 6-1 The electron micrographs of Col108 mini-fibrils; (i)A-E, the negatively stained mini-fibrils (A-E); (ii)A-B, positively stained mini-fibrils. Samples are prepared from the neutralized Col108 solution incubating for 6 hours at 37°C. Different measurements for d-period are shown in proportion to the scale bar are shown in all the images. The scale bars are 200nm (i)-A, D, and E and (ii)-A and B), 100nm (i)-B and C.
Figure 6-2 The electron micrographs of Col108 mini-fibrils; (i) A-D, the negatively stained mini-fibrils (A-D); (ii) A-D, positively stained mini-fibrils. Samples are prepared from the neutralized Col108 solution incubating for 24 hours at 37°C. Different measurements for d-period are shown in proportion to the scale bar are shown in all the images. The scale bars are 100nm (i)-A-D and (ii)-B-D, 0.2µm in (ii)-A.
Figure 6-3 The electron micrographs of Col108 mini-fibrils; (i)A-B, the positively stained mini-fibrils (A); the negatively stained mini-fibrils (B) Samples are prepared from the neutralized Col108 solution incubating for 6 hours at 26°C; (ii)A-B, positively stained mini-fibrils (A and B). Samples are prepared from the neutralized Col108 solution incubating for 24 hours at 26°C. Different measurements for d-period are shown in proportion to the scale bar are shown in all the images. The scale bars are 200nm (i)-A and (ii)-A and B, 100nm in (ii)-B. Small fibrils matching the dimensions of a triple helix can be seen in the inset marked with arrows.
6.3.2 The d-periodicity of mini-fibrils of Col108 characterized by AFM

The d-period of ~35nm observed in electron micrographs of Col108 fibrils was emulated nearly quantitatively by atomic force microscopy. The scanning image and the resulting contour profile for Col108 fibril samples after 24 hours of incubation at 37°C is shown in Figure 6-4. The mini-fibrils have similar morphological features to those seen under TEM. Irrespective of visible and bulky salt crystals on the scanned images, the unbranched and spindle-shaped fibrils of 800-1000nm length and 50-75nm in diameter were seen. The scanning profile of the fibrils revealed a periodicity of approximately ~38nm based on the measurements of 20 mini-fibrils. The d-period, in this case, is more than that estimated from TEM images because of the low resolution of the AFM images. The clear boundaries of gap and overlap regions were difficult to follow due to salt deposits as well as the small features of the d-period, itself. Also, the size of the gap/overlap, i.e. 25nm/10nm pattern, is close to the detection limit of the tip used in this study. Nonetheless, the prominent periodicity was evident.
Figure 6-4 AFM analysis of the Col108 mini-fibrils; A. The tapping mode topographic image of Col108 mini-fibrils. Samples were incubated for 24 hours at 37°C. Inset A is the phase retrace image of a region marked in circle on topographic image. B. surface profile of the mini-fibril marked by a short bar on the inset.
6.4 A Model for Col108 mini fibrils: a sequence unit staggered, lateral self-assembly of triple helix

6.4.1 The interaction curves

Similar to collagen fibrils, the assumption was made that molecular recognition during self-assembly of Col108 is mediated by the specific interactions of amino acid residues on the surface of the triple helices. These interactions can be translated directly from the amino acid sequence of the peptide, given the linear conformation of the triple helix.

In order to understand observed banding pattern in the TEM images, stained either positive or negative, and the AFM images, we constructed the stability curve based on the interaction graph using the method used for natural collagen. A computer program using in-house Perl script was created for this calculation, and the details of the calculation for Col108 is shown in Kaur et al. Basically, this program compared two amino acid sequences of Col108. One sequence was used as a reference while the other was shifted with respect to the reference one residue at a time. After each shift, the directly opposed amino acid and the immediately flanking neighbors in the two chains were compared. For each hydrophobic coupling and electrostatic interaction, a value of one was assigned, and the total interaction value was reported. Interaction values were normalized by the total number of possible interactions at each calculation cycle. Hydrophobic amino acids considered were Val, Met, Ile, Leu, Phe, and Pro. Positive amino acids were Lys and Arg. Negative amino acids were Asp and Glu. The stability curve was constructed by plotting the interaction value vs. the shift value. Calculations were run with both parallel and antiparallel chain alignment. Proline (Pro) is usually considered to have a weak hydrophobic interaction, and since it accounts for more than 20% of the X and Y residues in Col108, it was considered as a hydrophobic residue.
in the calculations. However, to check for imposed artifact calculations were also done under the assumption that proline does not belong to the hydrophobic class of the residues.

The foldon domain was treated as a ‘globular’ according to its conformation\textsuperscript{128}. Residues at the core were considered sequestered from solvent exposure and unlikely to be involved in molecular interactions during the self-assembly. Thus, the only surface residues Gly-Tyr-Ile-Pro-Glu-Ala-Pro-Arg-Asp-Asp-Gly-Glu-Trp along the linear trajectory extending from the triple helix, as determined by visual examination of a crystal structure of the T4 foldon domain (PDB Code 1RFO), were included for consideration during calculation cycles.

The innate periodicity of the sequence of Col108 is reflected clearly in the periodic nature of the calculated interaction graphs. In a parallel alignment of the helices, the set of peaks with shift values at multiple of 123 (n×123) stands out as a unique staggering arrangement presenting optimal electrostatic and hydrophobic interactions (Figure 6-5A and B). The peaks are especially pronounced in the total and hydrophobic interaction graphs. The interaction graphs with calculations including or excluding Proline as a hydrophobic residue were compared (Figure 6-5B). The set of peaks with n×123 is observed in calculations with and without the inclusion of proline. It was observed that the inclusion of Proline did not affect the periodic features of the outcome, but only increased the overall magnitude of the interaction because of the increased number of included residues.

Favorable interactions also exist at shift values of ~ n×24 but at a much lower magnitude. A closer inspection of the sequence indicates that the clusters of charged and hydrophobic residues are in approximate alignment when two helices are arranged with shift values of n×24 but this alignment is nearly completely out of phase when shift values are odd multiples of 12. The electrostatic interactions are particularly reflective of the favorable interactions at this small
stagger value, though the favorable interactions at n×123 shifting values are also evident. It is important to note here that the scale of the magnitude of the interactions is an arbitrary one. The peaks represent the potential of interaction, and the magnitude is in no way a representation of the absolute value of the stabilizing energy. The relatively lower magnitude of the electrostatic interactions can be related to the presence of fewer charged residues (75 residues) compared to hydrophobic ones (120 including Proline). Also, the contribution from interactions of like charges is set to be zero. The actual contribution of the electrostatic interactions to the stabilization of the fibrils may not be less than that of the hydrophobic interactions. It is the lack of a set of dominating peaks that set the electrostatic interactions apart from the hydrophobic interactions.

By comparison, the interaction graph of an antiparallel arrangement of helices remarkably lacks any dominating peaks (Figure 6-5C). No one particular shift value dominate the graph; pointing towards no particular dominating chain staggering conformation in an antiparallel arrangement.

The stability curve indicates when arranged with a mutual stagger of 123 residues the self-association will be maximally stabilized through both the hydrophobic and the electrostatic interactions of the neighboring triple helices. In Col108 mini-fibrils, the resulting offsets allowed us to implicate the sequence repeat unit in Col108, which is composed of a total of 120 residues-a set of (GPP)₄ and 108 residues long Col domain in proposing a model that was able to explain both positive and negative staining patterns of Col108 mini-fibrils.
Figure 6-5 The interaction curves for optimal interactions; A, Interaction curve showing total interactions for parallel chain arrangement and the constituent electrostatic interactions (as labeled). B, interaction curves of the constituent electrostatic interactions from A, with or without inclusion of Pro residues (as labeled). C, interaction curve showing total interactions for antiparallel chain arrangement.
6.4.2 The Unit-staggered model of Col108 mini-fibrils

Once a correlation was made between the observed banding pattern and pseudo-periodicity of the Col108 sequence, a self-assembly model of Col108 was proposed to reflect the banding pattern features of both negatively and positively stained mini-fibrils of Col108. (Figure 6-6).

A fully folded Col108 triple helix is about ~ 120 nm in length, 1.5 nm in diameter based on the crystal structures; each sequence unit corresponds to the length of ~35 nm. The foldon domain was a globular with a ~ 2-2.5 nm diameter (the diameter of triple helix is ~1.5 nm) and having an average of three negative charges on its surface based on the crystal structure. The C-terminal (GPP)$_4$, together with the Cys-knots at the C and N termini is expected to contribute to about ~7 nm. Overall the region outside of the three sequence units is ~12 nm, or 0.3X the length of one sequence unit.

In the proposed model, the ~35 nm $d$-period corresponds to a 123 residue pseudo-repeating sequence unit of Col108, and the C-terminal (GPP)$_4$ nucleation sequence and the foldon domain form the major part of the 0.3$d$ overhang. The mini fibrils formed through lateral association of the triple helices with a mutual staggering of one sequence unit at the N-terminus, creating an axially repeating $d$-periodicity of ~35 nm having a light/overlap zone of 0.3$d$ and dark/gap zone of 0.7$d$, as observed in micrographs of negatively stained mini-fibrils of Col108 (Figures 6-1, -2 and -3). Not only does the number of residues of a sequence unit correspond well in size to the observed ~35 nm axial-periodicity, an offset of one sequence unit would lead to a perfect alignment of the sequence units of neighboring helices and produce maximal interactions between the charged and the hydrophobic residues. This corresponds well with the set of $n \times 123$ peaks depicted in the interaction curve graph (Figure 6-5A). The distribution of the charged residues within one sequence unit can be roughly characterized by two high charge-density regions with one being
slightly wider in range than the other (Figure 6-6A). In the unit-stagging assembly the spatial arrangement of these charged regions will be preserved and be identified as three pairs of dark thick-thin bands, with intercalating light strips of charge-free zones every 105nm (the approximate length of 3 sequence units) in the micrographs of positively stained mini fibrils (Figures 6-1(ii) and 6-2(ii)). Thus, the model mini-fibril using a stagger of 1-sequence unit can explain both banding patterns of negatively stained fibrils as well as the positively stained ones.

While the repeating sequence architecture of Col108 warrants the alignment of the sequence units in any arrangement with an offset value of \( n \times \text{unit} \) \((n = 0, 1, 2, \text{or} 3)\), only the 1-unit staggering one will generate the \(~35\text{nm} \) \(d\)-periodicity. As shown in Fig., the exclusive 2-unit and 3-unit staggering arrangements will produce fibrils with an axial periodicity of 70 nm and 105 nm, respectively. These fibrils will have staining patterns easily distinguishable from those shown in Figure 6-6 B and C. An exclusive 0-unit offset assembly, on the other hand, will not form staggered, long fibrils but only end-on-end stacks with one dimension approximately the length of a Col108 monomer. Arrangements of triple helices with a mixture of different \(n\)-unit offset values would, therefore, produce fibrils with gaps and overlaps of varied sizes arranged axially in undefined formations. TEM images of such fibrils would have unrecognizable negatively-stained banding patterns. Thus, the clear and persistent presence of the \(~35\text{nm} \) \(d\)-period of the negatively stained mini-fibrils supports the 1-unit staggering as the predominant arrangement of the associating triple helices.
Figure 6-6: The model of the Col108 min-fibrils with ~35nm periodicity. A: the 1-unit staggering assembly of Col108 and the resultant negatively stained (top) and positively stained (bottom) banding patterns of the fibrils: the dark bars indicate stained area under TEM. The three sequence units of Col108 are shown in different shades of grey; the vertical lines show the locations of charged residues and the oval represents the foldon domain. A rod having three thick-thin charged zones indicated in black is superimposed on the Col108 to highlight the alignment of the charged residues. The double–headed arrows with solid lines mark one ~35nm d-period on the negatively stained fibril. The double-headed arrow with dotted lines mark the length of 3 sequence units (~ 105 nm) on the positively stained fibril. B: the model of 2-unit staggering mini-fibrils. The predicted negatively stained pattern of the fibrils is shown on top with the dark bar showing the stained gap region. The double–headed arrow marks one 70 nm d-period predicted for this arrangement. C: a model of fibrils of 3-unit staggering arrangement. The double–headed arrows mark one 105 nm d-period predicted for this arrangement. This 3-unit arrangement can also represent the fibrils driven by the inter-helical S-S cross-linking. The curved lines mark the locations of clusters of Cys residues.
6.5 Conclusion

Col108 in the neutralized solutions self-assembles into the mini-fibrils with axial periodicity upon incubation at a higher temperature. The $d$-period in Col108 mini-fibrils was measured to be ~35nm. It is important to mention here that for the formation of periodic mini-fibril of Col108, the triple helical conformation of Col108 was a pre-requisite. Fibrillogenesis trials using samples of Col108 that did not indicate the formation of the triple helix with distinct positive and negative peaks in CD spectrum of the starting sample resulted only in non-specific aggregates without any distinct banding pattern.

The proposed model is supported by the structural details of the folded Col18 and the dominating and isolated interaction peaks with a parallel staggering of 123 residues in the interaction graph. The unit-staggering arrangement brings the interacting residues together and affords both maximal molecular interactions and uniqueness in the association pattern during assembly of mini-fibril of Col108.

The reason for observed self-assembled collagen-like fibrils of Col108 with $d$-periodicity is most likely due to the sequence periodicity which may also exist implicitly in natural collagens 129 (Figure 6-7). The axial periodicity in Col108 fibrils is instigated by the built-in sequence periodicity and stabilized by the optimizing inter-helical interactions between the triple helices in a 1-unit staggered arrangement. Other factors in native collagen were also considered critical for fibril formation: the presence of hydroxyproline residues in the sequence and telopeptides. But, existing data has indicated the triple helices without Hyp and without telopeptides can also form D-periodic fibrils. The mini-fibrils of Col108 may indicate that the sequence features of the triple helical domain are sufficient to code for the axially repeating periodicity of fibrils since Col108 does not have either hydroxyproline residues or telopeptides. In summary, we can conclude that
self-assembly of Col108 proceeds via a lateral association between triple helices at a ~35nm stagger similar to the self-assembly of the native collagen at a stagger of 67 nm instead.
Figure 6-7 Comparison of self-assembly of Col108 and native collagen highlighting the similarity in assembly mechanism and differences in morphological features.
6.5 Discussion

The assembly of the mini-fibril is driven by non-covalent interactions between the triple helices, and the robust ~35nm periodicity indicates a highly specific molecular recognition mechanism. The residues of the Col domain were selected from type I collagen for their relatively high propensity for forming a triple helical conformation, yet they may also possess certain intrinsic properties of natural collagen which bring about strong molecular interactions between associating helices. The Col domain has a relatively high content of closely spaced negatively and positively charged residues. Sequences of charged residues such as KGE and KGD are likely involved in the formation of salt bridges between the three constituent chains of a triple helix while others may contribute to interactions and molecular recognition between triple helix monomers during self-assembly. The charged residues within the Col domain are often present as pairs with opposite charges. When the clusters of charged residues are ‘aligned’ between neighboring triple helices, the attractive interactions between oppositely charged residues likely dominate over the repulsion of like charges. A similar distribution of paired charged residues has also been noted for native collagen, although any attributions this distribution leads to the specific structure or functions of collagen were never explored. The periodic insertions of the (GPP)$_4$ sequence may also play a structural role favoring the 1-unit staggered assembly; their specific involvement awaits to be defined.

The role of the foldon in the mini-fibrils is intriguing. The conformation of the foldon has a diameter of 2.5 nm, which is slightly larger than the diameter of ~ 1.5 nm of a typical triple helix and is packed with a defined bent angle with respect to the triple helical domain in the crystal structure of GPP$_{10}$-foldon. However, the ‘kink’ between the two domains is likely caused by crystal packing, and the conformation is likely to be more flexible in solution. The smooth fibrils
of Col108 indicate the foldon can be accommodated, to a degree, in the packing of the helix. The fact that the foldon is located adjacent to the gap region may afford some spatial flexibility of the mini-fibrils to accommodate its larger size. On the other hand, the structural constraints of the foldon may impose certain destabilizing effects and impede the growth of the fibrils. The length of the mini-fibrils is limited to about 5-8 times the end-to-end lengths of the Col108 triple helix, which represents a limited extent of growth compared to that of collagen fibrils. In the modeling of the 1-unit staggered fibrils, the negatively charged foldon domain appears to be surrounded by two positively charged residues from neighboring triple helices, which likely affords some stability. Such interactions, however, are not a determining factor for the structural specificity of the ~35nm axial periodicity, since there exist multiple clusters of positively charged residues on the Col108 triple helix. It is inconceivable to consider the foldon selectively anchors on these two particular positive residues without a more profound stabilizing mechanism in place, which brings the neighboring triple helices together and positions the foldon in the vicinity of the two residues.

Distinct from native fibrillar collagens, Col108 contains no Hyp residues, nor does it have any telopeptides – two factors that have been under constant focus for their purported roles in the fibrillogenesis of native collagen. Despite its anticipated significance in collagen function and structure, the role of Hyp in fibrillogenesis has never been clearly defined. It was first proposed in the 1950s by Gustavson that Hyp may contribute to triple helix association through either a hydration network or direct intermolecular hydrogen-bonding interactions between triple helices. The differences in crystal packing of the triple helical peptide (Pro-Pro-Gly)_{10}, without Hyp residues, and peptides having a high Hyp content were taken as evidence that Hyp is involved in the organization of the water network surrounding the triple helix. It was further suggested that the formation of higher molecular aggregates of the (Pro-Hyp-Gly)_{10} peptide is mediated by
hydrogen-bonded hydration networks involving Hyp \textsuperscript{133}. The concept of this hydration network, or more precisely the lack of it, was used to explain the deficiency of fibril assembly of a recombinant collagen without Hyp, from transgenic plants, under physiological conditions. Nonetheless, this unhydroxylated collagen was found to retain the ability to form fibrils with a 67 nm banding pattern albeit only in buffers of low ionic strength \textsuperscript{45}. Another recombinant type I collagen from a yeast expression system – the rho collagen – was also reported to form ‘normal fibrils in vitro’ despite the absence of the canonical hydroxylation pattern of Pro residues \textsuperscript{133}. The Hyp content of rho collagen is only 50% of that of the type I collagen, and the locations of the Hyp are not specified. While the involvement of Hyp in a hydration network or a hydrogen-bonding network may contribute to the stabilization of fibrils, it is not an intrinsic factor for structural specificity or formation of collagen fibrils. Factors affected by the interactions involving Hyp are likely to include the kinetics and the optimal conditions of the fibrillogenesis, or even the diameter of the fibrils \textsuperscript{134}. Still, the absence of Hyp does not abolish the ability of a triple helix to form a periodic assembly. The ability of Col108 to self-assemble into a highly ordered, axially periodic structure is consistent with the conclusion that Hyp is not a requirement for D-staggered packing during the fibrillogenesis of collagen.

The role of telopeptides in fibrillogenesis is another debated issue. The involvement of telopeptides in the early stages of fibrillogenesis was extrapolated to involve the binding of one telopeptide to the triple helical region of a nearby collagen monomer to place ‘the monomers in quarter \textit{D-period} stagger,’ and to promote the formation of a structural nucleus essential for further growth of the fibrils \textsuperscript{46}. Removal of the telopeptides by proteolytic treatment of collagens isolated from tissues was found to affect the kinetics \textsuperscript{41, 89} and change the morphology of the fibrils \textsuperscript{135-137}, or even prevent the formation of fibrils having a banding pattern \textsuperscript{46}. Other studies, however,
attributed the negative impacts on fibril formation caused by the removal of the telopeptides to the proteolytic removal of portions of the triple helix domain itself. Subsequently, collagen was found capable of forming fibrils even after complete removal of the telopeptides. This latter finding challenges the notion that the telopeptides are part of the ‘code necessary for recognition’ during fibrillogenesis. While Col108 does not have the N- or C-telopeptide of collagen, it does contain a foldon domain at the C-terminus. Although we consider it unlikely, further study is needed to elucidate any possible anchoring roles of this non-collagenous domain of Col108 during the self-assembly of the mini-fibrils.

In comparison to other non-specific self-assemblies of triple helical peptides, the larger size and the rich collection of charged and hydrophobic residues of Col108, in addition to the repeating units of the sequences, may promote the structural specificity of the Col108 fibrils. The need for more diverse sequences than simple repeats of (Pro-Pro-Gly) or (Pro-Hyp-Gly) triads for the presence of structural features among self-assembled triple helices has long been recognized, and the effects of long-range sequence repeat beyond the (Gly-X-Y) triplets, is understudied. How to build such repeating units of sequences having a high variety of different residues into short synthetic peptides is challenging. The inclusion of multiple (Gly-Pro-Hyp) triplets is often necessary for the stability of short triple helical peptides. The larger size of Col108 may provide more extensive contacts between the associating helices. Similar to protein folding in general, a critical amount of stabilizing interactions is often required for a specific structure to be selected over other competing forms, and this leads to its dominance in the free energy-driven, spontaneous process. We do not know if the 120 residues of one sequence unit represent, or is close to, the minimum size necessary to support an axially repeating structure during assembly, or if the number three represents a ‘golden number’ for the number of repeating sequence units. The relatively
simple and yet effective Col108 system offers the opportunity to address these and an array of other fundamental questions about the fibrillogenesis of collagen and protein design.
Supplementary Information

### Index for TEM images of Col108 fibrils

<table>
<thead>
<tr>
<th>Incubation temperature</th>
<th>Salt Conc .</th>
<th>Incubation time</th>
<th>Trial</th>
<th>Number of files</th>
</tr>
</thead>
<tbody>
<tr>
<td>26°C</td>
<td>67.5mM NaCl</td>
<td>6 hours</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24 hours</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>37°C</td>
<td>67.5mM NaCl</td>
<td>6 hours</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24 hours</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>9</td>
</tr>
</tbody>
</table>

### Index for AFM images of Col108 fibrils

<table>
<thead>
<tr>
<th>Incubation temperature</th>
<th>Salt Conc .</th>
<th>Incubation time</th>
<th>Image folder</th>
<th>Number of files</th>
</tr>
</thead>
<tbody>
<tr>
<td>37°C</td>
<td>67.5mM NaCl</td>
<td>24 hours</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>9</td>
</tr>
</tbody>
</table>
Bibliography


106. Qiagen Ni-NTA agarose


108. Xu, K.; Nowak, I.; Kirchner, M.; Xu, Y., Recombinant Collagen Studies Link the Severe Conformational Changes Induced by Osteogenesis Imperfecta Mutations to the Disruption of a Set of Interchain Salt Bridges. *Journal of Biological Chemistry* **2008**, *283* (49), 34337-34344.


