RICE UNIVERSITY

Design, Self-Assembly and Applications of Heterotrimeric Collagen Mimics

by

Varun Gauba

A THESIS SUBMITTED
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

Approved, Thesis Committee

Jeffrey D. Hartgerink, Chair
Associate Professor, Chemistry
Rice University

Seiichi P. T. Matsuda
Professor, Chemistry
Rice University

K. Jane Grande-Allen
Associate Professor, Bioengineering
Rice University

HOUSTON, TEXAS

JANUARY 2009
Abstract

Design, Self-Assembly and Applications of Heterotrimeric Collagen Mimics

by

Varun Gauba

Collagen, a fibrous protein, is an essential structural component of all connective tissues, including cartilage, skin, tendon, ligaments and bone. Type I collagen is an AAB heterotrimer assembled from two identical α1 and one α2 chain. Missense mutations in either the α1 or α2 chains of type I collagen, which lead to the substitution of Gly in the ubiquitous X-Y-Gly repeat by bulky amino acids lead to Osteogenesis imperfecta (OI) of varying severity. However, the majority of studies on the effects of amino acid substitutions on triple helix stability have been performed on collagen-like peptides homotrimers. We report the design, synthesis, self-assembly and characterization of a series of peptides that self-assemble to form collagen-like heterotrimerics directed through electrostatic interactions. First, we utilize a series of peptides with net charge ranging from -10 to +10 to show the assembly of various AAB and ABC heterotrimers. We then analyze the ability of various charge pairs based upon naturally occurring amino acids, for instance E – R, E – K, D – R and D – K charge pairs, to stabilize a collagen triple helix. We report the synthesis of a surprisingly stable ABC heterotrimer, composed of (DOG)$_{10}$, (PKG)$_{10}$ and (POG)$_{10}$ chains (O = hydroxyproline), with a stability comparable to (POG)$_{10}$ homotrimer. This high stability heterotrimer is then used to develop a peptide model for OI, a hereditary disorder observed in type I collagen. We report the design of a novel peptide model that can mimic glycine mutations in either of the α1 or α2 chains of
type I collagen. This design utilizes an electrostatic recognition motif in three chains that can force the interaction of any three peptides, including AAA (all same) homotrimers, AAB (two same, one different) heterotrimers and ABC (all different) heterotrimers. The component peptides can be designed in such a way that the mutations are present in none, one, two or all three chains. We successfully report collagen mutants, for the first time, with the structure relevant to the native forms of OI. Furthermore, we are able to differentiate between four triple helices that differ from each other in the frequency of glycine mutations at a particular position. Thus, the ease of preparation of heterotrimers, coupled with our ability to separate single mutations, provides us with a tool to understand mutations in natural collagens that lead to various connective tissue disorders in general and OI in particular. We also introduce another peptide model based upon the ABC heterotrimer to understand the effect of proline hydroxylation and fluorination to the stability of a collagen triple helix, in a chain dependent manner.
Acknowledgments

I would like to thank my advisor Prof. Jeffrey D. Hartgerink for his support, encouragement, patience and guidance throughout my graduate life, and for accepting me as a student in his laboratory. It has been a fun and eventful journey and I sincerely hope our collaboration will continue long into the future.

Thanks a lot to Prof. K. Jane Grande-Allen and Prof. Seiichi P. T. Matsuda for being a part of my committee.

Thanks to my labmates over the last four years for your support, and sharing both my frustrations and happiness. I have enjoyed every moment of it.

Finally, thanks to my family for their love and having faith in me. All this work would have been impossible without their continuous encouragement and morale boosting.
# Table of Contents

Chapter 1  Thesis Overview ................................................................. 1
Chapter 2  Introduction ................................................................. 6
Chapter 3  Design of Collagen-like Heterotrimers .............................. 26
Chapter 4  Optimizing the Thermal Stability of Collagen-like Heterotrimers .... 48
Chapter 5  Collagen Heterotrimers: Peptide Model for Osteogenesis Imperfecta ... 66
Chapter 6  Collagen Heterotrimers: Peptide Model for Proline Hydroxylation and Fluorination ................................................................. 88
Chapter 7  Conclusions and Future Directions ..................................... 98
Chapter 8  Experimental ................................................................. 107
Appendix ...................................................................................... 127
List of Figures

Chapter 2

Figure 2-1 Chemical Structure of Gly-Pro-Hyp Triplet ......................... 6
Figure 2-2 Hydrogen Bonding in Collagen Triple Helix ......................... 7
Figure 2-3 Collagen Hierarchy .................................................. 11
Figure 2-4 Collagen D-Bands ..................................................... 12
Figure 2-5 Thermal Unfolding for (Pro-Flp-Gly)$_{10}$ Homotrimers ............. 17
Figure 2-6 Stabilization by Hydroxyproline .................................... 18
Figure 2-7 Assembly using Bipyridyl Ligands .................................. 20
Figure 2-8 Assembly using TRIS scaffold ..................................... 20
Figure 2-9 Chemical Structure of KTA and TREN .............................. 21
Figure 2-10 Cysteine-knot Strategy ............................................. 22
Figure 2-11 Circular Dichroism for (Pro-Hyp-Gly)$_{10}$ Homotrimers.......... 24

Chapter 3

Figure 3-1 Circular Dichroism for (Pro-Hyp-Gly)$_{10}$ Homotrimers .......... 30
Figure 3-2 Circular Dichroism for E3 and R3 Homotrimers .................... 31
Figure 3-3 Circular Dichroism for E5 and R5 Homotrimers .................... 32
Figure 3-4 Circular Dichroism for E10 and R10 Homotrimers ................. 33
Figure 3-5 Thermal Unfolding for a Mixture of Charged and Neutral Peptides
........................................................................................................... 36
Figure 3-6  Thermal Unfolding for a Mixture of Tricharged and Pentacharged Peptides ................................................................. 38
Figure 3-7  Thermal Unfolding for a Mixture of Decacharged Peptides ........ 38
Figure 3-8  Circular Dichroism for E10/POG and R10/POG Mixtures ....... 39
Figure 3-9  Thermal Studies on ABC Heterotrimer ................................................. 40

Chapter 4

Figure 4-1  Circular Dichroism for D and K Homotrimers ......................... 53
Figure 4-2  Thermal Unfolding for Heterotrimers Formed by Mixing D, K and O ........................................................................... 54
Figure 4-3  Circular Dichroism Analysis at pH7 and pH3 ........................... 55
Figure 4-4  Thermal Unfolding for Heterotrimers Formed by Mixing E, K and O ........................................................................... 56
Figure 4-5  Thermal Unfolding for Heterotrimers Formed by Mixing D, R and O ........................................................................... 59
Figure 4-6  Circular Dichroism for D/O and K/O Mixtures ...................... 61
Figure 4-7  Molecular Model for D•K•O heterotrimer ................................. 62

Chapter 5

Figure 5-1  A Schematic Representation of the Molecular Design ............. 71
Figure 5-2  Circular Dichroism for Homotrimeric Helices ........................ 73
Figure 5-3  Thermal Unfolding for 3/7, 3/1 and 7/1 Peptide Mixtures ....... 77
Figure 5-4  Thermal Unfolding for 3/8, 3/1 and 8/1 Peptide Mixtures ....... 77
Figure 5-5  Thermal Unfolding for 4/7, 4/2 and 7/2 Peptide Mixtures ........ 78
Figure 5-6  Thermal Unfolding for 4/8, 4/2 and 8/2 Peptide Mixtures ........ 78
Figure 5-7  Circular Dichroism for Heterotrimeric Helices .................. 80-81
Figure 5-8  Circular Dichroism for $A^*A^B$ and $A^*A^B$ Heterotrimers ...... 82
Figure 5-9  Refolding Analysis of Homotrimers and Heterotrimers ........... 86

Chapter 6

Figure 6-1  A Schematic Representation of the Molecular Design .......... 92
Figure 6-2  Circular Dichroism for Homotrimeric Helices ................. 94
Figure 6-3  Circular Dichroism for Heterotrimeric Helices ............... 96

Chapter 8

Figure 8-1  HPLC and MS for E3 ............................................. 111
Figure 8-2  HPLC and MS for R3 ............................................. 111
Figure 8-3  HPLC and MS for E5 ............................................. 112
Figure 8-4  HPLC and MS for R5 ............................................. 112
Figure 8-5  HPLC and MS for E10/E ......................................... 113
Figure 8-6  HPLC and MS for R10/R ......................................... 113
Figure 8-7  HPLC and MS for POG/O ....................................... 114
Figure 8-8  HPLC and MS for D ............................................... 114
Figure 8-9  HPLC and MS for K ............................................... 115
Figure 8-10 HPLC and MS for I .............................................. 115
Figure 8-11 HPLC and MS for 2 .............................................. 116
Figure 8-12  HPLC and MS for 3 ................................................. 116
Figure 8-13  HPLC and MS for 4 ................................................. 117
Figure 8-14  HPLC and MS for 5 ................................................. 117
Figure 8-15  HPLC and MS for 6 ................................................. 118
Figure 8-16  HPLC and MS for 7 ................................................. 118
Figure 8-17  HPLC and MS for 8 ................................................. 119

Appendix

Figure A-1  Thermal unfolding for A•B•A Heterotrimers ................. 130
Figure A-2  Thermal unfolding for B•A•A Heterotrimers .................... 131
List of Tables

Chapter 2
Table 2-1 Connective Tissue Disorders in Different Types of Collagen ...... 9

Chapter 3
Table 3-1 Sequence of Peptides ................................................. 28
Table 3-2 Melting Temperature of Homotrimers ......................... 34
Table 3-3 Melting Temperature of Heterotrimers ......................... 41

Chapter 4
Table 4-1 Sequence of Peptides ................................................. 51
Table 4-2 Melting Temperature of Heterotrimers ......................... 58

Chapter 5
Table 5-1 Sequence of Peptides ................................................. 69
Table 5-2 Composition of Triple Helices .................................... 70
Table 5-3 Thermal Stability and Refolding Half-lifes of Triple Helices .... 74

Chapter 6
Table 6-1 Sequence of Peptides ................................................. 91
Table 6-2 Composition of Triple Helices .................................... 91
Table 6-3 Thermal Stability of Triple Helices ................................ 95
## Appendix

<table>
<thead>
<tr>
<th>Table A-1</th>
<th>Sequence of Peptides Studied</th>
<th>127</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table A-2</td>
<td>Sequence of Peptides Studied (from Chapter 5)</td>
<td>128</td>
</tr>
<tr>
<td>Table A-3</td>
<td>Composition of Triple Helices</td>
<td>129</td>
</tr>
<tr>
<td>Table A-4</td>
<td>Thermal Stability of Triple Helices</td>
<td>132</td>
</tr>
</tbody>
</table>
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBTU</td>
<td>O-benzotriazole N,N,N',N'-tetramethyluronium hexafluorophosphate</td>
</tr>
<tr>
<td>HOBr</td>
<td>1-hydroxybenzotriazole hydrate</td>
</tr>
<tr>
<td>DiEA</td>
<td>N,N-diisopropylethylamine</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-dimethylformamide</td>
</tr>
<tr>
<td>DCM</td>
<td>Dichloromethane</td>
</tr>
<tr>
<td>TFA</td>
<td>Trifluoroacetic acid</td>
</tr>
<tr>
<td>HPLC</td>
<td>High pressure liquid chromatography</td>
</tr>
<tr>
<td>Fmoc</td>
<td>9-flourenyloxycarbonyl</td>
</tr>
<tr>
<td>Boc</td>
<td>N-tert-Butyloxycarbonyl</td>
</tr>
<tr>
<td>CD</td>
<td>Circular Dichroism spectroscopy</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
</tbody>
</table>
Chapter 1

Thesis Overview

The number of protein folding motifs that are well understood in terms of sequence-structure relationship is limited. One of the best-understood motifs is the α-helical coiled coil, which can be designed with a very high probability of success as both homomers or heteromers.¹⁻⁶ This understanding has had a broad impact in areas as diverse as protein folding,¹⁻⁶ catalyst design, nanotechnology⁷⁻⁹ and origins of life research.¹⁰⁻¹³ More recently, β-sheet design has also become relatively straightforward allowing greater understanding of the mechanism behind neurodegenerative diseases, nanotechnology and tissue engineering.¹⁴⁻¹⁸ The design of these protein motifs is largely based on hydrophobic packing interactions and encompasses two of the most prevalent protein secondary structures. Collagen, which is the most abundant protein in the human body, has been less intensively studied, particularly with respect to de novo design. The presented work is aimed at investigating the structure-function relationships in collagen-like heterotrimers. The thesis is sub-divided into various chapters, focusing on introduction (Chapter 2), the design (Chapter 3) and optimization (Chapter 4) of heterotrimers, the applications of optimized heterotrimers for mimicking osteogenesis imperfecta, a connective tissue disorder (Chapter 5) and for Pro hydroxylation and fluorination (Chapter 6), followed by conclusions and future work (Chapter 7).

Collagen, a fibrous protein, is an essential structural component of all connective tissues, including cartilage, skin, bones, tendon and ligaments. It is characterized by a unique tertiary structure, known as a collagen triple helix, in which three left-handed
poly-proline type II (PPII) helices wind around one another to form a right-handed super helix.\textsuperscript{19-27} Type I collagen is an AAB heterotrimer assembled from two identical α1 and one α2 chain.\textsuperscript{28,29} Gly mutations in collagen lead to variety of connective tissue disorders.\textsuperscript{28-32} Missense mutations in either of the α1 or α2 chains of type I collagen, which lead to the substitution of Gly in the ubiquitous X-Y-Gly repeat by bulky amino acids such as Arg, Asp, Glu, Cys, Ser, Ala or Val, lead to Osteogenesis imperfecta (OI) of varying severity.\textsuperscript{28-32} Furthermore, most of naturally occurring collagens are heterotrimers, either AAB or ABC, and only a few are homotrimers. However, the majority of studies on the effects of amino acid substitutions on triple helix stability have been performed on collagen-like peptides homotrimers.\textsuperscript{33-51} In a homotrimer, it is impossible to determine whether the contribution to stability is from the PPII helix propensity of amino acids or from interhelix amino acid interactions. Furthermore, the presence of amino acid substitution in three chains exaggerates their contribution. In contrast, in a heterotrimer, the individual chains can be tailored to have the amino acid substitutions in one, two or all three chains and thus can divulge specific information about any interaction based upon these substitutions. In chapter 3, I report the design, synthesis, self-assembly and characterization of a series of peptides that self-assemble to form collagen-like heterotrimers directed through electrostatic interactions. I utilize a series of peptides with net charge ranging from -10 to +10 to show the assembly of various AAB and ABC heterotrimers. I successfully demonstrate that mixing three peptides with individually negative, positive and neutral charge is the mechanism for the formation of specific ABC collagen-like heterotrimers. In chapter 4, with an goal to optimize the stability of a collagen-like heterotrimer, I analyze the ability of various
charge pairs based upon naturally occurring amino acids, for instance E – R, E – K, D – R and D – K charge pairs, to stabilize a collagen triple helix. I report the synthesis of a surprisingly stable ABC heterotrimer, composed of (DOG)$_{10}$, (PKG)$_{10}$ and (POG)$_{10}$ chains (O = 4(R)-hydroxyproline = Hyp). This heterotrimer has a stability comparable to (POG)$_{10}$ homotrimer, even though Asp and Lys amino acids occur twenty times in the heterotrimer and have been shown to significantly destabilize the triple helix as compared to Pro and Hyp imino acids respectively. Thus, I show that the stability of a heterotrimer cannot be anticipated by individual amino acid propensities for triple helix formation as reported for homotrimers. My results indicate that intermolecular electrostatic interactions can be utilized to direct heterotrimer formation. Additionally, amino acids with poor collagen triple helical propensity can be “rescued” in heterotrimers containing amino/imino acids with known high triple helical propensity. Furthermore, this allows for the introduction of a greater chemical diversity in collagen-like triple helices than what would be allowed in a homotrimer. In chapter 5, the high stability heterotrimer, composed of (DOG)$_{10}$, (PKG)$_{10}$ and (POG)$_{10}$ chains, is used to develop a peptide model for OI, a hereditary disorder observed in type I collagen. I report the design of a novel peptide model that can mimic Gly mutations in either of the $\alpha$1 or $\alpha$2 chains of type I collagen. This design utilizes an electrostatic recognition motif in three chains that can force the interaction of any three peptides, including AAA (all same) homotrimers, AAB (two same, one different) heterotrimers and ABC (all different) heterotrimers. The component peptides can be designed in such a way that the mutations are present in none, one, two or all three chains. I successfully report collagen mutants, for the first time, with the structure relevant to the native forms of OI. Furthermore, I am able to differentiate
between four triple helices that differ from each other in the frequency of Gly mutations at a particular position. Thus, the ease of preparation of heterotrimers, coupled with the ability to separate single mutations, provides me with a tool to understand mutations in natural collagens that lead to various connective tissue disorders in general and OI in particular. In chapter 6, I introduce another peptide model based upon the ABC heterotrimer composed of (DOG)$_{10}$, (PKG)$_{10}$ and (P$^*Y^*G$)$_{10}$ chains to understand the effect of Pro hydroxylation and fluorination to the stability of a collagen triple helix, in a chain dependent manner. Pro, Hyp or Flp (Flp = 4(R)-fluoroproline = Fp) replace 'Y' to synthesize a series of heterotrimers with either Pro to Hyp or Pro to Flp substitutions. The results suggest that there is a non-linear increase in thermal stability when Hyp or Flp replaces Pro in a chain-dependent manner.

The research performed over the last few years has resulted in the successful design of peptides to specifically form a high stability ABC heterotrimer in good yields, and elaborated on the applications of these heterotrimers for sequence-structure analysis and connective tissue disorders in native collagens found in human body. The work presented in this thesis has been published in the form of research and review articles in several scientific journals:


(Highlighted in Nature "News & Views" 2008, 453, 998-999)


Chapter 2

Introduction

2.1 Introduction

Collagen is the most abundant protein found in human body and is an essential structural component of all connective tissues such as cartilage, bone, tendon, ligament and skin. Collagen is one of the major components of extracellular matrix (ECM) and provides physical strength to the tissues. It is characterized by a unique tertiary structure, known as a collagen triple helix, in which three left-handed PPII helices wind around each other to form a right-handed super helix.\textsuperscript{19-27} A PPII helix is characterized by 3 residues/turn with a 3.1 Å rise/residue and the corresponding $\phi = -75^\circ$ and $\psi = 145^\circ$. However, when these PPII helices assemble to form a triple helix, these corresponding values change to 3.33 residues/turn with a 2.9 Å rise/residue and similar $\phi$ and $\psi$ angles.\textsuperscript{20,24,26} Thus, the chains pack more tightly when they are in a triple helical conformation as compared to a PPII helical conformation. This close packing of three PPII chains results in a steric constraint on every third amino acid, which is permitted by the presence of Gly, the smallest amino acid, as every third residue in a signature Gly-X-Y repeating motif, as shown in Figure 2-1.

![Chemical Structure](image)

\textbf{Figure 2-1.} The chemical structure of Gly-Pro-Hyp triplet is shown. Hyp is synthesized by the post-translational modification of Pro, and it has a hydroxyl group at $\gamma$ position in the (R) conformation.
The triple helix is further stabilized by an extensive network of CO\(\text{X} \cdot \text{NH}_{\text{Gly}}\) hydrogen bonds\(^{20,24,26}\), as highlighted in Figure 2-2. Additionally, the three chains in a triple helix are staggered by one residue relative to each other to allow for the presence of Gly in the interior of the triple helix.

![Figure 2-2](image)

**Figure 2-2.** The assembly of a collagen triple helix, as observed in the crystal structure of collagen-like peptides (PDB ID: 1CGD). Three left-handed PPII chains coil around each other to form a right-handed triple helix. The C atoms are shown in green, O in red, N in blue and H in white. Dashed lines represent the hydrogen bonds between CO\(\text{X} \cdot \text{NH}_{\text{Gly}}\) residues. These hydrogen bonds are perpendicular to the triple helical axis.

The side chains of the X and Y residues are oriented away from the helix core. Therefore, these side chains are exposed to the solvent\(^{52}\) and can take part in intra and intermolecular side chain interactions.\(^{53-55}\) X and Y positions can accommodate any amino acids, however, imino acids are generally favored because of similar phi and psi angles as to those found in a triple helix and their preferential formation of an extended
conformation. In mammalian collagen, approximately one third of X and Y positions are occupied by Pro and Hyp imino acid residues, respectively. Pro incorporated in the Y position in the collagen chain is post-translationally modified to Hyp by procollagen prolyl 4-hydroxylase in endoplasmic reticulum. The presence of imino acids and the hydroxylation of Pro to Hyp have a stabilizing effect on the thermal stability of triple helix. Hyp is suggested to form water bridges with unused carbonyl groups in the helix and has inductive effects that lead to enhanced stability of the triple helix.

2.2 Biological Significance

There are 27 distinct types of collagen found in humans and are classified into various subfamilies according to their function and supramolecular organization. Naturally occurring collagens form triple helices composed of either all identical (AAA, homotrimer), two different (AAB, heterotrimer) or three different (ABC, heterotrimer) chains. Types I, II, V and XI self-assemble to form fibrils with a D-period of 65-67 nm, as shown in Figure 2-4. Types I and V fibrils are the structural frameworks of bone and types II and XI fibrils are predominantly present in cartilage. Type IV forms basement membrane networks with non-collagenous components like laminin and perlecana while type VIII and X form hexagonal networks. Several other types of collagens are found on the surface of fibrils as fibril-associated collagen with interrupted triple helices and as membrane proteins. Even though there is a high structural and functional diversity between various collagen types, all the types have a common feature, the collagen triple helix. The triple helical domain has binding sites for different molecules including receptors, integrins, proteases, nucleic acids and other components of ECM enabling it to
serve a variety of functions. Collagens are also responsible for a variety of connective tissue disorders in humans and more than 1000 mutations have been documented in various types of collagen. Mutations in homotrimeric AAA collagens (type II, III, VII and X) lead to chondrodysplasia, Ehlers-Danlos syndrome, dystrophic epidermolysis bullosa and chondrocyte hyperthrophy respectively. Mutations in heterotrimeric AAB collagens (type I, IV and VIII) result in osteogenesis imperfecta (OI), Alport syndrome and corneal endothelial dystrophy respectively, while mutations in heterotrimeric ABC collagen (type VI) result in Bethlem myopathy. The diseases caused by various types of collagen are summarized in Table 2-1. Types of collagen that are heterotrimeric in nature are highlighted in red.

<table>
<thead>
<tr>
<th>Type</th>
<th>Trimer</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>AAB</td>
<td>Osteogenesis Imperfecta, Osteoporosis, Ehlers Danlos Syndrome (EDS) type VIIA and VIIIB</td>
</tr>
<tr>
<td>II</td>
<td>AAA</td>
<td>Severe chondrodysplasia</td>
</tr>
<tr>
<td>III</td>
<td>AAA</td>
<td>EDS type IV, Arterial aneurysms</td>
</tr>
<tr>
<td>IV</td>
<td>AAB, ABC</td>
<td>Alport Syndrome</td>
</tr>
<tr>
<td>V</td>
<td>ABC</td>
<td>EDS type I, EDS type II</td>
</tr>
<tr>
<td>VI</td>
<td>ABC</td>
<td>Bethlem Myopathy, Ullrich Congenital Muscular Dystrophy</td>
</tr>
<tr>
<td>VII</td>
<td>AAA</td>
<td>Dystrophic Epidermolysis Bullosa</td>
</tr>
<tr>
<td>VIII</td>
<td>AAB</td>
<td>Corneal Endothelial Dystrophy</td>
</tr>
<tr>
<td>IX</td>
<td>ABC</td>
<td>Multiple epiphyseal dysplasia, Osteoarthritis, Intervertebral disc diseases</td>
</tr>
<tr>
<td>X</td>
<td>AAA</td>
<td>Chondrocyte Hyperthrophy</td>
</tr>
<tr>
<td>XI</td>
<td>ABC</td>
<td>Endochondral Ossification</td>
</tr>
</tbody>
</table>

Table 2-1. Type I to XI collagen, their triple helical nature (homotrimer or heterotrimer) and the diseases caused by Gly mutations are reported. Collagen which are heterotrimeric in nature are highlighted in red.
Collagen type I is the most abundant among all types of collagens and is an AAB heterotrimer formed from two identical α1(I) and one α2(I) chains. Type I collagen forms more than 90% of organic mass of bone and is responsible for biomechanical properties concerning load bearing, tensile strength and torsional stiffness after calcification. Type I collagen has a triple helical domain of around 300 nm in length spanning over 1000 amino acids. OI, the most studied collagen disease, is primarily caused by missense mutations in either the α1 or α2 chains of type I collagen, which lead to the substitution of Gly in the ubiquitous X-Y-Gly repeat by bulky amino acids such as Arg, Asp, Glu, Cys, Ser, Ala or Val. OI phenotypes vary from mild to lethal forms depending on a variety of factors. These include the chain type mutated, the proximity to the C-terminus, the neighboring sequence of the mutated site, and the amino acid residue which substitutes for Gly. It is observed that the amino acid substitutions that are lethal when present in the α1 chain may not be lethal when present in α2 chain. Furthermore, the mutations present near the C-terminus are known to be more lethal than the mutations present near the N-terminus, which correlates with the C to N terminus folding of the triple helix. More than 150 mutations have been reported for type I collagen and they are present along the length of α1 and α2 chains. As the substitution of Gly by a bulkier amino acid leads to disruption of a tightly packed triple helix, it results in defective mineralization on collagen during bone formation. It has been reported that the defects in mineralization may be related to decrease in thermal stability, delayed folding, excess post-translational modifications, and abnormal fibril packing.
2.3 Molecular Structure of Collagen

Collagen has a very complicated structure with various levels of hierarchy starting from a triple helix, which assemble in a particular fashion to form fibrils which then further assemble with each other to form fibers,\textsuperscript{64,65} as shown in the Figure 2-3.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{collagen_hierarchy.png}
\caption{Schematic representation of the collagen hierarchy, ranging from the primary sequence of collagen, their assembly into a triple helix and their subsequent organization to form fibrils and fibers.\textsuperscript{65}}
\end{figure}

The collagen triple helices assemble in a highly oriented quarter-staggered fashion in which 300 nm monomers are separated by a 40 nm gap along the central axis, which results in a characteristic banding patterns known as D-period with a periodicity of 65-67 nm,\textsuperscript{28,29} as shown in Figure 2-4. Furthermore, the collagen triple helix structure is characterized by 3.33 residues per turn, which is denoted by a 10/3 helix symmetry (10 units in 3 turns) and has been observed in collagens from various species.\textsuperscript{28,29}
Figure 2-4. Schematic representation of collagen triple helix assembly in a highly oriented quarter-staggered fashion. 300 nm monomers are separated by a 40 nm gap along the central axis, which results in a characteristic banding patterns with a periodicity of 65-67 nm, as depicted in the image with alternate light and dark regions in the fiber.²⁸,⁶⁴

Studies on native collagen types I and III suggest that the folding initiates with the association of three procollagen chains which are rich in Pro-Hyp-Gly triplets at C terminal.⁶⁶-⁶⁹ These Pro-Hyp-Gly triplets at C terminal initiate a zipper like propagation along the triple helical axis from C to N termini. Collagen contains large amounts of
imino acids Pro and Hyp, which have a higher propensity to form cis amide bonds as compared to trans amide bonds. Thus, before the folding can take place, all the imino acid residues need to be converted to trans configuration. Therefore, this cis to trans conversion acts as a rate-limiting step for collagen folding. This conversion of imino acids from cis to trans configuration is referred to as a “zipper like propagation”.\textsuperscript{66-69} Furthermore, the rate of collagen folding is also dependent on the concentration of the monomeric polypeptide chains. At high concentrations of the monomer, the initial nucleation step occurs fast and the cis-trans isomerization is believed to be the rate-limiting step. However, at lower monomer concentrations, the cis-trans isomerization is the faster step and the nucleation is the rate-limiting step.\textsuperscript{66-69} Collagen-like peptides have been extensively used to understand collagen folding in greater detail and the findings are reported in the next section.

2.4 Tools to Study Collagen Triple Helix: Collagen-like Peptides

Collagen triple helices are involved in a wide-range of biological functions, apart from acting as scaffolding materials. However, the mechanisms of stability of collagen triple helices and their functions are not well understood in terms of sequence-structure relationships. It may be because the current techniques in molecular biology and protein biochemistry are designed for soluble proteins and do not work well with collagen which is insoluble and has a very high molecular weight. Collagen-like peptides that mimic the important structural features of a collagen triple helix could be useful models for understanding the underlying principles of stability and biological activity of collagen. In the remainder of this chapter, I will elaborate the current approaches including
homotrimeric\textsuperscript{33,34,36,38-40,43-47,49,51,70-77} and heterotrimeric peptide designs\textsuperscript{41,78-90} reported in the literature, and various experimental tools to characterize these assemblies.

The repeating unit in a collagen-like peptide is a Gly-X-Y motif. As the presence of Gly is crucial for the formation of a triple helix, it is invariant. Thus, only the amino/imino acid residues at X and Y position can be varied to understand the sequence-structure relationship in collagen like peptides. The initial work in this field, approximately four decades ago, relied solely on the incorporation of imino acids in the X and Y positions and (Pro-Pro-Gly)\textsubscript{n} and (Pro-Hyp-Gly)\textsubscript{n} peptides were synthesized to understand the triple helix formation.\textsuperscript{72-77} These initial studies revealed the importance on imino acids in stabilizing the triple helical structure. Recently, Brodsky and coworkers have used a systematic host-guest peptide approach to measure the propensity scale for all possible 20 amino acids in X or Y positions. (Gly-Pro-Hyp)\textsubscript{m}-(Gly-X-Y)\textsubscript{n}-(Gly-Pro-Hyp)\textsubscript{o} peptides were designed to provide information about the side chain interaction between the triple helices.\textsuperscript{43-46} In this design, one or two tripeptide units with various combinations of X and Y residues were incorporated in a stable Gly-Pro-Hyp framework and the thermal stabilities of the host-guest peptides were measured. Pro, Glu, Ala, Lys, Arg, Gln, and Asp were shown to be the most stabilizing for the X position in a Gly-X-Hyp triplet, while Hyp, Arg, Met, Ile, Gln, and Ala were observed to be most stabilizing for the Y position in Gly-Pro-Y triplet. The lease stabilizing residues for both X and Y positions were Gly or aromatic amino acids. This data provides the amino acid propensity for any given residue in a particular position. However, the propensity at any particular position may be affected by the neighboring sequence and the subsequent intramolecular interactions. A different set of host-guest peptides with Gly-X1-Y1-Gly-X2-Y2 guest in a
Gly-Pro-Hyp host framework were used to measure the propensity of groups of amino acids. The authors reported that both favorable ion pair and hydrophobic interactions are possible. High stability is observed for Gly-Arg-Asp, Gly-Lys-Asp, Lys-Gly-Glu and Lys-Gly-Asp sequences. The results from this study laid the groundwork for my research work, where I utilize ion pair interactions emanating from charged residues in different chains to direct the formation of a collagen-like heterotrimer. This groundbreaking work from Brodsky group allowed for the quantification of intrinsic propensities of various amino acids both present individually or in combination and are currently used as a basis for predicting stability in collagen-like homotrimers.

As previously mentioned, Gly mutations in collagen lead to a variety of connective tissue disorders both in homotrimeric and heterotrimeric collagens. Gly is the only amino acid with a hydrogen atom as side chain and does not face any steric hindrance in the center of a tight triple helix. Gly mutation by any other amino acid leads to the disruption of a closely packed triple helical structure and lowers the thermal stability of a collagen molecule. X-ray and thermal unfolding studies on the model collagen peptides containing a Gly to Ala substitution capped by Pro-Hyp-Gly triplets have been reported. This peptide forms homotrimers when the Gly to Ala mutations are absent. When the Gly to Ala mutation are present in all three chains, this peptide can still form a triple helix, however the thermal stability is 30 °C lower than the one with no mutations. X-ray studies reveal a disruption in the triple helix, which allowed additional water molecules to enter inside the helix and form water bridges. Brodsky and coworkers have also studied the relative thermal stability of various Gly mutations in a similar model peptide. The thermal stability decreased to varying extents depending on identity
of the amino acid residue replacing Gly. The residues, in the order of least to most disruptive were reported and the follow the following trend with Trp being the most destabilizing and Ser being the least destabilizing: Trp < Asp < Glu < Val < Arg < Cys < Ala < Ser.\(^9\) This order correlated with the known phenotype of OI and was further demonstrated by the study of two different Gly mutations at the same position in native collagen.

Studies to understand collagen folding have also been performed using collagen-like peptides.\(^8\) Brodsky and coworkers reported peptides which mimic of \(\alpha1\) chain of type I collagen at the N terminal and have a (Pro-Hyp-Gly)\(_4\) motif at the C terminal. The studies on these peptides revealed that the collagen folding is a third-order process. Furthermore, approximately 50% of the monomers are required to be in trans configuration to initiate the folding process.\(^9\) Additional studies on folding of collagen-like peptides in a broad range of concentrations revealed that the third-order kinetics changed to the first-order kinetics upon the increase in concentration of monomers.\(^9\) This suggested that the rate-limiting step at higher concentration is the cis-trans isomerization, and the rate-limiting step at lower concentration is the nucleation step.

Non-natural amino acids have also been incorporated in the X and Y positions of collagen-like peptides and the most significant results were reported by Raines group with the synthesis and self-assembly of a hyperstable (Pro-Flp-Gly)\(_{10}\) triple helix,\(^6\) with thermal stability higher than the (Pro-Hyp-Gly)\(_{10}\) and (Pro-Pro-Gly)\(_{10}\) homotrimers, as shown in Figure 2-5.
Figure 2-5. Thermal unfolding curves and first order derivatives for (Pro-Pro-Gly)$_{10}$ (blue), (Pro-Hyp-Gly)$_{10}$ (red) and (Pro-Flp-Gly)$_{10}$ homotrimerers, as observed by circular dichroism spectroscopy. The presence of Flp at Y position imparted hyperstability to the collagen triple helix.  

The mechanism of stabilization by Hyp is believed to be through water-mediated hydrogen bonding,$^{27,56-60}$ as confirmed by the crystal structures$^{26}$ of collagen like peptides where hydroxyl group in the Hyp acts as an anchoring point for inter- and intramolecular water bridges. Flp has a fluorine group instead of a hydroxyl group, which is a stronger electron-withdrawing group and has a lower tendency to form hydrogen bonds. Yet, the incorporation of Flp instead of Hyp results in an increase in thermal stability.$^{61}$ Therefore, it is believed that the inductive effect of fluorine in Flp at Y position and its subsequent effect on the exo ring pucker plays a role in the stabilization of triple helix. The crystals structures of Flp peptides have not yet been reported. However, in the crystal structures of other collagen-like peptides,$^{26}$ there is a clear preference of endo puckering for the imino acids in X position and exo puckering of imino acids in Y position. The exo pucker for Hyp is shown in Figure 2-6. Imino acid residues with 4(R)-substitutions (Hyp and Flp) stabilize the exo pucker while the imino acid residues with 4(S)-substitutions
stabilize the endo pucker. This observation is also supported by the fact that 4(S)-fluoroproline, which adapts an endo pucker, has a stabilizing effect on triple helix at X position and not in the Y position. As both Hyp and Flp adopt the exo pucker of the ring, their presence at the X position in a collagen-like peptide leads to a lower thermal stability and melting temperature of the triple helix. Furthermore, in naturally occurring collagens, only Pro residues at Y positions are post-translationally modified to Hyp residues. Thus, it is not clear whether Flp and Hyp stabilize the triple helix by same mechanism. It is likely that the inductive effects play a major role in the stabilization of triple helix in the case of Flp, and the stabilization for Hyp is a combination of inductive and water-bridge mediated hydrogen bonding, as depicted in Figure 2-6.

![Figure 2-6](image)

**Figure 2-6.** Proposed mechanisms for the contribution of Hyp to the thermal stability of a collagen triple helix. It can either stabilize the helix by orienting the hydroxyl group in exo pucker, favorable for the Y position or by enthalpic contribution from hydrogen bonding to water.
Several high-resolution crystal structures for collagen-like peptides forming a triple helix have been solved and are of particular interest.\textsuperscript{53-55,104-106} The most prominent of these are the crystal structures of host-guest peptide T3-785 with the sequence (Pro-Hyp-Gly)\textsubscript{3}-Ile-The-Gly-Ala-Arg-Gly-Leu-Ala-Gly-(Pro-Hyp-Gly)\textsubscript{4}\textsuperscript{53,55} and IBP peptide with (Pro-Hyp-Gly) repeats in the flanking regions and a central Gly-Phe-Hyp-Gly-Glu-Arg sequence.\textsuperscript{104,105} As mentioned before, the natural collagen triple helix has a 10/3 helix symmetry. However, most of the crystal structures reported for imino-rich collagen-mimic peptides showed a 7/2 helix symmetry.\textsuperscript{54,106} T3-785 and IBP were the first peptides to have both 7/2 and 10/3 symmetry as observed in the crystal structure.\textsuperscript{53,55,104,105} The central imino-poor region forms a 10/3 helix, similar to what is observed in nature and the imino-rich host region forms a 7/2 helix. These results suggest that collagen molecule may have non-uniform helical twist along its length and the amino acid rich regions may form a 10/3 helix while the imino acid rich regions may form a 7/2 helix. Host-guest peptides with varying guest sequences have been used in binding assays to establish the specificity of collagen-binding integrins and Gly-Glu-Arg (GER) has been identified as binding motif for integrins.\textsuperscript{107} Other studies have used host-guest peptides to find binding motifs for Hsp47, a collagen molecular chaperone.\textsuperscript{108}

It is widely believed that imino acid poor sequences greatly lower the stability of a triple helix\textsuperscript{44} and their incorporation in a collagen-mimic peptide require external stabilizers. The use of artificial covalent bridges, metal co-ordination, built-in sequences of known higher triple helical stability and preorganization by templated self-assembly have been reported. Koide and coworkers reported peptides with a N-terminal 2,2'-
bipyridyl (bpy) ligands that assembled into very stable triple helices in the presence of Fe(II) ion, which can form a Fe$^{II}$(bpy)$_3$ complex, as reported in Figure 2-7.\textsuperscript{109}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure2-7}
\caption{Schematic description of the assembly of collagen-like peptides using bipyridyl ligands on N termini in the presence of Fe(II) ions.\textsuperscript{109}}
\end{figure}

Similarly, Goodman and coworkers reported peptides with C-terminal catechol groups and N-terminal TRIS scaffold to form stable complexes and hence stable triple helices in the presence of Fe(III) ion. This approach led to the assembly of a triple helix tethered at both ends and results in the increase of thermal stability by 22 °C, as shown in Figure 2-8.\textsuperscript{36}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure2-8}
\caption{Schematic description of the assembly of collagen-like peptides using TRIS scaffold on N terminal and catechol groups on the C terminal, which assemble in the presence of Fe(III) ions.\textsuperscript{36}}
\end{figure}
Another approach relies on the use of preorganized scaffolds with three anchor groups where collagen-like peptides are attached. This model reduces the entropic cost of bringing the three chains together for the formation of a triple helix. Several scaffolds including tris(2-aminoethyl)amine (TREN) and Kemp triacid (KTA) have been reported, the structures of which are shown in Figure 2-9. These scaffolds function in the same way as shown for TRIS scaffolds in Figure 2-8. These scaffolds can induce triple helix formation even in very short peptides, which would otherwise not form a stable triple helix.

![Chemical structures of KTA and TREN scaffolds](image)

**Figure 2-9.** Chemical structures of KTA and TREN scaffolds.

Collagen homotrimers have been extensively used to study amino acid substitutions, their PPII helix propensity, their effect on stabilization by water-mediated hydrogen bonding or steric and stereoelectronic effects and from the inter-helix amino acid interactions. However most natural collagens, including the most abundant type I collagen, are heterotrimers, not homotrimers, and diseases caused by mutations in these sequences affect only one or two of the three chains which make up the helix. A number of studies have prepared collagen heterotrimers by cysteine-knot strategy. In this strategy, the component chains
of the triple helix are chemoselectively put together by stepwise disulfide bond formation, as shown in Figure 2-10.

![Diagram of collagen heterotrimer synthesis](image)

**Figure 2-10.** The synthesis of collagen heterotrimers by cysteine-knot strategy proceeds in 4 steps. Step 1 is the activation of the thiol group followed by heterodimerization in step 2. Step 3 is the selective deprotection of the second thiol group and its subsequent activation. The third peptide strand adds to the heterodimer in step 4 to form the collagen-like heterotrimer.¹¹³

This strategy has been used to study the importance of chain stagger for the interaction of integrin α1β1 with type IV collagen.³² Functional moieties from type I collagen have also been assembled in their native form using this methodology.⁷⁹,⁸⁰,⁸³,⁸⁹ This is a very effective system for studying heterotrimers, however, the synthetic process is cumbersome. Only three studies have been reported in literature about the non-covalent formation of collagen-like heterotrimers.⁸⁶,⁸⁷,⁹⁰ Two of these studies reported that mixing
(Pro-Pro-Gly)$_{10}$ with (Pro-Hyp-Gly)$_{10}$ forms heterotrimers. In another study, the authors mixed (Pro-Pro-Gly)$_7$ with (4(S)-fluoroproline-Flp-Gly)$_7$ to form a heterotrimer. However, in all the three studies reported, the heterotrimers formed are not stable over time and refold back to a mixture of both homotrimers and heterotrimers. Because of the importance of heterotrimers in natural collagens, it is important to design systems that have the potential to study these variations in their native form, whether they are AAB heterotrimers or ABC heterotrimers.

2.5 Experimental Characterization of Collagen-like Peptides

Experimental techniques that allow for the characterization of a collagen triple helix include circular dichroism (CD) spectroscopy, differential scanning calorimetry (DSC) and nuclear magnetic resonance (NMR). Other methods including X-ray crystallography are not very common as it is very tough to crystallize collagen-like peptides. The structure determination has been reported only for around a dozen peptides. Recently, Brodsky and coworkers reported the utility of FT-IR in analyzing collagen triple helices. CD spectroscopy is the most common technique used to study collagen-mimetic peptides. Typical CD spectra for a collagen triple helix is characterized by a negative peak around 200 nm and a positive peak in range of 220-227 nm. PPII helices also show a similar CD spectra with a negative peak around 205 nm and a positive peak around 228 nm. However, these two species can be differentiated by thermal unfolding studies. In this study, the peptide of interest is heated to a high temperature at a slow heating rate, around 10 °C/hour, and the CD signal at a wavelength where the maximal signal is observed is collected. For PPII helix, a constant decrease in the CD signal is
observed. However, the unfolding of a triple helix is co-operative in nature that results in a sharp transition in the CD signal in a narrow range from a triple helix to a monomer form. The first order derivative of this sharp transition is known as the melting temperature of a triple helix. In Figure 2-11, the representative data is shown for a (Pro-Hyp-Gly)$_{10}$ homotrimer, with a melting temperature of 67.5 °C.\textsuperscript{114}

\textbf{Figure 2-11.} Circular dichroism for homotrimeric triple helices of (Pro-Hyp-Gly)$_{10}$. The spectra goes from (a) a good PPII helix at 10 °C characterized by a maxima near 225 nm and a minima near 208 nm to (b) disordered structure at 90 °C with only single minima near 208 nm. (c) Thermal unfolding curve. (d) The first derivative of mean residual ellipticity is used to indicate a melting temperature, under the described conditions, of 67.5 °C for (Pro-Hyp-Gly)$_{10}$ homotrimers.\textsuperscript{114}
In the subsequent chapters, I report the design, synthesis, self-assembly, optimization and applications of stable collagen-like heterotrimers assembled by non-covalent electrostatic interactions.
Chapter 3

Design of Collagen-like Heterotrimers

3.1 Introduction

Collagen, a fibrous protein, is an essential structural component of all connective tissues, including cartilage, skin, bones, tendon and ligaments. It has a unique triple helical structure in which three left-handed PPII helices wind around one another to form a right-handed super helix.\textsuperscript{1,19-26} The stability of the triple helix is a result of an extensive network of CO\textsubscript{(X)}-NH\textsubscript{(Gly)} hydrogen bonds and the tightly packed nature of the triple helix permitted by the presence of Gly every third residue in an X-Y-Gly repeating motif. Gly is completely buried in the interior of the triple helix while the side chains of the X and Y residues are oriented away from the helix core where they are exposed to the solvent\textsuperscript{52} and can take part in intra and intermolecular side chain interactions.\textsuperscript{53-55} Hyp also plays an important role in triple helix stability as it is suggested to form water bridges with unused carbonyl groups in the helix\textsuperscript{26,27,56-59} and has inductive effects, which lead to enhanced stability of the helix.\textsuperscript{61}

Naturally occurring collagens form triple helices composed of either all identical (AAA, homotrimer), two different (AAB, heterotrimer) or three different (ABC, heterotrimer) polypeptides.\textsuperscript{28,29} For example, collagen type I is an AAB heterotrimer formed from two identical $\alpha$1(I) and one $\alpha$2(I) chains.\textsuperscript{28,29} However, the majority of studies on synthetic collagen-like peptides have been performed on self-assembling homotrimers.\textsuperscript{33-51} Amongst these, the amino acid propensity for triple-helix formation has been studied exhaustively.\textsuperscript{43-46} These studies have shown that the repeating unit (POG)$_n$
yields the most stable triple helix. They have also shown that Glu and Arg are the most common charged amino acids and are most stable in positions X and Y respectively. Recently, some results have become available for heterotrimers obtained by cystine-knot strategy, where cystine disulfide bonds covalently capture different polypeptides leading to the formation of heterotrimers.\textsuperscript{41,78-85,88,89} Only three studies have examined heterotrimer formation without covalently tethering the peptides together.\textsuperscript{86,87,90} Furthermore, there are no studies which demonstrate the designed formation of non-covalent ABC triple helices. Nonetheless, these types of heterotrimeric structures are of critical importance for understanding the extracellular matrix as many forms of natural collagen are of the type AAB or ABC.\textsuperscript{28,29}

In this chapter, I describe the design, synthesis and characterization of a series of collagen-like peptides which utilize electrostatically charged amino acids to bias their self-assembly into collagen-like heterotrimers. In addition to five AAB heterotrimers, for the first time I demonstrate the formation of an ABC heterotrimer utilizing only supramolecular interactions. I examine the melting temperature for each of these systems and find that the stability of a heterotrimer can not be anticipated by individual amino acid propensities for triple helix formation. Instead I observed that amino acids which are found to substantially destabilize the triple helix, such as Glu, can form triple helices of high stability when paired against oppositely charged amino acids such as Arg or, more surprisingly, when paired against other neutral amino acids with high stability such as Hyp. This results in the stability of the heterotrimeric triple helix to be almost indistinguishable from those containing only amino acids with high triple helix stability. In the most dramatic example I show that (EOG)$_{10}$ and (PRG)$_{10}$, which are unable to form
a triple helix in isolation, form high quality triple helices when mixed together. This work lays the groundwork for a better understanding of the thermal stability of heterotrimeric collagens, for the synthesis of a variety of synthetic extracellular matrix mimics\textsuperscript{49,115-120} and molecular recognition devices based on the interactions reported herein.

3.2 Molecular and Experimental Design

I synthesized a set of (POG)\textsubscript{10} based peptides, with a general sequence (POG)\textsubscript{n}, (EOGPOG)\textsubscript{n}(POG)\textsubscript{5-n} and (POG)\textsubscript{5-n}(PRGPOG)\textsubscript{n}(POG)\textsubscript{5-n}, where n represents the charge of each polypeptide chain, for n = 3, 5; and (EOG)\textsubscript{10} and (PRG)\textsubscript{10} for decacharged peptides. The polypeptides are abbreviated as POG, En or Rn throughout the remainder of this chapter, as shown in Table 3-1.

<table>
<thead>
<tr>
<th>#</th>
<th>Sequence</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(POG)\textsubscript{10}</td>
<td>POG</td>
</tr>
<tr>
<td>2</td>
<td>(POG)\textsubscript{2}(EOGPOG)\textsubscript{3}(POG)\textsubscript{2}</td>
<td>E3</td>
</tr>
<tr>
<td>3</td>
<td>(POG)\textsubscript{2}(PRGPOG)\textsubscript{3}(POG)\textsubscript{2}</td>
<td>R3</td>
</tr>
<tr>
<td>4</td>
<td>(EOGPOG)\textsubscript{5}</td>
<td>E5</td>
</tr>
<tr>
<td>5</td>
<td>(PRGPOG)\textsubscript{5}</td>
<td>R5</td>
</tr>
<tr>
<td>6</td>
<td>(EOG)\textsubscript{10}</td>
<td>E10</td>
</tr>
<tr>
<td>7</td>
<td>(PRG)\textsubscript{10}</td>
<td>R10</td>
</tr>
</tbody>
</table>

The N and C terminal of all peptides are acetylated and amidated respectively. O = hydroxyproline.

Glu, with a negative charge at neutral pH, replaces Pro at X positions. In natural collagen, Glu is found predominantly in the X position and is by far the most abundant of the two possible negatively charged amino acids.\textsuperscript{44} Arg, with a positive charge at neutral
pH, replaces Hyp at Y positions. Arg is found predominantly in the Y position in natural collagen and forms the most stable helix after Hyp.\textsuperscript{44} These two types of substitutions allow us to prepare oppositely charged peptides. Previously, it has been reported that ion pairs can be formed in homotrimeric helices between peptide strands with Glu in the X position of one polypeptide and Arg in the Y position of the adjacent polypeptide.\textsuperscript{70,121,122}

In my system, I expect the presence of opposite charges on the two polypeptides to favor heterotrimeric assembly over homotrimeric assembly. All the peptides used in the study are N-terminal acetylated and C-terminal amidated to avoid charge interactions between the termini.\textsuperscript{48}

### 3.3 Results

3.3.1 Homotrimers. POG, En and Rn homotrimers, at a concentration of 0.2 mM, were characterized by CD thermal unfolding studies at neutral pH using 10 mM phosphate buffer, as shown in Figures 3-1 to 3-4.
Figure 3-1. Circular dichroism for homotrimeric triple helices of POG. The spectra goes from (a) a good PPII helix at 10 °C characterized by a maxima near 225 nm and a minima near 208 nm to (b) disordered structure at 90 °C with only single minima near 208 nm. (c) Thermal unfolding curve. (d) The first derivative of MRE is used to indicate a melting temperature, under the described conditions, of 67.5 °C for POG homotrimers.
Figure 3-2. Circular dichroism for homotrimeric triple helices of E3 and R3. The spectra for both (a) E3 and (b) R3 goes from a good PPII helix at 10 °C characterized by a maxima near 224 nm and a minima near 199 nm and 205 nm for E3 and R3 respectively, to disordered structure at 90 °C with only single minima near 202 nm and 207 nm for E3 and R3 respectively. (c) Thermal unfolding curve. (d) The first derivative of MRE is used to indicate a melting temperature, under the described conditions, of 51.5 °C for E3 and 58.5 °C for R3 respectively.
Figure 3-3. Circular dichroism for homotrimeric triple helices of E5 and R5. The spectra for both (a) E5 and (b) R5 go from a good PPII helix at low temperature characterized by maxima near 224 nm and minima near 202 nm to a disordered structure at high temperature with only a single minima near 204 nm. (c) Thermal unfolding shows that the characteristic peak is eliminated in a cooperative transition which indicates the presence of a triple helix. (d) The first derivative of MRE is used to indicate a melting temperature, under the described conditions, of 46.5 °C and 55.5 °C for E5 and R5 respectively.
Figure 3-4. Circular dichroism for homotrimeric triple helices of E10 and R10. The spectra for (a) E10 goes from a PPII helix at 10 °C characterized by a downshifted maxima near 218 nm and a minima near 196 nm to disordered structure at 90 °C with only single minima near 200 nm. The spectra for (b) R10 show that a disordered structure is observed at both 10 °C and 90 °C with a minima near 199 nm and 203 nm respectively. Thermal unfolding for (c) E10 and (d) R10 show a linear decrease in ellipticity indicating that these peptides do not form triple helices.

As expected, POG forms the most stable triple helix, with a $T_m = 67.5$ °C which is consistent with the values reported for similar peptides.$^{43,87}$ The triple helix stability decreases as the number of substitutions at the X and Y positions with Glu and Arg,
respectively, increase. The melting temperature decreases to 58.5 °C and 55.5 °C when Arg residues replace three and five Hyp residues in POG respectively. The corresponding values when three and five Glu residues replace Pro residues in POG are 51.5 °C and 46.5 °C respectively, as shown in Figures 3-2 and 3-3. In the extreme example of E10 and R10, no triple helix formation was observed by CD studies, as shown in Figure 3-3 and 3-4. CD spectra for R10 does not show any positive maximum but shows a negative peak around 190 nm which resembles a PPII spectra. E10 spectra shows a blue shifted positive peak at 218 nm, instead of the regular 225 nm and the unfolding studies indicate a linear decrease in ellipticity at 218 nm, instead of the co-operative unfolding as observed for a normal triple helix. In summary, the CD spectra and the unfolding curve data for E10 and R10 show that these peptides do not form triple helices under the timeframe of the experiments but instead form weak PPII helices or disordered structures. The melting temperatures for all the homotrimers are summarized in Table 3-2.

Table 3-2. Homotrimeric melting temperatures.

<table>
<thead>
<tr>
<th>Homotrimer</th>
<th>Melting Temperature (Tm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>POG</td>
<td>67.5 °C</td>
</tr>
<tr>
<td>E3</td>
<td>51.5 °C</td>
</tr>
<tr>
<td>R3</td>
<td>58.5 °C</td>
</tr>
<tr>
<td>E5</td>
<td>46.5 °C</td>
</tr>
<tr>
<td>R5</td>
<td>55.5 °C</td>
</tr>
<tr>
<td>E10</td>
<td>no triple helix</td>
</tr>
<tr>
<td>R10</td>
<td>no triple helix</td>
</tr>
</tbody>
</table>
3.3.2 AAB Heterotrimers. 1:1 mixtures of E5 with POG and R5 with POG were prepared so that the final peptide concentration of the solution was 0.2 mM, and a neutral pH was maintained using 10 mM phosphate buffer. Thermal unfolding studies were performed with and without preheating. For preheating studies, peptides were mixed in desired ratios, heated to 85 °C and incubated for 15 min. The peptide solution was then slowly cooled to 25 °C at a rate of 1 °C/min and then incubated overnight at room temperature before performing the unfolding studies. These steps ensure that only the thermodynamically stable species are formed from the mixture of peptides upon preheating. For non-preheating studies, peptides were mixed in the desired ratios and the unfolding studies were performed immediately. The non-preheating step acts as a control to the preheating step. Without preheating, two separate transitions were observed in the unfolding curve, corresponding to the unfolding of E5 and POG homotrimers with a T_m of 46.5 °C and 67.5 °C respectively (Figure 3-5(a) – 3-5(c)). The two homotrimers are well resolved. After preheating, however, only one transition is observed in the unfolding curve, corresponding to a T_m of 64 °C, which is slightly lower than the T_m for POG homotrimer while significantly higher than the T_m for E5 homotrimer. Importantly, the maximum MRE of these samples are similar to one another indicating that the majority of the sample folds into a triple helix as opposed to one component (for example POG alone) folding into a helix while the other remains unfolded. A similar behavior is observed in the case of the R5 / POG mixture, with the corresponding T_m of 56 °C and 67 °C for non-preheating and 65 °C after preheating (Figure 3-5(d) – 3-5(f)). This shows that heterotrimeric species are observed when samples are preheated and equilibrated as
compared to the non-preheating case where well resolved individual homotrimers are observed.

Figure 3-5. Circular dichroism of charged peptides mixed with neutral peptide. (a) Thermal unfolding shows cooperative transition for POG, E5 and their mixture. (b) Mean residue ellipticity (MRE) versus temperature for POG, E5 and their mixture. (c) The first derivative of unfolding clearly show the individual homotrimeric species remain when preheating is not applied while these merge to a single heterotrimeric species after heating. (d) Thermal unfolding shows cooperative transition for POG, R5 and their mixture. (e) MRE versus temperature for POG, R5 and their mixture. (f) The first derivative of unfolding clearly show the individual homotrimeric species remain when preheating is not applied while these merge to a single heterotrimeric species after heating.

In another set of experiments, En and Rn peptides were mixed to analyze the effect of electrostatics on heterotrimer formation. E3 was mixed with R3 in 2:1, 1:1 and 1:2 ratios, followed by preheating and unfolding studies. T_m values of 56 °C, 56.5 °C and 57 °C were
observed for 2:1, 1:1 and 1:2 (E3:R3) respectively (Figure 3-6(a) – 3-6(c)), as compared to 51.5 °C and 58.5 °C for E3 and R3 homotrimers respectively. The E5/R5 series behaved in a similar fashion showing that preheating leads to the formation of a heterotrimeric species by allowing the system to equilibrate. \(T_m\) values observed were 54 °C, 54 °C and 55.5 °C for 2:1, 1:1 and 1:2 (E5:R5) respectively (Figure 3-6(d) – 3-6(f)), with values of 46.5 °C and 55.5 °C for E5 and R5 homotrimers respectively. The \(T_m\) for heterotrimers is observed to cluster near the \(T_m\) of the more stable Rn homotrimers, while maximum MRE is observed to be approximately the same in all cases. Finally, heterotrimers of E10 / R10 were prepared. Although neither E10 nor R10 form homomeric triple helices on their own, when they are mixed together and preheated they form a triple helix with a \(T_m\) of 41 °C (Figure 3-7(a) – 3-7(b)). This triple helix is a heterotrimer composed of both positively charged R10 and negatively charged E10 polypeptides which interact favorably by electrostatic interactions. When the pH of the E10 / R10 mixture is reduced to 3 the triple helical conformation is observed to be destroyed as indicated by the reduction of the maximum at 225 nm (Figure 3-7(c)).
Figure 3-6. Thermal denaturation studies on multiply charged heterotrimeric helices showing the fraction folded, MRE and first derivative plots versus temperature. (a) Fraction folded of E3 with R3. (b) MRE of E3 with R3. (c) The first derivative of unfolding for E3 and R3 mixture. (d) Fraction folded of E5 with R5. (e) MRE of E5 with R5. (f) The first derivative of unfolding for E5 and R5 mixture.

Figure 3-7. CD analysis of E10 / R10 mixtures. a) MRE versus temperature b) The first derivative of MRE versus temperature c) MRE versus wavelength at pH 7 and pH 3.
Next, unfolding studies were performed on E10/POG mix and R10/POG mix in 1:1 ratio. The unfolding curve in both the cases showed a single transition with a corresponding $T_m$ of 68.5 °C and 68 °C for E10/POG mix and R10/POG mix respectively (Figure 3-8). POG homotrimers show a nearly identical $T_m$ value of 67.5 °C. Importantly, MRE for these mixtures drops to approximately 1/2 of that observed for POG alone. Furthermore, there was essentially no difference between the CD spectra of E10/POG and R10/POG before or after preheating. Together with the $T_m$ value being approximately equal to POG alone indicates that no heterotrimeric species are formed in either the E10/POG or R10/POG cases.

![Figure 3-8. CD analysis of mixtures E10 / POG and R10 / POG. a) MRE vs. temperature b) first derivative of MRE vs. temperature c) MRE vs wavelength.](image)

3.3.3 ABC Heterotrimers: When E10, R10 and POG are mixed in 1:1:1 ratio a $T_m$ of 54 °C is observed after preheating (Figure 3-9). This melting temperature corresponds to an ABC triple helix composed of all three polypeptides. Again maximal MRE in the case of E10/R10/POG indicates that the majority of the sample has folded rather than one homotrimer (for example POG) leaving the other peptides unfolded. The E10 / R10 / POG mixture was further analyzed by comparing the non-preheated and preheated
mixtures. Without preheating, three transitions were observed in the unfolding curve, corresponding to POG homotrimer (68 °C), E10/R10 heterotrimer (38.5 °C) and an improperly folded E10/R10 heterotrimer (19 °C, see discussion below). On preheating, all of these transitions merged to a single \( T_m \) of 54 °C corresponding to an ABC type E10/R10/POG heterotrimer.

![Figure 3-9. Thermal denaturation studies on preheated and non-preheated E10 / R10 / POG mixtures at pH 7. (a) Fraction folded versus temperature. (b) MRE versus temperature. (c) The first derivative of MRE vs. temperature.](image)

The melting temperatures for all the heterotrimers reported in this chapter are summarized in Table 3-3.

### 3.4 Discussion

The thermal stability of homotrimers, shown in Table 3-2, compares favorably with the previous literature data.\(^{43,87}\) POG forms the most stable helix with a melting temperature of 67.5 °C; substitution with three Arg residues reduce the melting point to 58.5 °C while five Arg residues reduce the melting point to 55.5 °C. The use of ten Arg
residues destroys the ability of the peptide to form a triple helix within the time frames examined.

<table>
<thead>
<tr>
<th>Peptide/Peptide Mix</th>
<th>Melting Temperature (Tₘ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E3:R3 = 2:1</td>
<td>56 °C</td>
</tr>
<tr>
<td>E3:R3 = 1:1</td>
<td>56.5 °C</td>
</tr>
<tr>
<td>E3:R3 = 1:2</td>
<td>57 °C</td>
</tr>
<tr>
<td>E5:POG = 1:1*</td>
<td>46.5 °C and 67.5 °C</td>
</tr>
<tr>
<td>E5:POG = = 1:1</td>
<td>64 °C</td>
</tr>
<tr>
<td>R5:POG = 1:1*</td>
<td>55 °C and 67 °C</td>
</tr>
<tr>
<td>R5:POG = 1:1</td>
<td>65 °C</td>
</tr>
<tr>
<td>E5:R5 = 2:1</td>
<td>54 °C</td>
</tr>
<tr>
<td>E5:R5 = 1:1</td>
<td>54 °C</td>
</tr>
<tr>
<td>E5:R5 = 1:2</td>
<td>55.5 °C</td>
</tr>
<tr>
<td>E10:R10 = 1:1</td>
<td>41 °C</td>
</tr>
<tr>
<td>E10:R10:POG = 1:1:1*</td>
<td>19 °C, 38.5 °C, and 68 °C</td>
</tr>
<tr>
<td>E10:R10:POG = 1:1:1</td>
<td>54 °C</td>
</tr>
</tbody>
</table>

Peptides were preheated to ensure equilibrium unless marked with an asterix.

Glu substitution shows the same pattern but with greater destabilization per substitution. Using these results concerning the homotrimers, I was able to assess heterotrimeric helix assembly and stability. My hypothesis was that pairing Glu in the X position with Arg in the Y position would lead to the formation of a stabilized heterotrimer through charge paired hydrogen bonding. In homotrimeric systems, such an interaction has been shown to be possible if the charged residues are present in consecutive X and Y positions. Instead of incorporating both the residues in one polypeptide, I incorporated them in two different polypeptides in such a way that these two residues can
interact favorably only upon the formation of a heterotrimer. Mixtures of E3 with R3 appear to confirm this with the observation of a species that melts at a temperature intermediate to that of either E3 or R3 alone. However, the melting temperatures of E3, R3 and E3/R3 are close to one another (51.5 °C, 58.5 °C and 56.5 °C respectively) making analysis difficult. To help clarify the situation, E5 and R5 were prepared which have a wider separation of melting points. The mixture shows a melting point of 54.0 °C, surprisingly close to R5 alone (55.5 °C).

Mixtures of E5/POG and R5/POG also show melting points surprisingly close to POG alone and almost indistinguishable from one another. One possible interpretation of this is that the more stable peptide forms a homomeric triple helix while the less stable peptide remains unfolded. This can be ruled out by examining the MRE values of the various mixtures that do not dramatically change depending on components being mixed. AAB heterotrimers have two identical chains and one different chain. When peptides E5 and POG are mixed with each other and allowed to equilibrate, two kinds of AAB heterotrimers can be formed. One type will be composed of two E5 chains and one POG chain, abbreviated as E5•E5•POG and the other type will be composed of one E5 chain and two POG chains, abbreviated as E5•POG•POG. Furthermore, as observed from the non-preheating data for both E5/POG and R5/POG mixtures, the two transitions separated by 21 °C in the case of E5/POG are well resolved. The transitions observed for R5/POG mixture are separated by around 11 °C and are not as well resolved as the other case, with two peaks in the first order derivative convoluting with each other. Thus, it is reasonable to assume that the assigning two separate melting temperatures to two transitions separated by less than or close to 7 °C - 9 °C is impossible using CD analysis.
As I observe only one transition in the preheating case for both the mixtures, it can either mean that only one of the AAB type heterotrimerers are forming, or both the types are forming but have a melting temperature within 7 °C – 9 °C of each other thereby resulting in observation of a single transition. E5•POG•POG heterotrimer has two chains rich in imino acid residues that are believed to contribute significantly to the thermal stability of a triple helix. E5•E5•POG heterotrimer has only one such chain. Thus, I believe that E5•POG•POG heterotrimer is the major species formed, as it is more stable than the E5•E5•POG heterotrimerer A similar analysis can be applied to the mixtures of R5/POG, E3/R3, E5/R5 and E10/R10 peptides. However, there is no experimental evidence yet to suggest that this is indeed the case. Therefore, AAB and ABB heterotrimerers are grouped together for the subsequent analysis in this and the remainder of the chapters.

In recent studies,\(^{86,87}\) performed on (PPG\(_{10}^{}\))/(POG\(_{10}^{}\) mixtures (abbreviated here as PPG and POG), the T\(_m^{}\) value increased almost linearly in the order 3•PPG (44.9 °C), 2•PPG/POG(54.8 °C), PPG/2•POG (63.6 °C), and 3•POG (68.4 °C). My studies indicate that T\(_m^{}\) value increases non-linearly as the fraction of the more stable homotrimer is increased in the mixture (Table 3-3) with T\(_m^{}\) for all the mixtures clustered very close to the T\(_m^{}\) for the more stable homotrimeric helix. This behavior can be attributed to the differences in electrostatic interactions between the oppositely charged monomers, or between charged and neutral monomers, which are possible only in heterotrimeres. In the extreme example of E10 and R10, no homotrimeric helices are formed in the time frame of the experiment. However, mixing the two peptides results in the formation of a triple helix with a melting temperature of 41.0 °C. Adjusting the pH of the solution to 3 to eliminate the charge on E10 also eliminates possible charge pairing and helix formation.
Two factors must be considered while analyzing the formation of these heterotrimers: inherent helix propensity of the given amino acids and electrostatic interactions between them. It has been shown that P and O are preferred over E and R respectively and this predisposition dominates helix formation in most cases. Pairing E with R leads to the elimination of repulsive like-like charge and incorporation of attractive opposite charge interactions. However, there is a competition between the destabilization caused by the substitutions and the stability incurred by the attractive interactions and the final stability depends on the relative magnitude of both. Therefore the peptides E5 or R5 can be stabilized by mixing with POG, which eliminates repulsive interstrand interactions, or by mixing with one another where the interstrand interactions are favorable.

Mixing POG with E10 and R10 can theoretically result in the formation of seven types of triple helixes, including POG, E10 and R10 AAA homotrimers, E10/R10, E10/POG and R10/POG AAB heterotrimers and an E10/R10/POG ABC heterotrimer. This analysis groups together heterotrimers that are composed by the same units but in different ratios, for instance, AAB and ABB heterotrimers are grouped together. E10 and R10 homotrimers can be excluded from consideration by studies on these species in isolation, which show that they do not form helices. Likewise, unfolding studies on E10/POG and R10/POG mixtures show that neither E10 nor R10 interact with POG. Instead unfolding curves show a transition corresponding to POG homotrimers in both mixtures. Additionally, the MRE value is approximately half of the observed value for pure POG indicating the POG homotrimeric helix is forming and being observed at half the normal concentration while the charged peptide (either E10 or R10) remains unfolded. The other species (POG AAA homotrimer, E10/R10 AAB heterotrimer and
E10/R10/POG ABC heterotrimer) may be possible. A 1:1:1 mixture of these peptides without preheating results in the observation of three distinct transitions at 68 °C, 38.5 °C and 19 °C. These can be attributed to POG homotrimer, E10/R10 heterotrimer and a much less stable component which can not yet be identified but may be an out of register or misaligned triple helix. Preheating merges these three transitions into one new transition with T_m = 54.0 °C. This can only be attributed to the formation of an ABC heterotrimer. This ABC heterotrimer is significantly more stable than the E10/R10 heterotrimer because addition of a POG component allows the complete triple helix to have an overall neutral charge. It also has an MRE value almost double from what is observed for AAB heterotrimers in the E10/R10 mixture, which indicates that nearly all of the peptide in the solution forms the ABC helix.

As mentioned previously, the close packing of a collagen triple helix requires the presence of Gly in the center of the helix. This is allowed by staggering of the three chains in a triple helix by one residue relative to each other. In natural collagens, this register or the relative stagger of three chains in determined by the globular proteins. For a homotrimer, all the registers are same as the molecule is symmetrical. However, in an ABC heterotrimer, all the three chains are unique and there are six possible registers for this triple helix. These six registers include ABC, ACB, CAB, CBA, BCA and BAC for a heterotrimer composed of peptides A, B and C. As only one transition is observed in thermal unfolding analysis using CD spectroscopy, it means that the all these possible registers, if present, have melting temperature very close to each other and in a range of 7 °C – 9 °C (as explained earlier in the discussion for AAB heterotrimers). Furthermore, it is also possible that one of the register leads to appropriate positioning of the charged
residues very close to each other. To answer these questions, Jorge Fallas, another graduate student in the laboratory, performed extensive NMR studies and some of the preliminary results are reported in the next chapter.

Perhaps surprisingly, POG homotrimer is not observed in these mixtures despite it being the most stable individual triple helix. However, if POG homotrimer is to form, it must also lead to formation of the less stable E10/R10 AAB type heterotrimers as my studies show that E10/POG and R10/POG heterotrimers and E10 and R10 homotrimers do not form. The preferred formation of ABC type E10/R10/POG heterotrimer over a mixture of POG homotrimer and AAB type E10/R10 heterotrimer suggests that ABC type heterotrimer formation leads to greater system wide energy stabilization even though POG homotrimer is more stable in isolation. These results suggest that triple helix stability is not simply due to an amino acids propensity to adopt a particular conformation, but a convolution of amino acid conformational propensity, electrostatic attraction or repulsion and cross-strand pairing of amino acids. This allows highly stable motifs such as POG to significantly stabilize poor triple helix forming sequences such as EOG far beyond what one would expect from simply diluting out the bad amino acid. Similarly, oppositely charged amino acids E and R can form heterotrimers with stability comparable to R alone. I observe an effect in which a weak triple helix forming peptide can be "rescued" by mixing it with a stronger triple helix forming peptide. The extent of rescue depends on the individual polypeptides and their ability to interact with each other by favorable electrostatic interactions. Heterotrimer formation therefore allows the introduction of greater chemical diversity in the collagen triple helix as amino acids with weak individual propensity for triple helix formation can be paired with strong ones.
3.5 Conclusions

In this chapter, I have elaborated the design, synthesis and characterization of a series of peptides which utilize charged amino acids and the ubiquitous X-Y-Gly sequence to bias their self-assembly into collagen-like heterotrimeric helices. In addition to five novel AAB heterotrimers, I demonstrate for the first time the formation of an ABC heterotrimer utilizing only supramolecular interactions. Melting studies show that amino acids which substantially destabilize the triple helix, such as Glu, can form triple helices of high stability when paired against oppositely charged amino acids such as Arg or, more surprisingly, when paired against the neutral amino acid Hyp. This results in the stability of the heterotrimeric helix to be almost indistinguishable from those containing only amino acids with high triple helix stability. Heterotrimer formation therefore allows the introduction of greater chemical diversity in the collagen triple helix as amino acids with weak individual propensity for triple helix formation can be paired with - and rescued by - strong ones. These results lay the groundwork for a better understanding of the thermal stability of heterotrimeric collagens, for the synthesis of synthetic extracellular matrix mimics and molecular recognition devices based on these interactions.

In the next chapter, I optimize the formation of ABC heterotrimer by considering all the possible charge pairs and their contributions to the stability of a collagen-like heterotrimer.
Chapter 4

Optimizing the Thermal Stability of Collagen-like Heterotrimers

4.1 Introduction

Understanding the effect of amino acid substitution on collagen triple helix stability is crucial in analyzing the structural, mechanical and biological changes observed in mutant collagens which lead to diseases such as Osteogenesis Imperfecta (OI) and other deleterious sequence variations.\textsuperscript{29-32} OI is a brittle bone disease which is caused by mutations in type I collagen, an AAB heterotrimer consisting of two identical $\alpha_1$(I) and one $\alpha_2$(I) chains.\textsuperscript{29-32} Other naturally occurring heterotrimeric collagens like AAB type IV, VIII and ABC type V, VI and IX are involved in various diseases including Alport Syndrome, Ehlers-Danlos Syndrome and Bethlem Myopathy.\textsuperscript{29-32} As detailed in the introduction chapter, collagen homotrimers have been extensively used to study amino acid substitutions,\textsuperscript{43-46,48} their PPII helix propensity,\textsuperscript{44} their effect on stabilization by water-mediated hydrogen bonding\textsuperscript{26,27,56-59} or steric and stereoelectronic effects\textsuperscript{90,100,102,111,112} and from the inter-helix amino acid interactions.\textsuperscript{26,27,56-59} However most natural collagens, including the most abundant type I collagen, are heterotrimers, not homotrimers, and diseases caused by mutations in these sequences affect only one or two of the three chains which make up the helix. A number of studies have prepared collagen heterotrimers by cysteine-knot strategy,\textsuperscript{41,78-85,88,89} however, only a few have shown non-covalent assemblies that form heterotrimers.\textsuperscript{86,87,90} Because of the importance
of heterotrimeric in natural collagens, it is important to design systems that have the potential to study these variations in their native form, whether they are AAB heterotrimers or ABC heterotrimers. The synthesis of extracellular matrix mimics based upon collagen-like peptides has completely relied upon homotrimeric. Collagen-like heterotrimers, with their ability to assemble multiple functional groups close to each other and thereby increase chemical diversity, can lead to the synthesis of more sophisticated biomaterials which more closely mimic extracellular matrix\textsuperscript{49,115-120} when compared to those based on homotrimeric. Furthermore, the effect of amino acid substitution, for example Gly mutations, could be assessed in one, two or all three strands of the triple helix. In this chapter, I demonstrate the synthesis and characterize the stability of four ABC and four AAB collagen heterotrimeric and illustrate how results from the analysis of homotrimeric cannot be directly applied to heterotrimeric. I conclude that to more fully understand the structural implications of collagen structural mutants, these molecules must be prepared as heterotrimeric and not as homotrimeric.

As previously mentioned, collagen is characterized by a unique right-handed triple helical structure composed of three left-handed PPII helices.\textsuperscript{1,19-26} The primary sequence of collagen has a X-Y-Gly repeating motif, where X and Y are generally P and O respectively. The presence of G at every third position allows the formation of a tightly packed triple helix which, along with an extensive network of CO\(_{(X)}\)-NH\(_{(Gly)}\) hydrogen bonds, imparts stability to the triple helix. Side chains of amino acids in the X and Y positions are exposed to solvent to varying degrees\textsuperscript{52} and can take part in intra- and intermolecular interactions.\textsuperscript{53-55} Substitution of imino acids (P and O) by amino acids has been shown to cause destabilization of the triple helix.\textsuperscript{43-46} Charge pair interactions have
also been studied in contiguous triplets and they follow a pattern with GER causing the least destabilization followed by GDR, GEK and GDK when compared to parent polypeptide containing GPO. KGE and KGD triplets have also been shown to stabilize the triple helical structure as much as the parent OGP triplet\textsuperscript{43-46} and are found abundantly in natural collagens. These results can be used to predict the stability of charge pairs for homotrimers. However, the effect observed on stability will be a mixture of attractive and repulsive interactions between various charges, along with the PPII helix propensity of every substituted amino acid. Heterotrimers assembled by charge pair interactions may lead to a better estimate of the attractive interactions between the opposite charges because these interactions can be studied in isolation from other complicating interactions. I described the formation of ABC heterotrimeric system that consisted of negatively charged, positively charged and neutral polypeptide with the sequence (EOG)\textsubscript{10}*(POG)\textsubscript{10}*(PRG)\textsubscript{10} in the last chapter. Extending this strategy I can assess the magnitude of attractive interactions between E – R, D – R, E – K and D – K in heterotrimeric collagen helices.

The polypeptides prepared are abbreviated as $O$, $E$, $D$, $R$ and $K$ throughout the remainder of this chapter, as shown in Table 4-1. All the peptides used in the study are N-terminal acetylated and C-terminal amidated to avoid charge interactions between the termini.\textsuperscript{48} The triple helices are abbreviated, for example, $E\cdot R\cdot O$ for (EOG)\textsubscript{10}*(PRG)\textsubscript{10}*(POG)\textsubscript{10}. 
Table 4-1. Sequence of peptides studied.

<table>
<thead>
<tr>
<th>#</th>
<th>Sequence</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(POG)\textsubscript{10}</td>
<td>O</td>
</tr>
<tr>
<td>2</td>
<td>(EOG)\textsubscript{10}</td>
<td>E</td>
</tr>
<tr>
<td>3</td>
<td>(DOG)\textsubscript{10}</td>
<td>D</td>
</tr>
<tr>
<td>4</td>
<td>(PRG)\textsubscript{10}</td>
<td>R</td>
</tr>
<tr>
<td>5</td>
<td>(PKG)\textsubscript{10}</td>
<td>K</td>
</tr>
</tbody>
</table>

The N and C terminal of all peptides are acetylated and amidated respectively.

I report the formation of an ABC heterotrimer with stability comparable to that of \textit{O\cdot O\cdot O} homotrimer. This extremely stable ABC heterotrimer consists of \textit{D\cdot K\cdot O} as its triple helix and thus contains twenty imino to amino acid substitutions. The attractive interactions between the D and K side chains are strong enough to overcome the low PPII helix propensity of both these amino acids\textsuperscript{43-46} and are able to stabilize the heterotrimer to within 2.5 °C of the \textit{O\cdot O\cdot O} homotrimer. Thermodynamic analysis of this heterotrimer shows that its formation is enthalpically more favorable but entropically less favorable when compared to \textit{O\cdot O\cdot O} homotrimer, with an overall Gibbs free energy equivalent to that of the homotrimer. This work is a step closer to understanding of thermal stability of collagen heterotrimers, and lays the foundation for the detailed study of amino acid sequence – structure relationship in heterotrimers and the synthesis of extracellular matrix mimics\textsuperscript{49,115-120}.

4.2 Results and Discussion

Mixing a positive, negative and neutral peptide together can theoretically lead to ten possible collagen-like triple helices: three AAA homotrimers, six AAB type
heterotrimers and one ABC heterotrimer. To differentiate between these possible outcomes, the potential formation of homotrimers (AAA) and heterotrimers (AAB and ABC) was assessed separately by thermal unfolding experiments. Eliminating some of the possible outcomes makes analysis of the complete mixture easier.

My design utilizes the simple idea that like charges will repel each other while opposite charges will attract each other. Because of charge repulsion only one homotrimer (O•O•O) is able to form. This reduces the analysis of any A/B/C mixture to eight possible triple helices. Mixing all possible pairs of peptides allowed us to assess the potential of AAB heterotrimer formation. Results from all the unfolding experiments on AAB heterotrimers show that only pairs of oppositely charged peptides lead to a helix formation and these triple helices are of rather low stability. Therefore, four AAB heterotrimers composed of charged and neutral peptides can also be eliminated. This leaves us with two AAB heterotrimers composed of oppositely charged peptides. Within my experimental system, AAB and ABB triple helices are not distinguishable, thus I combine these two possibilities into one. Finally I am left with three possible triple helices to consider: The O•O•O homotrimer with high stability, AAB type heterotrimers that incorporate a positively charged peptide with a negatively charged peptide and are of relatively low stability, and a novel ABC heterotrimer with unknown stability.

4.2.1 Homotrimers. O•O•O, E•E•E, R•R•R, D•D•D and K•K•K homotrimers were studied at a concentration of 0.2mM in 10mM phosphate buffer at neutral pH. As expected, only O•O•O forms a triple helix within the time frame of the experiment. Unfolding studies for E•E•E, R•R•R, D•D•D and K•K•K homotrimers show a linear decrease in ellipticity, suggesting that they do not form triple helices, as shown in Figure
4-1 for D•D•D and K•K•K homotrimers. The data for E•E•E and R•R•R homotrimers was presented in the last chapter. Rather they are present as disordered or weak PPII helices.

Therefore, in the assessment of a mixture of three different peptides such as D, K and O peptides, the D•D•D and K•K•K homotrimers can be eliminated. This reduces the analysis to eight possible triple helices: the O•O•O homotrimer, six AAB type heterotrimers and the D•K•O ABC heterotrimer. Formation of AAB and ABC heterotrimers was then assessed separately by mixing the corresponding peptides.

![Graphs](image)

**Figure 4-1.** Circular dichroism for homotrimeric triple helices of D and K. Thermal unfolding for (a) D and (b) K shows a linear decrease in ellipticity indicating that these peptides do not form triple helices.

4.2.2 AAB Heterotrimers. Oppositely charged peptides were mixed in 1:1 ratio to achieve a final peptide concentration of 0.2mM in 10mM phosphate buffer at neutral pH. The mixtures were studied with and without preheating. I previously reported the results for E/R mixture, and showed that although the E•E•E and R•R•R homotrimers do not exist, they form an AAB heterotrimer with a T_m = 41 °C upon mixing and preheating at neutral pH.
$T_m$ values of 34.5 °C and 36 °C were observed for 1:1 $D/K$ mixture for preheating and non-preheating respectively (Figure 4-2(a) and 4-2(b)). The corresponding values of $T_m$ were 38 °C and 38.5 °C for preheated and non-preheated $E/K$ mixture respectively (Figure 4-4(a) and 4-4(b)). $D/R$ mixture showed the corresponding $T_m$ values of 42 °C and 36 °C for preheating and non-preheating respectively (Figure 4-5(a) and 4-5(b)). $T_m$ values for all the mixtures are summarized in Table 4-2.

![Graphs](image)

**Figure 4-2.** Circular dichroism for AAB and ABC heterotrimeric helices formed by mixing $D$, $K$ and $O$. (a) Thermal analysis shows co-operative unfolding for $D/K$ mixture in 1:1 ratio for the non-preheated and preheated case. (b) The first derivative of unfolding versus temperature. (c) Co-operative unfolding for $D/K/O$ mixture in 1:1:1 ratio for non-preheated and preheated case. (d) The first derivative of unfolding shows that $D/K$ AAB heterotrimer and $O\cdot O\cdot O$ homotrimer are observed without preheating, and these merge to a single transition corresponding to the $D\cdot K\cdot O$ ABC heterotrimer after preheating.
AAB heterotrimer formation is observed in all the cases with and without preheating. Furthermore, on lowering the pH of the various mixtures to 3, the positive peak decreased in intensity signaling that the triple helix is destroyed, as shown in Figure 4-3. At pH 3, R and K side chains have a positive charge but D and E side chains are neutral. Owing to the absence of negative charge on D and E side chains, the ion pairs can no longer form that leads to destabilization of the triple helix.

![Figure 4-3](image-url)

**Figure 4-3.** Circular dichroism analysis of peptide mixtures at pH7 and pH3. Triple helix formation is observed at pH7 and the assembly ceases to exist when the pH is lowered to 3. (a) D/K mixture. (b) E/K mixture. (c) D/R mixture.

It has been reported that the content of imino acids in human fibril-forming collagens [type I (α1 and α2), II (α1), III (α1), V (α1, α2 and α3) and X1 (α1 and α2)] is around 35%, and the content of triplets with E/D in the X position and R/K in the Y position is around 6-8% of all triplets. Using peptides with a sequence Ac-(GPO)_{3}-GXY-(GPO)_{4}-GG-CNH_{2}, Brodsky and coworkers reported on the quantification of pairwise interactions in collagen homotrimers. For triplets with E/D in X position and R/K in Y position, they observed that GER triplet caused the least amount of destabilization followed by GDR, GEK and GDK with \( T_m \) of 40.4 °C, 37.1 °C, 35 °C and 30.9 °C respectively (\( T_m = 44.5 \) °C for GPO). In the current study, \( T_m \) values observed for various
AAB heterotrimer approximately follow the same trend with the corresponding \( T_m \) values as follows: \( E/R \) mix (41 °C), \( D/R \) mix (42 °C), \( E/K \) mix (38 °C), \( D/K \) mix (36 °C). In both the previously published homotrimers\(^{45,121}\) and the AAB heterotrimers reported here, there is a combination of repulsive interactions between the same charges and the attractive interactions between opposite charges, along with the PPII helix propensity of amino acids. The combination of amino acid propensity and electrostatic interactions determine the final stability of the triple helix, and charge pairs in AAB heterotrimers closely follow the trend as reported in literature.

![Graphs](image)

**Figure 4.4.** Circular dichroism for AAB and ABC heterotrimeric helices formed by mixing \( E, K \) and \( O \). (a) MRE versus temperature for \( E/K \) mixture in 1:1 ratio. (b) The first derivative of unfolding versus temperature for (a). (c) MRE versus temperature for \( E/K/O \) mixture in 1:1:1 ratio. (d) The first derivative of unfolding versus temperature for (c).
Unfolding studies were also performed on \( D/O \) and \( K/O \) mixtures in 1:1 ratio to study the interactions between a charged and neutral polypeptide and their ability to form a heterotrimer. The melting curve for both showed a single transition with \( T_m \) of 67 °C and 68 °C for \( D/O \) and \( K/O \) mixture respectively, which is indistinguishable from the \( T_m \) value of 67.5 °C for the \( O\cdot O\cdot O \) homotrimer (Figure 4-6(b)). The MRE value for these mixtures is about one-half for what is observed for \( O\cdot O\cdot O \) alone (Figure 4-6(a)) indicating that only \( O\cdot O\cdot O \) contributes to the MRE of the mixture while neither the \( D\cdot D\cdot D, K\cdot K\cdot K \) homotrimers nor AAB heterotrimers form. It was also observed that the CD spectra for both \( D/O \) and \( K/O \) are very similar before and after preheating (Figure 4-6(c)). These results indicate that \( D \) and \( K \) peptides do not interact with \( O \). Instead, the \( D \) and \( K \) peptides are present as disordered or weak PPII helices. Results for \( E/O \) and \( R/O \) mixtures are similar, as reported in the last chapter.

Considering all the results from the unfolding experiments on AAB heterotrimers reveals that only pairs of oppositely charged peptides lead to helix formation and these triple helices are of rather low stability. Therefore, in the assessment of a mixture of \( D, K \) and \( O \) peptides, four AAB heterotrimers composed of charged and neutral peptides can be eliminated. This leaves us with two AAB heterotrimers composed of oppositely charged \( D \) and \( K \) peptides. Combining these results with the results from unfolding studies on homotrimers, the analysis can be reduced to three possible triple helices: the highly stable \( O\cdot O\cdot O \) homotrimer, AAB type heterotrimers that incorporate the positively charged peptide with negatively charged peptide and are of relatively low stability, and a novel \( D\cdot K\cdot O \) ABC heterotrimer with unknown stability.
Table 4-2. Melting temperatures (°C) of AAB and ABC heterotrimers.

<table>
<thead>
<tr>
<th>AAB Heterotrimer</th>
<th>Non-Preheated</th>
<th>Preheated</th>
<th>ABC Heterotrimers</th>
<th>Non-Preheated</th>
<th>Preheated</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D/K)</td>
<td>34.5</td>
<td>36</td>
<td>(D\cdot K\cdot O)</td>
<td>35, 68.5</td>
<td>65</td>
</tr>
<tr>
<td>(E/K)</td>
<td>38.5</td>
<td>38</td>
<td>(E\cdot K\cdot O)</td>
<td>38, 68</td>
<td>57.5</td>
</tr>
<tr>
<td>(E/R)</td>
<td>9, 41.5</td>
<td>41</td>
<td>(E\cdot R\cdot O)</td>
<td>19, 38.5, 68</td>
<td>54</td>
</tr>
<tr>
<td>(D/R)</td>
<td>36</td>
<td>42</td>
<td>(D\cdot R\cdot O)</td>
<td>34.5, 68</td>
<td>44.5</td>
</tr>
</tbody>
</table>

4.2.3 ABC Heterotrimers. Charged peptides were mixed with \(O\) in a 1:1:1 ratio and the mixtures were studied with and without preheating. As deduced above, mixing \(D\), \(K\) and \(O\) can lead to the formation of three triple helices: \(O\cdot O\cdot O\) homotrimer, AAB type heterotrimers that incorporate the positively charged peptide with negatively charged peptide, and a novel \(D\cdot K\cdot O\) ABC heterotrimer. In a 1:1:1 \(D/K/O\) mixture, under non-preheating conditions, two transitions were observed corresponding to \(O\cdot O\cdot O\) (68.5 °C) and \(D/K\) AAB heterotrimer (35 °C). The transition at 35 °C can be assigned to \(D/K\) AAB heterotrimer as a \(D/K\) mixture shows a transition corresponding to 34.5 °C under non-preheating conditions. Both of the transitions observed in non-preheated a \(D/K/O\) mixture merged to a single transition upon preheating corresponding to \(T_m\) of 65 °C (Figure 4-2(c) and 4-2(d)). This transition can only be attributed to the \(D\cdot K\cdot O\) ABC heterotrimer, which is significantly more stable than \(D/K\) AAB heterotrimer (\(\Delta T_m = +29 \, ^\circ\)C) and only slightly less stable than \(O\cdot O\cdot O\) (\(\Delta T_m = -2.5 \, ^\circ\)C). If \(O\cdot O\cdot O\) homotrimer were to form, and not the \(D\cdot K\cdot O\) ABC heterotrimer, it must also result in the formation of a less stable \(D/K\) AAB heterotrimers or unfolded species. A greater system-wide stability is expected when an ABC heterotrimer is formed as almost all the polypeptides are involved in a stable triple helix formation. This contrasts to the case in which both the \(O\cdot O\cdot O\) homotrimer, with
stability comparable to the ABC heterotrimer, and the \( D/K \) AAB heterotrimers, which have relatively low stability, are present. The same analysis applies to other ABC heterotrimers discussed in this article.

Figure 4-5. Circular dichroism for AAB and ABC heterotrimeric helices formed by mixing \( D, R \) and \( O \). (a) MRE versus temperature for \( D/R \) mixture in 1:1 ratio. (b) The first derivative of unfolding versus temperature for (a). (c) MRE versus temperature for \( D/R/O \) mixture in 1:1:1 ratio. (d) The first derivative of unfolding versus temperature for (c).

Analysis of the \( E/K/O \) mixture was similar. Two transitions were observed in the non-preheating case corresponding to \( O\cdot O\cdot O \) (68 °C) and \( E/K \) AAB heterotrimers (38 °C). A \( T_m \) of 38.5 °C was observed for the \( E/K \) mixture under non-preheating conditions, thus the transition at 38 °C in the \( E/K/O \) mixture could be assigned to the \( E/K \) AAB
heterotrimer. Only a single transition corresponding to the $E\cdot K\cdot O$ ABC heterotrimer with a $T_m$ of 57.5 °C was observed after preheating in case of the $E/K/O$ mixture (Figure 4-4(c) and 4-4(d)).

The $D/R/O$ mixture behaved in a different way when compared to other mixtures. Under non-preheating conditions, two transitions were observed at $T_m = 68$ °C and 34.5 °C corresponding to $O\cdot O\cdot O$ and $D/R$ AAB heterotrimers respectively. A $T_m$ of 36 °C was observed for the $D/R$ mixture under non-preheating conditions, thus the transition at 34.5 °C in the $D/R/O$ mixture could be assigned to a $D/R$ AAB heterotrimer. When the $D/R/O$ mixture was preheated, two transitions were still observed: a weak transition at $T_m = 68.5$ °C, corresponding to residual $O\cdot O\cdot O$ and a strong transition at 44.5 °C (Figure 4-5(c) and 4-5(d)). The corresponding $T_m$ value for $D/R$ mixture upon preheating was 42 °C. I believe that in case of a preheated $D/R/O$ mixture, the transition observed at around 44.5 °C corresponds to a mixture of $D/R$ AAB heterotrimers (probably a minority component based on the intensity of the $O\cdot O\cdot O$ transition) and the $D\cdot R\cdot O$ ABC heterotrimer. It seems that the ABC heterotrimer in this case is not much more stable compared to the AAB heterotrimer, leading to the formation of a mixture of both AAB and ABC heterotrimers. This idea is also supported by the fact that a weak transition corresponding to $O\cdot O\cdot O$ still appears in the preheated $D/R/O$ mixture. This mixture acts as a control to the other systems shown in this article as it demonstrates that residual $O\cdot O\cdot O$ can be observed even after preheating. All other mixtures show only a single transition in the preheated case and do not show the presence of any residual $O\cdot O\cdot O$ homotrimer. Melting temperatures for all the ABC heterotrimers is summarized in Table 4-2.
Figure 4-6. Circular dichroism analysis for D/O and K/O mixtures. (a) MRE versus temperature. (b) The first derivative of unfolding versus temperature for (a). (c) MRE versus wavelength. The corresponding data for E/O and R/O has been shown previously and it behaves in a similar fashion\textsuperscript{114}.

Based on the results from the AAB heterotrimers, the melting temperature follows a trend in terms of contribution of charge pairs to the stability of triple helix and it is in agreement with the results obtained for charge pairs previously reported\textsuperscript{45,121}. $T_m$ values observed for various AAB heterotrimer were as follows: $E/R$ mix = 41 °C, $D/R$ mix = 42 °C, $E/K$ mix = 38 °C, $D/K$ mix = 36 °C. However, this trend is completely reversed when the stability of ABC heterotrimers is assessed. The results show that the D•K•O heterotrimer is the most stable with $T_m = 65$ °C, followed by $E•K•O$ ($T_m = 57.5$ °C) and $E•R•O$ ($T_m = 54$ °C). For D/R/O mixture, it cannot yet be unambiguously identified whether the transition at approximately 44.5 °C corresponds to an AAB heterotrimer, ABC heterotrimer or a mixture of both but it is clearly weaker than the other systems. I believe that the trend observed from the ABC heterotrimers is a true indication of the stabilization caused by interaction between opposite charges, rather than what is observed in AAB heterotrimers or homotrimers. In AAB heterotrimers, as two chains are identical, there is always a net positive or negative charge. Homotrimers can be designed to be neutral overall, but consequently have no net intermolecular electrostatic interaction.
Thus, in both the cases, the stabilizing effect of opposite charge pairs are diluted or eliminated. An ABC heterotrimer, however, is composed of one neutral, one negatively charged and one positively charged chain at neutral pH. Thus, the repulsive electrostatic interactions between peptide chains can be eliminated and the effect of attractive interactions between oppositely charged chains can be evaluated. I observe that the favorable interactions between D and K side chains in the case of $D \cdot K \cdot O$ are able to increase its stability to nearly that of $O \cdot O \cdot O$. $D \cdot K \cdot O$ is surprisingly stable even though D and K amino acids have significantly less inherent propensity to form a triple helix than P and O imino acids and occur a total of twenty times in the triple helix. As $D \cdot K \cdot O$ has ten P and O imino acid residues substituted by D and K respectively as compared to $O \cdot O \cdot O$. Therefore the imino acid content of $D \cdot K \cdot O$ decreases to approximately 45% when compared to 67% for $O \cdot O \cdot O$. Molecular modeling of $D \cdot K \cdot O$ indicates that the side chains of Lys and Asp can be ideally positioned to form an ionic hydrogen bond with a N-O distance of 2.7 Å (Figure 4-7). This potential for the formation of this ideally situated hydrogen bonded charge pair may be the reason for the high stability of the triple helix.

![Figure 4-7. Molecular model for the $D \cdot K \cdot O$ heterotrimer. Only a portion of the heterotrimer is shown for clarity. $D$, $K$ and $O$ polypeptides are represented by red, blue and green chains respectively. Distances between nitrogen (blue) of K side chain and the closest oxygen (red) of D are shown.](image-url)
As mentioned in the previous chapter, the component chains of an ABC heterotrimer can assemble in six different registers. For example, an ABC heterotrimer composed of peptides $D$, $K$ and $O$ can have the following six registers: $D\cdot K\cdot O$, $D\cdot O\cdot K$, $K\cdot D\cdot O$, $K\cdot O\cdot D$, $O\cdot K\cdot D$ and $O\cdot D\cdot K$. Extensive NMR studies have been performed on the mixture of $D$, $K$ and $O$ peptides by Jorge Fallas, a graduate student in the laboratory. Total correlated spectroscopy (TOCSY), nuclear overhauser enhanced spectroscopy (NOESY) and 1-D heteronuclear single quantum coherence (HSQC) experiments have been performed on the mixture of $D$, $K$ and $O$ peptides, as $D\cdot K\cdot O$ heterotrimer forms the most stable triple helix. The results verify the presence of ABC collagen-like heterotrimer, and quite surprisingly a small amount of residual AAA homotrimer, even in the case of the most stable $D\cdot K\cdot O$ heterotrimer. I expect a similar result for all other peptide mixtures as well. These results are in stark contrast with the results from CD spectroscopy, where the presence of residual $O\cdot O\cdot O$ homotrimer was observed only in the mixture of $D$, $R$ and $O$ peptides that forms the least stable $D\cdot R\cdot O$ heterotrimer. This is probably because of limited sensitivity of CD spectroscopy, and it does not detect species present only in a small amount. Furthermore, the thermal stability of the $O\cdot O\cdot O$ homotrimer is only 2.5 °C higher than the $D\cdot K\cdot O$ heterotrimer, close enough not to be distinguishable by CD spectroscopy. However, 1-D HSQC experiments determine the ratio of $D\cdot K\cdot O$ heterotrimer and the $O\cdot O\cdot O$ homotrimer to be approximately 12:1. Furthermore, TOCSY and NOESY experiments can reveal the presence of a particular register, among six possible outcomes. The experiments to determine the register are under progress.

Many factors including the PPII helix propensity of amino acids,\textsuperscript{43-46} water mediated hydrogen bonds,\textsuperscript{26,27,56-59} steric and stereoelectronic effects\textsuperscript{90,100,102,111,112} and the
interactions between side chains of neighboring amino acids play a significant role in triple helix stability. Contributions of all these factors have been studied extensively on collagen homotrimers. I examined the formation of four ABC heterotrimers that utilize electrostatic interactions between D, E, K and R amino acid side chains. \( D\cdot K\cdot O \) consists of polypeptides with ten P to D substitutions, ten O to K substitutions and without any imino acid substitution. Despite these substitutions, the melting temperature of \( D\cdot K\cdot O \) is only 2.5 °C lower than \( O\cdot O\cdot O \) (67.5 °C vs. 65.0 °C). This highly stable complex is unexpected based on previous work that examined the effect of amino acid substitutions on the stability of homotrimeric helices and demonstrates the importance of examining helix stability in heterotrimers, particularly in ABC heterotrimers. The ability to selectively form an ABC triple helix with stability comparable to the most stable collagen triple helix formed from natural amino acids will have substantial implications in the field of collagen-like peptides.

4.3 Conclusions

In this chapter I have demonstrated three critical points: 1) Mixing three peptides with individually neutral, negative and positive charge is a general mechanism for forming specific ABC heterotrimeric collagen helices in high yield. The formation of a triple helix with net-neutral charge is the thermodynamically most favored assembly over all of the other nine possible combinations and will preferentially form when the solution is given the opportunity to come to equilibrium after preheating. 2) An amino acids contribution to the stability of a heterotrimeric collagen helix can not be extrapolated from similar studies on homotrimeric collagen helices. Previous work with homotrimers demonstrated
that substitutions of Pro in the X position or Hyp in the Y position with any other natural amino acid led to significant destabilization. In my work, I observed that even as many as 20 substitutions of Pro and Hyp in a triple helix can still lead to a highly stable helix. 3) Charge pairing between Asp and Lys leads to the formation of a triple helix of equivalent stability to helices containing exclusively the collagen consensus sequence P-O-G. Construction of such high stability collagen triple helices using solely natural amino acids and without the use of covalent tethers or cross-links will have great impact on our ability to probe heterotrimer structure and stability, particularly in model peptides which mimic diseases found in naturally occurring heterotrimeric collagens. Additionally this will allow us to begin designing more sophisticated collagen based constructs both for bioengineering applications as well as for nanoscience applications derived from well-controlled self-assembling architectures.

In the subsequent chapters, I elaborate the application of the D\textbullet{}K\textbullet{}O heterotrimer as a model system for osteogenesis imperfecta, a brittle bone disease and use these synthetic collagen heterotrimers as structural mimics of wild type and mutant collagen type I. Additionally, I also use the D\textbullet{}K\textbullet{}O heterotrimer as a model system for understanding the role of 4(R)-fluororpoline in the stability of a collagen triple helix in a chain dependent manner.
Chapter 5

Collagen Heterotrimers: Peptide Model for Osteogenesis

Imperfecta

5.1 Introduction

Collagen is a major structural component of all connective tissues such as skin, ligament, cartilage, bone and tendon. Collagen is characterized by a ubiquitous X-Y-Gly repeating motif, where X and Y are generally Pro and Hyp respectively while Gly is present every third amino acid.\(^{1,19-26}\) Three left-handed PPII helices coil around each other to form a right-handed triple helical structure that is stabilized by an extensive network of CO\(_{(X)}\)-NH\(_{(Gly)}\) hydrogen bonds and a tightly packed triple helix.\(^{1,19-26}\) The absence of a side chain in the Gly residue make it sterically possible to fit in the interior of the crowded triple helix. Substituting Gly by any other amino acid results in the distortion of the triple helix\(^{26,27}\) and leads to varying degrees of destabilization depending on the substitution. Gly mutations in natural collagens lead to a variety of connective tissue disorders. Mutations in homotrimeric AAA collagens (type II, III, VII and X) lead to chondrodysplasia, Ehlers-Danlos syndrome, dystrophic epidermolysis bullosa and chondrocyte hyperthrophy respectively.\(^{29-32}\) Mutations in heterotrimeric AAB collagens (type I, IV and VIII) result in osteogenesis imperfecta, Alport syndrome and corneal endothelial dystrophy respectively, while mutations in heterotrimeric ABC collagen (type VI) result in Bethlem myopathy.\(^{29-32}\) Osteogenesis imperfecta (OI), the most studied collagen disease, is primarily caused by missense mutations in either the α1 or α2 chains
of type I collagen, which lead to the substitution of Gly in the ubiquitous X-Y-Gly repeat by bulky amino acids such as Arg, Asp, Glu, Cys, Ser, Ala or Val. 30,31 OI phenotypes vary from mild to lethal forms depending on a variety of factors. These include the chain type mutated, the proximity to the C-terminus, the neighboring sequence of the mutated site, and the amino acid residue which substitutes for Gly. It is observed that the amino acid substitutions that are lethal when present in the α1 chain may not be lethal when present in α2 chain.30,31 Furthermore, the mutations present near the C-terminus are known to be more lethal than the mutations present near the N-terminus, which correlates with the C to N terminus folding of the triple helix.62,63 More than 150 mutations have been reported for type I collagen and they are present along the length of α1 and α2 chains.

Collagen-like peptide homotrimers have been widely used as models to study the effects of Gly mutations.33-51 Type I collagen, however, is an AAB heterotrimer and the mutations which lead to OI are present in either the α1 or α2 chains, not both. Furthermore, the sequences of α1 and α2 chains are not identical. Thus, collagen-like peptide homotrimers with Gly mutations in all the three chains, do not correctly mimic the mutations found in type I collagen and result in what is expected to be an exaggerated effect on stability and folding. I reported the synthesis and characterization of various ABC type heterotrimers, including a highly stable heterotrimer composed of three unique 30 amino acid peptides: neutral (POG)10, positive (PKG)10 and negative (DOG)10. It was found to have a thermal stability comparable to the (POG)10 homotrimer which is known to be the most stable peptide motif using natural amino acids.
Using this electrostatically controlled recognition motif, I show here the development of a system which can force the interaction of any three guest peptide sequences. The guest sequences can be all the same (AAA), two the same and one different (AAB), or all different from each other (ABC). These guest peptides can be used to understand the stabilities of heterotrimeric collagens which were previously synthetically inaccessible. In this study, I take advantage of this design flexibility to study mutations which give rise to OI. As long as the POG, PKG and DOG motifs are present in the three chains to guide the formation of an ABC heterotrimer, a natural sequence can be incorporated into the peptide design so that the effect of Gly mutations can be assessed in zero, one, two or all three chains. I demonstrate the synthesis of four ABC heterotrimers which incorporate the natural sequence from residues 242-250 from either the $\alpha 1$ or $\alpha 2$ chains of human type I collagen and the Gly to Ser mutation at position 247 in either one $\alpha 2$ chain,\textsuperscript{123} two $\alpha 1$ chains\textsuperscript{124} or all three chains. For the first time, I demonstrate the difference in thermal stability and refolding times between the triple helices with zero, one, two or three Gly mutations. I believe that using the peptide design proposed in this chapter, any natural mutated or normal sequence with either AAB or ABC organization can be incorporated in a synthetic system. I expect that Gly mutants will more accurately reflect natural conditions when prepared as $\alpha 1\cdot \alpha 1\cdot \alpha 2$ heterotrimers using the peptide design described in this paper, and not as $\alpha 1\cdot \alpha 1\cdot \alpha 1$ or $\alpha 2\cdot \alpha 2\cdot \alpha 2$ homotrimers. This ability to prepare heterotrimeric collagen helices in a highly flexible fashion is also expected to be of use when designing synthetic biomaterials which attempt to emulate the structure of collagen and, more generally, the extracellular matrix.$\textsuperscript{49,115-120}$
Eight polypeptides, shown in Table 5-1, were synthesized for use in this study. Their synthesis and characterization are described in the supporting materials. There are four peptides that use the sequence 242-250 of the α1 chain of human type I collagen and are numbered from 1 to 4. Similarly, there are four peptides that use the sequence 242-250 of the α2 chain of human type I collagen and are numbered from 5 to 8. All the peptides used in this study are N-terminally acetylated and C-terminally amidated to eliminate the charge repulsion between the termini.48

<table>
<thead>
<tr>
<th>#</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(POG)₅ (PQGPGGPPG) (POG)₅</td>
</tr>
<tr>
<td>2</td>
<td>(POG)₅ (PQGPQGPPG) (POG)₅</td>
</tr>
<tr>
<td>3</td>
<td>(PKG)₅ (PQGPGGPPG) (PKG)₅</td>
</tr>
<tr>
<td>4</td>
<td>(PKG)₅ (PQGPGQPPG) (PKG)₅</td>
</tr>
<tr>
<td>5</td>
<td>(POG)₅ (PVGAAGATG) (POG)₅</td>
</tr>
<tr>
<td>6</td>
<td>(POG)₅ (PVGAASATG) (POG)₅</td>
</tr>
<tr>
<td>7</td>
<td>(DOG)₅ (PVGAAGATG) (DOG)₅</td>
</tr>
<tr>
<td>8</td>
<td>(DOG)₅ (PVGAASATG) (DOG)₅</td>
</tr>
</tbody>
</table>

N and C termini of all the peptides are acetylated and amidated respectively.

The heterotrimeric helices prepared in this study are abbreviated in such a way that the first, second and third chains in the triple helix have neutral (POG)₅, positive (PKG)₅ and negative (DOG)₅ flanking regions respectively. For example, a non-mutated heterotrimer abbreviated as $A•A•B$ consists of polypeptides 1, 3 and 7 while a mutated
heterotrimer abbreviated as $A\cdot A\cdot B'$ consists of polypeptides 1, 3 and 8. The homotrimers are abbreviated as $3A$, $3B$, $3A'$ and $3B'$ for the triple helices formed from three copies of 1, 5, 2 and 6 polypeptides respectively. The abbreviations and composition of various triple helices used in this study are shown in Table 5-2.

Table 5-2. Composition of triple helices studied.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homotrimers</td>
<td></td>
</tr>
<tr>
<td>$3A$</td>
<td>1</td>
</tr>
<tr>
<td>$3A'$</td>
<td>2</td>
</tr>
<tr>
<td>$3B$</td>
<td>5</td>
</tr>
<tr>
<td>$3B'$</td>
<td>6</td>
</tr>
<tr>
<td>Heterotrimers</td>
<td></td>
</tr>
<tr>
<td>$A\cdot A\cdot B$</td>
<td>1, 3 and 7</td>
</tr>
<tr>
<td>$A\cdot A\cdot B'$</td>
<td>1, 3 and 8</td>
</tr>
<tr>
<td>$A'\cdot A'\cdot B$</td>
<td>2, 4 and 7</td>
</tr>
<tr>
<td>$A'\cdot A'\cdot B'$</td>
<td>2, 4 and 8</td>
</tr>
</tbody>
</table>

5.2 Results and Discussion

5.2.1 Molecular and Experimental Design. I designed peptides that utilize (POG)$_3$, (PKG)$_5$ and (DOG)$_5$ motifs to organize OI causing Gly mutations in synthetic collagen like heterotrimers similar to those observed in native type I collagen, as depicted schematically in Figure 5-1. As the flanking regions in the designed peptides are predisposed to selectively form ABC collagen heterotrimers directed through electrostatic interactions, I can drive the assembly of designed peptides containing native sequences from type I collagen into triple helices utilizing these interactions. In this study, I
assemble human collagen type I sequence 242-250 in the center of the triple helix in both its normal and mutated (Gly to Ser, at position 247) forms in an AAB type register. Two peptides with the (POG)$_3$ and (PKG)$_3$ flanking regions mimic the $\alpha 1$ chain (sequence: PQGPGGPPG), and one peptide with the (DOG)$_3$ flanking region mimics the $\alpha 2$ chain (sequence: PVGAAGATG) of type I collagen, giving rise to an $\alpha 1\cdot\alpha 1\cdot\alpha 2$ register in the central region of interest. The heterotrimer with no mutations is referred to as $A\cdot A\cdot B$, and the heterotrimers with mutations in either $\alpha 1$ chains or in the $\alpha 2$ chain are referred to as $A\cdot A\cdot B'$ or $A\cdot A\cdot B''$ respectively. For comparison, I also analyzed a heterotrimer with mutations in all three chains that is referred to as $A\cdot A'\cdot B'$. Homotrimers of the $\alpha 1$ and $\alpha 2$ chains were also analyzed, both in normal and mutated forms, and are referred to as $3A$, $3B$, $3A'$ and $3B'$.

![Flanking region](image)

**Figure 5-1.** A schematic representation of the molecular design. The flanking regions are assembled from (POG)$_3$ (green), (PKG)$_3$ (blue) and (DOG)$_3$ (red) motifs, which are predisposed to form an ABC heterotrimer. The central region mimics the sequence 242-250 of type I collagen in the AAB register with the Gly to Ser mutations in either $\alpha 1$ or $\alpha 2$ chains at position 247.
5.2.2 Homotrimers. 3A, 3B, 3A′ and 3B′ homotrimers were analyzed at a concentration of 0.2 mM in 10 mM PO₄ buffer at pH 7. All the homotrimers are composed of (POG)₅ flanking regions, with polypeptide A mimicking the normal and A′ mimicking the mutated form of the α₁ chain of type I collagen. In the same fashion, polypeptide B mimics the normal and B′ mimics the mutated form of the α₂ chain of type I collagen. 3A and 3B show melting temperature of 68 °C and 54 °C respectively, as shown in Figure 5-2. As expected, the Gly to Ser mutations show a decreased stability with melting temperatures of 45 °C and 33 °C for 3A′ and 3B′ respectively. This corresponds to a destabilization of 23 °C for mutations in the A homotramer and 21 °C for the B homotramer. This decreased stability is not surprising as the substitution of Gly with the bulkier amino acid Ser leads to a disruption of the tight packing of the triple helix. A similar drop in thermal stability of approximately 22 °C was reported for a lethal Gly to Ser mutation in homotrimers mimicking sequence from residues 904 - 921 from the α₁ chain of human type I collagen and with the mutation present at position 913.⁵⁰ However, a drop of only 11 °C was reported for a non-lethal mutation present at position 901 in a homotramer mimicking sequence from residues 904 - 921 from the α₁ chain demonstrating the importance of the neighboring sequences of the mutation site. Clearly, neighboring sequences modify the extent of destabilization differently. Additionally, a lower melting temperature for 3B and 3B′ when compared to 3A and 3A′ respectively is expected on the basis on reported literature where the authors report that 3α₂ homotrimer is thermally less stable than the 3α₁ homotrimer. 3A and 3A′ heterotrimers were observed to be more stable by 14 °C and 12 °C when compared to 3B and 3B′.
heterotrimers respectively. The thermal stability of the homotrimers is summarized in Table 5-3.

Figure 5-2. Circular dichroism for homotrimeric helices mimicking sequence 242-250 of the α1 and α2 chains of type I collagen, with and without Gly mutations at position 247. (a) Thermal analysis shows co-operative unfolding for both 3A and 3A' homotrimers mimicking the α1 chain. (b) The first derivative of unfolding versus temperature shows that the thermal stability decreases by 23 °C when the Gly to Ser mutation is present in all the three chains. (c) Co-operative unfolding for both 3B and 3B' homotrimers mimicking the α2 chain. (d) Thermal stability decreases by 21 °C when Gly to Ser mutations are present in all the three chains.

Type I collagen is an AAB heterotrimer, and the mutations that lead to OI are present only in either of the α1 or α2 chains. As homotrimers are composed of three identical chains, the mutation will either be absent in all the chains or present in all chains. Thus, by using synthetic homotrimers, the mutations cannot be accurately studied for type I
collagen or any other heterotrimeric collagens and are limited to analyzing only naturally occurring homotrimers. The presence of mutations in all three chains, and not in just one or two chains in synthetic collagen like peptides, exaggerates their destabilizing contribution. In contrast, synthetic heterotrimers are an ideal system for analyzing type I collagen or any other heterotrimer mutations in their native forms as the individual chains in a heterotrimer can be tailored to have the specific mutation in one, two or all three chains. The formation of various ABC heterotrimers, which mimic the mutations in either the α1 or α2 chains of type I collagen, was assessed by combining the corresponding polypeptides.

| Table 5-3. Thermal stability and refolding half-life of triple helices studied |
|---------------------------------|-----------------|-----------------|
| Abbreviation       | Melting Temperature (°C) | t₁/₂ (sec) |
| Homotrimers         |                      |                |
| 3A                 | 68                  | 580            |
| 3A'                | 45                  | 910            |
| 3B                 | 54                  | 345            |
| 3B'                | 33                  | 655            |
| Heterotrimers       |                      |                |
| A*A*B              | 60                  | 190            |
| A*A*B'             | 44.5                | 300            |
| A*A*B              | 42.5                | 325            |
| A*A*B'             | 36.5                | 470            |

5.2.3 ABC Heterotrimers. Various peptides with (DOG)₅, (PKG)₅ or (POG)₅ flanking regions and the sequence 242-250 (Gly to Ser mutation at position 247) from type I collagen in the center were mixed together in a 1:1:1 ratio to form ABC heterotrimers with zero, one, two or three mutations. The heterotrimer formation was confirmed by CD
analysis. Mixing a neutral, positive and negative collagen-like peptide can lead to the formation of seven possible triple helices: three AAA homotrimers, six AAB heterotrimers, and one ABC heterotrimer. The formation of these homotrimers and AAB heterotrimers was assessed separately by mixing the corresponding peptides.

Analyzing $A\cdot A\cdot B$ heterotrimer shows that mixing peptides 1, 3 and 7 can lead to the formation of the following species: three AAA homotrimers consisting of 1, 3 and 7 peptides respectively; three AAB heterotrimers consisting of mixtures of 3/7, 3/1 and 7/1 respectively and one ABC heterotrimer consisting of 1/3/7 ($A\cdot A\cdot B$ heterotrimer). Out of three possible AAA homotrimers, peptides 3 and 7 do not form a stable triple helix by themselves due to charge repulsion between the side chains in the flanking regions. Peptide 1 forms a homotrimer, abbreviated as 3A in the article, with a thermal stability of 68 °C as shown in Figure 5-2. Formation of various AAB heterotrimers was then analyzed and is reported in Figure 5-3. The results show that even though peptides 3 and 7 do not form a stable triple helix by themselves, they form a stable AAB heterotrimer with a thermal stability of 33.5 °C when mixed together in 1:1 ratio. Both 3/1 and 7/1 peptide mixtures in 1:1 ratio show a thermal stability of 67.5 °C, which is indistinguishable from the thermal stability of 3A homotrimer. Additionally, the MRE value for both the mixtures was observed to be almost half of that of 3A homotrimer. These two results show that only 3A homotrimers contribute to MRE and neither 3 and 7 homotrimers nor the AAB heterotrimers are formed. Combining all the results from analysis of homotrimers and AAB heterotrimers reveal that out of seven possible triple helical species, only three including 3A homotrimer, 3/7 AAB heterotrimer and 1/3/7 ($A\cdot A\cdot B$) ABC heterotrimer can form. Non preheated mixture of the three peptides shows
two transitions in the unfolding curve with thermal stabilities of 34.5 °C and 67.5 °C as shown in Figure 5-7. The transition at 34.5 °C corresponds to a AAB heterotrimer and the transition at 67.5 °C corresponds to a 3A homotrimer. Upon preheating, both the transitions merge to a single transition at 60 °C corresponding to a 1/3/7 (A•A•B) ABC heterotrimer. Thus, I can conclude from this analysis that when three neutral, positive and negative peptides are mixed together and allowed to equilibrate, they selectively lead to the formation of only an ABC heterotrimer and no other species is observed.

\[ A\cdot A\cdot B', A'\cdot A\cdot B \] heterotrimers were analyzed in the similar fashion as A•A•B heterotrimer and the results for two peptide mixtures are shown in Figure 5-4, 5-5 and 5-6 respectively. The results for three peptide mixtures for all the heterotrimers are shown in Figure 5-7.

\[ A\cdot A\cdot B' \] heterotrimer mixture behaved in a different way when compared to the other three mixtures. Upon preheating, in addition to a transition corresponding to the heterotrimer, a weak second transition was observed at 67.5 °C which corresponds to 3A homotrimer, as shown in Figure 5-7(d). It seems that the heterotrimer formed in this particular case in very unstable (23 °C) when compared to the homotrimer and thus a mixture of both species is formed upon the preheating and the subsequent equilibration of the peptides. In all other cases, the thermal stability of the heterotrimers is only slightly lower than that of the corresponding homotrimers and thus I do not observe the formation of any residual homotrimer. Additionally, the formation of 3A homotrimers in this particular case does not affect the interpretation of results in any way as both the transitions are well separated and can be easily attributed to either of the A•A•B' heterotrimer or 3A homotrimer.
Figure 5-3. Unfolding analysis by circular dichroism for AAB heterotrimers formed by 3/7, 3/1 and 7/1 peptide mixtures. (a) MRE versus temperature. (b) The first derivative of unfolding versus temperature shows a melting temperature of 33.5 °C, 67.5 °C and 67.5 °C for 3/7, 3/1 and 7/1 peptide mixtures respectively.

Figure 5-4. Unfolding analysis by circular dichroism for AAB heterotrimers formed by 3/8, 3/1 and 8/1 peptide mixtures. (a) MRE versus temperature. (b) The first derivative of unfolding versus temperature shows a melting temperature of 19 °C, 67.5 °C and 67.5 °C for 3/8, 3/1 and 8/1 peptide mixtures respectively.
Figure 5-5. Unfolding analysis by circular dichroism for AAB heterotrimers formed by 4/7, 4/2 and 7/2 peptide mixtures. (a) MRE versus temperature. (b) The first derivative of unfolding versus temperature shows a melting temperature of 18 °C, 45.5 °C and 45 °C for 4/7, 4/2 and 7/2 peptide mixtures respectively.

Figure 5-6. Unfolding analysis by circular dichroism for AAB heterotrimers formed by 4/8, 4/2 and 8/2 peptide mixtures. (a) MRE versus temperature. (b) The first derivative of unfolding versus temperature shows a melting temperature of 11 °C, 45.5 °C and 45 °C for 4/8, 4/2 and 8/2 peptide mixtures respectively.

The heterotrimer A•A•B, with no mutation and an α1•α1•α2 register in the central region showed the highest stability with a melting temperature of 60 °C, as shown in Figure 5-7. The thermal stability dropped by 15.5 °C when the single mutation was
introduced in the α2 chain with the heterotrimer $A\cdot A\cdot B^+$ showing a melting temperature of 44.5 °C. The unfolding curve for $A\cdot A\cdot B^+$ heterotrimer also showed a second transition at 67.5 °C which can be attributed to $34$ homotrimer. However, based on the intensity of the transition, the $34$ homotrimer is a minority component and the bulk of the peptides are present as $A\cdot A\cdot B^+$ heterotrimers. Importantly, the formation of a small fraction of homotrimer in this peptide mixture does not affect the interpretation of the results in any way as the homotrimer transition is well separated from the transition corresponding to the $A\cdot A\cdot B^+$ heterotrimer. The results from this particular peptide mixture are explained in a greater detail in the supplementary section. The heterotrimer $A'\cdot A'\cdot B$ with two mutations, both in the α1 chain, showed only a small additional decrease in stability, with a melting temperature of 42.5 °C. Thermal stability of the $A'\cdot A'\cdot B^+$ heterotrimer, with mutations in all three chains, decreases further to 36.5 °C. A continuous decrease in mean residual ellipticity (MRE) values was also observed for all the heterotrimers when the number of mutations increased from zero to three. This decrease in MRE value suggests that an increasing percentage of polypeptides fail to assemble into triple helices as the number of mutations increase. The thermal stability of all the heterotrimers discussed in this study is summarized in Table 5-3.
Figure 5-7. Circular dichroism for heterotrimeric helices mimicking sequence 242-250 of α1 and α2 chains of type I collagen and Gly to Ser mutation at position 247 in either of the two chains. (a) Thermal analysis for $A\cdot A\cdot B$ shows that two transitions are observed in the non-preheated case, which merges to a single transition upon preheating. (b) The first derivative of unfolding versus temperature for the unfolding curve in (a). (c) Thermal unfolding for $A\cdot A\cdot B'$ heterotrimer. (d) The first derivative of unfolding versus temperature for the unfolding curve in (c). (e) Thermal unfolding for $A\cdot A\cdot B$ heterotrimer. (f) The first derivative of unfolding versus temperature for the unfolding curve in (e). (g) Thermal unfolding for $A\cdot A\cdot B'$ heterotrimer. (h) The first derivative of unfolding versus temperature for the unfolding curve in (g). The thermal stability decreases as the number of mutations increase.

Thermal stability decreases non-linearly as the number of mutations in the triple helix increases. Incorporation of the Gly to Ser mutation in the α2 chain has a severe effect on the thermal stability and leads to a 15.5 °C drop in the melting temperature when compared to the non-mutated $A\cdot A\cdot B$ heterotrimer. The incorporation of two mutations in the α1 chain leads to a small additional decrease in thermal stability of about 2 °C when compared to the mutation in the α2 chain. Although $A\cdot A\cdot B'$ heterotrimer contains only one and $A\cdot A\cdot B$ heterotrimer contains two mutations respectively, the difference in thermal stability between the two heterotrimers does not correctly represent the destabilization caused by the incorporation of second mutation. The first mutation is present in a different sequence environment (α2 mimic peptide) when compared to the second mutation (α1 mimic peptide) and thus the difference in thermal stability is a combined effect of different environment and an additional mutation, thereby making a direct comparison impossible. To correctly quantify the decrease in thermal stability by individual Gly to Ser mutations, I synthesized $A\cdot A\cdot B$ and $A\cdot A\cdot B'$ heterotrimers where the first mutation is incorporated in either of the α1 mimetic peptides. Both $A\cdot A\cdot B$ and
$A\cdot A'\cdot B$ heterotrimers were assembled using the same principles as described for the other heterotrimers. $A'\cdot A\cdot B$ heterotrimer is composed of peptides 2, 3 and 7 and $A\cdot A'\cdot B$ heterotrimer is composed of peptides 1, 4 and 7. $A\cdot A'\cdot B$ heterotrimer shows a slightly lower MRE value than $A'\cdot A\cdot B$ heterotrimer suggesting that the amount of peptides that form triple helix is more for $A'\cdot A\cdot B$ heterotrimer. However, both $A'\cdot A\cdot B$ and $A\cdot A'\cdot B$ heterotrimers showed a thermal stability of 47 °C, as shown in the Figure 5-8, a drop of 13 °C when compared to the non-mutated $A\cdot A\cdot B$ heterotrimer. The incorporation of second mutation in the $\alpha$1 chain leads to a further decrease of 4.5 °C. When the mutations are present in all the three chains, a further destabilization of 6 °C is observed. This result indicates that it is comparatively less damaging to incorporate the second and third mutations in a triple helix when compared to the first mutation. I believe that the incorporation of the first mutation in triple helix is most disruptive and leads to a correspondingly large drop in stability while the addition of a second and third mutations to a already disrupted packing structure leads to only a small further decrease in stability.

**Figure 5-8.** Circular dichroism for $A'\cdot A\cdot B$ and $A\cdot A'\cdot B$ heterotrimers. (a) Unfolding analysis and (b) the corresponding first order derivative shows that both the heterotrimers show a melting temperature of 47 °C. The MRE value for $A'\cdot A'\cdot B$ is slightly lower than $A'\cdot A\cdot B$. 
Although the $A^\ast A^\ast B$ heterotrimer showed only a slightly lower thermal stability (by approximately 2 °C) than the $A^\ast A^\ast B'$ heterotrimer, the melting temperature observed for both the mutated heterotrimers is in agreement with the known clinical severity of the OI types caused by mutations in the $\alpha 1$\textsuperscript{124} or $\alpha 2$ chains.\textsuperscript{123} Mutations in the $\alpha 1$ chains lead to lethal OI type II while the mutation in the $\alpha 2$ chain lead to the relatively mild OI type IV.\textsuperscript{125} As there are two $\alpha 1$ chains and one $\alpha 2$ chain in type I collagen, the lethal phenotype for the mutation in the $\alpha 1$ chain is expected as there are two mutated chains instead of a single mutation present in the $\alpha 2$ chain. However, as the thermal stability difference between the two mutated heterotrimers in only 2 °C, the known severity of the disease may not be because of a drop in thermal stability but may be a result of abnormal collagen packing during fiber formation. Additionally, the comparatively longer flanking regions in my system (5 triplets on both the boundaries when compared to 3 guest triplets in the center) may be too stabilizing, thereby shielding the destabilizing effect caused by an increased number of mutations. To accurately identify the factors responsible for the different severity of the $\alpha 1$ or $\alpha 2$ chain mutations at position 247, I am currently exploring the simultaneous effects of shortened boundary regions and longer guest regions. When the mutation is present in all three chains, the thermal stability and the MRE values are the lowest showing that the triple helix formed is very weak. Apart from the non-linear decrease in thermal stability, for the first time, I show that the effect of just one Gly to Ser mutation in a particular chain can be measured. Thus, I am able to differentiate between various triple helices that differ from each other in only one amino acid residue. Additionally, I have designed a series of collagen like ABC heterotrimers that can mimic the mutations in two different chains in type I collagen, either present
individually in the α1 and α2 chains (A•A•B and A•A•B′ heterotrimers) or present in both the α1 and α2 chains (A'•A'A•B′ heterotrimer). To further understand the effects of mutations, refolding studies were performed on the heterotrimers and refolding half times (t1/2) were compared for all triple helices studied. Additionally, the refolding studies were performed on both normal and mutated homotrimers and the results were compared with both the heterotrimers reported in this chapter and the homotrimers reported elsewhere in the literature. Refolding half time, the time at which half of the polypeptides have recovered to form a triple helix, was calculated from the plots by monitoring when the fraction folded equaled half of the maximum equilibrium value as observed after overnight incubation on a repeated samples.

Refolding studies showed that all the heterotrimers recovered to approximately 85% of their respective fraction folded values in fewer than 50 min. The respective MRE values observed after 50 min and the refolding half times followed the same trend that was observed in the unfolding studies, with A•A•B showing the highest recovered MRE and best refolding rate followed by A•A•B′, A'•A'A•B and A'•A'A•B′. A refolding half time of 190 seconds, 300 seconds and 325 seconds was observed for A•A•B, A•A•B′ and A•A•B heterotrimers respectively, as shown in Figure 5-9 and Table 5-3. The A'•A'A•B′ heterotrimer, with mutation in all three chains, refolded much more slowly with a refolding half time of 470 seconds. All the homotrimers folded slower than the heterotrimers and refolding half times of 580 seconds and 910 seconds were observed for 3A and 3A′ respectively. 3B and 3B′ showed refolding half times of 345 seconds and 655 seconds respectively. Although 3B and 3B′ show a diminished thermal stability when compared to 3A and 3A′ respectively, they show a much faster recovery to triple helices.
Even though the refolding half times increase with the numbers of mutations for both heterotrimeric and homotrimeric, they are still comparatively faster than the reported values for mutated homotrimeric. Refolding half times of 1620 seconds were reported for non-mutated homotrimeric that increased to 2280 seconds upon Gly to Ala mutation at position 901 in a peptide mimicking sequence from residues 892-909 from the α1 chain of human type I collagen. These peptides were 30 amino acids long and had a (POG)₄ flanking region only on the C-terminal. Homotrimeric reported in my study are 39 amino acids long and have (POG)₃ flanking regions on both the terminals. Therefore, a comparatively faster folding is expected for homotrimeric when compared to the reported literature. It has been reported that (POG)₁₀ homotrimer has a refolding half time of 360 seconds. Refolding half time for both 3A and 3B fall in a similar range. Heterotrimeric, however, folded much faster than both the homotrimeric reported in this study and in the literature. I believe that the (DOG)₅, (PKG)₅ and (POG)₅ flanking regions on both the sides in my system help lead to a comparatively faster refolding. I believe that the electrostatic driving force for assembly in my system is responsible for this dramatic acceleration in folding, perhaps because of the long-range nature of these interactions as compared to hydrogen bonding. An additional factor that may be responsible for this difference is the higher percentage of amino rather than imino acids in my system. In summary, all heterotrimeric were found to reach 85% folding in approximately 50 min.

The non-mutated heterotrimer refolded fastest, followed by single mutation, double mutation and triple mutation heterotrimeric. Combining the results from the unfolding and refolding experiments, I can conclude that for my system the heterotrimer which folds the slowest is the least thermally stable and the heterotrimer which folds the fastest is the
most thermally stable. Both the refolding rates and thermal stability are affected by substituting Gly with a bulkier amino acid and the magnitude of these effects is entirely dependent upon the site and frequency of the Gly mutation. For future studies, I plan to explore the effects of shorter flanking regions and longer guest regions so as to accurately estimate the factors responsible for the known phenotypes of the mutations found in type I collagen, at position 247 in particular and any other position in general using the same design principle as described here.

Figure 5-9. Refolding analysis for various heterotrimers and homotrimers showed that the refolding half times increase as the number of mutations increase. Refolding spectra for the (a) heterotrimers and (d) homotrimers plotted as MRE versus time. Refolding spectra for the (b) heterotrimers and (e) homotrimers plotted as fraction folded versus time. (c) Refolding spectra in (b) zoomed in to clearly show $t_{1/2}$ for all the heterotrimers. (f) Refolding spectra in (e) zoomed in to clearly show $t_{1/2}$ for all the homotrimers.
5.3 Conclusions.

Substitution of Gly by bulky amino acid residues results in a multitude of disorders in natural homotrimeric and heterotrimeric collagens and leads to diseases such as OI, Alport syndrome and Ehlers-Danlos syndrome to name a few.\textsuperscript{29-32} Designing systems that can accurately mimic these mutations in synthetic analogs can provide us with a means to understand the structural and biological changes observed in mutant collagens. Various parameters including mutation site, chain type, amino acid residue replacing Gly and neighboring sequence can affect the phenotype of a disorder\textsuperscript{29-32,62,63} and these parameters need to be studied in their native form to estimate the true effect of these mutations. Synthetic homotrimers can have the mutation in either none or all three chains and thus are limited to the study of only naturally occurring homotrimers in their native forms. ABC heterotrimers can be designed to assemble three completely different sequences in the center of the triple helix and give us the ability to study naturally occurring heterotrimers in their native forms. Three different sequences can mimic a natural ABC heterotrimer while two similar sequences and a third different sequence can mimic a natural AAB heterotrimer. In the study reported here, I mimic the sequence 242-250 of type I collagen, with a Gly to Ser mutation at position 247 in either of the \(\alpha1\) or \(\alpha2\) chains. I am able to differentiate between four triple helices that vary only in the frequency of Gly mutations at a particular position. The ease of preparation of heterotrimers and this degree of resolution in terms of separating single mutations can have major implications in our understanding of the mutations in natural collagens that lead to various connective tissue disorders in general and OI in particular.
Chapter 6

Collagen Heterotrimers: Peptide Model for Proline

Hydroxylation and Fluorination

6.1 Introduction

Collagen is a major structural component of all connective tissues such as skin, ligament, cartilage, bone and tendon. Collagen is characterized by a ubiquitous X-Y-Gly repeating motif, where X and Y are generally Pro and Hyp respectively while Gly is present every third amino acid. It has a unique triple helical structure in which three left-handed PPII helices wind around one another to form a right-handed super helix.\textsuperscript{1,19-26} Non-natural amino acids have also been incorporated in the X and Y positions of collagen-like peptides and the most significant results were reported by Raines group with the synthesis and self-assembly of a hyperstable (Pro-Flp-Gly)$_{10}$ (Flp = 4($R$)-fluoroproline = Fp) triple helix,\textsuperscript{61} with thermal stability higher than the (Pro-Hyp-Gly)$_{10}$ and (Pro-Pro-Gly)$_{10}$ homotrimers. Initially, the mechanism of stabilization by Hyp was believed to be through water-mediated hydrogen bonding.\textsuperscript{26,27,56-59} Flp, with a stronger electron withdrawing fluorine group instead of a hydroxyl group, has a lower tendency to form hydrogen bonds. Yet, the incorporation of Flp instead of Hyp results in an increase in thermal stability.\textsuperscript{61} Therefore, it is believed that the inductive effect of fluorine in Flp at Y position and its subsequent stabilization of the exo ring pucker plays a role in the stabilization of triple helix.\textsuperscript{61,90,100,102,111,112} In the crystal structures of collagen-like peptides,\textsuperscript{26,27} there is a clear preference of endo puckering for the imino acids in X
positions and exo puckering of imino acids in Y positions, as shown in Figure 2-6. Both Hyp and Flp have 4(R)-substituents in the side chain, which have the appropriate stereochemistry to stabilize the exo pucker of the ring. In the case of 4(S)-hydroxyproline or 4(S)-fluoropropoline, the stereochemistry of the side chain substituent is opposite to Hyp and Flp, resulting in stabilization of the endo pucker. This observation is also supported by the fact that 4(S)-fluoropropoline, which adapts an endo pucker, has a stabilizing effect on triple helix at X position and not in the Y position.\textsuperscript{37,90,100-102} As both Hyp and Flp adopt the exo pucker of the ring, their presence at the X position in a collagen-like peptide leads to a lower thermal stability and melting temperature of the triple helix. Furthermore, (Pro-Flp-Gly)\textsubscript{10} triple helix has a higher thermal stability than the (Pro-Hyp-Gly)\textsubscript{10} triple helix,\textsuperscript{61} even though Flp has lower propensity to form hydrogen bonds when compared to Hyp. This creates significant doubts on the original stabilization theory based on water mediated hydrogen bonding. Instead, it is likely that the inductive effect plays a major role in the stabilization of triple helix in the case of Flp.\textsuperscript{61,90,100,102,111,112} The stabilization from Hyp is also likely inductive\textsuperscript{61,90,100,102,111,112} in nature, though water-bridge mediated hydrogen bonding\textsuperscript{26,27,56-59} can not be completely excluded. The amino acid propensity for both X and Y positions has been measured by Brodsky and coworkers,\textsuperscript{43,46,48} explained in detail in the introduction chapter. Among other peptides, they synthesized a series of host-guest peptides with X-Y-Gly guest in a Pro-Hyp-Gly host framework. They compared the thermal stability of Pro-Pro-Gly, Pro-Hyp-Gly and Pro-Flp-Gly triplets in a Pro-Hyp-Gly host and reported the thermal stability in the following order: Pro-Hyp-Gly > Pro-Pro-Gly > Pro-Flp-Gly.\textsuperscript{127} However, this is in contrast with the thermal stability of the homotrimers with repeating motifs and the
thermal stability is in the following order: (Pro-Flp-Gly)$_{10}$ > (Pro-Hyp-Gly)$_{10}$ > (Pro-Pro-Gly)$_{10}$. Additionally, in a different series of host guest peptides developed by Fields and coworkers, the authors reported that Pro-Flp-Gly triplet resulted in a more stable triple helix as compared to Pro-Hyp-Gly triplet, in direct contrast to the results obtained by Brodsky and coworkers. Furthermore, there are no reports on the effects of incorporation of Flp as a guest amino acid in more than one triplet, or in different chains to understand the contributions of Flp to stability of a triple helix.

In this chapter, I describe the design, synthesis, self-assembly and characterization of a series of heterotrimerers to understand the contribution of Pro hydroxylation and fluorination to the stability of a triple helix in a chain dependent manner. My design is based upon the ABC type heterotrimer reported earlier which is composed of three unique 30 amino acid peptides: neutral (POG)$_{10}$, positive (PKG)$_{10}$ and negative (DOG)$_{10}$. This (DOG)$_{10}$*(PKG)$_{10}$*(POG)$_{10}$ was found to have a thermal stability comparable to the (POG)$_{10}$ homotrimer which is known to be the most stable peptide motif using natural amino acids. Out of the three component chains of the heterotrimer, the negative and positive chains were preserved and the neutral chain was varied from (PPG)$_{10}$ to (POG)$_{10}$ and (PFpG)$_{10}$. The new peptides reported in this chapter, (PPG)$_{10}$, (POG)$_{10}$ and (PFpG)$_{10}$ were not protected at N and C terminals, unlike all the other peptides reported in this and other chapters. Thus, the heterotrimmers reported in this chapter are slightly different in the sense that one out of three chains has free N and C terminals. The reason for this change in termination is the fact that N and C protected (PFpG)$_{10}$ peptide in not soluble in water and precipitates out at the concentration desirable for my experiments. As my results are based upon comparison of different heterotrimmers, I decided to synthesize (PPG)$_{10}$ and
(POG)\textsubscript{10} peptides with free N and C terminals and incorporated them in an ABC collagen heterotrimer. This results in consistent N and C termination for all triple helices. The peptides used for this study are abbreviated in Table 6-1.

<table>
<thead>
<tr>
<th>#</th>
<th>Sequence</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(DOG)\textsubscript{10}</td>
<td>D</td>
</tr>
<tr>
<td>2</td>
<td>(PKG)\textsubscript{10}</td>
<td>K</td>
</tr>
<tr>
<td>3</td>
<td>(PPG)\textsubscript{10}</td>
<td>P</td>
</tr>
<tr>
<td>4</td>
<td>(POG)\textsubscript{10}</td>
<td>O</td>
</tr>
<tr>
<td>5</td>
<td>(PFpG)\textsubscript{10}</td>
<td>Fp</td>
</tr>
</tbody>
</table>

N and C termini of peptides 1 and 2 are acetylated and amidated respectively. All other peptides have free N and C termini.

The triple helices are abbreviated, for example, D•K•O for triple helices composed of D, K and O peptides. The abbreviations and composition of various triple helices used in this study are shown in Table 6-2.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>D•K•P</td>
<td>D, K and P</td>
</tr>
<tr>
<td>D•K•O</td>
<td>D, K and O</td>
</tr>
<tr>
<td>D•K•Fp</td>
<td>D, K and Fp</td>
</tr>
<tr>
<td>P•P•P</td>
<td>P, P and P</td>
</tr>
<tr>
<td>O•O•O</td>
<td>O, O and O</td>
</tr>
<tr>
<td>Fp•Fp•Fp</td>
<td>Fp, Fp and Fp</td>
</tr>
</tbody>
</table>
6.2 Results and Discussion

6.2.1 Molecular and Experimental Design. I designed peptides that evaluate the contribution of Hyp and Flp to the stability of a triple helix by substituting Pro in the P chain of D•K•P heterotrimer, as depicted schematically in Figure 6-1. As two of the component chains of the triple helix are same, the effects on stability of triple helix come only from the substitutions in the third chain. The difference between D•K•O and D•K•P is the substitution by Hyp for Pro in one of the chains. Likewise, the difference between D•K•Fp and D•K•P is the substitution by Flp for the Pro in one of the chains. The thermal stability is compared for the three heterotrimers and are reported in the remainder of the chapter. Additionally, the formation of P•P•P, O•O•O and Fp•Fp•Fp homotrimers was also assessed.

\[
\begin{align*}
(DOG)_{10} & \quad (DOG)_{10} & \quad (DOG)_{10} \\
(PKG)_{10} & \xrightarrow{P \to O} (PKG)_{10} & \xrightarrow{O \to Fp} (PKG)_{10} \\
(PPG)_{10} & \quad (POG)_{10} & \quad (PFpG)_{10} \\
D\cdot K\cdot P & \quad D\cdot K\cdot O & \quad D\cdot K\cdot Fp
\end{align*}
\]

Figure 6-1. A schematic representation of the molecular design. Two component chains of all the heterotrimers are kept constant and the third chain is varied with Pro substituted by either Hyp or Flp.

6.2.2 Homotrimers. P•P•P, O•O•O and Fp•Fp•Fp homotrimers were analyzed at a concentration of 0.2 mM in 10 mM PO₄ buffer at pH 7. P•P•P homotrimers show a weak melting transition at 29.5 °C, as shown in Figure 6-2. As expected, the P to O substitution led to an increase in thermal stability and a melting temperature of 60.5 °C was observed for O•O•O homotrimers. As mentioned before, the peptides used in this study have free
N and C termini, which leads to a destabilization of the triple helix as compared to the protected counterparts. The melting temperature for an O-O-O homotrimer with the protected N and C termini has been reported to be 67.5 °C, as mentioned in the earlier chapters. The drop in thermal stability of around 7 °C from protected termini to free termini can be accounted for by the repulsion between same charges on both C and N termini. When Fp residues replaced the P residues, a further increase in thermal stability was observed, with Fp-Fp-Fp homotrimers showing a melting temperature of 85 °C. This value is slightly lower than the thermal stability reported by Raines group for the same peptide with a temperature of 92 °C.\textsuperscript{61} The reason for this anomaly is likely the fact that I use a different heating rate from the one reported by the Raines group\textsuperscript{61} and the heating rate has been shown to affect the observed melting temperature. It has been reported that faster heating rates lead to higher melting temperature estimation. Raines group reported the increase in temperature by increments of 3 °C with 5-minute equilibration, which results in 36 °C/hour, as compared to a heating rate of 10 °C/hour used in my experiments. Thus, I observe an increase of 31 °C in thermal stability for a P to O substitution in all three chains, and an increase of 45.5 °C for a P to Fp substitution in all the three chains. The thermal stability of the homotrimers is summarized in Figure 6-2 and Table 6-3. Next, I analyze the effect of either P to O or P to Fp substitution on the stability of a triple helix in a chain dependent manner.
Figure 6-2. Circular dichroism for P•P•P, O•O•O and Fp•Fp•Fp homotrimers. (a) Thermal analysis shows co-operative unfolding for all the homotrimers. (b) The first derivative of unfolding versus temperature shows that the thermal stability increases by 31 °C when P is replaced by O and by 45.5 °C when P is replaced by Fp.

6.2.3 ABC Heterotrimers. D, K and P peptides were mixed in 1:1:1 ratio to form the D•K•P heterotrimer. Similarly, D, K and O peptides and D, K and Fp peptides were mixed in 1:1:1 ratio to form D•K•O and D•K•Fp heterotrimers respectively. The heterotrimer formation was confirmed by CD analysis, in the same way as explained for other heterotrimers in the earlier chapter. Mixing a neutral, positive and negative collagen-like peptide can lead to the formation of seven possible triple helices: three AAA homotrimers, six AAB heterotrimers, and one ABC heterotrimer. The formation of these homotrimers and AAB heterotrimers was assessed separately by mixing the corresponding peptides.
Table 6-3. Thermal stability of triple helices studied

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Melting Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D•K•P</td>
<td>54</td>
</tr>
<tr>
<td>D•K•O</td>
<td>61.5</td>
</tr>
<tr>
<td>D•K•Fp</td>
<td>65.5</td>
</tr>
<tr>
<td>P•P•P</td>
<td>29.5</td>
</tr>
<tr>
<td>O•O•O</td>
<td>60.5</td>
</tr>
<tr>
<td>Fp•Fp•Fp</td>
<td>85</td>
</tr>
</tbody>
</table>

D•K•P heterotrimer showed a thermal stability of 54 °C, as shown in Figure 6-3. When the Pro residues in the P peptide were replaced by Hyp residue to form D•K•O heterotrimer, the thermal stability increased by 7.5 °C to 61 °C. Instead, when the Pro residues in the P peptide were replaced by Flp to form D•K•Fp heterotrimer, the increase in thermal stability was 11.5 °C, as shown in Figure 6-3. Furthermore, for the analysis of homotrimers, it was observed that P•P•P homotrimer showed a weak transition at 29.5 °C. When the Pro residues in all the three P peptide were replaced by Hyp residue to form O•O•O homotrimer, the thermal stability increases by 31 °C. Therefore, substituting Pro by Hyp in a single peptide chain leads to an increase in thermal stability of 7.5 °C as observed for the heterotrimer, while the same substitution in all three chains leads to an increase of 31 °C as observed in the homotrimers. Thus, the contribution of Pro to Hyp substitutions in one chain to the increase in thermal stability is less than one third of the contribution of substitution in all three chains. This result suggests that there is a non-linear increase in thermal stability when Hyp replaces Pro in a chain-dependent manner.

Additionally, when the Pro residues in all the three P peptide chains of P•P•P homotrimer were replaced by Flp residue to form Fp•Fp•Fp homotrimer, the thermal
stability increased by 55.5 °C. As mentioned previously, the increase in thermal stability for Pro to Flp substitution in one chain is only 11.5 °C, which is less than one third of the increase for substitutions in all three chains. Therefore, even for Pro to Flp substitutions, there is a non-linear increase in thermal stability when Flp replaces Pro in a chain-dependent manner. Experiments to understand the effect of substitutions in two out of three chains on the stability of the triple helix need to be undertaken to completely understand the role of individual chains to the stability of a triple helix.

![Graphs showing unfolding analysis](image)

**Figure 6-3.** Unfolding analysis by circular dichroism for D•K•P, D•K•O and D•K•Fp ABC heterotrimers. (a) MRE versus temperature. (b) The first derivative of unfolding versus temperature shows a melting temperature of 54 °C, 61.5 °C and 65.5 °C for D•K•P, D•K•O and D•K•Fp heterotrimers.

### 6.3 Conclusions.

The role of Pro hydroxylation to the stability of triple helix is a highly debated topic. The mechanism of stabilization of a triple helix by Hyp was believed to be through water-mediated hydrogen bonding, as observed in the crystal structures of collagen like
peptides where hydroxyl group in the Hyp acts as an anchoring point for inter- and intramolecular water bridges. However, this mechanism of stability was questioned when Flp containing collagen-like peptides showed a much higher thermal stability than the corresponding Hyp containing peptides. Flp, with a fluorine group instead of a hydroxyl group, has a lower tendency to form hydrogen bonds and it is likely that the inductive effects play a major role in the stabilization of triple helix in Flp containing peptides. Furthermore, contrasting results were reported in literature\textsuperscript{127,128} about the contribution of a Pro-Flp-Gly triplet in a host-guest peptide design. In this study, I have designed a series of peptides which estimate the contribution of proline hydroxylation or fluorination to the stability of a triple helix by incorporating Hyp or Flp specifically in one of the three component chains of the helix. I successfully report that Flp peptides are more stable than the Hyp peptides which in turn are more stable than the Pro peptides, in agreement with the known results. I, for the first time, show the contributions of these substitutions to the stability of a collagen-triple helix in chain dependent manner and conclude that there is a non-linear increase in thermal stability as the increase in stability is less than 1/3\textsuperscript{rd} of the stability observed for the substitutions in all three chains.
Chapter 7

Conclusions and Future Directions

7.1 Conclusions

Collagen is an essential structural component of all connective tissues, including skin, tendon, ligaments and bone. It is a fibrous protein and has a unique triple helical structure in which three left-handed PII helices wind around each other to form a right-handed super helix. The triple helical structure is stabilized by an extensive network of hydrogen bonding between CO\(_{(X)}\) – NH\(_{(Gly)}\) residues in a X – Y – Gly repeating motif. X and Y are generally Pro and Hyp respectively and substitution of these imino acids by amino acids generally leads to a decrease in stability of the triple helix. Hyp is suggested to form water bridges with unused carbonyl groups in the helix and has inductive effects which lead to enhanced stability of the helix. Additionally, one amino acid stagger between the three chains allows for the presence of Gly in the center of a tightly packed triple helix, a crucial requirement for the formation of the helix as Gly is the only amino acid without a side chain and able to fit it in the interior of a crowded triple helix. Naturally occurring collagens form triple helices composed of all identical (AAA, homotrimer), two different (AAB, heterotrimer), or all different (ABC, heterotrimer) chains. Type I collagen, the most abundant of 28 types of collagen, is an AAB heterotrimer assembled from two identical α1 and one α2 chain. Gly mutations in collagen lead to variety of connective tissue disorders. Mutations in homotrimeric AAA collagens (type II, III, VII and X) lead to chondrodysplasia, Ehlers-Danlos syndrome, dystrophic epidermolysis bullosa and chondrocyte hyperthrophy
respectively.\textsuperscript{28-32} Mutations in heterotrimeric AAB collagens (type I, IV and VIII) result in osteogenesis imperfecta, Alport syndrome and corneal endothelial dystrophy respectively, while mutations in heterotrimeric ABC collagen (type VI) result in Bethlem myopathy.\textsuperscript{28-32} Osteogenesis imperfecta (OI), the most studied collagen disease, is primarily caused by missense mutations in either the $\alpha1$ or $\alpha2$ chains of type I collagen, which lead to the substitution of Gly in the ubiquitous X-Y-Gly repeat by bulky amino acids such as Arg, Asp, Glu, Cys, Ser, Ala or Val.\textsuperscript{30,31} A significant number of naturally occurring collagens are heterotrimer, however, the majority of studies on synthetic collagen-like peptides have been performed on homotrimers.\textsuperscript{33-51} Recently, some results have become available for collagen-like heterotrimers assembled by cystine-knot strategy, where cystine disulfide bonds covalently capture three different chains leading to the formation of a heterotrimer.\textsuperscript{41,78-85,88,89} Furthermore, only three studies have been reported which elaborate the formation of collagen-like heterotrimers using non-covalent interactions.\textsuperscript{86,87,90} However, this methodology always leads to the formation of a complex mixture of all possible homotrimers and heterotrimers. Furthermore, the heterotrimers formed using this approach are not stable over time and a purified heterotrimer from this complex mixture always converts back to all possible triple helices, thereby limiting the utility of these systems.

For my graduate research, I designed, synthesized and characterized a series of peptides that self-assemble to form collagen-like heterotrimers directed through electrostatic interactions. The thesis is divided into three main sections: design principles (chapter 3), optimization of the heterotrimers (chapter 4) and applications of the synthesized heterotrimers (chapter 5 and 6). For the design principles in chapter 3, I
synthesized a set of \((POG)_n\) based peptides, with a general sequence \((POG)_5\cdot\cdot\cdot n(EOGPOG)_n(POG)_5\cdot\cdot\cdot n(POG)_5\cdot\cdot\cdot n\), where \(n\) represents the charge of each polypeptide chain, for \(n = 3, 5\); and \((EOG)_n\) and \((PRG)_n\) for decacharged peptides. Oppositely charged peptides were mixed with each other and their ability to form collagen-like heterotrimers was assessed. I find that the stability of a heterotrimer cannot be anticipated by individual amino acid propensities for triple helix formation as reported for homotrimers. Instead, I find that amino acids which are found to substantially destabilize the triple helix, such as Glu, can form triple helices of high stability when paired against oppositely charged amino acids such as Arg or, more surprisingly, when paired against other neutral amino acids with high stability such as Hyp. This results in the stability of the heterotrimERIC triple helix to be almost indistinguishable from those containing only amino acids with high triple helix stability. In the most dramatic example, I show that \((EOG)_n\) and \((PRG)_n\) peptides, which are unable to form a triple helix in isolation, form high quality triple helices when mixed together. Thus, my results indicate that intermolecular electrostatic interactions can be utilized to direct heterotrimer formation. Additionally, amino acids with poor collagen triple helical propensity can be "rescued" in heterotrimers containing amino/imino acids with known high triple helical propensity. Furthermore, this allows for the introduction of a greater chemical diversity in collagen-like triple helices than what would be allowed in a homotrimer. These results lay the groundwork for a better understanding of thermal stability of collagen-like heterotrimers and analyzing the contributions of different charge pairs to the stability of the helix. The analysis on contributions of charge pairs forms the emphasis of chapter 3.
Chapter 4 elaborated on the optimization of thermal stability of heterotrimers. As the heterotrimer assembly is directed by electrostatic interactions, the contributions of different charge pairs among the naturally occurring amino acids were evaluated. There are two amino acids with acidic side chains, Glu and Asp and two amino acids with basic side chains, Lys and Arg at neutral pH. I evaluated the contributions of Glu – Arg charge pairs in chapter 3, and observed that the decacharged peptides resulted in the formation of both AAB and ABC heterotrimers, and did not form any homotrimers. Thus, I employed only decacharged peptides to evaluate the contributions of charge pairs. (DOG)_{10}(D) and (PKG)_{10}(K) peptides were synthesized, and were studied in combination with the (EOG)_{10}(E), (PRG)_{10}(R) and (POG)_{10}(O) peptides used in the chapter 3. The self-assembly and thermal stability of all four possible ABC heterotrimers, including E•R•O, E•K•O, D•R•O and D•K•O was evaluated. I report the synthesis of a surprisingly stable D•K•O heterotrimer, composed of (DOG)_{10}, (PKG)_{10} and (POG)_{10} chains. This heterotrimer has a stability comparable to O•O•O homotrimer, composed of three (POG)_{10} chains, even though Asp and Lys amino acids occur twenty times in the heterotrimer and have been shown to significantly destabilize the triple helix as compared to Pro and Hyp imino acid respectively. Thus, I show that the thermal stability of a heterotrimer cannot be deduced from the amino acid propensity for collagen-like heterotrimers. Furthermore, I demonstrated that mixing three peptides with individually negative, positive and neutral charge is the general mechanism for the formation of specific ABC collagen-like heterotrimers. The ABC heterotrimer is the most thermodynamically stable species among a total of seven possibilities, including the individual homotrimers and the AAB heterotrimers. Additionally, the ability to
synthesize such high stability collagen triple helices without the use of any covalent cross-links and using solely the naturally occurring amino acids will have great impact on our ability to probe triple helical structure and stability. These high stability D·K·O heterotrimers provide us with useful model systems to mimic connective tissue disorders in naturally occurring heterotrimeric collagen, including type I (AAB heterotrimer) collagen. D·K·O heterotrimers can also be used to understand the contributions of Flnp to the stability of a collagen triple helix in a chain dependent manner. These two topics are emphasized in chapters 5 and 6 respectively.

In chapter 5, I report the design of a peptide model for osteogenesis imperfecta (OI), a hereditary disorder also known as brittle bone disease and observed in type I collagen. Type I collagen is an AAB heterotrimer composed of two identical and one different chains. Gly mutations to a bulky amino acid in any of these chains lead to OI of varying severity. My design to mimic the mutations leading to OI utilizes an electrostatic recognition motif in three chains that can force the interaction of any three peptides, including AAA (all same), AAB (two same and one different), or ABC (all different) triple helices. I demonstrate that as long as POG, PKG and DOG motifs are present in the flanking regions of three peptides to guide the formation of an ABC heterotrimer, any natural sequence can be incorporated in the center of the triple helix. Furthermore, as the component peptides are tailor-made and different from each other, the effects of Gly mutations can be assessed in zero, one, two, or all three chains. I, for the first time, report collagen mutants, containing one or two Gly mutations, with the structure relevant to the native forms of OI. I successfully estimate the effect of Gly to Ser mutations at position 247 in both the chains of type I collagen, and my thermal unfolding and refolding results
are in agreement with the known phenotypes of these mutations. Additionally, I am able to differentiate between four triple helices that differ from each other in the frequency of Gly mutations at a particular position. Both the thermal unfolding and refolding effects are affected by substituting Gly with bulkier amino acids and the magnitude of these effects depends on the site and frequency of mutations. The heterotrimers which fold the fastest are the most thermally stable and the heterotrimers which fold the slowest are the least thermally stable. Thus, the ease of preparation of heterotrimers, coupled with our ability to separate single mutations, provides us with a tool to understand mutations in natural collagens that lead to various connective tissue disorders in general and OI in particular.

In chapter 6, I report another application of D•K•O based heterotrimers to understand the effect of Pro hydroxylation and fluorination to the stability of a collagen triple helix, in a chain dependent manner. The thermal stability of collagen-like homotrimers with repeating Flp, Hyp or Pro motifs is in the following order: (Pro-Flp-Gly)$_{10}$ > (Pro-Hyp-Gly)$_{10}$ > (Pro-Pro-Gly)$_{10}$. However, conflicting results have been reported for the thermal stability of Pro-Flp-Gly, Pro-Hyp-Gly and Pro-Pro-Gly triplets in a Pro-Hyp-Gly host and the thermal stability is in the following order: Pro-Hyp-Gly > Pro-Pro-Gly > Pro-Flp-Gly or Pro-Flp-Gly > Pro-Hyp-Gly. I report a model peptide system to evaluate the contributions of Hyp and Flp in a chain dependent manner. I compare the thermal stability of three heterotrimers with two common chains in their composition but a different third chain. The common chains are (DOG)$_{10}$ and (PKG)$_{10}$ for all the heterotrimers. However, the third chain in the heterotrimers is mutated from (PPG)$_{10}$ to (POG)$_{10}$ and (PFpG)$_{10}$. This provides us with a tool to directly compare the effects of
hydroxylation in one chain by comparing \((PPG)_{10}\) containing heterotrimer with \((POG)_{10}\) containing heterotrimer, or fluorination in one chain by comparing \((PPG)_{10}\) containing heterotrimer with \((PFpG)_{10}\) containing heterotrimer. I observe that the contribution of one chain to the increase in thermal stability is less than one third of the contribution of substitution in all the chains. The results suggest that there is a non-linear increase in thermal stability when Hyp or Flp replaces Pro in a chain-dependent manner.

7.2 Future Directions

The future work emanates from our current understanding of this system and will focus primarily on the elaborate characterization of \(D\cdot K\cdot O\) heterotrimers using NMR and X-ray crystallography, and further exploration of the peptide model for connective tissue disorders.

The NMR experiments are being done by Jorge Fallas, another graduate student in the library. The preliminary results (unpublished) confirm the presence of \(D\cdot K\cdot O\) heterotrimers, and efforts are underway to determine the register of this triple helix.

As previously mentioned, Gly mutations to bulky amino acids in type I collagen lead to varying phenotypes of OI. OI phenotypes vary from mild to lethal forms depending on a variety of factors. These include the chain type mutated, the proximity to the C-terminus, the neighboring sequence of the mutated site, and the amino acid residue which substitutes for Gly. It is observed that the amino acid substitutions that are lethal when present in the \(\alpha 1\) chain may not be lethal when present in \(\alpha 2\) chain. Furthermore, the mutations present near the C-terminus are known to be more lethal than the mutations present near the N-terminus, which correlates with the C to N terminus folding of the
triple helix.\textsuperscript{62,63} In the current work, I have only reported the peptide model to understand the effect of mutations in a chain dependent manner, in either $\alpha_1$ or $\alpha_2$ chains of type I collagen. The future work will focus on the other factors that affect the phenotype of OI. A similar peptide model can be developed to analyze the Gly substitution by bulky amino acids such as Arg, Asp, Glu, Cys, Ser, Ala or Val. The current study concentrates on Gly to Ser mutation. A series of peptides with Gly to Z, where $Z = \text{Arg, Asp, Glu, Cys, Ala or Val}$, can be incorporated in a ABC collagen heterotrimer and the effect of thermal unfolding and refolding can be observed. Another important factor in determining the phenotype is the neighboring sequences of the mutated Gly residues. In the current work, I analyzed the sequences from 242 – 250 in both the chains of type I collagen, as Gly mutations at position 247 are observed both in $\alpha_1$ and $\alpha_2$ chains. These mutations are present all over the length of type I collagen, and peptide models mimicking different regions on type I collagen can be synthesized. The results from this study could result in the generation of a stability map based upon the known Gly mutations over the full sequence of type I collagen. Furthermore, the same design principle can be used to study different functional motifs, like integrin binding domains and enzyme cleavage sites on type I collagen in their native forms. Additionally, a peptide model can be developed with flanking regions on only the C terminal, as opposed to the current system with flanking regions on both the termini. This will shed useful information on the ability of small collagen-like heterotrimers to fold from C to N terminal, similar to what is observed for natural collagens.

All the current work and the mentioned proposed future work focus on OI, a disorder in type I collagen. However, there are around than 28 types of collagen, both
homotrimeric and heterotrimeric, and Gly mutations and other mutations at X or Y positions lead to a variety of connective tissue disorders. Mutations in homotrimeric AAA collagens (type II, III, VII and X) lead to chondrodysplasia, Ehlers-Danlos syndrome, dystrophic epidermolysis bullosa and chondrocyte hypertrophy respectively.\textsuperscript{29-32} Mutations in heterotrimeric AAB collagens (type I, IV and VIII) result in osteogenesis imperfecta, Alport syndrome and corneal endothelial dystrophy respectively, while mutations in heterotrimeric ABC collagen (type VI) result in Bethlem myopathy.\textsuperscript{29-32} Peptide model following the same basic design principles can be used to mimic the mutations responsible for connective tissue disorders in any type of collagen.
Chapter 8

Experimental

8.1 Peptide Synthesis and Purification. All the peptides were synthesized on an Advanced ChemTech 396 multipeptide automated synthesizer using Fmoc solid-phase chemistry based on a 0.15mmole scale. A MBHA-rink amide resin was used. Amino acid couplings were 45 min long with the following amount of reagents: 4 equivalents of amino acids, 4 equivalents of 1-Hydroxybenzotriazole hydrate (HoBt), 4 equivalents of O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate (HBTU), 6 equivalents of N,N-diisopropyl-ethyl-amine (DiEA) with N,N-dimethylformamide (DMF) as a solvent. All Pro residues were double-coupled. Fmoc was deprotected with two 7 min treatments of 25% (by volume) piperidine in DMF. After completion of the peptide synthesis, the peptide was acetylated in the presence of acetic anhydride and DiEA at the molar ratio of 8.3:1 in dichloromethane (DCM). Cleavage of the peptide was accomplished by treatment of the resin with 20 ml of trifluoroacetic acid (TFA)/Triisopropylsilane/H2O (38:1:1 by volume) for three hours at room temperature for POG and En peptides and with 20 ml of TFA/Ethanedithiol/H2O/Triisopropylsilane (37.6:1:1:0.4 by volume) for three hours at room temperature for Rn peptides. This TFA solution was collected, followed by rinsing the resin twice with neat TFA. All washings were combined and rotoevaporated to a thick solution of approximately 5 ml. The peptide was triturated by addition of 50ml of cold diethyl ether. The precipitate was collected by centrifugation and the pellet was washed two times with cold ether. The pellet was then dried under vacuum overnight and redissolved in deionized water for purification by
reverse phase HPLC using a C-18 column with a linear gradient of acetonitrile and water containing 0.05% TFA.

8.2 Circular Dichroism (CD) Spectroscopy. CD measurements were performed with a Jasco J-810 spectropolarimeter, equipped with a Peltier temperature control system, using quartz cells with a path length of 0.1 cm. Thermal unfolding curves were obtained by monitoring the decrease in ellipticity in a desired temperature range at a wavelength where the CD spectra shows a positive maximum. The temperature ranged from 5°C to 95°C at a heating rate of 10°C/hour depending on the peptide mixture studied. Fraction folded was calculated for the melting curves according to the equation:

\[ F_T = \frac{[\theta]_T - [\theta]_{unfolded}}{[\theta]_{folded} - [\theta]_{unfolded}} \]  

(1)

where \( F_T \) is fraction folded at temperature T, \([\theta]_T\) is the Molar Residue Ellipticity (MRE) at temperature T, \([\theta]_{folded}\) and \([\theta]_{unfolded}\) are the MRE of the folded and unfolded forms respectively calculated as follows \([\theta] = (\theta * m) / (c * l * n_e)\), where \( \theta \) is ellipticity in mdeg, \( m \) is molecular weight in g/mol, \( c \) is concentration in mg/ml, \( l \) is pathlength of the cuvette in mm and \( n_e \) is the number of amino acid residues in the peptide.

Peptide solutions were mixed in desired ratios in such a way that the final total peptide concentration was 0.2mM, and a neutral pH was maintained by using 10mM phosphate buffer solution. Unfolding studies were performed with and without preheating. For preheating studies, peptides were mixed in desired ratios, heated to 85°C and incubated for 15 min. The peptide solution was then slowly cooled to 25°C at a rate of 1°C/min and then incubated overnight at room temperature before performing the unfolding studies. For non-preheating studies, peptides were mixed in the desired ratios and the unfolding
studies were performed immediately. The minimum of the derivative of the fraction folded plot indicates the steepest slope of the unfolding process and is used in this paper to indicate the melting temperature ($T_m$) under the conditions described above. This was calculated using the Jasco Spectra Manager software. All experiments were repeated at least once and $T_m$ values were found within $\pm 1 \, ^\circ C$ or less.

8.3 Refolding Studies. Refolding studies were performed using the Jasco J-810 spectropolarimeter, equipped with a Peltier temperature control system. Peptide mixtures were prepared as described above. The peptide mixture was heated to 85 $^\circ C$ in a glass vial and incubated for 15 min. It was then immediately transferred to a 0.1 cm cuvette which had been precooled to 5 $^\circ C$. Refolding data in the form of ellipticity was collected immediately afterwards with a data pitch of 5 seconds. The data was collected for 50 min. The dead time required to transfer the sample and start the data acquisition was approximately 15 seconds. Thus, for the data analysis, the zero time is actually 15 seconds after the peptide mixture is taken out of the heated glass vial. Refolding half time, the time at which half of the polypeptides have refolded into triple helices, was calculated by monitoring when the fraction folded equated half of the maximum equilibrated value as observed after overnight incubation on a repeat sample. The fraction folded was calculated using equation (1), where $[\theta]_T$ is the Mean Residue Ellipticity (MRE) at time $t$, $[\theta]_{folded}$ is the MRE of the maximally folded form measured at 5 $^\circ C$. $[\theta]_{unfolded}$ was measured by curve fitting of the refolding data by extrapolating to time zero. Repeated experiments indicate that the deviation in half-life is approximately $\pm 5$ seconds.
8.4 Molecular Modeling. Molecular model for the D*K*O heterotrimer was based on the atomic co-ordinates of (Pro-Hyp-Gly)$_4$(Leu-Hyp-Gly)$_4$(Pro-Hyp-Gly)$_3$ (http://www.rcsb.org/pdb, PDB Code: 2DRT). The model was build using HyperChem 1.0 and molecular mechanics calculations were done using Amber 99. In one chain, Asp substituted all Pro and Leu. For second chain, the central Leu was substituted by Pro, and all Hyp were substituted by Lys. Only the central Leu was substituted by Pro in the third chain so as to generate an ABC heterotrimer with (DOG)$_{10}$, (PKG)$_{10}$ and (POG)$_{10}$ chains. The backbone geometry was fixed and the amino acid side chain conformation was energy minimized using Polak-Ribiere conjugate gradient until the RMS gradient was 0.01 Kcal/Å*mol. The conformer with the minimum energy was used for further analysis and the distance between the nitrogen of K side chain and the closest oxygen of D side chain was measured.

8.5 HPLC and Mass Spectroscopy. All the peptides were purified by reverse phase HPLC using Microsorb C-18 column attached to a Varian Prostar system with a linear gradient of acetonitrile and water containing 0.5% TFA. Acetonitrile percent was increased at a rate of 3%/min. HPLC fractions were then analyzed by MALDI using Bruker autoflex II and a Prespotted Anchorchip PAC 384 HCCA (α-cyano-4-hydroxycinnamic acid).

The HPLC and MS for peptides utilized in chapter 3 are shown from Figure 8-1 to 8-7.
Figure 8-1. (a) HPLC and (b) MS data for E3. Expected: 2849.21[Na⁺], Observed: 2849.38[Na⁺].

Figure 8-2. (a) HPLC and (b) MS data for R3. Expected: 2860.42[Na⁺], Observed: 2860.55[Na⁺].
Figure 8-3. (a) HPLC and (b) MS data for E5. Expected: $2913.19[\text{Na}^+]$, Observed: $2913.39[\text{Na}^+]$.

Figure 8-4. (a) HPLC and (b) MS data for R5. Expected: $2946.53[\text{H}^+]$, Observed: $2946.75[\text{H}^+]$. 
Figure 8-5. (a) HPLC and (b) MS data for E10/E. Expected: 3073.14[Na⁺], Observed: 3072.60[Na⁺].

Figure 8-6. (a) HPLC and (b) MS data for R10/R. Expected: 3161.79[H⁺], Observed: 3161.97[H⁺].
Figure 8-7. (a) HPLC and (b) MS data for POG/O. Expected: 2753.24[Na⁺], Observed: 2753.44[Na⁺].

The HPLC and MS for peptides utilized in chapter 4 are shown from Figure 8-8 to 8-9.

Figure 8-8. (a) HPLC and (b) MALDI-TOF MS data for D. Expected: 2932.98[Na⁺], Observed: 2933.11[Na⁺].
Figure 8-9. (a) HPLC and (b) MALDI-TOF MS data for K. Expected: 2881.73[H⁺], Observed: 2881.79[H⁺].

The HPLC and MS for peptides utilized in chapter 5 are shown from Figure 8-10 to 8-17.

Figure 8-10. (a) HPLC and (b) MS data for peptide 1. Expected: 3513.57[K⁺], Observed: 3513.89[K⁺].
Figure 8-11. (a) HPLC and (b) MS data for peptide 2. Expected: 3543.58[K⁺], Observed: 3543.96[K⁺].

Figure 8-12. (a) HPLC and (b) MS data for peptide 3. Expected: 3626.09[H⁺], Observed: 3626.67[H⁺].
Figure 8-13. (a) HPLC and (b) MS data for peptide 4. Expected: 3656.1[H⁺], Observed: 3656.29[H⁺].

Figure 8-14. (a) HPLC and (b) MS data for peptide 5. Expected: 3434.59[Na⁺], Observed: 3434.54[Na⁺].
Figure 8-15. (a) HPLC and (b) MS data for peptide 6. Expected: 3464.6[Na⁺], Observed: 3464.59[Na⁺].

Figure 8-16. (a) HPLC and (b) MS data for peptide 7. Expected: 3592.35[H⁺], Observed: 3592.81[H⁺].
Figure 8-17. (a) HPLC and (b) MS data for peptide 8. Expected: 3660.3[K⁺], Observed: 3661.15[K⁺].
References


(23) Sasekaran, V. Acta Crystallogr. 1959, 12, 897-903.


(42) Persikov, A. V.; Pillitteri, R. J.; Amin, P.; Schwarze, U.; Byers, P. H.; Brodsky, B. Human Mut. 2004, 24, 330-337.


Appendix

The application of collagen-like heterotrimeric as a peptide model for OI was elaborated in chapter 5. The heterotrimeric helices prepared in that study were abbreviated in such a way that the first, second and third chains in the triple helix have neutral \((\text{POG})_5\), positive \((\text{PKG})_5\) and negative \((\text{DOG})_5\) flanking regions respectively. Furthermore, the peptides with \((\text{POG})_5\) and \((\text{PKG})_5\) flanking regions mimicked the \(\alpha_1\) chain of type I collagen, while the peptides with \((\text{DOG})_5\) flanking regions mimicked the \(\alpha_2\) chain of type I collagen. Thus, the heterotrimeres were abbreviated as \(A\cdot A\cdot B\), where the first and second peptides \((A)\) have \((\text{POG})_5\) and \((\text{PKG})_5\) flanking regions respectively, and the third peptide \((B)\) has \((\text{DOG})_5\) flanking region.

I synthesized a series of peptides, in addition to the ones reported in chapter 5, to understand the effects of mimicking \(\alpha_2\) chain of type I collagen in either of the chains. These include peptides with \((\text{DOG})_5\) flanking peptides mimicking the sequence 242-250 of the \(\alpha_1\) chain of type I collagen, and peptides with \((\text{PKG})_5\) flanking regions mimicking the sequence 242-250 of the \(\alpha_2\) chain of type I collagen, as abbreviated in Table A-2.

<table>
<thead>
<tr>
<th>#</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>((\text{DOG})_5) (PQGPGGPG) ((\text{DOG})_5)</td>
</tr>
<tr>
<td>10</td>
<td>((\text{DOG})_5) (PQGPGSPPG) ((\text{DOG})_5)</td>
</tr>
<tr>
<td>11</td>
<td>((\text{PKG})_5) (PVGAAAGATG) ((\text{PKG})_5)</td>
</tr>
<tr>
<td>12</td>
<td>((\text{PKG})_5) (PVGAASSATG) ((\text{PKG})_5)</td>
</tr>
</tbody>
</table>

N and C termini of all the peptides are acetylated and amidated respectively.
The heterotrimers were synthesized and abbreviated in the same fashion as described earlier and the first, second and third chains in the triple helix have neutral (POG)$_5$, positive (PKG)$_5$ and negative (DOG)$_5$ flanking regions respectively. In the current study, the $\alpha_2$ sequence is mimicked by either peptides with (POG)$_5$ or (PKG)$_5$ flanking regions. If the $\alpha_2$ mimic is present in (POG)$_5$ chain, then the triple helix is abbreviated as $B\bullet A\bullet A$. If the $\alpha_2$ mimic is present in (PKG)$_5$ chain, then the triple helix is abbreviated as $A\bullet B\bullet A$.

The peptides reported in chapter 5 and used in the analysis here are abbreviated in Table A-2.

<table>
<thead>
<tr>
<th>#</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(POG)$_5$ (PQGPGGPPG) (POG)$_5$</td>
</tr>
<tr>
<td>2</td>
<td>(POG)$_5$ (PQGPG$\delta$PPG) (POG)$_5$</td>
</tr>
<tr>
<td>3</td>
<td>(PKG)$_5$ (PQGPGGPPG) (PKG)$_5$</td>
</tr>
<tr>
<td>4</td>
<td>(PKG)$_5$ (PQGPG$\delta$PPG) (PKG)$_5$</td>
</tr>
</tbody>
</table>

$\alpha_1$ chain mimics ($A$)

| 5 | (POG)$_5$ (PVGAAGATG) (POG)$_5$       |
| 6 | (POG)$_5$ (PVGAASATG) (POG)$_5$       |
| 7 | (DOG)$_5$ (PVGAAGATG) (DOG)$_5$       |
| 8 | (DOG)$_5$ (PVGAASATG) (DOG)$_5$       |

$\alpha_2$ chain mimics ($B$)

N and C termini of all the peptides are acetylated and amidated respectively.

In a fashion similar to chapter 5, the heterotrimers prepared in this study are abbreviated in such a way that the first, second and third chains in the triple helix have neutral (POG)$_5$, positive (PKG)$_5$ and negative (DOG)$_5$ flanking regions respectively. For example, a non-mutated heterotrimers abbreviated as $A\bullet A\bullet B$ consists of polypeptides 1, 3 and 7, abbreviated as $A\bullet B\bullet A$ consists of polypeptides 1, 11 and 9, and abbreviated as
$B_{A}A_{A}$ consists of polypeptides 5, 3 and 9. All these triple helices have same overall composition. They differ in the presence of the $\alpha 2$ chain mimic ($B$) in chain one, chain two or chain three. Therefore, a mutated heterotrimer abbreviated as $A_{A}A_{A}B'$ consists of polypeptides 1, 3 and 8, abbreviated as $A_{A}B_{A}A$ consists of polypeptides 1, 12 and 9, and abbreviated as $B'_{A}A_{A}$ consists of polypeptides 6, 3 and 9. The abbreviations and composition of various triple helices used in this study are shown in Table A-3.

Table A-3. Composition of triple helices studied.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha 1\cdot\alpha 2\cdot\alpha 1$</td>
</tr>
<tr>
<td>$A_{A}B_{A}A$</td>
<td>1, 11 and 9</td>
</tr>
<tr>
<td>$A_{A}B_{A}A'$</td>
<td>1, 12 and 9</td>
</tr>
<tr>
<td>$A_{A}B'_{A}A'$</td>
<td>2, 11 and 10</td>
</tr>
<tr>
<td>$A_{A}B'_{A}A'$</td>
<td>2, 12 and 10</td>
</tr>
<tr>
<td></td>
<td>$\alpha 2\cdot\alpha 1\cdot\alpha 1$</td>
</tr>
<tr>
<td>$B_{A}A_{A}$</td>
<td>5, 3 and 9</td>
</tr>
<tr>
<td>$B'<em>{A}A</em>{A}$</td>
<td>6, 3 and 9</td>
</tr>
<tr>
<td>$B_{A}A'_{A}$</td>
<td>5, 4 and 10</td>
</tr>
<tr>
<td>$B'<em>{A}A'</em>{A}$</td>
<td>6, 4 and 10</td>
</tr>
</tbody>
</table>

Various peptides with (DOG)$_3$, (PKG)$_3$ or (POG)$_3$ flanking regions and the sequence 242-250 (Gly to Ser mutation at position 247) from type I collagen in the center were mixed together in a 1:1:1 ratio to form ABC heterotrimers with zero, one, two or three mutations. The formation of heterotrimer was analyzed in the same fashion as reported in chapter 5. The thermal unfolding curves for the heterotrimers with zero, one, two or three mutations are shown in Figure A-1 and A-2. Figure A-1 shows the data for the $\alpha 1\cdot\alpha 2\cdot\alpha 1$ register and Figure A-2 shows the data for the $\alpha 2\cdot\alpha 1\cdot\alpha 1$ register.
Figure A-1. Circular dichroism for heterotrimeric helices in α1·α2·α1 register. (a) Thermal analysis for $A\cdot B\cdot A$ shows that two transitions are observed in the non-preheated case, which merges to a single transition upon preheating. (b) The first derivative of unfolding versus temperature for the curve in (a). (c) Thermal unfolding for $A\cdot B\cdot A$ heterotrimer. (d) The first derivative of unfolding versus temperature for the curve in (c). (e) Thermal unfolding for $A\cdot B\cdot A'$ heterotrimer. (f) The first derivative of unfolding versus temperature for the curve in (e). (g) Thermal unfolding for $A\cdot B\cdot A'$ heterotrimer. (h) The first derivative of unfolding versus temperature for the curve in (g).
Figure A-2. Circular dichroism for heterotrimeric helices in α2•α1•α1 register. (a) Thermal analysis for $B•A•A$ shows that two transitions are observed in the non-preheated case, which merges to a single transition upon preheating. (b) The first derivative of unfolding versus temperature for the curve in (a). (c) Thermal unfolding for $B•A•A$ heterotrimer. (d) The first derivative of unfolding versus temperature for the curve in (c). (e) Thermal unfolding for $B•A•A'$ heterotrimer. (f) The first derivative of unfolding versus temperature for the curve in (e). (g) Thermal unfolding for $B•A•A'$ heterotrimer. (h) The first derivative of unfolding versus temperature for the curve in (g).
The thermal stability of the different triple helices, reported here and in Chapter 5, are abbreviated in Table A-4.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Melting Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A • B • A</td>
<td>60.5</td>
</tr>
<tr>
<td>A • B' • A</td>
<td>45</td>
</tr>
<tr>
<td>A' • B • A'</td>
<td>41.5</td>
</tr>
<tr>
<td>A' • B' • A'</td>
<td>36</td>
</tr>
<tr>
<td>B • A • A</td>
<td>Heterotrimers: α1•α2•α1</td>
</tr>
<tr>
<td>B' • A • A</td>
<td>60.5</td>
</tr>
<tr>
<td>B • A' • A'</td>
<td>39.5</td>
</tr>
<tr>
<td>B' • A' • A'</td>
<td>36.5</td>
</tr>
<tr>
<td>A • A • B</td>
<td>Heterotrimers: α1•α1•α2</td>
</tr>
<tr>
<td>A • A • B'</td>
<td>60</td>
</tr>
<tr>
<td>A' • A • B</td>
<td>44.5</td>
</tr>
<tr>
<td>A' • A' • B'</td>
<td>42.5</td>
</tr>
<tr>
<td>A' • A' • B'</td>
<td>36.5</td>
</tr>
</tbody>
</table>

Therefore, it can be concluded that the thermal stability of the heterotrimers is similar for all possible registers. Thus, the number, and not position, of mutations is the primary indicator of the thermal stability of a triple helix containing Gly mutations.