2-2019

The Determining Factors of the Self-Assembly of Collagen Mimetic Peptide

Fangfang Chen

The Graduate Center, City University of New York

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 Biochemistry Commons

Recommended Citation


This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
THE DETERMINING FACTORS OF THE SELF-ASSEMBLY OF COLLAGEN MIMETIC PEPTIDE

by

Fangfang Chen

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

2019
The Determining Factors of the Self-assembly of Collagen Mimetic Peptide

by

Fangfang Chen

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

__________________________
Date Yujia Xu
Chair of Examining Committee

__________________________
Date Richard S. Magliozzo
Executive Officer

Supervisory Committee:

Dr. Yujia Xu
Dr. Rein Ulijn
Dr. Mandë Holford
Dr. Peter Lipke
Dr. Shuiqin Zhou

THE CITY UNIVERSITY OF NEW YORK
Abstract

The Determining Factors of the Self-assembly of Collagen Mimetic Peptide

by

Fangfang Chen

Adviser: Dr. Yujia Xu

We recently reported that a designed triple helix Col108 self-associates to form collagen-like mini-fibrils having a $d$-periodicity of 35 nm. Recombinant protein Col108 consists of 378-residue triple helix domain organized into three repeating sequence units, and a C-terminal foldon domain. The 35 nm $d$-period of Col108 mini-fibrils is consistent with the one unit staggered arrangement of the associating helices. To further investigate if the sequence periodicity is crucial for collagen mimetic peptide fibril formation, we studied the self-assembly of four other designed triple helical peptides. Peptide 2U108 has two repeating sequence units; peptide Col877 consists of three repeating units but the amino acid sequence of each unit is very different from that of Col108; peptide 1U108 has only one sequence unit and peptide. Col108r has the same amino acid composition of that of Col108 but lacks the sequence periodicity. Both 2U108 and Col877 formed mini-fibrils having the same periodicity of 35 nm as that observed in Col108 under electron microscopy, but no fibril-like assemblies were observed for either 1U108 or Col108r. Since both 2U108 and Col877 have tandem repeats of sequence units while 1U108 and Col108r do not, these findings support the essential roles of the long-range repeating sequence units on the self-assembly of collagen-like, staggered fibrils and suggest a design-rule for the self-assembly of triple helices into collagen-like mini-fibrils.
Acknowledgement

First I would like to express my sincere thanks and deepest gratitude to my mentor Dr. Yujia Xu. Dr. Xu gave me a chance to come to the United State to pursue my dream of being a scientist. At the beginning of my Ph.D. study, I had a difficult time to get used to different circumstance and culture. She helped me get prepared for my study and fit in new life. She also helped me come up the topic of thesis and guided me when I met any difficult questions during research. Her suggestion helped me learn newest knowledge and think as a scientist. Her advice and comments in my thesis gave me a good opportunity to improve my writing skills. Without her advice and encourage, I cannot go through this rough road to finish my thesis.

I am also very grateful that Dr. Rein Ulijn, Dr. Peter Lipke, Dr. Mandë Holford and Dr. Shuiqin Zhou have been my thesis committee members for many years. They provided their valuable advice and helpful suggestions on my proposal and research presentation, which guide me to deeper thinking of my research from different perspectives.

I also would like to thank Mr. Jorge Morales and Dr. Zengyan Wei for TEM training and Dr. Tong Wang for assistance with cryo-EM, and Dr. Barney Yoo for LC-MS training. I am very thankful to our colleagues in Dr. Xu lab. The scientific discussion with them gave me some new ideas about my project. And their research experience is very helpful for me to solve the problems during research. Dr. Rebecca Strawn and Dr. Parminder Haven provided good suggestions and offered support on my difficult moment in research. Sam Wong helped me to learn new techniques and provided some advice on my thesis. Dr. Michele Kirchner taught me teaching skills. Our new lab
member Faizunnahar Dewan and undergraduates Michael Tankelevich, Carl Guenst, Julia Korobenko, Paula Laziuk and Mohamed Abdelhamid make the lab a great place to work.

Last but not least, I would like to thank my family for all their love and encouragement. My parents are always supporting me in all my pursuits. Finally, I am thankful to my husband for supporting my every decision.
# Table of Contents

Chapter 1. Introduction .................................................................................................................. 1

1.1 Collagen protein family ........................................................................................................... 1

1.2 Fibrillar collagens .................................................................................................................. 3

1.3 Hierarchical structure of collagen .......................................................................................... 4

1.4 Fibril-forming collagen biosynthesis ....................................................................................... 7

1.4.1 Intracellular step .................................................................................................................. 7

1.4.2 Extracellular step ................................................................................................................. 8

1.5 Collagen mimetic models ....................................................................................................... 9

1.6 Collagen mimetic mini-fibrils with D-period like banding pattern .............................................. 13

Chapter 2. The d-periodic mini-fibrils of collagen mimetic peptide Col108 .............................. 17

2.1 Introduction ............................................................................................................................. 17

2.2 Methods and materials .......................................................................................................... 17

2.2.1 The expression and purification of Col108 .......................................................................... 17

2.2.2 Preparing Col108 solutions ................................................................................................ 21

2.2.3 Conformation Studies ........................................................................................................ 21

2.2.4 Thermal stability studies .................................................................................................... 22

2.2.5 TEM characterization of Col108 mini-fibrils .................................................................... 22

2.2.6 AFM characterization of Col108 mini-fibrils ..................................................................... 23

2.2.7 Cryo-EM characterization of Col108 mini-fibrils ................................................................. 23

2.3 Results ..................................................................................................................................... 24

2.3.1 The purification of Col108 .................................................................................................. 24
2.3.2 Characterization of Collagen-Mimetic Peptide Col108 .......................................................... 27
2.3.3 The thermal stability of Col108 ............................................................................................. 28
2.3.4 The d periodicity of mini-fibrils of Col108 characterized by TEM .......................................... 29
2.3.5 The d periodicity of mini-fibrils of Col108 characterized by AFM .......................................... 31
2.3.6 Col108 mini-fibrils characterized by cryo-EM ...................................................................... 32

2.4 Summary .................................................................................................................................... 34

Chapter 3. The self-assembly of a triple helix with two sequence units – the sufficient conditions for the formation of mini-fibrils with the d-period ............................................. 35

3.1 Introduction ................................................................................................................................. 35
3.1.1 Sequence design of 2U108 and 1U108 .................................................................................. 35
3.2 Materials and Methods .............................................................................................................. 38
3.3 Results ........................................................................................................................................ 38
3.3.1 The triple helical peptides 2U108 and 1U108 ....................................................................... 38
3.3.2 Self-assembly of mini-fibrils ................................................................................................. 41
3.4 Summary .................................................................................................................................... 46

Chapter 4. Self-assembly of triple helical peptides Col877 and Col108R....................................... 47

4.1 Introduction ................................................................................................................................. 47
4.2 Methods and Materials ............................................................................................................. 47
4.2.1 The genes of Col877 and Col108R ......................................................................................... 47
4.2.2 Expression and Purification .................................................................................................. 48
4.2.3 The structural characterization and the study of the self-assembly ....................................... 51
4.3 Results ........................................................................................................................................ 51
4.3.1 The design of the peptides .................................................................................................... 51
4.3.2 The sequencing of Col877 and Col108R genes .......................................................... 54
4.3.3 The purification of Col877 .......................................................................................... 56
4.3.4 The purification of Col108R ...................................................................................... 63
4.3.5 The folding and the thermal stability of Col877 and Col108R ........................................ 70
4.3.5 Electron microscopy study of the self-assembly of Col877 and Col108R .................. 72
4.4 Summary ........................................................................................................................ 79

Chapter 5. Discussion ......................................................................................................... 81

APPENDIX ............................................................................................................................ 85

A. The result of Col877 DNA sequencing by GenScript ...................................................... 85
B. The result of Col877 DNA sequencing by GENEWIZ .................................................... 88
C. The result of Col108R DNA sequencing analyzed using Bioedit software ...................... 91
D. Protein Expression Evaluation of Col877 in E.coli done by GenScript ......................... 96

BIBLIOGRAPHY .................................................................................................................. 97
List of Figures

FIGURE 1-1 SCHEMATICS OF COLLAGEN FAMILY ................................................................. 2
FIGURE 1-2 THE PATTERN OF HYDROGEN BONDS IN TRIPLE HELIX ................................... 5
FIGURE 1-3 AXIAL STRUCTURE OF δ-PERIODIC COLLAGEN FIBRILS ................................... 7
FIGURE 1-4 THE MAIN PROCEDURES OF COLLAGEN TRIPLE HELICES SELF-ASSEMBLED INTO COLLAGEN FIBRILS ........................................................................................................... 9
FIGURE 1-5 SELF-ASSEMBLIES OF TRIPLE HELICAL PEPTIDES ........................................... 10
FIGURE 1-6 THE SEQUENCE OF COL108 .................................................................................. 14
FIGURE 1-7 THE TEM PICTURES OF COL108 MINI-FIBRILS .................................................. 15
FIGURE 1-8 COL108 MINI-FIBRIL MODEL SHOWING CHARGE DISTRIBUTION .................. 15
FIGURE 2-1 TIME PROGRAM GRAPH USED FOR PREPARATIVE COLUMN WITH PERCENTAGE GRADIENT ACETONITRILE ................................................................. 21
FIGURE 2-2 SDS-PAGE (12%) ANALYSIS OF THROMBIN CLEAVAGE OF COL108 FUSION PROTEIN ................................................................. 25
FIGURE 2-3 SDS-PAGE (12%) ANALYSIS OF FRACTION SAMPLES AFTER HPLC SEPARATION ........ 27
FIGURE 2-4 CD SPECTRA OF COL108 ...................................................................................... 28
FIGURE 2-5 CD MELTING CURVE OF COL108 TRIPLE HELIX FROM 4°C TO 65°C .................. 29
FIGURE 2-6 THE TEM IMAGES OF TYPE I COLLAGEN FIBRILS FROM RAT TAIL .................... 30
FIGURE 2-7 THE TEM IMAGES OF COL108 MINI-FIBRILS .................................................... 31
FIGURE 2-8 AFM TOPOGRAPHY IMAGES OF COL108 MINI-FIBRILS ..................................... 32
FIGURE 2-9 CRYO-EM IMAGE OF COL108 AGGREGATES ..................................................... 33
FIGURE 3-1 THE CONSTRUCT OF 2U108 AND 1U108 ............................................................. 36
FIGURE 3-2 THE AMINO ACID SEQUENCE ARCHITECTURE OF 2U108 .................................. 37
FIGURE 3-3 THE AMINO ACID SEQUENCE ARCHITECTURE OF 1U108 .................................. 38
FIGURE 3-4 CIRCULAR DICHROISM SPECTRA OF 2U108 AND 1U108 .................................. 39
FIGURE 3-5 TEMPERATURE MELTING CURVE OF COL108, 2U108 AND 1U108 IN 5MM ACETIC ACID FROM 4°C TO 55°C ...................................................................................... 40
FIGURE 3-6 THE TEM IMAGES OF 2U108 ............................................................................... 43
FIGURE 3-7 THE TEM IMAGES OF 2U108 MINI-FIBRILS ......................................................... 43
FIGURE 3-8 THE TEM IMAGES OF 1U108 ............................................................................... 45
FIGURE 4-1 SCHEMATIC EXPRESSION OF THE GENE CONSTRUCTS OF THE COLLAGEN MIMETIC PEPTIDES .................................................................................................................. 52
FIGURE 4-2 THE AMINO ACID SEQUENCE ARCHITECTURE OF COL877 ................................ 52
FIGURE 4-3 THE AMINO ACID SEQUENCE ARCHITECTURE OF COL108R ............................. 54
FIGURE 4-4 SDS-PAGE (12%) ANALYSIS OF PURIFICATION OF HIS-TAGGED COL877 FUSION PROTEIN EXPRESSED IN BL21(DE3) CELLS ................................................................. 57
FIGURE 4-5 SDS-PAGE (12%) ANALYSIS OF THROMBIN CLEAVAGE OF HIS-TAGGED COL877 FUSION PROTEIN ................................................................. 58
Chapter 1. Introduction

1.1 Collagen protein family

Collagen is the major structural protein in the extracellular matrix and the most abundant protein in animals. Collagen provides structural integrity and malleability to all internal organ (1). Most collagens are able to form supramolecular aggregates.

So far, 29 distinct types of collagens have been identified and divided into several subgroups based on their supramolecular structure and function (2,3). The subgroups of collagen include fibril forming collagen, fibril associated collagen containing interrupted triple helices (FACITs), collagen from beaded filament, collagen in basement membrane, and anchoring fibril and network forming and transmembrane collagens. The supramolecular structures of collagen subfamily are shown in Figure 1-1. Fibril forming collagens comprise major collagen type I, II, III, and minor type V, XI, XXIV and XXVII and compose 90% of proteins in extracellular matrix. FACITs and related collagens including collagens type IX, XII, XIV, XVI, XIX, XX, XXI, XXII and XXVI are relatively short proteins and do not form fibrils themselves, but are associated with the surface of fibrils of collagen fibrils (4). Membrane collagen and collagen-like membrane proteins including collagen XIII, XVII, XXIII and XXV, contain a transmembrane domain and do not form supramolecular structure. Multiplexins (multiple triple helix domains and interruptions) including types XV and XVIII have a triple helical domain with some interruptions. Collagen XV can form oligomeric assemblies and is found to bridge neighboring collagen fibrils near basement membranes (5). Anchoring fibrils and beaded-filament-forming collagens, known as microfibrillar collagens including VI, VII, XXVI and XXVIII, are widely spread in the body. Network-forming collagens including type IV, VI, VIII and X collagen are found in extracellular matrix which
supports epithelial cells, muscle fibers, and peripheral nerves (3).

Figure 1-1 Schematics of collagen family. (taken from van der Rest M et al (6)).
1.2 Fibrillar collagens

The fibril-forming collagens including collagen type I, II, III, V and XI are characterized by their ability to self-assemble into fibrils (3). The supramolecular structure of collagen fibrils is defined by a characteristic 67nm periodicity: a banding pattern characterized using electron microscopy. The collagen molecules are composed of a triple helix domain that contain around 1000 residues in Gly-X-Y repeating sequence flanked by short non-helical telopeptide at both ends. The individual triple helix in type I collagen is about 1.5 nm in diameter and 300 nm in length (6)

Type I collagen is most abundant protein and the major protein in bone, tendons, skin, ligaments, cornea, and many interstitial connective tissues except cartilage (3). The type I collagen triple helix is composed of two $\alpha_{1}(I)$ chains and one $\alpha_{2}(I)$ chain. Minor amounts of type I collagen are homotrimer composed of three $\alpha_{1}(I)$ chains in embryonic tissues (7). The molecular mechanism of forming heterotrimer vs homotrimer type I collagen remains unknown. Homotrimeric type I collagen is found to be resistant to all mammalian collagenases (8). The homotrimers can still assemble into fibrils with the same periodicity as that of heterotrimers. Type I Collagen molecules self-assemble to form striated fibrils with a periodicity of 67 nm (5). In the fibril, collagen molecules are arranged in staggered arrays. The stagger is equal to the $D$ period (67 nm). The length of single collagen molecule is $4.4D$. Thus, there is a 0.6 $D$ gap between the ends of neighboring molecules. The gap-overlap structure gives rise to $D$-periodic banding pattern of collagen fibrils. The $D$-periodic banding plays an important role in high tensile strength and bone mineralization (9,10).

The mutations in genes coding for type I collagen interfere with the folding of triple helix, assembly of collagen molecules and interaction with other molecules in extracellular matrix. The
common type I collagen-related diseases are osteogenesis imperfecta (OI) and Ehlers-Danlos syndrome (EDS). The characteristic features of OI are brittle bones, blue sclerae, abnormal teeth, thin skin, weak tendons and hearing loss (11). Most of the mutations are the substitution of glycine in Gly-X-Y pattern with other amino acid. Other mutations contain deletions, insertion and RNA splicing defects (11). EDS is characterized by hypermobility and skin abnormality (12). These disorders are caused by a failure to cleave N-termini propeptide due to a mutation at N-terminal cleavage site. Without cleavage of N-termini propeptide, collagen molecules are not able to self-assemble to fibrils. Some diseases related to homotrimeric type I collagen are cancer, osteoarthritis, osteoporosis and fibrosis (7).

Type II collagen is the major constituent of cartilage and a homotrimer formed by three α1(II) chains. Type II collagen covalently interacts with type IX and V/XI collagens to form heterotypic fibrils in the cartilage matrix (13). Type V/XI collagen is a heterotrimeric molecule composed of α1(XI), α2(XI), α3(XI) chains. Type III collagen mostly exists in soft connective tissues such as embryonic skin, lung and blood vessels. Type III collagen is a homotrimer of three α1(III) chains. Type III collagen assembles into fibrils with type I collagen in some tissues.

1.3 Hierarchical structure of collagen

The triple helix conformation of collagen is unique. All of the collagen molecules are made up with three polypeptide chains. The three α chains are supercoiled in a right hand manner in the triple helix and each is in a polyproline II like conformation. Each polypeptide chain contains the repeating Gly-X-Y pattern in which glycine residue occupies at every third position. Glycine is the smallest amino acid residue and the only residue that can fit tightly in the center of triple helix. X and Y can be any of the 21 amino acid residues but are frequently proline (at X position) and
4R-hydroxyproline (Hyp, at Y position). If glycine is replaced by other residue, the triple helix motif is interrupted. The triple helical structure was proposed by Ramachandran and Kartha (14).

These three polypeptide chains of the triple helix are stabilized by interchain hydrogen bond. N-H of glycine interacts in one chain with C=O of the residue in the X position of neighboring chain to form hydrogen bond (Figure 1-2). Hydrogen bonds are perpendicular to the helix axis. Not all backbone peptides of the repeating tripeptides are involved in hydrogen bonds in triple helix (15). Each repeating tripeptide form one hydrogen bond. If X or Y residues are not Pro (proline) or Hyp, there are second hydrogen bonds between N-H of the X position residue and the C=O of the Gly residue(16). Water molecules are involved in these hydrogen bonds. Water-mediated hydrogen bonds stabilize the collagen triple helices.

![Figure 1-2 The pattern of hydrogen bonds in triple helix (taken from reference (18)).](image)

The hierarchical structure of collagen starts with a long polypeptide chain – the alpha chain, consisting of the triple helix domain, and the N- C propeptides. The folding in ER generates
trimeric procollagen. The N- and C- globular domains are removed in the extracellular space to generate tropocollagen. Tropocollagen molecules self-associate to form fibrils with D-periodicity in appropriate conditions. Collagen fibrils have a diameter of 100 - 500 nm and their length is around micrometer. Multiple collagen fibrils aggregate spontaneously to form collagen fibers mediated by cross-linking macromolecules such as proteoglycans and FACIT collagens (9). The size and shape of collagen fibers vary depending on tissues and organs. The diameter of collagen fibers is 1 to 100 µm thick. The length of fibers is around milimeter (17).

The structure of collagen fibril is characterized by 67 nm axial periodicity, also called D periodicity shown in Figure 1-3 (18). The uninterrupted triple helix in fibril-forming collagen molecules is about 300 nm length and 1.5 nm in diameter (19). In the linear conformation of the triple helix having a uniform helix rise of 0.8 - 0.9 nm per Gly-X-Y triplet, a 234 residue segment correlates to a ~ 67 nm segment, Thus, each Collagen triple helix encompasses 4.4 D periods, each D-period is about 67nm and includes 234 residues. The D-periodicity of collagen fibrils is composed of overlap and gap regions in the staggered arrangement. Under electron microscope, the overlap region (light band) is ~ 27 nm and gap region (dark band) is ~ 40 nm in negatively stained fibrils (19). Studies using X-ray fiber diffraction have linked 67 nm D-period of collagen fibrils to 234 residues staggered pattern of triple helices (20). It is postulated that in the 234 residues staggered arrangement, hydrophobic and electrostatic interaction is maximized (21). However, it remains unclear what properties of the triple helix make this staggered arrangement a unique conformation during fibril formation. It is also not clear what specific amino acid residues are involved in these interactions and if the specific placement of these residues along the triple helix has any connection with the 234-residues staggered arrangement.
Figure 1-3 Axial structure of D-periodic collagen fibrils (taken from reference (30)). (a) Schematic representation of the axial arrangement of triple helical collagen molecules in a fibril. (b) Negatively stained fibril shown under electron microscopy.

1.4 Fibril-forming collagen biosynthesis

Most collagens assemble into supramolecular architectures by self-assembly or by interacting with other collagen molecules. The specific pathways of Fibril-forming collagen biosynthesis are discussed below (Figure 1-4).

1.4.1 Intracellular step

All individual collagen chains are synthesized on the ribosome and translocated into the lumen of rough endoplasmic reticulum (ER). A series of post-translational modifications happen including hydroxylation of proline at Y-position and selected lysine residues (22). In fibrillar collagens, nearly all Pro on Y position are hydroxylated. Some hydroxylysine residues are further glycosylated to form galactosyl-hydroxylysine and glucosyl-galactosyl hydroxylysine. Hydroxylysine also plays crucial roles in stabilizing triple helix by forming cross-links between collagen molecules in fibrils.
Following the hydroxylation of proline and lysine and glycosylation of hydroxylysine, individual collagen chains associate together to form the triple helix, which is often referred to as the procollagen. In fibril-forming collagens, C terminus propeptide plays an important role in selecting and trimerizing three appropriate α chains to form triple helix but prevent fibril formation (23). Collagen molecules are nucleated and elongated in C- to N- direction after initiation by C propeptide. During the trimerization, intramolecular and intermolecular disulfide bonds are formed by protein disulfide isomerase (PDI) (24). Type I collagen does not have disulfide bonds. All fibrillar collagens are synthesized to soluble procollagen with N- and C- terminal propeptide domains. Once post-translational modifications and folding are completed, procollagen molecules are transported to the Golgi apparatus. In Golgi, procollagen molecules form triple helices. The triple helices are secreted into extracellular parts.

1.4.2 Extracellular step

After procollagen molecules are secreted into extracellular space and N- and C-propeptides are cleaved by procollagen N-proteinase and C-proteinase respectively, collagen molecules aggregate to form fibrils. In the process of fibrillogenesis, specific lysines and hydroxylysines in the N- and C- telopeptides are oxidatively deaminated by lysyl oxidase to allysine and hydroxyallysine respectively (25). Aldehyde groups of allysine and hydroxyallysine react with amine groups of lysine and hydroxylysine in the telopeptides to form intra- and inter-molecular cross-links. The role of cross-links is critical in stabilizing fibrils and providing tensile strength and mechanical stability to connective tissues (26,27).
Figure 1-4 The main procedures of collagen triple helices self-assembled into collagen fibrils (taken from reference (19)). Three α chains are selected and trimerized by carboxyl propeptides to form triple helix. The N- and C-propeptides of triple helices are cleaved by proteinases in the extracellular matrix. The triple helices are staggered to form cross-striated fibrils.

1.5 Collagen mimetic models

Collagen is very useful as a biomaterial for medical and nanotechnological application due to molecular properties such as low antigenicity, biodegradability and high compatibility (28). However, collagen isolated from natural sources and by expression is difficult to modify, lack of purity, affected by post-translational modifications and can induce pathogenic and immunological side effects (29). Studies utilize “bottom-up” approach for molecular design focused on generating self-assembly from peptides. The different types of fibrils are shown in Figure 1-5. Collagen
mimetic peptides are designed to mimic multiple levels of collagen’s structural hierarchy and study the details of the molecular mechanism of fibril formation. The peptides may also provide a tool to study interactions with other proteins and cells that require higher ordered structure.

Figure 1-5 Self-assemblies of triple helical peptides. (A) Self-assembled structure of (Pro-Hyp-Gly)_{10} is from Dr. Brodsky’s lab. (taken from Kar et al. (26)). (B) CPB:CPC samples that assemble into fibrils at mixing ratio 6:8 are from Dr. Nanda’s lab (taken from Xu et al. (34)). (C) The well defined periodic microfibrils are from Dr. Chaikof’s lab. The gap region is around 6.7 nm. The overlap region is around 11.2 nm. (taken from Przybyla et al. (4)) (D) The synthetic peptide that assembles into nanofibrils is from Dr. Hartgerink’s lab. The fibril is much thinner than natural collagen fibrils. (taken from O’Leary et al. (38))

Raines et al. have used triple helical propensity to study the assembly of synthetic collagen peptides with sticky ends (30). They designed and constructed two synthetic collagen fragments with overhanging (Pro-Yaa-Gly)_{5}. Collagen fragments self-assemble to thin and long fibrils via
intermolecular triple helix formation. Some fibrils are longer than natural collagen triple helices. Also, they found that the length of fibrils is modulated by thermal stability, amino acid composition, temperature and solvent. The fibrils formed from synthetic collagen peptide through sticky ends is different from the staggered arrangement of natural collagen molecules. These collagen-like fibrils formed through the intermolecular assembly of the overhanging (Pro-Yaa-Gly)$_5$ segments. Natural collagen molecules assemble into long fibrils with $D$ periodicity through 234 residues staggered arrangement.

Brodsky et al. have reported a triple helix peptide (Pro-Hyp-Gly)$_{10}$ that self-associates to disordered aggregates with branched, filamentous structures under defined temperature, pH and concentration (18). In this model, intermediate loosened state of the triple helix and water molecules associated with triple helix are thought to be crucial factors for self-association. The $T_m$ of the peptide (Pro-Hyp-Gly)$_{10}$ is $\sim 60^\circC$. The hydrated triple helix transfers to loosened state with temperature higher than $35^\circC$. And then loosened triple helix interacting with water starts to self-associate to supermolecular form or a denatured monomer state. There are a number of similarities between fibril formation process of this peptide and natural collagen discussed below. Both Collagen fibril formation and the self-association of this peptide are optimal in neutral pH and do not occur in low pH. With increasing temperature, collagen fibril formation and the association of this peptide are faster (31). Similar to collagen fibril formation, the assembly of the peptide is a reversible process. However, this peptide forms a disorganized structure in contrast to higher order structure of natural collagen fibrils.

Yoshizumi et al. reported a recombinant triple helix consisting of two tandem collagen domains, the peptide CL-CL, formed fibrillar structures at neutral pH (32,33). The fibril structure was characterized as bundled in-register arrays of the triple helices, but has no $D$-period like axial
structures. They attributed the lack of $D$-periodicity to the insufficient size of the CL-CL triple helix.

Nanda et al. presented that synthetic peptides (CPB and CPC) through non-covalent electrostatic interaction assembled into fibrils and hydrogel (34,35). The amino acid sequence of CPB is PEGEOGROGROGROGPRGPRGPRGPRG. The amino acid sequence of CPC is PRGROGROGPRGPRGPRGPRGPRGPRG. The morphology of high order assembly depends on concentration and mixing the stoichiometry of CPB and CPC (34). They also indicate that triple helical intermediate is a prerequisite for the formation of high order structure. Individual synthetic peptides are associating perpendicular to the fiber axis. No periodic banding pattern is found in these fibrils. Recently Nanda’s lab found that a new synthetic peptide H4 self-assembles into discs driven by hydrophobic interactions between the central (LIG)$_2$ domains of triple-helices (36). The thickness of discs are ~ 10 nm that is equal to the length of H4 triple helix. The discs co-assemble with natural collagen by hydrophobic interaction. This peptide scaffold contributes to study the molecular mechanism of collagen fibril formation.

Chaikof and co-workers designed a fibril-forming collagen-like peptide that includes 36 amino acid residues with the sequence (Pro-Arg-Gly)$_4$(Pro-Hyp-Gly)$_4$(Glu-Hyp-Gly)$_4$ (37). This collagen-like peptide self-assembles into microfibrils with 18 nm periodicity through electrostatic interaction. The fibrillogenesis of collagen-like peptide is through axially staggered assembly between oppositely charged residues Arg and Glu of triple helices. However, the temperature for fibrils formation is higher than melting temperature of triple helices. It indicates that triple helices are unfolded during self-assembly of collagen-like peptide. The predicted model of collagen-like peptide fibrillogenesis assumes the microfibril is formed assembly between oppositely charged N- and C-terminal ends of triple helices.
Hartgerink et al. constructed a new synthetic collagen mimetic peptide based on the analysis of different electrostatic interactions between charged amino acids (38). The difference between this peptide and the work of Chaikof and Conticello is that Hartgerink’s group replaced Arg with Lys and Glu with with Asp. The charged pairs of Lys and Asp form stronger salt-bridge hydrogen bonds than charged pairs of Arg and Glu. The sequence of this designed peptide is (Pro-Lys-Gly)₄(Pro-Hyp-Gly)₄(Asp-Hyp-Gly)₄. The results show that collagen mimetic peptides assembled into nanofibers by triple helical nucleation. Axial charged pairs between (PKG)₄ and (DOG)₄ facilitate to form sticky ended triple helices. The Lys and Asp side chains that do not participate in intrahelical salt bridges promote to form triple helix bundles through interhelix interactions.

1.6 Collagen mimetic mini-fibrils with D-period like banding pattern

In our lab, we created a new collagen mimetic peptide called Col108. Col108 was obtained from a recombinant gene expressed in E. coli. The sequence of Col108 has 417 residues that include triple helix domain and foldon domain (27 residues) (Figure 1-6). The triple helix domain has three repeating units named U1, U2 and U3 respectively. In each unit, there is a 108-residue col domain and a N-terminal (Gly-Pro-Pro)₄ sequence. The 108 residues of the col domain were taken from different regions of α1 chain of human type I collagen: residue 242 to 256, 296 to 322, 434 to 478, 515 to 535. These residues were selected based on their relatively high propensity for triple helix stabilization according to the work of Brodsky’s lab (39). A Cys-knot sequence consisting of Gly-Pro-Cys-Cys was incorporated at N-and C-termini of the triple helical domain to covalently link the three polypeptide chains of a triple helix through a set of interchain disulfide bonds (40). The C terminal 27-residue-long foldon domain taken from bacteriophage T4 fibritin
was included as a nucleation domain during the folding of the triple helix.

Figure 1-6 The sequence of Col108. Grey boxes are restriction enzyme sites. The three sequence units are marked by the double-head arrows. Numbers above the residues indicate the positions of residues in human α1(I) chain. (taken from Kaur et al (40)).

The self-assembled Col108 mini-fibril was characterized by an axial periodicity of 35 nm (40). The banding pattern in Col108 fibrils is similar to that of natural type I collagen. Under electron microscope, the overlap zone (light band) is ~ 10 nm and gap zone (dark band) is ~ 25 nm in negatively stained Col108 fibrils (Figure 1-7). The 35 nm d-period of Col108 is consistent with a one-unit staggering arrangement of Col108 triple helices in the mini-fibrils. A sequence unit of Col108 is about 123 residues including the col domain and the (GPP)₄ insert. In the linear conformation of the triple helix having a uniform helix rise of 0.8 – 0.9 nm per Gly-X-Y triplet, a 123 residue segment correlates to a ~ 35 nm segment, highly agreeable with the observed d-periodicity of the mini-fibrils (40). The predicted one-unit staggered pattern of Col108 mini-fibril is shown in Figure 1-8. In this model, the sequence units of the associating triple helices are in register, thus the hydrophobic residues of one triple helix have potential to interact another hydrophobic residues of the neighboring triple helices. While the charged residues interact only
with those having an opposite charge, the characteristic sequence placement of Col108 having clusters of oppositely charged residues next to each other can potentially create zones of charged residues and ultimately reinforce the electrostatic interactions between the helices. Thus, the one-unit staggered pattern brings residues in optimal alignment to maximize the interactions between associating triple helices and makes the $d$-periodic mini-fibrils the unique, most stable conformation.

Figure 1-7 The TEM pictures of Col108 mini-fibrils. A, the negatively stained mini-fibrils after 6h incubation at 37 °C. B, the negatively stained mini-fibrils after 24 h incubation at 37 °C. (taken from Kaur et al (40)).

Figure 1-8 Col108 mini-fibril model showing charge distribution. The three sequence units of the peptide are separated by black vertical bar; Blue vertical bars show positive charge residues; Red vertical bars show negative charge residues; The oval represents foldon domain.
Although Collagen mimetic peptide Col108 self-associate into mini-fibrils with \( d \)-banding periodicity, we still do not know what the determining factor for self-assembly of Col108 is. My thesis work focuses on the hypothesis that the periodicity of Col108 mini-fibrils is instigated by the repeating sequence units of Col108. We propose that a one-unit staggered arrangement of the triple helices would maximize both the hydrophobic and electrostatic interactions of associating helices and thus, stabilize the fibrils. I plan to test this hypothesis by studying the minimum number of repeating sequence units and investigating the roles of the sequence architecture.
Chapter 2. The \( d \)-periodic mini-fibrils of collagen mimetic peptide Col108

2.1 Introduction

Previous studies in our lab lead to the discovery of the self-assembly of Col108 to form mini-fibrils having a 35 nm axial repeating structure similar like that of the 64 nm \( D \)-period of fibrillar collagen. We call this axial structure of Col108 mini-fibrils the \( d \)-period. A previous work on the characterization of mini-fibrils by analytical centrifugation (AUC) demonstrate that molecular weight average of the Col108 samples in TES buffer drastically increased. This AUC data supports Col108 self-assemble in solution during fibrillogenesis other than in dry condition (40). I started my work on the further characterization of the structure of the Col108 mini-fibrils using TEM, atomic force microscopy and cryo-EM.

2.2 Methods and materials

2.2.1 The expression and purification of Col108

Col108 was expressed in \textit{Escherichia coli} strain BL21(DE3) using a customary pET32a(+) plasmid which carries the His-tagged thioredoxin-Col108 fusion protein (Trix-Col108). The fusion protein was first purified using Ni-NTA metal affinity resin (Qiagen Cat# 30210). The His-tagged thioredoxin was subsequently removed by thrombin cleavage, and separated from Col108 by reverse phase high-performance liquid chromatography (RP-HPLC). Some details of the purification procedure are described in reference (40).

During the first inoculation of the purification, a colony from the ampicillin plate of
transformed BL21(DE3) cell was picked and grew in 10 mL LB medium containing 0.05 mg/mL antibiotic ampicillin (Amp+) overnight in the shaker at 200 rpm at 37°C. Then, the 10 mL cell culture were grown in 1 L LB media containing 0.05 mg/mL Amp+ at 37°C until OD$_{600}$ reached a value of 0.6. The expression of protein was induced by adding 100 µL of 2 M isopropyl-β-D thiogalactoside (IPTG) to a final concentration of 0.2 mM IPTG. The culture was, subsequently incubated at 25°C in the shaker at 200 rpm overnight (around 16 h).

The cells of one liter liquid culture were harvested by centrifugation at 5000 rpm for 20 min at 4°C (Beckman Coulter centrifuge with JA-12 conical rotor, serial #03U1169). The supernatant media was removed and cell pellet was transferred into a 50 mL disposable centrifuge tube. The cell pellets were resuspended in 20 mL of pre-chilled pH 7.4 wash buffer (50 mM Tris, 300 mM NaCl). 1 mL of 20 mg/mL of fresh prepared lysozyme solution was added to lyse cell walls.

To prevent the proteolysis of protein, 200 µL of 1 M PhenylMethlysulfonyl Fluoride (PMSF) freshly made solution was added into cell lysate to a final concentration of 0.2 mM. 4 µL of 5 kU Pierce™ Universal Nuclease (Cat# 88700) was also added to degrade DNA. The cells were ruptured by sonication under cooling (Vibracell, six 60 second pulses, machine output 3, duty cycle 30%, microtip limit 4). The cell lysate was spun down at 10000 rpm for 20 min at 4 °C with JA-12 conical rotor. After centrifugation, the supernatant was poured gently into a 50 mL centrifuge tube. The resin was prepared by taking 4 mL of re-suspended Ni-NTA agarose resin slurry into a gravity flow Column (Fisher, Cat# K420401-1505). The resin was washed three times (50 mM Tris, 300 mM NaCl, pH 7.4). The resin was then added into the centrifuge tube containing cell lysate sealed with parafilm. The mixture was agitated at 4°C for 3 hours for His-tagged protein binding to the resin. And then the mixture was transferred into that gravity flow column. After the
resin was settled in the column, the supernatant was taken out from the top and labeled as non-binding sample.

The resin in the column was first washed using 20 mL wash buffer containing 50 mM Tris and 300 mM NaCl (pH 7.4). The resin in the column was washed the second time by 20 mL wash buffer containing 10 mM imidazole (pH 7.4). And then the resin was washed the third time by 20 mL wash buffer including 20 mM imidazole. The fusion protein was then eluted by 10 mL elution buffer (50 mM Tris, 300 mM NaCl and 300 mM imidazole, pH 7.4), and collected in fractions. Every fraction has around 1 mL. The collected samples were analyzed by running 12% polyacrylamide gel electrophoresis following the protocol in reference (41).

The eluted fractions showing target protein were pooled together about 7 mL. Combined sample was dialyzed against 50 mM Tris, 300 mM NaCl wash buffer with a pH of 7.4 to remove imidazole. The dialysis was carried out by using dialysis cassette (ThermoFisher Scientific Cat # 66110) in 4°C refrigerator. The buffer was changed once after around 4 hours and then left overnight at 4°C. The ratio of the volume of dialyzed sample and dialysis buffer was 1: 100.

After dialysis, the sample was cleavage using thrombin enzyme. The powder form of human Thrombin enzyme was purchased from Thermo Fisher (Cat number T6884). Each vial contains 120.4 NIH units of the enzyme, and each NIH unit equals to 0.324 μg of the enzyme (120.4 NIH=39 μg). The enzyme was dissolved in 100 μL of 50 mM Tris, 150 mM NaCl (pH 7) buffer. Based on the instruction manual, cleavage of fusion proteins can be carried out at a thrombin to fusion protein ratio of 1:500. One vial of human thrombin enzyme powder can cleave ~19 mg of protein. 20 μL dissolved thrombin was added into ~ 7 mL dialyzed sample and left at 4°C refrigerator for 24 h. The cleaved sample was run on a 12% polyacrylamide gel to if fusion protein was cleaved completely.
Cleaved protein and His-tagged thioredoxin were separated by high-performance liquid chromatography (Beckmann Coulter System Gold 126 solvent module, detector 168, SC100 fraction collector running 32 Karat 7.0 software) with reverse phase C8 preparative column (Vydac Cat# 208TP1010). 5 mL of the 7 mL combined sample was run using a time program graph where the initial acetonitrile gradient changes from 5% to 30% is done in 15 minutes followed by the gradient change from 30% to 45% in 30 minutes (0.5% per minute) and from 45% to 70% in 25 minutes (1% per minute). The elution gradient created for Col108 separation is shown (Figure 2-1). The 50 elution fractions were collected every 40 seconds after 22 nd minute of the run. The presence of protein in collected fractions was confirmed using 12% SDS-PAGE. The small portion of sample (~ 80 μL) were taken from each collected fraction sample and left at vacuum to dry for 3 days. The fractions that have the protein were pooled together and lyophilized, and stored as lyophilized powder.
2.2.2 Preparing Col108 solutions

The lyophilized protein powder was dissolved in 5 mM Acetic Acid (pH 4) to a final concentration of 1 mg/mL. The concentration was determined using NanoDrop 1000 Spectrophotometer with extinction coefficient of 0.232 for Col108 at 280 nm. The extinction coefficient of Col108 proteins was calculated by ProtParam online tool (https://web.expasy.org/protparam/). The samples were equilibrated in 4°C refrigerator for 7 days.

2.2.2 Conformation Studies

All CD experiments were performed using AVIV CD spectrometer (Biomedical, Model 202-01) with a temperature control system (a thermal controller Thermo Neslab Merlin M33 connected to a water bath), and quartz cuvettes with 1 mm optical path. Spectra were taken at 4°C, wavelength range 190 – 300 nm. Ellipticity measurements were corrected for buffer baseline using the same cuvette. CD data were analyzed using the Origin software. The raw CD data (in millidegree) are normalized to mean residue molar ellipticity (MRE):

$$[\theta] = \frac{\theta \times m}{c \times l \times n_r}$$

$\theta$ is the ellipticity in millidegree. $m$ is the molecular weight in g/mol. $c$ is the concentration in mg/ml. $l$ is the path length of the cuvette in mm. $n_r$ is the number of amino acid residues in the peptide.
2.2.3 Thermal stability studies

Thermal stability is determined using temperature melt experiment. The thermal melting profiles were obtained by monitoring CD value at a wavelength of 225 nm, as the temperature was increased from 4 °C to 65 °C and equilibration time of 2 min at each temperature (equivalent to an average heating rate of 0.3 °C/min). The fraction of folded sample was calculated based on the equation in Reference (42). The equation for calculation of fraction of folded is shown below:

\[
\text{Fraction of folded} = \frac{\theta_{\text{observed}} - \theta_{\text{monomer}}}{\theta_{\text{trimer}} - \theta_{\text{monomer}}}
\]

\(\theta_{\text{observed}}\) is observed ellipticity. \(\theta_{\text{trimer}}\) is calculated from the best linear fit equation for the trimer part of the melting curve. \(\theta_{\text{monomer}}\) is calculated from the best linear fit equation for the monomer part of the melting curve.

2.2.4 Fibril formation

To start fibril formation, Protein samples at 1 mg/ml in pH 4 buffer at 4°C were mixed with an equal volume of pre-chilled, double strength neutralization buffer (60 mM TES, 60 mM NaHPO₄, and 135 mM NaCl, pH 7.4). Thus, the final concentration of Col108 was 0.5mg/mL and the final composition of the fibrillogenesis buffer after mixing was 2.5mM acetic acid, 30 mM TES, 30 mM Na₂HPO₄, and 67.5mM NaCl, pH 7.4 (I= 0.09); this buffer will be called TES buffer for later discussion. Col108 in TES buffer was incubated in water bath at 37°C for 24 h. All solutions and buffers were made using ultrapure water.

2.2.5 TEM characterization of Col108 mini-fibrils

To prepare for TEM: After incubation at 37 °C, 3 µL of incubated sample was placed on a
400 mesh formvar carbon-coated copper grid (Electron Microscopy Sciences, Cat # FCF400-Cu). After 1 min, the grid was drained slowly with filter paper. And then 6 µL of 1% sodium phosphotungstate (the staining solution) was immediately added on the grid, after 4 min, excess staining solution was removed using filter paper, and then rinsed with deionized water. The grid was air-dried for at least one hour and then examined by Zeiss 902 electron microscopy or JEM-2100 electron microscope (Jeol Corp.) to visualize the morphology of higher order structures.

2.2.6 AFM characterization of Col108 mini-fibrils

For tapping mode AFM imaging, samples were prepared on a freshly peeled mica support. The upper layer of mica was peeled using a piece of standard Scotch tape. This step was repeated a couple times until we have a nice layer of mica on the tape indicating an entire layer of mica was exfoliated. And then 10 µL protein sample was added on the mica surface as soon as possible because if the fresh mica is exposed it starts reacting with air. The sample was incubated on mica for 2 min at room temperature and rinsed with 10 mL ultrapure water. After that the sample was dried quickly with Nitrogen gas (at least for 5 min). The mica supports were sent to Dr. Laurent Kreplak’s lab (Department of Physics and Atmospheric Science, School of Biomedical Engineering, Dalhousie University, Halifax, Canada) for imaging. All the imaging was carried out using a tapping mode AFM and scanner (Agilent 5500). All image analysis and measurements were performed using SPIP 6.3.3 software (Image Metrology). The work was done in collaboration with Dr. Kreplak’s lab.

2.2.7 Cryo-EM characterization of Col108 mini-fibrils

The Cryo-electron microscopy (cryo-EM) study was carried out in ASRC CUNY, in collaboration with Dr. Tong Wang. Cryo-EM grids were prepared in an FEI Vitrobot at 20 °C with
the relative humidity set to 100% and the blotting force set to zero. The samples (3 µL) of peptides after incubation in TES buffer were pipetted onto a freshly glow-discharged 400 mesh lacey carbon grid (TED PELLA INC. Cat#01824). The sample solution was incubated on the EM grid for 2 minute, blotted for 4.5 seconds before being plunged into liquid ethane that was pre-cooled by liquid nitrogen. The cryo-EM grids were then transferred to and stored in liquid nitrogen. When needed, they were transferred in liquid nitrogen into a Gatan 626 cryo-specimen holder and then inserted into the microscope. The specimen temperature was maintained at −170 °C during the data collection. Cryo-EM imaging was performed in an FEI TITAN Halo TEM operating at 300 kV.

2.2.8 Sample preparation of commercial collagen

Commercial collagen (type I collagen from rat tail, Sigma-Aldrich C7661) was first dissolved in 5 mM acetic acid to concentration of ~ 1 mg/mL and then mixed with the same volume of double strength neutralization buffer. The sample was prepared for fibril formation with the same procedure of Col108. The difference is that commercial collagen in TES buffer was incubated in water bath at 37°C for 2 h. The EM grids were prepared as described for Col108.

2.3 Results

2.3.1 The purification of Col108

The purification of Col108 is shown in Figure 2-2. The fusion Col108 monomer band (black box) and fusion Col108 trimer band (purple box) were seen in samples taken before dialysis and taken after dialysis. The position of fusion Col108 monomer and trimer are usually higher than their expected molecular weight position. After thrombin cleavage, fusion Col108 monomer band and fusion Col108 trimer band disappeared, and Col108 monomer band (red box) and Col108
trimer band (green box) were seen on the gel. 52 kDa fusion Col108 monomers (fusion Col108 trimer ~156 kDa) were efficiently cleaved into 38.3 kDa Col108 monomers (Col108 trimer ~114.9 kDa) by thrombin enzyme. On the gel, the migration of Col108 bands does not correlate with their molecular weights markers because triple helical protein is rod-like structure, and moves differently from globular proteins. The position of globular protein thioredoxin (13.7 kDa) highlighted by yellow box in lane 5 of Figure 2-2 is near its expected molecular weight position on the gel. There are additional bands presumably due to the nonspecific binding of Ni-NTA resin. We saw the bands of Trix-Col108 trimer and Col108 monomer on the gel because the reduction of protein trimers using reducing agent DTT is not complete and some triple helices are cross-linked through disulfide bonds.

Figure 2-2 SDS-PAGE (12%) analysis of thrombin cleavage of Col108 fusion protein. Samples before cleavage are in lane 3 and 4. Sample after cleavage is in lane 5. Lane 1 and 2 are high range protein makers and low range protein makers, respectively. Black box and purple box respectively highlight Col108
fusion protein monomer and trimer. Red box and green box respectively highlight Col108 monomer and trimer form. Yellow box labels thioredoxin.

After separation from the His-tagged thioredoxin using HPLC, the purified Col108 was confirmed by SDS-PAGE (Figure 2-3); only Col108 monomer (red box), dimer and trimer (green box) bands were seen on the gel. The molecular weight of His-tagged thioredoxin and thrombin enzyme are ~ 13.7 kDa and ~ 37.4 kDa, respectively. No bands that correlate with thrombin enzyme are present on the gel. The gel in Figure 2-3 was over-run. We have other data indicating that thioredoxin band is not present in these samples. No other contaminants are shown either. This result demonstrates that the protein sample was very pure.
2.3.2 Characterization of Collagen-Mimetic Peptide Col108

The triple helix conformation of collagen is characterized by a positive maximum peak at 225 nm and a deep negative peak at 197 nm on a CD spectrum (43). Lyophilized Col108 powder was dissolved in 5mM Acetic Acid (HAc, pH 4) at a concentration of ~1 mg/mL. The sample was equilibrated about one week in 4°C refrigerator to form triple helix structure. The concentration used for CD is diluted to 0.4 mg/mL. The triple helix conformation of Col108 was confirmed by the CD spectrum showing a small positive peak at 225 nm and a deep negative peak at 197 nm at 4°C (open circle, Figure 2-4). When temperature was increased to 47°C, CD spectrum of Col108 resembled that of a random coil having a shallow negative peak ~190 nm, and no positive peak at 225 nm (filled circle, Figure 2-4). The Col108 is unfolded at this temperature and there is no triple helix conformation in Col108 at 47°C.
2.3.3 The thermal stability of Col108

The temperature induced change of CD signal at 225 nm which is the characteristic maximum for triple helix was used to study the thermal stability of Col108. As Col108 sample is heated, the mean residue ellipticity decreases indicating a transition from triple helices to random coils. The concentration of the sample used for this study is 0.4 mg/mL Col108, with a temperature range from 4°C to 65°C. The melting temperature \( T_m \) which is the temperature at the mid-point of the transition, was determined to be 41°C (Figure 2-5). This result suggests that Col108 triple helix is stable at 37°C (the incubation temperature) during mini-fibril formation.

Figure 2-4 CD spectra of Col108. CD spectra of 0.4 mg/mL Col108 in 5 mM acetic acid were respectively taken at 4°C (open circle) and 47°C (filled circle).
2.3.4 The \( d \) periodicity of mini-fibrils of Col108 characterized by TEM

In order to better understand the nature of the \( d \)-periodic mini-fibrils I first studied the \( D \)-periodic fibrils of natural collagen using TEM. The cross-striated type I collagen fibrils with 67 nm \( D \)-period were observed under Zeiss 902 electron microscopy (Figure 2-6). The \( D \)-period includes \( \sim 27 \) nm the overlap region (light band) and \( \sim 40 \) nm gap region (dark band). The fibrils are \( \sim 60 \) nm to \( \sim 120 \) nm in diameter. The length of fibrils is difficult to identify because the fibrils are very long and overlapped under TEM.
The axial repeating structure of Col108 mini-fibrils looked very much like that of the D-period of natural collagen fibrils when examined using TEM, but smaller in size. The banding patterns of Col108 mini-fibrils were clearly observable under JEM-2100 electron microscopy (Figure 2-7). The periodicity of banding pattern is ~ 35 nm estimated from the negatively stained fibrils: including a dark band of ~ 25 nm and a light band of ~ 10 nm. The periodic mini-fibrils examined using TEM are ~800 nm to ~ 1 µm in length and ~ 50 nm to ~ 75 nm in diameter.

Figure 2-6 The TEM images of type I collagen fibrils from rat tail. (A-B) the negatively stained commercial collagen fibrils after 2 h incubation at 37 °C. The magnifications are 5000 X and 3000 X, respectively.
2.3.5 The $d$ periodicity of mini-fibrils of Col108 characterized by AFM

The study of the Col108 mini-fibrils by AFM were carried out in collaboration with Dr. Laurent Kreplak’s lab in Canada. Three micas of Col108 were prepared and the surface of micas were examined by AFM. All of three micas showed plenty of mini-fibrils with $d$ periodicity. The topography images of Col108 fibril samples after 24 hours of incubation at 37°C are shown in Figure 2-8. The mini-fibrils have similar morphological features to those observed using TEM. In the AFM micrographs, the mini-fibrils are ~ 800 nm to ~ 1 µm in length and ~ 25 nm in diameter.
according to the scale bar. The diameter of mini-fibrils in AFM micrograph is smaller than that in TEM micrograph. The TEM image is not high resolution and involved a lot of staining. The diameter of mini-fibrils measured by AFM is more reliable. The $d$-periodic band of mini-fibrils is strongly apparent in the phase channel, compared with height image. The $d$-periodicity banding pattern is estimated ~ 35 nm observable in AFM images. These AFM data support that Col108 under appropriate condition forms cross-striated mini-fibrils with $d$-periodicity.

Figure 2-8 AFM topography images of Col108 mini-fibrils. Acquired height and phase images are of Col108 in TES buffer deposited onto freshly split mica. Scale bar is 500 nm.

### 2.3.6 Col108 mini-fibrils characterized by cryo-EM

Cryo-EM is a technique to prepare samples in a more native-like hydrated state. In the preparation, the protein samples were brought to the solid state without dehydration by fast freezing, and the sample was maintained at low temperature during transferring and observation in the EM (44). I examined six grids made from different solutions and two purification samples. There are no discernable mini-fibrils and single triple helices on the first two of grids observed by
cryo-EM. It may be caused by low concentration of Col108 mini-fibrils samples. I concentrated the Col108 samples in TES buffer from another protein purification by centrifugation at 2000 rpm for 5 min and taking samples from the bottom. Four grids were made for cryo-EM examination. Many flattened bundles of single triple helices are observed in cryo-EM micrographs (Figure 2-9). We think the mini-fibrils had formed, but fell apart during freezing to create the imaged flattened-bundles of triple helix. The dissociation of Col108 mini-fibrils might be induced by the degradation of mini-fibrils by freezing ((45), personal communication with Dr. Joachim Frank). Finding a way to stabilize the mini-fibrils during fast freezing is a solution to improve the characterization of mini-fibrils by cryo-EM. We are planning to continue this work in collaboration with Dr. Tong Wang at CUNY ASRC.

![Cryo-EM image of Col108 aggregates. Scale bar is 200 nm.](image)

Figure 2-9 Cryo-EM image of Col108 aggregates. Scale bar is 200 nm.
2.4 Summary

The self-assembly of collagen mimetic peptide Col108 into higher order structure under appropriate conditions using TEM, AFM and cryo-EM. The prerequisite of fibril formation is that Col108 forms triple helical structure. The data of thermal stability indicate that Col108 triple helix is stable at fibril formation temperature. The AFM data reproduced the earlier findings (40) but at higher resolution, and with better quality. The cryo-EM appeared to capture the disassembly of the mini-fibrils. We are working to find a way to stabilize the mini-fibrils or different freezing approaches for further investigation.

While this part of the work confirmed the self-assembly of Col108 mini-fibrils, the mechanism of the fibril assembly is not clear. We need to explore if the repeating sequence unit is a crucial factor for periodic mini-fibrils and what the minimum number of repeating units for the self-assembly of the $d$-period fibrils is. The major part of my thesis focused on addressing these two questions.
Chapter 3. The self-assembly of a triple helix with two sequence units – the sufficient conditions for the formation of mini-fibrils with the \(d\)-period

3.1 Introduction

3.1.1 Sequence design of 2U108 and 1U108

The study of Col108 have shown that, under appropriate buffer condition, Col108 self-assembles into mini-fibrils possessing the characteristic 35 nm \(d\)-periodicity through a unit-staggered arrangement (2). The triple helix of Col108 has a built-in sequence periodicity that contains three pseudo-identical repeating sequence units. A unit-staggered model was proposed to indicate that because of the identical, repeating sequence units, a mutual staggering of one-sequence unit of associating triple helices represent the unique, most stable structure of self-assembly. Based on the unit-staggered model, peptide with two units should self-assemble into mini-fibrils with the same periodicity of Col108 mini-fibrils. While peptide without any long-range repeating sequences will not be able to form the \(d\)-staggered mini-fibrils.

Two new peptides 2U108 and 1U108, consisting of, respectively, two and one sequence repeats were generated for this study. The gene constructs of 2U108 and 1U108 were generated by a graduate student (Sam Wong) in our lab using the original Col108 plasmid in reference (46). The schematic diagram of the generation of the expression constructs is shown in Figure 3-1.
Figure 3-1 The construct of 2U108 and 1U108. For generating 2U108, the second collagenous unit and (Gly-Pro-Pro)_4 were removed. The third collagenous unit and (Gly-Pro-Pro)_4 were deleted from 2U108 to make 1U108. Cysteine knots (purple), (Gly-Pro-Pro)_4 (green), Col108 domain (blue) and foldon domain (red) are highlighted in box. The restriction enzyme sites (BamHI, XbaI, KpnI and EcoRI) are shown by black arrows.

The whole amino acid sequence of 2U108 and 1U108 are shown in Figure 3-2 and Figure 3-3, respectively. Collagen mimetic peptide 2U108 contains 294 amino acid residues that includes two identical sequence units in tandem with a nucleation domain, a sequence for Cys-knot (Gly-Pro-Cys-Cys) at both the C- and the N- termini of the triple helix domain and a C-terminal foldon domain. Peptide 1U108 contains 174 amino acid residues that is composed of one sequence unit ((Gly-Pro-Pro)_4 and Col-domain) with a nucleation domain, a sequence of (Gly-Pro-Cys-Cys) incorporated at both ends of triple helical domain and foldon domain at C-terminus. The main difference between 2U108 and 1U108 sequence is that 2U108 has two repeating sequence units and 1U108 only have one repeating sequence unit. The cysteine residues are used to form interchain disulfide bonds to link peptide monomer chain together to form triple helix structure.
The foldon domain at C terminus serves as trimerization domain for triple helix folding. 2U108 is a 27 kDa triple helical peptide. 1U108 is a 16 kDa triple helical peptide, comprised of a singular sequence unit, with a C-terminal foldon domain.

Figure 3-2 The amino acid sequence architecture of 2U108. (I) The overall sequence architecture of 2U108. The repeating sequence units are marked with double headed arrows. The amino acid sequence of the Col domain is shown in an expanded view. The residues introduced by restriction enzyme sites are highlighted by red box. The cysteine knots are shown in pink, (Gly-Pro-Pro)_4 in green, and foldon in red, respectively. (II) The entire amino acid sequence of 2U108, with the cysteine knots, (Gly-Pro-Pro)_4 and foldon highlighted in pink, green and red, respectively.
Figure 3-3 The amino acid sequence architecture of 1U108. (I) The overall sequence architecture of 1U108. The repeating sequence units of 1U108 are marked with double headed arrows. The amino acid sequence of the Col domain is shown in an expanded view. The residues introduced by restriction enzyme sites are highlighted by red box. The cysteine knots are shown in pink, (Gly-Pro-Pro)₄ in green, and foldon in red, respectively. (II) The entire amino acid sequence of 2U108, with the cysteine knots, (Gly-Pro-Pro)₄ and foldon highlighted in pink, green and red, respectively.

3.2 Materials and Methods

The expression and purification of 2U108 and 1U108 were under the similar conditions as for with Col108 as described in chapter 2 and in reference (40). The extinction coefficient of 0.322 for 2U108 and 0.538 for 1U108 at 280 nm are calculated using the online tool Protparam. The characterization, including the folding, the thermal stability study, and fibril formation, and the characterization using TEM are all described in Materials and Methods of Chapter 2 and in reference (40).

3.3 Results

3.3.1 The triple helical peptides 2U108 and 1U108

The samples were examined using circular dichroism spectrometer (CD) to check for the conformation of triple helix. The proteins were equilibrated to ensure the folding into triple helix. 1 mg/mL equilibrated proteins solution were diluted to 0.4 mg/mL or 0.5 mg/mL for examining the folding. CD spectra of 2U108 and 1U108, recorded from 300 nm to 190 nm, show typical triple helical spectra with a positive peak at 225 nm and a minimum peak at 197 nm (Figure 3-4). We diluted a stock solution of 1U108 of ~1mg/mL in acetic acid buffer to 0.5 mg/mL to test the
Because we did not have enough 2U108 sample for the study of fibril formation, we did not dilute the 1mg/mL 2U108 solution to 0.5 mg/mL. The data of Col108 triple helix is included for comparison. The Rpn values (the ratio of the value of positive peak to that of the negative peak) of 2U108 (~ 0.09) and 1U108 (~ 0.09) are comparable to that of Col108 (40).

![Circular dichroism spectra of 2U108 and 1U108](image)

Figure 3-4 Circular dichroism spectra of 2U108 and 1U108. CD spectra of Col108 (open black circle), 2U108 (red square) and 1U108 (blue triangle) in 5 mM acetic acid were taken at 4°C. The concentration of Col108, 2U108 and 1U108 are 0.4mg/mL, 1mg/mL and 0.5 mg/mL, respectively.

The thermal stability of 2U108 and 1U108 was estimated to be 41°C (Figure 3-5) for both peptides. The thermal stability of Col108 is around 41°C and included for comparison. These three peptides have different size, and different number of amino acid residues, but their thermal stabilities are similar. Thermal stability depends on amino acid sequence but no direct relation to
the length of peptides, when the peptide is larger than 45 residues (33,47). Thermal stability of human type I collagen is \( \sim 37^\circ C \) and contains hydroxyproline in the Y position of \((Gly-X-Y)_n\) tripeptide (48). Foldon domain at C-terminal end of the mimetic peptides, cysteine knots at both ends of triple helix domain and the high content of the charged residues may contribute to the increased thermal stability (49). The measurement of conformation and thermal stability of these two collagen mimetic peptides indicate both peptides are sufficiently stable for the study of self-assembly.

Figure 3-5 Temperature melting curve of Col108, 2U108 and 1U108 in 5mM acetic acid from 4°C to 55°C. The concentrations of Col108, 2U108 and 1U108 are 1 mg/mL. The thermal unfolding of Col108 (open black circle), 2U108 (red square) and 1U108 (blue triangle) triple helices was monitored by at 225 nm, and normalized to fraction of folded.
3.3.2 Self-assembly of mini-fibrils

First, I studied the self-assembly of 2U108. For TEM examination, 20 grids were prepared from one purification of 2U108. Under electron microscopy, we observed that 2U108 formed smooth mini-fibrils, similar to those observed for Col108 (40). The mini-fibrils are about 500 nm to 1 µm in length and ~ 25 nm to ~ 75 nm in the diameters of the central part of the mini-fibrils. The images of 2U108 mini-fibrils shown in Figure 3-6 were examined by Zeiss 902 electron microscopy (City College, CUNY). The banding pattern of these mini-fibrils are visible, but cannot be clearly measured because the resolution of these images is low. To get a better resolution of the structure, I moved on to use JEM-2100 electron microscopy. I prepared 10 grids from another purification of 2U108. Under high resolution TEM, it is clear that 2U108 mini-fibrils have the same 35 nm $d$-periodicity as that of Col108. The light band (overlap region) is ~ 10 nm and dark band (gap region) is ~ 25 nm in negatively stained 2U108 fibrils (Figure 3-7). The TEM image of 2U108 in 5 mM acetic acid before transferring to the TES buffer was also examined for comparison (Figure 3-6(F)). This image shows a uniform background of 2U108 triple helices with no mini-fibrils. 2U108 single triple helix is expected to be ~ 85 nm in length and ~ 1-2 nm in diameter. Some 2U108 monomers may exist according to the size of the structure.
Figure 3-6 The TEM images of 2U108. (A-B) the negatively stained 2U108 mini-fibrils after 24 h incubation at 37 °C at the magnification of 30000 X. (C-D) the negatively stained 2U108 mini-fibrils after 24 h incubation at 37°C at the magnification of 50000 X. (E) the negatively stained 2U108 mini-fibrils after 24 h incubation at 37°C at the magnification of 20000 X. (F) the TEM image of 2U108 in 5mM HAc.

Figure 3-7 The TEM images of 2U108 mini-fibrils. A-B, the negatively stained 2U108 mini-fibrils examined by JEM-2100 electron microscopy after 24 h incubation at 37°C at the magnification of 10000 X.
For the study of 1U108, I prepared ~ 20 grids from two different purifications of 1U108 and examined them using Zeiss 902 electron microscopy. No fibril-like assembles were found in any grids (Figure 3-8). A 1U108 triple helix is expected to have a length of about 50 nm, and ~ 1-2 nm in diameter. Some aggregates of 1U108 formed in different shapes can been seen via TEM. These aggregates are thicker than single 1U108 triple helical molecule and ~ 200 nm in length. They are likely formed by lateral association of the triple helices, but do not follow any specific pattern. Compared to the TEM images of 2U108 at the same magnification, the size of 1U108 aggregate is smaller than that of 2U108 mini-fibrils. Because of the lack of any specific, identifiable structural features of the aggregates of 1U108, we consider them as non-specific aggregates.
Figure 3-8 The TEM images of 1U108. (A-B) the negatively stained aggregates after 24 h incubation at 37 °C at the magnification of 30000 X. (C-D) the negatively stained aggregates after 24 h incubation at 37 °C at the magnification of 50000 X.
3.4 Summary

The unit-staggered model requires a triple helix with at least two repeating units like 2U108 to produce mini-fibrils with the \(d\)-period. One-unit staggered arrangement maximizes the interactions between the associating helices by optimal alignment of interacting residues to stabilize the staggered structure, and thus, makes the arrangement the most stable structure compared to other staggered arrangements. As expected, our results show collagen mimetic peptide 2U108 with two repeating sequence units self assembles into mini-fibrils with 35 nm \(d\)-periodicity. The formation of the 2U108 mini-fibrils demonstrated that there are robust interactions between repeating sequence between the associating helices during to support one-unit staggering self-assembly mechanism. In contrast, 1U108 which has only one sequence unit failed to form any fibril assemblies having the \(d\)-period. The interactions between triple helices of 1U108 might lead to different kinds of aggregates without periodic banding pattern. The \(d\)-periodic mini-fibrils of 2U108 offers an alternative sequence architecture to design new collagen mimetic peptide. Being smaller, 2U108 also appears to have higher expression yields and improve the efficiency of our research. The finding of 2U108 and 1U108 confirmed our understanding of Col108 \(d\)-periodic mini-fibrils. In the next, we will study if any specific residues of the tandem repeating units play deterministic roles in the self-assembly of the collagen mimetic peptide.
Chapter 4. Self-assembly of triple helical peptides Col877 and Col108R

4.1 Introduction

Col108 having three repeating sequence units and 2U108 having two repeating sequence units self-assemble into identical mini-fibrils with $d$-periodicity, while 1U108 with one sequence unit does not. The length of triple helix formed by one sequence unit of Col108 and 2U108 is approximately 35 nm ($d$-periodicity). The unit-staggered arrangement explains the size of the overlap and gap region of mini-fibrils. At least two repeating units are necessary for the unit-staggered model to form periodic mini-fibrils. Col108 and 2U108 share the same repeating sequence unit - the (Gly-Pro-Pro)$_4$-Col-domain. To further understand the involvement of specific residues in the Col-domain during the self-assembly of Col108 and 2U108, we created two new collagen mimetic peptides.

Col108R which has the same amino acid composition of that of Col108 but lacks the sequence periodicity; and Col877 which has the same repeating sequence architecture of that of Col108 but has a completely different sequence in each repeating sequence unit. Based on our one-unit staggered arrangement, Col877 with three repeating sequence units may self-associate to mini-fibrils with $d$-periodicity. Col108R which does not have sequence periodicity should not form fibrils.

4.2 Methods and Materials

4.2.1 The genes of Col877 and Col108R

The genes of Col877 and Col108R were synthesized with two restriction enzymes sites
BamHI and EcoRI at 5’ and 3’ ends, respectively, by GenScript Corporation. The genes were codons optimized for bacterial expression. We provided the expression plasmid. The synthesized genes were directly cloned into the expression vector pET32a(+) using a 5’ BamHI and a 3’ EcoRI restriction enzymes. The sequences of the genes were confirmed by GenScript Corporation (Appendix). After we received the plasmids, we double checked the sequence of the genes. The plasmids were transformed into DH5α E. coli competent cells (Invitrogen) and purified using Wizard Plus SV Minipreps DNA purification kit (Promega Cat# A1330). The concentration of plasmids was measured using Nanodrop 1000 Spectrophotometer (Thermo Scientific). The ratio of 260/280 is ~ 1.9. The Col877 plasmid was sequenced using the forward primer of T7 promotor and the reverse primer of T7 terminator (provided by Genewiz Corp.). The Col108R plasmid was purified and sequenced using two synthesized primers: primer 5’-ATCCGTGGTATCCCGACTCT-3’ and primer 5’-GGGCGATACCGGCCCGGA-3’ targeting thioredoxin region and second repeating sequence unit region, respectively. Both primers were synthesized by Macrogen Corporation and the sequencing was done by Macrogen. Upon receiving the sequence results, the DNA sequencing data were further analyzed and confirmed using bioedit software. The products of the genes were fusion proteins in the form of His-tagged thioredoxin-Col877 (Trix-Col877) or His-tagged thioredoxin-Col108R (Trix-Col108R). There is a thrombin cleavage site between the Trix and the cloned protein which is subsequently used to remove the His-tagged thioredoxin.

4.2.2 Expression and Purification

The GenScript provided the initial expression and analysis of Col877 (result in appendix). Small scale purification was used to analyze and optimize the expression of Col108R. The plasmid was transformed into bacterial BL21(DE3) following BL21(DE3) transformation protocol. To test
the expression, an isolated colony of the transformed BL21 was first inoculated in 2mL LB cell medium containing 0.05 mg/mL ampicillin (Amp+ LB) and grew overnight in the shaker at 200 rpm at 37°C. During the second inoculation, the 200 µL cell cultures were inoculated into 10 mL Amp+ LB media and grown at 37°C. The cells were induced by adding 1 µL of 2 M IPTG to a final concentration of 0.2 mM IPTG when the optical density (600 nm) reached 0.4 and 0.6. After induction, the cells were grown in the shaker at 200 rpm overnight (around 16 h) at ~20°C. The 10 mL cell cultures were spun down at 7500 rpm for 10 min at 4°C. The supernatant was discarded and cell pellets were resuspended in 200 µL wash buffer (50 mM Tris, 300 mM NaCl) and 10 µL of 20 mg/mL of fresh prepared lysozyme solution was added to lyse cell wall. To prevent the proteolysis of protein, 0.4 µL of 0.1 M freshly made PMSF solution was added into cell lysate to a final concentration of 0.2 mM. 0.2 µL of 5 kU Universal Nuclease was also added into lysed cells. The cells were further ruptured by sonication (Vibracell, fifteen second pulses for twice, machine output 3, duty cycle 30%). The cell lysate was spun down at 5000 rpm for 10 min at 4°C. The supernatant was separated for purification.

HisPur Ni-NTA Spin Columns (0.2mL resin bed, ThermoFisher Scientific, Cat # 88224) were used to purify His-tag fusion protein. The protein extracts were added into the column and mixed on an orbital shaker for 30 minutes at 4°C. The columns were centrifuged at 700 x g for 2 minutes and the flow-through was collected in centrifuge tube as non-bonding. The resin was washed with 400 µL wash buffer (50 mM Tris, 300 mM NaCl) wash buffer and centrifuged again, the spin-down solution is collected and labeled “first wash”. The resin was washed with another 400 µL wash buffer additioning 10 mM imidazole, centrifuged and collected as “second wash”. The resin was washed by 400 µL wash buffer containing 20 mM imidazole and collected as the “third wash”. Then, the His-tagged proteins were eluted from the resin by adding 200 µL elution buffer (50 mM Tris, 300 mM NaCl and 300 mM imidazole) and centrifuged at 700 × g for 2
minutes. This step was repeated two more times and each fraction was collected in a separate tube and labeled as “elution”.

The same small scale procedure was also used to analyze the expression in JM109(DE3) cells.

Based on the small scale purification results, BL21(DE3) cells were used for further expression and purification both Col877 and Col108R. The gene products are, respectively, His-tagged thioredoxin Col877 and His-tagged thioredoxin Col108R. The fusion protein was first purified using Ni-NTA metal affinity resin (Qiagen Cat# 30210). The thioredoxin was subsequently removed by thrombin cleavage, and separated from protein by reverse phase HPLC. Details of the purification procedure are described in reference (40) and methods and materials of Chapter 2.

There are two differences between the purification of Col108 and the purification of Col877 and Col108R. First, for Col108, the culture was grown at ~ 25°C in a shaker after induction. For Col877 and Col108R, after the expression of protein was induced by 0.2 mM IPTG, the culture was, then grown at ~ 20°C in a shaker cooled with ice at 200 rpm overnight (around 16 h). (The shaker does not have cooling system. The temperature started at 15°C and maintained for a few hours and then increased to ~20°C.) Second, for Col108, 5 mL of ~ 7 mL cleaved protein sample was run on RP-HPLC. For Col877 and Col108R, cleaved protein sample was concentrated by centrifugal filter tube with a molecular weight cutoff membrane of 30 kDa (Amicon Ultra centrifugal filter, Cat #UFC903024) in a fixed angle rotor at 4000 x g for 10 minutes at 4°C to decrease the total volume ~ 5 mL to 3 mL. 3 mL concentrated sample was run on RP-HPLC using the same time gradient used for Col108 purification (Figure 2-1 of Chapter 2).
4.2.3 The structural characterization and the study of the self-assembly

The folding and thermal stability of Col877 and Col108R, and characterization of Col877 mini-fibrils and Col108R were followed similar procedures with that of Col108 shown in methods and materials of Chapter 2.

4.3 Results

4.3.1 The design of the peptides

The Col877 peptide was designed to have three repeating sequence units. In each sequence unit, the col-domain in Col108 was replaced by the C877 domain which consists of the segment of residues 877-984 of the α1 chain of type I collagen (Figure 4-1). The complete amino acid sequence is given in Figure 4-2. The C877 domain has 108 residues, which is the same as that of Col-domain. The region 877-984 was chosen because the segment of natural collagen contains no Hyp and is thus, a better candidate for modeling using bacterial expression system. Additionally, our lab previously studied the residues in the 877-939 region extensively in another recombinant construct and have first-hand knowledge of its thermal stability and conformational properties (42). Overall, Col877 has the same periodic sequence architecture as that of Col108 but different amino acid composition. To relate the amino acid sequence of a collagen triple helix to its thermal stability, we used the online tool by Dr. Persikov (http://compbio.cs.princeton.edu/csc/) (39). The thermal stability of the 108 residue C877 domain was predicted to be around 37.4°C. This temperature does not reflect the true thermal stability for Col877, which has three C877 domains and other residues, but it is used as an indication of the relative stability for the peptide. For comparison, the estimated $T_m$ of Col-domain of Col108 is 38.2°C. We, thus, expect the Col877 should have similar stability as that of Col108.
Figure 4-1 Schematic expression of the gene constructs of the collagen mimetic peptides. Cysteine knots in purple, (Gly-Pro-Pro)₄ in green, C877 domain in yellow, Col domain in blue, randomized Col domain in orange, reversed Col domain in white and foldon domain in red are shown. The Col-domain, C877 domain, reversed Col-domain and randomized Col-domain all have 108 amino acid residues.

Figure 4-2 The amino acid sequence architecture of Col877. (I) The overall sequence architecture of Col877. The repeating units are marked with double headed arrows. The residues introduced by restriction enzyme sites are highlighted by red box. The cysteine knots are shown in pink, (Gly-Pro-Pro)₄ in green, and foldon in red, respectively. (II) The entire amino acid sequence of Col877, with the cysteine knots, (Gly-Pro-Pro)₄ and foldon highlighted in pink, green and red, respectively.
Peptide Col108R was designed to have the same acid composition as that of Col108 but without the long range periodicity of the sequences. The gene of Col108R is generated replacing the second and the third Col-domain of the original Col108 with, respectively, a randomized sequence of the Col-domain, and the reversed sequence of the original Col-domain (Figure 4-1); the complete amino acid sequence is given in Figure 4-3. The randomized sequence of the Col-domain of Col108R was created by randomizing the residues at X and Y positions while keeping the Gly at every third position. Cautions were also made to avoid region with extensive residues of the same charges in order to avoid the charge repulsion in the structure. The randomized sequence of Col-domain was further optimized to keep predicted thermal stability around 37.6°C using the online tool by Persikov. The reversed sequence of original Col-domain was generated by reversing the residues at X and Y residues within each triplet. The estimated $T_m$ of reversed sequence is $\sim$36.6°C. In both peptides, the (Gly-Pro-Pro)$_4$ sequence in each sequence unit are kept. Thus, while Col108R has the same amino acid composition as that of Col108, the sequences in the three sequence are different from each other. Most importantly the sequence of Col108R loses the periodic features of Col108 (exclude the (Gly-Pro-Pro)$_4$ inserts).
Figure 4-3 The amino acid sequence architecture of Col108R. (I) The overall sequence architecture of Col108R. The three units of Col108R are marked with double headed arrows. The residues introduced by restriction enzyme sites are highlighted by red box. The cysteine knots were shown in pink, (Gly-Pro-Pro)$_4$ in green, randomized Col domain in orange, reversed Col domain in grey and foldon in red. (II) The entire amino acid sequence of Col108R, with the cysteine knots, (Gly-Pro-Pro)$_4$ and foldon highlighted in pink, green and red, respectively.

4.3.2 The sequencing of Col877 and Col108R genes.

The genes were synthesized and cloned directly into the expression plasmid using the service by GenScript. We provided the expression plasmid, and the gene was inserted using a 5’ BamHI and a 3’ EcoRI restriction enzyme sites. After receiving the plasmid of the two peptides, we sequenced them to confirm the sequence. Because of the large size of the genes (1170 bp) the direct sequencing of Col877 was not complete. The sequencing confirmed 870 base pairs of the Col877 genes shown in Appendix B. 300 base pairs in second repeating sequence unit of Col877 were not confirmed. Since GenScript provided a complete sequencing results of Col877 when the cloned plasmid was delivered, we accepted this sequencing results as being sufficient to confirm the accuracy of the plasmid. The complete sequence result of Col877 was shown in Appendix A.

To obtain a complete coverage of the Col108R gene by sequencing, we designed two forward primers. One primer was designed to start ‘forward’ chain reaction from the middle of thioredoxin so the sequencing would cover from the beginning region of Col108R to second
sequencing unit of Col108R. The other primer was designed to start ‘forward’ chain reaction from the second sequencing unit of Col108R. The design of this primer took into account that the sequencing results of Col877 showing the amplification stopped at second sequencing unit. The sequencing of Col108R using two designed inside primers resulted in complete coverage and thus, confirmed the plasmid shown in Appendix C.
4.3.3 The purification of Col877

The expression of Col877 was tested by GenScript using Western blot (data in Appendix D). We moved on to express it in BL21(DE3) host cell directly without further testing of the expression. The fusion protein Trix-Col877 was first purified using His-tag affinity resin. The SDS-PAGE gel was run to check if the results of elution using 300 mM imidazole (Figure 4-4). The samples of supernatant, cell pellet, non-binding, wash1-wash3 were included to help identify the target protein and to check the binding. The Trix-Col877 was not lost in cell pellet, non-binding and wash1-wash3 because there were no bands of the target protein shown in these fractions. No Trix-Col877 bands in cell pellets demonstrates that the target protein is soluble and also the cells were completely broken by sonication to release the protein into the supernatant. No Trix-Col877 bands in non-binding solution indicates that the binding of His-tag resin and Trix-Col877 are good enough. The particular supernatant sample was very sticky, and has low resolution; the Trix-Col877 could not be clearly identified in supernatant on the gel. Trix-Col877 were present in elution fractions shown in Lane8-Lane14 of Figure 4-4. The position of Trix-Col877 is higher than their expected molecular weight position. On the gel, migration of the protein bands does not correlate with their molecular weight markers because triple helical protein has a rod shape, and move differently from globular proteins. Some non-specific binding proteins were eluted with Trix-Col877. They were further removed by later purification procedures. The elutions that contains fusion protein Trix-Col877 were combined, and the imidazole was removed through dialysis.
Figure 4-4 SDS-PAGE (12%) analysis of purification of His-tagged Col877 fusion protein expressed in BL21(DE3) cells. On the gel, lane 2 is supernatant after sonication. Lane 3 is cell pellet after sonication. Lane 4 is non-bonding sample. Lane 5 is first wash using solution containing 50 mM Tris and 300 mM NaCl (pH = 7.4). Lane 6 is second wash using solution containing 50 mM Tris, 300 mM NaCl and 10mM Imidazole (pH = 7.4). Lane 7 is third wash using solution containing 50 mM Tris, 300 mM NaCl and 20mM Imidazole (pH = 7.4). Lane 8-14 are elution fractions. Lane 1 and 15 are low range protein makers and high range protein markers, respectively. Fusion Col877 monomer is labeled by blue box. Fusion Col877 trimer is identified by purple box.

Thrombin cleavage was used to remove the His-tagged thioredoxin. After thrombin cleavage, instead of a 51 kDa fusion Col877 monomer, we would expect to see a 38.2 kDa Col877 monomer and a 14 kDa His-tagged thioredoxin. The completion of the digestion was examined using electrophoresis (Figure 4-5). The fusion Col877 monomer band appeared to be present after 12 h cleavage (Figure 4-5, lane 2). After incubation of 24 h, the cleavage appeared to be complete;
the fusion protein is no longer present. Interestingly, there is another band at ~26 kD location, which also disappeared upon incubation with thrombin. We suspect this band represents an incomplete chain of Col877. Col877 may be degraded during purification; or Col877 proteins may be not fully translated during expression due to the limited number of t-RNA for Pro and/or Gly residues. Collagen like peptides have an unusual high content of Pro and Gly. We plan to use another proteinase inhibitor Cocktail (Clontech Cat# 635672), which is a mixture of protease inhibitors that contain Aprotinin, Bestatin, Leupeptin, Pepstatin A and PMSF. This Cocktail is likely more effective to protect proteins from being digested by endogenous proteases than PMSF alone. We are also trying to use a different host cell, BL21 Codon-plus, which has multiple copies of Pro-, Arg- and Gly-tRNA genes to improve the translation of Col877. The work is on-going.

Figure 4-5 SDS-PAGE (12%) analysis of thrombin cleavage of His-tagged Col877 fusion protein. SDS-PAGE was used to follow the thrombin cleavage reaction. Lane 1 is sample before cleavage. Lane 2 is
sample after 12 h cleavage by thrombin enzyme. Lane 3 is sample after 24 h cleavage by thrombin enzyme. Lane 4 is high range protein marker. Blue box highlights Col877 fusion protein monomer. Red box highlights Col877 monomer form. Yellow box labels thioredoxin.

In the final stage of purification, the Col877 was further separated from the His-tagged thioredoxin and thrombin enzyme by HPLC. The chromatograph using absorbance spectrum at a wavelength of 280 nm and 214 nm are shown in Figure 4-6 and Figure 4-7, respectively.

Figure 4-6 HPLC spectrum of Col877 after thrombin cleavage at 280 nm. Col877 elutes at 37% acetonitrile (30th min). His-tagged thioredoxin elutes at 58% acetonitrile (59th min). A peak related to Col877 on the spectrum is shown in a red box. A peak corresponding to His-tagged thioredoxin is shown in a yellow box.
Figure 4-7 HPLC spectrum of Col877 after thrombin cleavage at 214 nm. Col877 elutes at 37% acetonitrile (30th min). His-tagged thioredoxin elutes at 58% acetonitrile (59th min). A peak related to Col877 on the spectrum is shown in a red box. A peak corresponding to His-tagged thioredoxin is shown in a yellow box.

The SDS-PAGE of the elutions (Figure 4-8) confirmed that Col877 eluted starting at 37% acetonitrile (30th min - shown in red box). No bands that correlates with thrombin enzyme, or His-tagged thioredoxin are present are present on the fractions of Col877 (Figure 4-8 and Figure 4-9A).
The purification results of another purification of Col877 was shown in (Figure 4-9B). No His-tagged thioredoxin is shown at 14 kDa location. Again, the sample was pure. There is no His-tagged thioredoxin in the Col877 protein after HPLC separation.

Figure 4-8 SDS-PAGE (12%) analysis of fraction samples after HPLC separation. After thrombin cleavage, Col877 was further purified by HPLC. Lane 3-15 are fraction samples (1-13) after collection starting at 22 min. Red box highlights Col877 monomer form.
Figure 4-9 SDS-PAGE analysis of fraction samples after HPLC separation. (A) (continuing from Figure 4-10) Fractions 14 – 26 are in lane 3 – 15. Col877 monomer form is labeled by red box. (B) SDS-PAGE of fraction samples from another purification after HPLC separation. We expected to see thioredoxin around 14.2 kDa location of protein ladder.
4.3.4 The purification of Col108R

Small scale purification of Trix-Col108R was first conducted to improve the yield under different expression conditions using different cell lines (JM109 and BL21(DE3) cells). The results were analyzed using SDS-PAGE. No bands that correlate with Trix-Col108R were present in “non-banding”, “wash1” and “wash2” (Figure 4-10A). This result demonstrates that the binding of His-tagged resin and Trix-Col108R were good and no Trix-Col108R were lost in non-binding and wash for resin. The eluted fusion Col108R proteins were shown in lane 6 -13 (Figure 4-10B). The band of fusion protein under the condition of OD = 0.6 in BL21(DE3) cells was thicker than others. This result indicates that Col108R has better expression yield with induction at OD 0.6 in BL21(DE3) cells. We therefore, chose to use this condition for large scale purification.
Figure 4-10 SDS-PAGE (12%) analysis of small scale purification of Col108R in different conditions. (A) Lane 2 and 3 are low range and high range protein marker, respectively. Lane 4, 8 and 12 are respectively non-bonding, first wash and second wash samples expressed in JM109 and induced when OD = 0.4. Lane 5, 9 and 1 are non-bonding, first wash and second wash samples expressed in JM109 and induced at OD = 0.6. Lane 6, 10 and 14 are non-bonding, first wash and second wash samples expressed in BL21(DE3) and induced at OD = 0.4. Lane 7, 11 and 15 are non-bonding, first wash and second wash samples expressed in BL21(DE3) and induced at OD = 0.6. (B) Lane 2, 6 and 10 are third wash, first elution, and second elution samples expressed in JM109 and induced when OD = 0.4. Lane 3, 7 and 11 are third wash, first elution, and second elution samples expressed in JM109 and induced at OD = 0.6. Lane 4, 8 and 12 are third wash, first elution, and second elution samples expressed in BL21(DE3) and induced at OD = 0.4. Lane 5, 9 and 13 are third wash, first elution, and second elution samples expressed in BL21(DE3) and induced at OD = 0.6. Lane 1 is low range protein marker.

The purification results of Col108r using 300 mM imidazole are shown in Figure 4-11. Trix-Col108R were present in elution using 300 mM imidazole (lane1-lane8 of Figure 4-11). Trix-Col108R moved similarly on the gel as Trix-Col877 and Trix-Col108. The position of Trix-
Col108R band on the gel is higher than their expected molecular weight position, which is common for rod-like triple-helix proteins.

Figure 4-11 SDS-PAGE (15%) analysis of purification of His-tagged Col108R expressed in BL21(DE3) cells. SDS-PAGE was run to check the purification results. Lane 1-8 are elution samples. Lane 9 are low range protein makers. Trix-Col108R monomer is labeled by blue box. Trix-Col108R trimer is highlighted by purple box.

The completion of the cleavage of Trix-Col108R was examined using electrophoresis (Figure 4-12). After 24 h cleavage, the band of 52 kDa fusion protein Col108R monomer
disappeared and the band of 38.5 kDa Col108R monomer was present. The gel was run out of the bottom line so we cannot see the band of Trix at 14.2 kDa location in lane 3 of Figure 4-12. Two bands around 40 kDa position of protein ladder disappeared after thrombin cleavage. This may be caused by incomplete translation during expression or degradation by proteinase. This situation of Col108R is similar as Col877. The solution has been mentioned in the result section of Col877 purification. After thrombin cleavage, an unknown band was shown around 72 kDa location of protein ladder in lane 3 of Figure 4-12. We cannot identify what this band is.

![SDS-PAGE Image](image)

Figure 4-12 SDS-PAGE (15%) analysis of thrombin cleavage of His-tagged Col108R protein. SDS-PAGE was used to follow the thrombin cleavage reaction. Sample before dialysis is in lane 2. Sample after 24 h cleavage by thrombin enzyme is in lane 3. Lane 1 and 4 are low range and high range protein markers,
respectively. Blue box highlights Col108R fusion protein monomer. Red box highlights Col108R monomer form.

The elution gradient for Col108R purification using HPLC is the same with that of Col877. The chromatograph using spectrum at a wavelength of 280 nm and 214 nm are shown in Figure 4-13 and Figure 4-14.

Figure 4-13 HPLC spectrum of Col108R after thrombin cleavage at 280 nm. Col108R elutes at 37% acetonitrile (30th min). His-tagged thioredoxin elutes at 57% acetonitrile (58th min). A peak corresponding to Col108R on the spectrum is shown in a red box. A peak corresponding to His-tagged thioredoxin is shown in a yellow box.
Figure 4-14 HPLC spectrum of Col108R after thrombin cleavage at 214 nm. Col108R elutes at 37% acetonitrile (30th min). His-tagged thioredoxin elutes at 57% acetonitrile (58th min). A peak corresponding to Col108R on the spectrum is shown in a red box. A peak corresponding to His-tagged thioredoxin is shown in a yellow box.

The SDS-PAGE of the elution (Figure 4-15 and Figure 4-16) confirmed that Col108R peptide eluted started at ~ 37% acetonitrile (30th min - shown in red box). The His-tagged thioredoxin peak in Col108R absorbance spectrum is smaller than that in Col877 absorbance spectrum because Col108R sample was concentrated by centrifugal filter tube with a molecular weight cutoff membrane of 30 kDa. The molecular weight of His-tagged thioredoxin is ~ 14 kDa. Most of His-tagged thioredoxin is separated from Col108R through centrifugal filter tube. No bands that correlates with thrombin enzyme or His-tagged thioredoxin were present on the
fractions of Col108R after HPLC separation (Figure 4-16). There are no other contaminants on the gel. This result demonstrates that the Col108R sample was pure.

Figure 4-15 SDS-PAGE (15%) analysis of fraction samples after HPLC separation. Lane 1-9, 11 and 12 are fraction samples 1-8 and 15-17 after collection starting at 18 min. Lane 10 is low range protein marker.
Figure 4-16 (continuing from Figure 4-17) Fractions 18 – 24 are in lane 5 – 11. Col108R monomer form is labeled by red box. Col108R trimer and dimer are highlighted by green box. Thioredoxin is highlighted by yellow box. Lane 2 and lane 3 are samples taken from before thrombin cleavage and after cleavage, respectively. Trix-Col108R monomer and Col108R monomer are shown in blue and red, respectively. Lane 1 and 4 are low range and high range protein markers.

4.3.5 The folding and the thermal stability of Col877 and Col108R

CD spectra of Col877 and Col108R similar with CD spectra of Col108 demonstrated that Col877 and Col108R formed triple-helical structure (Figure 4-17). The thermal stability was monitored by the change of CD signal at 225 nm with the temperature from 4°C to 55°C. The temperature of the thermal transition ($T_m$) of Col877 is about 39°C (Figure 4-18). The $T_m$ of Col108R is about 38°C (Figure 4-18). Compared to $T_m$ of Col108, which has a $T_m$ of 41°C, thermal stability of Col877 and Col108R is slightly lower than that of Col108. While, Col877 triple helices
are stable at incubation temperature of fibril formation (37°C or 26°C), Col108R triple helices may not be stable at incubation temperature of 37°C, and for this reason, the fibril formation was done at 26°C.

Figure 4-17 Circular dichroism spectra of Col877 and Col108R. CD spectra of Col877 (red filled squares) and Col108R (blue filled triangles) were recorded in 5 mM acetic acid at 4°C and final concentration of 1mg/mL. CD spectrum of 0.4 mg/mL Col108 (black open circles) in 5 mM acetic acid 4°C is as a comparison.
Figure 4-18 Temperature melting curve of Col877 and Col108R from 4°C to 55°C. The thermal unfolding of 1mg/mL Col877 (red filled squares) and 1mg/mL Col108R (blue filled triangles) triple helix was monitored by at 225 nm, and normalized to fraction of folded. Temperature melting curve of 1 mg/mL Col108 (black open circles) is as a comparison.

4.3.5 Electron microscopy study of the self-assembly of Col877 and Col108R

4.3.5.1 Self-association of Col877

The fibrillogenesis was carried out in the same way as that of Col108. I have examined ~36 grids made from different solutions and from two different purifications of Col877. All staining solutions were fresh made within one week. I examined 20 grids made from protein samples of one purification. I have observed the same mini-fibrils from most of the 20 grids observed. It is
not easy to find the mini-fibrils on the grids due to the small quantity of mini-fibrils. If the concentration of protein samples is increased, it might solve this problem. For those I failed to see mini-fibrils I found large aggregates, but the banding pattern is not clear. This could be caused by ineffective staining. All results are shown in Figure 4-19(A-D). In Figure 4-19(A-B), there are some mini-fibrils with obscure banding pattern. The mini-fibrils are ~ 650 nm to 1µm in length and 40 nm to 95 nm in diameter. The exact length of mini-fibrils is difficult to determine because the ends of individual long mini-fibrils overlap together. The good TEM micrographs of mini-fibrils revealing clear banding pattern are in Figure 4-19(C-D). Overall, the Col877 mini-fibrils have the same banding pattern as that of Col108 (Figure 2-7) when examined using electron microscopy with negative staining under different magnification. The d-periodicity of banding pattern is ~ 35 nm estimated from the negatively stained fibrils, consisting of a dark band of ~ 25 nm (0.7d) and a light band of ~ 10 nm (0.3d).

I examined ~ 16 grids made from Col877 sample of another purification at two different incubation temperatures. The mini-fibrils are still found in most of these grids in both incubation temperature. For the incubation temperature of 37°C, the TEM images showing clear banding pattern are in Figure 4-19(E-H). The Col877 mini-fibrils have similar features with that of sample from previous purification. The mini-fibrils are about 700 nm – 1 um in length, with the diameters of the mini-fibrils varying between ~ 40 to ~ 90 nm. This result indicates that self-assembly of Col877 into mini-fibrils is reproducible. For the incubation temperature of 26°C, the mini-fibrils with the same banding pattern were found in TEM micrographs (Figure 4-20). The banding pattern on a few TEM images is not clear. This might be caused by ineffective staining. The mini-fibrils after incubation at 26°C are ~600 nm to 1µm in length and 35 nm to 95 nm in diameter.
Figure 4-19 The TEM images of Col877 mini-fibrils incubated at 37°C. (A-H) the negatively stained Col877 mini-fibrils after 24 h incubation at 37°C. A-D made from Col877 sample of one purification. E-H made from Col877 sample of the other purification. B, C, D, F and H are at magnification 10000 X. A, E and G are at magnification 6000 X.

Figure 4-20 The TEM images of Col877 mini-fibrils incubated at 26°C. (A-B) the negatively stained Col877 mini-fibrils after 24 h incubation at 26°C. A is at magnification 10000 X. B is at magnification 5000X.

Col877 in 5 mM acetic acid at 4°C was conducted as a control. No mini-fibril-like structure is found under the examination of TEM (Figure 4-21). An individual Col877 triple helix is approximately 120 nm long and 1.5 nm in diameter. Some thread looking structures are observed on the TEM images of Col877 in 5 mM acetic acid. These structures appear to be Col877 monomers based on their size. Also, there are a few aggregates shown on TEM micrographs. These
aggregates are a bit larger than Col877 monomer and have very different appearances from Col877 mini-fibrils.

Figure 4-21 The TEM images of Col877 in 5 mM acetic acid. (A-B) Col877 in 5 mM acetic acid at 4°C. A and B are at magnification 20000 X.
4.3.5.2 Self-association of Col108R

Col108R in TES buffer were carried out at different incubation time of 6 h, 12 h and 24 h at 37°C, and also at incubation time of 24 h at 26°C since its melting temperature is ~ 38°C, lower than those of Col108 and Col877. I have examined ~30 grids made from different samples prepared using of the peptide from two purifications, with incubation times 6 h, 12 h or 24 h at 37°C. No fibril-like assembles were observed on any of all grids (Figure 4-22). Under TEM, some aggregates were observed with 24h incubation time (Figure 4-22(A-B)). The diameter of the aggregates is ~10 nm, which is smaller than that of Col877 mini-fibrils. There is no banding pattern in the aggregates. Some aggregates that have non-regular morphology exist on the TEM images with another incubation time of 6 h and 12 h (Figure 4-22(C-F)). None of these aggregates looks like periodic mini-fibrils.

I also prepared the samples for self-association incubated at 26°C. I examined ~ 10 grids from two different purification samples. Under TEM, still no mini-fibrils-like aggregates were found on all grids (Figure 4-23). Peptide Col108 was shown to form d-periodicity mini-fibrils at both 37°C and 26°C incubation temperatures (40). All Col108R aggregates that found under TEM after incubation at 26°C are obviously different from the mini-fibrils of Col108 or Col877. This result support our hypothesis that the repeating sequence units are important for fibril formation.
Figure 4-22 The TEM images of Col108R at different incubation times. (A-B) No fibrils-like aggregates of Col108R after 24 h incubation at 37 °C. (C-D) Col108 aggregates after 12 h incubation at 37 °C. (E-F)
Col108 aggregates after 6 h incubation at 37 °C. A, C and E are at magnification 6000 X. B and D are at magnification 10000 X. F is at magnification 12000 X.

Figure 4-23 The TEM images of Col108R at the other incubation temperature. (A-B) Col108R after 24 h incubation at 26 °C. A is at magnification 6000 X. B is at magnification 10000 X.

4.4 Summary

Two new collagen mimetic peptides (Col877 and Col108R) were designed and built to test if sequence periodicity is crucial factor for the self-assembled mini-fibrils of Col108. Col877 has built-in repeating sequence units in its primary structure. Col108R does not have long range sequence periodicity in its sequence. The results indicate that Col877 triple helix can self-assemble into mini-fibrils with the same d-periodicity of those of Col108 mini-fibrils under appropriate conditions. Col108R triple helix, on the other hand, form stable triple helix but could not self-assemble to fibrils-like assemblies, even after reducing the incubation temperature to 26°C. These data support our hypothesis that the periodicity in the sequence architecture in Col108 are important for mini-fibril formation.
The $d$-period of the Col877 mini-fibrils indicate that the self-assembly is driven by the repeating sequence architecture, and not specific residues in the repeating sequence units. Col877 which has the same repeating sequence architecture as that of Col108 but having a completely different sequence also form mini-fibrils with the identical periodicity as that of Col108 mini-fibrils. The $\sim35\text{nm}$ $d$-period corresponds to a 123 residue pseudo-repeating sequence unit of Col877, and the overlap region ($0.3d$) of mini-fibrils is from C terminal foldon domain, $(\text{Gly-Pro-Pro-Pro})_4$ nucleation domain and Cysteine knots. This one-unit staggered assembly is most stable arrangement because hydrophobic residues interact with hydrophobic residues of triple helices and charged residues interact with charges residues. In the contrary, Col108R triple helices that lose long range sequence periodicity and did not show any sign of self-assembled mini-fibrils with $d$-periodicity. This also supports our proposed model.
Chapter 5. Discussion

We created four collagen mimetic peptides 2U108, 1U108, Col877 and Col108R to study the determining factors for self-assembly of collagen mimetic mini-fibrils. 2U108 has two repeating sequence units and 1U108 has only one sequence unit. Col877 has three repeating sequence units but different amino acids composition with Col108. Col108R which has the same amino acid composition of that of Col108 lacks repeating sequence architecture. Both 2U108 and Col877 triple helices self-assembles into mini-fibrils with identical $d$ periodicity, which agrees in size of their repeating sequence unit. But no fibrils-like assemblies are found in Col108R or 1U108 since both peptides lack the long-range sequence periodicity. The one-unit staggering model appears to apply for both Col877 and 2U108, which produces in-register alignment of interacting residues of neighboring triple helices as shown in Figure 5-1A and in reference (46), respectively. The white parts represent some large hydrophobic residues (valine, methionine, leucine, isoleucine and phenylalanine); blue and red lines represent the positive and negative charged residues. In one-unit staggered arrangement, these residues form the triple helices will be placed in the close vicinity of comparable residues from the adjacent helices to promote stabilizing interaction. The interactions of hydrophobic residues, and charged residues between neighboring triple helices are maximized in this arrangement. These interactions stabilize the staggered structure. On the other hand, the one-unit staggered pattern of Col108R fibril model is shown in Figure 5-1B. The in-register alignments of the charged and hydrophobic residues of the neighboring triple helices are lost. Therefore, the one-unit staggered arrangement of Col108R does not represent an optimally stabilized conformation and does not lead to the formation of the one-unit staggered fibrils.
Figure 5-1 Predicted assembled triple helices models. (A) predicted Col877 fibril model showing charge distribution. (B) predicted Col108r fibril model showing charge distribution. The three sequence units of the peptide are separated by black vertical bar; Blue vertical bars show positive charge residues; Red vertical bars show negative charge residues; The oval represents foldon domain.

Overall, three collagen mimetic peptides Col108, 2U108 and Col877 self-associate into identical \(d\)-period mini-fibrils. These peptides have similarities in their primary structures. The prominent feature is that all of them have at least two repeating sequence units. 1U108 and Col108R assemble into non-specific aggregates other than \(d\)-periodic mini-fibrils because these two peptides lack tandem repeats of sequence units. Therefore, lacking sequence periodicity in the primary sequence will induce non-specific aggregates.

The foldon domain at C terminal end contributes to stabilizing interactions of 2U108, Col877 and Col108 mini-fibrils. In one-unit staggered arrangement, the residues on the surface of foldon domain may interact with neighboring helices and contribute to the stability of the mini-fibril assembly. But the foldon domain is not a determining factor for self-assembly of mini-fibrils. 1U108 and Col108R that have foldon domain at C terminus only formed non-specific aggregates.
In the contrary, because the size of the foldon domain its presence may limited the growth of the mini-fibrils. Foldon has a beta hairpin conformation. Although smooth mini-fibrils formed indicates that foldon domain can be accommodated in the packing of triple helices, foldon domain has a diameter of 2.5 nm (50), which is a bit larger than that of a triple helix (1.5 nm). The foldon domains inside the mini-fibrils may induce steric tension that impedes the growth of fibril formation.

Another critical factor to have for the formation of the alternating gap-and-overlap structure of the mini-fibrils is the inclusion of the overhang region of the peptides, which consists of the foldon, the C-terminal (Gly-Pro-Pro)$_4$, and the Cys-knot sequences at the N- and the C-termini. The 0.3$d$ overhang contributes to the overlap region of the $d$-period. Yoshizumi et al reported that a collagen-like construct CL-CL which has two collagenous repeating sequence units self-associate into fibrillar structures (32,33). But this fibrillar structures do not have any periodicity. This may be caused by construct CL-CL lacking an overhang unit based on our analyses of our collagen mimetic peptides. In our collagen mimetic peptides, the 0.3$d$ overhang unit consists of a N-terminal cysteine knot, a C-terminal nucleation domain, a C-terminal cysteine knot and a C-terminal foldon domain. If they add an overhang unit into CL-CL, CL-CL molecules may self-associate into fibrils with periodicity.

For the further study, it will be interesting to replace foldon with another trimerization domain. Globular protein V domain is another trimerization domain of bacterial collagen and is at N terminal end (51), which can be a good candidate. This domain can subsequently be removed by protease digestion. Finding the replacement of foldon domain will contribute to generate stable and may be longer collagen mimetic fibrils. Based on our design rule, we can design other collagen mimetic peptides with two repeating sequence units, an overhang unit consisting of partial
collagenous domain and V domain at N terminal end to associate into periodic fibrils. The development of collagen mimetic peptides will lead to a better understanding of the fibrillogenesis of collagen, and the development of new biomaterials for the application in medicine, pharmaceuticals and cosmetics.
APPENDIX

A. The result of Col877 DNA sequencing by GenScript
B. The result of Col877 DNA sequencing by GENEWIZ

<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTCUCACTATTGAAGTTAAATATTTGAACATCGGATCAGTACGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>520</td>
<td>540</td>
<td>560</td>
<td>580</td>
<td>600</td>
<td>620</td>
<td>640</td>
</tr>
<tr>
<td>GTCUCACTATTGAAGTTAAATATTTGAACATCGGATCAGTACGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Col877 4-T7.ab1 (93>608)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 5-T7.ab (92>654)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 7-T7.ab (89>669)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 2-T7.ab (91>719)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 1-T7.ab (90>708)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 6-T7.ab (122>673)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 3-T7.ab (125>694)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 7-T7.ab (89>669)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 5-T7.ab (92>654)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 4-T7.ab1 (93>608)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 1-T7.ab (90>708)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 6-T7.ab (122>673)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 3-T7.ab (125>694)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 2-T7.ab (91>719)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 1-T7.ab (90>708)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 6-T7.ab (122>673)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 3-T7.ab (125>694)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 2-T7.ab (91>719)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 1-T7.ab (90>708)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 6-T7.ab (122>673)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 3-T7.ab (125>694)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 2-T7.ab (91>719)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 1-T7.ab (90>708)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 6-T7.ab (122>673)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 3-T7.ab (125>694)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 2-T7.ab (91>719)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 1-T7.ab (90>708)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 6-T7.ab (122>673)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 3-T7.ab (125>694)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC

**Col877 seg(1>1263)**
- GCCGATTCGCAAGAAGGCCGATTCGCAAGTACGAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATC
The results of seven Col877 plasmids using forward primer of T7 promotor and reverse primer of T7 terminator are showed in green arrow and red arrow, respectively. The region that was not covered is highlighted using a red box. The bases are not identified are labeled as N. Col877.seq(1>1263) is the Col877 DNA sequence provided by GenScript.
C. The result of Col108R DNA sequencing analyzed using Bioedit software
The result of one Col108R plasmid using two designed forward primers are showed as 108R1_vseqF and 108R1_1seq1, respectively. The bases are not identified are labeled as N. 108R is the Col108R DNA sequence provided by GenScript.
D. Protein Expression Evaluation of Col877 in E.coli done by GenScript

Summary

E. coli BL21 (DE3) was transformed with the recombinant plasmid. A single colony was inoculated into LB medium containing ampicilin; cultures were incubated in 37 °C at 200 rpm. Once cell density reached to OD=0.6-0.8 at 600 nm, IPTG was introduced for induction. Western blot was used to monitor the expression (Figure D-1).

Results

![Western blot image](image)

Figure D-1 Col877 expression analyzed by Western blot (using anti-His antibody, GenScript Cat.No. A00186). Lane 1: Cell lysate with induction for 16 h at 15°C. Lane 2: Cell lysate with induction for 4 h at 37°C. Lane 3: Supernatant of cell lysate with induction for 16 h at 15°C. Lane 4: Pellet of cell lysate with induction for 16 h at 15°C. Lane 5: Supernatant of cell lysate with induction for 4 h at 37°C. Lane 6: Pellet of cell lysate with induction for 4 h at 37°C

Conclusion

The Western blot result showed that Col877 was expressed with degradation in E.coli.
BIBLIOGRAPHY


42. Xu, K., Nowak, I., Kirchner, M., and Xu, Y. (2008) Recombinant collagen studies link the severe conformational changes induced by osteogenesis imperfecta mutations to the disruption of a set of interchain salt bridges. *J Biol Chem* **283**, 34337-34344


