RICE UNIVERSITY

Multi-Hierarchical Self-Assembly of Collagen Mimetic Peptides into AAB Type Heterotrimers, Nanofibers and Hydrogels Driven by Charged Pair Interactions

by

Lesley Russell O'Leary

A THESIS SUBMITTED

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Jeffrey D. Hartgerink, Chair
Associate Professor, Chemistry

Zachary T. Ball
Assistant Professor, Chemistry

K. Jane Grande-Allen
Associate Professor, Bioengineering

HOUSTON, TEXAS

JANUARY 2012

ABSTRACT

Multi-Hierarchical Self-Assembly of Collagen Mimetic Peptides into AAB Type Heterotrimers, Nanofibers and Hydrogels Driven by Charged Pair Interactions

by

Lesley Russell O'Leary

Replicating the multi-hierarchical self-assembly of collagen (peptide chain to triple helix to nanofiber and, finally, to a hydrogel) has long attracted scientists, both from the fundamental science perspective of supramolecular chemistry and for the potential biomedical applications perceived in tissue engineering. In terms of triple helical formation, collagen is the most abundant protein in the human body with at least 28 types, yet research involving collagen mimetic systems has only recently began to consider the innate ability of collagen to control helix composition and register. Collagen triple helices can be homotrimeric or heterotrimeric and while some types of natural collagen form only one specific composition of helix, others can form multiple. It is critical to fully understand and, if possible, reproduce the control that native collagen has on helix composition and register. In terms of nanofiber formation, many approaches to drive the self-assembly of synthetic systems through the same steps as natural collagen have been partially successful, but none have simultaneously demonstrated all levels of structural assembly. In this work, advancements in the ability to control helix composition and replicate the multi-hierarchical assembly of collagen are described. Both positive and negative design for the assembly of AAB type collagen heterotrimers were

utilized by promoting heterotrimer formation though the use of charged amino acids to form intra-helix electrostatic interactions, while simultaneously discouraging homotrimers, resulting in the identification of multiple peptide systems with full control over the composition of the resulting triple helix. Similar salt-bridged hydrogen bonds between charged residues were incorporated into nanofiber forming peptides, one of which successfully assembled into sticky-ended triple helices, nanofibers with characteristic triple helical packing visible in the solution state, and strong hydrogels that are degraded by collagenase at a similar rate to natural collagen. Together, these results provide a better understanding of the self-assembly of collagenous sequences as well as a novel design scheme for synthetic extracellular matrix mimetics with potential applications in regenerative medicine and drug delivery.

ACKNOWLEDGMENTS

There are many people that I would like to thank for the support and strength that they have given me over the course of my graduate school work. The first of which is my advisor Jeffrey Hartgerink who guided me over the past four years with a constant positive balance of freedom and supervision. Thank you for allowing me to take charge of my research from an early stage while consistently reminding me that one person cannot do everything.

Next, I would like to thank the Hartgerink lab both past and present for your guidance when I was a first year graduate student and your willingness to help with research or life outside of graduate school whenever necessary. Specifically, I would like to thank Jorge Fallas. Thank you for everything. It has been a roller coaster but I could not imagine a better person to help me get through it all.

Last, I want to thank my family, which believed in me from the start and never let me lose sight of my goals. Meredith, thank you for constantly leading through example, as only an older sister can, and showing me that achieving any dream is possible given work ethic and perseverance. Mom, thank you for always reminding me to focus on the bigger picture and that every failure can lead to victory if taken as an opportunity for learning. Dad, thank you for passing on to me your inquisitive nature and your passion for learning things first hand via experience. One of the best lessons in graduate school is to never accept results until you have tested them yourself and you instilled this passion for knowledge in me at an early age. And George, thank you for the daily unconditional love that you have given me over the past five years. You supported my dreams when we first met and continue to remind me that the sky is the limit. Thank you.

TABLE OF CONTENTS

List of Tables.....x

List of Figures.....xi

Abbreviations.....xvi

Chapter 1: Introduction.....1

1.1. Supramolecular Chemistry: from Biology to Synthetic Mimics.....1

1.2. Self-Assembly of Biological Systems.....2

1.3. Synthetic Peptide Assembly.....5

1.3.1. Alpha-helical Coiled-Coils.....6

1.3.2. Beta-Sheet Nanofibers.....8

1.3.2.1. Cyclic Peptide Nanofibers.....9

1.3.2.2. Peptide Amphiphiles.....10

1.3.2.3. Multi-Domain Peptides.....11

1.3.3. Drawbacks to Alpha-Helical and Beta-Sheet Nanofibers.....14

1.4. Collagen.....15

1.4.1. Collagen Assembly.....16

1.4.2. Collagen Mimetic Peptides: Synthesis and Characterization 17

1.4.3. Homotrimeric Synthetic Collagen Mimics.....21

1.4.3.1. Triple Helical Structure.....21

1.4.3.2. Triple Helical Stability and Folding.....26

1.4.4. Heterotrimeric Collagen Mimetics.....35

1.4.4.1. Covalently-Linked Heterotrimers.....36

1.4.4.2. Self-Assembled Heterotrimers.....39

1.4.5. Higher Order Assemblies of Collagen Mimetics.....44

1.5. Conclusion and Preview of Thesis.....52

References.....55

Chapter 2: AAB Heterotrimer Formation Using Positive Design.....67

2.1. Peptide Design.....67

2.2. Homotrimer Formation.....69

2.3. Heterotrimer Formation.....74

2.3.1. A/B System.....75

2.3.2. A/C System.....77

2.3.3. B/C System.....79

2.4. Conclusions.....81

2.5. Experimental.....82

References.....84

Chapter 3: AAB Heterotrimer Formation Incorporating Negative Design.....85

3.1. Peptide Design.....85

3.2. Peptide Library.....86

3.3. Homotrimer Stability.....86

3.4. Proof of Design Principles.....92

3.5. CD Analysis of Heterotrimer Formation.....94

3.5.1. Arginine-Aspartate Pairing.....95

3.5.2. Arginine-Glutamate Pairing.....99

3.5.3. Lysine-Glutamate Pairing.....103

3.5.4. Lysine-Aspartate Pairing.....107

3.6. DSC Analysis on Selected Systems.....110

3.7. NMR Analysis on Selected Systems.....114

3.7.1. NMR Analysis of $(PRG)_{10}$•$2(EOGPOG)_5$ in Phosphate.....115

3.7.2. NMR Analysis of $2(PKGPOG)_5$•$(EOG)_{10}$ and $2(PKGPOG)_5$•$(DOG)_{10}$...120

3.7.2.1. NMR Analysis of $2WG(PKGPOG)_5$•$WG(EOG)_{10}$ in Tris.....124

3.7.2.2. NMR Analysis of $2WG(PKGPOG)_5$•$WG(DOG)_{10}$ in Phosphate....128

3.8. Conclusions on Designed AAB Systems.....131

3.9. Experimental.....133

References.....140

Chapter 4: Collagen Mimetic Nanofiber Formation Driven by Hydrophobic Interactions.....142

4.1. Peptide Design.....143

4.2. CD Analysis of Triple Helical Stability.....148

4.2.1. CD Analysis of Model I.....148

4.2.2. CD Analysis of Model III.....151

4.2.2.1. CD Analysis of Model III: $(LOGPOG)_5$.....152

4.2.2.2. CD Analysis of Model III: $(POGPIG)_5$.....154

4.2.2.3. CD Analysis of Model III: $(LOGPIG)_5$.....157

4.3. DSC Analysis of Triple Helix Re-folding on Selected Systems.....159

4.3.1. DSC Analysis of Model I.....160

4.3.2. DSC Analysis of Model III: $(LOGPOG)_5$.....162

4.3.3. DSC Analysis of Model III: $(POGPIG)_5$.....163

4.4. Nanofiber Formation.....164

4.5. Conclusions and Drawbacks.....167

4.6. Experimental.....168

References.....172

Chapter 5: Collagen Mimetic Nanofiber Formation Driven by Electrostatic Interactions.....174

5.1. $(PRG)_4(POG)_4(EOG)_4$ Replication.....175

5.1.1. $(PRG)_4(POG)_4(EOG)_4$: Triple Helical Stability.....175

5.1.2. $(PRG)_4(POG)_4(EOG)_4$: Nanofiber Formation.....179

5.1.3. Drawbacks of $(PRG)_4(POG)_4(EOG)_4$ Peptide.....182

5.2. Modifications to $(PRG)_4(POG)_4(EOG)_4$.....183

5.2.1. $(EOG)_3(POG)_3(PRG)_3$.....183

5.2.2. $(DOG)_4(POG)_4(PKG)_4$ and $(DOG)_6(POG)_6(PKG)_6$.....186

5.2.3. $(PKG)_4(POG)_4(DOG)_4$ Assembly Based on Chaikof Method.....190

5.3. Hierarchical Self-Assembly of $(PKG)_4(POG)_4(DOG)_4$.....192

5.3.1. $(PKG)_4(POG)_4(DOG)_4$: Triple Helical Stability.....192

5.3.2. $(PKG)_4(POG)_4(DOG)_4$: Nanofiber Formation.....194

5.3.3. $(PKG)_4(POG)_4(DOG)_4$: Hydrogel Properties.....200

5.3.4. $(PKG)_4(POG)_4(DOG)_4$: Enzyme Degradation.....204

5.3.5. $(PKG)_4(POG)_4(DOG)_4$: X-Ray Diffraction.....206

5.3.6. $(PKG)_4(POG)_4(DOG)_4$: Proposed Mechanism of Assembly.....208

5.3.7. $(PKG)_4(POG)_4(DOG)_4$: Assembly in Additional Buffers.....210

5.4. Conclusions.....217

5.5. Experimental.....218

References.....223

Chapter 6: Conclusions.....226

Appendix 1: Peptide Library.....228

Appendix 2: Peptide Purification.....230

LIST OF TABLES

Chapter 3

3.1 Peptide library and melting temperatures of homotrimers.....87

3.2 Peptide mixtures and melting temperatures measured by CD.....95

3.3 Glycine packing cross-peaks expected from the three possible registers of $(PRG)_{10}$•$2(EOGPOG)_5$.....119

Chapter 4

4.1 Designed peptide models for hydrophobicity driven nanofiber formation.....147

Chapter 5

5.1 Collagenase mediated degradation of $(PKG)_4(POG)_4(DOG)_4$ and rat-tail collagen hydrogels.....205

LIST OF FIGURES

Chapter 1

1.1 Scheme depicting the lock and key model for enzyme-substrate interactions.....2

1.2 Structures of a-, b- and z-type DNA helices.....3

1.3 Schematic representations of an α-helix and a β-sheet.....5

1.4 Schematic representation of an α-helical coiled-coil.....7

1.5 Current research within the coiled-coil field.....8

1.6 Cyclic peptide nanotubes designed by Hartgerink *et al*.....9

1.7 Schematic of the covalent capture peptide amphiphile system.....11

1.8 Proposed model of MDP nanofiber self-assembly.....13

1.9 Multi-hierarchical self-assembly of collagen type I.....17

1.10 Triple helical structure of $(POG)_{10}$ collagen mimetic peptide.....22

1.11 Triple helical symmetry.....25

1.12 Triple helical stabilization mechanisms.....27

1.13 Computer models of pairwise interactions between lysine and glutamate.....32

1.14 Chemical structures of small molecules used for the templated assembly of triple helical peptides.....33

1.15 Regioselective cystein chemistry for heterotrimeric collagen mimetic peptide synthesis.....37

1.16 Strategy for the self-assembly of ABC heterotrimeric collagen mimetic peptides utilizing electrostatic interactions.....42

1.17 Solution structure of an ABC heterotrimer.....43

1.18 Scheme describing the progress that collagen mimetic peptides have made in the replication of the hierarchical self-assembly of collagen.....46

1.19 Hierarchical self-assembly of synthetic collagen systems compared to rat-tail collagen.....49

Chapter 2

2.1 Scheme depicting the three types of lysine-aspartate interactions.....68

2.2 Possible heterotrimeric species formed by the A/B, A/C and B/C systems.....69

2.3 CD experiments performed on the homotrimers of peptides A, B and C.....71

2.4 Homotrimers formed by peptides A, B and C.....73

2.5 CD thermal unfolding curves for A/B system.....76

2.6 CD thermal unfolding curves for A/C system.....78

2.7 CD thermal unfolding curves for B/C system.....80

Chapter 3

3.1 Structures of phosphate and Tris.....87

3.2 CD thermal unfolding curves for all homotrimers with a ±10 charge.....88

3.3 CD thermal unfolding curves for all homotrimers with a ±5 charge.....90

3.4 DSC melting profiles for $(PRG)_{10}$.....91

3.5 CD melting experiments on mixtures of $(PRG)_{10}$ and $(EOGPOG)_5$ at different peptide ratios.....93

3.6 CD thermal unfolding curves for the $(PRG)_{10}$•$2(POGDOG)_5$ system.....96

3.7 CD thermal unfolding curves for the $2(PRGPOG)_5$•$(DOG)_{10}$ system.....98

3.8 CD thermal unfolding curves for the $(PRG)_{10}$•$2(EOGPOG)_5$ system.....101

3.9 CD thermal unfolding curves for the $(PRG)_{10}$•$2(POGEOG)_5$ system.....102

3.10 CD thermal unfolding curves for the $2(PRGPOG)_5$•$(EOG)_{10}$ system.....103

3.11 CD thermal unfolding curves for the $(PKG)_{10}$•$2(EOGPOG)_5$ system.....104

3.12 CD thermal unfolding curves for the $(PKG)_{10}$•2$(POGEOG)_5$ system.....106

3.13 CD thermal unfolding curves for the 2$(PKGPOG)_5$•$(EOG)_{10}$ system.....107

3.14 CD thermal unfolding curves for the $(PKG)_{10}$•2$(POGDOG)_5$ system.....109

3.15 CD thermal unfolding curves for the 2$(PKGPOG)_5$•$(DOG)_{10}$ system.....110

3.16 DSC melting profile for $(PRG)_{10}$•2$(EOGPOG)_5$ system in Tris.....111

3.17 DSC melting profiles for the 2$(PKGPOG)_5$•$(EOG)_{10}$ system.....113

3.18 DSC melting profile for 2$(PKGPOG)_5$•$(DOG)_{10}$ in phosphate.....114

3.19 Solution NMR experiments on $(PRG)_{10}$•2$(EOGPOG)_5$ in Tris.....116

3.20 Graphical depiction of NOE assignments for $(PRG)_{10}$•2$(EOGPOG)_5$ in Tris....118

3.21 $^1H,^{15}N$-HMQC NMR spectra of 2WG$(PKGPOG)_5$•WG$(EOG)_{10}$ and 2WG$(PKGPOG)_5$•WG$(DOG)_{10}$.....122

3.22 $^1H,^1H$-edited NOESY NMR spectrum showing NH-NH resonances in 2WG$(PKGPOG)_5$•WG$(EOG)_{10}$ system.....126

3.23 $^1H,^1H$-edited NOESY NMR spectrum showing interaction between charged residues in 2WG$(PKGPOG)_5$•WG$(EOG)_{10}$ system.....128

3.24 $^1H,^1H$-edited NOESY NMR spectrum showing glycine packing interactions in 2WG$(PKGPOG)_5$•WG$(DOG)_{10}$ system.....130

3.25 $^1H,^1H$-edited NOESY NMR spectrum showing interaction between charged residues in 2WG$(PKGPOG)_5$•WG$(DOG)_{10}$ system.....131

Chapter 4

4.1 Images of $(POG)_{10}$ homotrimer and $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ heterotrimer...142

4.2 Schematic of the driving forces from triple helical and nanofiber assembly.....143

4.3 Schematic representation of the four models designed for the formation of heterotrimeric nanofibers via hydrophobic interactions.....144

4.4 Schematic representation of the proposed nanofiber assembly for the designed hydrophobic models.....146

4.5 CD thermal stability for the component peptides of Model I.....149

4.6 CD thermal stability for the 1:1 mixtures of the peptides of Model I.....150

4.7 CD melting studies on the 1:1:1 mixture of the peptides in Model I.....151

4.8 CD melting experiments for $(PKG)_{10}$, $(DOG)_{10}$ and $(LOGPOG)_5$.....152

4.9 CD melting experiments for the 1:1 mixtures of $(PKG)_{10}$, $(DOG)_{10}$ and $(LOGPOG)_5$.....153

4.10 CD melting experiments for the 1:1:1 mixture of $(PKG)_{10}$, $(DOG)_{10}$ and $(LOGPOG)_5$.....154

4.11 CD melting experiments for $(PKG)_{10}$, $(DOG)_{10}$ and $(POGPIG)_5$ and the 1:1 mixtures of the peptides.....156

4.12 CD thermal unfolding studies for the 1:1:1 mixture of $(PKG)_{10}$, $(DOG)_{10}$ and $(POGPIG)_5$.....157

4.13 CD thermal unfolding curves for $(PKG)_{10}$, $(DOG)_{10}$ and $(LOGPIG)_5$.....158

4.14 CD thermal unfolding curve for the 1:1:1 mixture of $(PKG)_{10}$, $(DOG)_{10}$ and $(LOGPIG)_5$.....159

4.15 DSC melting profiles for Model I.....161

4.16 DSC melting profile for $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPOG)_5$.....162

4.17 DSC melting profile for $(PKG)_{10}$•$(DOG)_{10}$•$(POGPIG)_5$.....163

4.18 TEM images of hydrophobic Model I.....165

4.19 Dry-TEM images of $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPOG)_5$.....166

4.20 TEM images of hydrophobic model $(PKG)_{10}$•$(DOG)_{10}$•$(POGPIG)_5$.....167

Chapter 5

5.1 CD thermal stability for $(PRG)_4(POG)_4(EOG)_4$ reported by Chaikof.....176

5.2 CD melting studies for $(PRG)_4(POG)_4(EOG)_4$ in different pHs.....178

5.3 Dry-TEM images of $(PRG)_4(POG)_4(EOG)_4$ in Buffer 8.....180

5.4 Cryo-TEM images of $(PRG)_4(POG)_4(EOG)_4$ in Buffer 8.....181

5.5 CD thermal unfolding curves for $(PRG)_4(POG)_4(EOG)_4$ and $(EOG)_3(POG)_3(PRG)_3$ in Buffer 8.....184

5.6 TEM images of $(EOG)_3(POG)_3(PRG)_3$ in Buffer 8.....186

5.7 CD melting curves for $(PRG)_4(POG)_4(EOG)_4$, $(DOG)_4(POG)_4(PKG)_4$ and $(DOG)_6(POG)_6(PKG)_6$ in Buffer 8.....188

5.8 Dry and cryo-TEM images of $(DOG)_4(POG)_4(PKG)_4$ and $(DOG)_6(POG)_6(PKG)_6$ in Buffer 8.....189

5.9 CD thermal unfolding studies for $(PRG)_4(POG)_4(EOG)_4$ and $(PKG)_4(POG)_4(DOG)_4$ in Buffer 8.....190

5.10 Cryo-TEM images of $(PKG)_4(POG)_4(DOG)_4$ in Buffer 8.....191

5.11 CD spectra of $(PKG)_4(POG)_4(DOG)_4$ in phosphate buffer.....194

5.12 Dry-TEM images of $(PKG)_4(POG)_4(DOG)_4$ in phosphate buffer.....196

5.13 Cryo-TEM images of $(PKG)_4(POG)_4(DOG)_4$ in phosphate buffer.....197

5.14 AFM images of $(PKG)_4(POG)_4(DOG)_4$ in phosphate buffer.....198

5.15 SEM images of $(PKG)_4(POG)_4(DOG)_4$ in phosphate buffer.....199

5.16 Time course rheological studies on $(PKG)_4(POG)_4(DOG)_4$ in phosphate.....201

5.17 Rheological studies on $(PKG)_4(POG)_4(DOG)_4$ in phosphate buffer.....203

5.18 Proposed mechanism of $(PKG)_4(POG)_4(DOG)_4$ fiber self-assembly.....207

5.19 CD thermal unfolding curves for $(PKG)_4(POG)_4(DOG)_4$ in different buffers...211

5.20 AFM of $(PKG)_4(POG)_4(DOG)_4$ nanofibers in water.....213

5.21 AFM of $(PKG)_4(POG)_4(DOG)_4$ nanofibers in Tris.....214

5.22 AFM of $(PKG)_4(POG)_4(DOG)_4$ nanofibers in PBS.....215

5.23 Rheology of $(PKG)_4(POG)_4(DOG)_4$ in different buffers.....216

ABBREVIATIONS

AFM	atomic force microscopy
CD	circular dichroism spectroscopy
CMP	collagen mimetic peptide
DiEA	N,N,-diisopropylethylamine
DMF	N,N,-dimetylformamide
DSC	differential scanning calorimetry
ECM	extracellular matrix
ESI-TOF	electrospray ionization time of flight mass spectrometry
Fmoc	9-fluoronylmethoxycarbonyl
HBTU	O-benzotriazole N,N,N',N'-tetramethyluronium hexafluorophosphate
HOBt	1-hydroxybenzotriazole hydrate
HPLC	high pressure liquid chromatography
MALDI-TOF	matrix desorption assisted time of flight mass spectrometry
MRE	molar residual ellipticity
NMR	nuclear magnetic resonance spectroscopy
NOESY	nuclear Overhauser effect spectroscopy
PBS	phosphate buffered saline
SEM	scanning electron microscopy
TEM	transmission electron microscopy
TFA	trifulouroacetic acid
TOCSY	total correlated spectroscopy
Tris	tris(hydroxymethyl)aminomethane

Chapter 1: Introduction

1.1. Supramolecular Chemistry: from Biology to Synthetic Mimics

Over the past several decades, the boundaries between the traditional disciplines of chemistry (organic, inorganic, physical and analytical) and even between chemistry, physics and biology have become blurred by the emergence of new areas within the scientific fields; one of the most notable of these being supramolecular chemistry. Similar to many of the other contemporary disciplines such as nanotechnology, supramolecular chemistry gained notoriety with the awarding of a Nobel Prize. In 1987, Donald J. Cram, Jean-Marie Lehn and Charles J. Peterson received the Nobel Prize in Chemistry for their independent, but collectively important, research involving the design, development and use of molecules containing highly selective structure-specific interactions.[1] However, the birth of this new field has widely been credited to Emil Fischer for his "Lock and Key" model of enzyme-substrate interactions proposed in the 1890's. In his model, Fischer hypothesized that the specificity and selectively of enzymes for their respective substrates is driven by the complementary shapes of the two molecular structures, similar to the way that a lock is opened with a key (Figure 1.1).[1] The recognition of the importance of molecular shape even before molecular structure was fully defined makes Fischer's model the inspiration behind all subsequent work within the field of supramolecular chemistry.

By definition, supramolecular chemistry is the study of intermolecular (non-covalent) interactions and the structures built using these forces.[1] Covalent bonds are the backbone for intramolecular interactions due to their high energies, directionality and

facilitation for the achievement of stable electron configurations by sharing electrons. Intermolecular interactions, specifically hydrogen bonding, electrostatic interactions, π-π interactions and the hydrophobic effect, have increased distances compared to covalent bonds allowing for a longer range of impact as well as reversibility which allows for triggered formation and disassembly of structures designed using these interactions. Such structures utilize the "bottom up" approach for molecular design in which smaller sub-sets are combined in order to make and analyze a larger superstructure. This method is in contrast to the more historical "top down" approach in which larger systems are used and broken down to understand the smaller sub-units. Biological systems utilize the "bottom up" methodology therefore this approach has emerged as a strong method for scientific research.

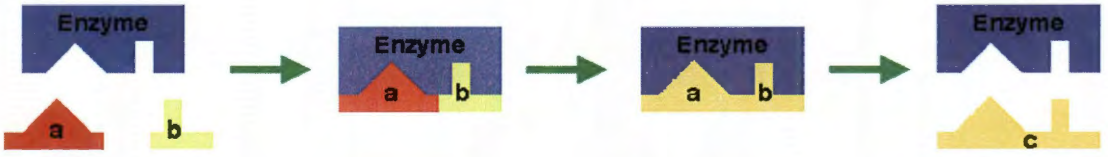

Figure 1.1. Scheme depicting the lock and key model for enzyme-substrate interactions. The substrates, labeled as **a** and **b**, dock into the enzyme, are converted into product **c** and the product is then released from the enzyme.

1.2. Self-Assembly of Biological Systems

A basic example of supramolecular assembly within the human body is DNA, the polynucleotide that contains the entire genetic code. The double helix structure, initially solved by Watson and Crick,[2] is stabilized by the combination of π-π stacking between nucleic acids along the axis of the double helix and hydrogen bonding perpendicular to

the helical axis between complementary nucleic acids. The structure of the double helix is shown in Figure 1.2. Once the DNA double helix is formed, it is assembled further in the presence of histones[3] into tightly packed chromatin structures,[4] which are the state that the DNA will remain in until it is needed for transcription.

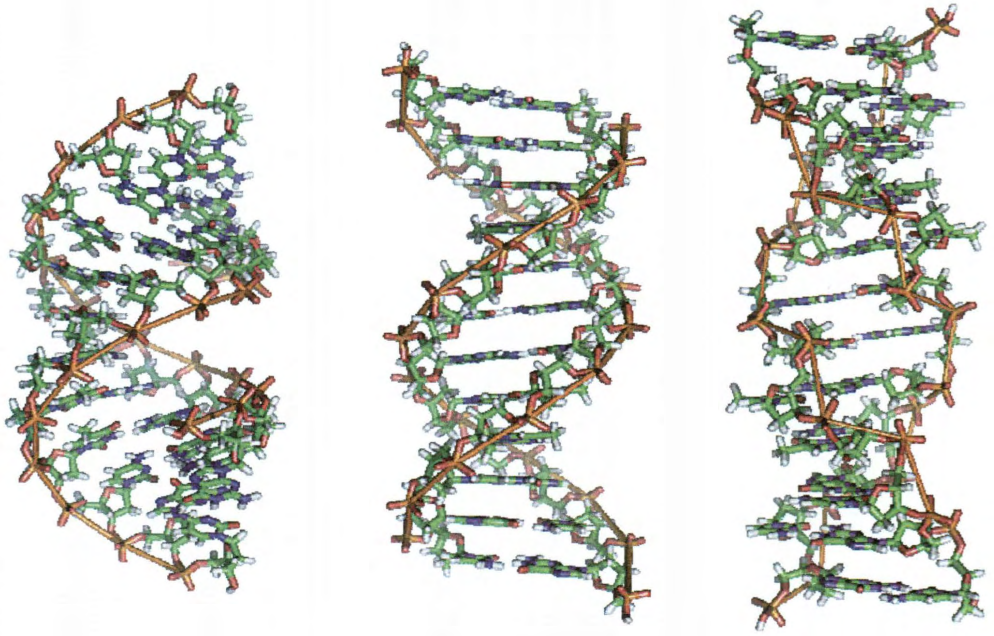

Figure 1.2. Structures of **a-**, **b-** and **z**-type DNA helices respectively. Adapted from reference 5.[5]

On a larger scale, the importance of supramolecular interactions can be seen through the phospholipid bilayer that forms the plasma membrane around human cells. Briefly, the hydrophobic lipid portion of the phospholipid interdigitates with the lipid chain of other phospholipids creating a 2-dimensional elongation of the bilayer with the lipids creating the core and the hydrophilic phosphate heads forming the top and bottom edges of the membrane. Assembly occurs in such a way because of the aqueous environment of the human body. However, if a plasma membrane is exposed to a

hydrophobic solvent such as hexane, the bilayer would invert so that the lipids would be exposed to the solvent and the phosphate heads would be shielded in the core.

The most diverse example of supramolecular chemistry within the body is the self-assembly of proteins. Proteins are polypeptide chains composed of hundreds of amino acids that undergo multiple levels of assembly beginning with the amino acid sequence, the primary structure, and ending with the 3-D morphology comprised of multiple peptide subunits, the quaternary structure. The amino acid sequence dictates the secondary structure of the polypeptide based on the placement of charged, hydrophilic, hydrophobic, aromatic and imino acid residues within the chain. The most common secondary structures include α-helices, beta β-sheets and collagen triple helices which are all stabilized by hydrogen bonding interactions between specific amino acids within the peptide chain. Figure 1.3 depicts the hydrogen bonding in α-helices and β-sheets. The hydrogen bonding in both structures is between amides and carbonyls in the peptide backbone however for an α-helix, it is parallel to the helical axis and is specifically between amino acids in positions *i* and *i+4* in the same peptide chain. In β-sheets, individual peptide chains are called β-strands and they form β-sheets by hydrogen bonding to adjacent β-strands perpendicular to the peptide backbone axis. If the β-strands are aligned so that the N-termini are on the same side of the β-sheet, it is called a parallel β-sheet and if one end of the β-sheet has alternating N and C termini between β-strands, it is referred to as anti-parallel. The details of the collagen triple helix will be described in detail below.

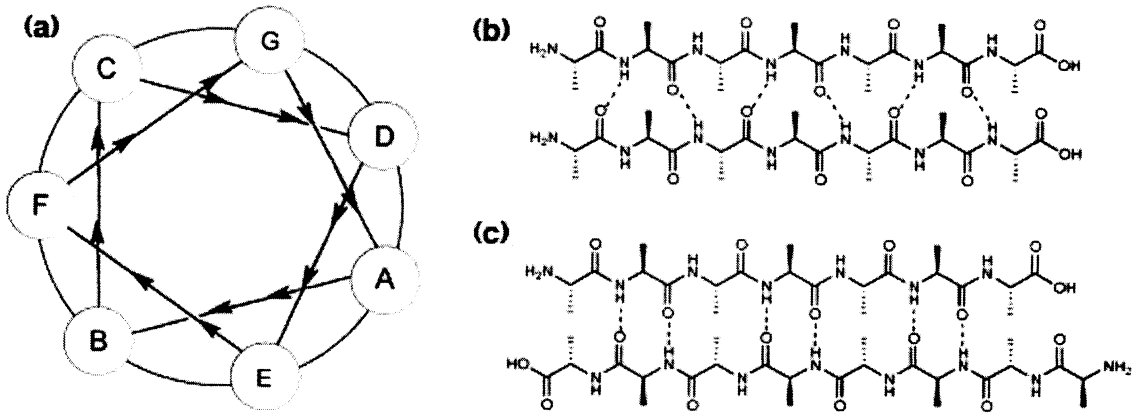

Figure 1.3. Schematic representations of (a) an α-helix with the alphabet used to depict the amino acid sequence and (b) a parallel β-sheet and (c) an anti- parallel β-sheet.

1.3. Synthetic Peptide Self-Assembly

Based on the diversity of protein self-assembly, it has become the subject of much research in an attempt to synthetically replicate it. The building blocks of proteins, amino acids, can be categorized into different subsections based on the nature of the residue side chain and these differences can account for the variety of secondary and tertiary structures seen in native proteins. The functionality of the side chains can be aromatic, charged, hydrophilic, and hydrophobic or a combination of these traits. Aromatic amino acids participate in π-π stacking interactions, charged residues can contribute to hydrogen bonding or partake in ion-ion or ion-dipole interactions, hydrophilic amino acids can participate in hydrogen bonding or form interactions with water and hydrophobic residues pack together in aqueous systems to reduce the number of water molecules that are interacting with the hydrophobic surface. Each category of amino acids has the power to drive the assembly of peptides and each play an integral role in native protein folding. Therefore, the implementation of different types of residues in different positions within

a peptide chain can lead to diverse secondary structures and nano-morphologies. Synthetic systems based upon the properties of amino acids can be categorized based on the type of secondary structure that they trigger: α-helical, β-sheet and collagen triple helices.

1.3.1. Alpha-Helical Coiled-Coils

The α-helix, as mentioned previously, is one of the basic secondary structures of peptides. The structure is defined by hydrogen bonds between amino acids in positions *i* and *i+4*. These hydrogen bonds dictate a twisting of the peptide backbone resulting in the name α-helix. In order to accommodate the stabilizing hydrogen bonds between positions *i* and *i+4*, the side chains of other amino acids in the α-helix are forced into an outward position. By manipulating the identity of the residues in these solvent exposed positions, the self-assembly of multiple α-helices into coiled-coil structures can be induced. Specifically, when the repeating unit of α-helical forming peptides is considered, $(abcdefg)_n$, placing hydrophobic amino acids in positions **a** and **d** and hydrophilic residues in positions **e** and **g** caused coiled-coils composed of two α-chains to form. These coiled-coils create a left-handed super-helix composed of the right-handed α-helices. A schematic of such an assembly is shown in Figure 1.4. By implementing these design parameters, synthetic systems of dimers, trimers, tetramers, pentamers and hexamers have been reported.[6-9]

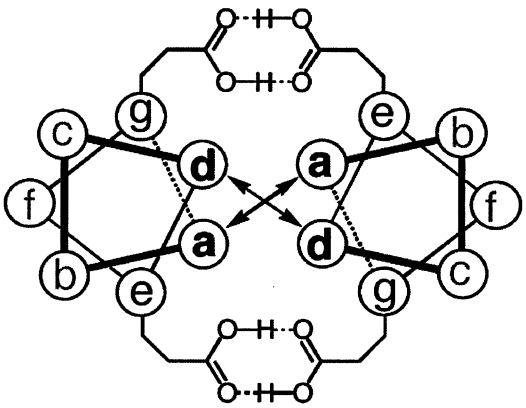

Figure 1.4. Schematic representation of an α-helical coiled-coil. Adapted from Figure 1 in reference 7. Hydrophobic amino acids are shown in black, hydrophilic residues that can participate in hydrogen bonding are in green and charged amino acids that can participate in salt bridges are in red.[7]

Through optimization, specific amino acids have been identified that drive α-helices to assemble into coiled-coil structures and even into extended coiled-coil nanofibers. Potekhin and coworkers, through a series of papers, examined coiled-coils composed of varying numbers of α-helices including a study on the kinetics and thermodynamics associated with the folding and unfolding of trimeric coiled-coils[10] and the design of pentameric coiled-coil nanofibers (Figure 1.5a).[11-13] Woolfson and coworkers have explored the fiber forming abilities of coiled-coil motifs and have thus designed a library of such peptides,[9,14-20] including a subset of sticky-ended coiled-coils that assemble into coiled-coil nanofibers[15] and another subclass of peptides that changes properties based on their environment.[14] Through this research, Woolfson has bridged the gaps between synthetic peptide replication and protein biology. Using an algorithm they designed called SOCKET, Woofson and co-workers can identify the coiled-coil domains within the structures of natural proteins.[18] Lastly, within the Hartgerink lab,[7,8,21] the limits of coiled-coil formation were explored through the assembly of homo- and heterodimeric

coiled-coils composed of short peptide chains[7] and the placement of hydrophobic amino acids within the repeating amino acid motif to convert α-helical coiled-coils to into β-sheet secondary structures.[8]

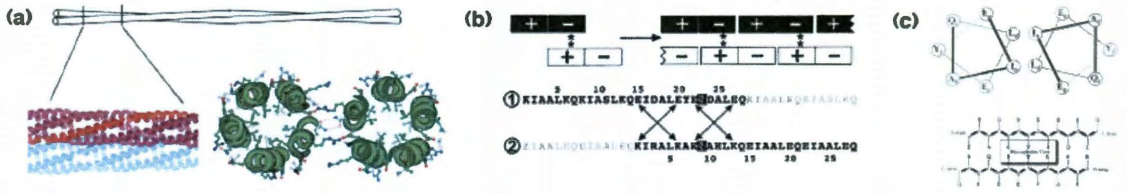

Figure 1.5. Current research within the coiled-coil field. (a) Pentameric coiled-coil system designed by Potekhin *et al.* shown as fibril-containing filaments (top), an enlarged section within the fibril (bottom left) and an axial projection of the fibril (bottom right). Adapted from Figure 5 in reference 11.[11] (b) Sticky-ended coiled-coils designed by Woolfson *et al.* that elongate into coiled-coil nanofibers where the design is shown on top and the amino acid sequence is shown below. Adapted from Figure 1 in reference 15.[15] (c) Coiled-coils that can convert into β-sheets designed by Dong and Hartgerink. Adapted from Figure 1 in reference 8.[8]

1.3.2. Beta-Sheet Nanofibers

As described above, β-sheets are peptide secondary structures in which the peptide backbones are in a mostly-extended conformation and the hydrogen bonding between backbone carbonyls and amides is perpendicular to the peptide backbone axis. By manipulating the identity and chirality of amino acids within the peptide chains, many different self-assembled systems can be made with β-sheets at the core. Specifically, there are three major types within current research: cyclic peptide nanotubes, peptide amphiphiles and multi-domain peptides.

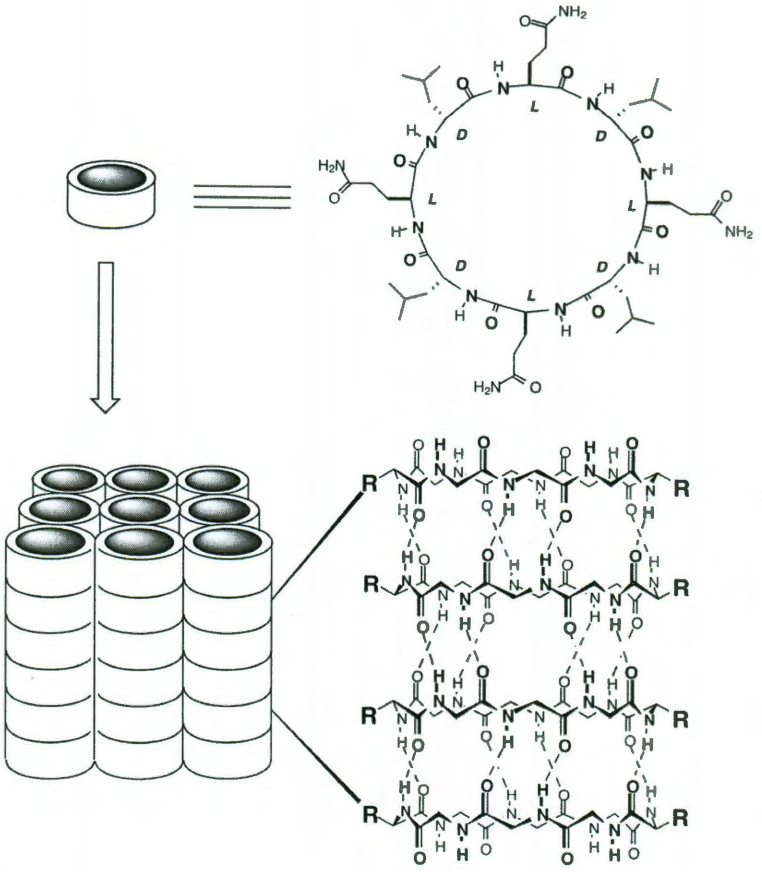

Figure 1.6. Cyclic peptide nanotubes designed by Hartgerink *et al.* Adapted from Figure 1 in reference 24.[24]

1.3.2.1. Cyclic Peptide Nanotubes

In standard β-strands, the side chains of the amino acids within the chain will alternate sides as you move along the peptide backbone. This is due to the fact that all natural occurring amino acids are L-amino acids. When amino acids in a β-sheet forming peptide are alternated between L and D chirality, the side chains of each amino acid will be on one face of the β-sheet and the resulting chain will have a propensity to curve and form a cyclic structure. This type of system was first reported by Hartgerink *et al.* (Figure 1.6) and since then, much research has continued on the optimization of peptide identity

within the peptide chain, number of amino acids in the chain and propensity to assemble into nanofibers.[22-26] The nanotubes are held together by backbone hydrogen bonding between the top face of one cyclic peptide and the bottom face of the cyclic peptide on top of it. In addition, the side chains on the peptides can be designed with specific functionality to pack nanotubes together through hydrophobicity or hydrogen bonding.[22]

1.3.2.2. Peptide Amphiphiles

The next major type of β-sheet driven synthetic peptide nanofibers is peptide amphiphiles. In short, there are three regions within a peptide amphiphile: a hydrophobic alkyl tail, a glycine linker region and a head group at the C-terminus. In an aqueous environment, the hydrophobic tails of the peptide amphiphiles will pack together to reduce the surface area of contact with the water molecules.[27-29] By adjusting the ratio of the length of the amphiphilic tail with the length of the glycine linker and head regions, the nanostructure of the resulting assembly can be adjusted from a spherical micelle to cylindrical micelles, vesicle and bilayers.[27] In order to replicate the extracellular matrix within the body, peptide amphiphiles with cylindrical micelle morphologies have been most thoroughly pursued. Through the addition of cysteine residues between the hydrophobic tail and glycine linker region, disulfide bonds can be used to covalently capture peptide amphiphile nanofibers and prevent the previously reversible assembly.[29] A schematic of the covalent capture design is shown in Figure 1.7. Since the initial reports on cylindrical peptide amphiphiles, the properties of these structures has been explored for the driving forces of stability[30-33] and functionalized for many different purposes including silica nanotube formation,[34] cell entrapment,[35] dental tissue

engineering scaffolding,[36] inhibition of cancer cell proliferation[37] and ischemic tissue repair.[38]

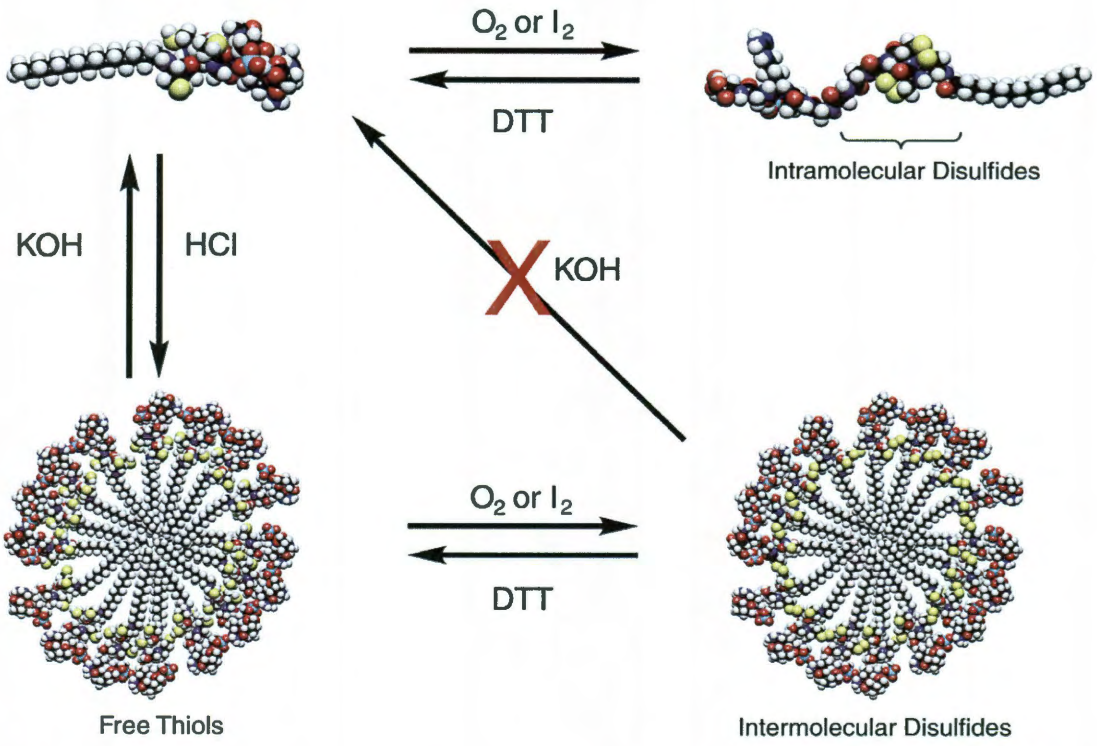

Figure 1.7. Schematic of the covalent capture peptide amphiphile system designed by Hartgerink *et al.* that reversibly forms such structures based on the pH of the sample but once oxidized, the peptide amphiphile is stable over a broad pH range. Adapted from Figure 5 in reference 29.[29]

1.3.2.3. Multi-Domain Peptides

The last subsection of β-sheet driven peptides are multi-domain peptides (MDPs). These peptides alternate amino acids with hydrophobic and hydrophilic side chains within the core of the peptide in order to create two distinct faces along the resulting β-sheet: a hydrophilic face and a hydrophobic face. In an aqueous solution, the hydrophobic faces of two MDPs will pack together to reduce the contact area of hydrophobic residues

with water. Peptide nanofibers utilizing these β-sheet conformations have been synthesized and explored Zhang, Pochan and Hartgerink.

Zhang and coworkers pioneered work in this field in 1993 through the synthesis of the peptide named EAK16 in which alanine is the hydrophobic amino acid and a combination of glutamate and lysine are used in the hydrophilic face.[39,40] This peptide, in water, forms sandwiched dimers that assembly perpendicular to the peptide backbone into nanofibers that are stabilized by alanine residues in the interior and salt bridges between lysine and glutamate in the exterior. Modifications of this initial design resulted in the peptide RADA16, composed of arginine, aspartate and alanine, which has been extensively studied in vitro and is commercially available as Puramatrix.[41,42]

Pochan and Schneider modified the design published by Zhang to create β-hairpin structures that further assembled into nanofibers. Their original peptide, MAX1, utilized valines and lysines to form the hydrophobic and hydrophilic faces respectively.[43] At physiological pH, charge repulsion between the lysine residues prevents β-hairpin assembly. However, at pH 9, the lysine side chain is deprotonated allowed β-hairpin and subsequent nanofiber assembly.[43] In 2007, the sequence was modified by the replacement of one lysine residue with a glutamate and the subsequent peptide, MAX8, forms hydrogels with higher strengths than MAX1.[44] Due to the differences in the hydrogel properties of the two peptides, they have both been explored for biomedical applications including biomineralization and continue to be studied for potential tissue engineering applications.[45]

Hartgerink took a different approach than either Zhang or Pochan in the assembly of MDPs. By using polar amino acids instead of charged residues in the hydrophilic face

of the peptides, the resulting peptide assembles into insoluble β-sheets with the hydrogen bonding axis perpendicular to the peptide axis. However, when charged amino acids were placed at the N and C termini of the core region, the resulting peptide was water-soluble and assembly into nanofibers was triggered by the addition of charge-screening salts.[46] A schematic of the basic MDP design is shown in Figure 1.8. The incorporation of aromatic amino acids into the hydrophobic region, Hartgerink's MDPs can be used as carbon nanotube surfactants due to π-π stabilizing interactions.[47-49] These peptides are also easily functionalized for tissue engineering applications though the inclusion of cell adhesion or matrix-metalloprotein cleavage sequences within the MDP design[50-53] and form strong hydrogels which can serve as viable extracellular matrix substitutes.[54]

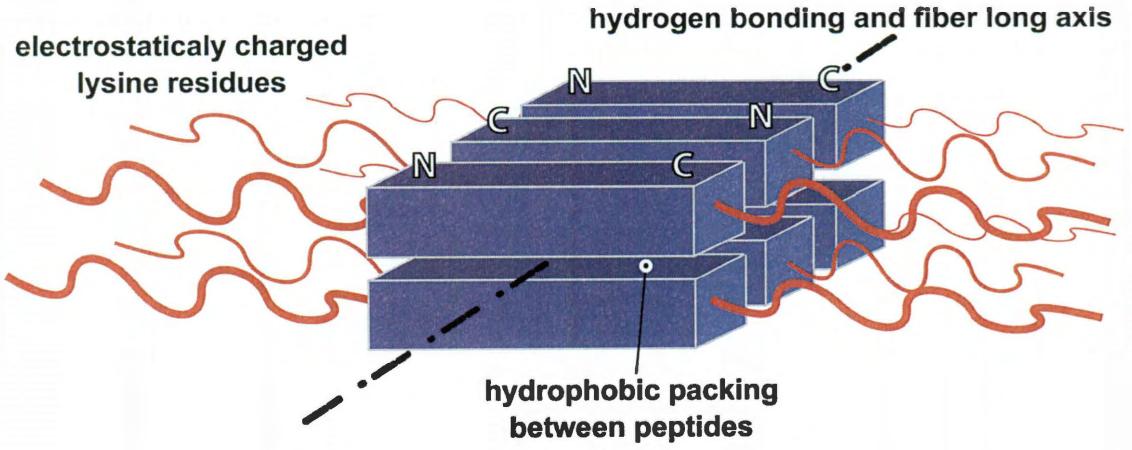

Figure 1.8. Proposed model of MDP nanofiber self-assembly including hydrophobic packing region, hydrogen bonding axis and repulsive charged peptide termini. Adapted from Figure 1 in reference 46.[46]

1.3.3. Drawbacks to Alpha-Helical and Beta-Sheet Nanofibers

Despite the ability of α-helical and β-sheet based systems to replicate the properties of the extracellular matrix through nanofiber formation, there are major drawbacks to the use of these types of peptide assemblies. Amyloid fibrils have been associated with the diagnosis and progression of neurodegenerative diseases.[55] These assemblies are β-sheet based and once formed, they are very difficult to disassemble and remove; which is part of the problem they pose in diseases such as Alzheimer's. Therefore any synthetic assembly that is considered for use in tissue engineering applications is judged based on the similarity of the system to amyloid fibrils. Due to the presence of β-sheets within cyclic nanotubes, peptide amphiphiles and multi-domain peptides, these systems are at risk of forming amyloid-like species *in vivo*. Similarly, since α-helical coiled-coils can be converted into β-sheets based on the placement of hydrophobic residues within the peptide chains, these systems are also at risk. Therefore, synthetic peptide systems that are not β-sheet forming but are rather modeled after structural proteins within the extracellular matrix, such as collagen, may provide a more viable option for future tissue engineering applications.

1.4. Collagen*

Collagen, the most abundant fibrous protein in the human body, plays a major role in the structural stability of many different tissues within the body including the skin, tendons, cardiovascular system, cartilage, basement membranes and many others.[56,57] A main reason why collagen is capable of constituting this variety of tissue types due to the size of its protein family: there are at least 28 known types of collagen.[58] These large multi-domain proteins reside within the extracellular matrix (ECM) and are assembled into multiple types of supramolecular structures depending on the specific function of the protein. For example, collagen forms planar sheet-like networks within the basement membrane and then it assembles into organized fibrils and fibers in the ECM of cardiovascular tissue and tendons.[59]

Five sub-classes of collagen exist to organize the broad protein family and divide the collagen types based on their supramolecular morphology. The five classes are fiber forming, fibril associated containing interrupted triple helices (referred to as FACITs), beaded filament, anchoring fibril and network forming and transmembrane collagens.[60] These classes differ greatly in the architectures that they adopt however, all proteins within the collagen family contain a common structural domain called the collagen triple helix which will be further described in the following section. The triple helical nature of collagen draws much attention to the protein family from a biochemical perspective due collagen's participation in cell-ECM interactions[61,62] and ECM remodeling mediated by

* This section, including all sub-sections, are based on the following publication that was written in collaboration with Jorge Fallas of the Chemistry Department, who is advised by Prof. Jeff Hartgerink: Jorge A. Fallas, Lesley E. R. O'Leary and Jeffrey D. Hartgerink. Synthetic collagen mimics: self-assembly of homotrimers, heterotrimers and higher order structures. *Chem. Soc. Rev.*, **2010**, 39, 3510-3527.

matrix-metalloproteins (MMPs).[63] In both of these instances, the presence of a triple helical conformation within the ECM is necessary for function whether it's associated with homeostatic[64] or pathologic[65] conditions. From a supramolecular chemistry perspective, the ability of collagen and it's sub-classes to assemble into a broad variety of multi-hierarchical supramolecular structures makes this protein an interesting template for synthetic replication.

1.4.1. Collagen Assembly

In order to thoroughly describe the multi-step assembly of collagen, one must begin with the composition of collagen peptide chains. The amino acid sequence of natural collagen has a distinct repeating pattern X-Y-Gly where the X position is most commonly proline and the Y position is predominately 4*R*-hydroxyproline, an amino acid that is post-translationally modified from proline by the addition of a hydroxyl group on the γ-carbon. The repeating triplet of collagen drives the staggered assembly of peptides within helices such that the second peptide is offset from the first by one amino acid and the third is offset by two amino acids. This staggering allows for glycine residues to inhabit the center of the triple helix throughout its length. The high levels of proline within collagen peptides promote a poly-proline type II helix secondary structure that further supercoils with the other two peptide chains into a collagen triple helix. If all three peptide chains, also called α chains, are identical in sequence, the triple helix is called a homotrimer (AAA). A triple helix is a heterotrimer if it is composed of two distinct α chains (AAB) or three unique amino acid sequences (ABC). In native collagen, five triple helices pack in a quasi-hexagonal lattice to form fibrils that further self-assemble both

linearly and laterally into mature collagen fibers and an organized hydrogel network.[66-69] The staggered packing of fibrils within the fibers yields a characteristic banding pattern, called D-banding. D-bands can have a length of 64-67 nm depending on the helical twist of the fibrils within the fibers.[66-69] A schematic of this multi-step assembly is shown in Figure 1.9.

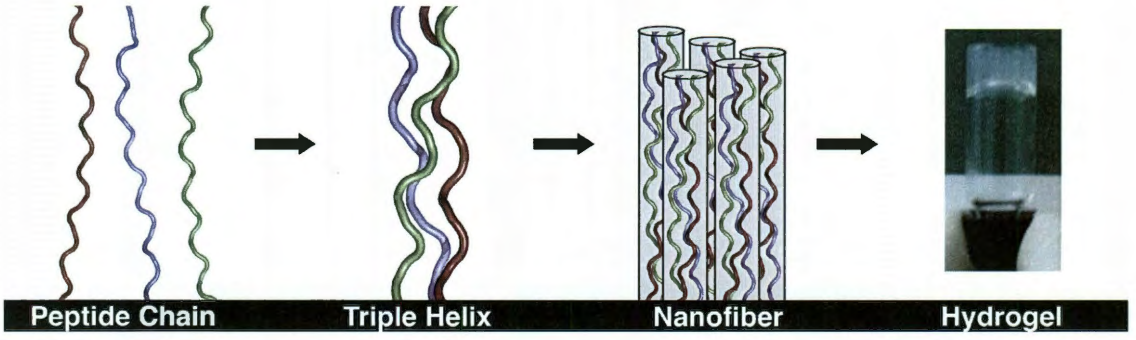

Figure 1.9. Multi-hierarchical self-assembly of collagen type I. Adapted from Figure 1 in reference 70.[70] The peptide chains (shown in red, blue and green) fold into triple helices that laterally pack in a staggered manner forming nanofibers (shown in grey). The hydrogel pictured is from a rat-tail collagen sample.[70]

1.4.2. Collagen Mimetic Peptides: Synthesis and Characterization

The largest hindrance to the duplication of collagen's supramolecular morphologies is the lack of knowledge about the mechanism of assembly by which these proteins assemble. The difficulty in studying the structure, assembly and biochemistry of collagen is largely based in the size of the collagenous proteins that can range from a few hundred amino acids per α chain to over one thousand residues per chain and the intrinsic heterogeneity of the environment that native collagen can be found.[71] In order to overcome these difficulties, short (less than 100 amino acids in length) synthetic peptides that are triple helical in nature have been designed and developed to examine the

molecular structure,[72] stability[73] and biochemistry[74] of the collagen triple helix and to further understand the higher-order self-assembly of collagen superstructures.[75] The use of the smaller synthetic peptides, referred to as collagen mimetic peptides (CMPs), has introduced high-resolution structural techniques such as X-ray crystallography[76-79] and nuclear magnetic resonance (NMR) spectroscopy,[80] which are not suited for the large native collagen, to the collagen family resulting in major advances within the field such as the first high-resolution triple helical structure[81] and the interaction mechanism of cell surface receptors such as the I-domain of the α2β1 integrin with collagenous proteins.[82]

Due to the unique amino acid pattern in collagen peptides, mimics of collagen are frequently examined using the triplet Pro-Hyp-Gly (POG in single letter code; proline is P, hydroxyproline is O and glycine is G) as a template where hydroxyproline is depicted using the three-letter code Hyp. As mentioned previously, hydroxyproline is a post-translationally modified amino acid therefore 4*R*-hydroxyproline is not encoded in the standard genome and the systems available for protein biosynthesis lack the enzyme responsible for the hydroxylation, prolyl hydroxylase. It is for this reason that the majority of collagen mimetic peptide (CMP) research has been produced by chemical synthesis. In the 1960's, the first CMPs were synthesized utilizing the then novel technology of solid phase peptide synthesis (SPPS).[83] One of the first CMPs examined, (PPG)10, exhibited behavior similar to natural collagen in terms of the temperature dependence of the peptide's optical rotation properties (ORD).[84,85] In continued studies, synthetic peptides produced single crystals,[86] an ability that is not shared with natural collagen, that yielded diffraction patterns[87] consistent with previously reported fiber diffraction patterns obtained for natural collagen.[88] The combination of the similarity in

ORD properties and diffraction patterns for CMPs compared to natural collagen validated the use of synthetic peptides as mimics of collagenous proteins.

Today, CMPs are typically synthesized using standard *N*-(9-fluorenyl)methoxycarbonyl- (Fmoc) based SSPS which includes the use of benzotriazole coupling reagents and piperidine for Fmoc deprotection. Minor modifications are necessary in order to maximize the yield of difficult couplings such as the sequential coupling of imino acids. These modifications include the coupling triplets instead of amino acids,[89] increasing the amino acid coupling time, utilizing a mixture of diaza(1,3)bicyclo[5.4.0]undecane and piperidine during the Fmoc deprotection steps[90] and the implementation of double couplings for imino acids in the X position when the Y position is occupied by proline or a proline derivative.[91] In some instances such as expensive sequences that incorporate isotopically enriched amino acids or for particularly difficult sequences, manual amino acid couplings that allow for the monitoring of the success of each deprotection and coupling as well as the use of HATU as a coupling reagent can be implemented.

Once synthesis is complete and the peptide has been cleaved from the solid support, peptides can be purified by reversed phase high-pressure liquid chromatography (HPLC) using C18 or biphenyl columns with a water-acetonitrile gradient.[92] Peptide purity is assessed by time of flight (TOF) mass spectrometry, specifically matrix desorption assisted- (MALDI-TOF) or electrospray ionization- (ESI-TOF) techniques. When utilizing MALDI-TOF mass spectrometry, an appropriate matrix for measurements in the mass range of most CMPs is α–cyano-4-hydroxycinnamic acid.

Pure peptides are initially characterized by circular dichroism (CD) spectroscopy in order to assess the folding of the peptides into collagen triple helices. In order to determine whether CMP forms a triple helix, CD experiments must be performed: a wavelength spectrum and a thermal unfolding curve. Collagen triple helices have a signature CD spectral profile that consists of a maximum at 225 nm and a minimum near 200 nm, which is indicative of a poly-proline type II helix. Depending on the sequence of the CMP examined, the exact position of the extrema, specifically the maximum, may vary within a few nanometers between different peptide assemblies. The thermal unfolding experiment monitors the spectral maximum as temperature is increased which, when a triple helix is present, shows a cooperative transition.

Other common methods used to study the supramolecular assembly of CMPs are solution NMR and x-ray crystallography as well as microscopy techniques such as atomic force microscopy (AFM), transmission electron microscopy (TEM) and scanning electron microscopy (SEM). For NMR studies, CMPs up to ten triplets can be readily characterized using homonuclear experiments however, the difficulty associated with obtaining correlation spectra greatly increases for longer peptides due to the rational diffusion anisotropy for the larger species.[93] In order to overcome this hindrance, ^{15}N- and/or ^{13}C- labeled amino acids can be incorporated into the longer peptides to increase the signal intensity. For X-ray crystallography studies, CMPs have characteristically been difficult to crystallize. However, recent developments, which include the use of polyethylene glycol as a co-precipitant, have been successful in producing well-ordered crystals.[76-79] The microscopy techniques listed, AFM, TEM and SEM, are general

techniques for assessing the nano-morphologies of the CMPs and will be discussed further in the Higher Order Assemblies section below.

1.4.3. Homotrimeric Synthetic Collagen Mimics

Over the past few decades, the vast majority of CMP research has utilized homotrimeric CMPs and it is through this work that much progress has been made in understanding the intricacies of the collagen protein family specifically the triple helical structure, stability and folding of collagen.

1.4.3.1. Triple Helical Structure

Prior to the advent of SPPS, early attempts to determine the molecular structure of collagen were based on fiber diffraction data of stretched kangaroo[94] and rat-tail[95] tendon collagen fibers. Ramachandran proposed the first triple helical model[96] based on diffraction patterns that estimated a fiber period of 28.6 Å.[97] This model, although correct in it's basic features involving the super-coiling of three left-handed chains around each other, overestimated the number of hydrogen bonds present per tripeptide unit. Rich and Crick[98] later revised the Ramachandran model, based on their findings for the poly-glycine helix,[99] and it is this Rich and Crick model that is currently accepted.[100] Simultaneously, Cowan *et al.*[95] proposed a model that was very similar to the Rich and Crick model, in the context of their work on the poly-proline type II helix.[101] Many years later, the structure proposed by Rich and Crick and Cowan *et al.* was found to be very similar to the structure of the synthetic CMP $(PPG)_{10}$, which was refined to a low

resolution electron density map based on single crystal diffraction data.[81] The first high-resolution structure (1.9 Å) of a collagen triple helix was not available until Brodsky and co-workers[102] solved the structure of a triple helical peptide, validating the Rich and Crick model (Figure 1.10a), in the early 1990's.

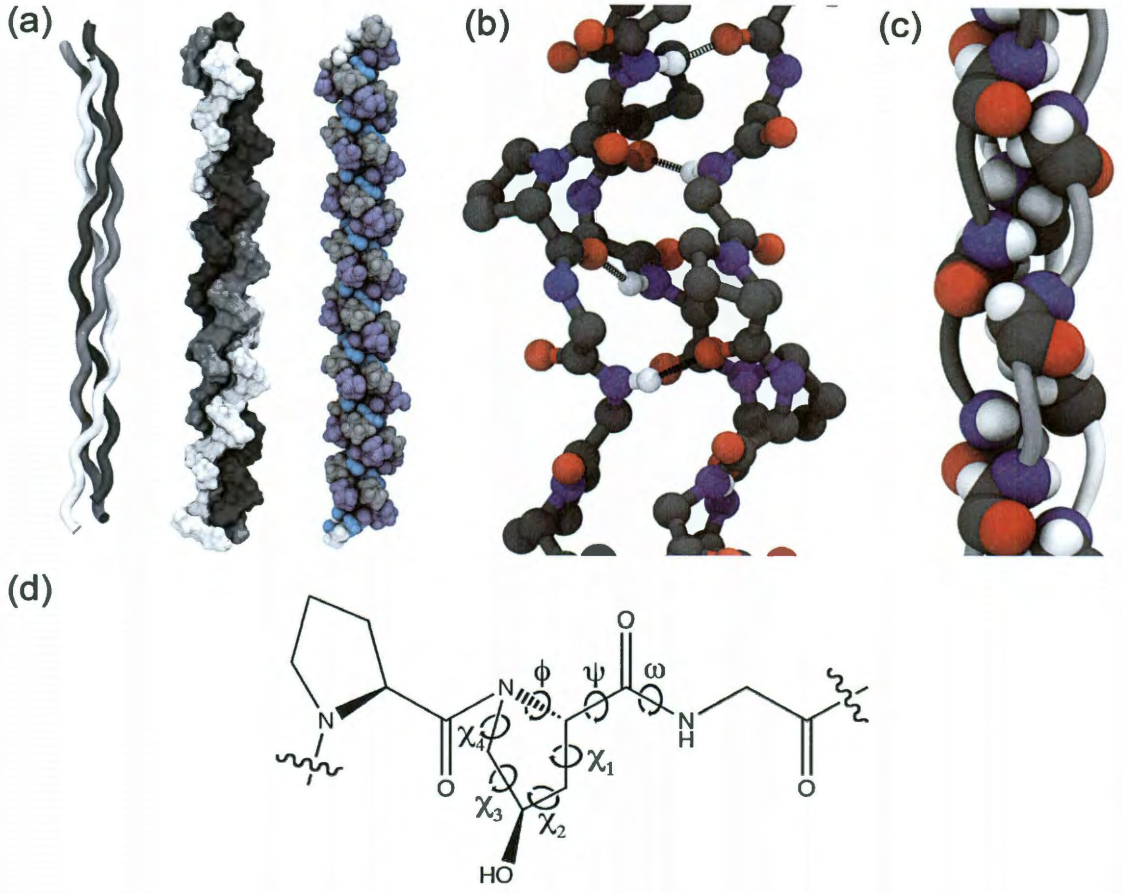

Figure 1.10. Triple helical structure of $(POG)_{10}$ CMP. Adapted from Figure 1 in reference 103.[103] (a) Cartoon, surface and space filling models of the $(POG)_{10}$ structure (gray = P, purple = O, cyan = G). (b) Hydrogen bonding network along the backbone of the CMP, highlighted by dotted lines. (c) Glycine interactions at the core of the triple helix. The chains are colored as follows in (a) and (c): leading strand = dark gray, middle strand = light gray, trailing strand = white. (d) Schematic of the prototypical POG sequence highlighting the dihedral angles. All structural models are based on the $(PPG)_{10}$ crystal structure (PDB 1K6F)[104] and modified, following the protocol described in Fallas *et al.*,[105] to include hydroxyproline.[103]

As briefly mentioned in the Collagen Assembly section above, the primary structure of the triple helix contains a three amino acid repeating unit (X-Y-Gly), requiring glycine as every third amino acid. Each of the component α chains forms a polyproline type II left-handed helix that tightly packs with two other chains to form a right-handed super-helix that is stabilized van der Waals interactions as well as through a hydrogen-bonding network. The hydrogen-bonding network follows along the backbone of the peptide chains, perpendicular to the helical axis, and each bond is between the backbone carbonyl group of the amino acid in the X position in one chain and the amide nitrogen of glycine in a different chain (Figure 1.10b). To allow for the tight packing of the α chains within the triple helix and to maximize the contact area for the van der Waals interactions, the glycine residue, which is every third amino acid based on the three amino acid repeating unit of collagen, is always facing towards the core of the helix (Figure 1.10c). In order for glycine to be present in every cross-section taken perpendicular to the helical axis, the peptide chains are forced to assemble in a staggered manner, offset by a single amino acid.

The large predominance of imino acids in the X and Y positions of the collagen repeating triplet can be understood in terms of its structural consequences. Within proline and its derivative, the pyrrolidine ring can adopt two different conformations (*endo* or *exo*) that fix the $\chi 1$ and ϕ dihedral angles. The $\chi 1$ dihedral is the torsion angle that describes the rotation along the C_α-C_β bond in the amino acid side chains while the ϕ dihedral describes the rotation around the N-C_α bond in the peptide backbone. For the *endo* conformer of the pyrrolidine ring, the dihedral values for $\chi 1$ and ϕ are 19° and -75° respectively. The *exo* conformer has a $\chi 1$ value of -6° and a ϕ value of -68.[106] Within the

repeating triplet of collagen, residues in the X position display ϕ values coinciding with the *endo* conformer and conversely, amino acids in the Y position exhibit ϕ values that agree with the *exo* conformer. Based on these observations, it is not surprising to find that the pyrrolidine ring in proline, the most common amino acid in the X position of collagens, has a slight preference to adopt the *endo* conformation and that the *exo* conformation is preferentially adopted by 4*R*-hydroxyproline, the most common amino acid found in the Y-position in collagens. Figure 1.10d depicts a schematic representation of a POG triplet highlighting the dihedrals discussed above.

Despite sharing the same structural elements, triple helical models derived from native collagen show a significant historical difference in the helical symmetry when compared to those derived from CMPs. The Rich and Crick model that was proposed on the basis of fiber diffraction data is a 10/3 helix, a helix with 10 scattering units (30 amino acids) and 3 helical turns in each axial repeat,[88] with an axial repeat (or fiber period) of 28.6 Å. Contrastingly, the single-crystal diffraction structures derived from CMPs exhibit 7/2 helical symmetry, 7 scattering units (21 amino acids) and 2 helical turns in each axial repeat,[88] with an axial repeat of 20 Å. Within the two models, both helices have the same unit height, or translation along the helical axis per tripeptide unit of 2.86 Å,[107] but the two models have different unit twists with the 7/2 model being more tightly wound than its 10/3 counterpart (Figure 1.11). A major issue with the difference in helical symmetry is that the fiber diffraction data does not contain enough reflections to unambiguously determine the axial repeat and the data can be explained with either the 10/3 or 7/2 model,[108] making both structures suitable candidates for the correct structure of native collagen.

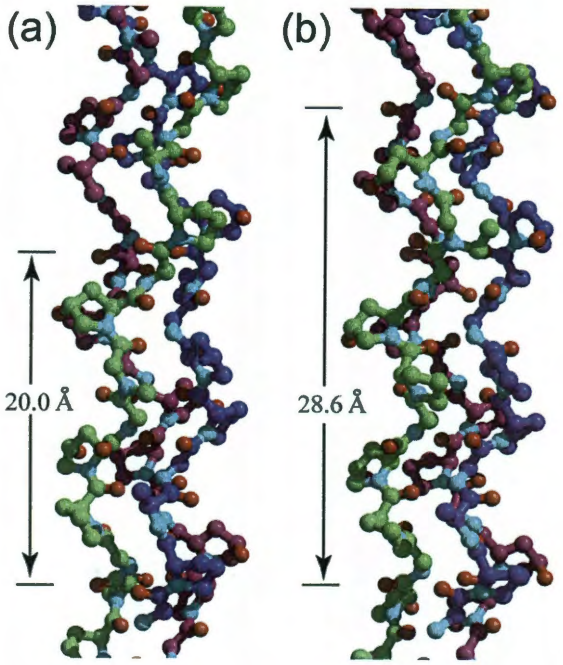

Figure 1.11. Triple helical symmetry. Adapted from Figure 2 in reference 103. (a) 7/2 helix and (b) 10/3 helix in which the three peptide chains within the triple helix are shown in pink, blue and green.[103]

Some light was shed on the question of helical symmetry with the solution of high-resolution structures of CMPs that included sequences taken from natural collagens. These peptides contain N- and C-terminal flanking regions to drive the formation of stable triple helical assemblies and, in the center of the peptide, include any amino sequence taken from a natural collagen. An example of such a peptide is T3-785[109] that contains amino acids 785-796 from human collagen type III, with the sequence ITGARGLAG, and is flanked by three repeats of the POG sequence at the N- and C-termini. The guest region of this peptide, which contains a very low content of imino acids, agrees more closely with the looser 10/3 model than the 7/2 model for helical symmetry.[110] Conversely, the imino-acid rich POG flanking regions show a 7/2 helical symmetry which complements the imino acid rich peptides that were used for previous

crystallographic studies.[76,77,104] Since the publication of the T3-785 peptide, other crystal structures that include guest sequences from natural collagens have been solved and show a similar helical symmetry pattern.[111,112] Therefore, both models for helical symmetry are correct based on experimental evidence and the helical symmetry within the collagen triple helix varies along the length of the fiber depending on the imino acid content of a particular region.

The main method for analysis of the conformation of CMPs in solution is multi-dimensional NMR experiments. Using the prototypical POG sequence as a template, Brodsky *et al.*[106] identified two sets of spin systems for each residue type with 80% of the triplets being in a similar chemical environment and thus having identical chemical shifts. Furthermore, using nuclear Overhauser effect spectroscopy (NOESY), the topology of the molecule was examined and when comparing the NOE cross-peaks measured for these triplets with the ones expected based on the fiber diffraction model, the authors concluded that the solution conformation of the peptide is similar to that present in the natural fibers.[106] The remaining 20% of the triplets that did not exhibit identical chemical shifts were assigned to the termini of the molecule, a less ordered form of the helix.

1.4.3.2. Triple Helical Stability and Folding

In the same way that CMPs have been integral in uncovering key features of the triple helical structure, they have also helped reveal the major stabilizing forces within triple helices. The central feature of CMPs used to compare their stabilities is the unfolding temperature, which has most commonly been measured through CD, ORD melting studies or calorimetry experiments. As previously mentioned, proline and 4*R*-

hydroxyproline are the two most abundant amino acids in the X- and Y-positions of the collagen repeating triplet, respectively. The importance of hydroxylation, specifically hydroxyproline and not hydroxylated lysine (another common posttranslational modification found in natural collagens) in triple helical stability was initially reported over 25 years ago in thermal melting studies performed on digested collagen type I.[113] Shortly after, the necessity for the hydroxyproline residue to be in the Y-position within the collagen triplet as opposed to the X position was proven by the inability of the CMP $(OPG)_{10}$ to form a triple helix despite the fact that both $(PPG)_{10}$ and $(POG)_{10}$ form stable helices.[114] Following this work, two leading arguments emerged to explain the stabilizing forces provided by the presence of hydroxyproline in the Y-position of the collagen triplet: stereoelectronic effects and hydrogen bonding. Schematic representations of the two theories are shown in Figure 1.12.

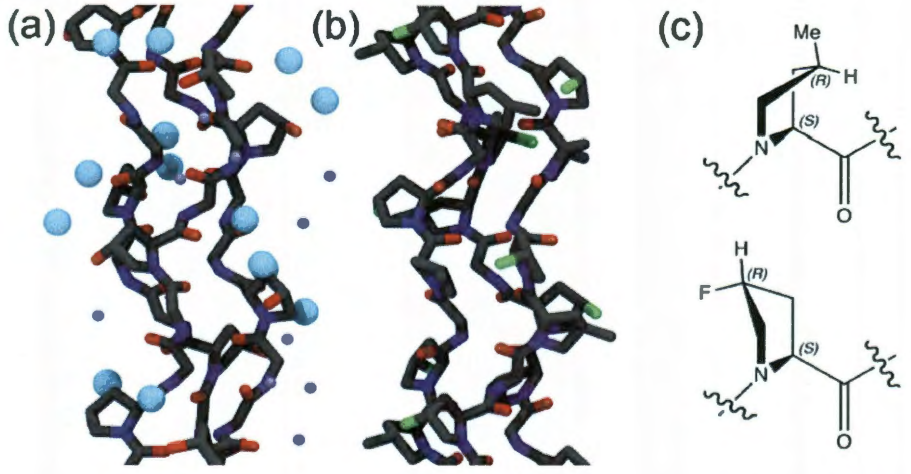

Figure 1.12. Triple helical stabilization mechanisms. Adapted from Figure 3 from reference 103. (a) Hydration network of a triple helical peptide containing hydroxyproline (PDB 1CGD).[115] Water molecules involved in inter-strand bridges are represented as large cyan spheres and water molecules involved in intra-strand water bridges are represented by small purple spheres. (b) Structure of a triple helical peptide containing 4*R*-methylproline and 4*R*-fluoroproline (PDB 3IPN).[116] (c) The *endo* ring pucker can be observed in 4*R*-methylproline residues and the *exo* ring pucker in 4*R*-fluoroproline residues (fluorine is depicted green).[103]

The first leading explanation for the stabilizing effects of hydroxyproline is the presence of hydrogen bonding networks in hydroxyproline containing CMPs. First proposed in 1973 by Ramachandran *et al.* and later confirmed in 1980, the existence of hydrogen bonding networks in native collagen is well known.[117,118] However in order to better understand the role of these networks, CMPs were employed to examine the amino acids involved. In 1994, Brodsky reported a high resolution x-ray crystal structure of a derivative of $(POG)_{10}$ containing a single glycine to alanine mutation in the center of the peptide.[102] Along with several observations on the puckering of the proline and hydroxyproline residues, *endo* and *exo* respectively and the unraveling of the peptide at the termini, the most significant observation from the 1.9 Å resolution structure was a cylinder of hydrogen bonding in which hydroxyproline bonds with water instead of simply with other amino acids. Therefore, it was suggested that the hydroxyl group present in hydroxyproline allows for hydrogen bonding with water thus stabilizing the CMP triple helix. In 2001, the x-ray crystal structure of a CMP containing a sequence from collagen type III confirmed the existence of water-mediated hydrogen bonds that connect the N-H of the amino acid in position X of the collagen triplet to the carbonyl oxygen of the glycine residue.[109] Thus through a series of x-ray crystal structures on CMPs, the importance of the hydrogen bonding networks in triple helical systems were confirmed and more importantly, that they are enhanced by water-mediated interactions and the ability of hydroxyproline to form bonds with water, not just other amino acids.[119]

In contrast, Raines has argued that the stabilizing effect of hydroxyproline is largely a stereoelectronic effect.[120-123] To understand his work, one must note that the rate-limiting step for the folding of collagen triple helices is the cis-trans isomerization of

all proline and proline derivative amino acids.[69] Therefore, a CMP designed to drive the equilibrium of isomerization towards the trans state will have faster folding of the peptide chains into triple helices.[119] Through a series of papers comparing proline, 4*R*-hydroxyproline and 4*R*-fluoroproline (Flp), Raines explored the inductive effects that stabilize CMPs hypothesizing that the presence of electron-withdrawing groups such as hydroxyl or fluorine on proline drives a structural preference for the trans conformation of the amino acid thus speeding up the folding of triple helices. Beginning in 1996 with a study exploring the acetylated, methyl ester derivatives of the three amino acids, Raines proved that the presence of an electron-withdrawing substituent on proline induces structural changes which are observable as shifts in the amide I vibrational mode seen in the infrared (IR) spectra and reduces the isomerization energetic barrier which has been calculated by NMR spectroscopy.[121] In 2010, the difference in cis/trans ratios for different proline derivatives was explained by the presence of a favorable interaction between the oxygen on the carbonyl of amino acid *i* and the carbonyl carbon of amino acid *i*+1.[116,120,124] The effect of cis-trans isomerization and hydrogen bonding in triple helical assemblies has also been explored by the use of isosteric alkenes in place of amide bonds.[125] Despite locking the conformation around the double bond, the isosteres induce changes in the hydrogen-bonding network and the results of the experiments are complex and difficult to interpret.[126,127]

The next logical step in examining the stereoelectronics in CMPs was the incorporation of fluoroproline into CMPs to compare the thermal stability to $(PPG)_{10}$ and $(POG)_{10}$. As expected, the stability of the CMPs was highest in $(Pro\text{-}Flp\text{-}Gly)_{10}$ with a melting temperature of 91 °C and lowest in $(PPG)_{10}$ which melted at 41 °C, with $(POG)_{10}$

in-between the two at 69 °C.[122,123] As mentioned in section 1.1 the pucker of the pyrolidine ring in imino acids controls the ϕ backbone dihedral and can adopt two different conformations, one, *endo*, that suits the requirements of the X position and one, *exo*, that is more apt for the Y position of the triple helix. Proline shows a slight tendency to adopt the *endo* conformer with a ϕ dihedral corresponding to the X position. This tendency can be reinforced by either adding a bulky substituent, such as a methyl group, to the γ-carbon in the R configuration (Figure 1.12c) or an electronegative group, like fluorine or a hydroxyl group, in the S configuration. On the other hand, electron-withdrawing groups, like the hydroxyl and fluoro groups in 4*R*-hydroxyproline and 4*R*-fluoroproline with R configuration favor an *exo* ring pucker (Figure 1.12c) that pre-organizes the ϕ value towards the Y position of the triple helix. Overall, by choosing the right stereochemistry in the proline derivatives to drive the *endo* and *exo* conformations in the X and Y positions respectively, one can pre-organize the backbone dihedrals and reduce the entropic penalty for triple helical formation leading to a stabilization of the folded state. While the merits of hydrogen bonding versus stereoelectronics have been argued as competing theories for the stabilization of the collagen triple helix, both likely play a role with stereoelectronics as the main component contributing to the higher stability seen in hydroxyproline containing systems.

While examining the hydrogen-bonding network surrounding hydroxyproline in CMPs, Brodsky *et al.* explored how single amino acid mutations affect the stability of CMPs. In a series of papers, the effect that single mutations within the $(POG)_8$ CMP template have on the triple helical melting temperature was investigated for all natural amino acids in the X and Y positions within the collagen triplet.[128,129] Most notable from

these results is first, the correlation between stabilities seen and the occurrence in fibril-forming collagens of the residue in the specific position of the collagen triplet and second, the predominance for negatively and positively charged amino acids to have higher melting temperatures in the X and Y positions respectively. The X position is known to be more solvent exposed[130,131] and thermodynamic calculations performed on the 1996 melting data of hydrophobic residues showed that the preference for the X position is entropically driven.[129] However, the driving force for the prevalence of negatively charged residues in this position was initially unclear. In order to explore this question, CMPs containing two mutations in the $(POG)_9$ template were synthesized using charged and hydrophobic amino acids.[132] The experimental thermal stabilities collected from CD melting studies were compared to the predicted unfolding temperature based on the contributions from individual triplets. The most surprising result from this study was the high stability seen for the CMPs containing GPKGEO and GPKGDO. In both peptides, lysine is the positively charged amino acid in the Y position and a negatively charged residue, glutamate or aspartate, is in the X position. Computer modeling on the GPKGEO peptide (shown in Figure 1.13) revealed the close proximity of the side chains of the charged amino acids allowing for cross-chain hydrogen bonding. This work laid the framework for a large portion of current research on synthetic collagen due to proof-of-principle success on engineering CMPs that can form side-chain stabilizing interactions.

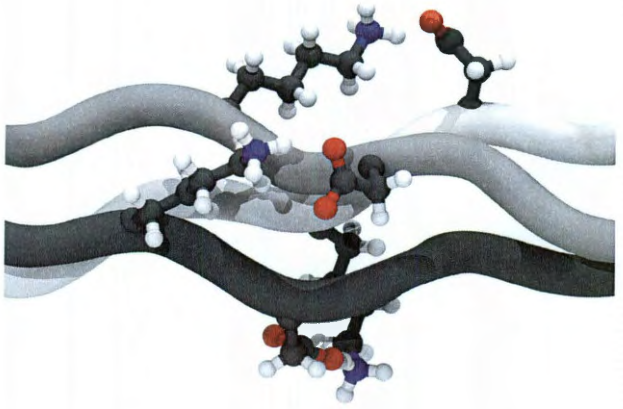

Figure 1.13. Computer models of pairwise interactions between lysine and glutamate made to explain the increased thermal stability of GPKGEO-containing homotrimeric CMPs.[132] Adapted from Figure 4 in reference 103 that referenced the original publication of this cartoon as Figure 4 in reference 126.[103,132] The model is based on the $(PPG)_{10}$ crystal structure (PDB 1K6F),[104] modified following the protocol described in Fallas *et al.*,[105] to include charged residues. [103,132]

A vastly different approach that has been successful in studying sequences with low triple helical propensities is the templated assembly of triple helices. This method utilizes small organic molecules attached to the peptides to drive the trimerization of the helix by forcing them into close proximity with one another. Organic molecules that contain three reactive moieties such as cis,cis-1,3,5-trimethylcyclohexane-1,3,5-tricarboxylic acid (KTA),[133,134] tris(2-aminoethyl)amine (TREN),[135,136] tris(hydroxymethyl)aminomethane (TRIS),[137,138] and triserine lactone (TSL)[139] have been used to facilitate the close proximity of peptide chains and promote their assembly into triple helices (see Figure 5 for chemical structures). An interesting result from the use of such scaffolds, particularly KTA, has been the successful incorporation of peptoid N-isobutylglycine in the X and Y position of a triple helix which demonstrates that the rigid triple helix is capable of accommodating peptoids in its structure.[140-144] Other related approaches include the use of disulfide bridges,[145,146] lysine-lysine cross-linking,[147] metal

ion coordination sites[148,149] and non-covalent induction of aggregation by alkyl chains[28,150] to promote triple helical nucleation.

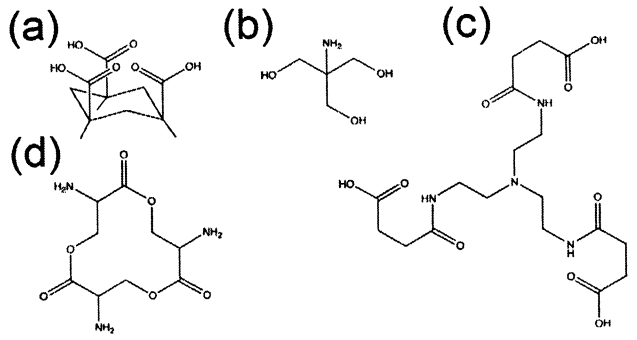

Figure 1.14. Chemical structures of small molecules used for the templated assembly of triple helical peptides. Adapted from Figure 5 in reference 103. (a) KTA, (b) TRIS, (c) TREN coupled with succinic acid and (d) Triserine Lactone.[103]

In addition to gaining insight on the structure and stabilization of collagen, CMPs have also been used to study the folding pathway of triple helical assemblies. Major findings in this area include the slow folding time scale for the triple helix, ranging from minutes to hours, and the concentration dependence of the free energy landscape of triple helical peptides. Early attempts that used CD spectroscopy for analysis determined third order kinetics for the folding of the $(PPG)_{10}$ peptide[151] as well as for a 36 residue fragment derived from rat-skin collagen analyzed after cyanogen bromide digestion.[152] The sequence dependence of the folding kinetics was demonstrated using host-guest CMPs that contained a guest sequences in the fourth triplet of a $(GPO)_8$ template.[153] These host-guest experiments also showed third order kinetics at low concentrations and a dependence of the rate-limiting step on the sequence of the guest triplet. Peptides containing hydroxyproline in the guest triplet demonstrated the fastest folding rate. Later studies utilized heteronuclear NMR to follow the triple helical folding which allowed for

more realistic models due to the enrichment of ^{15}N at strategic positions along the guest region of the T3-785 peptide (described above) to monitor the appearance of triple helical peaks and the decrease in the monomer peaks.[154] The data could be fit to a simple nucleation/growth mechanism that demonstrated a zipper-like folding where the peptides nucleate at the C- or N-terminus and then propagate towards the opposite end. The nucleation step for folding in the T3-785 peptide was assigned second order kinetics in which a two-step nucleation mechanism was proposed beginning with the fast formation of an unfolded dimer followed by the addition of the third chain as the rate limiting step.[155] The propagation of the nucleated species was assigned first order kinetics, with the cis-trans isomerization of proline peptide bonds as the rate-limiting step.[156]

Further analysis using the ^{15}N-enriched heteronuclear NMR approach was carried out with peptides that only contained a nucleation sequence (four POG repeats) at the C-terminus, followed by a natural sequence taken from collagen type I α1 chain, the T1-892 peptide with the sequence $(GPA)_2GPVGPAGARGPA(GPO)_4GV$.[157] At high concentrations, the kinetics of T1-892 was shown to be biphasic and the folding to be initiated at the C-terminus and directional. The initial phase was fast in the NMR time-scale but the slow phase could be fit to first order kinetics for both the nucleation and propagation steps, indicating that the *cis-trans* isomerization is the rate-limiting step for both events because only peptides with all *trans* conformation in the C-terminal domain are able to nucleate[158] In fact, the population of folded peptides in the first data point agrees well with the amount of peptide in all *trans* conformation that is expected to be present in the unfolded ensemble, indicating that the high-order nucleation step is fast in comparison to the *cis-trans* isomerization at high concentrations.[157] The situation is

different at low concentrations (see below).[157] NMR experiments also showed the formation of off-path folding intermediates for the T1-892 peptide, a partially unfolded metastable state that disappears when the sample reaches thermal equilibrium.[157] Accompanying CD experiments were performed at lower concentrations and showed an all-or-none third order reaction, where all peptide bonds in the nucleation domain must be in the trans conformation to allow the peptide to fold.[69] Because of the lower concentration, the *cis-trans* inter-conversion is faster than the nucleation event and thus not the rate limiting. The C- to N-terminus directionality and *cis-trans* isomerization barrier[159] observed for the T1-892 peptide matches the folding pathway proposed for type III collagen.[160]

1.4.4. Heterotrimeric Collagen Mimetics

Despite some of the most abundant collagens in nature being heterotrimeric, including types I, IV and VIII, the first heterotrimeric CMPs were not reported until the mid 1990's using Lys-Lys covalent tethering,[161] about thirty years later than their homotrimeric counterparts. In the last ten years, different strategies have been used to synthesize heterotrimeric CMPs to study collagen degradation, integrin binding and connective tissue diseases. The major approaches to heterotrimer synthesis and findings from these studies will be discussed in the next two sections.

1.4.4.1. Covalently-Linked Heterotrimers

The most common approach for the covalent synthesis of heterotrimeric CMPs has been regioselective disulfide bridging. The synthetic procedure has been described in detail elsewhere[162] and is depicted in Figure 1.15. For this approach, a single cysteine residue is required in the leading and in the lagging peptide strands, as well as two cysteines in the middle strand. The general procedure involves the protection of the two cysteines in the middle chain using the tertbutylthiol (StBu) and acetamidomethyl (Acm) groups, meanwhile the cysteines in the leading and lagging strands are protected with the StBu functionality. The StBu derivatives can be readily deprotected by reducing the disulfide bond using tributylphosphine (PBu_3) to yield free thiols. The free thiol in the middle strand is then activated using 2,2'-dithio-di(5-nitro)pyridine ($Npys_2$) to yield the 5-nitropyridil-2sulefenyl (Npys) cysteine derivative which can readily react with a free thiol from another strand leading to the covalent cross-linking of two of the peptides. The Acm group in the second position is subsequently deprotected using Npys-Cl to yield the Npys derivative in the second position, which in turn readily reacts with the free thiol in the third strand completing the heterotrimer. This approach has been successful in synthesizing heterotrimers with POG triplets at the N-terminus as triple helical nucleating sequences followed by guest sequences from heterotrimeric collagens with the register fixed by the C-terminal cysteine knot.

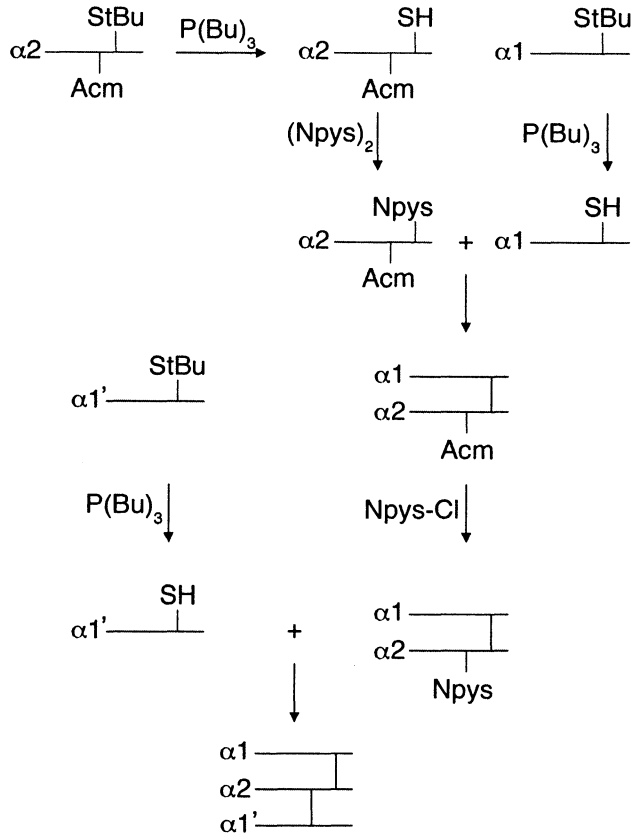

Figure 1.15. Regioselective cysteine chemistry for heterotrimeric CMP synthesis. Adapted from Figure 10 in reference 103 that was originally published in reference 163. Full names of protecting groups and reagents are available in the Covalently Linked Heterotrimers section.[103,163]

One of the sequences studied using this approach corresponds to the matrix metalloprotein-1 (MMP-1) and MMP-8 cleavable site from type I collagen, which includes residues 772-784. Collagenases, which include the MMPs listed, perform a highly selective and conformationally dependent scission of the peptide bonds between glycine and leucine in the α1 chain and glycine and isoleucine in the α2 chain (amino acids 775 and 776). Peptides were designed to include the type I cleavage sequence at the C-terminus and an N-terminal nucleation domain, $(POG)_n$, with values for n of 3 and 5. The peptides were fixed in the α1α2α1 register using the cysteine knot strategy described

above and then characterized via CD spectroscopy and NMR.[164,165] All the heterotrimers fold into triple helices however the thermal stability ranges from 9 to 41 °C depending on the length of the nucleation sequence. The NMR experiments demonstrated the partially ordered structure of the C-terminal domain of the longer peptide[166] and enzyme digestion assays performed at room temperature on the peptides showed that MMP-8 proteolysis of the partially-folded CMP is efficient but the completely unfolded trimer shows a very slow process. Conversely, MMP-2, also known as gelatinase A, cleaves the unfolded peptide preferentially.[167,168]

The integrin binding sequence of the most abundant form of collagen type IV, an AAB heterotrimer, is located between residues 457-468 and amino acids in different chains are thought to be important for binding. Peptides containing the type IV sequences in the center flanked by three POG repeats at the N-terminus and two at the C-terminus were synthesized to study the binding of this type IV sequence to the $\alpha_2\beta_1$ integrin.[169] Using the cysteine knot approach described above, two different heterotrimer registers were synthesized and tested for their integrin binding affinity. The experiments showed the slightly higher binding affinity for the α2α1α1 register than the α1α2α1 register.[170] These results complemented fluorescence resonance energy transfer (FRET) measurements of cyanogen-bromide derived fragments extracted from type IV collagen that proposed the preference for the α2α1α1.[171] To add on to the superiority of the α2α1α1 register, the thermal stability and folding rate was also found to be register dependent with the α1α2α1 register showing a higher melting temperature and higher rate constant.[172] The difference in melting temperature between registers can be partially explained by the work on self-assembled heterotrimers driven by electrostatic interactions

that will be described in the next section. The sequences used to study $\alpha_2\beta_1$ integrin binding contain several lysine and aspartate residues, however only when in the α1α2α1 register do they have the correct three-dimensional arrangement to form an inter-chain salt-bridge similar to the ones that drive the formation of the heterotrimers discussed below.

1.4.4.2. Self-Assembled Heterotrimers

Self-assembly of heterotrimers have two challenges that are not faced by comparable homotrimeric assemblies: control over composition and control over register. As described in the Collagen Assembly section above, heterotrimers can be either AAB or ABC depending on the number of unique peptide chains used. In both cases, compositional control is required such that the desired triple helix composition is favored while the undesired composition is disfavored. For example, peptides A and B may form the two homotrimers, A_3 and B_3, or they can associate to form either A_2B or AB_2 heterotrimers. For ABC heterotrimer formation, a mixture of A, B and C peptides is even more complicated, potentially leading to any of ten different triple helical compositions including three homotrimeric helices (A_3, B_3, and C_3), any of six possible two component heterotrimers (A_2B, AB_2, A_2C, AC_2, B_2C, and C_2B) or the desired ABC heterotrimer. Compiling onto the barrier of compositional control is issue of register. The three peptides of the collagen triple helix are offset from one another by a single amino acid creating leading, middle and lagging strands. In a homotrimer, since all peptides are identical, it makes no difference which peptide is in which position. However for heterotrimers, each composition can be formed in several different registers. For example

a triple helix with the composition AAB could be any of three different registers (AAB, ABA or BAA) while a triple helix with the composition ABC could be any of six different registers (ABC, ACB, BAC, BCA, CAB, or CBA).

Early studies performed on heterotrimeric self-assembled CMPs had some major shortcomings including low thermal stability[173] and lack of specificity with respect to composition and register. An annealed mixture of $(POG)_{10}$ and $(PPG)_{10}$ peptides shows the formation of heterotrimers however there is no control over the composition of the assemblies.[174] Brodsky *et al.* synthesized a host-guest peptide system with POG flanking regions and sequences from the α1 and α2 chains of collagen type IV in the interior that forms an AAB type heterotrimer visible by CD, DSC (differential scanning calorimetry) and NMR. Although this was one of the first examples of a self-assembled AAB type collagen triple helix, a significant drawback with this design was the low thermal stability of the system, 14.5 °C. Longer peptides containing type IV guest sequences that used C-terminal cysteine knots to stabilize the self-assembled heterotrimer through disulfide bridges showed a mixture of several registers when analyzed by solution NMR.[175] Raines *et al.* were able to form a self-assembled heterotrimer from $(PPG)_7$ and a peptide containing fluoroproline derivatives with the appropriate stereochemistry for the X and Y positions.[176] Despite having a large degree of pre-organization, the fluoroproline containing peptide failed to fold into a stable triple helix due to unfavorable steric interactions arising from the fluorine atoms. The inclusion of $(PPG)_7$ within the system allows the mitigation of this effect by increasing the spacing between fluorine atoms and the two peptides form a triple helix in a 2:1 ratio of the fluoroproline containing peptide

to $(PPG)_7$.[176] While having control over the composition, the ability of the system to control register is still unknown.

A different approach to the problems of compositional and register control uses oppositely charged amino acids to drive the self-assembly of a stable heterotrimeric CMPs has been successful in the Hartgerink group to minimize many of these issues. Experiments mixing three peptides, one with charge $+\eta$, one with charge $-\eta$ and the third with a neutral peptide, show that a zwitterionic ABC triple helix can be formed (Figure 1.16). The effects of different values for η were inspected for CMPs with 10 triplets using arginine and glutamate as the charged moieties and the prototypical POG sequence as the neutral species.[177] The peptides were designed to include the positive amino acid in the Y position of the $(X\text{-}Y\text{-}Gly)_n$ triplet and the negative amino acid in the X position because this arrangement is most commonly found in natural collagens.[128] From these experiments, it was found that the peptides with $\eta = 10$ including 20 mutations from the prototypical $(POG)_{10}$ sequence formed the most stable ABC heterotrimers. In order to optimize the heterotrimeric assembly, other plausible charged pairs were studied following the same design pattern and it was found that lysine–aspartate interactions provide the most stable heterotrimer, with a melting temperature only a few degrees below that of a $(POG)_{10}$ homotrimer, followed by lysine-glutamate interactions then arginine-glutamate interactions and arginine-aspartate interactions forming the least stable heterotrimers.[177,178]

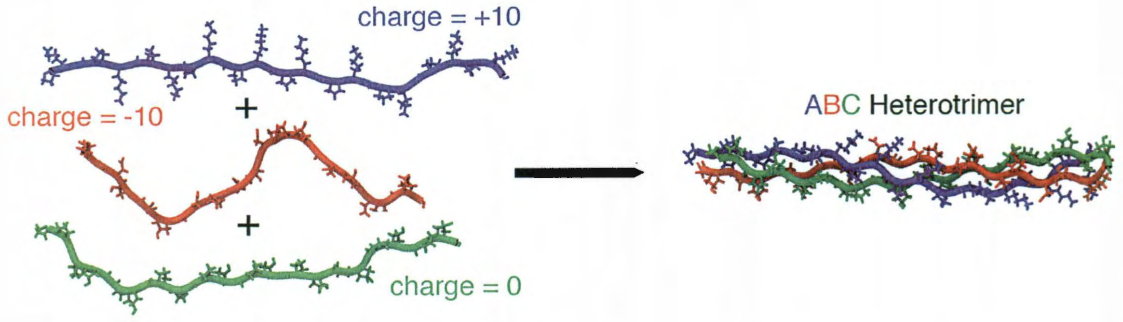

Figure 1.16. Strategy for the self-assembly of ABC heterotrimeric CMPs utilizing electrostatic interactions. Adapted from Figure 11 in reference 103.[103]

In order to do a more thorough analysis of the high stability $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ system, solution NMR experiments were carried out to determine the composition and fold of an annealed mixture of these peptides. Using ^{15}N- and ^{13}C-uniformly-labelled amino acids at strategic positions in the core of the helix as spectroscopic probes, the experiments showed that the major species was the expected $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ heterotrimer with some residual 3•$(POG)_{10}$ homotrimer as well as monomers of the charged species were present in small quantities.[105] The NOESY spectrum showed features characteristic of triple helical assemblies, which confirmed the molecular topology of the assembly. Furthermore, the NMR experiments showed that the long lysine side chains adopt a conformation that is ideal to form ionic hydrogen bonds with the aspartate in a neighboring chain. Specifically, the salt-bridge conformation goes from the lysine in triplet *n* to the aspartate in triplet *n+1* of the neighboring chain. An unexpected, but important, result is that only *one* of six possible collagen registers was observed. Because the amide nitrogen and hydrogen chemical shifts in triple helices are very sensitive to changes in their hydrogen-bonding network,[174,175] more heterogeneous spectra would be expected if more than one register of the ABC heterotrimer was present

in the mixture. The register of the helix was determined to be $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$, with the $(PKG)_{10}$ occupying the leading position (N-terminal), the $(DOG)_{10}$ chain in the middle position and the $(POG)_{10}$ chain in the lagging (C-terminal) position. The information gathered from the NMR experiments lead to a full structural determination, leading to the first solution structure of a heterotrimeric CMP (Figure 1.17).

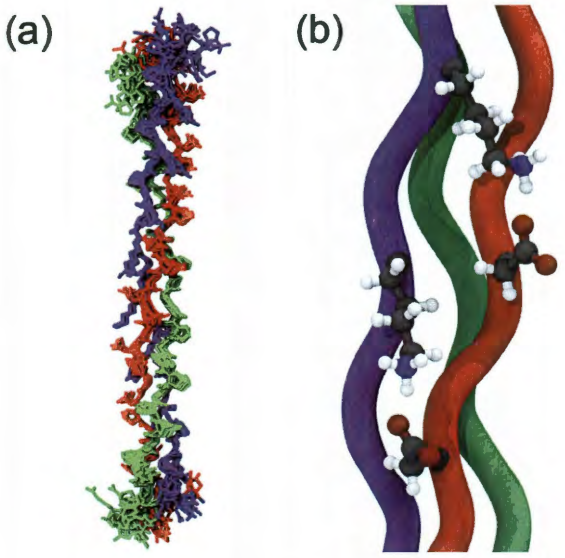

Figure 1.17. Solution structure of an ABC heterotrimer. Adapted from Figure 12 in reference 103. (a) 15 lowest energy conformers superimposed (PDB 2KLW).[105] (b) Section of the helix highlighting inter-chain ionic hydrogen bonds. Peptide chains in (a) and (b) are shown as $(PKG)_{10}$ in blue, $(DOG)_{10}$ in red and $(POG)_{10}$ in green.[103]

The $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ heterotrimer has also been used as the basis to design a novel host-guest system to study the effects of *OI* mutations on the thermal stability and folding rate of heterotrimeric triple helices, which mimic collagen type I.[179] In the new host-guest system, the flanking regions are composed of five triplets of the designed sequences, which direct triple helical self-assembly though electrostatic interactions, and the guest region consists of nine amino acids from type I collagen. This

new model system allows for both α_1 and α_2 sequences to be included and thus the effects of one or two mutations in the triple helix to be assessed. This new host-guest system contrasts homotrimeric models, which can only display zero or three mutations, making the new heterotrimeric host-guest model a more accurate representation of natural collagens.[180] The results of the initial *OI* study showed that the first mutation causes a drastic drop in thermal stability but subsequent mutations, although still lowering the thermal stability, have a less pronounced effect.

In a more recent publication, the use of computational protein design was applied to CMPs to create peptides that will selectively form a desired heterotrimer composition over any undesired homotrimer or heterotrimeric species.[181] The initial work utilizing this design focused on arginine and glutamate containing peptides and unfortunately did not display success in driving the selective formation of the desired heterotrimer. However, through the incorporation of solution NMR data, specifically from the $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ system,[105] the computational design technique for CMP creation recently yielded a lysine-aspartate based system that formed an ABC type heterotrimer.[182] Based on these results, the computational technique has a large potential for yielding results that can demonstrate both compositional and register control.

1.4.5. Higher Order Assemblies of Collagen Mimetics

In native collagen, five triple helices pack in a quasi-hexagonal lattice to form fibrils that further self-assemble both axially and laterally into mature collagen fibers.[68,183] The staggered packing of fibrils within the fibers yields a characteristic banding pattern, called D-banding. D-bands can have a length of 64-67 nm depending on

the helical twist of the fibrils within the fibers. This highly organized multi-step assembly is unique to native collagen and has been one of the largest obstacles in synthetic collagen research. Figure 1.18 is a scheme that describes the three types of triple helices that collagen forms and what progress has been made in replicating the hierarchical self-assembly of collagen into nanofibers and hydrogels using CMPs. Research that has already been published is shown in black and areas for current and future research are shown in red. Although the row corresponding to homotrimeric systems is shown in black, no previously reported system has demonstrated discretely each level of collagen self assembly. Nonetheless, the large amount of red in the scheme leads to the conclusion that the hierarchical assembly of CMPs is a wide-open area within synthetic collagen research and is at the forefront of the field, especially for heterotrimeric systems.

CMP systems have been previously shown to self-assemble into thermally stable homotrimeric and heterotrimeric (both AAB and ABC types) triple helices. At a high enough concentration and after a long enough incubation, all peptide solutions will eventually precipitate out no matter the buffer. The structure of these precipitates can be determined using atomic force microscopy (AFM), scanning electron microscopy (SEM) and transmission electron microscopy (TEM). In 2006, Brodsky *et al.* reported the high order structure of $(POG)_{10}$ including an electron micrograph (shown in Figure 1.19b) in the publication.[184] It is clear that multiple triple helices aggregated together to form the mesh-like assemblies seen, however, the absence of a driving force to organize the self-assembly of $(POG)_{10}$ prevents it from being able to mimic the hierarchical assembly of natural collagen and instead forms a relatively disorganized structure. The simple incorporation of sequences from native collagen (such as type IV collagen) into CMPs

with POG flanking regions is not enough to make ordered nanofibers despite the presence of charged and hydrophobic amino acids within these systems.[185] One hypothesis for the inability of these CMPs to form fibrils and fibers is the lack of "sticky-ended" structures, which have been used successfully in α-helical coiled-coil design to drive fiber formation.[9]

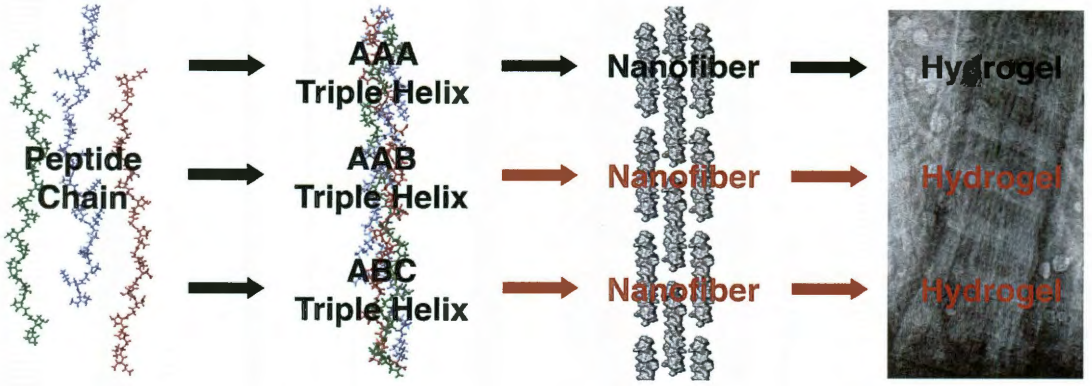

Figure 1.18. Scheme describing the three types of triple helices that collagen forms and what progress has been made in replicating the hierarchical self-assembly of collagen into nanofibers and hydrogels using CMPs. Adapted from Figure 13 in reference 103. Published research is indicated by the black text and areas for current and future research are highlighted in red.[103]

A method that bypasses the need for a sticky-ended moiety, and instead relies on the fibrillar nature of native collagen to serve as a template, is the work by the Yu lab at Johns Hopkins. In a series of papers, they showed the ability of CMPs with the prototypical $(POG)_x$ sequence to infiltrate the partially unfolded portions of collagen. The CMPs have been covalently linked to polyethylene glycol (PEG), a common polymer used in tissue engineering that prevents cell adhesion, to demonstrate the ability of the CMPs to bind to collagen films and change the surface properties of collagen films.[186] Later, the exact position of CMP binding along collagen fibers was discovered by using

gold-sulfur bonds to link the CMPs to gold nanoparticles and examining the system using TEM.[187] Yu concluded that there were twice as many nanoparticles in the gap regions of the collagen fibers as compared to the overlap regions. Most recently, the CMPs were copolymerized with PEG to form hydrogels that were shown to increase the collagen and glycosaminoglycan content of the hydrogel after a two week incubation with chondrocytes when compared to a PEG hydrogel without CMP modification.[188] Although this method does not rely on the fiber forming ability of the CMPs themselves, it does explore the ability of the CMPs to enhance other fibrous systems and enhance the biological compatibility of common tissue engineering scaffolds.

Many different approaches for driving organized fiber formation have been pursued. The incorporation of hydrophobic residues at the termini of CMPs have resulted in the assembly of nanoscale fibers in multiple systems.[189,190] One report examines the changes in fiber formation seen in systems driven by the presence of tyrosine and phenylalanine (Figure 1.19c) versus those containing L-pentafluorophenylalanine. The fluorine-containing system shows assembly that is stunted at the fibril phase. One explanation for this difference in nanostructure is the difference in the electronic properties of the hydrophobic residues in the presence and absence of electron-withdrawing fluorine atoms. In the non-fluorinated case, π interactions with proline and hydroxyproline were proposed as the main driving force for nanofibers seen in uranyl acetate stained TEM. Another report expanded on the use of hydrophobicity to drive fiber formation and assessed the thrombogenic potential of the CMPs to drive platelet aggregation due to the fact that platelets are known to adhere to exposed collagen in blood vessels during tissue repair. Although no AFM or light microscopy images of the

CMPs demonstrated long-range organized fiber-like nature of the self-assembly, the CMPs did have the ability to trigger platelet aggregation demonstrating the collagen-like nature of the system. The incorporation of biological experiments in the analysis of this CMP system expands the criteria for success that one may consider when analyzing a designed peptide system.

Another straightforward design technique for driving the formation of collagen nanostructures is the use of cysteine residues to form cystine knots that covalently link CMPs together. One approach placed cysteines in the interior of a long POG-based CMP that, when oxidized, covalently cross-linked the peptides together forming a hydrogel.[191] In this system, although gelation was seen, the presence of organized fibers was not confirmed. Another incorporation of cysteine into CMP design placed the residue at the C-termini of a POG-containing CMP for the first two peptides in the collagen triple helix and as a linker between two POG-containing sequences for the third peptide.[192] The concept was to oxidize the cysteines and form disulfide bridges between the three peptides that would form triple helices with a single peptide chain to serve as a "sticky-end" upon which another triple helix can assemble. The nanostructures seen in TEM resembled that of the $(POG)_{10}$ aggregates and are shown in Figure 1.19d. Utilizing a different approach, our group incorporated a cysteine residue at the N-terminus and thioester at the C-terminus to use native chemical ligation instead of disulfide bond formation to drive polymerization.[91] This allowed for selective head to tail polymerization of the peptides to occur and yielded highly uniform CMP nanofibers confirmed by negative stain TEM and is shown in Figure 1.19e.

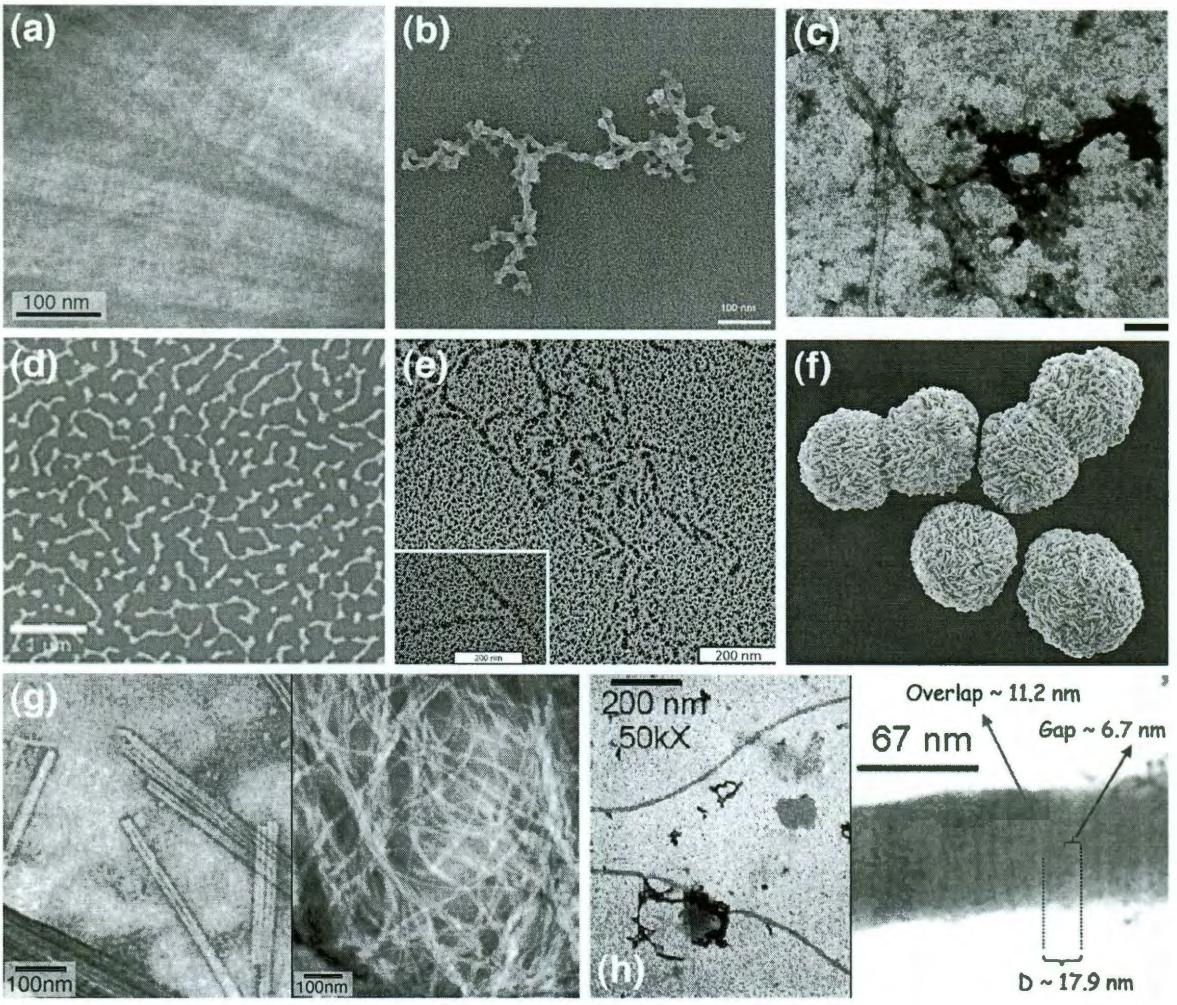

Figure 1.19. Hierarchical self-assembly of synthetic collagen systems compared to rat-tail collagen. Adapted from Figure 14 in reference 103.[103] (a) rat-tail collagen, (b) $(POG)_{10}$,[184] (c) phenylalanine and tyrosine containing CMP,[190] (d) hydrophobic containing CMP that triggered platelet aggregation,[189] (e) sticky-ended POG-based CMP with cystine knot,[192] (f) microflorettes assembled from CMPs,[193] (g) nanofibers formed by native chemical ligation[91] and (h) $(PRG)_4(POG)_4(EOG)_4$.[194]

Metal-ion mediated self-assembly, although a common technique for the assembly and cross-linking of many other peptide and polymer systems including α-helical coiled-coils[195] and multi-domain peptides,[46,54] had previously been unexplored in CMP design and assembly until Chmielewski reported the use of divalent metals to trigger the head-to-tail self-assembly of POG-based CMPs into micro-florettes.[193,196] By

incorporating nitrilotriacetic acid at the N-terminus and two histidine residues at the C-terminus, the $(POG)_9$ CMP, which was used as the peptide backbone for the system, assembles into organized nanostructures with unique morphologies depending on the metal ion identity and concentration (shown in Figure 1.19f). This method, however, deviates from the native assembly of collagen triple helices by the formation of microflorette structures instead of fibers.

Recently, a hydroxyproline-lacking CMP was reported to self-assemble into nanofibers using both electrostatic interactions and cystine knots to drive its formation.[197] The peptide design included charged residues within the CMP that were chosen based upon previous work on electrostatic interactions within homotrimers and cysteine residues placed at the C-terminus. Previously, CMPs designed without hydroxyproline either did not form triple helices or formed homotrimers with very low thermal stabilities. Therefore the design and success of this peptide system is promising due to the uniformity and organization of the nanofibers.

Finally, electrostatic interactions can be incorporated into homotrimeric design in order to drive fiber formation. In work by Chaikof and Conticello, the peptide, $(PRG)_4(POG)_4(EOG)_4$, was designed with a sticky-end motif such that at physiological pH, the positively charged arginine-containing N-terminal region would form electrostatic interactions with the negatively charged glutamate-containing C-terminal region.[194] The mechanism for fiber formation, as described by Chaikof, proposed that a stable triple helix forms at acidic pH and by adjusting the pH to 7 via the addition of a high salt buffer, the triple helices interdigitate to form nanofibers. The most impressive aspect of this system was the high level of organization of the nanofibers as observed in

uranyl acetate stained TEM images. This peptide was the first CMP designed and reported with D-periodicity (shown in Figure 1.19g). Although the measured D-period of 18 nm was far smaller than that of native collagen (64-67 nm), the authors argue that the difference can be attributed to the length of the CMP as compared to the length of peptides in native collagen. This system provides much insight on the driving forces for long-range organization within synthetic collagen homotrimers and is a template upon which much future work may be based.

As shown in Figure 1.18, there is much work yet to be done on synthetic collagen fiber assembly. While some successes, particularly the work of Chaikof and Conticello, have been reported for homotrimer fibers, this work is still in its early stages. In our opinion, the greatest opportunities for improvement in fiber forming CMPs are the following: 1) the fibers reported and shown using dry TEM or AFM images constitute a minor component of the system while the major component seen are amorphous aggregates, 2) the inability to replicate results seen in dried AFM and TEM using hydrated techniques such as vitreous ice cryo-TEM, and 3) the lack of heterotrimeric systems that assemble beyond the triple helix. For example, while the collagen fibers reported by Chaikof and Conticello are the first to show a repeating D-banding in the uranyl acetate stained samples, the TEM images of the CMP fibers show a high presence of irregular aggregates which are not present in vitreous ice cryo-TEM or dry stained TEM samples of re-constituted rat-tail collagen (cryo-TEM shown in Figure 1.19a). In fact the majority of fibers reported using CMPs show a high occurrence of amorphous aggregates within the TEM images, which suggest poor uniformity of the assembled materials. The system that has shown the highest ratio of organized fibers to aggregates is

the native chemical ligation polymerization technique, however this system is severely limited by the yield of the polymerization step. Additionally, none of the reported systems have been able to tackle the obstacle of replicating results seen in dry TEM using vitreous ice cryo-TEM. The major advantage of vitreous ice techniques is the insight that they provide on the structure of a system in solution without drying effects. Since most CMP systems are studied in buffers such as phosphate, tris or PBS (phosphate buffered saline), as a sample dries the salt and peptide concentrations steadily increase which can drastically change the environment surrounding the CMPs and the nature of the structures found in solution. For these reasons, the most conclusive form of characterization of collagen nanostructure will be vitreous ice cryo-TEM. Only when a system shows high-order long range fibers using this technique can one be sure of the state in solution for the peptide system. As of yet, there are no reports of higher order assemblies from heterotrimeric CMPs. However, heterotrimers may in the end become a superior platform for further self-assembly because the surface chemistry of the triple helix can be more effectively modified to create other modes of molecular recognition. Fibers from heterotrimeric CMPs would be particularly useful in their ability to mimic type I and type IV collagen, both of which would have substantial use of biomimetic materials for a variety of tissue engineering strategies.

1.5. Conclusion and Preview of Thesis

The self-assembly of molecules, specifically biologically relevant systems, has emerged as a promising field both for the understanding of proteins and other molecules within the body, but also for designing potential biological mimetics that can be used for

tissue engineering applications. The use of peptide-based systems for such applications provides much promise due to the diversity of functionality possible because of the broad amino acid library. Many of these systems, such as α-helical coiled-coils and β-sheet peptide amphiphiles, have made substantial progress in the formation of nanofibers that mimic the functionality of the extracellular matrix.[6-54] However, the similarity of these systems to amyloid fibrils makes them less-favorable for potential tissue engineering applications. Systems based on the replication of natural components of the extracellular matrix, such as collagen, has a greater potential for success when optimized and incorporated into *in vivo* applications.

In order to better replicate the multi-step assembly of collagen, a thorough understanding of the stabilizing forces and first stages of assembly, such as the collagen triple helix, is required. Despite having gained considerable knowledge on the structure and stabilization of the collagen triple helical folding motif, there is still much left unknown. Significantly, much published work uses homotrimeric collagen mimics, which are good models for some collagen types like fibril forming collagen type II and III found in cartilage and skin respectively. However, very little work is available on heterotrimeric triple helices primarily because, until recently, there was no straightforward method for their assembly.[176-182] Some of the most abundant proteins within the collagen family are heterotrimers, such as type I found in bones and teeth (AAB heterotrimer) and type IV, the major component of basement membranes (both AAB and ABC varieties exist), making heterotrimer research very promising to gain a deeper understanding of these important proteins. Another field where there is significant room for progress is the assembly of synthetic triple helices into higher-order supramolecular structures. Although

there have been several attempts to promote the hierarchical self-assembly of triple helical peptides into large supramolecular structures, only limited success has been achieved. Particularly, no synthetic system based on triple helical peptides has been shown to self-assemble into nanofibers and form a hydrogel. Such a synthetic construct would have potential application in areas such as tissue engineering because it could be tailored to be an ECM mimic allowing for collagen mediated cell-ECM interactions and remodeling.

In this thesis, two methods for the design and optimization of AAB type collagen triple helices as well as two designs for the assembly of collagen mimetic nanofibers are described. Chapter 2 illustrates the design of peptides containing both lysine and aspartate for the purpose of driving the formation of AAB heterotrimers. Though these peptides did not form distinct AAB triple helices, they demonstrated the strength of lysine-aspartate interactions and their ability to drive heterotrimer, as well as homotrimer ,formation. Chapter 3 expands on results from Chapter 2 and incorporates negative design into the system in order to drive AAB type heterotrimers whose component peptides do not form homotrimers. The results from this chapter marked the first synthetic collagen heterotrimers that demonstrated compositional control within the systems. In Chapter 4, based on the success of lysine-aspartate interactions in driving and stabilizing the formation of collagen heterotrimers, a design scheme for the formation of ABC type collagen nanofibers through the incorporation of hydrophobic residues into the peptides is depicted. Though the systems were successful in demonstrating the formation of nanofibers in dry microscopy techniques, the lack of reducibility in the nanofiber results and an inability to control peptide aggregation led to the overall breakdown of this

peptide design scheme. Lastly, Chapter 5 describes replication of $(PRG)_4(POG)_4(EOG)_4$, a peptide designed by Chaikof *et al.* that showed D-periodicity in dry-TEM, and modification to the Chaikof design intended to improve upon their nanofiber results. Through the sequence adjustments performed, the peptide $(PKG)_4(POG)_4(DOG)_4$ was identified that discretely demonstrates each step of the multi-hierarchical assembly of collagen. This system was the first reported synthetic peptide to display such properties and through initial biological testing, the peptide shows promise as a mimic of the extracellular matrix for tissue engineering applications.

References

(1) Steed, J. W.; Atwood, J. L. *Supramolecular Chemistry*; John Wiley & Sons, Ltd: Chichester, 2000; pp 1-34.

(2) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 964-967.

(3) Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251-260.

(4) Kornberg, R. D. *Annu. Rev. Biochem.* **1977**, *46*, 931-954.

(5) Lynch, J. *Story of Science*; Octopus Publishing Group: London, 2010;

(6) Apostolovic, B.; Danial, M.; Klok, H. A. *Chem. Soc. Rev.* **2010**, *39*, 3541-3575.

(7) Dong, H.; Hartgerink, J. D. *Biomacromolecules* **2006**, *7*, 691-695.

(8) Dong, H.; Hartgerink, J. D. *Biomacromolecules* **2007**, *8*, 617-623.

(9) Woolfson, D. N. *Adv. Protein Chem.* **2005**, *70*, 79-112.

(10) Dragan, A. I.; Potekhin, S. A.; Sivolob, A.; Lu, M.; Privalov, P. L. *Biochemistry* **2004**, *43*, 14891-14900.

(11) Kajava, A. V.; Potekhin, S. A.; Corradin, G.; Leapman, R. D. *J. Pept. Sci.* **2004**, *10*, 291-297.

(12) Melnik, T. N.; Villard, V.; Vasiliev, V.; Corradin, G.; Kajava, A. V.; Potekhin, S. A. *Protein Eng.* **2003**, *16*, 1125-1130.

(13) Potekhin, S. A.; Melnik, T. N.; Popov, V.; Lanina, N. F.; Vazina, A. A.; Rigler, P.; Verdini, A. S.; Corradin, G.; Kajava, A. V. *Chem. Biol.* **2001**, *8*, 1025-1032.

(14) Banwell, E. F.; Abelardo, E. S.; Adams, D. J.; Birchall, M. A.; Corrigan, A.; Donald, A. M.; Kirkland, M.; Serpell, L. C.; Butler, M. F.; Woolfson, D. N. *Nat. Mater.* **2009**, *8*, 596-600.

(15) Pandya, M. J.; Spooner, G. M.; Sunde, M.; Thorpe, J. R.; Rodger, A.; Woolfson, D. N. *Biochemistry* **2000**, *39*, 8728-8734.

(16) Papapostolou, D.; Smith, A. M.; Atkins, E. D.; Oliver, S. J.; Ryadnov, M. G.; Serpell, L. C.; Woolfson, D. N. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10853-10858.

(17) Smith, A. M.; Acquah, S. F. A.; Bone, N.; Kroto, H. W.; Ryadnov, M. G.; Stevens, M. S. P.; Walton, D. R. M.; Woolfson, D. N. *Angew. Chem. Int. Edit.* **2005**, *44*, 325-328.

(18) Walshaw, J.; Woolfson, D. N. *J. Mol. Biol.* **2001**, *307*, 1427-1450.

(19) Woolfson, D. N. *Biopolymers* **2010**, *94*, 118-127.

(20) Woolfson, D. N.; Mahmoud, Z. N. *Chem. Soc. Rev.* **2010**, *39*, 3464-3479.

(21) Dong, H.; Paramonov, S. E.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 13691-13695.

(22) Brea, R. J.; Reiriz, C.; Granja, J. R. *Chem. Soc. Rev.* **2010**, *39*, 1448-1456.

(23) Buriak, J. M.; Ghadiri, M. R. *Mat. Sci. Eng. C-Bio. S.* **1997**, *4*, 207-212.

(24) Hartgerink, J. D.; Granja, J. R.; Milligan, R. A.; Ghadiri, M. R. *J. Am. Chem. Soc.* **1996**, *118*, 43-50.

(25) Karlstrom, A.; Unden, A. *Biopolymers* **1997**, *41*, 1-4.

(26) Konig, B. *Angew. Chem. Int. Edit.* **1997**, *36*, 1833-1835.

(27) Israelachvili, J. N. *Intermolecular & Surface Forces, Second Edition*; Academic Press: London, 1992;

(28) Yu, Y. C.; Berndt, P.; Tirrell, M.; Fields, G. B. *J. Am. Chem. Soc.* **1996**, *118*, 12515-12520.

(29) Hartgerink, J. D.; Beniash, E.; Stupp, S. I. *P. Natl. Acad. Sci. USA* **2002**, *99*, 5133-5138.

(30) Niece, K. L.; Hartgerink, J. D.; Donners, J. J. J. M.; Stupp, S. I. *J. Am. Chem. Soc.* **2003**, *125*, 7146-7147.

(31) Stendahl, J. C.; Rao, M. S.; Guler, M. O.; Stupp, S. I. *Adv. Funct. Mater.* **2006**, *16*, 499-508.

(32) Paramonov, S. E.; Jun, H. W.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2006**, *128*, 7291-7298.

(33) Jun, H. W.; Paramonov, S. E.; Dong, H.; Forraz, N.; McGuckin, C.; Hartgerink, J. D. *J. Biomater. Sci. Polym. Ed.* **2008**, *19*, 665-676.

(34) Yuwono, V. M.; Hartgerink, J. D. *Langmuir* **2007**, *23*, 5033-5038.

(35) Beniash, E.; Hartgerink, J. D.; Storrie, H.; Stendahl, J. C.; Stupp, S. I. *Acta Biomater.* **2005**, *1*, 387-397.

(36) Galler, K. M.; Cavender, A.; Yuwono, V.; Dong, H.; Shi, S.; Schmalz, G.; Hartgerink, J. D.; D'Souza, R. N. *Tissue Eng. Part A* **2008**, *14*, 2051-2058.

(37) Aulisa, L.; Forraz, N.; McGuckin, C.; Hartgerink, J. D. *Acta Biomater.* **2009**, *5*, 842-853.

(38) Webber, M. J.; Tongers, J.; Newcomb, C. J.; Marquardt, K. T.; Bauersachs, J.; Losordo, D. W.; Stupp, S. I. *P. Natl. Acad. Sci. USA* **2011**, *108*, 13438-13443.

(39) Zhang, S. G.; Holmes, T.; Lockshin, C.; Rich, A. *P. Natl. Acad. Sci. USA* **1993**, *90*, 3334-3338.

(40) Zhang, S. G.; Lockshin, C.; Cook, R.; Rich, A. *Biopolymers* **1994**, *34*, 663-672.

(41) Misawa, H.; Kobayashi, N.; Soto-Gutierrez, A.; Chen, Y.; Yoshida, A.; Rivas-Carrillo, J. D.; Navarro-Alvarez, N.; Tanaka, K.; Miki, A.; Takei, J.; Ueda, T.; Tanaka, M.; Endo, H.; Tanaka, N.; Ozaki, T. *Cell Transplant* **2006**, *15*, 903-910.

(42) Zhang, S. G.; Holmes, T. C.; DiPersio, C. M.; Hynes, R. O.; Su, X.; Rich, A. *Biomaterials* **1995**, *16*, 1385-1393.

(43) Schneider, J. P.; Pochan, D. J.; Ozbas, B.; Rajagopal, K.; Pakstis, L.; Kretsinger, J. *J Am Chem Soc* **2002**, *124*, 15030-15037.

(44) Haines-Butterick, L.; Rajagopal, K.; Branco, M.; Salick, D.; Rughani, R.; Pilarz, M.; Lamm, M. S.; Pochan, D. J.; Schneider, J. P. *P Natl Acad Sci Usa* **2007**, *104*, 7791-7796.

(45) Gungormus, M.; Branco, M.; Fong, H.; Schneider, J. P.; Tamerler, C.; Sarikaya, M. *Biomaterials* **2010**, *31*, 7266-7274.

(46) Dong, H.; Paramonov, S. E.; Aulisa, L.; Bakota, E. L.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 12468-12472.

(47) Bakota, E. L.; Aulisa, L.; Tsyboulski, D. A.; Weisman, R. B.; Hartgerink, J. D. *Biomacromolecules* **2009**, *10*, 2201-2206.

(48) Tsyboulski, D. A.; Bakota, E. L.; Witus, L. S.; Rocha, J.-D. R.; Hartgerink, J. D.; Weisman, R. B. *J. Am. Chem. Soc.* **2008**, *130*, 17134-17140.

(49) Witus, L. S.; Rocha, J.-D. R.; Yuwono, V. M.; Paramonov, S. E.; Weisman, R. B.; Hartgerink, J. D. *J. Mater. Chem.* **2007**, *17*, 1909-1915.

(50) Bakota, E. L.; Aulisa, L.; Galler, K. M.; Hartgerink, J. D. *Biomacromolecules* **2011**, *12*, 82-87.

(51) Galler, K. M.; Aulisa, L.; Regan, K. R.; D'Souza, R. N.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2010**, *132*, 3217-3223.

(52) Bakota, E. L.; Wang, Y.; Danesh, F. R.; Hartgerink, J. D. *Biomacromolecules* **2011**, *12*, 1651-1657.

(53) Wang, Y.; Bakota, E.; Chang, B. H. J.; Entman, M.; Hartgerink, J. D.; Danesh, F. R. *J. Am. Soc. Nephrol.* **2011**, *22*, 704-717.

(54) Aulisa, L.; Dong, H.; Hartgerink, J. D. *Biomacromolecules* **2009**, *10*, 2694-2698.

(55) Hwang, W.; Zhang, S. G.; Kamm, R. D.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12916-12921.

(56) Di Lullo, G. A.; Sweeney, S. M.; Korkko, J.; Ala-Kokko, L.; San, A., JD. *J. Biol. Chem.* **2002**, *277*, 4223-4231.

(57) Kadler, K. E.; Baldock, C.; Bella, J.; Boot-Handford, R. P. *J. Cell Sci.* **2007**, *120*, 1955-1958.

(58) Heino, J. *Bioessays* **2007**, *29*, 1001-1010.

(59) van der Rest, M.; Garrone, R. *FASEB J.* **1991**, *5*, 2814-2823.

(60) Khoshnoodl, J.; Cartailler, J.-P.; Alvares, K.; Cels, A.; Hudson, B. G. *J. Biol. Chem.* **2006**, *281*, 38117-38121.

(61) Tuckwell, D. S.; Ayad, S.; Grant, M. E.; Takigawa, M.; Humphries, M. J. *J. Cell Sci.* **1994**, *107*, 993-1005.

(62) Vandenberg, P.; Kern, A.; Ries, A.; Luckenbill-Edds, L.; Mann, K.; Kuhn, K. *J. Cell Biol.* **1991**, *113*, 1475-1483.

(63) Lauer-Fields, J. L.; Juska, D.; Fields, G. B. *Biopolymers* **2002**, *66*, 19-32.

(64) Kalluri, R. *Nat. Rev. Cancer* **2003**, *3*, 422-433.

(65) Lauer-Fields, J. L.; Tuzinski, K. A.; Shimokawa, K.; Nagase, H.; Fields, G. B. *J. Biol. Chem.* **2000**, *275*, 13282-13290.

(66) Hulmes, D. J. S.; Miller, A. *Nature* **1979**, *282*, 878-880.

(67) Miller, A.; Tocchetti, D. *Int. J. Biol. Macromol.* **1981**, *3*, 9-18.

(68) Ottani, V.; Raspanti, M.; Ruggeri, A. *Micron* **2001**, *32*, 251-260.

(69) Xu, Y.; Bhate, M.; Brodsky, B. *Biochemistry* **2002**, *41*, 8143-8151.

(70) O'Leary, L. E. R.; Fallas, J. A.; Bakota, E. L.; Kang, M. K.; Hartgerink, J. D. *Nature Chem.* **2011**, *3*, 821-828.

(71) Orgel, J. P.; Irving, T. C.; Miller, A.; Wess, T. J. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9001-9005.

(72) Okuyama, K.; Wu, G.; Jiravanichanun, N.; Hongo, C.; Noguchi, K. *Biopolymers* **2006**, *84*, 421-432.

(73) Shoulders, M. D.; Raines, R. T. *Annu. Rev. Biochem.* **2009**, *78*, 929-958.

(74) Farndale, R. W.; Lisman, T.; Bihan, D.; Hamaia, S.; Smerling, C. S.; Pugh, N.; Konitsiotis, A.; Leitinger, B.; de Groot, P. G.; Jarvis, G. E.; Raynal, N. *Biochem. Soc. T.* **2008**, *36*, 241-250.

(75) Brodsky, B.; Thiagarajan, G.; Madhan, B.; Kar, K. *Biopolymers* **2008**, *89*, 345-353.

(76) Nagarajan, V.; Kamitori, S.; Okuyama, K. *J. Biochem.* **1998**, *124*, 1117-1123.

(77) Nagarajan, V.; Kamitori, S.; Okuyama, K. *J. Biochem.* **1999**, *125*, 310-318.

(78) Mohs, A.; Silva, T.; Yoshida, T.; Amin, R.; Lukomski, S.; Inouye, M.; Brodsky, B. *J. Biol. Chem.* **2007**, *282*, 29757-29765.

(79) Kramer, R. Z.; Venugopal, M. G.; Bella, J.; Mayville, P.; Brodsky, B.; Berman, H. M. *J. Mol. Biol.* **2000**, *301*, 1191-1205.

(80) Li, Y.; Brodsky, B.; Baum, J. *J. Biol. Chem.* **2008**, *282*, 22699-22706.

(81) Okuyama, K.; Okuyama, K.; Arnott, S.; Takayanagi, M.; Kakudo, M. *J. Mol. Biol.* **1981**, *152*, 427-443.

(82) Emsley, J.; Knight, C. G.; Farndale, R. W.; Barnes, M. J.; Liddington, R. C. *Cell* **2000**, *101*, 47-56.

(83) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149-&.

(84) Sakakibara, S.; Kishida, Y.; Kikuchi, Y.; Sakai, R.; Kakiuchi, K. *B. Chem. Soc. Jpn.* **1968**, *41*, 1273.

(85) Berg, R. A.; Olsen, B. R.; Prockop, D. J. *J. Biol. Chem.* **1970**, *245*, 5759-5763.

(86) Sakakibara S, .; Tanaka, N.; Kakudo, M.; Okuyama, K.; Ashida, T.; Kishida, Y. *J. Mol. Biol.* **1972**, *65*, 371.

(87) Okuyama, K.; Tanaka, N.; Ashida, T.; Kakudo, M.; Sakakibara, S.; Kishida, Y. *J. Mol. Biol.* **1972**, *72*, 571-572.

(88) Okuyama, K.; Takayanagi, M.; Ashida, T.; Kakudo, M. *Polym. J.* **1977**, *9*, 341-343.

(89) Fields, G. B.; Prockop, D. J. *Biopolymers* **1996**, *40*, 345-357.

(90) Fields, G. B. *Org. Biomol. Chem.* **2010**, *8*, 1237-1258.

(91) Paramonov, S. E.; Gauba, V.; Hartgerink, J. D. *Macromolecules* **2005**, *38*, 7555-7561.

(92) Fields, C. G.; Grab, B.; Lauer, J. L.; Fields, G. B. *Anal. Biochem.* **1995**, *231*, 57-64.

(93) Xiao, J. X.; Baum, J. *J. Am. Chem. Soc.* **2009**, *131*, 18194-18195.

(94) Cohen, C.; Bear, R. S. *J. Am. Chem. Soc.* **1953**, *75*, 2783-2784.

(95) Cowan, P. M.; McGavin, S.; North, A. C. *Nature* **1955**, *176*, 1062-1604.

(96) Ramachandran, G. N.; Kartha, G. *Nature* **1955**, *176*, 593-595.

(97) Ramachandran, G. N.; Ambady, G. K. *Curr. Sci.* **1954**, *23*, 349-350.

(98) Rich, A.; Crick, F. H. C. *Nature* **1955**, *176*, 915-916.

(99) Crick, F. H. C.; Rich, A. *Nature* **1955**, *176*, 780-781.

(100) Rich, A.; Crick, F. H. C. *J. Mol. Biol.* **1961**, *3*, 483-506.

(101) Cowan, P. M.; McGavin, S. *Nature* **1955**, *176*, 501-503.

(102) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. *Science* **1994**, *266*, 75-81.

(103) Fallas, J. A.; O'Leary, L. E.; Hartgerink, J. D. *Chem. Soc. Rev.* **2010**, *39*, 3510-3527.

(104) Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A. *Protein Sci.* **2002**, *11*, 262-270.

(105) Fallas, J. A.; Gauba, V.; Hartgerink, J. D. *J. Biol. Chem.* **2009**, *284*, 26851-26859.

(106) Li, M. H.; Fan, P.; Brodsky, B.; Baum, J. *Biochemistry* **1993**, *32*, 7377-7387.

(107) Okuyama, K. *Connect. Tissue. Res.* **2008**, *49*, 299-310.

(108) Okuyama, K.; Xu, X. Z.; Iguchi, M.; Noguchi, K. *Biopolymers* **2006**, *84*, 181-191.

(109) Kramer, R. Z.; Bella, J.; Brodsky, B.; Berman, H. M. *J. Mol. Biol.* **2001**, *311*, 131-147.

(110) Kramer, R. Z.; Bella, J.; Mayville, P.; Brodsky, B.; Berman, H. M. *Nat. Struct. Biol.* **1999**, *6*, 454-457.

(111) Emsley, J.; Knight, C. G.; Farndale, R. W.; Barnes, M. J. *J. Mol. Biol.* **2004**, *335*, 1019-1028.

(112) Schumacher, M. A.; Mizuno, K.; Bachinger, H. P. *J. Biol. Chem.* **2006**, *281*, 27566-27574.

(113) Berg, R. A.; Prockop, D. J. *Biochem. Biophys. Res. Commun.* **1973**, *52*, 115-120.

(114) Inouye, K.; Kobayashi, Y.; Kyogoku, Y.; Kishida, Y.; Sakakibara, S.; Prockop, D. J. *Arch. Biochem. Biophys.* **1982**, *219*, 198-203.

(115) Bella, J.; Brodsky, B.; Berman, H. M. *Structure* **1995**, *3*, 893-906.

(116) Shoulders, M. D.; Satyshur, K. A.; Forest, K. T.; Raines, R. T. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 559-564.

(117) Ramachandran, G. N.; Bansal, M.; Bhatnagar, R. S. *Biochim. Biophys. Acta.* **1973**, *322*, 166-171.

(118) Suzuki, E.; Fraser, R. D. B.; Macrae, T. P. *Role of hydroxyproline in the stabilization of the collagen molecule via water-molecules*; 1980; pp 54-56.

(119) Xu, Y.; Hyde, T.; Wang, X.; Bhate, M.; Brodsky, B.; Baum, J. *Biochemistry* **2003**, *42*, 8696-8703.

(120) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. *J. Am. Chem. Soc.* **2001**, *123*, 777-778.

(121) Eberhardt, E. S.; Panasik, N.; Raines, R. T. *J. Am. Chem. Soc.* **1996**, *118*, 12261-12266.

(122) Holmgren, S. K.; Bretscher, L. E.; Taylor, K. M.; Raines, R. T. *Chem. Biol.* **1999**, *6*, 63-70.

(123) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. *Nature* **1998**, *392*, 666-667.

(124) Jakobsche, C. E.; Choudhary, A.; Miller, S. J.; Raines, R. T. *J. Am. Chem. Soc.* **2010**, *132*, 6651-6653.

(125) Jenkins, C. L.; Vasbinder, M. M.; Miller, S. J.; Raines, R. T. *Org. Lett.* **2005**, *7*, 2619-2622.

(126) Dai, N.; Wang, X. J.; Etzkorn, F. A. *J. Am. Chem. Soc.* **2008**, *130*, 5396-5397.

(127) Dai, N.; Etzkorn, F. A. *J. Am. Chem. Soc.* **2009**, *131*, 13728-13732.

(128) Persikov, A. V.; Ramshaw, J. A.; Kirkpatrick, A.; Brodsky, B. *Biochemistry* **2000**, *39*, 14960-14967.

(129) Shah, N. K.; Ramshaw, J. A.; Kirkpatrick, A.; Shah, C.; Brodsky, B. *Biochemistry* **1996**, *35*, 10262-10268.

(130) Fraser, R. D.; MacRae, T. P.; Suzuki, E. *J. Mol. Biol.* **1979**, *129*, 463-481.

(131) Jones, E. Y.; Miller, A. *J. Mol. Biol.* **1991**, *218*, 209-219.

(132) Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B. *Biochemistry* **2005**, *44*, 1414-1422.

(133) Melacini, G.; Feng, Y. B.; Goodman, M. *J. Am. Chem. Soc.* **1996**, *118*, 10359-10364.

(134) Feng, Y. B.; Melacini, G.; Taulane, J. P.; Goodman, M. *J. Am. Chem. Soc.* **1996**, *118*, 10351-10358.

(135) Li, Y.; Mo, X. A.; Kim, D.; Yu, S. M. *Biopolymers* **2011**, *95*, 94-104.

(136) Kwak, J.; De Capua, A.; Locardi, E.; Goodman, M. *J. Am. Chem. Soc.* **2002**, *124*, 14085-14091.

(137) Kinberger, G. A.; Cai, W. B.; Goodman, M. *J. Am. Chem. Soc.* **2002**, *124*, 15162-15163.

(138) Cai, W.; Wong, D.; Kinberger, G. A.; Kwok, S. W.; Taulane, J. P.; Goodman, M. *Bioorg. Chem.* **2007**, *35*, 327-337.

(139) Cai, W.; Taulane, J. P.; Sorto, N. A.; Oganesyan, A.; Guitierrez, C. G.; Goodman, M. *Lett. Org. Chem.* **2007**, *4*, 96-101.

(140) Goodman, M.; Melacini, G.; Feng, Y. B. *J. Am. Chem. Soc.* **1996**, *118*, 10928-10929.

(141) Melacini, G.; Feng, Y. B.; Goodman, M. *J. Am. Chem. Soc.* **1996**, *118*, 10725-10732.

(142) Feng, Y. B.; Melacini, G.; Taulane, J. P.; Goodman, M. *Biopolymers* **1996**, *39*, 859-872.

(143) Feng, Y. B.; Melacini, G.; Goodman, M. *Biochemistry* **1997**, *36*, 8716-8724.

(144) Melacini, G.; Feng, Y. B.; Goodman, M. *Biochemistry* **1997**, *36*, 8725-8732.

(145) Barth, D.; Musiol, H.; Schutt, M.; Fiori, S.; Milbradt, A. G.; Renner, C.; Moroder, L. *Chem.-eur. J.* **2003**, *9*, 3692-3702.

(146) Barth, D.; Kyrieleis, O.; Frank, S.; Renner, C.; Moroder, L. *Chem.-eur. J.* **2003**, *9*, 3703-3714.

(147) Henkel, W.; Vogl, T.; Echner, H.; Voelter, W.; Urbanke, C.; Schleuder, D.; Rauterberg, J. *Biochemistry* **1999**, *38*, 13610-13622.

(148) Cai, W. B.; Kwok, S. W.; Taulane, J. P.; Goodman, M. *J. Am. Chem. Soc.* **2004**, *126*, 15030-15031.

(149) Kinberger, G. A.; Taulane, J. P.; Goodman, M. *Inorg. Chem.* **2006**, *45*, 961-963.

(150) Yu, Y. C.; Tirrell, M.; Fields, G. B. *J. Am. Chem. Soc.* **1998**, *120*, 9979-9987.

(151) Suto, K.; Noda, H. *Biopolymers* **1974**, *13*, 2477-2488.

(152) Piez, K. A.; Sherman, M. R. *Biochemistry* **1970**, *9*, 4134-4140.

(153) Ackerman, M. S.; Bhate, M.; Shenoy, N.; Beck, K.; Ramshaw, J. A.; Brodsky, B. *J. Biol. Chem.* **1999**, *274*, 7668-7673.

(154) Liu, X. Y.; Siegel, D. L.; Fan, P.; Brodsky, B.; Baum, J. *Biochemistry* **1996**, *35*, 4306-4313.

(155) Baum, J.; Brodsky, B. *Fold. Des.* **1997**, *2*, R53-60.

(156) Baum, J.; Brodsky, B. *Curr. Opin. Struct. Biol.* **1999**, *9*, 122-128.

(157) Buevich, A. V.; Dai, Q. H.; Liu, X. Y.; Brodsky, B.; Baum, J. *Biochemistry* **2000**, *39*, 4299-4308.

(158) Buevich, A.; Baum, J. *Philos. T. Roy. Soc. B* **2001**, *356*, 159-168.

(159) Bachinger, H. P. *J. Biol. Chem.* **1987**, *262*, 17144-17148.

(160) Bachinger, H. P.; Bruckner, P.; Timpl, R.; Prockop, D. J.; Engel, J. *Eur. J. Biochem.* **1980**, *106*, 619-632.

(161) Fields, C. G.; Grab, B.; Lauer, J. L.; Miles, A. J.; Yu, Y. C.; Fields, G. B. *Lett. Pept. Sci.* **1996**, *3*, 3-16.

(162) Ottl, J.; Moroder, L. *Tetrahedron Lett.* **1999**, *40*, 1487-1490.

(163) Renner, C.; Sacca, B.; Moroder, L. *Biopolymers* **2004**, *76*, 34-47.

(164) Ottl, J.; Moroder, L. *J. Am. Chem. Soc.* **1999**, *121*, 653-661.

(165) Ottl, J.; Musiol, H. J.; Moroder, L. *J. Pept. Sci.* **1999**, *5*, 103-110.

(166) Fiori, S.; Sacca, B.; Moroder, L. *J. Mol. Biol.* **2002**, *319*, 1235-1242.

(167) Ottl, J.; Gabriel, D.; Murphy, G.; Knauper, V.; Tominaga, Y.; Nagase, H.; Kroger, M.; Tschesche, H.; Bode, W.; Moroder, L. *Chem. Biol.* **2000**, *7*, 119-132.

(168) Muller, J. C. D.; Ottl, J.; Moroder, L. *Biochemistry* **2000**, *39*, 5111-5116.

(169) Sacca, B.; Moroder, L. *J. Pept. Sci.* **2002**, *8*, 192-204.

(170) Sacca, B.; Sinner, E. K.; Kaiser, J.; Lubken, C.; Eble, J. A.; Moroder, L. *Chembiochem* **2002**, *3*, 904-907.

(171) Golbik, R.; Eble, J. A.; Ries, A.; Kuhn, K. *J. Mol. Biol.* **2000**, *297*, 501-509.

(172) Sacca, B.; Renner, C.; Moroder, L. *J. Mol. Biol.* **2002**, *324*, 309-318.

(173) Madhan, B.; Xiao, J. X.; Thiagarajan, G.; Baum, J.; Brodsky, B. *J. Am. Chem. Soc.* **2008**, *130*, 13520-13521.

(174) Slatter, D. A.; Miles, C. A.; Bailey, A. J. *J. Mol. Biol.* **2003**, *329*, 175-183.

(175) Slatter, D. A.; Foley, L. A.; Peachey, A. R.; Nietlispach, D.; Farndale, R. W. *J. Mol. Biol.* **2006**, *359*, 289-298.

(176) Hodges, J. A.; Raines, R. T. *J. Am. Chem. Soc.* **2005**, *127*, 15923-15932.

(177) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 2683-2690.

(178) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 15034-15041.

(179) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 7509-7515.

(180) Brodsky, B.; Baum, J. *Nature* **2008**, *453*, 998-999.

(181) Xu, F.; Zhang, L.; Koder, R. L.; Nanda, V. *Biochemistry* **2010**, *49*, 2307-2316.

(182) Xu, F.; Zahid, S.; Silva, T.; Nanda, V. *J. Am. Chem. Soc.* **2011**, *133*, 15260-15263.

(183) Ottani, V.; Martini, D.; Franchi, M.; Ruggeri, A.; Raspanti, M. *Micron* **2002**, *33*, 587-596.

(184) Kar, K.; Amin, P.; Bryan, M. A.; Persikov, A. V.; Mohs, A.; Wang, Y. H.; Brodsky, B. *J. Biol. Chem.* **2006**, *281*, 33283-33290.

(185) Kar, K.; Wang, Y. H.; Brodsky, B. *Protein Sci.* **2008**, *17*, 1086-1095.

(186) Wang, A. Y.; Mo, X.; Chen, C. S.; Yu, S. M. *J. Am. Chem. Soc.* **2005**, *127*, 4130-4131.

(187) Mo, X.; An, Y. J.; Yun, C. S.; Yu, S. M. *Angew. Chem. Int. Edit.* **2006**, *45*, 2267-2270.

(188) Lee, H. J.; Lee, J. S.; Chansakul, T.; Yu, C.; Elisseeff, J. H.; Yu, S. M. *Biomaterials* **2006**, *27*, 5268-5276.

(189) Cejas, M. A.; Kinnney, W. A.; Chen, C.; Vinter, J. G.; Almond, H. R.; Balss, K. M.; Maryanoff, C. A.; Schmidt, U.; Breslav, M.; Mahan, A.; Lacy, E.; Maryanoff, B. E. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 8513-8518.

(190) Kar, K.; Ibrar, S.; Nanda, V.; Getz, T. M.; Kunapuli, S. P.; Brodsky, B. *Biochemistry* **2009**, *48*, 7959-7968.

(191) Yamazaki, C. M.; Asada, S.; Kitagawa, K.; Koide, T. *Biopolymers* **2008**, *90*, 816-823.

(192) Kotch, F. W.; Raines, R. T. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3028-3033.

(193) Pires, M. M.; Chmielewski, J. *J. Am. Chem. Soc.* **2009**, *131*, 2706-2712.

(194) Rele, S.; Song, Y.; Apkarian, R. P.; Qu, Z.; Conticello, V. P.; Chaikof, E. L. *J. Am. Chem. Soc.* **2007**, *129*, 14780-14787.

(195) Dublin, S. N.; Conticello, V. P. *J. Am. Chem. Soc.* **2008**, *130*, 49-51.

(196) Pires, M. M.; Ernenwein, D.; Chmielewski, J. *Biomacromolecules* **2011**, *12*, 2429-2433.

(197) Krishna, O. D.; Kiick, K. L. *Biomacromolecules* **2009**, *10*, 2626-2631.

Chapter 2: AAB Heterotrimer Formation Utilizing Positive Design

As mentioned in Chapter 1, many of the most common types of collagen are heterotrimers meaning that at least one of the three peptide chains constituting the triple helix is unique. These heterotrimers can be ABC in which all peptide chains have unique amino acid sequences, collagen types V, IX and XI form these types of triple helices. Heterotrimers can also be AAB, collagen types I and IV assemble in such a way, in which two of the peptide chains are identical and the third is unique.[1] Electrostatically driven assembly of ABC heterotrimers was previously reported by the Hartgerink lab and was described in the heterotrimer formation section of Chapter 1.[2,3] The highest stability system within this study was composed of the three peptides $(PKG)_{10}$, $(DOG)_{10}$ and $(POG)_{10}$ and 2D solution NMR on the system reported the presence of direct electrostatic interactions between the lysine (K) in triplet *n* and the aspartate (D) in triplet *n+1*.[4] Therefore we hypothesized that peptide design based on the incorporation of lysine-aspartate interactions could drive the formation of AAB type heterotrimers, a subset of heterotrimers that had previously only been reported as low stability systems.[5-8]

2.1. Peptide Design

The $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ system provided a template for the synthesis of thermally stable heterotrimeric collagen triple helices and by incorporating the NMR work by Fallas that determined the position of stabilizing electrostatic interactions between lysine and aspartate within the system, these same interactions can be replicated in a peptide system that drives AAB type heterotrimer formation. In order to drive AAB

triple helix formation, peptides were designed that maximized the number of K-D interactions from the leading to middle, middle to lagging or lagging to leading strands when assembled in an AAB heterotrimer. Figure 2.1 highlights these three types of interactions.

Figure 2.1. Scheme depicting the three types of lysine-aspartate interactions possible in a collagen triple helix based on the work by Fallas *et al.*[4] using single letter amino acid codes. (a) Leading to middle strand interaction, (b) middle to lagging strand interaction and (c) lagging to leading strand interaction where the lysines are highlighted in blue, aspartates are shown in red and the interaction is depicted with a yellow lasso. The other amino acids present in the scheme are proline (P), hydroxyproline (O) and glycine (G).

Using this design principle, many peptide systems were designed and discussed as potential options for forming AAB heterotrimers. Three peptides were identified as the most promising based on the number of K-D salt bridges they provided when in an AAB heterotrimer: $(PKGDOG)_5$, $(POGDKG)_5$ and $(DKGPOG)_5$. The peptides will be referred to as **A**, **B** and **C** respectively so, for example, the mixture of $(PKGPOG)_5$ and $(POGDKG)_5$ will be referred to as the **A/B** system. The most stable organizations of peptides for each peptide mixture based on favorable electrostatic interactions are shown in Figure 2.2 using amino acid single-letter codes. The lysine and aspartate residues involved in direct electrostatic interactions that stabilize each triple helix are bolded in blue and red respectively with an example interaction in each system depicted by a yellow lasso. Through the incorporation of lysine and aspartate in each peptide, we hoped to form directed K-D interactions when the peptides were mixed. Based on the designs, the maximum number of lysine-aspartate salt bridges are possible in the AAB

heterotrimer therefore we hypothesized that mixtures of the peptides would selectively form the species with the most K-D interactions, the heterotrimer.

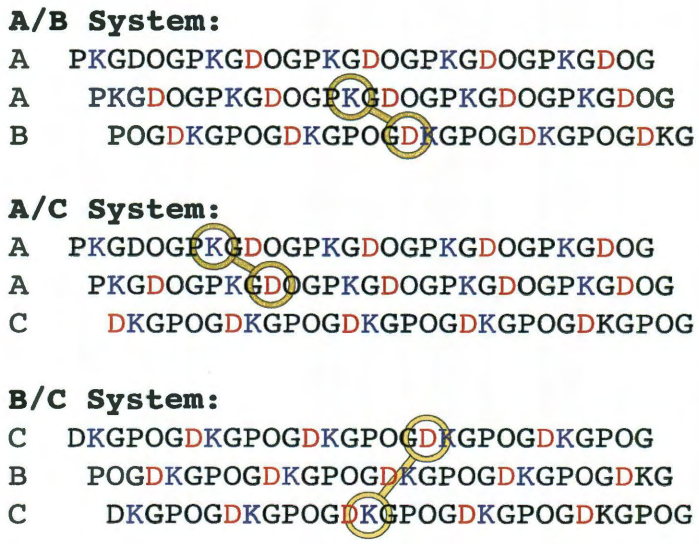

Figure 2.2. Heterotrimeric species that can be formed by mixing two peptides where peptide **A** = $(PKGDOG)_5$, peptide **B** = $(POGDKG)_5$ and peptide **C** = $(DKGPOG)_5$. Lysine and aspartate residues that participate in electrostatic interactions are shown in blue and red respectively. Examples of these interactions are highlighted in yellow lassos.

2.2. Homotrimer Formation

All three peptides within the proposed systems were successfully synthesized and purified (for details see the experimental section below and Appendix 2 for representative HPLC chromatograms and mass spectra). Once purified, the triple helical nature of the peptides was examined using circular dichroism (CD). Procedures for sample preparation and CD analysis are given in the experimental section below. Before mixtures of the peptides could be explored, the ability of each peptide to form stable homotrimers was assessed.

In order to determine whether a collagen mimetic peptide (CMP) forms a triple helix, two CD experiments must be performed: a wavelength spectrum and a thermal unfolding curve. Collagen triple helices have a signature CD spectral profile that consists of a maximum at 225 nm and a minimum near 200 nm. This profile is indicative of a poly-proline type II helix, the secondary structure adopted by the individual peptide chains within a triple helix. The thermal unfolding experiment monitors the spectral maximum as the temperature is increased which, when a triple helix is present, shows a cooperative transition. The first derivative of the thermal unfolding curve is then used to identify the melting temperature of the sample. In the Hartgerink group, we define the melting temperature of a system as the temperature corresponding to the minimum value in the first derivative of the melting curve. This method differs slightly from other research groups who define the melting temperature of a peptide as the temperature associated with the halfway point between the fully folded and unfolded species. Despite the different techniques, the resulting melting temperatures for peptide systems calculated by each method will be within ± 2 °C of each other. All CD spectra are reported as molar residual ellipticity (MRE), which normalizes the data for peptide concentration, peptide length and pathlength (see experimental section for more details).

The CD wavelength spectrum for each peptide displayed the predicted maximum at 225 nm and a minimum around 200 nm, which is typical for polyproline type II helices (Figure 2.3 left column). Similarly, the thermal unfolding curves for each peptide were sigmoidal in shape indicating that each peptide forms a homotrimer that unfolds in a cooperative manner (Figure 2.3 center column). In the right column of Figure 2.3, the first derivative of the melting curve is shown which was used to decipher the thermal

stability of each peptide. Based on our method of melting temperature determination, the homotrimers thermally unfolded at 38, 48, and 50 °C for **A•A•A**, **B•B•B**, and **C•C•C** respectively. All melting studies were repeatable with similar peak intensities and identical melting temperatures (T_m).

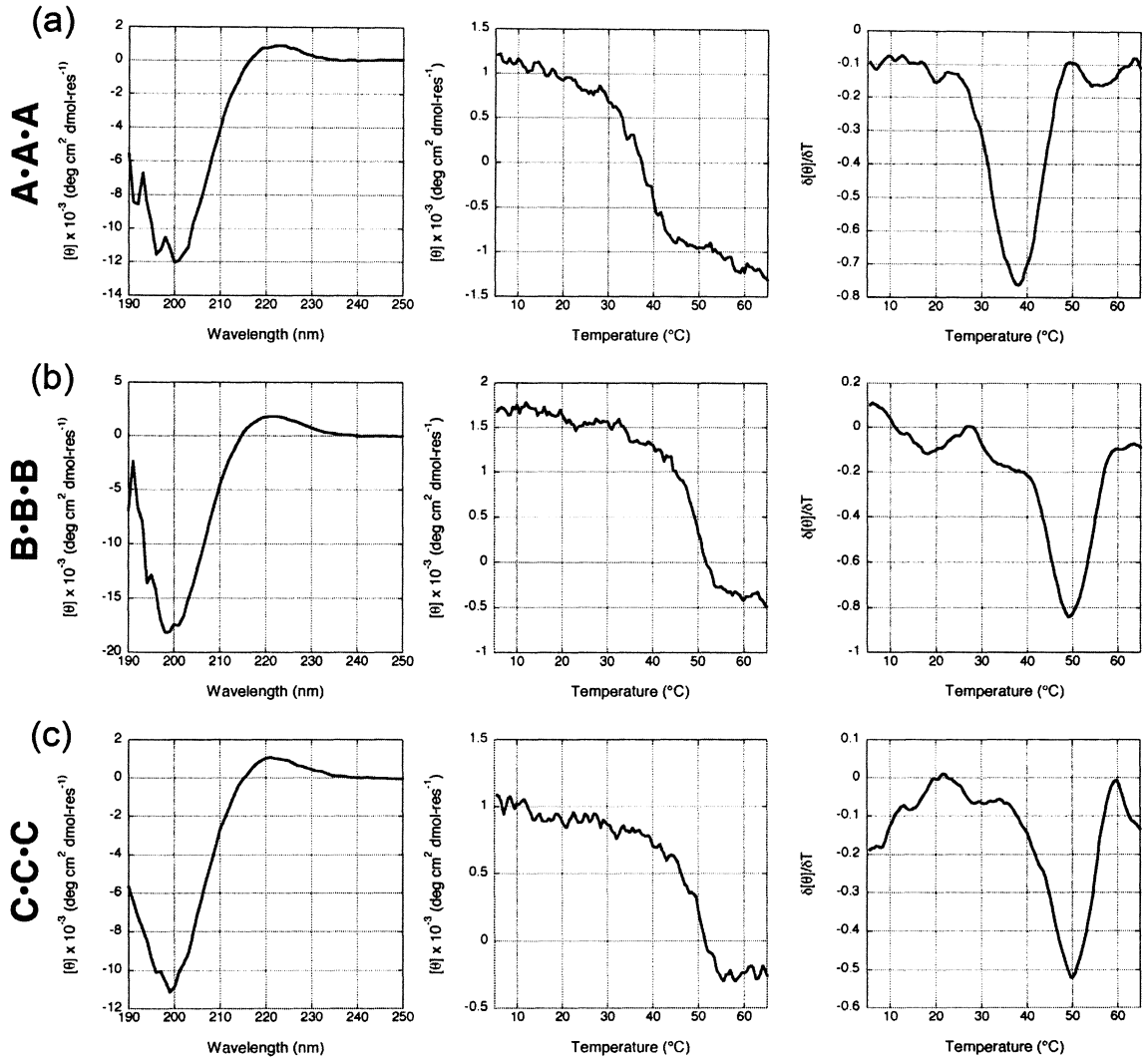

Figure 2.3. CD experiments performed on the homotrimers of (a) peptide **A** = $(PKGDOG)_5$, (b) peptide **B** = $(POGDKG)_5$ and (c) peptide **C** = $(DKGPOG)_5$. The left column displays the CD wavelength spectrum as MRE vs. wavelength, the center column is the thermal unfolding curve shown as MRE vs. temperature and the right column is the first derivative of the melting curve given as the first derivative of the MRE vs. temperature.

The ability of each peptide to form homotrimers was not unexpected based on the amino acid sequences. The presence of lysine and aspartate residues in each peptide allowed for potential K-D interactions within homotrimers thus stabilizing these species. When the match up of lysine and aspartate within each homotrimer was considered, a substantial number of salt bridges were possible in the **A•A•A** homotrimer (Figure 2.4a). In contrast, for **B•B•B** and **C•C•C**, the alignment of K-D interactions is very poor within the triple helix. The only interactions possible between lysine and aspartate were lateral interactions (between the same triplets in adjacent strands), which have been seen in solution NMR for collagen mimetic systems studied by Jorge Fallas, but these interactions are less stabilizing than the salt bridges between triplets *n* and *n+1*. However, when the peptide chains within **B•B•B** and **C•C•C** are staggered by one triplet, the number of K-D salt bridges substantially increases. A depiction of the two homotrimers in the traditional and staggered alignments is shown in Figure 2.4b and 2.4c. The additional K-D interactions that are possible by staggering the strands within the homotrimer could be responsible for the higher thermal unfolding temperatures seen for the **B•B•B** and **C•C•C** homotrimers as compared to the **A•A•A** homotrimer. It must be noted that no collagen mimetic systems have proven the presence of a staggered triple helical arrangement. In order for a staggered assembly to occur, one of the glycine backbone hydrogen bonds, that is known to stabilize the triple helical conformation, must be sacrificed. Therefore the electrostatic interactions gained by the stagger must compensate for the loss of the glycine hydrogen bond in order for the staggered arrangement to be energetically favorable. Additionally, both **B•B•B** and **C•C•C** contain five POG triplets within the amino acid sequences while **A•A•A** does not contain any.

Hence, the presence of these POG triplets in addition to the lateral interactions possible in the traditional staggering arrangement could be the reason for the higher stability **B•B•B** and **C•C•C** compared to **A•A•A**.

Figure 2.4. Homotrimers formed by (a) $(PKGPOG)_5$, (b) $(POGDKG)_5$ and (c) $(DKGPOG)_5$. Lysine and aspartate residues that contribute to K-D salt bridges are shown in blue and red respectively. (b and c) Peptides are shown in the traditional arrangement on the left and the yellow arrow indicated the staggering of the peptide chains by one triplet resulting in the staggered arrangement on the right.

Between the **B** and **C** peptides, $(POGDKG)_5$ and $(DKGPOG)_5$, the amino acid sequences only differ by the order of the triplets within the repeating unit where peptide **B** has the POG triplet first and peptide **C** has the DKG triplet first. However, there was a 2 °C difference in thermal stability seen for the two peptides. In order to account for this difference, two principles of CMP design must be explained. The first is that the more amino acid mutations made within a peptide chain that deviate it from the canonical $(POG)_{10}$, the lower the expected thermal stability will be.[2,9] The second is that the N- and C- termini of triple helices are more loosely folded than the interior of the helix. In $(POGDKG)_5$, an aspartate and a lysine are in the triplet at the C-terminus of the peptide. Based on K-D pairing, the aspartate can be involved in direct electrostatic interactions but

the lysine will remain unpaired. In contrast, $(DKGPOG)_5$ has an aspartate and a lysine in the N-terminal triplet where the lysine can be involved in K-D pairing and the aspartate remains unpaired. Due to the fact that the $(POGDKG)_5$ has a homotrimeric melting temperature that is 2 °C lower than that for $(DKGPOG)_5$, we concluded that the unpaired lysine in the former peptide is more destabilizing in the triple helix than the unpaired aspartate in the latter peptide.

2.3. Heterotrimer Formation

In order to drive the formation of AAB type heterotrimers, the peptides must be combined in ratios that would trigger such an assembly. For example, in a 2:1 mixture of peptides **A:B**, if the two peptides do not interact and solely form homotrimers, then the unfolding studies should result in two thermal transitions from the **A•A•A** and **B•B•B** homotrimers. However, if the peptides interact and a third transition peak emerges, the new peak corresponds to a heterotrimeric form of the peptides. Each system was studied using three initial ratios: 2:1, 1:1 and 1:2. These ratios were chosen based on the idea that the triple helix is composed of three peptide chains so if a heterotrimer is composed of two chains of peptide A and 1 chain of peptide B, then the ideal ratio for that system would be a 2:1 ratio of **A:B**. Once the results of the initial peptide ratios were analyzed, extra ratios such as 3:1 or 1:3 were added in order to provide an excess of the peptide that composes two-thirds of the triple helix, giving rise to the re-emergence of the homotrimer peak in the CD melting studies.

In addition to using the ratio of peptides within a system to drive heterotrimer formation, thermal annealing was also implemented during sample preparation in order to

unfold any kinetically trapped species and drive the folding of the peptides into the most thermodynamically stable species. Therefore, CD melting curves for peptide mixtures are shown as non-annealed and annealed depending on whether the system has undergone the thermal annealing process, details of which are given in the experimental section. Since each peptide formed a homotrimer, the presence of the homotrimers in the non-annealed samples was expected.

2.3.1. A/B System

A/B heterotrimer forming mixtures were made at the following ratios: 3:1, 2:1, 1:1 and 1:2. The thermal unfolding curves and the first derivatives of the melting curves for each ratio are shown in Figure 2.5. In the non-annealed cases of each ratio (the blue curves within Figure 2.5), two peaks were seen at 40 °C and 50 °C. The intensity of each peak complemented the ratio of **A:B** within the sample where the peak at 50 °C was more intense when more **B** peptide is present. Based on this observation and the fact that **B•B•B** melts at 48 °C, this peak was determined to correspond to the homotrimer of **B**. No peak was seen in the non-annealed samples at 38 °C to complement **A•A•A**. Instead, a peak at 40 °C existed which indicated the presence of a low-stability **A/B** heterotrimeric triple helix composed of an unknown combination of **A** and **B** peptides.

When the annealed versions of each ratio were examined, a different story emerged (red curves within Figure 2.5). In samples where the ratio of **A:B** is large (i.e. 3:1 ratio of **A:B**), a peak at 38 °C was seen which corresponded to **A•A•A**. In all other ratios, an undefined shoulder was seen at ~ 38 °C, which could indicate the presence of **A•A•A** or could be an artifact. However, in all annealed samples, despite the peptide ratio

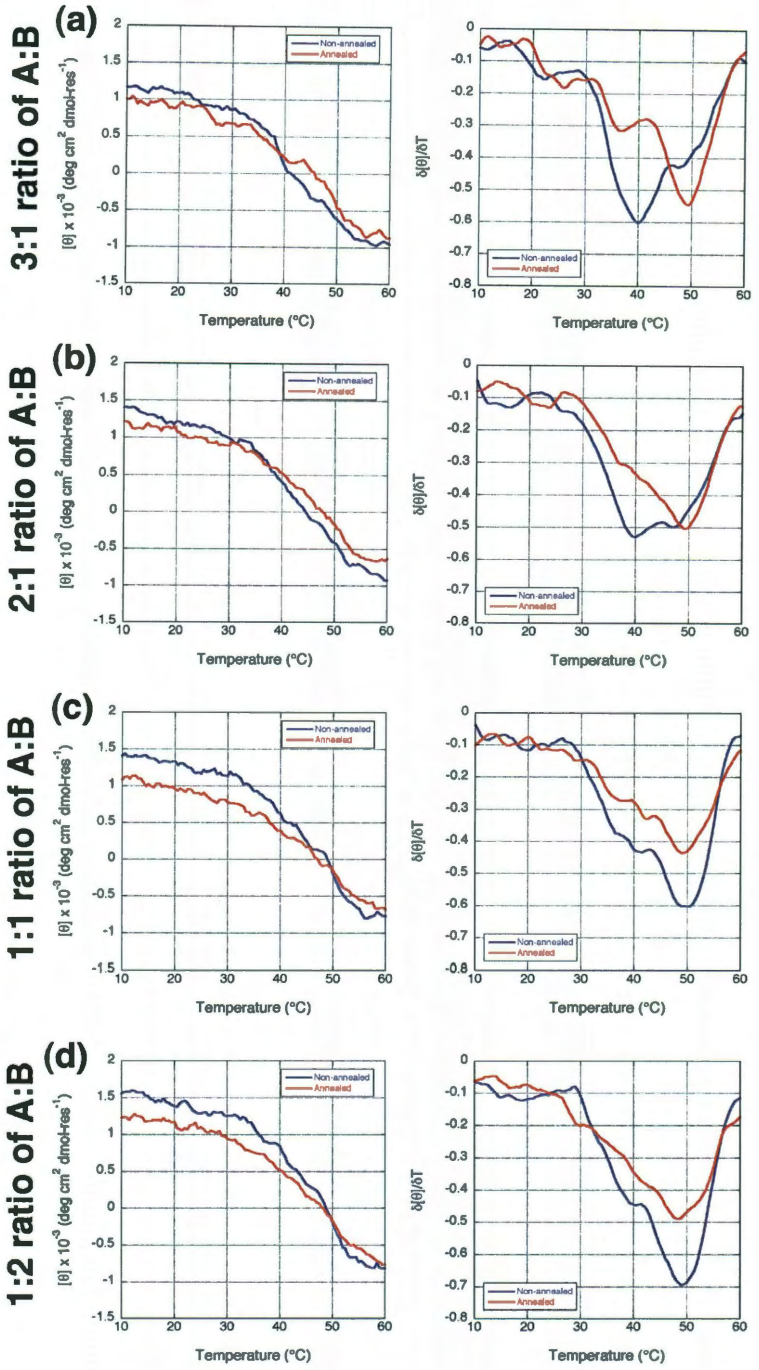

Figure 2.5. CD thermal unfolding curves for **A/B** system where peptide **A** is $(PKGPOG)_5$ and peptide **B** is $(POGDKG)_5$. The left column is the melting curve shown as MRE vs. temperature and the right column is the first derivative of the melting curve shown as the first derivative of MRE vs. temperature. Curves for the non-annealed samples are given in blue and the annealed samples are shown in red. The ratios of **A:B** examined are (a) 3:1, (b) 2:1, (c) 1:1 and (d) 1:2.

within the system, a strong peak at 50 °C was apparent. In the samples with a 3:1 ratio of **A:B**, the peak at 50 °C had a higher intensity in the annealed case than it does in the non-annealed (Figure 2.5a) suggesting the formation of a high-stability **A/B** heterotrimer post-annealing. In order to determine the composition of the high-stability heterotrimer, the intensity of the 50 °C peak in annealed samples of all **A:B** peptide ratios was examined. The peak intensity was directly proportional to the amount of **A** peptide present in the system therefore suggesting that the heterotrimer was composed of two **A** peptides and one **B** peptide. However, none of the ratios examined can be concluded to have a single species present: all showed the presence of homotrimers.

2.3.2. A/C System

Mixtures of the **A** and **C** peptides were made at the following ratios: 2:1, 1:1, 1:2 and 1:3. CD spectra from the thermal unfolding curves and the first derivative of the melting curves are shown in Figure 2.6. Similar results to the **A/B** mixtures were seen for the **A/C** non-annealed samples: two peaks were present in all ratios. The two peaks each overlapped with the melting temperature for one of the peptide homotrimers, ~ 40 °C for **A•A•A** and ~ 50 °C for **C•C•C**. However, when the ratio of peaks in the non-annealed samples was considered, especially the 1:1 ratio of peptides, the peak at 50 °C had a much higher intensity than the peptide ratio should dictate. Therefore the non-annealed samples indicate the possibility that an **A/C** heterotrimer existed with a T_m around 50 °C.

In the annealed samples, a peak at 38 °C corresponding to **A•A•A** was defined in the 2:1 and 1:1 ratios. In the 1:2 and 1:3 ratios, a shoulder around 35 to 40 °C could be seen but it was not defined. Additionally, all annealed samples showed overlapping major

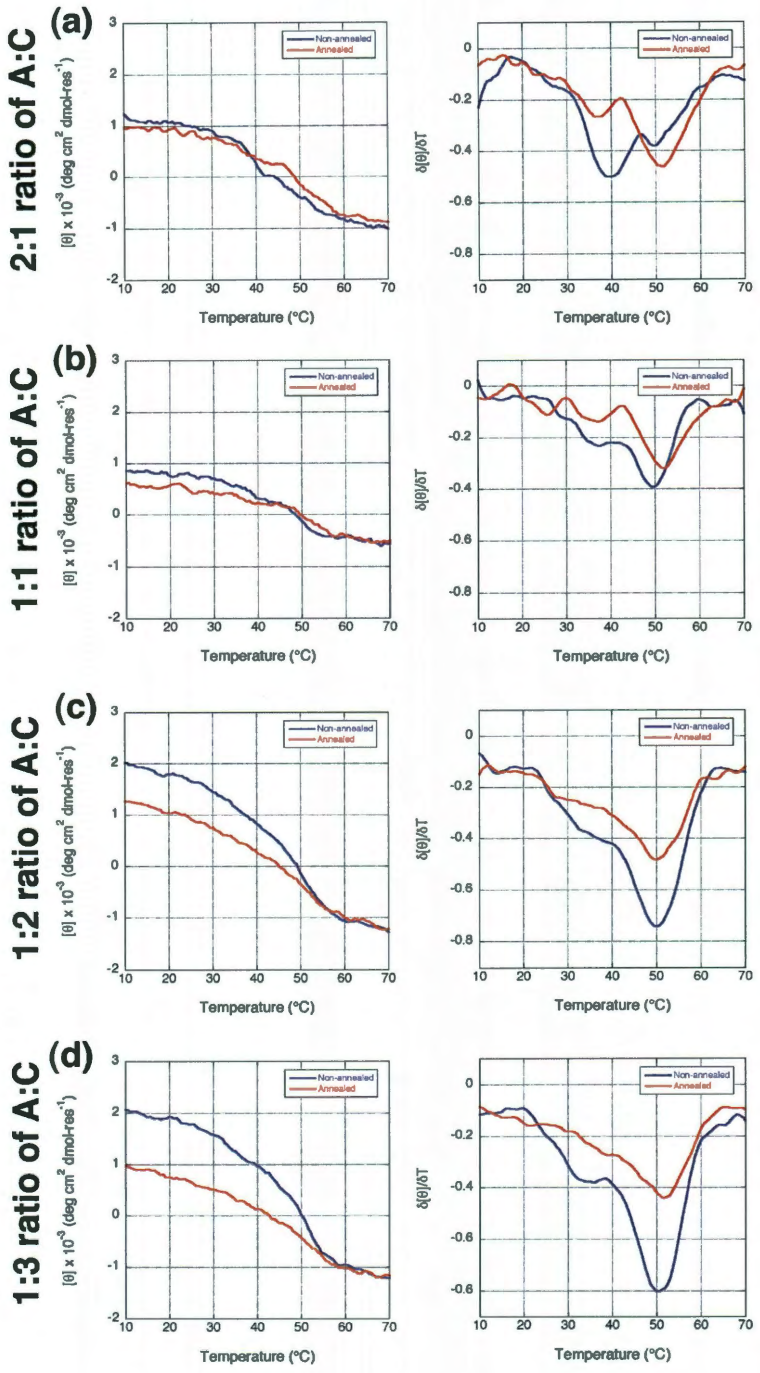

Figure 2.6. CD thermal unfolding curves for the **A/C** system where peptide **A** is $(PKGPOG)_5$ and peptide **C** is $(DKGPOG)_5$. The left column is the melting curve shown as MRE vs. temperature and the right column is the first derivative of the melting curve shown as the first derivative of MRE vs. temperature. Curves for the non-annealed samples are given in blue and the annealed samples are shown in red. The ratios of **A:C** examined are (a) 2:1, (b) 1:1, (c) 1:2 and (d) 1:3.

major peaks at 52 °C, a slightly higher value than the unfolding temperature of **C•C•C**, and the peak increased in intensity as the amount of peptide **C** in the system increased with a maximum intensity in the 1:2 ratio. The peak in the 1:3 ratio had a similar intensity to that seen in the 1:2 ratio. Therefore, the composition of the **A/C** heterotrimer could be deduced to be one chain of peptide **A** and two chains of peptide **C**. Unfortunately, similarly to the **A/B** system, all ratios studied show a shoulder or a well-defined peak at lower temperatures alluding to the fact that multiple species existed within the system.

2.3.3. B/C System

The final system, mixtures of the **B** and **C** peptides, was the least successful of the three designs systems. Samples were made in three initial ratios of peptide **B** to **C**, 2:1, 1:1 and 1:2, and none of them showed the assembly of a heterotrimeric species. CD thermal unfolding curves for each ratio are shown in Figure 2.7. Since both **B•B•B** and **C•C•C** had thermal unfolding transitions at 50 °C, a single peak in the non-annealed samples was expected. This peak was visible in all ratios however, a second peak could be seen in the 1:1 and 1:2 ratios of the peptide, which could correspond to an AAB type heterotrimer. However, when the annealed samples were examined, very low MRE values and broad transitions were seen. The lower MRE values in the unfolding curves for the annealed samples versus the non-annealed indicate that a smaller portion of the population within the mixture was folded into triple helices in the annealed cases. For the 2:1 ratio of **B:C**, a broad, shallow melting transition was seen between 35 and 55 °C (centered at 45 °C). In the annealed versions of the 1:1 and 1:2 ratios, a more distinct peak centered at 50 °C

was seen and the peak at 38 °C that seemed promising in the non-annealed samples, was no longer clear. Based on the fact that none of the three initial ratios examined indicated that an AAB heterotrimer might be present, no further ratios were examined.

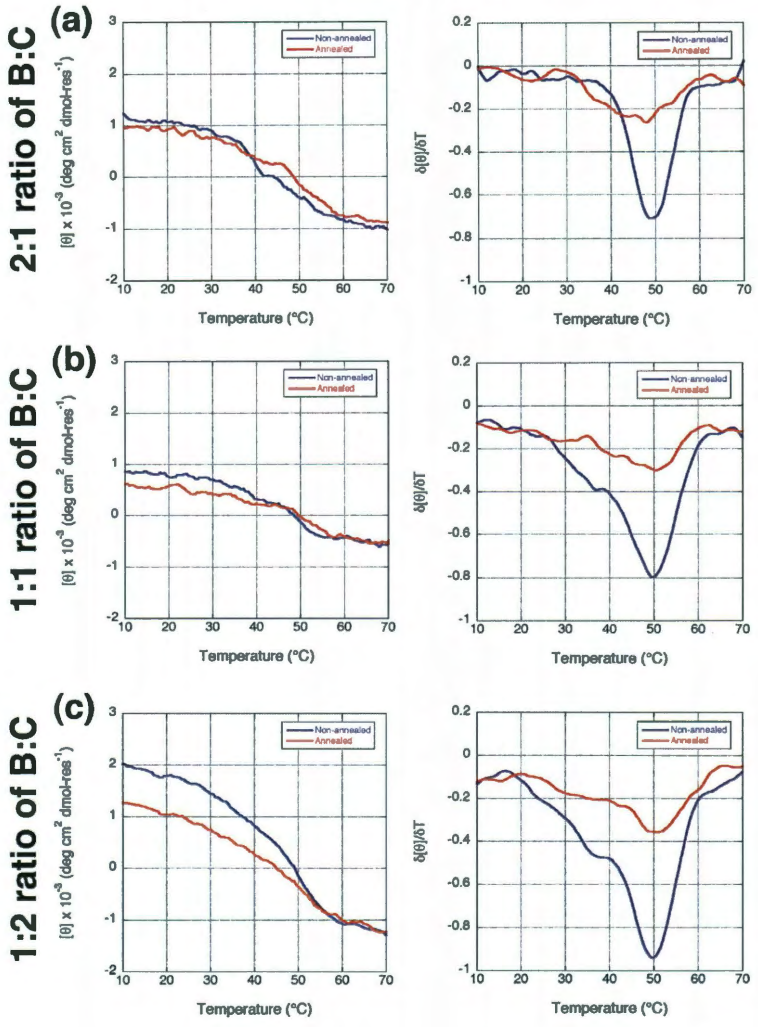

Figure 2.7. CD thermal unfolding curves for the **B/C** system where peptide **B** is $(POGDKG)_5$ and peptide **C** is $(DKGPOG)_5$. The left column is the melting curve shown as MRE vs. temperature and the right column is the first derivative of the melting curve shown as the first derivative of MRE vs. temperature. Curves for the non-annealed samples are given in blue and the annealed samples are shown in red. The ratios of **B:C** examined are (a) 2:1, (b) 1:1 and (c) 1:2.

2.4. Conclusions

Three peptides were designed and synthesized for the purpose of forming AAB type heterotrimers stabilized by lysine-aspartate interactions. In the design scheme, the peptides formed the highest number of K-D salt bridges when in an AAB heterotrimer. In combinations of the three peptides designed, $(PKGDOG)_5$, $(POGDKG)_5$ and $(DKGPOG)_5$, two of the three peptide mixtures formed AAB type heterotrimers. The first system, a mixture of $(PKGDOG)_5$ and $(POGDKG)_5$, had a heterotrimeric thermal stability of 50 °C and based on the peak intensity in different peptide ratios, the composition of the heterotrimer was hypothesized to be two chains of $(PKGDOG)_5$ and one chain of $(POGDKG)_5$. The second system, a mixture of $(PKGDOG)_5$ and $(DKGPOG)_5$, showed a T_m for the heterotrimer of 52 °C and based on the peptide ratios, the AAB heterotrimer was composed of one chain of $(POGDKG)_5$ and two chains of $(DKGPOG)_5$. The third system, a mixture of $(POGDKG)_5$ and $(DKGPOG)_5$, did not show the formation of an AAB type heterotrimer in the annealed samples of any ratio examined.

Despite the presence of AAB type heterotrimers in the first two systems, residual homotrimers were seen for both mixtures. Therefore, neither were viable heterotrimer systems. The ability of the lysine and aspartate residues within each peptide to form K-D interactions within a homotrimer proved to be the major drawback to this peptide design scheme; the presence of lysine and aspartate within each peptide was not necessarily the issue. By attempting to maximize the number of K-D interactions present in an AAB heterotrimer, we inadvertently created lysine-aspartate bridges within homotrimers thus stabilizing these undesired species. Therefore, in future peptide design schemes for the

formation of AAB heterotrimers, the ability of each peptide to form homotrimers must be discouraged.

2.5. Experimental

Peptide Synthesis. All peptides were synthesized using an Advanced Chemtech Apex 396 multipeptide automated synthesizer. Standard Fmoc chemistry for solid-phase peptide synthesis was used to synthesize all peptides at a 0.15 mM scale on Wang resin pre-loaded with glycine. Amino acids were added in a 4:1 molar ratio to the growing peptide chain using the coupling agents *O*-benzotriazole-*N*,*N*,*N*',*N*'-tetramethyluroniumhexafluorophosphate (HBTU), 1-hydroxybenzotriazole hydrate (HoBt), and *N*,*N*-diisopropylethylamine (DiEA) in dimethylformamide (DMF) at molar ratios of 4:4:6 respectively. Once coupled to the peptide chain, amino acids were deprotected using a 25 % (by volume) solution of piperidine in DMF. The peptide was cleaved from the resin with a 38:1:1 mixture of trifluoroacetic acid (TFA), triisopropylsilane, and water.

Mass Spectrometry. Post synthesis, all peptides were examined on a Bruker Autoflex mass spectrometer in positive ion mode to verify that the peptides were synthesized correctly. Spectra were analyzed using FlexAnalysis software.

Peptide Purification. Purification was performed on a Varian PrepStar220 HPLC using a preparative reverse phase C-18 column. The two HPLC solvents referred to as solvents A and B are water and acetonitrile respectively, each containing 0.05 % TFA.

The solvents were eluted through the column with a linear gradient ranging from a 1 to 3 % increase in concentration of solvent B per minute. Once collected, the HPLC fractions were rotovapped down to remove the acetonitrile fraction and then lyophilized.

Sample Preparation. After all peptides were purified and lyophilized, stock solutions for each peptide were made with a 2 mM peptide concentration (measured by mass). Samples were then made with a total peptide concentration of 0.2 mM in 10 mM sodium phosphate buffer, pH 7. Non-annealed samples were immediately incubated at 10 °C overnight before any characterization was performed. Annealed samples were made, preheated for 15 minutes at 85 °C, then incubated at 10 °C overnight. All samples were allowed to incubate for at least 12 hours before CD analysis was performed.

Circular Dichroism. All CD experiments were performed on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature control system. Spectra were taken from 250-190 nm with a pitch of 1 nm and a bandwidth of 2 nm. Thermal unfolding curves monitoring the wavelength of the maximum seen in the pre-melting spectra for a given sample (between 222 and 225 nm) were obtained in a range of 5 to 75 °C at a heating rate of 10 °C per hour. The first derivative of the melting curve was taken in order to find the transition temperature of the sample. The molar residual ellipticity (MRE) was calculated from the measured ellipticity from the equation:

$$[\theta] = \frac{\theta \times m}{10 \times c \times l \times n_r}$$

where θ is the ellipticity in mdeg, m is the molecular weight in g/mol, c is the concentration in mg/mL, l is the path length of the cuvette in cm, and n_r is the number of

amino acids in the peptide. All data reported for melting studies was taken every 0.5 °C and smoothed with a 5.5 °C width in order to emphasize the thermal transitions seen.

References

(1) Heino, J. *Bioessays* **2007**, *29*, 1001-1010.

(2) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 2683-2690.

(3) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 15034-15041.

(4) Fallas, J. A.; Gauba, V.; Hartgerink, J. D. *J. Biol. Chem.* **2009**, *284*, 26851-26859.

(5) Hodges, J. A.; Raines, R. T. *J. Am. Chem. Soc.* **2005**, *127*, 15923-15932.

(6) Madhan, B.; Xiao, J. X.; Thiagarajan, G.; Baum, J.; Brodsky, B. *J. Am. Chem. Soc.* **2008**, *130*, 13520-13521.

(7) Slatter, D. A.; Foley, L. A.; Peachey, A. R.; Nietlispach, D.; Farndale, R. W. *J. Mol. Biol.* **2006**, *359*, 289-298.

(8) Slatter, D. A.; Miles, C. A.; Bailey, A. J. *J. Mol. Biol.* **2003**, *329*, 175-183.

(9) Persikov, A. V.; Ramshaw, J. A.; Kirkpatrick, A.; Brodsky, B. *Biochemistry* **2000**, *39*, 14960-14967.

Chapter 3: AAB Heterotrimer Formation Incorporating Negative Design*

The main drawback from the AAB heterotrimer systems designed in Chapter 2 was the ability of the designed peptides to form stable homotrimers and this observation has been the major drawback for heterotrimeric systems designed from CMPs.[1-6] Therefore in order to improve upon the previous design strategy and the competing designs reported in literature, negative design, that discouraged the formation of homotrimers, was incorporated into the design of AAB type heterotrimers.

3.1. Peptide Design

Homotrimer formation observed with $(PKGDOG)_5$, $(POGDKG)_5$ and $(DKGPOG)_5$ was largely stabilized by the presence of large numbers of lysine and aspartate residues within each peptide chain. This allowed for the formation of stabilizing K-D salt bridges within the homotrimer. In order to improve upon this design, the new strategy separated the positively and negatively charged amino acids onto two distinct peptide chains. Hence, one peptide chain had a net positive charge and the other, a net negative, which discouraged homotrimer formation due to charge repulsion. In order to keep the AAB heterotrimer formed by these peptides neutral, the more abundant peptide

* The work in this chapter was done in collaboration with Jorge Fallas of the Chemistry Department, who is advised by Prof. Jeff Hartgerink. Jorge performed all of the NMR and molecular modeling experiments. This chapter is based on the following publications:

1) Lesley E. Russell, Jorge A. Fallas and Jeffrey D. Hartgerink. Selective Assembly of a High-Stability AAB Collagen Heterotrimer. *J. Am. Chem. Soc.*, **2010**, 132, 3242-3243.
2) Lesley E. R. O'Leary, Jorge A. Fallas and Jeffrey D. Hartgerink. Positive and Negative Design Leads to Compositional Control in AAB Collagen Heterotrimers, *J. Am. Chem. Soc.*, **2011**, 133, 5432-5443.

within the heterotrimer has a charge ½ and opposite of the other peptide, resulting in an electrostatically neutral AAB heterotrimer.

Lastly, the placement of charged residues within the POG-containing peptides was implemented in the peptides from Chapter 2 and was continued within the new peptide design scheme in order to discourage homotrimer formation. In general for CMPs, the more POG repeats present in a peptide chain, the more likely it will be to form a homotrimer.[7] Therefore, in order to prevent homotrimer formation, POG triplets were spaced out within the designed peptides.

3.2. Peptide Library

The isolation of the positively and negatively charged residues within different peptide chains allowed for the incorporation of arginine and glutamate into the peptide library, not just lysine and aspartate. Therefore the entire peptide library includes positively charged peptides containing either arginine or lysine with a net charge of +10 or +5 and negatively charged peptides containing either aspartate or glutamate and a net charge of −10 or −5. The library of peptides and their homotrimeric melting temperatures (T_m) are given in Table 3.1 using single letter amino acid codes (arginine = R, aspartate = D, glutamate = E, glycine = G, lysine = K, proline = P and hydroxyproline = O).[7,8]

3.3. Homotrimer Stability

Before mixtures of peptides could be examined, the ability of each peptide to form homotrimers was assessed. All peptides were studied in three buffers, all pH 7: 10

mM sodium phosphate (phosphate), 10 mM tris(hydroxymethyl)-aminomethane, (Tris), and 10 mM Tris 150 mM sodium chloride (Tris/NaCl).[7,8]

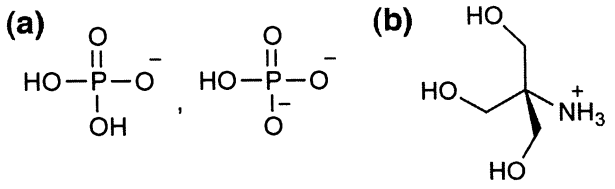

Figure 3.1. Structures of (a) phosphate and (b) Tris at pH 7. Phosphate is shown as monobasic (left) and dibasic (right).

Previously, we primarily used phosphate buffers in heterotrimeric systems.[1-3,9] Recently, we began exploring the differences in homotrimer and heterotrimer stabilities in different buffers as we observed that in some instances, triple helical stability can be highly dependent both on buffer composition and overall ionic strength, as would be expected for systems whose assembly is highly dependent on charged pair interactions.[7] More specifically, we investigated phosphate versus Tris to explore the effect of anionic versus cationic buffers and how the addition of NaCl to increase the ionic strength affects melting temperatures. The structures of phosphate and Tris are depicted in Figure 3.1.

Peptide	Phosphate	Tris	Tris/NaCl
$(PRG)_{10}$	none	none	37 °C
$(PKG)_{10}$	none	none	none
$(EOG)_{10}$	none	none	none
$(DOG)_{10}$	none	39.5 °C	37.5 °C
$(PRGPOG)_5$	55.5 °C	56 °C	56.5 °C
$(PKGPOG)_5$	none	none	none
$(EOGPOG)_5$	46 °C	46 °C	46.5 °C
$(POGEOG)_5$	43 °C	43 °C	43.5 °C
$(POGDOG)_5$	35.5 °C	33 °C	35.5 °C

Table 3.1. Peptide Library and T_m of corresponding homotrimer formation as determined by CD. Adapted from Table 2 in reference 7.[7]

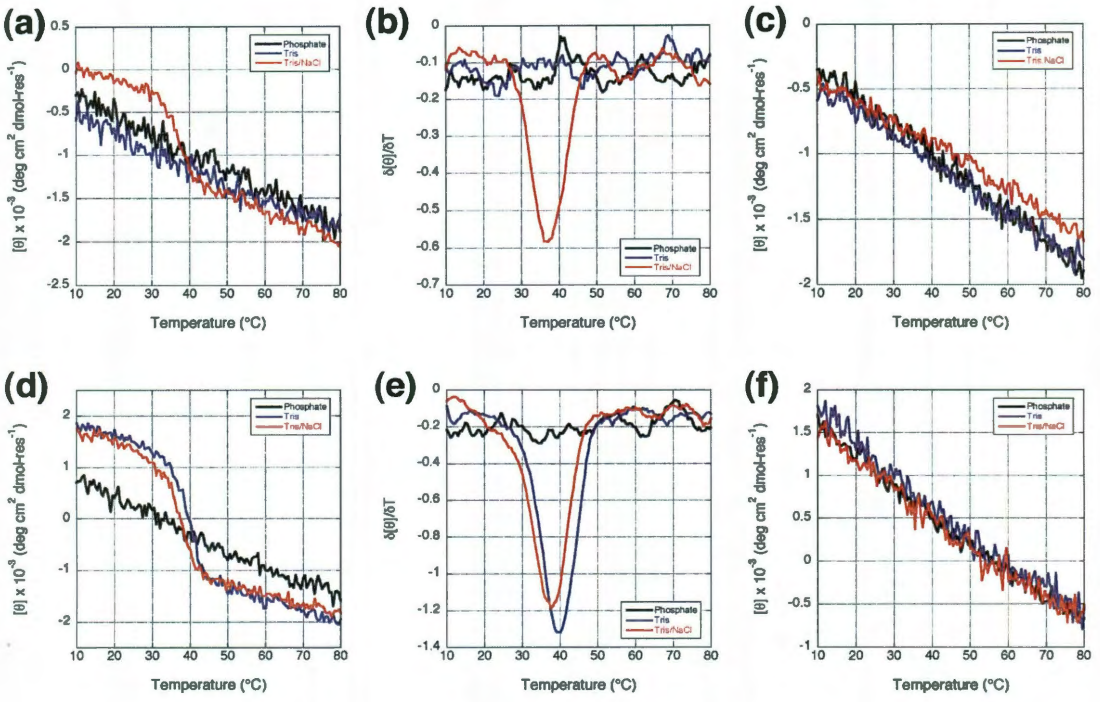

Figure 3.2. CD thermal unfolding curves shown as MRE vs. temperature and first derivative of MRE vs. temperature for all homotrimers with a ±10 charge in phosphate (black), Tris (blue) and Tris/NaCl (red): (a) $(PRG)_{10}$ melting curve, (b) $(PRG)_{10}$ first derivative, (c) $(PKG)_{10}$ melting curve, (d) $(DOG)_{10}$ melting curve, (e) $(DOG)_{10}$ first derivative and (f) $(EOG)_{10}$ melting curve. Adapted from Figure S7 in reference 7.[7]

As mentioned above, a negative design technique was implemented that reduced the number of POG repeats present in a peptide chain in order to discourage homotrimer formation as much as possible. It is for this reason that we do not expect ± 10 charged peptides to form homotrimers, but are not surprised when ± 5 charged peptides assemble into homotrimeric helices. Detailed melting profiles for all peptides are given in Figure 3.2 (peptides with ± 10 charge) and Figure 3.3 (peptides with ± 5 charge). CD spectra are reported as molar residual ellipticity (MRE), which normalized the data for peptide concentration, peptide length and pathlength (see experimental section for more details). When examining Table 3.1, three observations can be made: 1) $(PRG)_{10}$ and $(DOG)_{10}$ did

form homotrimers in at least one of the buffer systems despite the putative charge repulsion preventing such assembly, 2) $(EOGPOG)_5$ and $(POGEOG)_5$ had identical amino acid compositions yet their homotrimer stabilities differed by 3 °C, 3) $(PKGPOG)_5$ did not form a homotrimer in any buffer system tested despite the presence of five POG triplets within the peptide and having just ½ the charge of the above peptides.[7]

Based on previous results,[1-3,9] we did not expect for any of the peptides with ± 10 charges to form homotrimers in any of the low ionic strength buffers. When mixed with high ionic strength buffers, host-guest peptides containing one or more PRG triplets had been previously shown to form a homotrimer in PBS (10 mM phosphate with 150 mM NaCl), pH 7, and, in the same study, $(PRG)_8$ formed a homotrimer in 10 mM sodium phosphate, 2 M NaCl, pH 7.[10] Complementing those studies, we found that $(PRG)_{10}$ formed a homotrimer with a melting temperature of 37 °C in Tris/NaCl, another high ionic strength buffer. In differential scanning calorimetry (DSC) experiments on $(PRG)_{10}$, shown in Figure 3.4, a small population of peptide formed a homotrimer in phosphate and in Tris/NaCl.[8] The peak for the homotrimer was seen in the first peptide scan and was absent in all subsequent scans indicating that the $(PRG)_{10}$ homotrimer was weak in stability and re-folded in a timescale larger than that of the DSC experiment.[8] An understanding of these results was not established until NMR experiments on peptide mixtures were performed and full details of the conclusions will be described in the NMR section below. Briefly, the guanadinium within the side chain of arginine interacted with a backbone carbonyl of hydroxyproline in an adjacent strand stabilizing the homotrimer. This interaction was masked by charge repulsion in low ionic strength buffers. However in PBS and Tris/NaCl, the charges on the arginine side chains were screened by the high

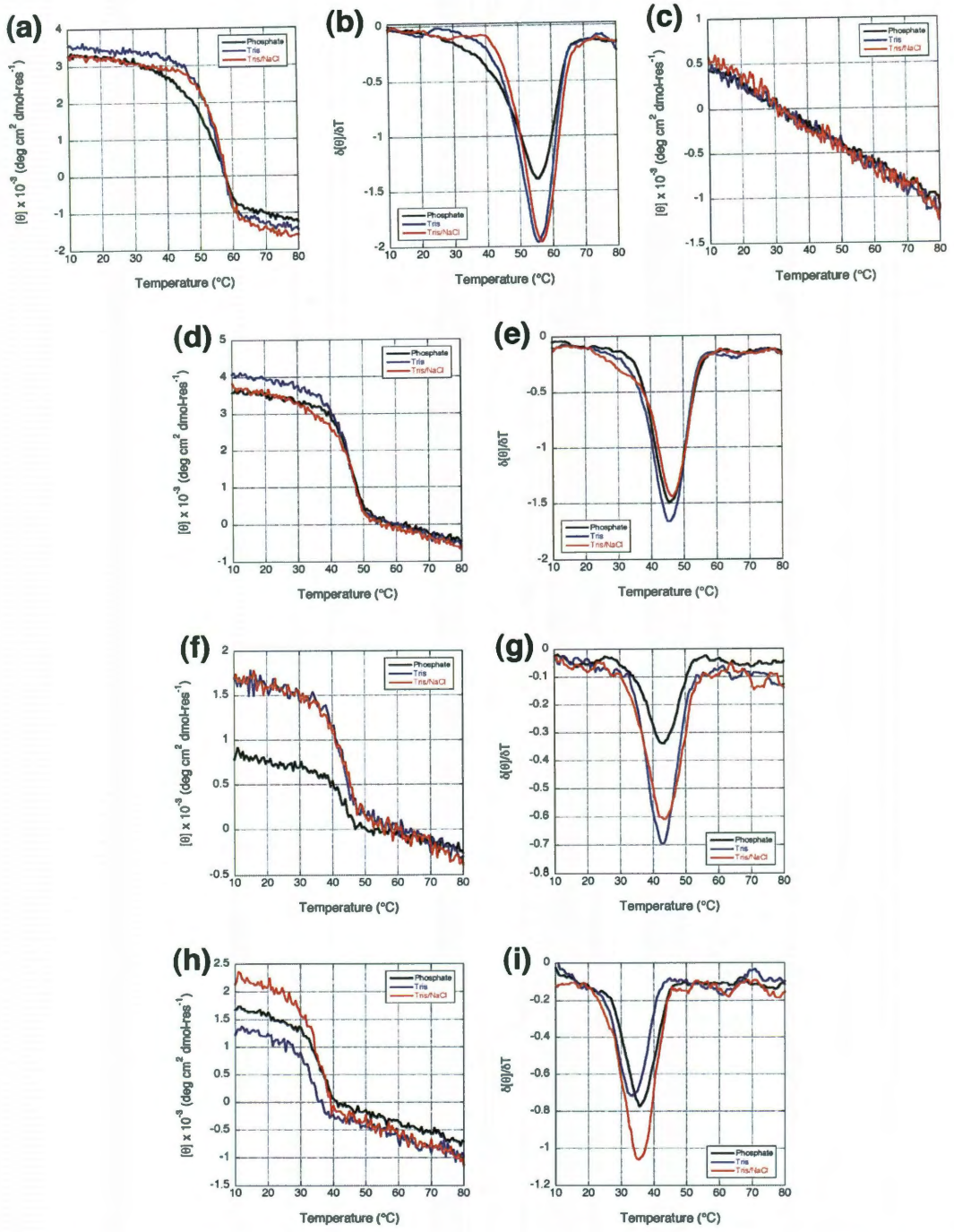

Figure 3.3. CD thermal unfolding curves shown as MRE vs. temperature and first derivative of MRE vs. temperature for all homotrimers with a ±5 charge in phosphate (black curve), Tris (blue curve) and Tris/NaCl (red curve): (a) $(PRGPOG)_5$ melting curve, (b) $(PRGPOG)_5$ first derivative, (c) $(PKGPOG)_5$ melting curve, (d) $(EOGPOG)_5$ melting curve, (e) $(EOGPOG)_5$ first derivative, (f) $(POGEOG)_5$ melting curve, (g) $(POGEOG)_5$ first derivative, (h) $(POGDOG)_5$ melting curve and (i) $(POGDOG)_5$ first derivative. Adapted from Figure S8 in reference 7 and Figure 1 in reference 8.[7,8]

salt concentration, allowing for some of the peptide population to form homotrimers.[8]

The more surprising result within the homotrimer study on ± 10 charged peptides was that $(DOG)_{10}$ formed a homotrimer in both Tris and Tris/NaCl buffers. A hypothesis for this observation was that that the cationic nature of the Tris buffer might allow for a specific stabilizing interaction (as opposed to simple charge screening) between Tris and the negatively charged aspartate preventing side chain charge repulsion and allowing for triple helix formation in the lower ionic strength buffer.[7] This interaction was then weakened with the addition of NaCl reducing the T_m of the homotrimer by 2 °C.

The second observation is the difference in homotrimer stabilities between $(EOGPOG)_5$ and $(POGEOG)_5$. A similar result was described in Chapter 2 with the 2 °C difference in melting temperatures between $(POGDKG)_5$ and $(DKGPOG)_5$. Since $(EOGPOG)_5$ had a melting temperature of 46 °C and $(POGEOG)_5$ had a T_m of 43 °C, we concluded that the glutamate in the triplet at the N-terminus of $(EOGPOG)_5$ was less destabilizing than the glutamate in the final triplet at the C-terminus of $(POGEOG)_5$. This would result in a slightly higher thermal stability for the former peptide compared to the latter which complements the experimental findings.

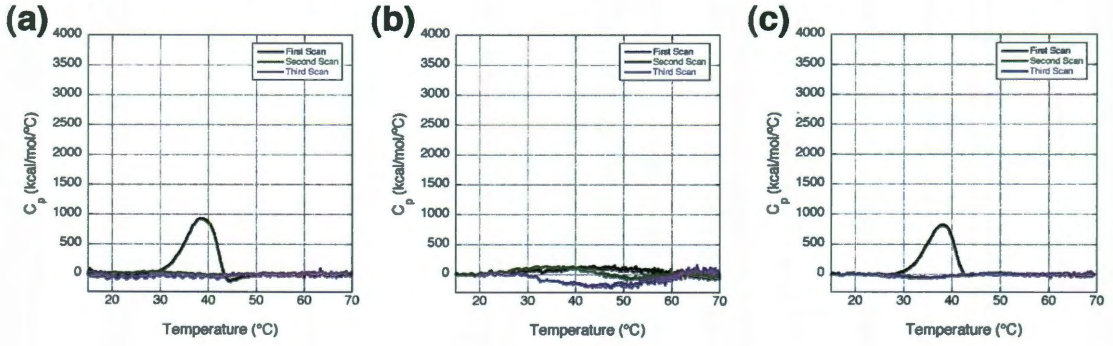

Figure 3.4. DSC melting profiles for $(PRG)_{10}$ in (a) phosphate, (b) Tris and (c) Tris/NaCl. Adapted from Figure S5 in reference 8. The first three peptide scans for each buffer are shown with the first in black, the second in green and the third in blue.[8]

The third observation and arguably the most notable result within this table, is the inability of $(PKGPOG)_5$ to form a homotrimer in any of the three buffers formed. Based on previous work by Gauba *et al.*, in 2007,[2,3] peptides with ± 5 charges were expected to fold in phosphate buffer with T_m values between 30 and 40 °C. One possible explanation for the current observation was based on the amino acid propensity for forming stable collagen triple helices which had shown that lysine in the Y-position of a collagen triplet had a far lower stability than arginine, causing it to have a lower propensity for triple helical formation.[7] From the perspective of negative design, the inability of this peptide to form a stable homotrimer made it unique within the library as the lone peptide that, when combined with a −10 charged peptide, could potentially form an AAB heterotrimer in the absence of either peptide forming a homotrimer.[7]

3.4. Proof of Design Principles

Before extensive studies were carried out to explore all of the possible systems from the peptide library, a single peptide mixture was chosen to test the design principles. The mixture of $(PRG)_{10}$ and $(EOGPOG)_5$ was examined in ratios of 2:1, 1:1 and 1:2 respectively in order assess whether the preferred ratio of 1:2 based on the peptide design does in fact form a heterotrimer. CD melting experiments were performed on each ratio in Tris buffer and the results are shown in Figure 3.5. Both the non-annealed and annealed samples for all three ratios showed a strong peak at 56 °C, which did not correspond to the $(EOGPOG)_5$ homotrimer (T_m of 46 °C).[8] The intensity of this peak was larger in the annealed samples versus the non-annealed samples and it increased in size with the increasing concentration of $(EOGPOG)_5$. In the 1:2 ratio of peptides, both the

non-annealed and annealed samples showed a single transition at 56 °C corresponding to an AAB type heterotrimer. In depth analysis of the heterotrimer formed from the mixture of these peptides is given below. However, in brief, this peptide system was the first reported AAB type heterotrimer with a thermal stability higher than the homotrimers of any of the component peptides.[8] Therefore, the peptide design based on the mixture of oppositely charged peptides to generate zwitterion AAB triple helices was validated.

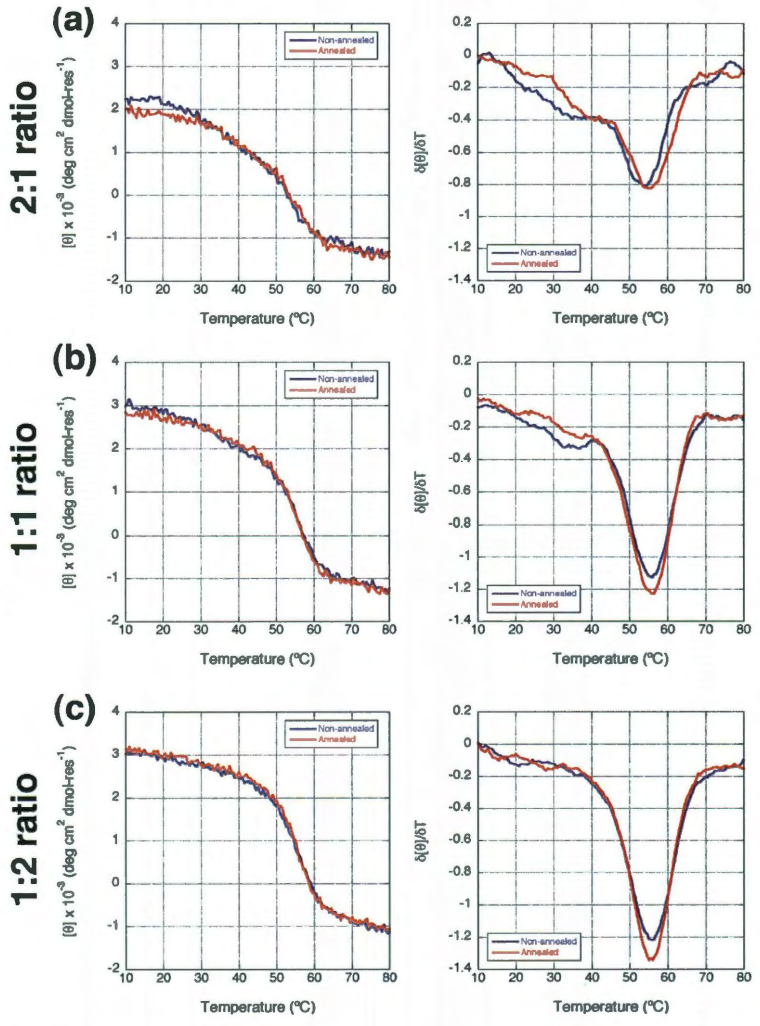

Figure 3.5. CD melting experiments on mixtures of $(PRG)_{10}$ and $(EOGPOG)_5$ at (a) 2:1, (b) 1:1 and (c) 1:2 ratios. The melting curve, shown as MRE vs. temperature, is in the left column and the first derivative of the melting curve vs. temperature is shown in the right column. Non-annealed samples are in blue and annealed samples are in red.[8]

3.5. CD Analysis of Heterotrimer Formation

Once the peptide design was validated, peptides were mixed in a 2:1 ratio such that the more abundant peptide had a charge ½ and opposite of the less abundant peptide, resulting in the formation of a zwitterionic, neutral AAB heterotrimer.

There are three major points to discuss before an in depth analysis of each charged pairing progresses. First, all of the peptide mixtures examined that form heterotrimers had their highest melting temperatures in Tris buffer, reiterating the ideas proposed during the homotrimer discussion that Tris had a stabilizing effect on negatively charged peptides.[7] Second, the direct comparison of results in Tris versus Tris/NaCl exposed the charge shielding that resulted from a higher ionic strength buffer which, depending on the system, either hid or unveiled the presence of an AAB heterotrimer by altering the relative thermal stability of homo- versus heterotrimers.[7] This characteristic demonstrated the versatility of the design system for AAB heterotrimer formation and the challenges associated with it because, by adjusting the ionic strength of the buffer used, visible heterotrimers unfolding transitions could be seen for three out of the four amino acid pairings examined in the proper buffer composition. Third, the ability of two of the peptide systems to form heterotrimers when none of the component peptides formed homotrimers strongly suggested that we have compositional control over triple helix assembly based on the CD melting studies.[7] This point was confirmed using 2D solution NMR experiments (described below), making these the first reported heterotrimeric system in which there was complete control over heterotrimer composition.

All ten peptide systems explored using this design will be described below based on the charged amino acids used within the system: arginine-aspartate (R-D), arginine-

glutamate (R-E), lysine-glutamate (K-E) and lysine-aspartate (K-D). All peptide mixtures and the transitions seen in CD melting studies in all three buffers are listed in Table 3.2 and are organized based on the charged residue pairing.[7,8] If the unfolding transition seen in mixtures overlapped any homotrimer transitions within the range of ± 2 °C, "overlap" is written in the table. If the transition did not overlap, the melting temperature is listed. Last, if multiple transitions were seen in CD melting studies indicating the inability of a system to form an AAB heterotrimer, "mult. pks." is listed in the table.

Peptide Mixture	T_m Phosphate	T_m in Tris	T_m in Tris/NaCl
$(PRG)_{10}$•2$(POGDOG)_5$	mult. pks.	mult. pks.	overlap
2$(PRGPOG)_5$•$(DOG)_{10}$	mult. pks.	mult. pks.	mult. pks.
$(PRG)_{10}$•2$(EOGPOG)_5$	51 °C	56 °C	overlap
$(PRG)_{10}$•2$(POGEOG)_5$	47 °C	51.5 °C	overlap
2$(PRGPOG)_5$•$(EOG)_{10}$	52.5 °C	53 °C	47 °C
$(PKG)_{10}$•2$(EOGPOG)_5$	43 °C	overlap	40.5 °C
$(PKG)_{10}$•2$(POGEOG)_5$	40 °C	overlap	39 °C
2$(PKGPOG)_5$•$(EOG)_{10}$	**45 °C**[a]	**45 °C**[a]	**42 °C**[a]
$(PKG)_{10}$•2$(POGDOG)_5$	44 °C	48 °C	43 °C
2$(PKGPOG)_5$•$(DOG)_{10}$	**46 °C**[a]	48 °C	43 °C

[a]Heterotrimer formation when neither of the component peptides forms homotrimers

Table 3.2. Peptide mixtures and melting temperatures measured by CD. Adapted from Table 2 in reference 7.[7]

3.5.1. Arginine-Aspartate Pairing

Beginning with the arginine-aspartate pairing, $(PRG)_{10}$•2$(POGDOG)_5$ and 2$(PRGPOG)_5$•$(DOG)_{10}$, it is immediately apparent from Table 3.2 that neither of the two

systems examined using the combination of these residues formed a clean heterotrimer: multiple peaks are seen in all peptide mixtures.[7]

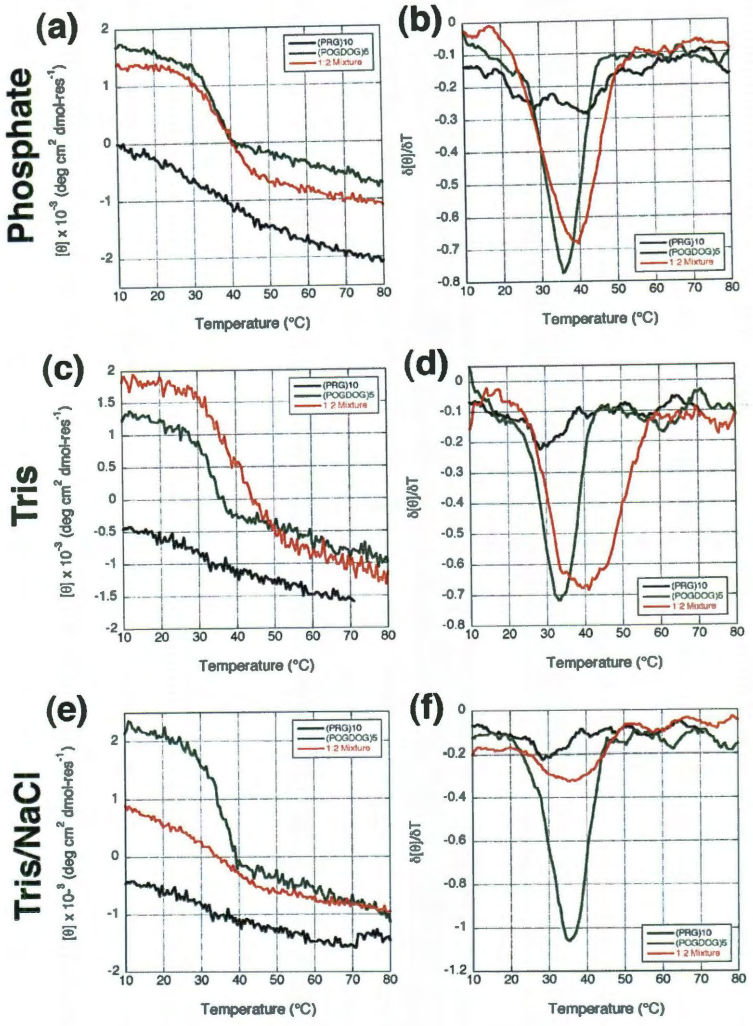

Figure 3.6. CD thermal unfolding curves shown as MRE vs. temperature (left column) and first derivative of MRE vs. temperature (right column) for the $(PRG)_{10}$•2$(POGDOG)_5$ system in phosphate (a and b), Tris (c and d) and Tris/NaCl (e and f). Adapted from Figure S9 in reference 7. Data for component peptides are shown as the black and green curves and the 1:2 mixture of the peptides is shown as the red curve.[7]

Figure 3.6 gives the results from the CD melting studies for the $(PRG)_{10}$•2$(POGDOG)_5$ in all three buffers. The component peptides, $(PRG)_{10}$ and

$(EOGPOG)_5$, and the annealed 1:2 mixtures of the peptides are shown in black, green and red respectively. In phosphate and Tris, the peaks seen for the annealed mixture overlapped with the $(EOGPOG)_5$ homotrimer, but are also very broad indicating that multiple species were present within the system. In Tris/NaCl, the intensity for the mixture drastically decreased in comparison to the other buffers implying the lack of a stable AAB heterotrimer formed in this buffer.

Distinct multiple peaks were seen in all three buffers for the $2(PRGPOG)_5•(DOG)_{10}$ system, shown in Figure 3.7. One peak corresponded to the $(PRGPOG)_5$ homotrimer at 56 °C and the other at about 45 °C possibly corresponding to an AAB heterotrimer. Unlike the previous system, $(PRG)_{10}•2(POGDOG)_5$, the heterotrimer peak was distinct from any homotrimeric species and had it's largest intensity in phosphate buffer. However, in all samples, a residual homotrimer peak was present therefore $2(PRGPOG)_5•(DOG)_{10}$ was not a viable AAB heterotrimeric system.

In order to understand the results for both R-D containing systems, we turn to our conclusion for homotrimer stabilization in PRG containing peptides: the interaction between the guanadinium in the arginine side chain with the backbone carbonyl of hydroxyproline in an adjacent strand, which will be further explained in the NMR section below.[8] When a heterotrimer containing these peptides was formed, we believe that this interaction was still present which prevented the arginine terminal amines from forming a salt bridge with the carboxyl group on aspartate.[8] It is for this reason that, even if a heterotrimer peak was seen in mixtures of arginine and aspartate containing peptides, homotrimer peaks could still be seen. This has a deleterious effect on both positive and negative design: the desired heterotrimer is destabilized as optimal conformations

allowing arginine-aspartate interactions are prevented while simultaneously stabilizing unwanted homotrimers through arginine-backbone hydrogen bonding.[8]

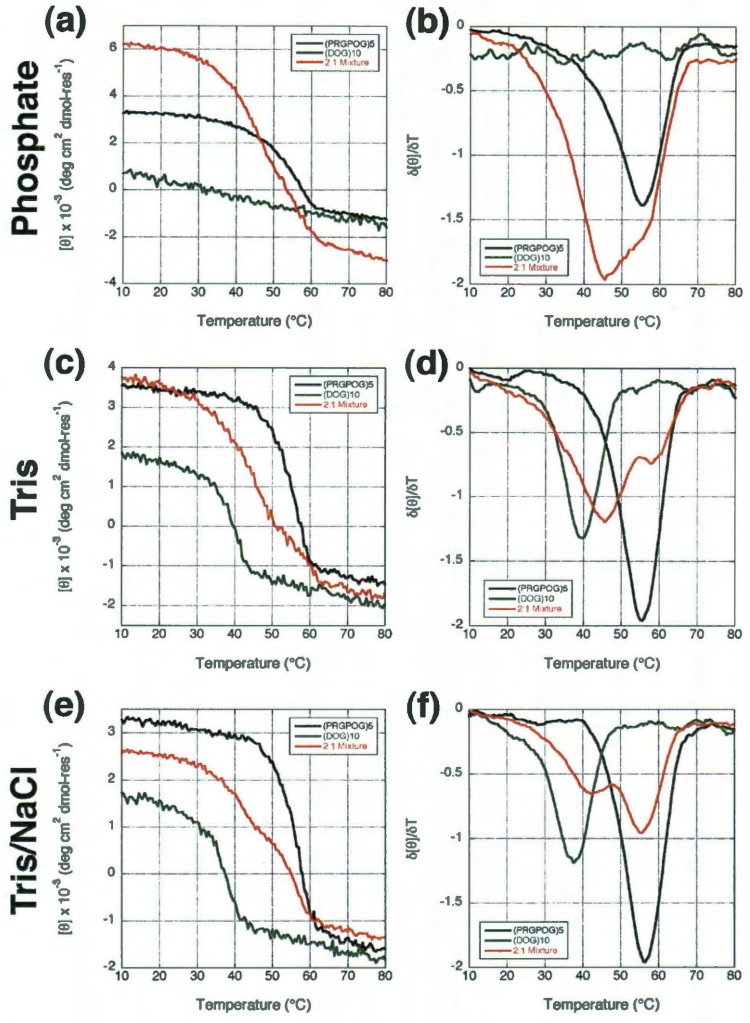

Figure 3.7. CD thermal unfolding curves shown as MRE vs. temperature (left column) and first derivative of MRE vs. temperature (right column) for the $2(PRGPOG)_5 \bullet (DOG)_{10}$ system in phosphate (a and b), Tris (c and d) and Tris/NaCl (e and f). Adapted from Figures 1 and S10 in reference 7. Data for component peptides are shown as the black and green curves and the 2:1 mixture of the peptides is shown as the red curve.[7]

3.5.2. Arginine-Glutamate Pairing

When we considered the arginine-glutamate pairing, a drastic difference could be seen when compared to the arginine-aspartate coupling: heterotrimer formation could be seen in at least one buffer for all three systems.[7,8] The CD thermal unfolding studies in all three buffers for $(PRG)_{10}$•2$(EOGPOG)_5$, $(PRG)_{10}$•2$(POGEOG)_5$ and 2$(PRGPOG)_5$•$(EOG)_{10}$ are given in Figures 3.8, 3.9 and 3.10 respectively.

Beginning with the system $(PRG)_{10}$•2$(EOGPOG)_5$ in phosphate and Tris, a heterotrimer was clearly indicated with a melting temperature of 51 °C and 56 °C respectively. The result in Tris buffer marked the first synthetic peptide system that selectively formed a collagen-like heterotrimer with a thermal stability higher than any of its component peptides: 10 °C higher than the $(EOGPOG)_5$ homotrimer.[8] In Tris/NaCl, the heterotrimer was destabilized by the higher ionic strength causing the melting temperature of the heterotrimer to overlap with the $(EOGPOG)_5$ homotrimer. Due to the novelty of this system as the first to form with a heterotrimeric stability higher than it's component homotrimers, further analysis in DSC and NMR was performed, which will be discussed in the respective sections below.

The $(PRG)_{10}$•2$(POGEOG)_5$ mixture behaved very similarly to the $(PRG)_{10}$•2$(EOGPOG)_5$ systems described above. A clear peak corresponding to a heterotrimer could be seen in the phosphate and Tris buffers while the results in Tris/NaCl overlap with the $(POGEOG)_5$ homotrimer.[7] The melting temperature for the heterotrimer in Tris was higher than the component homotrimers however only by 8.5 °C, a smaller margin than the $(PRG)_{10}$•2$(EOGPOG)_5$ system. The similarity between the two systems could be attributed to the similarity between the two negatively charged peptides:

they only differed in the order or the repeating triplets. Insight on how this difference in order affected triple helix stability was given in the homotrimer section above and we hypothesized that a similar argument applies here. The final item that must be noted was the lack of clear heterotrimer formation for the $(PRG)_{10}$-containing systems in Tris/NaCl: $(PRG)_{10}$•2$(EOGPOG)_5$ and $(PRG)_{10}$•2$(POGEOG)_5$. The increase in ionic strength destabilized the non-specific charge interaction between the arginine and glutamate residues causing a decrease in the melting temperature of the triple helix, making it comparable to the T_m for the $(EOGPOG)_5$ and $(POGEOG)_5$ homotrimers respectively.[7] This masking and unveiling of heterotrimer formation based on the ionic strength of buffers will be discussed further below. CD thermal unfolding studies for the $(PRG)_{10}$•2$(POGEOG)_5$ system are shown in Figure 3.9.

The last system within the R-E pairing is 2$(PRGPOG)_5$•$(EOG)_{10}$ and in contrast to the first two systems, a defined heterotrimer could be seen in all three buffers. The CD thermal unfolding curves for this system are shown in Figure 3.10. The heterotrimers had melting temperatures of 52.5 °C, 53 °C and 47 °C in phosphate, Tris and Tris/NaCl respectively.[7] Although the system did show distinct heterotrimers in all buffers, the $(PRGPOG)_5$ homotrimer had a higher stability in each case. Therefore, none of the heterotrimers formed in this system were an improvement on previously reported systems[4-6] or the $(PRG)_{10}$•2$(EOGPOG)_5$,[8] which is the system within the R-E pairing with the highest stability. The success in heterotrimer formation for systems containing R-E as compared to R-D systems previously discussed was attributed to the fact that glutamate had one more methylene group than aspartate which allowed for closer interaction between the oppositely charged amino acids and therefore better shielding of side chain

charges while still maintaining the hydrogen bond between the guanidinium group of arginine and the hydroxyproline backbone carbonyl.[7]

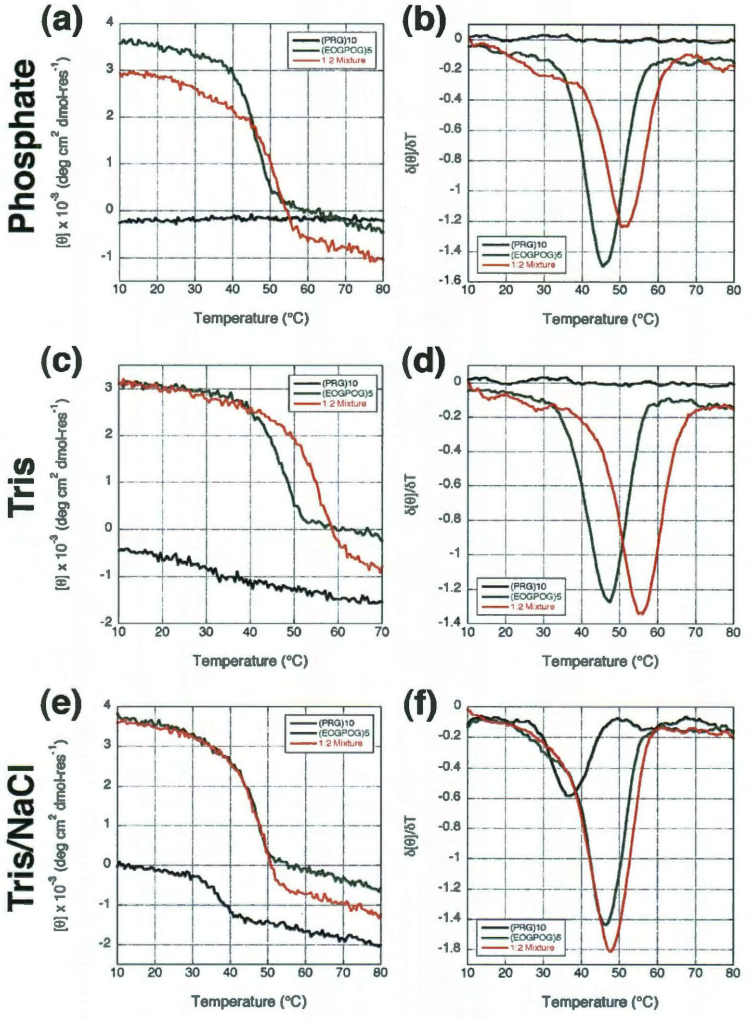

Figure 3.8. CD thermal unfolding curves shown as MRE vs. temperature (left column) and first derivative of MRE vs. temperature (right column) for the $(PRG)_{10}$•$2(EOGPOG)_5$ system in phosphate (a and b), Tris (c and d) and Tris/NaCl (e and f). Adapted from Figures 1 and S3 in reference 8. Data for component peptides are shown as the black and green curves and the 1:2 mixture of the peptides is shown as the red curve.[8]

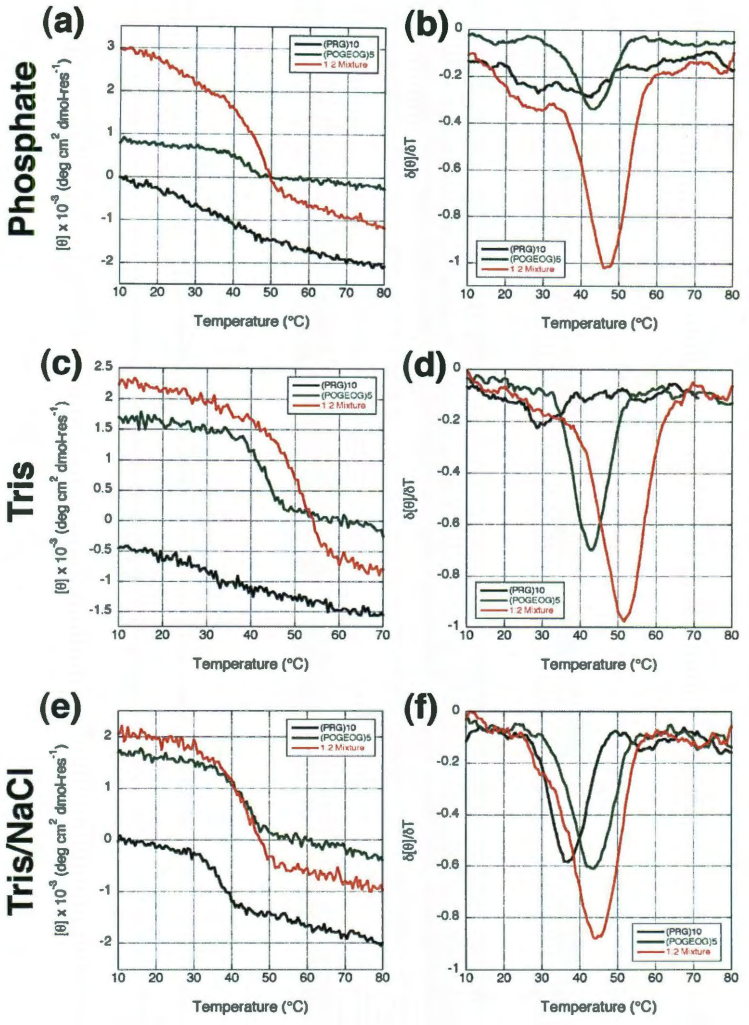

Figure 3.9. CD thermal unfolding curves shown as MRE vs. temperature (left column) and first derivative of MRE vs. temperature (right column) for the $(PRG)_{10}$•2$(POGEOG)_5$ system in phosphate (a and b), Tris (c and d) and Tris/NaCl (e and f). Adapted from Figure S11 in reference 7. Data for component peptides are shown as the black and green curves and the 1:2 mixture of the peptides is shown as the red curve.[7]

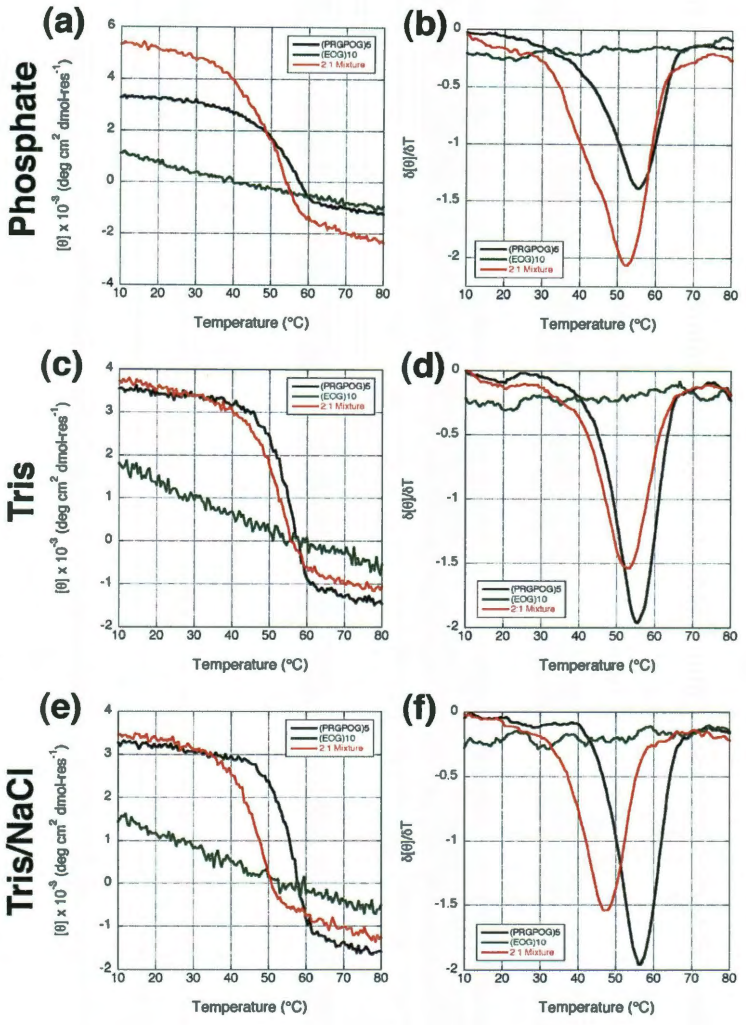

Figure 3.10. CD thermal unfolding curves shown as MRE vs. temperature (left column) and first derivative of MRE vs. temperature (right column) for the 2(PRGPOG)$_5$•(EOG)$_{10}$ system in phosphate (a and b), Tris (c and d) and Tris/NaCl (e and f). Adapted from Figure S12 in reference 7. Data for component peptides are shown as the black and green curves and the 2:1 mixture of the peptides is shown as the red curve.[7]

3.5.3. Lysine-Glutamate Pairing

Moving to peptide mixtures containing lysine residues, the pairing of lysine and glutamate will be discussed first. Similar to the results with arginine and glutamate, all three peptide systems formed a distinct heterotrimer in at least one buffer. In a similar

fashion to the arginine-containing systems, the K-E systems will be discussed individually beginning with $(PKG)_{10}$•$2(EOGPOG)_5$ and $(PKG)_{10}$•$2(POGEOG)_5$ (Figures 3.11 and 3.12 respectively) then turning to $2(PKGPOG)_5$•$(EOG)_{10}$ (Figure 3.13).

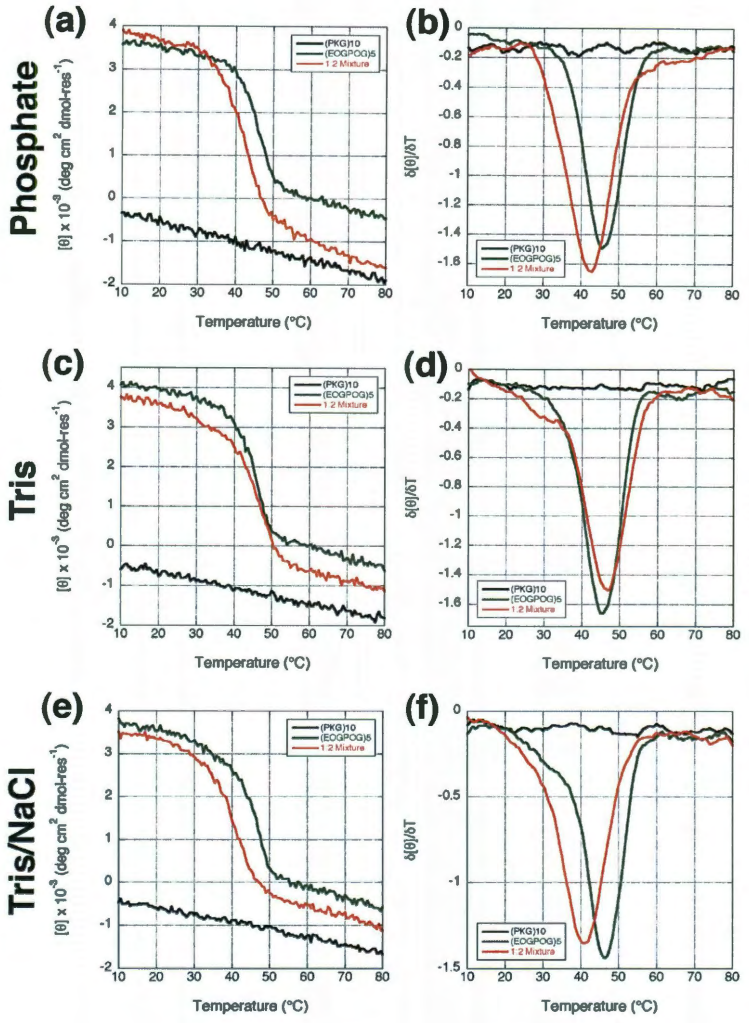

Figure 3.11. CD thermal unfolding curves shown as MRE vs. temperature (left column) and first derivative of MRE vs. temperature (right column) for the $(PKG)_{10}$•$2(EOGPOG)_5$ system in phosphate (a and b), Tris (c and d) and Tris/NaCl (e and f). Data for component peptides are shown as the black and green curves and the 2:1 mixture of the peptides is shown as the red curve.

Both $(PKG)_{10}$•2$(EOGPOG)_5$ and $(PKG)_{10}$•2$(POGEOG)_5$ resulted in heterotrimers with a lower stability than the homotrimer in phosphate buffer.[7] In Tris, heterotrimer formation was masked by the overlapping melting temperatures between the peak of the peptide mixture and the $(EOGPOG)_5$ and $(POGEOG)_5$ homotrimers. Upon the addition of NaCl, the heterotrimer peak became visible with a melting temperature of 40.5 °C for $(PKG)_{10}$•2$(EOGPOG)_5$ and 40.5 °C for $(PKG)_{10}$•2$(POGEOG)_5$, which was lower than the homotrimer, demonstrating the heterotrimer unmasking that can occurred by adjusting the ionic strength of the buffer.[7]

When we inverted the charged pair going from +10/−5 to +5/−10 with 2$(PKGPOG)_5$•$(EOG)_{10}$, a heterotrimer was formed in all samples (T_m of 45 °C in phosphate and Tris and T_m of 42 °C in Tris/NaCl) while none of the two individual peptides formed homotrimers in any of the three buffers. This was the first reported collagen mimetic system that formed a high-stability AAB triple helix when neither of the component peptides formed a triple helix, demonstrating the successful implementation of both positive and negative design parameters.[7] Due to the fact that neither peptide formed a homotrimer, any triple helix seen in 2:1 mixtures of the peptides must have resulted from stabilizing interactions between peptides upon heterotrimer formation. The melting profile and first derivative for this system in all three buffers are given in Figure 3.13. Since this system resulted in heterotrimers without homotrimer formation in all three buffers, we continued analysis of all 2$(PKGPOG)_5$•$(EOG)_{10}$ mixtures by performing DSC and NMR experiments which will be discussed in the corresponding sections below.

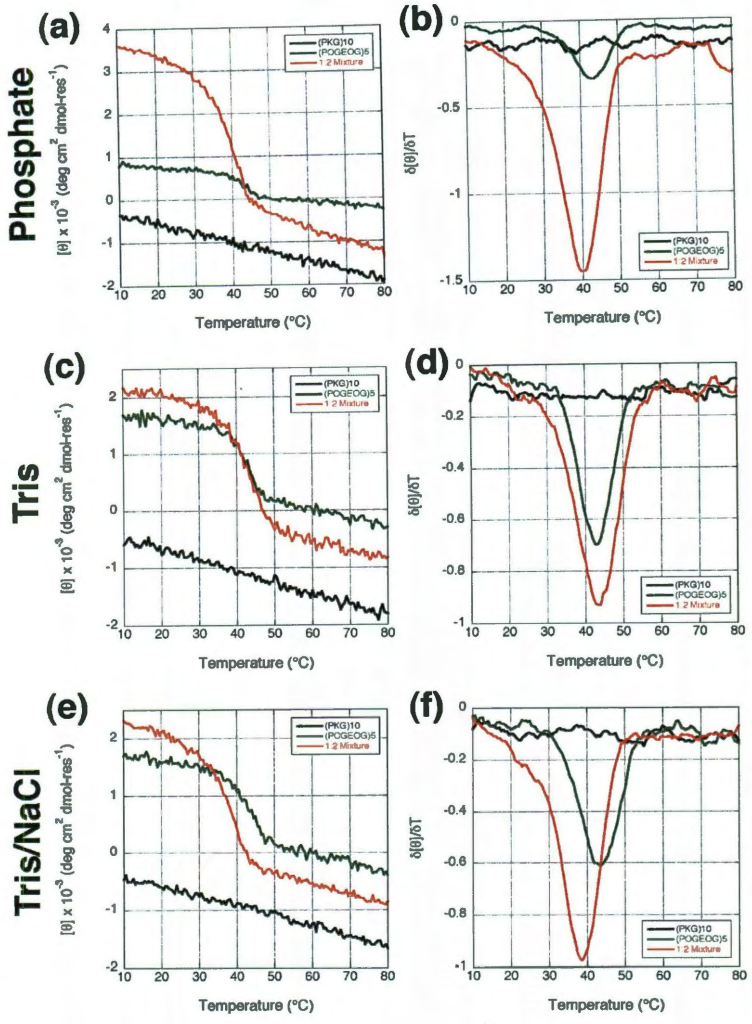

Figure 3.12. CD thermal unfolding curves shown as MRE vs. temperature (left column) and first derivative of MRE vs. temperature (right column) for the $(PKG)_{10}$•2$(POGEOG)_5$ system in phosphate (a and b), Tris (c and d) and Tris/NaCl (e and f). Adapted from Figures 2 and S15 in reference 7. Data for component peptides are shown as the black and green curves and the 1:2 mixture of the peptides is shown as the red curve.[7]

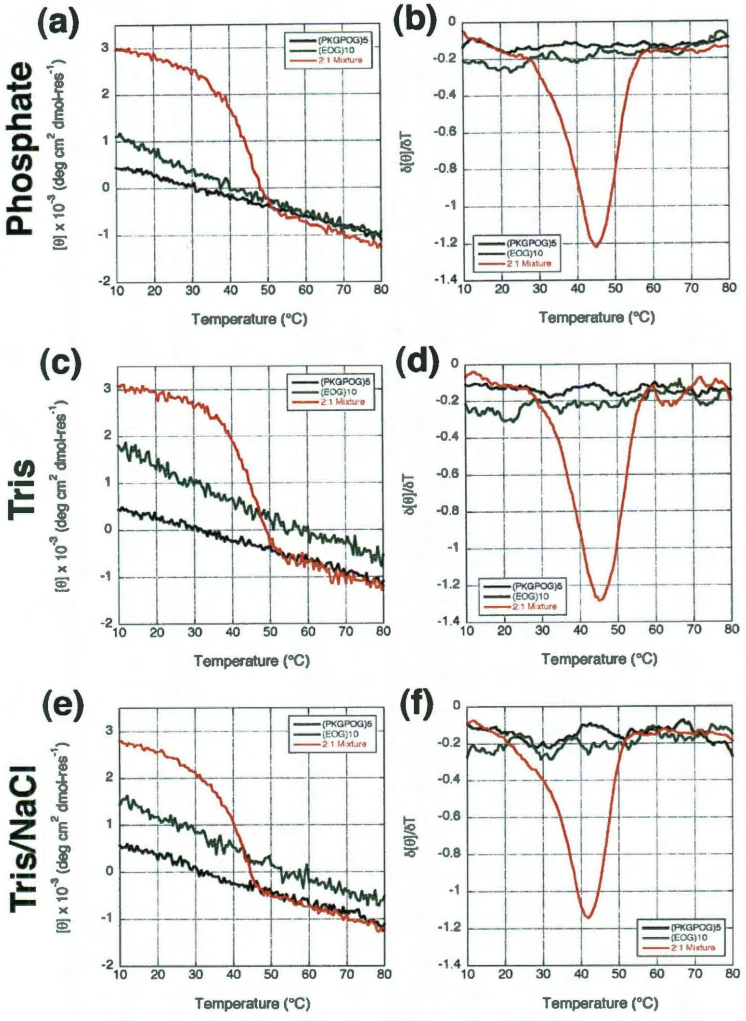

Figure 3.13. CD thermal unfolding curves shown as MRE vs. temperature (left column) and first derivative of MRE vs. temperature (right column) for the $2(PKGPOG)_5 \bullet (EOG)_{10}$ system in phosphate (a and b), Tris (c and d) and Tris/NaCl (e and f). Adapted from Figure 3 in reference 7. Data for component peptides are shown as the black and green curves and the 2:1 mixture of the peptides is shown as the red curve.[7]

3.5.4. Lysine-Aspartate Pairing

The last amino acid pair that we analyzed is the combination of lysine and aspartate, which had previously been shown to form direct electrostatic interactions within a self-assembled ABC heterotrimer.[1,3] The first major observation about the results

on $(PKG)_{10}$•$2(POGDOG)_5$ and $2(PKGPOG)_5$•$(DOG)_{10}$ was the presence of thermal transitions for AAB heterotrimers in all buffers regardless of the charge combination being +10/−5 or +5/−10.[7] This made the lysine-aspartate charged pairing different from, and superior to, the other three charge combinations tested. The CD melting profiles for $(PKG)_{10}$•$2(POGDOG)_5$ and $2(PKGPOG)_5$•$(DOG)_{10}$ in all three buffers are shown in Figures 3.14 and 3.15 respectively. In addition, all of the heterotrimers observed had higher thermal stabilities than any homotrimer in the systems.[7] Again, this was the only amino acid pairing with such results. Before discussing the possible driving forces for this heterotrimeric stability, a second observation must be made: in all three buffers, the melting temperatures between the systems are within two degrees of each other, even though the stabilities of the homotrimers vary significantly.[7] We hypothesized that there was direct electrostatic bridging between the lysine and aspartate residues, which was possible with lysine-aspartate pairing as opposed to the previously discussed arginine-aspartate combinations due to the structural freedom of the lysine side chain. This direct salt bridge was previously reported on an ABC system and was shown to occur between lysines of triplet *n* and aspartates of triplet *n+1*.[1,3] In these AAB mixtures, compared to the previously reported ABC system, only 5 bridges would be possible based on the peptide design, instead of the 10 bridges seen for the ABC heterotrimer. In order to confirm this theory, further analysis using DSC and 2D solution NMR was required. For such analysis, we selected the $2(PKGPOG)_5$•$(DOG)_{10}$ system in phosphate due to the fact that it formed an AAB heterotrimer without either component peptide forming a homotrimer, allowing for cleaner analysis and again demonstrating the successful implementation of positive and negative design strategies to achieve a desired result.[7]

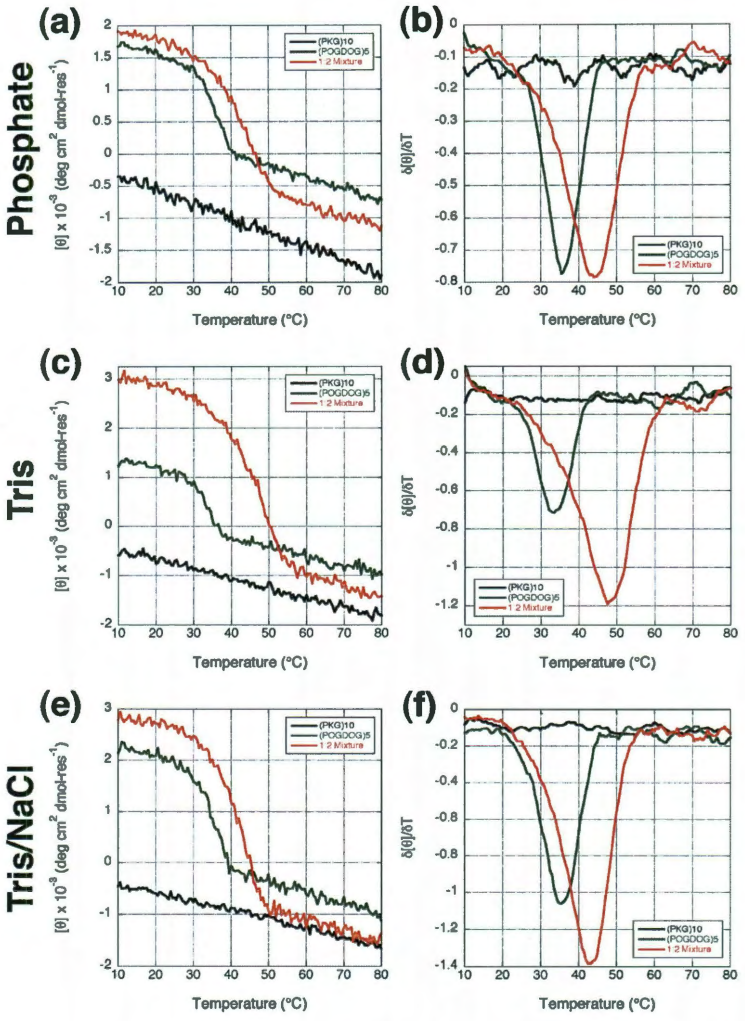

Figure 3.14. CD thermal unfolding curves shown as MRE vs. temperature (left column) and first derivative of MRE vs. temperature (right column) for the $(PKG)_{10}$•2$(POGDOG)_5$ system in phosphate (a and b), Tris (c and d) and Tris/NaCl (e and f). Adapted from Figure S13 in reference 7. Data for component peptides are shown as the black and green curves and the 1:2 mixture of the peptides is shown as the red curve.[7]

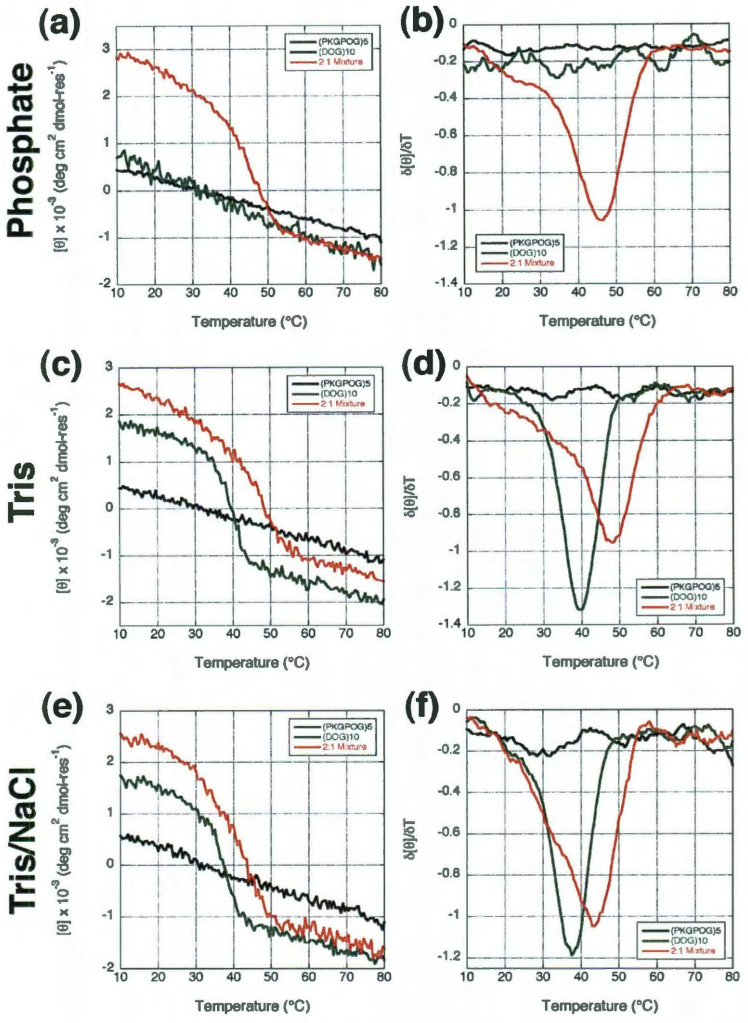

Figure 3.15. CD thermal unfolding curves shown as MRE vs. temperature (left column) and first derivative of MRE vs. temperature (right column) for the $2(PKGPOG)_5 \bullet (DOG)_{10}$ system in phosphate (a and b), Tris (c and d) and Tris/NaCl (e and f). Adapted from Figures 3 and S14 in reference 7. Data for component peptides are shown as the black and green curves and the 2:1 mixture of the peptides is shown as the red curve.[7]

3.6. DSC Analysis on Selected Systems

After CD melting experiments were performed on all amino acid charged pair combinations, three peptide systems were highlighted for differential scanning

calorimetry (DSC) analysis: $(PRG)_{10}$•$2(EOGPOG)_5$ in Tris, $2(PKGPOG)_5$•$(EOG)_{10}$ in phosphate, Tris and Tris/NaCl and $2(PKGPOG)_5$•$(DOG)_{10}$ in phosphate.[7,8] The DSC melting experiments for each system gave an alternative and more sensitive measure of the melting temperature, which confirmed results from CD.[7] Additionally, DSC gave information about the thermal recovery, or lack thereof, for each heterotrimer. The profiles for each of these systems are shown in Figures 3.16, 3.17 and 3.18 respectively. In short, the major peak seen in the first peptide scan for all samples matched the T_m seen in CD melting studies given in Table 3.2. As done for the CD thermal stability analysis, the systems will be discussed individually.

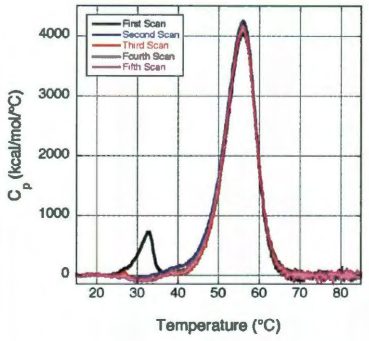

Figure 3.16. DSC melting profile for the $(PRG)_{10}$•$2(EOGPOG)_5$ system in Tris. Adapted from Figure S4 in reference 8. The first five peptide scans are shown in black, blue, red, green and pink sequentially.[8]

Beginning with $(PRG)_{10}$•$2(EOGPOG)_5$ in Tris, shown in Figure 3.16, the first peptide scan showed two peaks: a minor peak at 32 °C, that disappeared in all subsequent scans, and a major peak at 56 °C that agreed with the CD melting data.[8] Based on the DSC scans on $(PRG)_{10}$ described in the homotrimers section above, the minor peak could be residual $(PRG)_{10}$ homotrimer present within the system. However, since $(PRG)_{10}$ did not show any homotrimer formation in Tris buffer, only in phosphate and Tris/NaCl, the

minor peak at 32 °C could also be assigned to a different peptide register which was not detectable by CD and which cannot refold during the DSC cycle time.[8] The reproducibility of multiple DSC scans were consistent with previous reports that heterotrimeric species stabilized by electrostatic interactions have a shorter refolding half-life than neutral homotrimers and thus, can be analyzed repeatedly after only a short refolding interval.[2,8]

If we look at $2(PKGPOG)_5 \bullet (EOG)_{10}$, strong contrasts can be seen in the DSC melting profiles in each of the three buffers (Figure 3.17). The peptide system in phosphate showed a clean single peak in the first peptide scan however, in the second and third scans, a shoulder at 30 °C was visible.[7] Since neither peptide formed a homotrimer visible by CD or DSC in phosphate, based on these DSC results alone we hypothesized that a kinetically trapped AAB heterotrimer may be forming in all subsequent melting scans.[7] When the profile in phosphate was compared to that seen in Tris, a substantial difference could be seen. The profile in Tris showed a single transition in the first peptide scan that was repeated in all subsequent scans indicating that any species present within the system refolded within the timescale of the experiment.[7] In the last buffer tested, Tris/NaCl, a total breakdown of the AAB heterotrimer occurred after the first thermal scan.[7] The second scan had a large shoulder at 30 °C and the major peak had a much lower intensity that continues to decrease in the third scan. Although this result was disappointing for the thermal recovery of the peptide system, it was not unexpected due to the high salt concentration of the buffer. In such an environment, the high salt was expected to largely prevent the charged residues from forming significant interactions and thus slowing the refolding time of the peptide mixture causing it to greatly exceed

that of the DSC experiment, which resulted in the decrease of the AAB heterotrimer population with each subsequent scan.[7] Re-folding CD experiments of this peptide system in all three buffers were performed and complemented the DSC results (Figure 3.17d): samples in phosphate and Tris refolded > 80% of the population within an hour whereas the Tris/NaCl sample required 4 hours to refold to 80% of the peptide mixture.

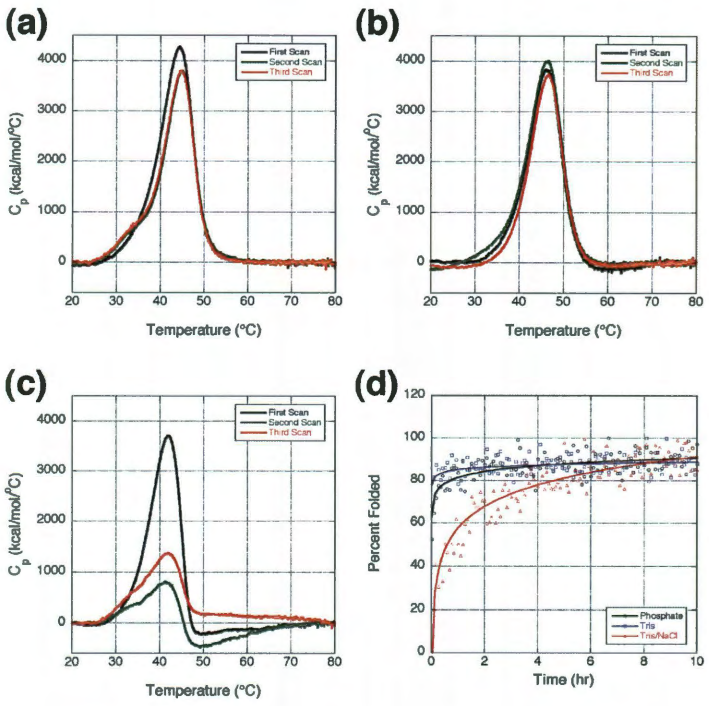

Figure 3.17. DSC melting profiles and for $2(PKGPOG)_5•(EOG)_{10}$ in (a) phosphate, (b) Tris and (c) Tris/NaCl. Adapted from Figure 3 in reference 7. The first three peptide scans are shown in black, green and red sequentially. (d) CD re-folding times for the peptide systems shown as percent folded vs. time in phosphate (black), Tris (blue) and Tris/NaCl (red). Adapted from Figure S17 in reference 7. Logarithmic fits were created for each buffer to highlight the general trend of folding for each.[7]

When we examined the last system, $2(PKGPOG)_5•(DOG)_{10}$ in phosphate (Figure 3.18), DSC experiments were similar to that for the $2(PKGPOG)_5•(EOG)_{10}$ in phosphate.[7] The first peptide scan is a clean single peak that corresponded to the melting temperature

seen in CD melting studies. However, in all subsequent scans, a shoulder could be seen and the intensity of the major peak decreased significantly. Analogous to the $2(PKGPOG)_5•(EOG)_{10}$ system, the second and third scans overlapped each other suggesting that the refolding time of the mixture exceeded that of the DSC experiment such that a large portion of the system was able to refold within the time scale of the experiment.[7]

Therefore from DSC experiments, only the $2(PKGPOG)_5•(EOG)_{10}$ system in Tris buffer was capable of complete thermal recovery within the timescale of the experiments.[7]

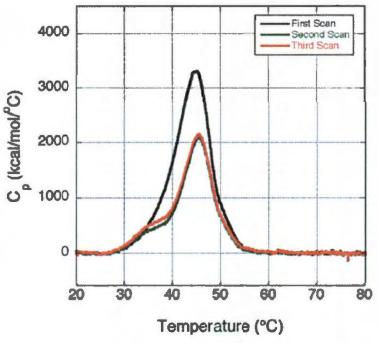

Figure 3.18. DSC melting profile for $2(PKGPOG)_5•(DOG)_{10}$ in phosphate. Adapted from Figure 3 in reference 7. The first three peptide scans are shown as black, green and red sequentially.[7]

3.7. NMR Analysis on Selected Systems

After CD and DSC analysis, $(PRG)_{10}•2(EOGPOG)_5$ in Tris, $2(PKGPOG)_5•(EOG)_{10}$ in phosphate, Tris and Tris/NaCl and $2(PKGPOG)_5•(DOG)_{10}$ in phosphate all formed visible AAB heterotrimers. The thermal recovery of many of the mixtures was beyond the timescale of the DSC experiments, but a single transition at the

T_m for the heterotrimer was seen in each system in at least one of the DSC peptide scans. In order to understand the identity of the minor species seen in DSC, solution NMR was required. Additionally, despite the fact that $2(PKGPOG)_5•(EOG)_{10}$ in Tris showed single transitions in all CD and DSC data, multiple species could still be present within the solution. Since CD and DSC measure the ellipticity and specific heat respectively as a function of temperature, species present that have overlapping thermal stabilities will appear as a single species in both types of experiments. Therefore, NMR is required to verify whether a peptide mixture forms a single heterotrimeric species or not.

Solution NMR experiments were run to 1) confirm the triple helical topology of the molecular assembly giving rise to the cooperative thermal transition, 2) to confirm the AAB composition of the helix, 3) to determine how the side chains on the charged amino acids are interacting and 4) to assess the register, or relative stagger, between peptide chains within the triple helix.[7,8]

3.7.1. NMR Analysis of $(PRG)_{10}•2(EOGPOG)_5$ in Phosphate

Nuclear Overhauser effect spectroscopy (NOESY) and total correlated spectroscopy (TOCSY) experiments were recorded at 25 °C to sequence the spin systems. The experiments showed that a 1:2 mixture of the $(PRG)_{10}$ and $(EOGPOG)_5$ peptides contained several species including monomeric forms of both peptides and a small quantity of $(EOGPOG)_5$ homotrimer, while the major component corresponded to an AAB heterotrimer.[8] The AAB assembly showed typical NOEs expected from a triple helix, such as the glycine packing interactions at the core of the helix. Using these peaks in conjunction with arginine-glutamate backbone cross-peaks, it was possible to

determine the register of the dominant species to be $(PRG)_{10}$•$(EOGPOG)_5$•$(EOGPOG)_5$.[8] A secondary heterotrimeric spin system with weaker peaks was identified but could not be unambiguously sequenced due to spectral overlap. This spin system may have arisen from less ordered regions at the termini of the triple helices or from a different peptide register. In either case, it was a minor component.[8]

Using the information from the intra-residue and inter-chain NOEs of the arginine and glutamate side-chains, a set of conformational restraints was built and a model of the AAB triple helix minimized using experimental constraints was made starting from the crystal structure of the triple helical peptide $(PPG)_{10}$, pdb identifier: 1k6f.[8,11] The minimization was done following the procedure reported by Fallas, *et al.*[1] in the amber09 package[12] using conformational restraints for the arginine and glutamate side chains derived from the NOESY spectrum.[8]

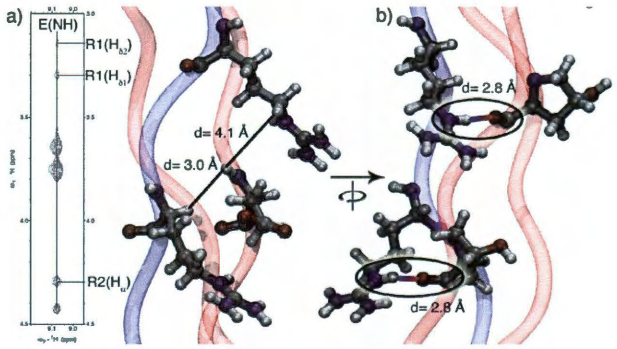

Figure 3.19. Solution NMR experiments on $(PRG)_{10}$•$2(EOGPOG)_5$ in phosphate. Adapted from Figure 2 in reference 8. (a) NOESY spectrum and molecular model highlighting the cross-peaks between the arginine and glutamate side-chains and the atoms giving rise to the NOEs, sequential and intra-residue peaks are not labeled for clarity. (b) Alternate view of the model highlighting the hydrogen bonds between the guanidinium groups and hydroxyproline backbone carbonyls using colored arrows.[8]

The NOE results (Figure 3.19a) showed two main arginine side chain conformers with fixed dihedrals, leading to a unique chemical shift for all diastereotopic methylene

protons along the side-chain.[8] While the arginine generated NOE contacts with the glutamate amide proton, the predominant configuration did not appear to include direct hydrogen bonded contacts between the charged moieties. Instead, the guanidinium groups were positioned to form optimal hydrogen bonds with the hydroxyproline carbonyls of $(EOGPOG)_5$, which was positioned in the second register (Figure 3.19b).[8] The glutamate side chains were positioned between alternate arginines to screen the positive charges. Furthermore, the $\chi 2$ dihedral of these glutamates adopted a dynamic conformation, as evidenced by the chemical shift equivalence of both γ-protons. This observation suggested the formation of a weak salt bridge between the glutamate and the second arginine conformer, although this cannot be confirmed due to lack of information on the $\chi 4$ dihedral of the arginine side-chain.[8] This conformation differed from the aspartate-lysine salt bridges previously determined via NMR in an ABC heterotrimer,[1] but was not completely unexpected as there are crystal structures of triple helical peptides that show a similar interaction.[13,14] Interestingly, the $(EOGPOG)_5$ in the third peptide register was chemically distinct with its glutamates oriented away from arginine side chains.[8] A schematic representation of the modeling and NOE assignments is shown in Figure 3.20.

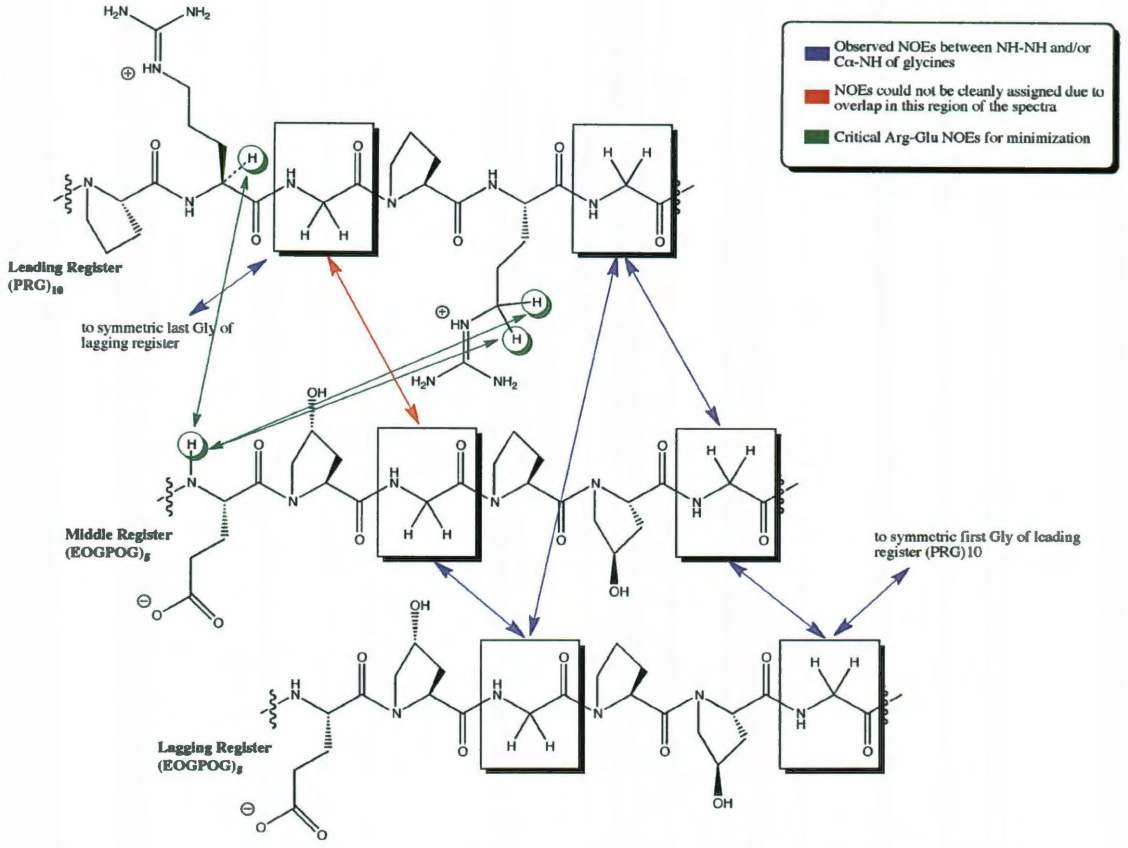

Figure 3.20. Graphical depiction of key NOE assignments used to assign the helix register and to restrain the helix model for the $(PRG)_{10}$•2$(EOGPOG)_5$ system. Adapted from Figure S6 in reference 8.[8]

To determine the register of the AAB heterotrimer, the glycine packing cross-peaks at the core of the helix were considered.[8] In order to provide a large surface area for van der Waals contacts and to maintain the hydrogen-bonding network that stabilized the assembly, there was a glycine residue at every cross section taken perpendicular to the helical axis. This tight packing allowed for the alpha and amide protons of different glycine residues in different chains to be close enough to generate inter-chain NOEs.[8] Different registers would give rise to different NOE patterns because the topology of the helix constrained the cross-peaks to be only between amino acids that were sequential in

space. Table 3.3 shows the glycine packing cross-peaks expected from each of the three possible registers of an AAB triple helix given the $(PRG)_{10}$•2$(EOGPOG)_5$ system, in addition to backbone cross-peaks between arginine and glutamate.[8] The subscripts define the triplet from which the glycine stems, for the $(EOGPOG)_5$ chains (labeled "E"), the numbers are chain identifiers and for the $(PRG)_{10}$ chain (labeled "R"), the numbers refer to the first and second triplets in the PRG sextet, as the amino acids had different chemical shifts in subsequent triplets.[8] In the table, a "•" is placed for peaks expected from the REE register, a "n" is placed for peaks expected from the ERE register and a "✚" is placed for the cross-peaks expected for the EER register.[8] Experimental NOEs are highlighted by black boxes. Based on the NOE cross-peaks observed, the dominant register was determined to be $(PRG)_{10}$•$(EOGPOG)_5$•$(EOGPOG)_5$.[8]

	G_{EOG1}	G_{POG1}	G_{EOG2}	G_{POG2}	G_{PRG1}	G_{PRG2}	E_{EOG1}	R_{PRG2}
G_{EOG1}			• ✚	n	• n			
G_{POG1}			n	• ✚	✚	• n		
G_{EOG2}					n ✚	•		
G_{POG2}					•	n ✚		
G_{PRG1}								
G_{PRG2}								
E_{EOG1}								• n
R_{PRG2}								

Table 3.3. Glycine packing cross-peaks expected from each of the three possible registers of the $(PRG)_{10}$•2$(EOGPOG)_5$ AAB heterotrimer. Adapted from Table S1 in reference 8. The subscripts specify the triplet. For the E chains, the numbers are chain identifiers. For the R chain, the numbers refer to the first and second triplets in the PRG sextet. A "•" is placed for peaks expected from the REE register, a "n" is for peaks expected from the ERE register and a "✚" is for peaks expected from the EER register. Observed NOEs are highlighted by black boxes. G_{PRG1}-G_{EOG1} (• n): While these NOEs were not able to be assigned, peaks existed in this region, which may correspond to the expected NOE for the REE register. However, they could not be definitively assigned due to overlap[8]

3.7.2. NMR Analysis of $2(PKGPOG)_5•(EOG)_{10}$ and $2(PKGPOG)_5•(DOG)_{10}$

In order to study the composition and supramolecular topology of the $(PKGPOG)_5$ containing triple helical assemblies that gave rise to the cooperative transitions seen in the CD and DSC spectra, a set of peptide derivatives of the $2(PKGPOG)_5•(DOG)_{10}$ and $2(PKGPOG)_5•(EOG)_{10}$ systems were synthesized.[7] The derivatives included a tryptophan at the N-terminus of all peptides for accurate concentration determination and a glycine linker between the spectroscopic tag and the triple helical sequence. In all previous samples, peptide concentration was determined by mass. However, this method can be very inaccurate due to the presence of residual salts within samples from peptide synthesis and purification. Even when peptides have been dialyzed against deionized water, a small amount of salt will remain in the sample, especially for highly charged peptides. Therefore, in order to have accurate peptide concentration determination, a spectroscopic tag such as tryptophan must be added to the peptide sequence. Furthermore, the sixth triplet of the $WG(PKGPOG)_5$ peptide contained an ^{15}N-enriched glycine (amino acid 20). $2WG(PKGPOG)_5•(EOG)_{10}$ and $2WG(PKGPOG)_5•(DOG)_{10}$ were both characterized using homonuclear $^1H,^1H$-NOESY and $^1H,^1H$-TOCSY experiments as well as $^1H,^{15}N$-HMQC and $^1H,^1H$-planes of 3D HNHA (2D HNHA) and NOESY-^{15}N-HSQC (edited NOESY). All experiments for the $2WG(PKGPOG)_5•WG(DOG)_{10}$ system were carried out in phosphate buffer while the $2WG(PKGPOG)_5•WG(EOG)_{10}$ was studied in deuterated Tris.[7]

The $^1H,^{15}N$-HMQC spectrum for each system showed two pairs of cross-peaks of equal intensity (Figure 3.21).[7] Because the ^{15}N-labelled amino acid was present in the peptide chain with the lower overall charge, and therefore twice in each triple helix, two

distinct cross-peaks were expected for each register of the desired AAB triple helix since the chemical environment was, in principle, not identical for both chains with identical sequence in the heterotrimer. On the other hand, a homotrimer or an AAB heterotrimer of the opposite stoichiometry would give rise to a single cross-peak. The spectra obtained showed that both the $2WG(PKGPOG)_5 \bullet WG(DOG)_{10}$ and $2WG(PKGPOG)_5 \bullet WG(EOG)_{10}$ systems populated two distinct sets of chemical environments.[7] Furthermore, each of those environments included two labeled glycines with equivalent populations but distinct chemical shifts as we expected for an AAB heterotrimer comprised of two positive chains and one negative chain. We believe that those two environments were best accounted for by two competing registers of the desired heterotrimer and that no triple helical assembly of a different composition was significantly populated.[7] With this data, we concluded that this was the first time that a self-assembled heterotrimeric triple helical system showed control over composition as was suggested by the CD melting studies.[7] This means that within this peptide mixture, no competing homotrimers or alternative composition of heterotrimers were formed. Despite having control over the composition of the helix, the self-assembled AAB triple helices lacked complete control over their register.[7] The relative population of each register could be obtained by comparing the cross-peak intensity in the spectra. The integration lead to a 2.6:1 ratio between the major and minor registers for $2WG(PKGPOG)_5 \bullet WG(EOG)_{10}$ (72% of the triple helical population corresponded to the major register and 28% to the minor register). A similar result was obtained for the $2WG(PKGPOG)_5 \bullet WG(DOG)_{10}$ system where the ratio stands at 2.2:1 (69% of the triple helical population corresponded to the major register and 31% to the minor register.[7]

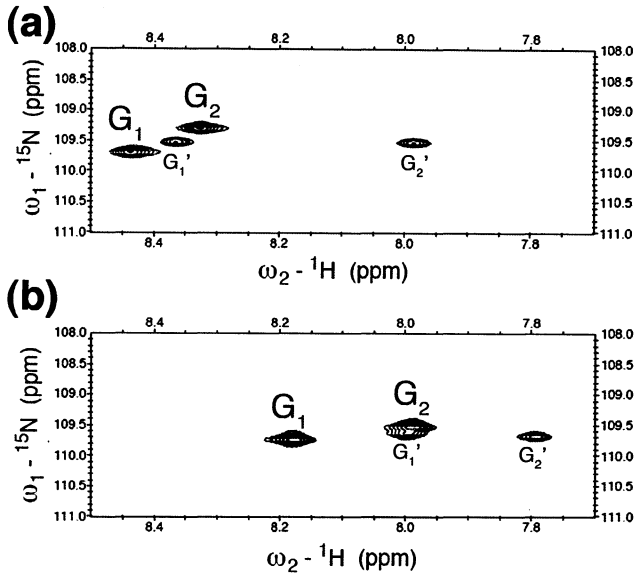

Figure 3.21. $^{1}H,^{15}N$-HMQC spectra of (a) 2WG(PKGPOG)$_5$•WG(EOG)$_{10}$ and (b) 2WG(PKGPOG)$_5$•WG(DOG)$_{10}$. Adapted from Figure 4 in reference 7. The cross-peaks corresponding to the minor register are denoted by `.[7]

The NOESY and TOCSY spectra of each system showed the structure expected from a triple helical assembly.[7] Because of the symmetry of the helix and the periodicity of the sequence, only one set of cross-peaks was observed for each amino acid in the structural repeating unit of each chain. In the case of the negative chain, the structure of the positive chain's sequence caused two consecutive triplets to be chemically distinct even though they were equivalent in sequence, making the repeating unit DOGDOG and EOGEOG instead of DOG and EOG respectively. This differed from what had been previously observed for ABC heterotrimers,[1] where the repeating unit corresponds to a triplet, but agrees with the results for the (PRG)$_{10}$•2(EOGPOG)$_5$ reported above, where the repeating unit corresponds to a sextet.[8] The backbone chemical shifts of the PKG triplet and EOGEOG and DOGDOG sextets were identified using the homonuclear sequential assignment procedure.[1,7,8] For the POG triplets within the WG(PKGPOG)$_5$

peptide, a combination of the 2D HNHA and edited NOESY experiments was successful in identifying the glycine and hydroxyproline backbone chemical shifts, but the sequential following of two imino acids made the identification of the proline H_α frequency for the POG triplets very difficult.[7] Furthermore, no sequential links were available between the PKG and POG triplets that made up the repeating unit in the positive chains. NOEs between the glycines of these triplets were used to determine which chain each glycine belonged to. Because of constraints in the triple helical structure, no NOEs between the two consecutive glycines of the same chain were possible. Thus, NOEs between the glycines of PKG and POG triplets must be inter-chain and could be used to determine the chemical shift of the repeating units for the two chemically distinct positive chains. Due to spectral overlap in the aliphatic region, only a partial assignment of the amino acid's side chain resonances was possible and the imino acid side chain assignment was not attempted.[7]

Both the $^1H,^1H$-NOESY and $^1H,^1H$-edited NOESY spectra showed the cross-peaks expected from triple helical peptides.[7] Such peaks include inter-chain proline delta to glycine amide correlations[15] as well as glycine amide-alpha and amide-amide resonances[1] due to the tight packing of glycines in the core of the helix. Other interesting features included resonances between lysine ε-protons and the acidic residue's amide proton, suggesting an interaction between the oppositely charged amino acids.[7] This information, in principle, should have sufficed for the determination of the register of the triple helices, but the task was complicated by the chemical shift overlap observed. In the next two sections, the distinctive features observed for each of the systems will be discussed.

3.7.2.1. NMR Analysis of 2WG(PKGPOG)$_5$•WG(EOG)$_{10}$ in Tris

This system presented a particular challenge because the chemical shifts of the glutamate amide protons overlapped with some of the glycines.[7] Furthermore, some of the amide protons of the minor register overlapped with the major register. For instance, the glycine amide proton chemical shift of the second EOG triplet in the major register overlapped with the glycine amide proton chemical shift of one of the POG triplets of the minor register.[7] Thanks to the ^{15}N-label, this assignment could be made unambiguously. However, in the regular NOESY spectrum, the cross-peaks corresponding to the minor register are obscured by those of the major register.[7] In addition, the chemical shift of most of the glycine α-protons was degenerate. This led to considerable spectral crowding in the region corresponding to the glycine amide – alpha proton resonances and made the register determination impossible from this area. Instead, we focused on studying the relatively weak amide-amide resonances arising from glycine packing.[7] In the regular NOESY spectrum, the area corresponding to the amide-amide cross-peaks was dominated by sequential and diagonal peaks. On the other hand, the edited NOESY spectrum provided a clear view of the region and even though only information on the POG triplets could be gained, this provided enough information to determine the register.[7] Figure 3.22a showed the edited NOESY spectrum for this system, where the chemical shifts of the labeled amino acids were marked by vertical lines and labeled as G_{O1} and G_{O2}. Each of those glycines presented cross-peaks to two other glycines, whose chemical shifts were marked by horizontal lines. From this information, we deduced the register of the heterotrimeric triple helix.[7] Starting at the G_{O1} chemical shift, two NOEs could be observed: one going to the glycine in the second triplet of EOGEOG repeating

unit, labeled as G_{E2}, and one going to the PKG triplet in the second positive chain, labeled as G_{K2}.[7] Now, considering the NOEs observed for G_{O2}, a cross-peak to G_{E2} could also be observed indicating that the second triplet of the negative chain was flanked by the two labeled glycines in the POG triplets of the positive chains, positioning it as the middle chain in the peptide register and making the register $WG(PKGPOG)_5$•$WG(EOG)_{10}$•$WG(PKGPOG)_5$.[7] Figure 3.22b and 3.22c showed a sequence repeat to clarify the naming conventions used in the discussion and a heterotrimeric triple helix model. Furthermore, the observed NOEs between the glycine amide protons are highlighted in Figure 3.22c using colored arrows that match them to colored circles in Figure 3.22a.[7] Most peaks from the minor register were below the level of noise but a particular resonance between both POG triplets indicated that their glycines are in close proximity. A possible register that would agree with such an arrangement is $WG(PKGPOG)_5$•$WG(PKGPOG)_5$•$WG(EOG)_{10}$,[7] which would be predicted based on a recent theoretical paper showing that the charged pair interactions between different chains are not equivalent and the K-E charged pair is most stable between the middle and lagging chains.[16]

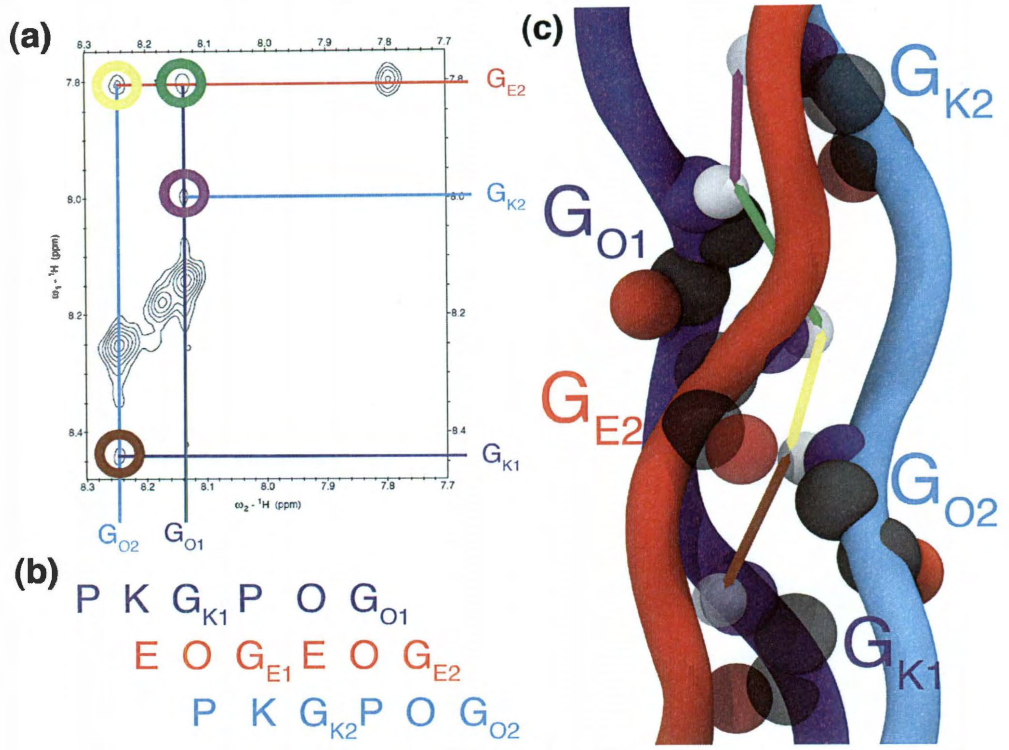

Figure 3.22. (a) ^{1}H,^{1}H-edited NOESY spectrum showing NH-NH resonances between chains, (b) molecular model highlighting the glycine packing interactions at the core of the helix and (c) sextet repeat of $WG(PKGPOG)_5 \bullet WG(EOG)_{10} \bullet WG(PKGPOG)_5$. Adapted from Figure 5 in reference 7. In (a), the chemical shift of the ^{15}N-labelled amino acids is highlighted by vertical lines and the chemical shifts of other amino acids by horizontal lines. Resonances relevant to the register determination are highlighted by colored circles in (a) and colored arrows in (c).[7]

The NOESY spectrum of this system also allowed the study of the interaction between oppositely charged amino acids, which we rationalized as the driving force behind the self-assembly process.[7] Of particular interest were the cross-peaks between the lysine ε-protons and both the lysine and glutamate amide protons (Figure 3.23a). Fallas *et al.* identified a cross-peak between the last methylene group of the basic side chain and the amide proton of the acidic residue located in the adjacent strand, two amino acids down in sequence.[1] This resonance arose because of the extended conformation of the positively charged residue, which was adopted in order to interact efficiently with the

negatively charged residue. Another peak, that was barely observable in a previously published ABC heterotrimer but was strong in this collagen assembly, arose between the same methylene group and its own amide proton.[7] Such a cross-peak was barely noticeable in the ABC system with a ratio between the inter-strand and intra-residue NOEs of 5. In the case of the $WG(PKGPOG)_5$•$WG(EOG)_{10}$•$WG(PKGPOG)_5$ register, the ratio between these two peaks decreased to 0.85, indicating that the distance between the lysine ε-methylene and its own amide proton is about the same as the distance to the acidic residue on the opposite strand.[7] Figure 3.23b shows a model of the charged residue's side chain conformation for this system satisfying constraints derived from the NMR data. In this conformation, the amino group in the lysine residue was not able to effectively interact with either of the carboxylates of the negative chain.[7] We attributed the observed resonances to a dynamic equilibrium of two possible charged paired states between lysine and the two successive glutamates (Figure 3.23c). Such a frustrated interaction could be used to rationalize the low thermal stability of these systems (compared to previously published electrostatically driven heterotrimers containing stable hydrogen bonding interactions)[1-3] and the fact that there was a lack of control over the register of the peptides.[7]

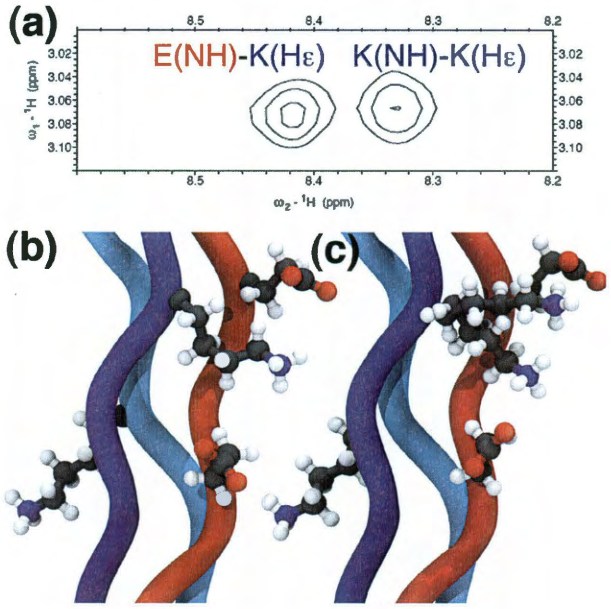

Figure 3.23. (a) $^{1}H,^{1}H$-NOESY spectrum and (b),(c) molecular models highlighting the interaction between charged residues in $WG(PKGPOG)_5 \bullet WG(EOG)_{10} \bullet WG(PKGPOG)_5$. Adapted from Figure 5 in reference 7. The model in (b) satisfies conformational constraints from (a) but prevents the formation of salt bridges and is depicted as the average between two possible hydrogen bonded conformations, shown in (c).[7]

3.7.2.2. NMR Analysis of $2WG(PKGPOG)_5 \bullet WG(DOG)_{10}$ in Phosphate

As in the previous system, there was still serious overlap in the amide-glycine alpha region of the NOESY spectrum of this system.[7] For this reason, and as mentioned in the previous discussion, the register for this system was determined using the amide-amide cross-peaks presented in the edited NOESY spectrum. The edited NOESY spectrum for this system (Figure 3.24a), showed a similar cross-peak pattern to the one observed for $WG(PKGPOG)_5 \bullet WG(EOG)_{10} \bullet WG(PKGPOG)_5$ (Figure 3.22a).[7] Once again, the chemical shifts of the labeled amino acids are marked by vertical lines, while the glycines that they interacted with are marked by horizontal. The register of the triple helix could be determined in a similar manner following the same naming convention. If

we start considering G_{O1}, two NOEs could be observed: one going to the glycine in the second triplet of the DOGDOG repeat, labeled as G_{D2}, and one going to the PKG triplet in the second positive chain, G_{K2}.[7] G_{O2} also showed a cross-peak to G_{D2}, proving that this system also chose an ABA arrangement with the register being $WG(PKGPOG)_5$•$WG(DOG)_{10}$•$WG(PKGPOG)_5$, since other registers would require a cross-peak between the two labeled amino acids.[7] Figure 3.24b and 3.24c show a sequence repeat to clarify the naming conventions used in the discussion and a model of the heterotrimeric triple helix. Furthermore, the observed NOEs between the glycine amide protons are highlighted in Figure 3.24c using colored arrows that match them to colored circles in Figure 3.24a.[7] As in the previous case, not enough resonances were observed to fully determine the minor register, but the presence of the cross-peak between both POG glycines suggested the $WG(PKGPOG)_5$•$WG(PKGPOG)_5$•$WG(DOG)_{10}$ register, similar to the previous system.[7]

The interaction between the charged residues could also be studied using the NOESY spectrum of this system. As in the previously discussed heterotrimer, cross-peaks between the lysine ε-protons and the lysine and aspartate amide protons were present (Figure 3.25a).[7] The ratio between the inter-strand and intra-residue NOEs was 1.5 for this system, as compared to 5 in the ABC heterotrimer directed by lysine-aspartate charged pairs,[1] indicating a different conformation of the side chains for the AAB system. In Figure 3.25b, a model of the charged residue's side chain conformation for this system satisfying constraints derived from the NMR data is depicted.[7] This conformation prevented an effective interaction between the oppositely charged moieties and thus we attributed the observed resonances to a dynamic equilibrium of two possible charged

paired states between lysine and the two successive aspartates (Figure 3.25c). Therefore, the charged pair interactions present in this system were more similar to the $2WG(PKGPOG)_5 \bullet WG(EOG)_{10}$ AAB system discussed in the previous section, in which the negatively charged amino acid corresponds to glutamate, than to the previously studied ABC system driven by K-D ionic hydrogen bonds.[1,7] It should be noted that the previously published ABC heterotrimer containing lysine-aspartate charged pairs did not exhibit compositional control, as a homotrimer was formed by one of the peptides, but it did produce a single register heterotrimer.[1]

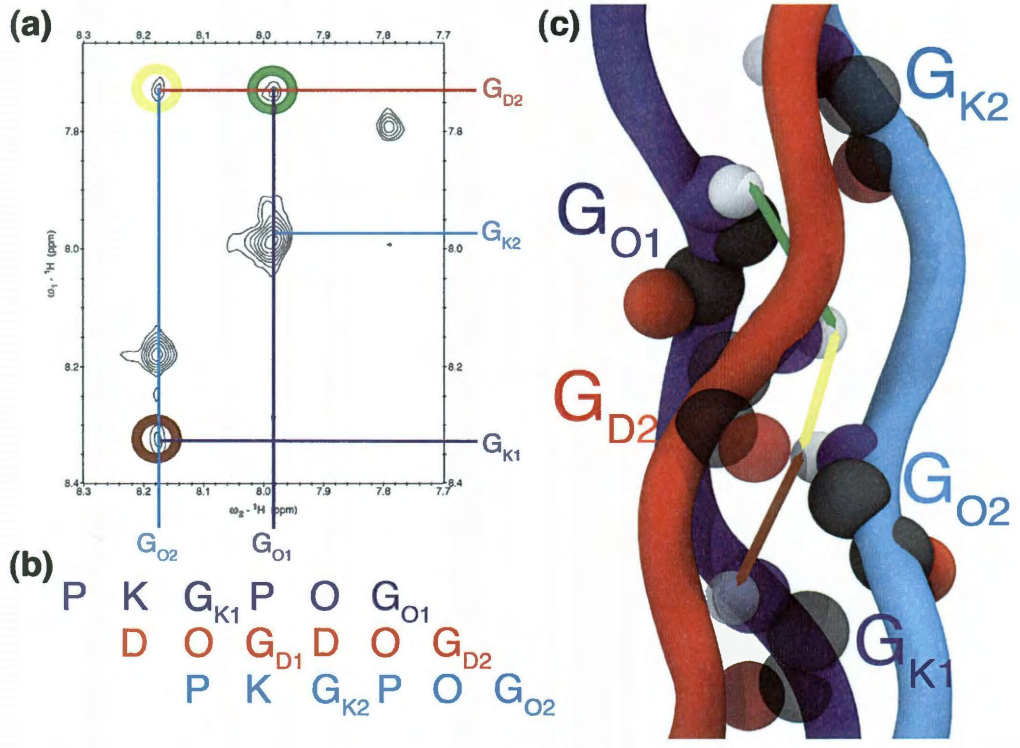

Figure 3.24. (a) $^{1}H,^{1}H$-edited NOESY spectrum, (b) molecular model highlighting the glycine packing interactions at the core of the helix and (c) sextet repeat of $WG(PKGPOG)_5 \bullet WG(DOG)_{10} \bullet WG(PKGPOG)_5$. Adapted from Figure 7 in reference 7. In (a), the chemical shift of the ^{15}N-labelled amino acids is highlighted by vertical lines and the chemical shifts of other amino acids by horizontal lines. Resonances relevant to the register determination are highlighted by colored circles in (a) and colored arrows in (b).[7]

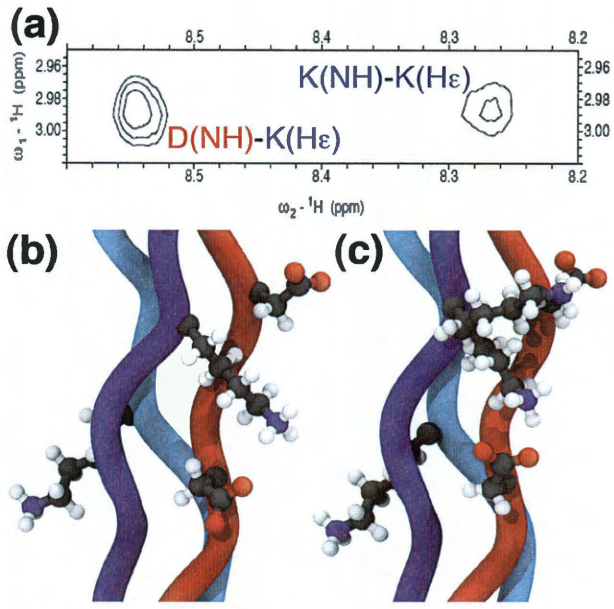

Figure 3.25. (a) $^{1}H,^{1}H$-NOESY spectrum and (b and c) molecular models highlighting the interaction between charged residues in $WG(PKGPOG)_5•WG(DOG)_{10}•WG(PKGPOG)_5$. Adapted from Figure 8 in reference 7. The model in (b) satisfies conformational constraints from (a) but prevents the formation of salt bridges and is depicted as the average between two possible hydrogen bonded conformations, shown in (c).[7]

3.8. Conclusions on Designed Systems

A design scheme for AAB heterotrimer formation in which oppositely charged peptides were mixed in a 1:2 ratio where the more abundant peptide had a charge ½ and opposite of the less abundant peptide was implemented and it examined triple helices with arginine-aspartate (R-D), arginine-glutamate (R-E), lysine-aspartate (K-D) and lysine-glutamate (K-E) based salt bridges.[7,8] The heterotrimers utilized both positive and negative design in which desired heterotrimers are reinforced by favorable interactions of oppositely charged amino acids while undesirable homotrimers were minimized through the reduction of stabilizing POG triplets in each peptide in addition to the incorporation

of charge repulsion.[7,8] The inclusion of negative design within the peptide systems was implemented after the lack of success in the systems described in Chapter 2.

The arginine-aspartate charged pair was the only pairing that did not form a heterotrimer in any of the tested systems.[7] We concluded that this was due to the interaction between the arginine side chain with a backbone carbonyl, which prevented the arginine from adopting the conformation necessary to optimally hydrogen bond with aspartate.[7,8] In contrast, the R-E, K-E and K-D pairings formed heterotrimers in all buffers. The composition and ionic strength of examined buffer systems played a large role in determining heterotrimer stability in these systems and in some instances, a change in buffer allowed for the presence of a heterotrimer to be unveiled. From these pairings, three systems were identified for further analysis with DSC and NMR due to the promising CD results: $(PRG_{10}•2(EOGPOG)_5$, $2(PKGPOG)_5•(EOG)_{10}$ and $2(PKGDOG)_5•(DOG)_{10}$.[7,8]

The peptide system $(PRG)_{10}•2(EOGPOG)_5$ was the first collagen mimetic peptide system that selectively formed a high stability AAB heterotrimer with a thermal melting temperature higher than that of any of its homotrimers analyzed via CD and DSC unfolding studies.[8] 2D NMR experiments confirmed the triple helical nature of the system and identified the dominant species to consist of a single register $(PRG)_{10}•(EOGPOG)_5•$ $(EOGPOG)_5$. The success of this system was likely due to the high content of arginine in the triple helix, which in other studies has been shown to form high stability triple helices in host-guest studies due to the ability of the arginine side chain to hydrogen bond with backbone carbonyls of other peptide chains.[17] However, the

presence of residual homotrimer in the 2D NMR experiments prevented this system from demonstrating compositional control.[8]

In contrast, $2(PKGPOG)_5•(EOG)_{10}$ and $2(PKGPOG)_5•(DOG)_{10}$ were the first reported high-stability collagen-like heterotrimers to form when none of the potential peptides form homotrimers demonstrating the success of the negative design aspect of this method.[7] 2D solution NMR results on these peptide systems demonstrate that, for the first time, a heterotrimeric system was reported in which there is control over heterotrimer composition such that all species within the system were of a single composition.[7] In contrast, all previous reports on POG-containing heterotrimeric systems utilized peptides that formed homotrimers of various quantities. The major component in both peptide mixtures had registers of $(PKGPOG)_5•(EOG)_{10}•(PKGPOG)_5$ and $(PKGPOG)_5•(DOG)_{10}•(PKGPOG)_5$ determined by NMR. Furthermore, all systems containing the combination of lysine and aspartate, regardless of the charge distribution, formed heterotrimers. This suggests the possibility of a direct interaction between the charged residues similar to previous reports on an ABC heterotrimer and a special role for these types of interactions in collagen stabilization.[1]

Together, these results provided a novel design scheme for synthetic extracellular matrix mimetics with the ability to control triple helical composition, a novel result in the CMP field.[7]

3.9. Experimental

Peptide Synthesis and Purification. All peptides were synthesized on an Advanced Chemtech Apex 396 multi-peptide automated synthesizer using standard Fmoc

chemistry for solid phase peptide synthesis. Rink Amide MHBA resin was used for all peptides, which yielded a C-terminal amide group on the peptide. The synthesis was performed at a 0.15 mM scale. Amino acids were added in a 4:1 molar ratio to the growing peptide chain using the coupling agents *O*-benzotriazole-*N*,*N*,*N*',*N*'-tetramethyluroniumhexafluorophosphate (HBTU), 1-hydroxybenzotriazole hydrate (HoBt), and *N*,*N*-diisopropylethylamine (DiEA) in dimethylformamide (DMF) at molar ratios of 4:4:6 respectively. Once coupled to the peptide chain, amino acids were deprotected using a 25 % (by volume) solution of piperidine in DMF. The peptide was acetylated at the N-terminus with an actetylation cocktail of 0.7 mL acetic anhydride and 0.15 mL of diisopropylethylamine in dichloromethane. The peptide was cleaved from the resin with a 74:2:2:1:1 mixture of trifluoroacetic acid (TFA), water, ethylene diamine, triisopropylsilane, and anisole.

Mass Spectrometry. Post synthesis, all peptides were either examined by MALDI/TOF mass spectrometry on a Bruker Autoflex mass spectrometer in positive ion mode or by ESI/TOF mass spectrometry on a Bruker microTOF to verify that the peptides were synthesized correctly. Spectra were analyzed using FlexAnalysis software.

Peptide Purification. Purification was performed on a Varian PrepStar220 HPLC using a preparative reverse phase C-18 column. The two HPLC solvents referred to as solvents A and B are water and acetonitrile, each containing 0.05 % TFA. The solvents were eluted through the column with a linear gradient ranging from a 1 to 3 % increase in concentration of solvent B per minute. Once collected, the HPLC fractions were

rotovapped down to remove the acetonitrile fraction and then lyophilized resulting in a peptide powder.

Sample Preparation. After all peptides were purified and lyophilized, stock solutions for each peptide were made with a 2 mM peptide concentration (measured by mass). Samples were then made with a total peptide concentration of 0.2 mM in one of three buffers, all at pH 7: 10 mM sodium phosphate buffer, 10 mM tris(hydroxymethyl)-aminomethane (Tris) or 10 mM Tris 150 mM sodium chloride. Once made, the solutions were incubated at 5 °C overnight before any characterization was performed. For tryptophan containing samples, a 2 mM peptide concentration (measured by mass) was made. The exact peptide concentration was determined by measuring the sample absorbance using UV-Vis spectroscopy at 280 nm and calculating the concentration using the equation $A=\varepsilon lc$ where A is the absorbance, ε is the molar absorptivity (5502 $cm^{-1}M^{-1}$), l is the pathlength (cm) and c is the concentration (M). Using the calculated concentration, samples were made for the stock solutions with a total peptide concentration of 0.2 mM in one of the three buffers listed above.

Circular Dichroism. All CD experiments were performed with a Jasco J-810 spectropolarimeter equipped with a Peltier temperature control system using quartz cells with a pathlength of 0.1 cm. Samples were heated to 85 °C for 15 minutes and subsequently incubated at 10 °C overnight before spectra and melting experiments were performed. Spectra were taken from 190-250 nm and the wavelength of the maximum seen in the spectra, between 223 and 225 nm, was monitored during thermal unfolding

curves. Melting experiments were performed from 5 to 85 °C with a temperature increase of 10 °C/hr. The first derivative of the melting curve was taken in order to determine the melting temperature (T_m) of the sample. The molar residual ellipticity (MRE, $[\theta]$) is calculated from the measured ellipticity using the equation:

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

where θ is the ellipticity in mdeg, m is the molecular weight in g/mol, c is the concentration in mg/mL, l is the path length of the cuvette in cm, and n_r is the number of amino acids in the peptide.

Differential Scanning Calorimetry. All DSC experiments were performed on a VP-DSC MicroCalorimeter from MicroCal using the same temperature parameters as the CD experiments (range of 5 to 85 °C with a scan rate of 10 °C/hr). After reaching the maximum temperature, the sample was rapidly cooled to 5 °C and equilibrated at that temperature for one hour before beginning the next scan. All samples were dialyzed for three days in buffer prior to each experiment. The DSC curves of the dialysis buffer were used as the baseline and subtracted from each peptide curve prior to data analysis. Heat capacity (C_p) baseline before and after unfolding was also subtracted resulting in a baseline value of zero. During data analysis, the curves were normalized to the triple helix concentration by dividing the measured total peptide concentration (determined by mass) by 3. The melting temperature of the system was defined as the temperature at which the maximum measured C_p was observed.

Nuclear Magnetic Resonance (NMR) for $(PRG)_{10}$•2$(EOGPOG)_5$ System. NMR samples were prepared in a 9:1 ratio of H_2O to D_2O and a 10 mM phosphate buffer to maintain a neutral pH. Two samples were prepared, one containing exclusively the E5 peptide with a concentration of 1.2 mM, determined by mass. A second sample including both the E5 and R10 peptides was prepared in a 1:1:1 ratio, with a total peptide concentration of 3.7 mM, annealed at 85 °C for 15 minutes and then incubated for at least 18 hours at room temperature before beginning the NMR measurements. TSP was used as an internal proton standard in both samples

All NMR experiments were recorded in an 800 MHz Varian spectrometer equipped with a cryogenic probe. The spectra were processed using the NMRpipe[18] software and analyzed using Sparky[19] and ccpnmr.[20] TOCSY spectra with a 75 ms spinlock were acquired for each at 25 °C. For the homotrimer sample, a total of 1918 complex points were recorded in 16 scans for the directly acquired dimension with 360 increments in the sates mode for the indirect dimension while for the heterotrimer sample 1918 complex points were recorded in 8 scans for the directly acquired dimension and 480 increments were recorded in the sates mode for the indirect dimension. NOESY spectra with a 75 ms mixing time were recorded at 25 °C. For the homotrimer sample a total of 3269 complex points were recorded in 8 scans for the directly acquired dimension with 360 increments in the sates mode for the indirect dimension while for the heterotrimer sample 3269 complex points were recorded in 8 scans for the directly acquired dimension and 480 increments were recorded in the sates mode. A square spectral window of 9600 Hz was used for all experiments. Square Cosine bell windows were used as apodization functions and the data was zero-filled to the next power of two

in both dimensions for all experiments. Linear baseline corrections and digital solvent suppression schemes were applied when necessary.

Using the homotrimer sample as a reference, the peaks corresponding to the heterotrimeric species were easily identified. However, the relative amount of homotrimer present could not be determined without the use of labeled backbone atoms due to spectral overlap. The spin systems belonging to each chain of the AAB heterotrimer were distinct but showed identical chemical shifts for every amino acid of every sextet along the sequence. The systems were sequenced using the TOCSY and NOESY spectra. Intra-residue connectivity was readily identified in the TOCSY spectrum and possible inter-residue NOEs from the NH of residue *i* to the $C_\alpha H$ of residue *i-1* were present. Given the high imino acid content of our sequences only the PRG and OGE stretches could be sequenced in a straightforward manner. The glycine spin systems in the EOG triplets were identified using inter-chain NOEs with the arginine side chains, analog to cross-peaks previously observed for long side chains in triple helical conformations or glycine packing interactions. Due to a large overlap of peaks in the aliphatic region of the TOCSY and NOESY spectra the proline spin systems were not unambiguously determined and were not included in the analysis.

NMR for $2(PKGPOG)_5$•$(EOG)_{10}$ and $2(PKGPOG)_5$•$(DOG)_{10}$ Systems. NMR samples were prepared in a 9:1 ratio of H_2O to D_2O at pH 7. Peptides were mixed in a 2:1 ratio to a total concentration of 3.7 mM, determined by tryptophan absorption at 280 nm. The $2WG(PKGPOG)_5$•$WG(DOG)_{10}$ mixture was prepared in 10 mM phosphate buffer and the $2WG(PKGPOG)_5$•$WG(EOG)_{10}$ in 10 mM deuterated Tris (tris(hydroxymethyl)-

aminomethane) buffer. The WG(PKGPOG)$_5$ contains a single ^{15}N-labelled glycine (amino acid 20), purchased from isotech, in the sixth triplet.

All NMR experiments were recorded in a 600 MHz Varian Inova spectrometer at 25 °C unless otherwise noted, processed using the NMRpipe software and analyzed using Sparky. Both systems were characterized using homonuclear ^{1}H,^{1}H-NOESY and ^{1}H,^{1}H-TOCSY experiments. Also, ^{1}H,^{15}N-sofast-HMQC[21] spectra were acquired without ^{15}N decoupling and with a 1 s acquisition time. The in-phase and anti-phase spectra were combined and shifted by $\frac{1}{2}J_{NH}$ in opposite directions to reconstruct the non-split spectra as described by Brutscher et al.[21] Furthermore, ^{1}H,^{1}H-planes of a 3D HNHA and NOESY-^{15}N-HSQC were recorded for each sample by keeping the chemical shift evolution constant in the heteronuclear dimension. For ease of discussion, we will refer to the HNHA experiment as a 2D HNHA and the NOESY-^{15}N-HSQC as an edited NOESY spectrum. The latter was acquired on an 800 MHz Varian spectrometer.

Acquisition and processing parameters for both systems studied were identical for each experiment. Data was zero-filled to next power of two prior to Fourier transforming. Square cosine and cosine bell apodization functions were used in the direct and indirectly detected dimension respectively. *^{1}H,^{1}H-NOESY* – A total of 1366 complex points were acquired in 8 scans for the directly detected dimension and 320 increments in the states mode for the indirect dimension. A spectral width of 8000 Hz was used for both dimensions. *^{1}H,^{1}H-TOCSY* – A total of 2048 complex points were acquired in 8 scans for the directly detected dimension and 420 increments in the states mode for the indirect dimension. A spectral width of 11190 Hz was used for both dimensions. *^{1}H,^{15}N-HMQC* – A total of 12002 complex points were acquired in 16 scans with a spectral width of 6000

Hz for the directly detected dimension and 200 increments in the states mode with a spectral width of 600 Hz for the indirect dimension. The data was acquired using an interleaved in-phase, anti-phase filter and sorted prior to processing. Details on the processing scheme are available in the experimental section. *HNHA* – The experiment was recorded as a $^{1}H,^{1}H$-2D spectrum by holding the heteronuclear evolution period constant. A total of 514 complex points were acquired in 16 scans for the directly detected dimension and 100 increments in the states mode for the indirect dimension. A spectral width of 6000 Hz was used for both dimensions. *NOESY-^{15}N-HSQC* – The experiment was recorded as a $^{1}H,^{1}H$-2D spectrum at 800 MHz by holding the heteronuclear evolution period constant. A total of 1364 complex points were acquired in 16 scans for the directly detected dimension and 420 increments in the states mode for the indirect dimension. A spectral width of 8000 Hz was used for both dimensions.

References

(1) Fallas, J. A.; Gauba, V.; Hartgerink, J. D. *J. Biol. Chem.* **2009**, *284*, 26851-26859.

(2) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 2683-2690.

(3) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 15034-15041.

(4) Li, Y.; Mo, X. A.; Kim, D.; Yu, S. M. *Biopolymers* **2011**, *95*, 94-104.

(5) Madhan, B.; Xiao, J. X.; Thiagarajan, G.; Baum, J.; Brodsky, B. *J. Am. Chem. Soc.* **2008**, *130*, 13520-13521.

(6) Xu, F.; Zhang, L.; Koder, R. L.; Nanda, V. *Biochemistry* **2010**, *49*, 2307-2316.

(7) O'Leary, L. E. R.; Fallas, J. A.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2011**, *133*, 5432-5443.

(8) Russell, L. E.; Fallas, J. A.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2010**, *132*, 3242-3243.

(9) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 7509-7515.

(10) Yang, W.; Chan, V. C.; Kirkpatrick, A.; Ramshaw, J. A. M.; Brodsky, B. *J. Biol. Chem.* **1997**, *272*, 28837-28840.

(11) Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A. *Protein Sci.* **2002**, *11*, 262-270.

(12) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668-1688.

(13) Boudko, S. P.; Engel, J.; Okuyama, K.; Mizuno, K.; Bachinger, H. P.; Schumacher, M. A. *J. Biol. Chem.* **2008**, *283*, 32580-32589.

(14) Kramer, R. Z.; Bella, J.; Brodsky, B.; Berman, H. M. *J. Mol. Biol.* **2001**, *311*, 131-147.

(15) Li, M. H.; Fan, P.; Brodsky, B.; Baum, J. *Biochemistry* **1993**, *32*, 7377-7387.

(16) Gurry, T.; Nerenberg, P. S.; Stultz, C. M. *Biophys. J.* **2010**, *98*, 2634-2643.

(17) Persikov, A. V.; Ramshaw, J. A.; Kirkpatrick, A.; Brodsky, B. *Biochemistry* **2000**, *39*, 14960-14967.

(18) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277-293.

(19) Goddard, T. D.; Kneller, D. G. *SPARKY 3, University of California, San Francisco*

(20) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, P.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins* **2005**, *59*, 687-696.

(21) Schanda, P.; Kupce, E.; Brutscher, B. *J. Biomol. NMR* **2005**, *33*, 199-211.

Chapter 4: Collagen Mimetic Nanofiber Formation Driven by Hydrophobic Interactions

Heterotrimeric collagen mimetic systems have been shown to self-assemble into thermally stable ABC type triple helices.[1,2] These triple helices are stabilized by electrostatic interactions between positively and negatively charged amino acids. Within these charged systems, the peptide combination of $(PKG)_{10}$, $(DOG)_{10}$ and $(POG)_{10}$ resulted in a high stability ABC heterotrimer with a defined register and direct electrostatic interactions between the lysine in triplet *n* and the aspartate in triplet *n+1*.[3] Despite the success in triple helical assembly of these systems, none of them assembled into organized nanofibers, even after months of incubation. In some cases, precipitation occurs at high concentrations, but TEM images of the aggregates showed mesh-like assemblies instead of organized nanofibers, which were similar to those reported for $(POG)_{10}$ homotrimers (Figure 4.1).[4]

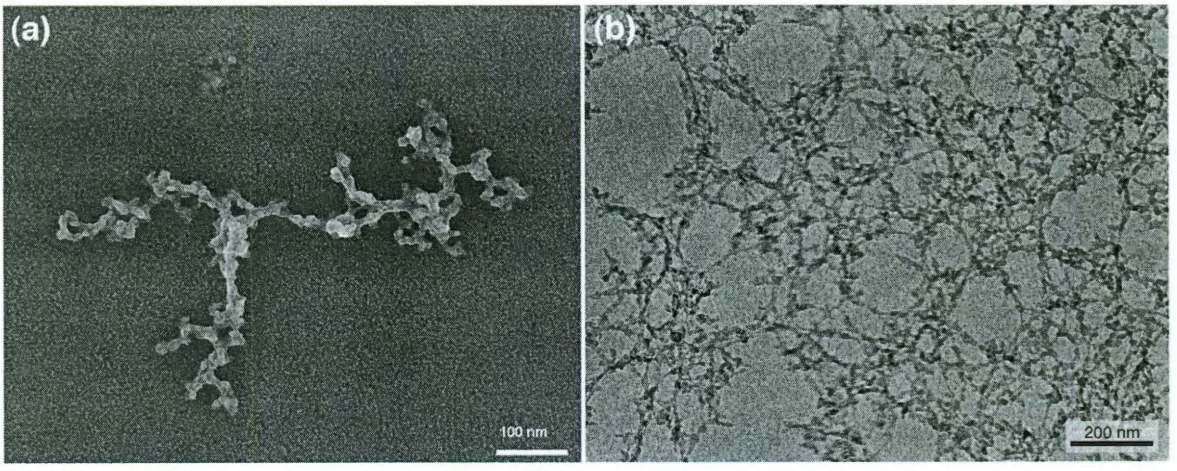

Figure 4.1. Images of (a) $(POG)_{10}$ homotrimer and (b) $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ heterotrimer. The $(POG)_{10}$ electron micrograph is adapted from Figure 6 in reference 4. The cryo-TEM image of the $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ heterotrimer was taken at 12,000X magnification on a 3.6 mM peptide concentration sample that was incubated for one month.[4]

One hypothesis for the inability of $(POG)_{10}$ and the $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ heterotrimer to form fibrils and fibers was the lack of "sticky-ended" structures within the peptide designs that can drive assembly beyond the triple helix. The principle idea of the project described in this chapter was to incorporate hydrophobic sites into peptide design to serve as the sticky-ended driving force to direct the self-assembly of collagen mimetic nanofibers. Therefore, the assembly of the designed peptides utilized two discrete driving forces for each step of self-assembly: charged pair interactions to stabilize the triple helix and hydrophobic packing to guide nanofiber formation. Figure 4.2 highlights the design scheme.

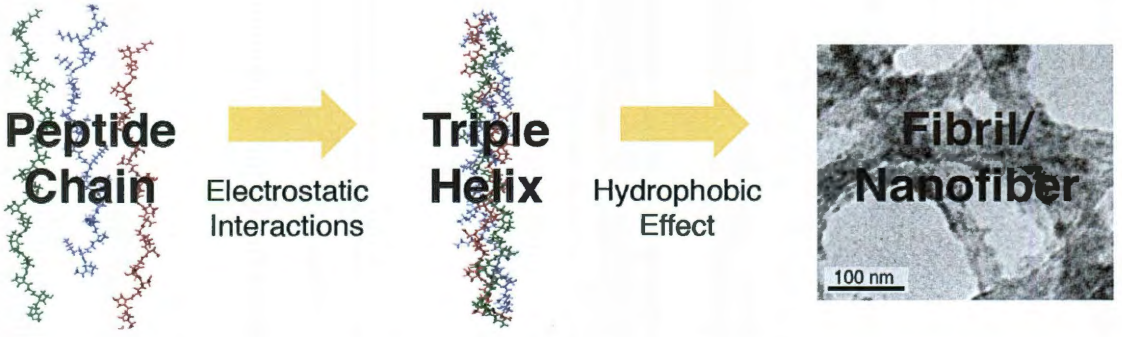

Figure 4.2. Schematic of the driving forces for triple helical and nanofiber assembly within the designed peptide models.

4.1. Peptide Design

Peptides for nanofiber formation used the $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ system as a template where all three peptides were modeled after $(PKG)_{10}$, $(DOG)_{10}$, and $(POG)_{10}$ respectively. Incorporating ideas from fiber formation of coiled coil peptides,[5-13] hydrophobicity was used to drive the packing of triple helices into fibrils and fibers. Four models were designed that created hydrophobic stripe-like regions in unique geometric

patterns. In order to be consistent between models, the amino acid composition (leucine and isoleucine), number of hydrophobic mutations, position of the hydrophobic mutations within the X-Y-Gly amino acid repeating motif and number of lysine-aspartate electrostatic interactions within each system was kept as uniform as possible. With these attempts at standardization, the only variable between models was the three-dimensional placement of the hydrophobic residues within the designed triple helices. The use of leucine in the X-position and isoleucine in the Y-position was based on sequence analysis of collagen types I, III, V and VI[14-17] as well as studies on the stability of POG-containing CMPs when proline and hydroxyproline were mutated.[18,19]

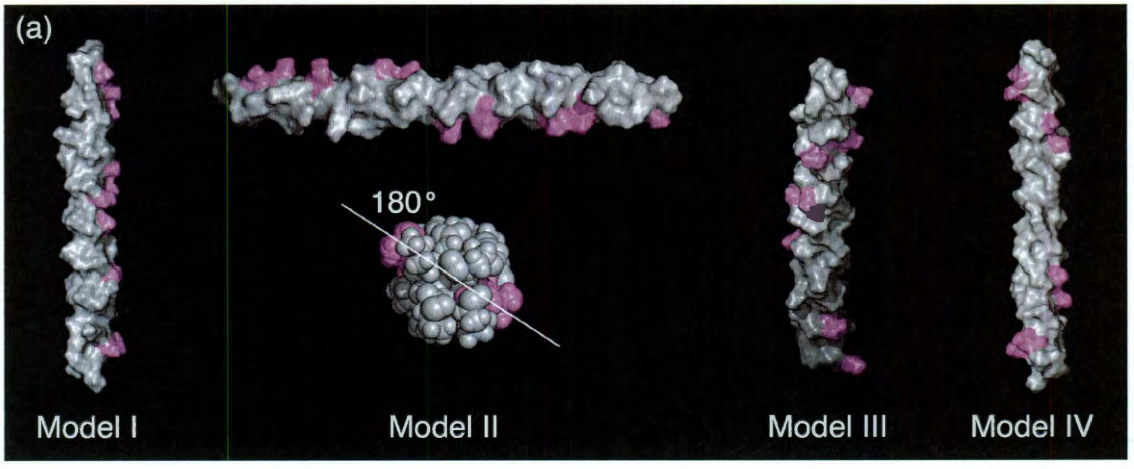

Figure 4.3. Schematic representation of the four models designed for the formation of heterotrimeric nanofibers via hydrophobic interactions. Hydrophobic residues are shown in magenta.

Schematics of all four models are shown in Figure 4.3 with the hydrophobic patches shown in magenta. Model I placed all hydrophobic amino acids on the same face of the triple helix so that looking down the long axis of the heterotrimer, a single linear stripe of hydrophobic residues was present. Model II organized the hydrophobic residues

into two faces within the helix where the hydrophobic stripe in the top half of the triple helix was 180° from the stripe on the bottom half. Model III, replaced amino acids only in the $(POG)_{10}$ template peptide so that the hydrophobic stripe spiraled around the triple helix. Model IV had four hydrophobic patches positioned so that two were on one face of the triple helix, one near each termini, and two closer to the center of the helix, 180° from the first two patches.

Based on the four designed geometries, nanofibers with different axes relative to the triple helical axis were expected. For Model I, the linear stripe of hydrophobicity was capable of blunt-ended dimer formation composed of two triple helices. However, based on results seen for coiled-coil nanofibers,[8] increasing the peptide concentration can drive staggered assembly and the formation of nanofibers along the triple helical axis. Figure 4.4a depicts the hypothesized staggered assembly for Model I. For Model II, the separation of the two hydrophobic patches by 180° suggested that fiber formation would occur neither parallel nor perpendicular to the helical axis, but rather at an angle (Figure 4.4b). Model III contained hydrophobic residues spiraling down the helical axis therefore the direction of fiber formation within this model was unclear. For the last model, Model IV, fiber formation can occur both parallel and perpendicular to the fiber axis due to the presence of four hydrophobic patches distributed over two faces of the helix, shown in Figure 4.4c.

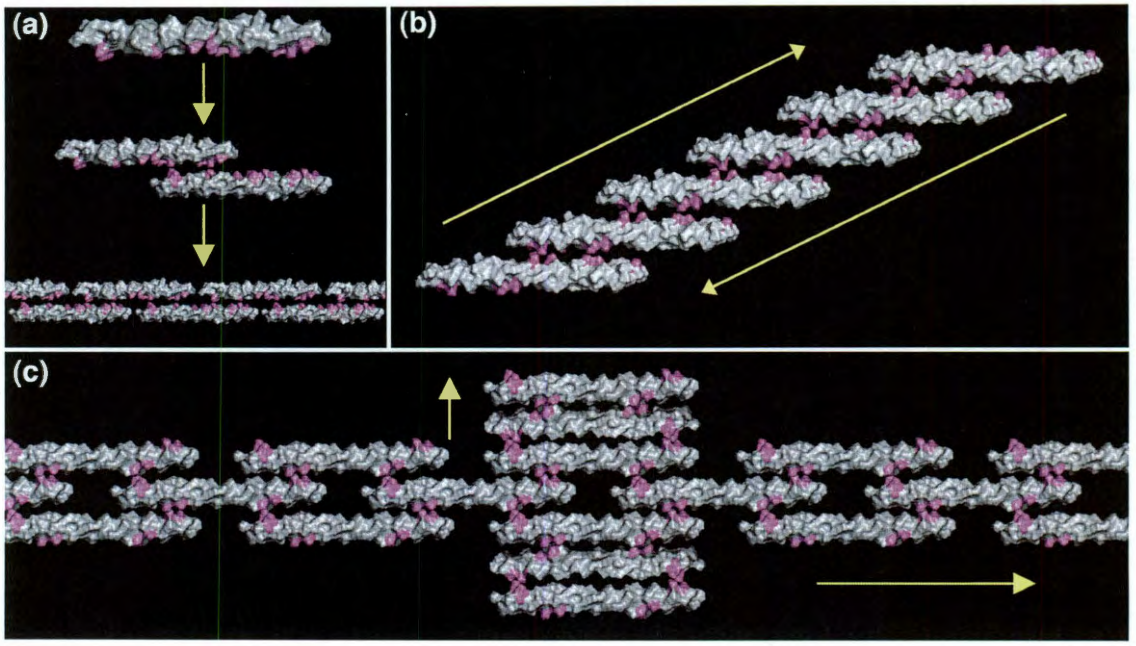

Figure 4.4. Schematic representation of the proposed nanofiber assembly for (a) Model I, (b) Model II and (c) Model IV. The hypothesized axes for nanofiber formation are shown with yellow arrows and the hydrophobic residues are highlighted in magenta.

The sequences for all models examined are given in Table 4.1 in single letter code with the hydrophobic residues (L for leucine and I for isoleucine) in bold. The peptides are labeled based on the model number and the peptide template that the sequence was based off of: **K** for $(PKG)_{10}$, **D** for $(DOG)_{10}$ and **O** for $(POG)_{10}$. The only exception is for Model III in which $(PKG)_{10}$ and $(DOG)_{10}$ are included for all three systems within this model due to the fact that all hydrophobic residues resided within the $(POG)_{10}$ chain. Therefore, the different systems within Model III are named based on the identity of the $(POG)_{10}$-based chain. Model III is the only designed model in which the number of hydrophobic residues was used as a variable, all other systems have 10 total leucines and isoleucines. Due to the simplicity of synthesis within Model III because only one peptide differs from the $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ template, it was easiest to use this model to

examine the effect that adjusting the number of hydrophobic residues had on assembly. Therefore for Model III, systems with 5 and 10 total hydrophobic amino acids were studied.

Model	Abbreviation	Peptide Sequence
Model I	I-K	PKGLKG(PKG)$_7$LIG
	I-D	(DOG)$_3$LOG(DIG)$_2$(DOG)$_4$
	I-O	PIG(POG)$_4$(LOG)$_2$PIG(POG)$_2$
Model II	II-K	(PKG)$_5$LKG(PKG)$_2$LKGLIG
	II-D	LOGDIG(DOG)$_3$DIG(DOG)$_4$
	II-O	(POG)$_2$LOGPIG(POG)$_2$LOG(POG)$_3$
Model III	-	(PKG)$_{10}$
	-	(DOG)$_{10}$
	-	(LOGPOG)$_5$
	-	(POGPIG)$_5$
	-	(LOGPIG)$_5$
Model IV	IV-K	PKGLKG(PKG)$_7$LIG
	IV-D	(DOG)$_7$LIG(DOG)$_2$
	IV-O	PIGPOGLIG(POG)$_7$

Table 4.1. Designed peptide models for hydrophobicity driven nanofiber formation.

Although four models were designed, only two models were examined. Models I and III were chosen for initial studies due to the parallel nature of the fiber axis with respect to the triple helical axis in Model I (similar to the assembly of native collagen) and the ease of synthesis for Model III. Based on the results seen for these systems, no further models were synthesized. A complete discussion for this decision was given in the conclusions section below.

Models I and III were examined for their ability to form stable heterotrimers using circular dichroism (CD) first and then differential scanning calorimetry (DSC) to examine the re-folding of the heterotrimers. Once the presence of stable ABC heterotrimers was confirmed, the nanomorphology of each system was analyzed using

transmission electron microscopy (TEM). The success at each level of assembly will be described sequentially below.

4.2. CD Analysis of Triple Helical Stability

In a similar manner as in Chapters 2 and 3, the analysis of the triple helical stability for a heterotrimer first began with the ability of each peptide to form homotrimers. In addition, since the designed peptide systems are mixtures of three different peptides, the results for the 1:1 mixtures of each component was also analyzed before the 1:1:1 mixture of the three peptides could be examined. These steps were followed for each model sequentially. All peptide systems examined were in phosphate buffer and for the 1:1:1 mixtures, thermal annealing was included in the sample preparation to drive the formation of the most thermodynamically stable species. Samples that did not undergo annealed are referred to as non-annealed and the latter as annealed. CD spectra are reported as molar residual ellipticity (MRE), which normalizes the data for peptide concentration, peptide length and pathlength. Details of sample preparation are given in the experimental section below.

4.2.1. CD Analysis of Model I

The triple helical stability of each of the peptides within Model I was the first set of CD experiments that were performed. As predicted based on the high content of charged amino acids in peptides **I-K** and **I-D**, neither peptide formed a homotrimer: each showed a linear transition in the CD melting experiments. Peptide **I-O** formed a

homotrimer with a melting temperature of 41 °C. This value was far below the melting temperature of 67 °C for $(POG)_{10}$, the template peptide, however due to the presence of four hydrophobic residues distributed within the **I-O** sequence, the reduced thermal stability was expected. The CD melting curves and first derivatives of the melting curves are given in Figure 4.5a and 4.5b respectively.

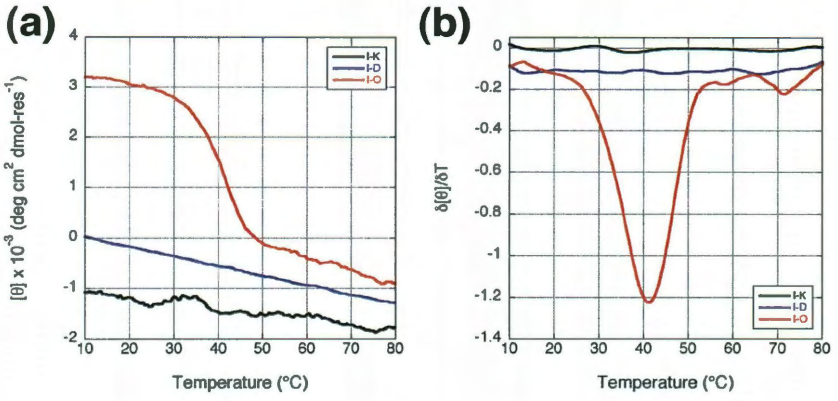

Figure 4.5. CD thermal stability for the component peptides of Model I. (a) Melting curve shown as MRE versus temperature and (b) the first derivative of MRE versus temperature. Data for **I-K** is shown in black, **I-D** is in blue and **I-O** is in red.

The next step in the analysis of ABC heterotrimeric systems was the assessment of any two component peptides to form AAB type heterotrimers. To accomplish this, 1:1 mixtures of peptides were examined. Specifically, annealed mixtures of **I-K•I-D**, **I-D•I-O** and **I-K•I-O** were analyzed. The mixtures were thermally annealed in order to unfold any kinetically trapped species and drive the formation of the most thermodynamically stable species. The CD melting curves and first derivative of the melting curve are shown in Figure 4.6a and 4.6b respectively. The mixture of the two charged peptides, **I-K** and **I-D**, resulted in a single transition at 13 °C. Since neither peptide formed a homotrimer, the

peak must correspond to an AAB heterotrimer. The other two 1:1 mixtures, **I-D•I-O** and **I-K•I-O**, only showed peaks at 41 °C, which overlapped with the **I-O** homotrimer T_m. Therefore, the only combination of peptides that formed an AAB heterotrimer was the mixture of **I-K•I-D**.

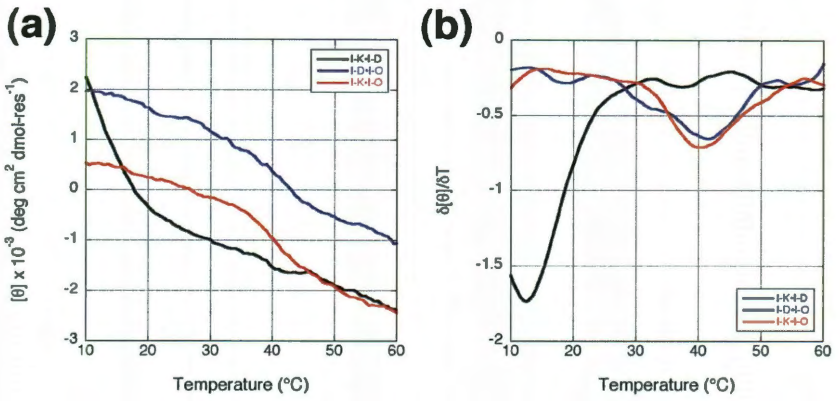

Figure 4.6. CD thermal stability for the 1:1 mixtures of the component peptides of Model I. (a) Melting curve shown as MRE versus temperature and (b) the first derivative of MRE versus temperature. Data for **I-K•I-D** is shown in black, **I-D•I-O** is in blue and **I-K•I-O** is in red.

Once the potential homotrimers and AAB heterotrimers resulting from the combination of the peptides in Model I were analyzed and cataloged, a 1:1:1 mixture of the three peptides was examined. The non-annealed sample of this mixture showed two peaks at 13 and 40 °C. Based on the 1:1 mixture and homotrimer results, these peaks were assigned to an AAB heterotrimer composed of **I-K•I-D** and the **I-O** homotrimer respectively. When the 1:1:1 mixture of the peptides was thermally annealed and then analyzed, a single transition at 36 °C was visible. Due to the fact that a peak at this value was not seen in the homotrimer or 1:1 mixture results, this peak could be assigned as an ABC heterotrimeric triple helix. Therefore, Model I formed an ABC heterotrimer with a

thermal stability visible in CD slightly lower than the **I-O** homotrimer. CD melting studies for the non-annealed and annealed samples for the 1:1:1 mixture are shown in Figure 4.7.

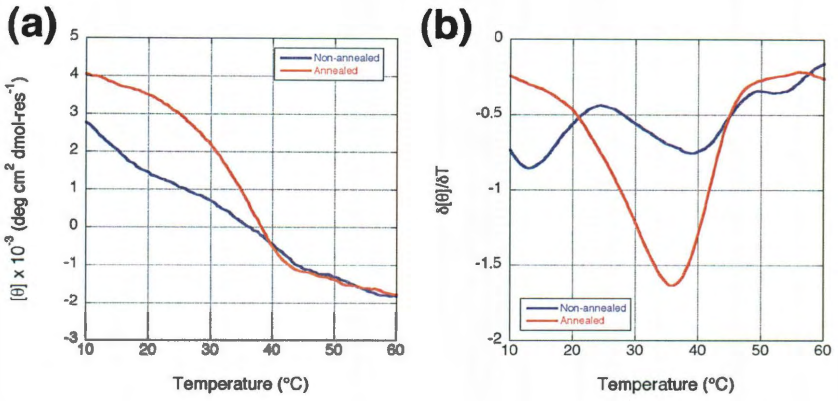

Figure 4.7. CD melting studies on the 1:1:1 mixture of the peptides in Model I. (a) CD thermal unfolding experiment shown as MRE versus temperature. (b) The first derivative of MRE versus temperature. The non-annealed sample is displayed in blue and the annealed sample is shown in red.

4.2.2. CD Analysis of Model III

Peptides for Model III, a ribbon-like stripe of hydrophobicity around the triple helix, were previously synthesized by Varun Gauba for a different project. In the model, these peptides, $(LOGPOG)_5$, $(POGPIG)_5$, and $(LOGPIG)_5$, were mixed with $(PKG)_{10}$ and $(DOG)_{10}$ to form ABC heterotrimers so that the stripe of hydrophobicity was contained within a single peptide chain. The three systems within Model III will be discussed individually below in sections named based on the component hydrophobic peptide.

4.2.2.1. CD Analysis of Model III: $(LOGPOG)_5$

In the same way that Model I was sequentially analyzed beginning with homotrimers, then 1:1 mixtures of component peptides and finally the 1:1:1 mixture of all peptides, $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPOG)_5$ was analyzed via CD melting experiments. Gauba *et al.* reported in 2007 that $(PKG)_{10}$ and $(DOG)_{10}$ do not form homotrimers in phosphate buffer[2] and we repeated these results. The $(LOGPOG)_5$ peptide did form a homotrimer visible by CD in phosphate buffer with a melting temperature of 37 °C. The lowered T_m compared to that of $(POG)_{10}$ was not unexpected due to the presence of five leucine residues within the peptide sequence that each destabilize the homotrimer. CD melting experiments for $(PKG)_{10}$, $(DOG)_{10}$ and $(LOGPOG)_5$ are given in Figure 4.8.

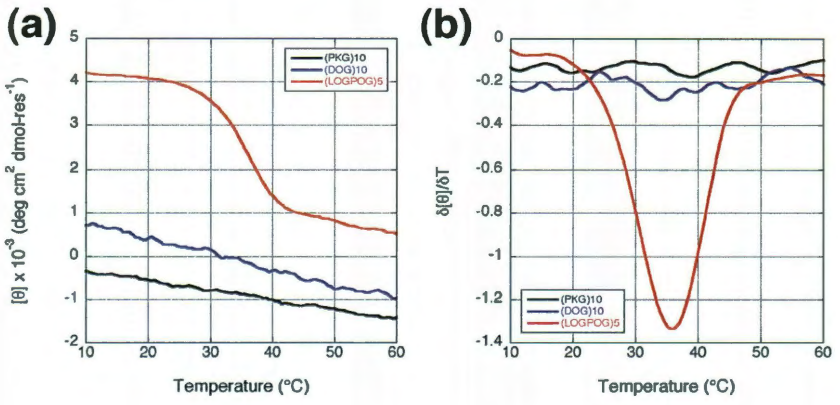

Figure 4.8. CD melting experiments for $(PKG)_{10}$, $(DOG)_{10}$ and $(LOGPOG)_5$ shown in black, blue and red respectively. (a) Thermal unfolding experiment displayed as MRE versus temperature. (b) The first derivative of the unfolding experiments versus temperature.

To continue analysis of the Model III $(LOGPOG)_5$ system, the 1:1 mixtures of the component peptides were examined in CD. The 1:1 mixture of $(PKG)_{10}$ and $(DOG)_{10}$

resulted in a single transition at 36 °C. Similar to the results for the 1:1 mixtures in Model I, the 1:1 mixtures containing the $(POG)_{10}$-based peptide, $(LOGPOG)_5$, only showed transitions that overlapped the homotrimeric T_m of $(LOGPOG)_5$, 37 °C. The CD melting studies for the 1:1 mixtures are shown in Figure 4.9.

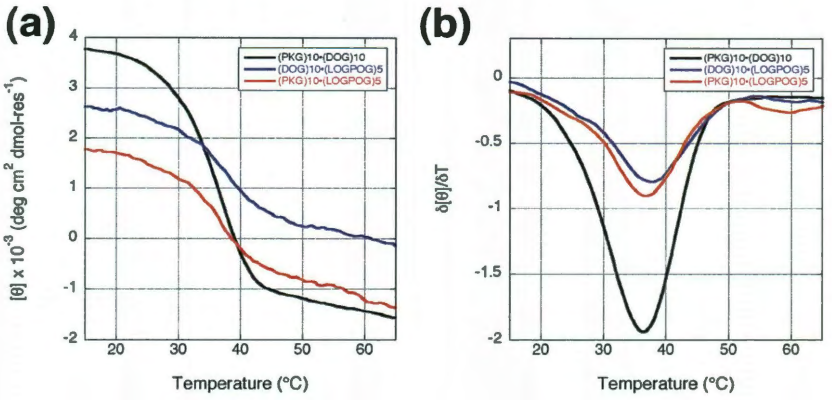

Figure 4.9. CD thermal unfolding studies for the 1:1 mixtures of $(PKG)_{10}$, $(DOG)_{10}$ and $(LOGPOG)_5$. (a) CD melting experiment shown as MRE versus temperature and (b) the first derivative of MRE versus temperature. Curves for the peptide mixtures $(PKG)_{10}$•$(DOG)_{10}$, $(DOG)_{10}$•$(LOGPOG)_5$ and $(PKG)_{10}$•$(LOGPOG)_5$ are displayed in black, blue and red respectively.

After analyzing the ability of the component peptides to form homotrimers and AAB heterotrimers, in order to assess whether the 1:1:1 mixture of $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPOG)_5$ formed an ABC heterotrimer, a single peak at a temperature other than 36-37 °C must be seen. The non-annealed sample for this mixture showed a minor peak at 30 °C and a strong peak at 56 °C. Since neither peak corresponded to homotrimers or AAB heterotrimers, they must be attributed to different ABC heterotrimeric species. When the annealed sample of the mixture was examined, the peak at 30 °C disappeared and only a single transition at 56 °C was visible. The CD

thermal unfolding for the non-annealed and annealed samples are given in Figure 4.10. Based on the differences between the two spectra, the peak at 30 °C was determined to be a low stability ABC heterotrimer that was unfolded during annealing resulting in the high stability ABC heterotrimer that has a T_m of 56 °C. This system was the first observed ABC heterotrimer to have a higher melting temperature than that for the homotrimer of any component peptide.

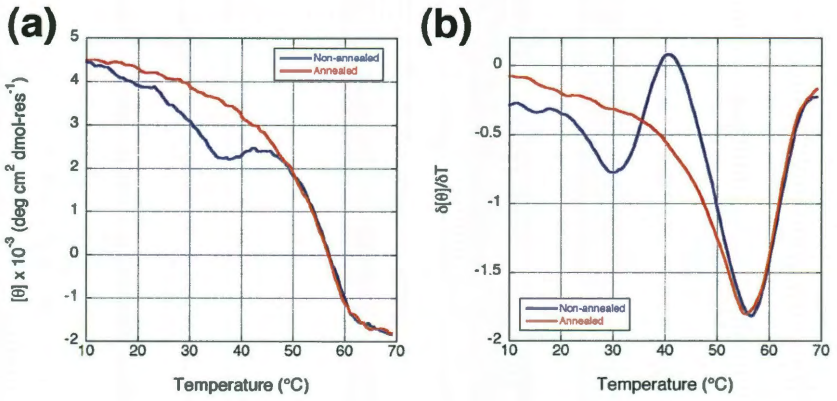

Figure 4.10. CD melting experiments for the 1:1:1 mixture of $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPOG)_5$. (a) CD thermal unfolding curve shown as MRE versus temperature. (b) The first derivative of the thermal unfolding curve versus temperature. Data for the non-annealed sample is shown in blue and the annealed sample is in red.

4.2.2.2. CD Analysis of Model III: $(POGPIG)_5$

Analysis for the Model III system containing $(POGPIG)_5$ complemented that seen for the $(LOGPOG)_5$-containing system in terms of the ability of the component peptides to form homotrimers and based on the 1:1 mixtures of peptides. Since $(PKG)_{10}$ and $(DOG)_{10}$ were also used in this system, the homotrimeric and 1:1 data for these two peptides and their mixture mirrored what was described in the previous section: neither

peptide formed a homotrimer and the 1:1 mixture of the two resulted in an AAB heterotrimer with a T_m of 36 °C. $(POGPIG)_5$ formed a stable homotrimer with a thermal stability of 28 °C. The drastic decrease in stability for the $(POGPIG)_5$ homotrimer as compared to the $(POG)_{10}$ homotrimer was not predicted to be as severe. Based on the studies by Persikov et al., the mutation of a leucine within the X position of the X-Y-Gly triplet should have resulted in a lower stability homotrimer compared to the mutation of an isoleucine residue in the Y position. However, the T_m for $(LOGPOG)_5$ is 37 °C, 10 °C higher than that for $(POGPIG)_5$. Lastly, the 1:1 mixtures of $(DOG)_{10}$•$(POGPIG)_5$ and $(PKG)_{10}$•$(POGPIG)_5$ resulted in a single transition for each mixture whose temperature overlapped that for the $(POGPIG)_5$ homotrimer. Therefore, the only homotrimeric or AAB heterotrimeric species that could form from these component peptides were the $(POGPIG)_5$ homotrimer and the $(PKG)_{10}$•$(DOG)_{10}$ heterotrimer. The CD melting studies for the homotrimers are given in Figure 4.11a and 4.11b and that for the 1:1 mixtures are shown in Figure 4.11c and 4.11d.

When the 1:1:1 mixtures of the $(POGPIG)_5$ system within Model III were examined, a very different story emerged than that for Model I and the $(LOGPOG)_5$ system in Model III. The non-annealed and annealed samples both showed a single transition at 51 °C. This melting temperature was 15 °C higher than the $(PKG)_{10}$•$(DOG)_{10}$ heterotrimer and was 23 °C higher than the $(POGPIG)_5$ homotrimer. However, the novelty in this system was not in its ability to form a heterotrimer with a higher stability than the homotrimer of any component peptide, the $(LOGPOG)_5$ system described above accomplished that. The fact that the peak at 51 °C, which corresponded to an ABC heterotrimer, was the only visible peak in both the non-annealed and annealed samples

made the $(POGPIG)_5$ system distinct from the previously described models. The mixture of $(PKG)_{10}$, $(DOG)_{10}$ and $(POGPIG)_5$ selectively assembled into an ABC heterotrimer without the requirement of thermal annealing, even though one of the component peptides formed a homotrimer. Therefore, the ABC heterotrimer was the most stable and favorable species within the peptide system. The CD melting studies for the 1:1:1 mixtures of $(PKG)_{10}$, $(DOG)_{10}$ and $(POGPIG)_5$ are given in Figure 4.12.

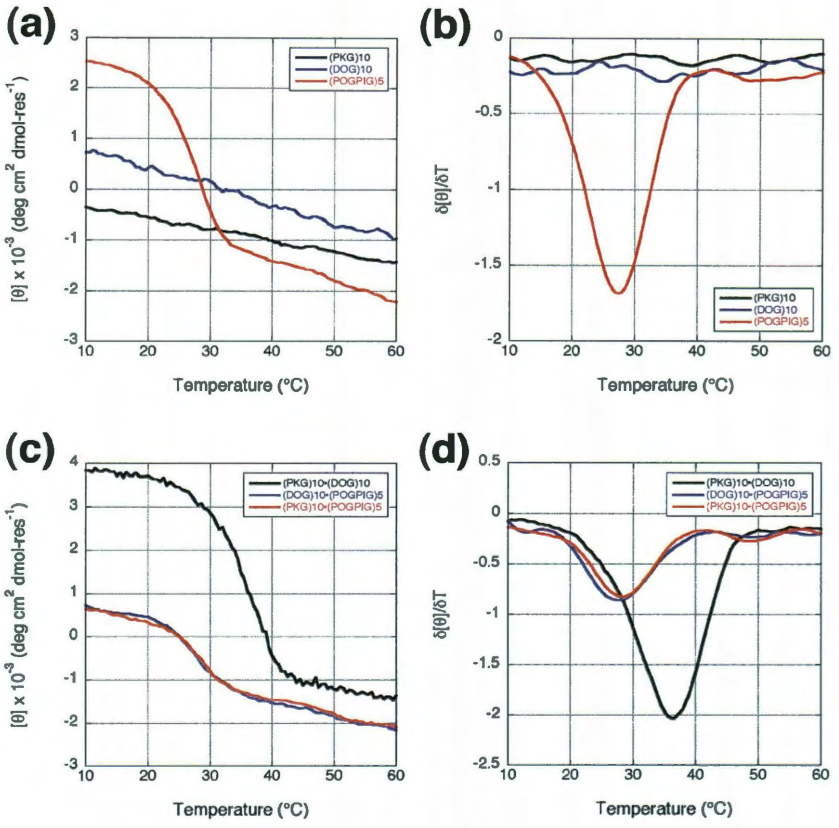

Figure 4.11. (a and b) CD melting experiments for the component peptides of $(PKG)_{10}$, $(DOG)_{10}$ and $(POGPIG)_5$, shown in black, blue and red respectively, displayed as MRE versus temperature (a) and the first derivative of the thermal unfolding curve versus temperature (b). (c and d) CD thermal stability studies for the 1:1 mixtures of $(PKG)_{10}$•$(DOG)_{10}$ in black, $(DOG)_{10}$•$(POGPIG)_5$ in blue and $(PKG)_{10}$•$(POGPIG)_5$ in red. (c) Thermal unfolding curve shown as MRE versus temperature and (d) the first derivative of the thermal unfolding curve versus temperature.

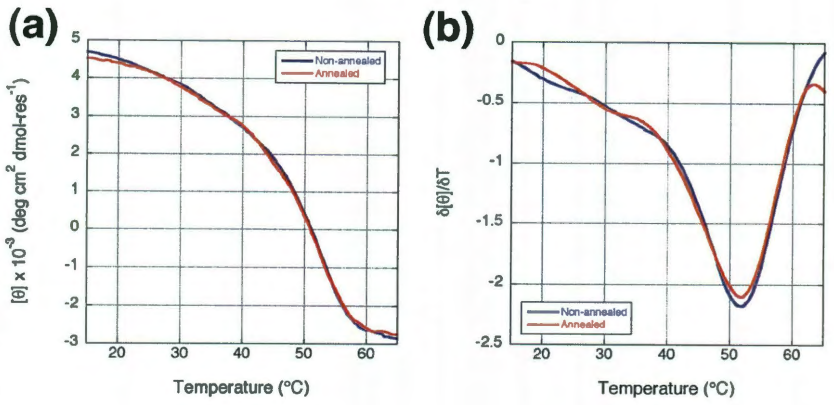

Figure 4.12. CD thermal unfolding studies for the 1:1:1 mixture of $(PKG)_{10}$, $(DOG)_{10}$ and $(POGPIG)_5$ shown as (a) MRE versus temperature and (b) the first derivative of the unfolding curve versus temperature. The non-annealed sample is highlighted in blue and the annealed sample is in red.

4.2.2.3. CD Analysis of Model III: $(LOGPIG)_5$

The last peptide system within Model III used $(LOGPIG)_5$ as the $(POG)_{10}$-based peptide, which contained 10 hydrophobic residues, the same number distributed throughout all of the peptides in Model I and double the amount of amino acids in the $(LOGPOG)_5$ and $(POGPIG)_5$ systems. Therefore, this system was designed to be a comparison with the other Model III systems to assess the effect that a higher number of hydrophobic residues within the triple helix had on thermal stability and on nanofiber growth. In addition, $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPIG)_5$ could be compared with Model I due to the fact that both systems had 10 total hydrophobic residues, the only difference was the geometry of the hydrophobic patches around the triple helix.

To begin analysis of this system, the potential for the component peptides to form homotrimers was analyzed in the exact same manner as the previous systems. $(PKG)_{10}$ and $(DOG)_{10}$ have repeatedly been shown to not form homotrimers in phosphate buffer

and for the first time within this peptide design scheme, the $(POG)_{10}$-based peptide $(LOGPIG)_5$ did not form a homotrimer either. The thermal unfolding curves for all three peptides are given in Figure 4.13.

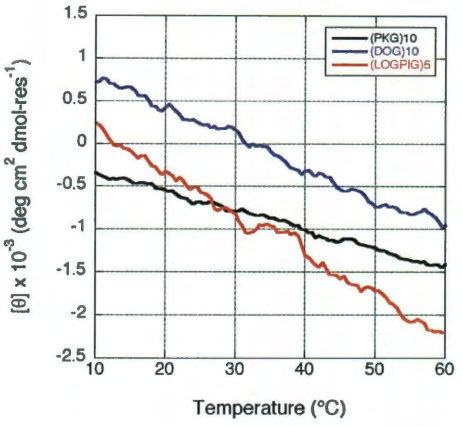

Figure 4.13. CD thermal unfolding curves shown as MRE versus temperature for $(PKG)_{10}$ in black, $(DOG)_{10}$ in blue and $(LOGPIG)_5$ in red.

Due to the fact that none of the component peptides formed homotrimers, the 1:1 mixtures of component peptides were not analyzed, but rather the immediate assessment of the ability of the 1:1:1 mixture of peptides to form an ABC heterotrimer was examined. Disappointingly, neither the non-annealed nor annealed samples for the $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPIG)_5$ mixture showed triple helical formation: both CD melting curves were linear. The CD unfolding curves for both samples are given in Figure 4.14. The inability of this system to form an ABC heterotrimer can be attributed to the high content of hydrophobic residues within the $(POG)_{10}$-based chain. Lysine and aspartate had a lower propensity for the formation of triple helices therefore in order for a system containing $(PKG)_{10}$ and $(DOG)_{10}$ to form a triple helix, the third chain required POG-triplets present in order to template the triple helix formation. This explanation accounts

for the observation of ABC heterotrimers for the $(LOGPOG)_5$ and $(POGPIG)_5$ systems and the absence of a heterotrimer in the $(LOGPIG)_5$ system. Due to the inability of $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPIG)_5$ to form an ABC heterotrimer, no further analysis was performed on this system.

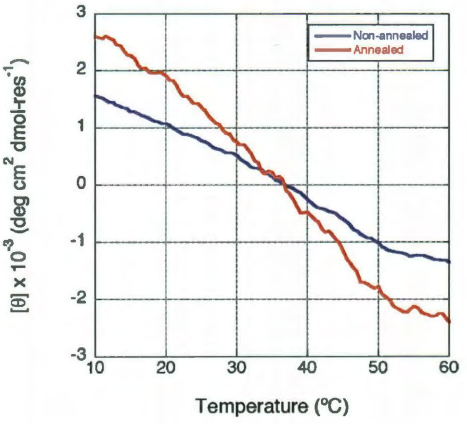

Figure 4.14. CD thermal unfolding curve for the 1:1:1 mixture of $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPIG)_5$ shown as MRE versus temperature. The non-annealed sample is highlighted in blue and the annealed sample is in red.

4.3. DSC Analysis of Triple Helix Re-folding on Selected Systems

After CD analysis on Model I and the three systems within Model III, only one system was eliminated from further analysis: $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPIG)_5$. The peptide mixture did not form a heterotrimer in the 1:1:1 mixture of the peptides indicating the system's inability to form stable triple helices. The next step of analysis for the remaining three systems was differential scanning calorimetry (DSC). DSC melting experiments give an alternative and more sensitive measure of the melting temperature for a system and can indicate the presence of multiple species when the CD studies show a single transition. Additionally, DSC gives information about the thermal recovery, or lack

thereof, for each heterotrimer. As done for the CD thermal stability analysis above, the systems will be discussed individually.

4.3.1. DSC Analysis of Model I

Beginning with Model I, the first DSC experiment replicated the parameters from the CD melting studies in which the system was equilibrated for 15 minutes at the initial temperature (5 °C) before the thermal unfolding was started. The DSC curve for this experiment is given in Figure 4.15a. The first peptide scan for this experiment showed a single peak at 37 °C, the same temperature as the ABC heterotrimer seen in CD studies. However, in all subsequent scans, the strong single peak was replaced with a double peak with maxima at 32 and 37 °C. The latter, as previously identified, corresponded to the ABC heterotrimer however, the new peak at 32 °C did not correspond to any known homotrimer or AAB heterotrimer. The **I-O** homotrimer unfolded at 41 °C and the 1:1 mixture of **I-K•I-D** had a T_m of 13 °C, thus neither could validate the peak at 32 °C. In addition, when the height, width and shape of the peak in the first peptide scan was compared with that of the subsequent scans, the possibility that the peak at 32 °C was present in the first peptide scan but was masked by the height of the peak at 41 °C became apparent. Therefore, we concluded that the peak at 32 °C was most likely an ABC heterotrimer with a different register than the heterotrimer with a T_m of 37 °C, however this conclusion cannot be confirmed based on DSC data alone.

In order to try and prevent the breakdown of the Model I peptide system in DSC, another DSC experiment was setup in which the peptide system was equilibrated for one hour before the thermal unfolding began. The idea behind this modification in the

parameters was that the extra equilibration time would allow the peptide system to properly refold, resulting in a single transition at 37 °C in all peptide scans. The DSC curve for this experiment is given in Figure 4.15b. Despite the extra equilibration time, the peptide scans for the second DSC experiment overlapped that for the first experiment with a single peak in the first peptide scan and a double peak in all subsequent scans. The temperatures associated with these peaks were identical to the 15 minute equilibration experiment, 37 °C for the single peak and then 32 and 37 °C for the double peak. Therefore, the additional equilibration time was not sufficient enough to allow the system to properly refold. Based on the sample preparation parameters in which peptides are mixed, annealed and then incubated for at least 8 hours before analysis and a single peak was seen for the CD studies and the first peptide scan in the DSC experiments, Model I required greater than one hour of equilibration before thermal unfolding but is completely refolded within 8 hours.

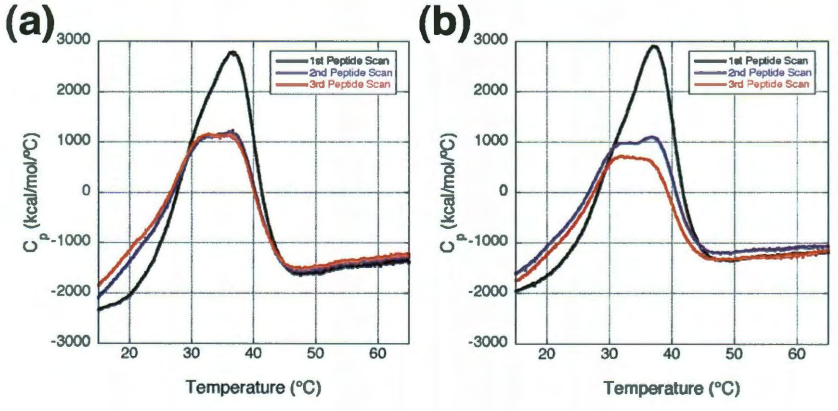

Figure 4.15. DSC melting profiles for Model I in which the sample was equilibrated for (a) 15 minutes between peptide scans and (b) one hour between peptide scans. The first three peptide scans are shown as black, blue and red respectively for each experiment.

4.3.2. DSC Analysis of Model III: $(LOGPOG)_5$

The next system analyzed via DSC was the Model III system containing $(LOGPOG)_5$. As a reminder, this system was composed of a mixture of $(PKG)_{10}$, $(DOG)_{10}$ and $(LOGPOG)_5$. Based on the results seen for Model I, the DSC experiment on this system was setup with an hour long pre-scan equilibration. The DSC curve for the system is shown in Figure 4.16. The first peptide scan showed a single peak at 56 °C and based on the previously described CD results, this peak was attributed to the ABC heterotrimer. In all subsequent scans, the DSC profile maintained a single peak. Therefore the $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPOG)_5$ system refolded within the one hour pre-scan equilibration to a single ABC heterotrimeric species.

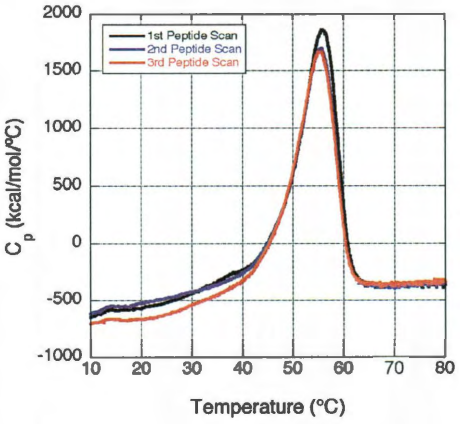

Figure 4.16. DSC melting profile for $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPOG)_5$ in which the sample was equilibrated for one hour between peptide scans. The first three peptide scans are shown in black, blue and red respectively.

4.3.3. DSC Analysis of Model III: $(POGPIG)_5$

The last system examined in DSC was the mixture of $(PKG)_{10}$, $(DOG)_{10}$ and $(POGPIG)_5$. Similar to the analysis for the $(LOGPOG)_5$-containing system, DSC experiments were only run on this system with a pre-scan equilibration of one hour before thermal unfolding (Figure 4.17). The first peptide scan shows a single peak at 51 °C with a small shoulder around 30 °C. The major peak overlapped with the ABC heterotrimeric peak seen in 1:1:1 mixtures in CD. The $(POGPIG)_5$ homotrimer had a melting temperature of 27 °C therefore, the shoulder at about 30 °C could be attributed to residual homotrimer. In the second, third and subsequent scans, the large peak at 51 °C was still visible and the shoulder disappeared. A similar result was seen for the $(PRG)_{10}$•2$(EOGPOG)_5$ system discussed in Chapter 3, where residual homotrimer was present in the first peptide scan but was absent in all subsequent scans due to the relatively long refolding time of the homotrimer compared to the DSC timescale.[20] We concluded that the $(PKG)_{10}$•$(DOG)_{10}$•$(POGPIG)_5$ system behaved similarly.

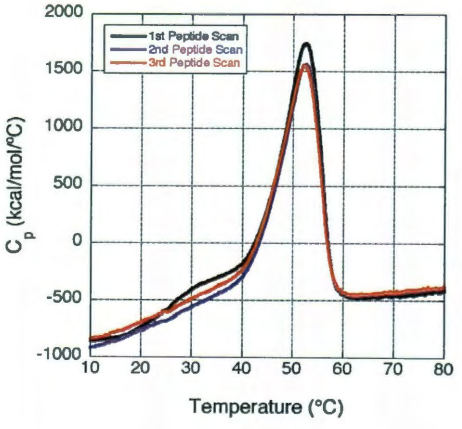

Figure 4.17. DSC melting profile for $(PKG)_{10}$•$(DOG)_{10}$•$(POGPIG)_5$ in which the sample was equilibrated for one hour before thermal unfolding. The first three peptide scans are shown in black, blue and red respectively.

4.4. Nanofiber Formation

Through CD and DSC melting experiments, the ability of Model I, $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPOG)_5$ and $(PKG)_{10}$•$(DOG)_{10}$•$(POGPIG)_5$ to form ABC heterotrimers was confirmed and the re-folding capabilities of each was determined. Therefore, the next step in the analysis of the multi-step assembly of the designed hydrophobic systems was to assess the nanomorphology of the systems using transmission electron microscopy (TEM). Samples at peptide concentrations ranging from 5 to 15 mM with buffer conditions spanning from 10 to 100 mM phosphate, pH 7, were analyzed so that visible aggregation could be used as an indication of assembly beyond the triple helix. The increase in peptide concentration, as alluded to in the peptide design section above, was included to drive the assembly of sticky-ended moieties over blunt-ended structures. The range of ionic strengths within the buffers was explored under the hypothesis that a higher ionic strength would shield the inter-helix charge repulsions and allow for aggregation of the helices into nanostructures. Once the timeline for sample precipitation at a given concentration was established, samples for TEM were prepared before aggregation was visible by the eye. Details of the TEM sample preparation and imaging are given in the experimental section below.

Beginning with Model I, sample aggregation was seen in 5 mM total peptide concentration samples prepared in 10 mM phosphate that were allowed to evaporate to about half volume in a desiccators When analyzed using dry-TEM techniques, large linear fibers were visible, the magnitude of which was almost beyond the scale of TEM. Dry TEM samples were negatively stained with uranyl acetate in order to enhance the contrast of the peptide system and an example TEM image is shown in Figure 4.18a. In

order to properly assess the nanomorphology for peptide systems, cryo-TEM must be used so that the morphology in solution state is visible. In cryo-TEM, the peptide sample is frozen in vitreous ice and then it is imaged at the temperature of liquid nitrogen. Therefore, no heavy metal stain can be used to enhance the contrast between the peptide system and the carbon-coated TEM grid and resulting cryo-TEM images can be faint in appearance. An example cryo-TEM image for Model I is shown in Figure 4.18b. Large aggregates of liquid ethane were visible within the image, which were artifacts from the sample freezing process (details of sample preparation are given in the experimental section below). However, underneath the ethane, a peptide meshwork was visible. The lack of organization and directionality within the peptide meshwork was disappointing when compared to the dry-TEM images taken for Model I, which showed large linear fibers. Therefore, Model I did not form organized nanofibers in the solution state and did not replicate the assembly of collagen.

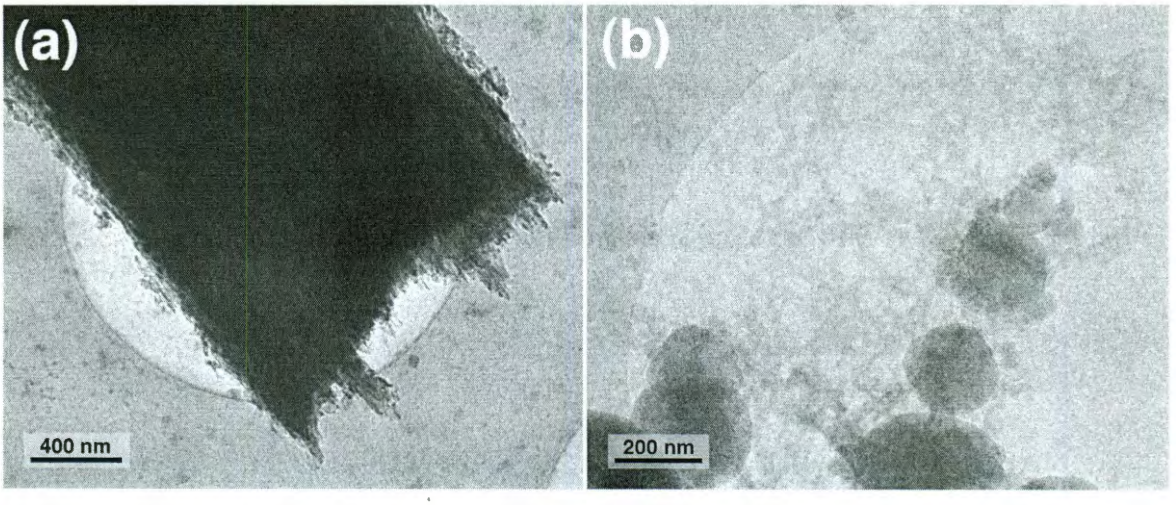

Figure 4.18. TEM images of hydrophobic Model I. (a) Uranyl acetate stained TEM image taken at 8,000X magnification. (b) Cryo-TEM image taken at 12,000X magnification.

Although Model I did not form visible nanofibers in the solution state, we were optimistic about analyzing the systems from Model III in TEM due to the superior CD and DSC profiles seen for these systems compared to Model I. Starting with the $(LOGPOG)_5$-containing mixture, only dry-TEM was taken for this system. Due to the complexity of sample preparation and imaging, no cryo-TEM images were acquired to determine the solution state nanomorphology for $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPOG)_5$. The uranyl acetate stained images taken showed long linear nanofibers. The fiber sizes were noticeably smaller than those seen for Model I, and a possible reason for this was the reduced number of hydrophobic residues present in the $(LOGPOG)_5$ compared to Model I. Example dry-TEM images for the $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPOG)_5$ system are shown in Figure 4.19.

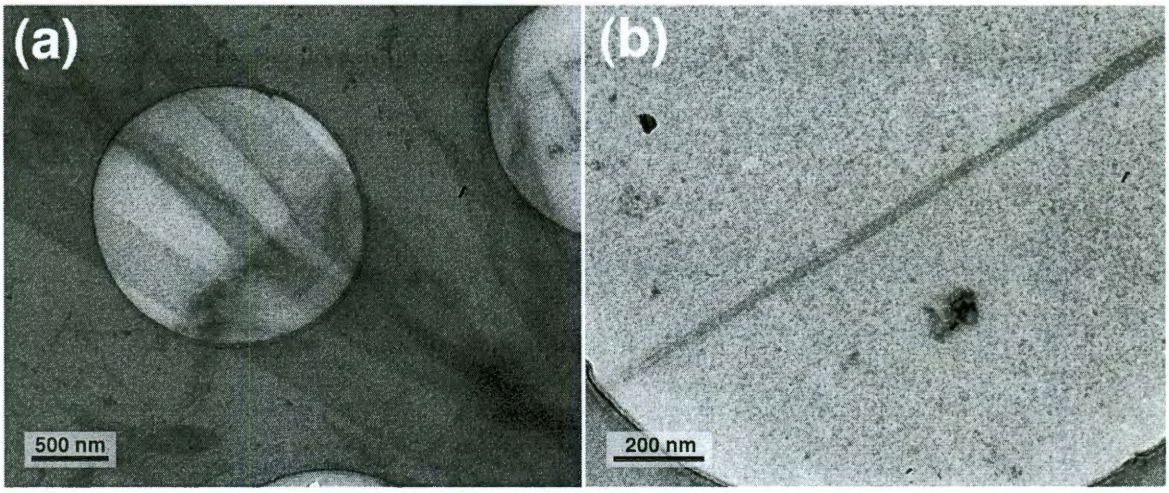

Figure 4.19. Uranyl acetate stained dry-TEM images of $(PKG)_{10}$•$(DOG)_{10}$•$(LOGPOG)_5$ at magnifications of (a) 8,000X and (b) 12,000X.

Finally, $(PKG)_{10}$•$(DOG)_{10}$•$(POGPIG)_5$ was examined via dry and cryo-TEM. Examples images of each are shown in Figure 4.20. The uranyl acetate stained samples for this system differed from those seen for Model I and the $(LOGPOG)_5$-containing

system: no large fiber precipitates were seen. In contrast, smaller fibers were visible amidst large uranyl acetate aggregates. In Figure 4.20a, the contrast was low for the image due to the large uranyl acetate particles within the image. Based on dry-TEM alone, the $(PKG)_{10}$•$(DOG)_{10}$•$(POGPIG)_5$ system appeared unique compared to the other hydrophobic systems designed. However, in cryo-TEM, a peptide meshwork similar to that seen for Model I and $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ shown in Figure 4.18b and 4.1b respectively. The lack of organized nanofibers within the solution state prevented the $(PKG)_{10}$•$(DOG)_{10}$•$(POGPIG)_5$ from mimicking the higher order assembly seen in natural collagen.

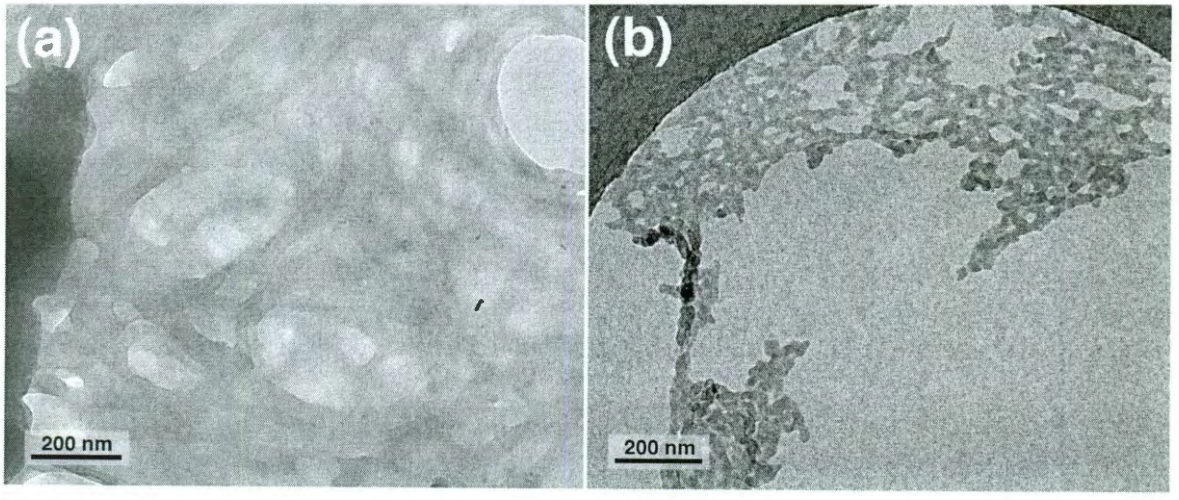

Figure 4.20. TEM images of hydrophobic model $(PKG)_{10}$•$(DOG)_{10}$•$(POGPIG)_5$. (a) Uranyl acetate stained TEM image taken at 12,000X magnification. (b) Cryo-TEM image taken at 12,000X magnification.

4.5. Conclusions

A series of peptides based on the $(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ system were designed with the intent of driving nanofiber formation. Electrostatic interactions between lysine and aspartate were utilized to assemble the peptide mixtures into ABC

heterotrimers and then hydrophobic residues incorporated into the peptide sequences were included to facilitate assembly beyond the triple helix, into nanofibers. The use of hydrophobic residues within the peptide design was to create separate mechanisms for each step of the multi-hierarchical assembly of collagen. Four models were designed that differed in the geometry of the hydrophobic amino acids within the triple helix and from those, two were thoroughly examined in CD and DSC for the triple helical stability of heterotrimers and in TEM for the nanomorphology. Despite the fact that multiple systems formed ABC heterotrimers, two of which had higher melting temperatures than homotrimers of the component peptides, none of the examined systems formed nanofibers visible in the solution state. Large fiber aggregates were seen in dry-TEM that were not easily reproducible. Therefore, the inability to control assembly as well as the lack of reproducibility in the fiber results for these peptide models led to the termination of this design scheme.

4.6. Experimental

Peptide Synthesis and Purification. All peptides were synthesized using standard Fmoc chemistry for solid state peptide synthesis on an Advanced Chemtech Apex 396 multi-peptide automated synthesizer. Peptides were grown in a Wang resin pre-loaded with a glycine residue resulting in a free carboxyl on the C-terminus. The synthesis was performed at a 0.15 mM scale and amino acids were added in a 4:1 molar ratio to the growing peptide chain using the coupling agents *O*-benzotriazole-*N*,*N*,*N*',*N*'-tetramethyluroniumhexafluorophosphate (HBTU), 1-hydroxybenzotriazole hydrate (HoBt), and *N*,*N*-diisopropylethylamine (DiEA) in dimethylformamide (DMF) at molar

ratios of 4:4:6 respectively. Amino acids were deprotected once coupled to the peptide chain using a 25% (by volume) solution of piperidine in DMF. The peptide was cleaved from the resin with a 38:1:1 mixture of trifluoroacetic acid (TFA), water and triisopropylsilane.

Mass Spectrometry. Post synthesis, all peptides were examined by MALDI/TOF mass spectrometry on a Bruker Autoflex mass spectrometer in positive ion mode to verify that the peptides were synthesized correctly. Spectra were analyzed using FlexAnalysis software.

Peptide Purification. Purification was performed on a Varian PrepStar220 HPLC using a preparative reverse phase C-18 column. The two HPLC solvents referred to as solvents A and B are water and acetonitrile, respectively, each containing 0.05% TFA. The solvents were eluted through the column with a linear gradient ranging from a 1 to 3% increase in concentration of solvent B per minute. Once collected, the HPLC fractions were rotovapped down to remove the acetonitrile fraction and then lyophilized resulting in a peptide powder.

Sample Preparation. After all peptides were purified and lyophilized, stock solutions for each peptide were made with a 2 mM peptide concentration (measured by mass). Samples were then made with a total peptide concentration of 0.2 mM in10 mM sodium phosphate buffer, pH 7. Annealed samples were pre-heated at 85 °C for 15 minutes and then incubated at 10 °C overnight before any characterization was

performed. Non-annealed samples were simply mixed and then incubated. When higher concentration samples were made for TEM, a similar procedure was followed in which the peptide stock solutions had a ten-fold higher concentration than the total peptide concentration desired in the sample.

Circular Dichroism. All CD experiments were performed with a Jasco J-810 spectropolarimeter equipped with a Peltier temperature control system using quartz cells with a pathlength of 0.1 cm. Samples were heated to 85 °C for 15 minutes and subsequently incubated at 10 °C overnight before spectra and melting experiments were performed. Spectra were taken from 190-250 nm and the wavelength of the maximum seen in the spectra, between 223 and 225 nm, was monitored during thermal unfolding curves. Melting experiments were performed from 5 to 85 °C with a temperature increase of 10 °C/hr. The first derivative of the melting curve was taken in order to determine the melting temperature (T_m) of the sample. The molar residual ellipticity (MRE, $[\theta]$) is calculated from the measured ellipticity using the equation:

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

where θ is the ellipticity in mdeg, m is the molecular weight in g/mol, c is the concentration in mg/mL, l is the path length of the cuvette in cm, and n_r is the number of amino acids in the peptide.

Differential Scanning Calorimetry. All DSC experiments were performed on a VP-DSC MicroCalorimeter from MicroCal using the same temperature parameters as the CD experiments (range of 5 to 85 °C with a scan rate of 10 °C/hr). After reaching the

maximum temperature, the sample was rapidly cooled to 5 °C and equilibrated at that temperature for either 15 minutes or one hour before beginning the next scan. All samples were dialyzed for three days in buffer prior to each experiment. The DSC curves of the dialysis buffer were used as the baseline and subtracted from each peptide curve prior to data analysis. Heat capacity (C_p) baseline before and after unfolding was also subtracted resulting in a baseline value of zero. During data analysis, the curves were normalized to the triple helix concentration by dividing the measured total peptide concentration (determined by mass) by 3. The melting temperature of the system was defined as the temperature at which the maximum measured C_p was observed.

Transmission Electron Microscopy (TEM). Samples for TEM imaging were prepared on Quantifoil® R1.2/1.3 holey carbon mesh on copper grids. For dry TEM, uranyl acetate (UA) was used to stain the TEM grids using a positive staining technique. A 0.5 % (by weight) solution of UA was made (pH 3.5) bi-weekly and syringe filtered prior to use in order to remove heavy metal aggregates. For dry TEM sample preparation, the peptide solution was added to the carbon side of a TEM grid, allowed to dry for one minute, then indirectly blotted with filter paper to remove excess solution. For positive staining, UA solution was drop-wise added to the grid for 10 seconds and then immersed in water two times. The grid was then allowed to dry in ambient conditions overnight.

Vitreous ice TEM samples were prepared very differently. First, the TEM grids were glow discharged for one minute with a 5 mA discharge. The next stages of sample preparation were all performed using a Vitrobot type FP5350/60. The peptide solution

was added to the grid and immediately blotted for one second before being immersed in liquid ethane. The grid was then manually transferred from the liquid ethane to liquid nitrogen where it was stored until imaging.

All TEM imaging was performed on a JEOL 2010 microscope (200 kV) and cryo-imaging was taken at a temperature of -176 °C using low dose conditions.

References

(1) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 2683-2690.

(2) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 15034-15041.

(3) Fallas, J. A.; Gauba, V.; Hartgerink, J. D. *J. Biol. Chem.* **2009**, *284*, 26851-26859.

(4) Kar, K.; Amin, P.; Bryan, M. A.; Persikov, A. V.; Mohs, A.; Wang, Y. H.; Brodsky, B. *J. Biol. Chem.* **2006**, *281*, 33283-33290.

(5) Apostolovic, B.; Danial, M.; Klok, H. A. *Chem. Soc. Rev.* **2010**, *39*, 3541-3575.

(6) Dong, H.; Hartgerink, J. D. *Biomacromolecules* **2006**, *7*, 691-695.

(7) Dong, H.; Hartgerink, J. D. *Biomacromolecules* **2007**, *8*, 617-623.

(8) Dong, H.; Paramonov, S. E.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 13691-13695.

(9) Dragan, A. I.; Potekhin, S. A.; Sivolob, A.; Lu, M.; Privalov, P. L. *Biochemistry* **2004**, *43*, 14891-14900.

(10) Melnik, T. N.; Villard, V.; Vasiliev, V.; Corradin, G.; Kajava, A. V.; Potekhin, S. A. *Protein Eng.* **2003**, *16*, 1125-1130.

(11) Potekhin, S. A.; Melnik, T. N.; Popov, V.; Lanina, N. F.; Vazina, A. A.; Rigler, P.; Verdini, A. S.; Corradin, G.; Kajava, A. V. *Chem. Biol.* **2001**, *8*, 1025-1032.

(12) Woolfson, D. N. *Adv. Protein Chem.* **2005**, *70*, 79-112.

(13) Woolfson, D. N. *Biopolymers* **2010**, *94*, 118-127.

(14) Chu, M. L.; Conway, D.; Pan, T.; Baldwin, C.; Mann, K.; Deutzmann, R.; Timpl, R. *J. Biol. Chem.* **1988**, *263*, 18601-18606.

(15) Greenspan, D. S.; Cheng, W.; Hoffman, G. G. *J. Biol. Chem.* **1991**, *266*, 24727-24733.

(16) Imamura, Y.; Scott, I. C.; Greenspan, D. S. *J. Biol. Chem.* **2000**, *275*, 8749-8759.

(17) Myers, J. C.; Loidl, H. R.; Stolle, C. A.; Seyer, J. M. *J. Biol. Chem.* **1985**, *260*, 5533-5541.

(18) Persikov, A. V.; Ramshaw, J. A.; Kirkpatrick, A.; Brodsky, B. *Biochemistry* **2000**, *39*, 14960-14967.

(19) Shah, N. K.; Ramshaw, J. A.; Kirkpatrick, A.; Shah, C.; Brodsky, B. *Biochemistry* **1996**, *35*, 10262-10268.

(20) Russell, L. E.; Fallas, J. A.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2010**, *132*, 3242-3243.

Chapter 5: Collagen Mimetic Nanofiber Formation Driven by Electrostatic Interactions*

The use of hydrophobicity to drive the formation of collagen mimetic nanofibers proved ineffective due to the lack of reproducibility of the nanofiber results. Triple helical stability was successful however fiber formation could not be stopped once it had begun resulting in large aggregates that precipitated out of solution. A successful collagen mimetic system must simultaneously demonstrate all levels of the structural assembly of collagen: peptide chain to triple helix to nanofibers and finally a hydrogel. In 2007, Chaikof *et al.* published a homotrimeric system that formed D-periodic nanofibers visible in dry-TEM.[1] The peptide, $(PRG)_4(POG)_4(EOG)_4$, was designed with a sticky-end motif such that at physiological pH, the positively charged arginine-containing N-terminal region would form electrostatic interactions with the negatively charged glutamate-containing C-terminal region. Based on these results, the assembly of $(PRG)_4(POG)_4(EOG)_4$ was replicated in order to assess the advantages and drawbacks of the system. Once analyzed, modifications to the system were made in order to test the positioning of the charged amino acids (positive at the N-terminus versus C-terminus) and the choice of amino acids within these regions.

---

* The work in this chapter was done in collaboration with Jorge Fallas , Erica Bakota and Marci Kang, all of the Chemistry Department and all advised by Prof. Jeff Hartgerink. Jorge performed the fiber diffraction experiments, Erica contributed the SEM sample preparation and imaging and Marci carried out the biocompatibility tests. This work was published as: Lesley E.R. O'Leary, Jorge A. Fallas, Erica L. Bakota, Marci K. Kang and Jeffrey D. Hartgerink. Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. *Nature Chem.*, **2011**, 3, 821-828.

5.1. $(PRG)_4(POG)_4(EOG)_4$ Replication

In the report by Chaikof *et al.*, nanofiber formation was only seen in a small window of sample preparation techniques.[1] The peptide, $(PRG)_4(POG)_4(EOG)_4$, was dissolved in 10 mM hydrochloric acid (HCl) and then combined with a high ionic strength buffer system commonly used for rat-tail collagen fiber formation to give a final peptide concentration of 0.071% by weight (0.2 mM). The addition of this buffer system (30 mM Na_2HPO_4, 30 mM TES (N-[tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid) and 135 mM NaCl, pH 7.4), referred to as Buffer 8, adjusted the pH of the peptide from below the pK_a of glutamate to above it, which produced a net negative charge in the glutamate region of the peptide. Chaikof *et al.* proposed that a stable triple helix formed at acidic pH and the adjustment to pH 7 allowed the triple helices to interdigitate forming nanofibers. However, fibers with D-periodicity were only reported for samples that were annealed (incubated at 70 °C for 40 minutes then gradually cooled to room temperature) after the addition of the buffer cocktail, which is a process used to drive the formation of the most thermodynamically stable species. In order to understand the reported success with the peptide $(PRG)_4(POG)_4(EOG)_4$, replication of the Chaikof results and analysis at each step of assembly was performed.

5.1.1. $(PRG)_4(POG)_4(EOG)_4$: Triple Helical Stability

In the 2007 paper, Chaikof and co-workers tested the triple helical stability of $(PRG)_4(POG)_4(EOG)_4$ in many different buffer systems with ionic strengths ranging from deionized water to the high ionic strength Buffer 8 described above. The reported first

derivative of the melting curves for these buffers are shown in Figure 5.1. The peptide system had different melting temperatures (T_m) in each buffer ranging from 37 to 48 °C, with the lowest stability in Buffer 8 and the highest stability in PBS (phosphate buffered saline).[1] One thing to note about the data shown in Figure 5.1 is that the intensities of the peaks in the first derivative of the melting curves are very small compared to those for other triple helical systems presented in the previous chapters. For example, the ABC triple helix formed by Model I in Chapter 3 had melting transition with an intensity of 1.0 for the first derivative. In general, a sample is considered more triple helical than another if it's melting transition has a larger intensity than that of the other sample.

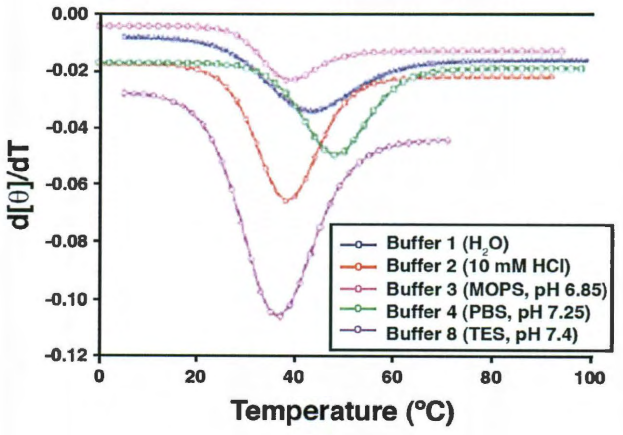

Figure 5.1. First derivative of the melting curve versus temperature for $(PRG)_4(POG)_4(EOG)_4$ in different buffers reported by Chaikof *et al.* Adapted from Figure 2b in reference 1.[1]

Due to the fact that D-periodic nanofibers were only reported in Buffer 8, triple helical stability was only replicated in that specific buffer cocktail. Following the sample preparation procedure described above and in the experimental section below,[1] samples were made for CD melting analysis. In the thermal unfolding curves, a strong transition at

14 °C was seen as well as two minor transitions at 43 °C and about 70 °C. The melting curve and the first derivative of the melting experiment are shown in green in Figure 5.2. When the intensities of the observed transitions were considered, the peak at 14 °C has an intensity far greater than that seen in the CD melting curves reported by Chaikof *et al.*[1] However, the peak at 43 °C overlapped the melting temperatures reported and it had an intensity very similar to that seen for Buffer 8 in the published CD study. Therefore, it is feasible that this transition reflects that seen by Chaikof. The observe T_m was 6 °C higher than the literature value in Buffer 8 however, the broad nature of the reproduced CD melting study suggested that the true T_m could be anywhere between 37 and 45 °C. As for the last transition, at about 70 °C, the melting experiment reported in the publication ended around 70 °C therefore this transition would not have been seen in the published work. Based on these results, I was only partially able to replicate the CD thermal stability seen for $(PRG)_4(POG)_4(EOG)_4$ in Buffer 8. The small transition at 43 °C in the replicated work could correspond to the published T_m for the peptide in Buffer 8, but the stronger transition seen at 14 °C was not visible in any buffer in Figure 5.1. This inconsistency raised a question about the validity of the reported data.

Despite the discrepancy between the observed and reported CD results for $(PRG)_4(POG)_4(EOG)_4$ in Buffer 8, an additional series of CD melting studies were performed at different pH values to analyze the proposed self-assembly mechanism for the peptide. Peptide samples were made at pH 2 and pH 12.7, values that are below the pK_a of glutamate and above the pK_a of arginine respectively, and the thermal unfolding studies for each are shown in Figure 5.2. Beginning with the high pH sample, shown in blue, a strong transition was seen with a T_m of 46 °C. This value was very similar to the

minor peak seen in the reproduced Buffer 8 study discussed above however, its intensity is far greater. The low pH sample, shown in red, had a much lower intensity than that seen for the high pH sample, however its intensity was comparable to those reported by Chaikof *et al.* The peak was centered at 38 °C, which overlapped the reported peak. Based on Chaikof's hypothesis for peptide assembly, $(PRG)_4(POG)_4(EOG)_4$ should have formed a stable triple helix at low pH. Instead, the peptide showed a strong transition at high pH and a low intensity peak at low pH. Additionally, the published CD thermal stabilities for Buffer 8 were more similar to the reproduced studies at a pH below the pK_a of glutamate. Therefore, the reported triple helical stability was most likely attributed to the assembly of blunt-ended structures instead of sticky-ended super-structures composed of triple helices whose charged termini are loosely folded and interdigitated with other helices.

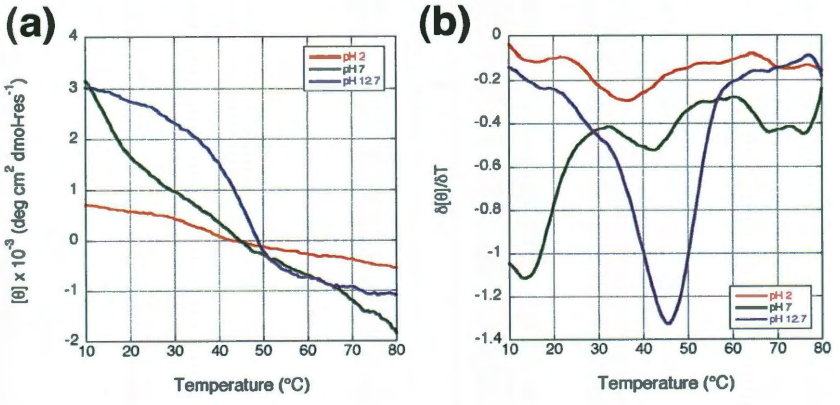

Figure 5.2. CD melting studies for $(PRG)_4(POG)_4(EOG)_4$ at pH 2 (red), 7 (green), and 12.7 (blue). (a) Thermal unfolding curves for samples at given pHs shown as MRE versus temperature. (b) The first derivative of MRE versus temperature, which yields the transition temperature.

5.1.2. $(PRG)_4(POG)_4(EOG)_4$: Nanofiber Formation

Despite the inconsistent results seen in the replication of the triple helical stability for $(PRG)_4(POG)_4(EOG)_4$, analysis on the nanofiber forming ability of the peptide was assessed. Since the reported fibers had an observed D-periodicity, a novel feature for collagen mimetic peptides, the replication of the assembly at the nanofiber level was more desirable than reproducing the triple helical studies described above. Samples made following the Chaikof sample preparation technique in Buffer 8 at peptide concentrations of 0.071% by weight (0.2 mM)[1] and allowed to incubate at both room temperature and at 5 °C precipitated out of solution within a week. Samples prepared in acidic and basic pHs never formed visible aggregates after months of incubation. TEM analysis of Buffer 8 samples using dry techniques revealed the self-assembly of fibers that continued to aggregate into large fibers on the micron scale once the sample showed visible precipitation. Dry-TEM was performed using a uranyl acetate positive stain in order to increase contrast between the peptides and the carbon-coated TEM grids. Full details of TEM sample preparation are given in the experimental section below. Example dry TEM images stained with uranyl acetate are shown in Figure 5.3. Linear fiber morphology can be seen in Figures 5.3a-d that were taken after samples were incubated for seven days. When samples that were allowed to incubate for three weeks so that substantial visible aggregation was seen, TEM images revealed micron-scale fibers like those seen in Figures 5.3e-f. At both time points, amorphous aggregates were seen in dry TEM images, which complemented the images reported by Chaikof *et al.*[1] Nonetheless, in all dry TEM samples examined, D-periodicity was never seen.

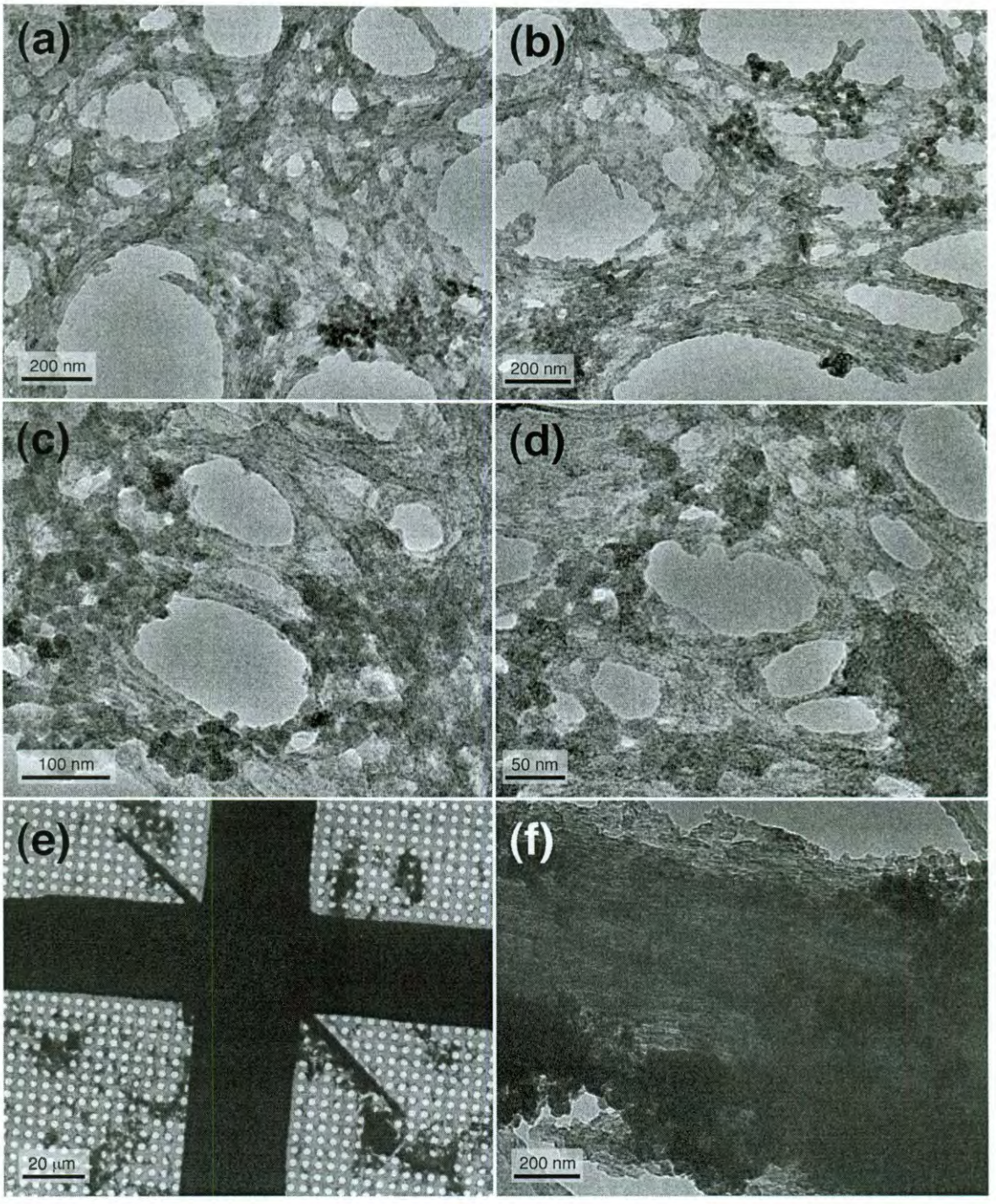

Figure 5.3. Uranyl acetate stained TEM images of $(PRG)_4(POG)_4(EOG)_4$ taken after incubation at room temperature for 7 days (a-d) and 3 weeks (e and f). Images were taken at 12,000X (a, b, f), 30,000X (c), 40,000X (d) and 100X (e) magnification.

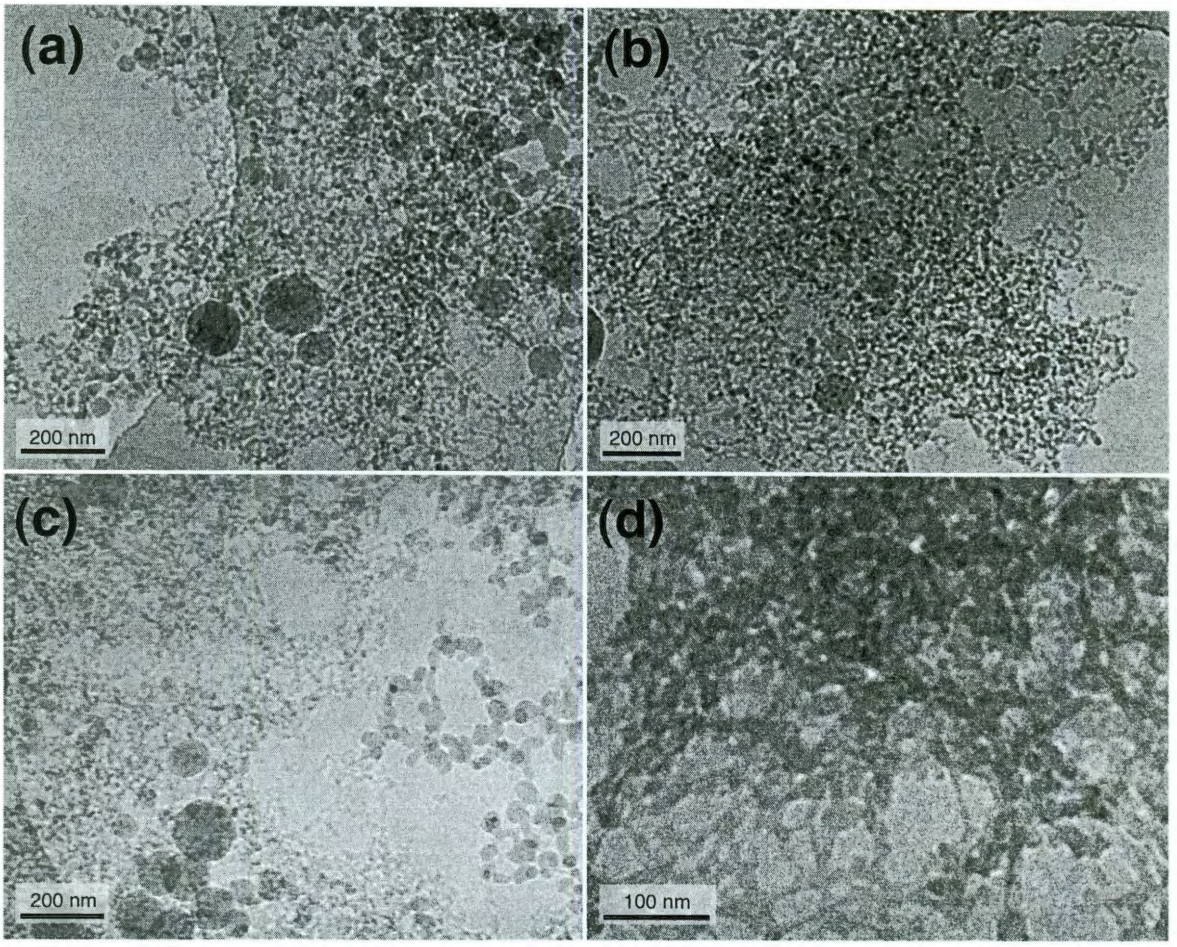

Figure 5.4. Cryo-TEM images of $(PRG)_4(POG)_4(EOG)_4$ taken 2 days after sample preparation. Images have magnifications of 12,000X (a-c) and 30,000X (d).

In order to determine whether the nanofiber morphology was present in the solution state or was simply a drying effect, cryo-TEM was performed on $(PRG)_4(POG)_4(EOG)_4$ even though it was not reported in the publication.[1] Although nanofibers were seen in stained TEM images taken of samples incubated at multiple time points, cryo-TEM images taken of the same samples did not show nanofibers, only small, unorganized aggregates. Figure 5.4 displays vitreous ice cryo-TEM images taken of $(PRG)_4(POG)_4(EOG)_4$ samples incubated for two days. The aggregates seen in these images were very similar to those shown in other collagen mimetic systems such as

$(PKG)_{10}$•$(DOG)_{10}$•$(POG)_{10}$ (Figure 4.1). Due to the absence of organized fibers in cryo-TEM, the solution state structure of $(PRG)_4(POG)_4(EOG)_4$ could not be confirmed. Without this proof, the nanofiber assembly of this peptide seen in dry-TEM could simply be a drying artifact, not assembly based on the peptide design.

5.1.3. Drawbacks of the $(PRG)_4(POG)_4(EOG)_4$ Peptide

The fact that both the reported triple helical thermal stability for $(PRG)_4(POG)_4(EOG)_4$ in Buffer 8 and the D-periodic nanofibers seen in dry-TEM were not reproducible caused the results in the report by Chaikof *et al.* to be questioned.[1] The thermal stability of the peptide at different pH values did not support the proposed mechanism of assembly that hypothesized inter-helical electrostatic interactions. Rather, the results at low pH most closely overlapped the CD spectra reported for $(PRG)_4(POG)_4(EOG)_4$ in Buffer 8, alluding to the fact that helical assembly was blunt-ended. In addition, the dry-TEM images exposed a mixed composition of fibers and amorphous material where the observed fibers were a minor component of the peptide system, not the major species. All samples at physiological pH phase separated as nanofibers were formed instead of forming an organized hydrogel network. For use in tissue engineering applications, a hydrogel structure is preferred. Precipitation was seen in all peptide concentrations and buffers explored. Lastly, Chaikof *et al.* reported the requirement of a narrow window of peptide concentration and buffer composition for fiber formation. When outside this range, the quality of the peptide assembly degraded or failed all together.[1] A more robust system that has a larger window within which assembly can be seen is desirable. Therefore, based on the reported results and those seen

in the replication of $(PRG)_4(POG)_4(EOG)_4$, the designed peptide fell short of success for replicating the natural assembly of collagen.

5.2. Modifications to $(PRG)_4(POG)_4(EOG)_4$

Although $(PRG)_4(POG)_4(EOG)_4$ was not successful in replicating the assembly of natural collagen, we hypothesized that certain modifications to the peptide design might result in a more viable system. Specifically, we explored the importance of the placement of the positively charged region at the N-terminus and the negatively charged region at the C-terminus by reversing this order and assessing the results. Varun Gauba had previously synthesized the peptide $(EOG)_3(POG)_3(PRG)_3$ therefore this peptide was available for analysis and comparison to $(PRG)_4(POG)_4(EOG)_4$. Additionally, the impact that increasing the number of repeats within the regions of the peptide design was questioned so two peptides, also previously synthesized by Gauba, were included in the study: $(DOG)_4(POG)_4(PKG)_4$ and $(DOG)_6(POG)_6(PKG)_6$. Last, the peptide $(PKG)_4(POG)_4(DOG)_4$ was synthesized to complete the peptide library so that the lysine-aspartate interaction could be compared to the arginine-glutamate pairing utilized in $(PRG)_4(POG)_4(EOG)_4$. Through these modifications, we hoped to better understand the Chaikof peptide design and to improve upon its successes.

5.2.1. $(EOG)_3(POG)_3(PRG)_3$

The first modification to the $(PRG)_4(POG)_4(EOG)_4$ sequence that was explored was the reversal of the location of the charged regions within the peptide so that the

positive region was at the C-terminus and the negative region was at the N-terminus: $(EOG)_3(POG)_3(PRG)_3$. By keeping the charged residues the same and only changing the order of the regions within the peptide, the sequence dependence of triple helix stability and nanofiber formation could be examined. Samples were made with a peptide concentration of 0.07% by weight in Buffer 8 following the same protocol as used for $(PRG)_4(POG)_4(EOG)_4$.[1] CD melting studies for the peptide system are shown in Figure 5.5 and the first derivative of the melting curve revealed a single transition at 27 °C. The T_m for $(EOG)_3(POG)_3(PRG)_3$ was higher than the major transition seen for $(PRG)_4(POG)_4(EOG)_4$ in the replication studies, but was lower than the melting temperature reported by Chaikof.[1]

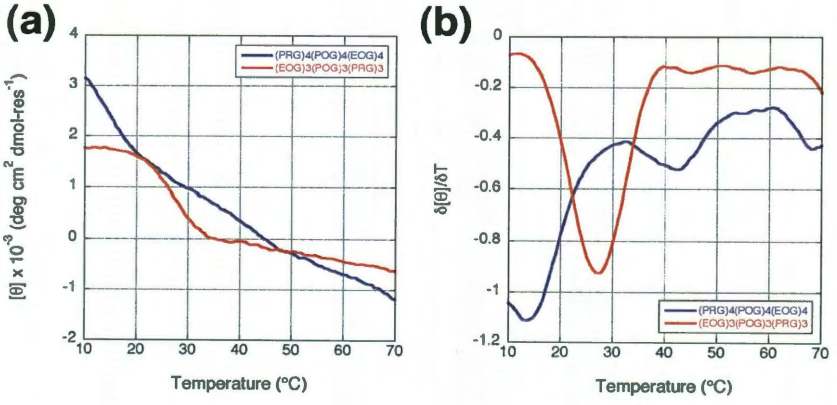

Figure 5.5. CD thermal unfolding curves for $(PRG)_4(POG)_4(EOG)_4$ and $(EOG)_3(POG)_3(PRG)_3$ in Buffer 8 shown in blue and red respectively. (a) Melting studies shown as MRE versus temperature and (b) the first derivative of the melting curve versus temperature.

Once the ability of $(EOG)_3(POG)_3(PRG)_3$ to form triple helices was confirmed, the peptide system was examined using TEM to assess the nanomorphology of the designed peptide. In uranyl acetate stained dry-TEM, peptide aggregates were the only

visible morphologies seen (Figures 5.6a and 5.6b). This contrasted with the results seen for $(PRG)_4(POG)_4(EOG)_4$ in which organized fibers were present in dry-TEM images. Although these nanofibers were the minor component in the images with amorphous aggregates as the major component, linear fibers were visible and reproducible in TEM images taken at multiple time points. When $(EOG)_3(POG)_3(PRG)_3$ was analyzed in cryo-TEM, no clear peptide organization was visible (Figure 5.6c). In the image, large ethane artifacts were seen, which resulted from sample preparation (see cryo-TEM sample preparation for details). Beneath the ethane, small clusters of peptide aggregates were visible. Therefore in the solution state, $(EOG)_3(POG)_3(PRG)_3$ behaved similarly to $(PRG)_4(POG)_4(EOG)_4$: neither peptide forms organized nanofibers.

Finally, $(EOG)_3(POG)_3(PRG)_3$ was examined in phosphate and water, not just in Buffer 8. In these buffers, peptide precipitation occurred more rapidly than in the higher ionic strength Buffer 8, in conjunction with work on $(PKG)_4(POG)_4(DOG)_4$ described later in this chapter. Specifically, a sample made with a peptide concentration of 0.2% by weight in 10 mM phosphate buffer, pH 7, precipitated out of solution overnight. The triple helical nature of the peptide within these buffers was not assessed due to this rapid precipitation. Based on the results for the peptide system, the reversal of the charged domains within the arginine and glutamate containing peptide sequence did not provide a substantial improvement in triple helical and nanofiber properties compared to $(PRG)_4(POG)_4(EOG)_4$.

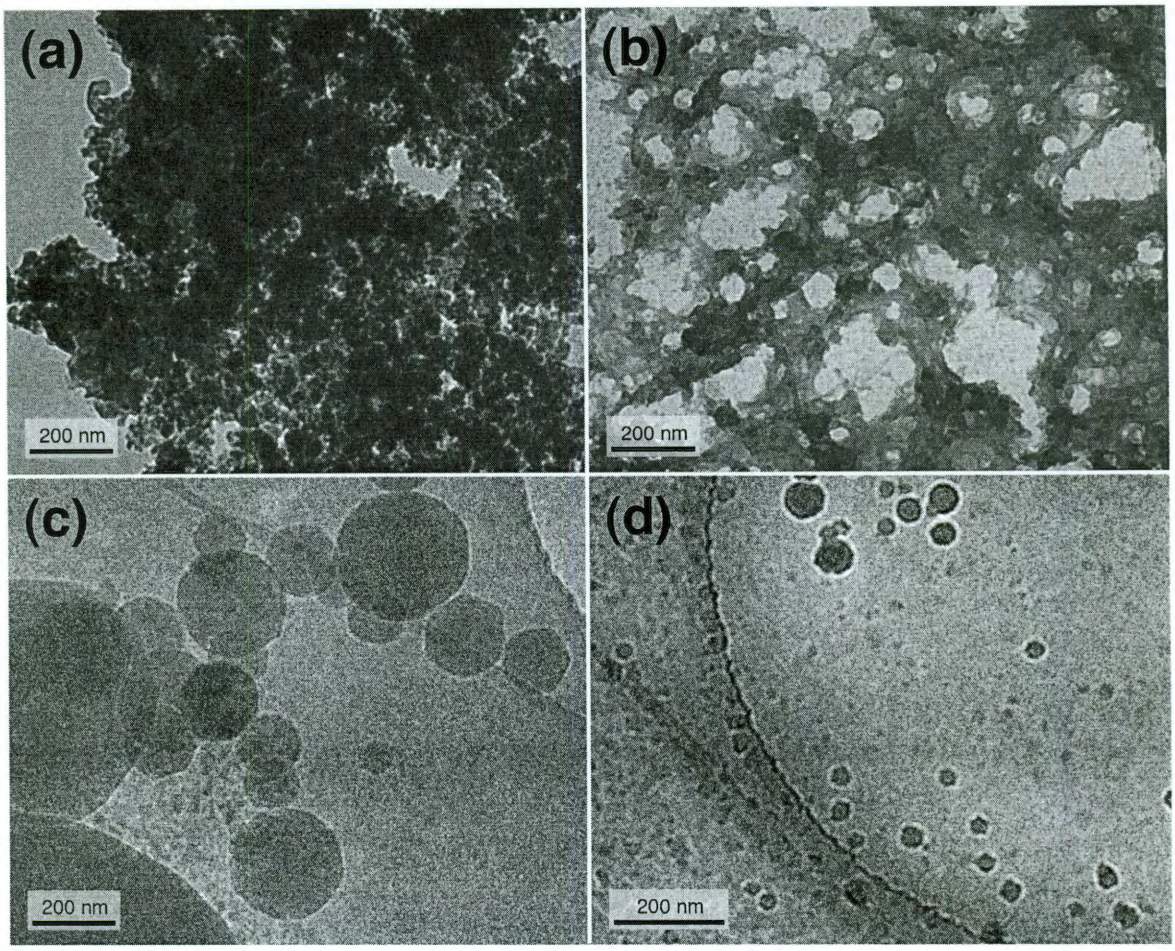

Figure 5.6. TEM images of $(EOG)_3(POG)_3(PRG)_3$ prepared in Buffer 8. (a and b) Uranyl acetate stained dry-TEM images taken at a magnification of 12,000X after (a) 3 weeks of incubation and (b) 9 days of incubation. (c and d) Cryo-TEM images of the peptide taken after 7 days of incubation at magnifications of (c) 12,000X and (d) 15,000X.

5.2.2. $(DOG)_4(POG)_4(PKG)_4$ and $(DOG)_6(POG)_6(PKG)_6$

Since $(EOG)_3(POG)_3(PRG)_3$ was examined without successful improvement over $(PRG)_4(POG)_4(EOG)_4$, the next adjustment to the Chaikof peptide sequence that was explored in an attempt to improve the peptide properties was the replacement of arginine and glutamate with lysine and aspartate respectively. Since the peptides $(DOG)_4(POG)_4(PKG)_4$ and $(DOG)_6(POG)_6(PKG)_6$ were previously synthesized by Varun

Gauba, they were examined first before synthesizing any new lysine and aspartate containing peptides. The CD thermal unfolding studies for both peptides as well as $(PRG)_4(POG)_4(EOG)_4$ in Buffer 8 are shown in Figure 5.7. $(DOG)_4(POG)_4(PKG)_4$ had a melting temperature of 25 °C and $(DOG)_6(POG)_6(PKG)_6$ had a T_m of 40 °C. At first glance, it was easy to see that the additional peptide triplets present in the latter peptide facilitated triple helical folding resulting in a higher thermal stability seen for $(DOG)_6(POG)_6(PKG)_6$ compared to $(DOG)_4(POG)_4(PKG)_4$. When the melting temperatures were contrasted to that of $(PRG)_4(POG)_4(EOG)_4$ (black curve in Figure 5.7), both lysine and aspartate-containing peptides showed higher thermal stabilities than the major transition seen in CD for the arginine and glutamate-containing peptide. Based on the amino acid propensity for triple helical formation alone, we expected $(PRG)_4(POG)_4(EOG)_4$ to have a higher stability than either $(DOG)_6(POG)_6(PKG)_6$.[2] In order to understand these results, we turned to the homotrimer stabilities discussed in Chapter 3 for AAB heterotrimer formation. In this study, $(DOG)_{10}$ formed a homotrimer in 10 mM Tris buffer (tris(hydroxymethyl)-aminomethane) visible in CD melting experiments. The ability of this highly charged peptide to form a triple helix in the low ionic strength buffer was attributed to the ability of the cationic Tris buffer to form a specific interaction with the negatively charged aspartate residues thus stabilizing the homotrimer. One of the component of Buffer 8 is TES, a sulfonic acid derivative of Tris. Therefore, it was possible that the TES in Buffer 8 stabilized the aspartate region of $(DOG)_4(POG)_4(PKG)_4$ and $(DOG)_6(POG)_6(PKG)_6$ resulting in homotrimer stabilities slightly higher than expected.

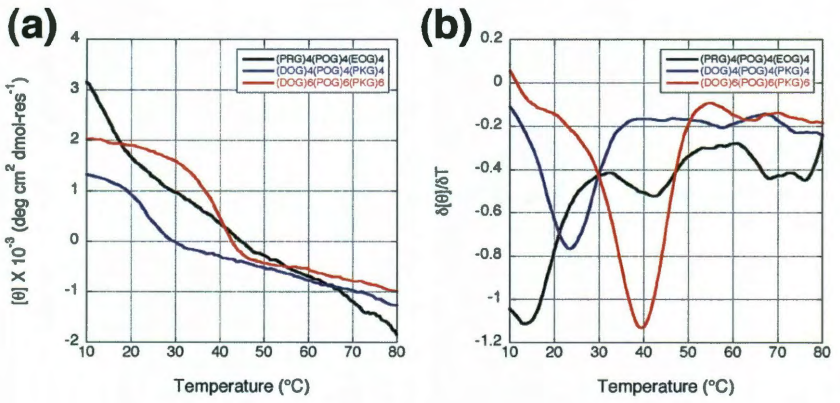

Figure 5.7. CD melting curves for $(PRG)_4(POG)_4(EOG)_4$, $(DOG)_4(POG)_4(PKG)_4$ and $(DOG)_6(POG)_6(PKG)_6$ shown in black, blue and red respectively. (a) Thermal unfolding studies shown as MRE versus temperature and (b) the first derivative of the melting curve versus temperature.

The next step of analysis for $(DOG)_4(POG)_4(PKG)_4$ and $(DOG)_6(POG)_6(PKG)_6$ was TEM imaging of possible nanostructures formed by each peptide. In uranyl acetate stained dry-TEM (Figure 5.8a), $(DOG)_4(POG)_4(PKG)_4$ showed peptide aggregates in all samples. When the peptide was analyzed in cryo-TEM (Figure 5.8b), small amorphous peptide networks could be seen that resembled the morphologies seen for $(PRG)_4(POG)_4(EOG)_4$. The TEM images for $(DOG)_6(POG)_6(PKG)_6$ resulted in a slightly different story. Figures 5.8c and 5.8d are example uranyl acetate stained dry-TEM images for this peptide. Cryo-TEM images were never taken. In Figure 5.8c, amorphous peptide aggregates were seen which assembled into large fibers such as the one shown in Figure 5.8d. Within the peptide library explored in Chapter 5, $(DOG)_6(POG)_6(PKG)_6$ was the only peptide other than $(PRG)_4(POG)_4(EOG)_4$ that assembled into large nanofibers in Buffer 8, similar to those seen for the hydrophobic nanofibers designed in Chapter 4. The other peptides only formed small peptide aggregates.

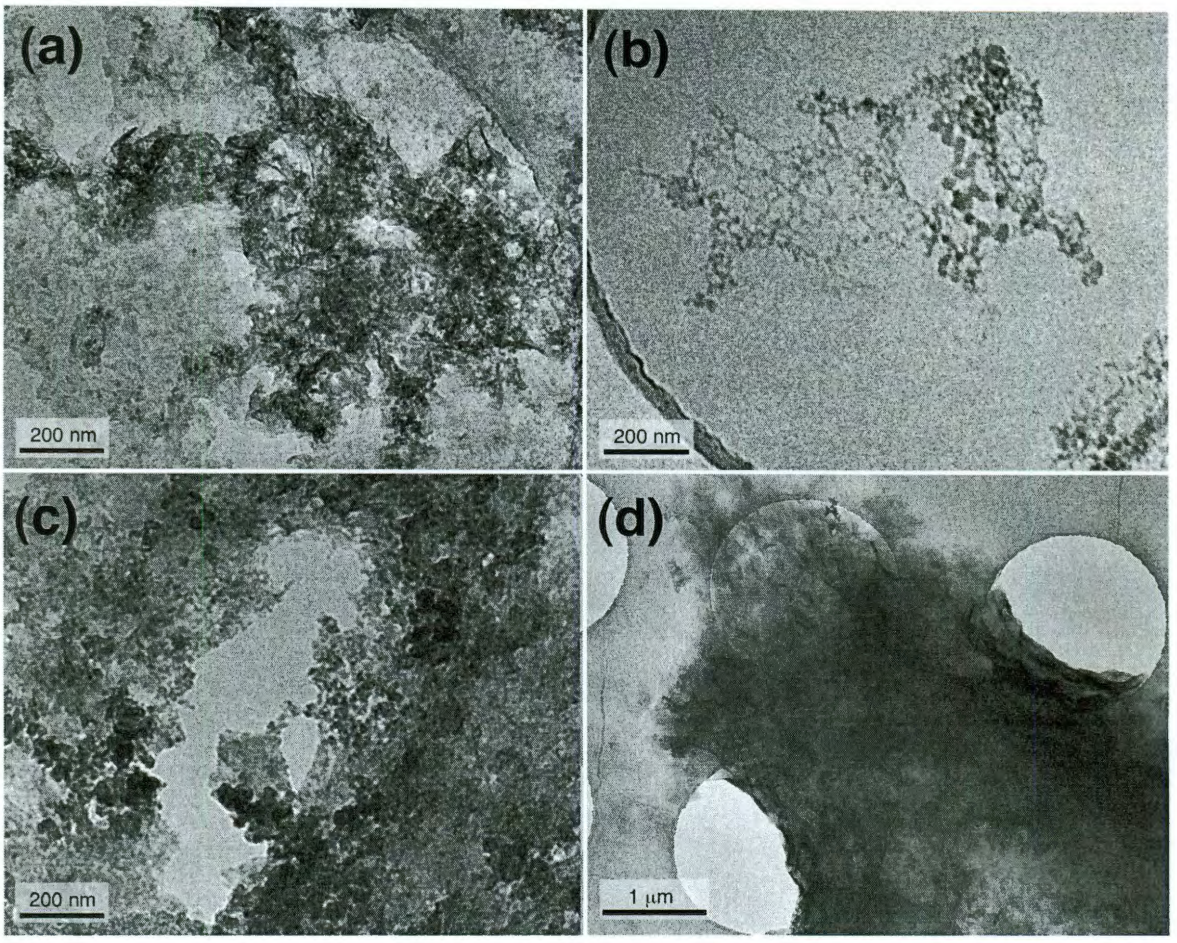

Figure 5.8. TEM images of (a and b) $(DOG)_4(POG)_4(PKG)_4$ and (c and d) $(DOG)_6(POG)_6(PKG)_6$ prepared in Buffer 8. (a) Uranyl acetate stained dry-TEM image of $(DOG)_4(POG)_4(PKG)_4$ taken at a magnification of 12,000X. (b) Cryo-TEM image of $(DOG)_4(POG)_4(PKG)_4$ taken at a magnification of 12,000X. (c and d) Uranyl acetate stained dry-TEM images of $(DOG)_6(POG)_6(PKG)_6$ taken at (c) 12,000X and (d) 4,000X magnifications.

Despite any nanostructures seen for the peptides, $(DOG)_4(POG)_4(PKG)_4$ and $(DOG)_6(POG)_6(PKG)_6$ precipitated out of solution over time. For the latter peptide, precipitation occurred within hours depending on the peptide concentration and for the former peptide, precipitation occurred within the week scale. Such precipitation was seen in other buffer systems, not just Buffer 8. In deionized water and 10 mM sodium phosphate, both pH 7, the peptides precipitated out of solution along the same time scales

as for Buffer 8. Therefore these peptides did not serve as improvements to the $(PRG)_4(POG)_4(EOG)_4$ peptide design despite the presence of lysine and aspartate amino acids instead of arginine and glutamate.

5.2.3. $(PKG)_4(POG)_4(DOG)_4$ Assembly Based on Chaikof Method

The last peptide examined within the $(PRG)_4(POG)_4(EOG)_4$ peptide library was $(PKG)_4(POG)_4(DOG)_4$, which was synthesized to directly complement the former peptide with the only difference being the identity of the charged residues used. Synthesis and purification details are given in the experimental section below and a representative mass spectrum and HPLC chromatogram are shown in Appendix 2. In Buffer 8, CD melting experiments performed on $(PKG)_4(POG)_4(DOG)_4$ showed a single transition at 17 °C (Figure 5.9) This T_m was very similar to the melting temperature recorded for $(PRG)_4(POG)_4(EOG)_4$ in the replication studies described above however, the intensity of the transition was much lower than that for $(PRG)_4(POG)_4(EOG)_4$ in Buffer 8.

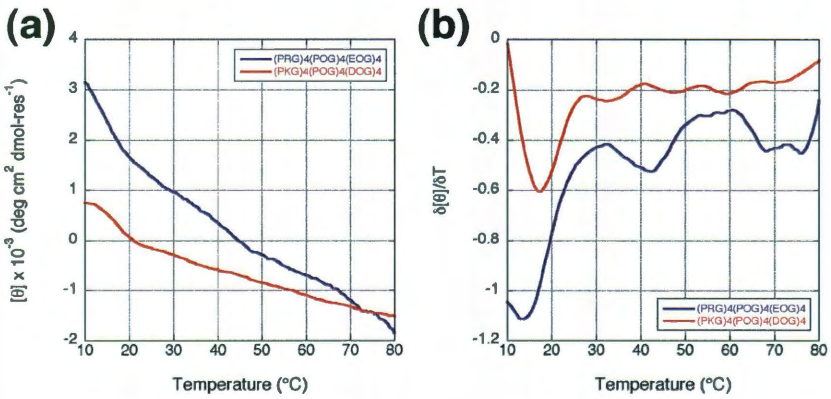

Figure 5.9. CD thermal unfolding studies for $(PRG)_4(POG)_4(EOG)_4$ shown in blue and $(PKG)_4(POG)_4(DOG)_4$ displayed in red in Buffer 8. (a) Melting curves shown as MRE versus temperature and (b) the first derivative of the melting curves versus temperature.

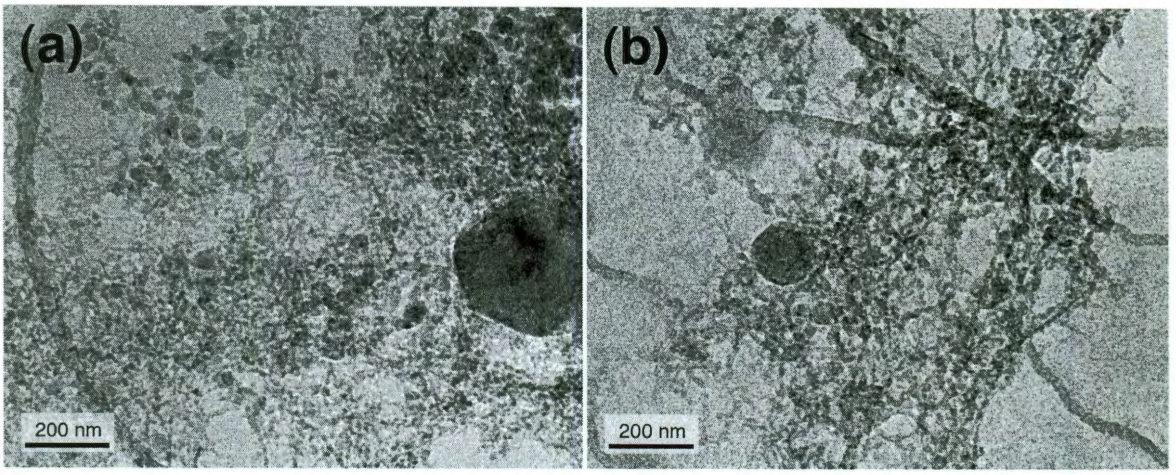

Figure 5.10. Cryo-TEM images of $(PKG)_4(POG)_4(DOG)_4$ in Buffer 8 taken at 12,000X magnification.

When the nanomorphology of $(PKG)_4(POG)_4(DOG)_4$ in Buffer 8 was analyzed by cryo-TEM, the resulting images were very similar to those taken for $(PRG)_4(POG)_4(EOG)_4$: small populations of amorphous peptide networks. Representative cryo-TEM images for $(PKG)_4(POG)_4(DOG)_4$ are given in Figure 5.10. Based on the triple helical stability and solution state morphology of $(PKG)_4(POG)_4(DOG)_4$ compared to $(PRG)_4(POG)_4(EOG)_4$ in Buffer 8, the two peptides appeared to be very similar. However, one major observation was made for the latter peptide that was never seen for the former: sample precipitation. Depending on the peptide concentration, $(PRG)_4(POG)_4(EOG)_4$ precipitated out of solution within hours or, at the most, over the span of a week. In contrast, $(PKG)_4(POG)_4(DOG)_4$ never showed visible precipitation, even after months of incubation. In order to understand this, further analysis of this peptide was performed in different buffer systems in order to decipher the assembly of the peptide and understand why it did not precipitate out of solution when the other four peptides within the library did. During these experiments, a sample of $(PKG)_4(POG)_4(DOG)_4$ at a concentration of 2 mM dissolved in water formed a hydrogel

overnight. Based on this result, extensive analysis of the peptide was performed and will be described in detail below.

5.3. Hierarchical Self-Assembly of $(PKG)_4(POG)_4(DOG)_4$

Once observation of gelation of the $(PKG)_4(POG)_4(DOG)_4$ was seen in water at a concentration of 2 mM, samples were made at specified concentrations between 0.2% (0.6 mM) and 1.0% (3 mM) by weight.[3] Multiple buffer systems were explored with varying ionic strengths (see Assembly in Additional Buffers section below) however, discussion will begin with 10 mM sodium phosphate buffer at pH 7 (referred to as phosphate) due to that fact that the majority of analysis was performed in this buffer. In phosphate, all samples made at concentrations of 0.5% (1.5 mM) by weight or higher formed hydrogels within a few hours.[3] In order to thoroughly analyze the assembly of this peptide, we systematically characterized the peptide at each level of self-assembly: triple helix, nanofiber and hydrogel.

5.3.1. $(PKG)_4(POG)_4(DOG)_4$: Triple Helical Stability

As mentioned previously, in order to determine whether a collagen mimetic peptide forms a triple helix, two circular dichroism (CD) experiments must be performed: a wavelength spectrum and a thermal unfolding curve. For the peptide $(PKG)_4(POG)_4(DOG)_4$, CD spectra taken at all concentrations showed a strong maximum at 225 nm.[3] The spectra for 0.2%, 0.5% and 1.0% by weight in phosphate are shown in Figure 5.11a. Note the size of the maximum in the spectrum increased as the peptide

concentration increased despite the fact that the data was normalized for concentration.[3] This indicated that an increased percentage of the peptide was folded at higher concentrations. The spectrum for 1.0% by weight is only shown from 250 – 205 nm due to the increase in background noise for higher concentration samples at lower wavelengths. When CD thermal unfolding experiments were performed from 5 to 85 °C on samples at 0.2%, 0.5% and 1.0% by weight concentrations, all samples exhibited a cooperative transition in the melting profile.[3] Additionally, transitions for samples at higher peptide concentrations were stronger and more clear than those for lower concentrations, again indicating that the higher concentration of peptide helped to drive triple helix formation. The thermal unfolding curve and the first derivative of the curve for 0.2%, 0.5% and 1.0% by weight samples in phosphate are shown in Figure 5.11b and 5.11c respectively. A major transition was seen in the first derivative curve for all three concentrations at 40 – 41 °C, corresponding to the melting temperature for the peptide.[3] However, a broad, minor transition was also visible between 10 and 30 °C in the samples with peptide concentrations of 0.5% and 1.0% by weight. The minor transition may be due to increased helicity upon fiber elongation and lateral packing since it was is more pronounced in samples that formed visible hydrogels.[3] However, a more detailed explanation for this will be given in the Hydrogel section below.

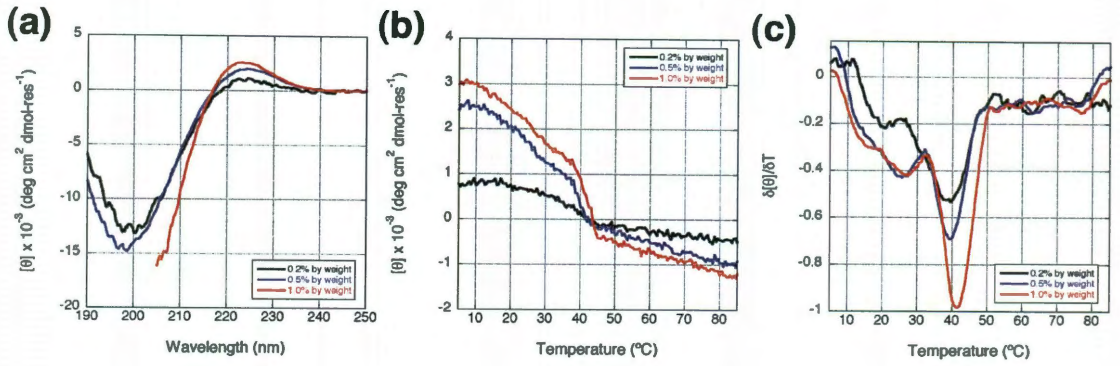

Figure 5.11. CD spectra highlighting the triple helical nature of $(PKG)_4(POG)_4(DOG)_4$ in phosphate at peptide concentrations of 0.2% (black), 0.5% (blue) and 1.0% (red) by weight. Adapted from Figures 3 and S3 in reference 2. (a) CD spectrum shown as molar residual ellipticity (MRE) vs. wavelength, (b) thermal unfolding curves shown as MRE vs. temperature and (c) first derivative of MRE vs. temperature.[3]

5.3.2. $(PKG)_4(POG)_4(DOG)_4$: Nanofiber Formation

Once the triple helical nature of the peptide in phosphate was confirmed, the next step was to understand the nanostructure of the self-assembled peptide. Multiple microscopy techniques were used including transmission electron microscopy (TEM), atomic force microscopy (AFM) and scanning electron microscopy (SEM). TEM is an integral technique for viewing the morphology and measuring the length and width of structures on the nanoscale. It is most commonly a dry technique that, for viewing carbon-based materials, requires the sample to be stained with a heavy metal such as phosphotungstic acid (PTA). For this peptide, a 1.0% by weight concentration sample in phosphate was prepared and negatively stained with PTA. Images of these stained samples (Figure 5.12) revealed long nanofibers present both as single fibers and as fiber bundles.[3] The dry-TEM images exhibited a variety of fiber widths present within this system when dried and stained. These fibers were the major species within the TEM

sample, in contrast to previously reported collagen mimetic nanofibers, such as $(PRG)_4(POG)_4(EOG)_4$, that showed large aggregates and a variety of other non-fibrous structures in the TEM images.[1,3] Figure 5.12b revealed the twisting nature of some of the nanofibers in contrast to fibers with a smoother morphology.[3] Although the negatively stained TEM images showed the presence of nanofibers for this peptide system, drying artifacts can cause samples to appear more densely packed or with a completely different structure than what is present in the hydrated state, as was seen for $(PRG)_4(POG)_4(EOG)_4$ described above. In addition, the use of a heavy metal stain added an additional level of uncertainty in assessing fiber size and morphology. For these reasons, imaging the system in a hydrated environment using vitreous ice cryo-TEM is the only way to prove the presence of nanofibers in a solution state.

The sample preparation for cryo-TEM differed greatly from dry TEM due to the fact that cryo-TEM required a thin aqueous film of sample on the TEM grid before it was flash frozen in ethane slush. Representative TEM images from this preparation are given in Figure 5.13. In contrast to the dry TEM images, the fibers seen in the vitreous ice cryo-TEM sample had uniform widths from 4 – 5 nm and fiber lengths from several hundred nanometers to many microns.[3] However, similar to the dry TEM images, the observed fibers in cryo-TEM were the majority of the peptide population in the sample. (The spherical species seen in the cryo-TEM image were ethane artifacts that resulted from sample preparation, not peptide aggregates). Therefore, in both dry and cryo-TEM, the presence of nanofibers was confirmed and they were observed to be the major species within the system.[3]

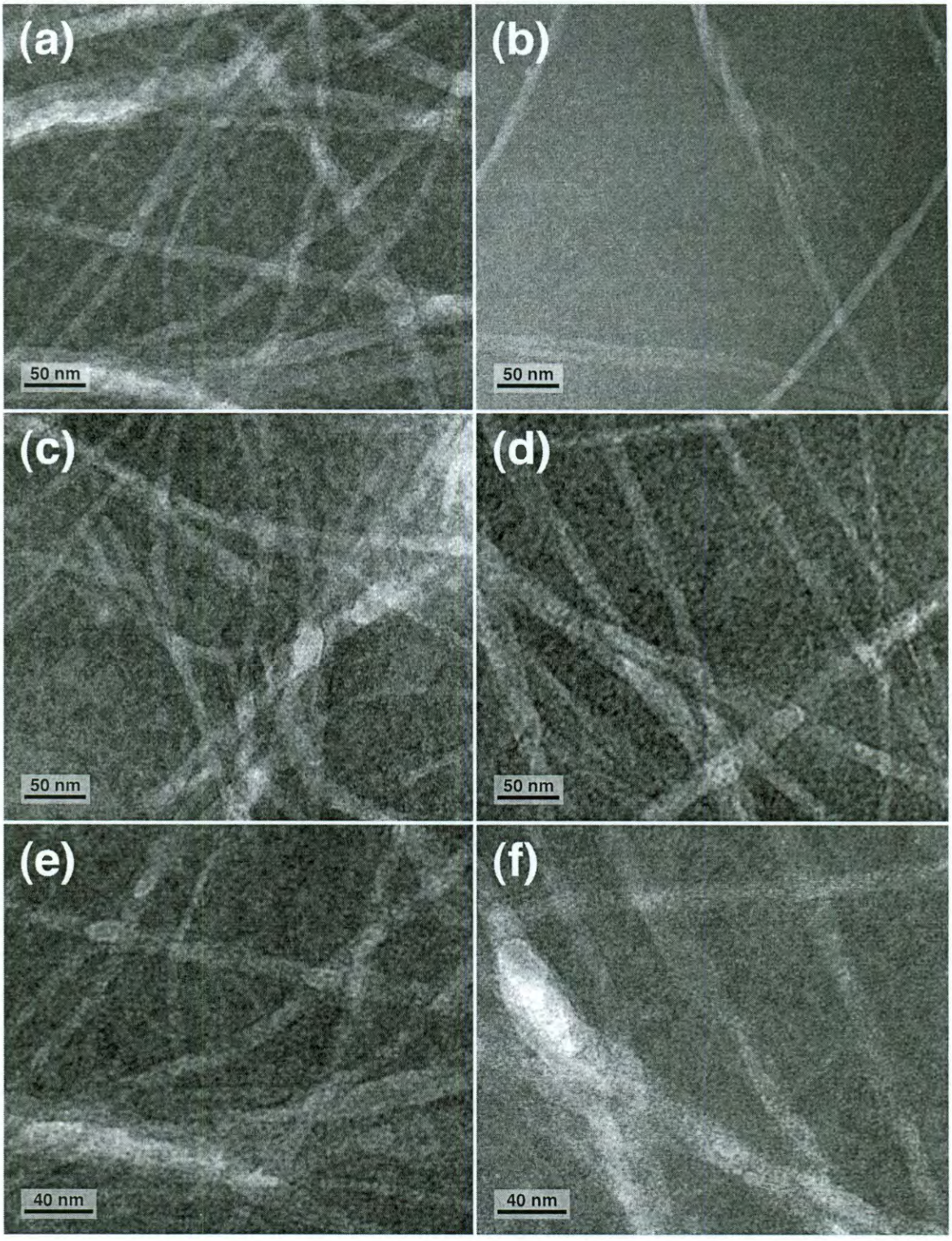

Figure 5.12. Negatively stained TEM images of $(PKG)_4(POG)_4(DOG)_4$ at a concentration of 1.0% by weight in phosphate buffer, pH 7, that were stained using 2.0% by weight PTA, pH 6. Adapted from Figures 4 and S4 in reference 2. Magnifications shown are 40,000X (a-d) and 60,000X (e and f).[3]

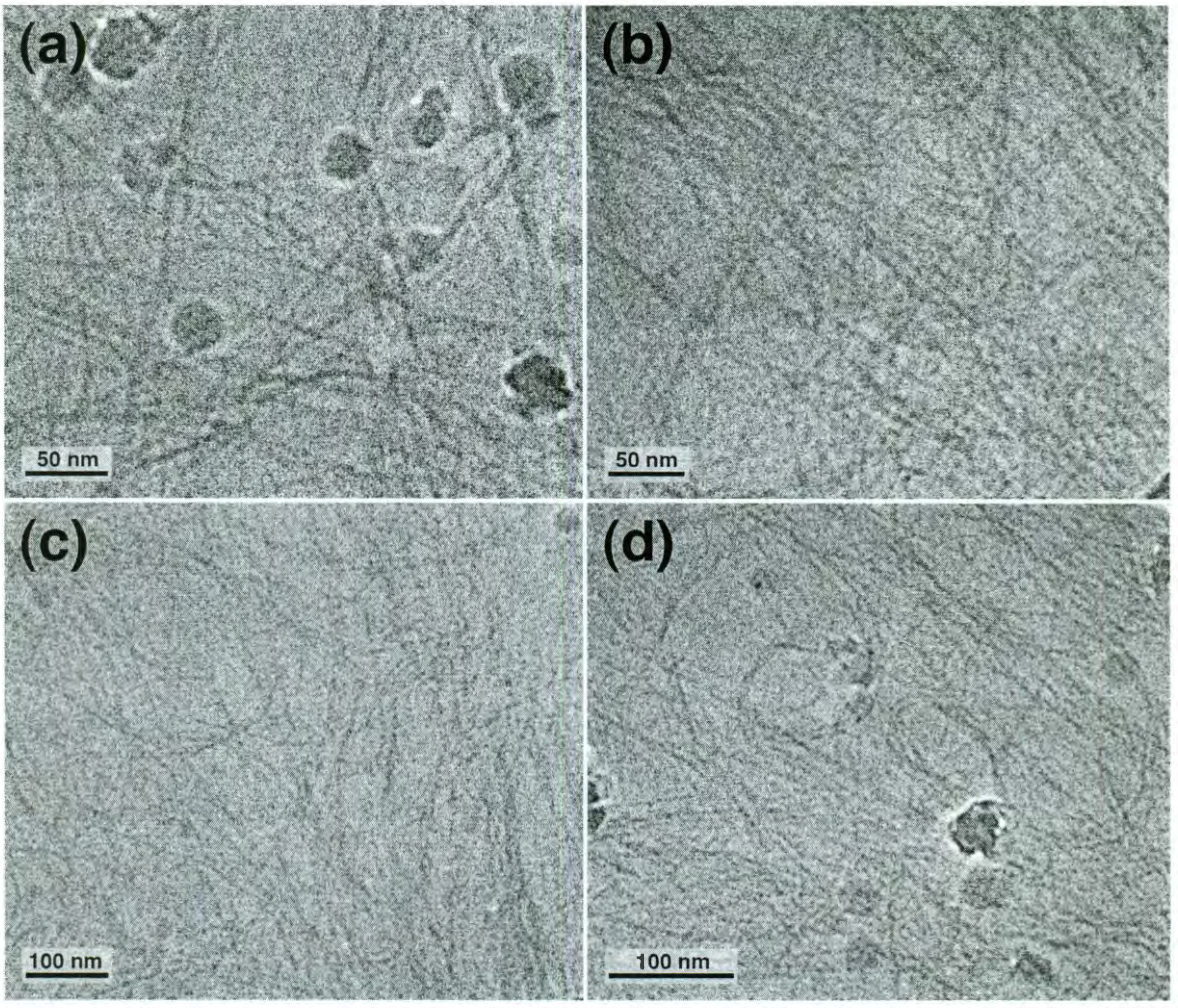

Figure 5.13. Vitreous ice cryo-TEM images of $(PKG)_4(POG)_4(DOG)_4$ taken in phosphate buffer at a concentration of 0.25% by weight that was diluted from a 0.5% by weight sample. Adapted from Figures 4 and S4 in reference 2. Magnifications shown are 40,000X (a and b), 20,000X (c) and 30,000X (d).[3]

Once the length and width of the nanofibers formed from the peptide was determined from TEM, the height of the fibers was needed in order to understand the mechanism of fiber formation. Tapping mode AFM is the most efficient method for acquiring this data. Figure 5.14 displays AFM images taken of 1.0% and 0.5% by weight samples, respectively, in phosphate buffer.[3] Nanofibers were seen in both images with the higher concentration sample exhibiting a thicker network of nanofibers. The measured

height profile in phosphate buffer from the AFM images was 1.2 ± 0.3 nm.[3] This value was much lower than the fiber width of 4 – 5 nm measured from TEM and the observed fiber lengths seen in AFM also appear smaller than those seen in TEM. A hypothesis for this difference will be discussed below in the Proposed Mechanism of Assembly section. One advantage of these images was that due to their lower magnification, a larger area was observable and the uniformity of the population of self-assembled nanofibers was more apparent than in TEM images.

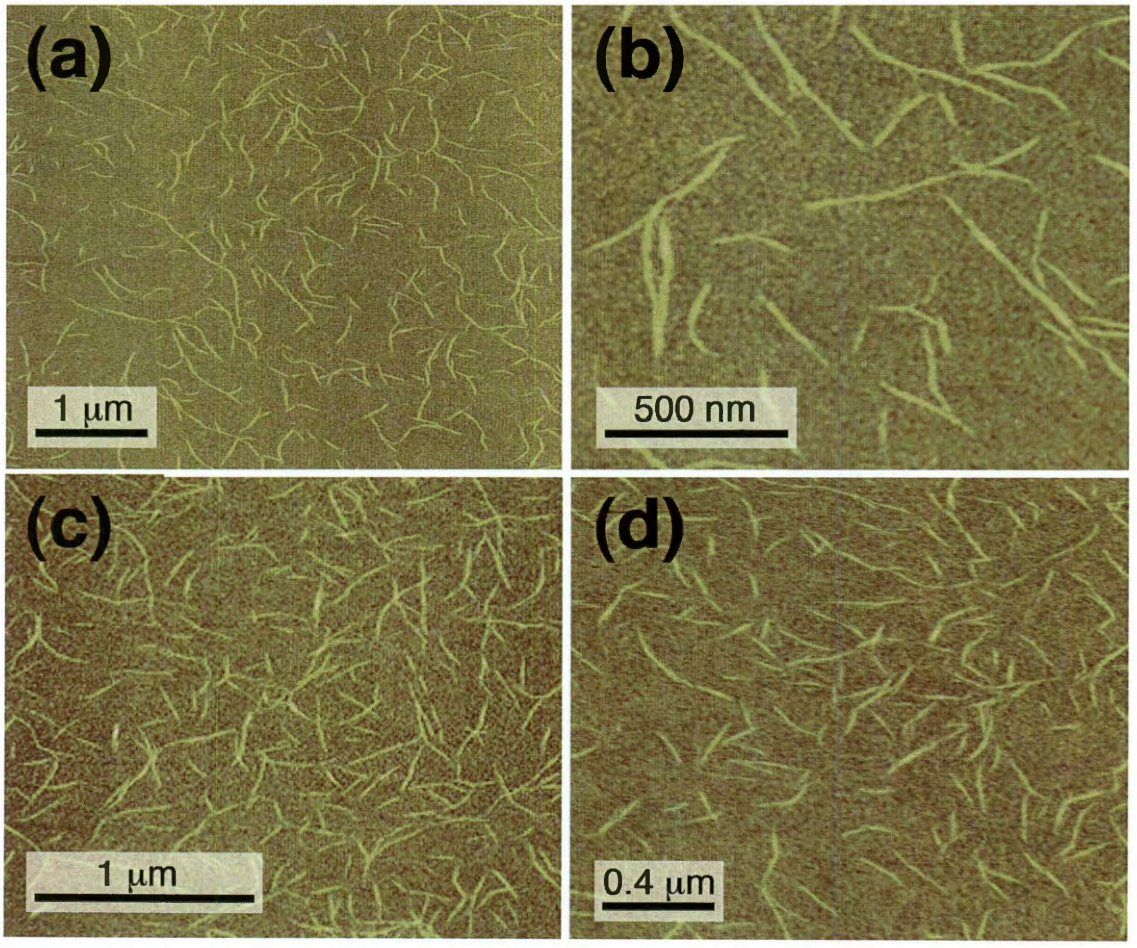

Figure 5.14. AFM images of $(PKG)_4(POG)_4(DOG)_4$ in phosphate buffer, pH 7, as observed after spin coating onto freshly cleaved mica at concentrations of 0.5% (a and b) and 1.0% (c and d) by weight. Adapted from Figures 4 and S5 in reference 2.[3]

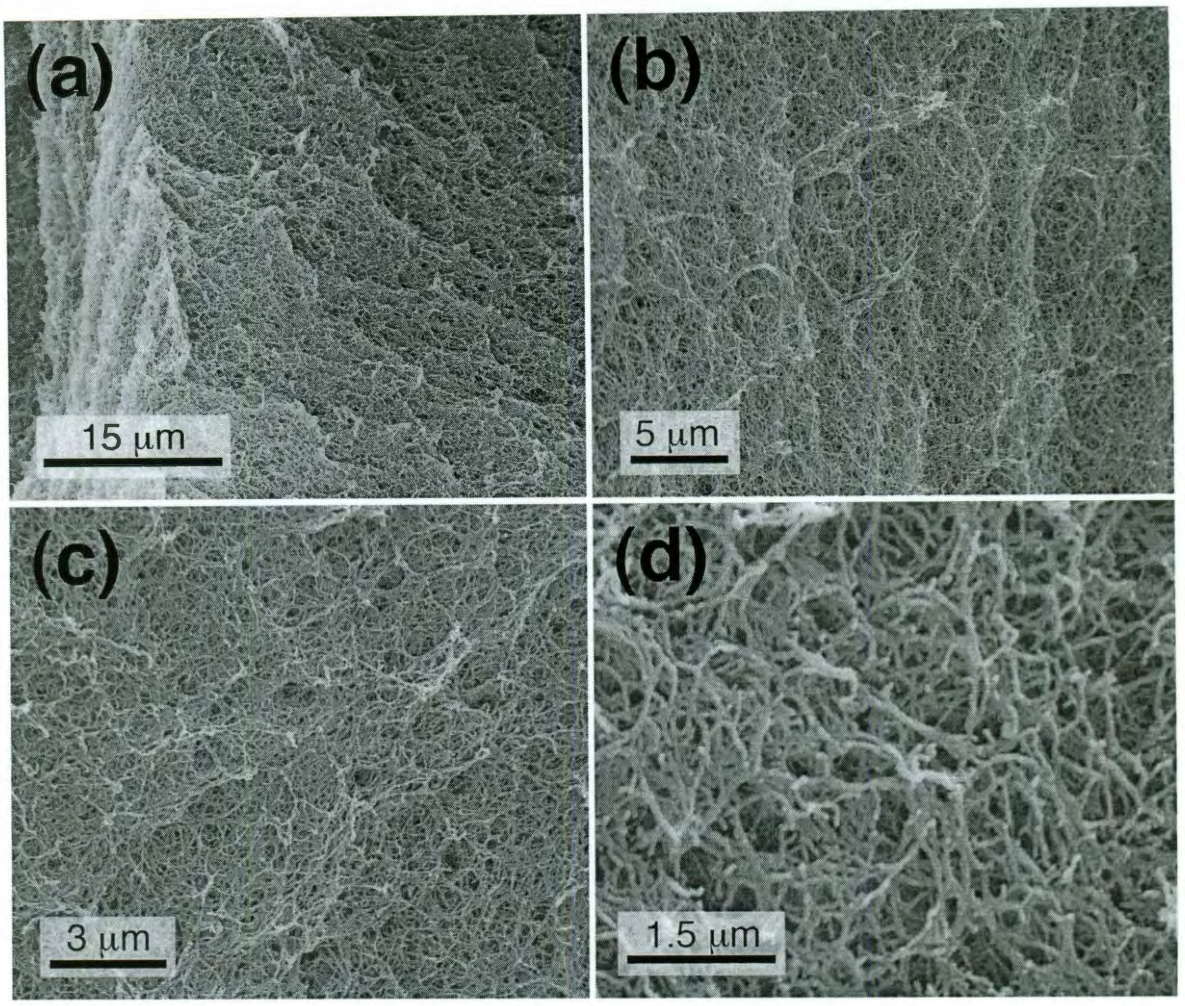

Figure 5.15. SEM images of a critical point dried $(PKG)_4(POG)_4(DOG)_4$ hydrogel with a peptide concentration of 1.0% by weight. Adapted from Figures 4 and S6 in reference 2. Magnifications shown are 3,100X (a), 5,000X (b), 10,000X (c) and 30,000X (d).[3]

One final microscopy method, SEM, is important in understanding the qualitative long-range nanoscale behavior of the system. Samples imaged by SEM had a peptide concentration of 1.0% by weight in phosphate buffer (for more details refer to the experimental section below). SEM images at varying magnifications are shown in Figure 5.15. At lower magnifications (Figure 5.15a) a dense fiber network that was homogeneous and extended tens of microns was apparent.[3] When the magnification was increased, the uniform nature of the nanofibers within the network could be more readily

seen. These results directly complemented the fiber morphologies observed by TEM and AFM and also gave an indication of the three-dimensional structure of the hydrogel.[3]

Through the use of multiple microscopy techniques, the nano-morphology of $(PKG)_4(POG)_4(DOG)_4$ in phosphate was determined to be nanofibers of relatively uniform dimensions with observed lengths of at least several hundred nanometers, widths of 4 – 5 nm, measured heights of 1.2 ± 0.3 nm and a uniform long-range behavior visible in the hydrated state.[3]

5.3.3. $(PKG)_4(POG)_4(DOG)_4$: Hydrogel Properties

With the first two levels of self-assembly confirmed, the final layer of analysis needed to describe the multi-hierarchical assembly of $(PKG)_4(POG)_4(DOG)_4$ in phosphate was the assessment of the visco-elastic properties of the formed hydrogel. Visually, the gels maintained their shape when they were removed from their containers, including the visible sustainability of the gel's sharp edges.[3] The image in Figure 5.16c depicts the visual properties of the hydrogel. To quantitatively analyze the peptide hydrogels, rheological studies were performed that measure the storage modulus (G') and loss modulus (G"), which correspond to the elastically stored energy and energy lost as heat within the hydrogel respectively. A sample is considered to be a hydrogel if the following properties are upheld: (1) the G' is greater than the G" over a linear region and (2) after a certain level of strain is applied to the sample, the gel breaks down so that the G" exceeds the G'. In order to assess the time required for complete gelation, a time course rheological study was run at 25 °C on a freshly annealed $(PKG)_4(POG)_4(DOG)_4$ sample with a concentration of 1.0% by weight.[3] The sample was annealed, immediately

placed on the rheometer stage with a humidity chamber present and the experiment was begun. Assembly was deemed complete when the G' leveled off. Figures 5.16a and 5.16b depict the time course rheological studies. Initial gel assembly (Figure 5.16a) showed a small discontinuity at approximately 0.4 hours, which was due to the addition of mineral oil to prevent the hydrogel from dehydrating during the prolonged measurement. When the gelation was monitored for a longer period of time (Figure 5.16b), gelation was deemed complete after 8 hours. Therefore, peptide assembly for $(PKG)_4(POG)_4(DOG)_4$ from unfolded peptide chains to an organized hydrogel network occurred within 8 hours.[3]

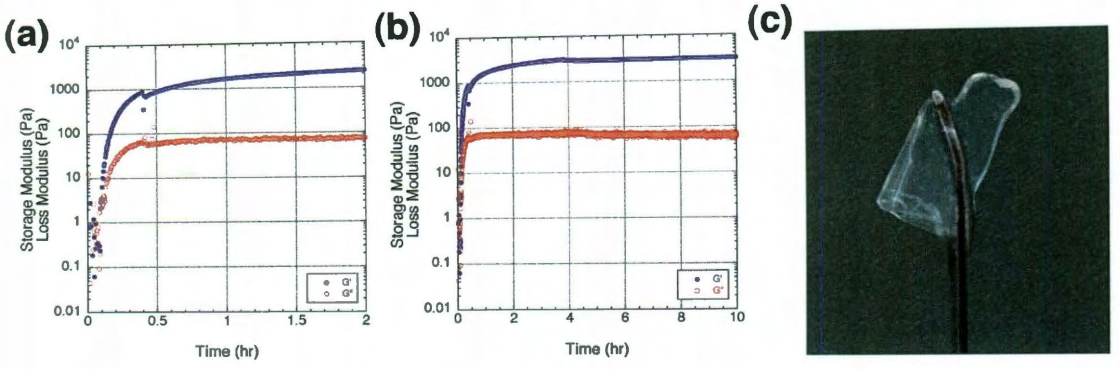

Figure 5.16. Rheological studies used to assess the time necessary for $(PKG)_4(POG)_4(DOG)_4$ to completely assemble in phosphate buffer at a concentration of 1.0% by weight. Adapted from Figures 5 and S7 in reference 2. (a) Initial and (b) complete gel assembly seen via a time course rheological study run at 25 °C, 1 rad/s and 0.1% strain. (c) Photo of the shape-persistent nature of the gel with a concentration of 1.0% by weight in phosphate. The gel was prepared at a volume of 0.5 ml. Note the sustainability of the sharp gel edges.[3]

Strain and frequency sweep experiments were performed to assess the gel properties and specifically, the storage modulus (G') and loss modulus (G") which measure the elastically stored energy and energy lost as heat within the hydrogel respectively. Representative graphs of each type of experiment are shown in Figures

5.17a and 5.17b respectively and the first observation that can be made is that the G' was substantially larger than the G" for both 0.5% and 1.0% by weight concentrations of the peptide in phosphate buffer.[3] Therefore, $(PKG)_4(POG)_4(DOG)_4$ formed a hydrogel in phosphate buffer at 0.5% by weight concentrations and higher. It should be noted that the observed G' of this collagen mimetic system was similar to that typically observed for a collagen hydrogel formed from natural sources, such as rat tail collagen, despite the fact that our peptide is approximately thirty times shorter (36 amino acids as compared to 1,000).[4] It was also higher than Matrigel[5] and comparable to popular β-sheet hydrogels described in the literature.[6-11]

The collagen mimetic hydrogel was found to be temperature sensitive. From the CD melting studies, we know that the triple helix unfolded at 40 – 41 °C therefore a temperature ramp rheological experiment from 20 to 60 °C with a ramp of 0.5 °C/min was used to demonstrate the melting of the hydrogel.[3] Figure 5.17c displays the temperature ramp rheological study run on a $(PKG)_4(POG)_4(DOG)_4$ sample with a concentration of 1.0% by weight in phosphate with parameters of 0.1% strain and 1 rad/s. The G' values decreased beginning at 40 °C and by 50 °C, the G" values exceeded the G' values, which indicated that the gel had disassembled.[3] Figure 5.17d is a bar graph representation of the G' values for 0.5% and 1.0% by weight gels in phosphate at 20, 30 and 37 °C. The temperatures examined were included in order to gain insight on the behavior of the system before the gel melts. As shown in Figure 5.17d, the gels showed their highest G' at 30 °C and 37 °C and a substantially lower observed storage modulus at 20 °C.[3] The CD melting profile showed a minor transition of the peptide from 10 to 30 °C prior to the actual triple helix unfolding of the system. When we combined the

temperature dependent rheological results with the CD data, it suggested that as the peptide slightly unfolds between 10 and 30 °C, the unfolded regions of fiber may interdigitate with other nanofibers resulting in the strengthening of the hydrogel.[3]

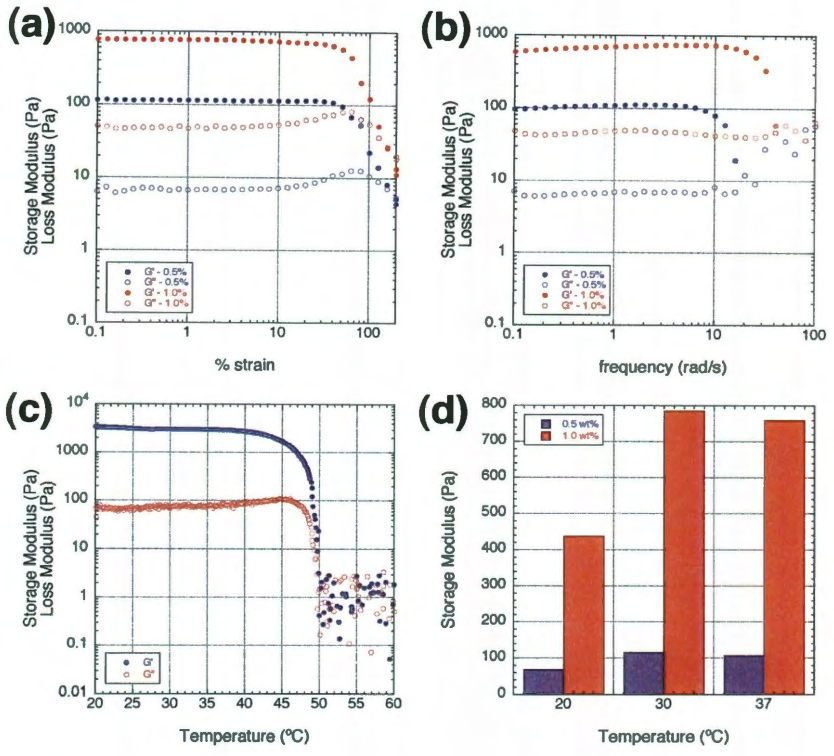

Figure 5.17. Rheology of $(PKG)_4(POG)_4(DOG)_4$ in phosphate demonstrating the temperature dependent strength of the hydrogel. Adapted from Figures 5 and S7 in reference 2. (a) Strain sweep at 0.5% and 1.0% by weight peptide concentration in phosphate buffer at a temperature of 30 °C and a frequency of 1 rad/s shown as storage modulus (G') and loss modulus (G''). (b) Frequency sweep at 0.5% and 1.0% by weight peptide concentration in phosphate at a temperature of 30 °C and 1% strain shown as storage modulus (G') and loss modulus (G''). (c) Temperature ramp rheological study run from 20 to 60 °C with a ramp of 0.5 °C/min and with parameters of 0.1% strain and a frequency of 1 rad/s.[3]

5.3.4. $(PKG)_4(POG)_4(DOG)_4$: Enzyme Degradation

As a simple functional test of the $(PKG)_4(POG)_4(DOG)_4$ hydrogel, we compared its ability to be broken down by collagenase (type IV, Invitrogen) to that of rat-tail collagen. The primary component of collagenase type IV is MMP2, a protease known to specifically cleave between the X and Gly residues of an X-Y-Gly repeat found in a triple helix.[12] $(PKG)_4(POG)_4(DOG)_4$ hydrogels were prepared at a concentration of 2.0% by weight in phosphate buffer and treated with either collagenase in HBSS (Hank's Balanced Salt Solution) or HBSS alone. Hydrogels of rat-tail collagen were prepared in the same fashion, with and without collagenase. All samples were incubated at room temperature (approximately 20 °C), 30 °C or 37 °C and the state of the hydrogels were assessed at the following time points after collagenase addition: 0, 1, 4, 6, 12, 24 and 48 hours.[3] Details of sample preparation are given in the experimental section below. As shown in Table 5.1, hydrogels prepared from our self-assembling peptide and rat-tail collagen degraded at similar rates: samples of both types of hydrogels treated with collagenase were found to be fully dissolved after 1 hour at 37 °C or 4 hours at room temperature and 30 °C while untreated controls were not.[3] However, the $(PKG)_4(POG)_4(DOG)_4$ control without collagenase degraded at 37 °C at the 24 hour time point, which was discouraging for future biomedical applications.

Hydrogel		Time (hours)						
		0	1	4	6	12	24	48
$(PKG)_4(POG)_4(DOG)_4$	20 °C	intact	intact	intact	intact	intact	intact	intact
	30 °C	intact	intact	intact	intact	intact	intact	intact
	37 °C	intact	intact	intact	intact	intact	dissolved	dissolved
$(PKG)_4(POG)_4(DOG)_4$ with collagenase	20 °C	intact	intact	dissolved	dissolved	dissolved	dissolved	dissolved
	30 °C	intact	intact	dissolved	dissolved	dissolved	dissolved	dissolved
	37 °C	intact	dissolved	dissolved	dissolved	dissolved	dissolved	dissolved
Rat-tail Collagen	20 °C	intact	intact	intact	intact	intact	intact	intact
	30 °C	intact	intact	intact	dispersed	dispersed	dispersed	dispersed
	37 °C	intact	intact	intact	intact	intact	intact	intact
Rat-tail Collagen with collagenase	20 °C	intact	intact	dissolved	dissolved	dissolved	dissolved	dissolved
	30 °C	intact	intact	dissolved	dissolved	dissolved	dissolved	dissolved
	37 °C	intact	dissolved	dissolved	dissolved	dissolved	dissolved	dissolved

Table 5.1. Collagenase mediated degradation of $(PKG)_4(POG)_4(DOG)_4$ and rat-tail collagen hydrogels. Adapted from Table S1 in reference 2. Samples were tested at room temperature (approximately 20 °C), 30 °C and 37 °C with and without collagenase. The table is organized by hydrogel, temperature and time point. At a given time point, if the gel appeared unchanged, intact was written in the table. If the gel degraded so that the sample appeared homogeneously fluid-like, dissolved was written in the table. If the gel outlines were no longer visible but gel-like regions were visible within the sample, dispersed was written in the table.[3]

5.3.5. $(PKG)_4(POG)_4(DOG)_4$: X-Ray Diffraction

To learn more about the packing morphology of the $(PKG)_4(POG)_4(DOG)_4$ self-assembled nanofibers, x-ray diffraction studies were carried out on a dried sample of $(PKG)_4(DOG)_4(DOG)_4$ in phosphate (see experimental section below for sample preparation). As was apparent from the microscopy images, neighboring fibers lacked a common orientation axis.[3] In order to partially align the fibers, the drying peptide solution was placed in a strong magnetic field to promote alignment during the drying process. This methodology has been shown to produce highly aligned protein fibers,[13] but had only limited success in this system. Figure 5.18d shows the recorded diffraction pattern. The dried pellet exhibited some alignment and the data was analyzed by performing a radial integration of the diffraction pattern to yield a plot of the observed intensities as a function of D-spacing (Figure 5.18e). The plot showed three distinct features: a weaker, sharp line near 2.8 Å, a diffuse intense reflection near 4.3 Å and a strong well-defined band near 11.5 Å.[3] The spacing of the observed lines agreed well with that observed for collagen from stretched kangaroo-tail tendon.[14]

Based on this, we assigned the 11.5 Å band to the distance between two triple helices inside the nanofibers, the diffuse reflection at 4.3 Å to the distance between peptide chains inside a triple helix and the reflection at 2.8 Å to the translation per triple helical triplet.[14] This suggested that our collagen-like peptide fibers were packing in a fashion similar to natural collagen.

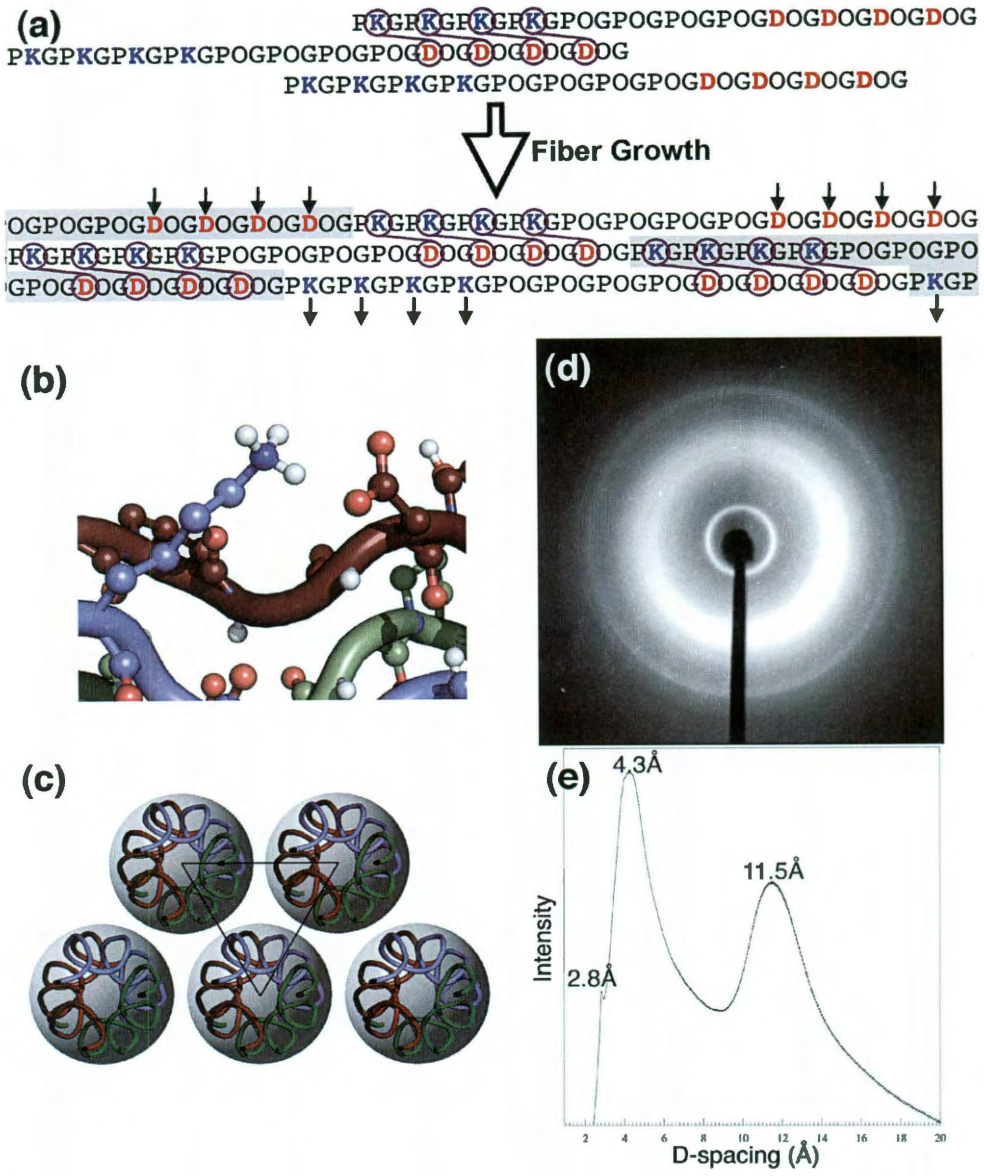

Figure 5.18. Proposed mechanism of fiber self-assembly. Adapted from Figure 5 in reference 2. (a) Peptide sequence shown as single letter amino acid code with P for proline, K for lysine, G for glycine, O for hydroxyproline and D for aspartate. Minimum repeating unit of the triple helical fiber has extensive "sticky" ends. As additional peptides (shaded grey) add to the minimum repeating unit, the percentage of amino acids forming a high quality triple helical structure rapidly increases. Positively charged lysine residues are in blue, negatively charged aspartates are in red and satisfied intrahelical electrostatic interactions are indicated by purple lassos. Available interhelical charged pair hydrogen bonds are indicated by small arrows. (b) Lysine - Aspartate interaction between triplets *n* and *n+1* of adjacent peptide strands. (c) Quasi-hexagonal packing of growing fibers results in a bundle approximately 2 by 4 nm based on a triple helical cross section of 1.2 nm. (d) Fiber diffraction pattern and (e) its radially averaged intensity.[3]

5.3.6. $(PKG)_4(POG)_4(DOG)_4$: Proposed Mechanism of Assembly

As mentioned previously, the charged pairing of lysine and aspartate had been shown to form direct electrostatic interactions in collagen mimetic peptides.[15] Specifically, lysine's side chain reaches in a C-terminal direction to make an intimate salt-bridge hydrogen bond with an aspartate on an adjacent, lagging peptide offset by three amino acids (Figure 5.18b). Since $(PKG)_4(POG)_4(DOG)_4$ formed a homotrimer, there was a potential for these charged amino acid salt bridges to form between peptide strands and create an offset, sticky-ended triple helix.[3] Similar sticky-ended assemblies have been designed and reported, particularly for alpha-helical coiled coils.[16,17] Figure 5.18a shows the proposed repeating unit of peptide self-assembly.[3] Lysine-aspartate interactions are highlighted with purple lassos. This favorable interaction forced a dramatic sticky-ended triple helix in which only one third of the possible lysine-aspartate pairs were satisfied. However, as additional peptides were added to extend the triple helical system, the fraction of satisfied charged pairs increased. For example, adding just one more peptide increased the fraction of satisfied charged pairs to one half and an infinite length triple helical fiber will have two thirds of the salt-bridges satisfied through intra-helical interactions.[3] In addition, for our collagen mimetic system, fiber elongation satisfied a larger percentage of inter-peptide backbone hydrogen bonds donated from glycine which are known to stabilize collagen triple helices.[18-23] In the three peptide nucleation center, only 50% of the glycine residues were capable of forming these inter-peptide interactions however as the fiber grew, the percentage of glycines participating in hydrogen bonds approached 100%.[3]

As observed by TEM, SEM and AFM, the nanofibers formed had dimensions

greater than that of a single collagen triple helix.[3] Therefore, several triple helices must bundle together to form the observed nanofibers. This was backed up by fiber diffraction data, which clearly displayed the characteristic triple helix packing band at 11.5Å (Figures 5.18d and 5.18e). The lysine and aspartate side chains not participating in intrahelix salt-bridges (indicated by small arrows in Figure 5.18a) were available for inter-helix interactions, which promoted helix bundling.[3] In natural collagen, five helices are believed to pack in a quasi-hexagonal fashion to form fibrils that continue to assemble into mature fibers.[24,25] Based on the measured height and width for the $(PKG)_4(POG)_4(DOG)_4$ nanofibers measured from AFM and cryo-TEM respectively and a helix packing distance from fiber diffraction, we hypothesized that our peptide system assembled in a similar fashion.[3] A schematic of this packing is given in Figure 5.18c. The calculated nanofiber height from AFM was found to be 1.2 ± 0.3 nm and the observed nanofiber width from cryo-TEM was 4 – 5 nm.[3] Both of these measured values are within reason for our proposed quasi-hexagonal packing however, some additional comment on the fiber height should be made. The value measured by AFM appeared to be significantly less than expected. There were several possible explanations for this. First, it is known that in AFM, soft organic materials often have measured heights less than expected due to flattening from surface forces or from the AFM tip itself.[26] Another possible explanation was that the triple helices not in direct contact with the mica surface are removed during the washing step leaving behind collagen ribbons only one triple helix high and shorter in length.[3] In fact, our AFM measured height was very nearly exactly what would be expected from a single triple helix. Nevertheless, the bundled fibrous structure was well supported by our x-ray diffraction data and the variances

between cryo-TEM, stained TEM, AFM, SEM and x-ray diffraction could be attributed to necessary differences in sample preparation.[3]

5.3.7. $(PKG)_4(POG)_4(DOG)_4$: Assembly in Additional Buffers

As alluded to above, we explored the self-assembly of $(PKG)_4(POG)_4(DOG)_4$ in multiple buffers.[3] The buffer library attempted to include a range of ionic strengths and all buffers were made at pH 7. The buffers examined were deionized water (these samples were pH adjusted prior to final dilution in order to ensure accurate pH), Tris (10mM tris(hydroxymethl)-aminomethane, pH 7) and PBS (10 mM phosphate, 150 mM sodium chloride, pH 7). Samples at 0.5% and 1.0% by weight concentrations of peptide were prepared in each of these buffers and examined for triple helical stability using CD and nanofiber formation via AFM.[3] In addition, for the samples that formed hydrogels, rheological studies were performed to assess the gel properties. The results from the CD experiments, the AFM analysis and the rheological studies are described below sequentially.[3]

Thermal unfolding studies showed the peptide capable of forming stable triple helices in all three buffers at both peptide concentrations.[3] In water (Figures 5.19a and 5.19b), the melting temperature for the peptide was 40 – 41 °C, the same T_m seen for samples made in phosphate. In Tris (Figures 5.19c and 5.19d), the 0.5% by weight sample unfolded at 23 °C while the higher concentration sample had a melting temperature that overlapped the water and phosphate samples. Last, in PBS (Figures 5.19e and 5.19f), both peptide concentrations unfolded at 23 °C. When the CD data for all three buffers was combined with the observations of gelation for each sample, a

correlation between the melting temperature and the presence of a hydrogel emerged: gel-forming samples showed a melting transition at 40 – 41 °C while samples that did not form gels had a melting temperature around 23 °C.[3]

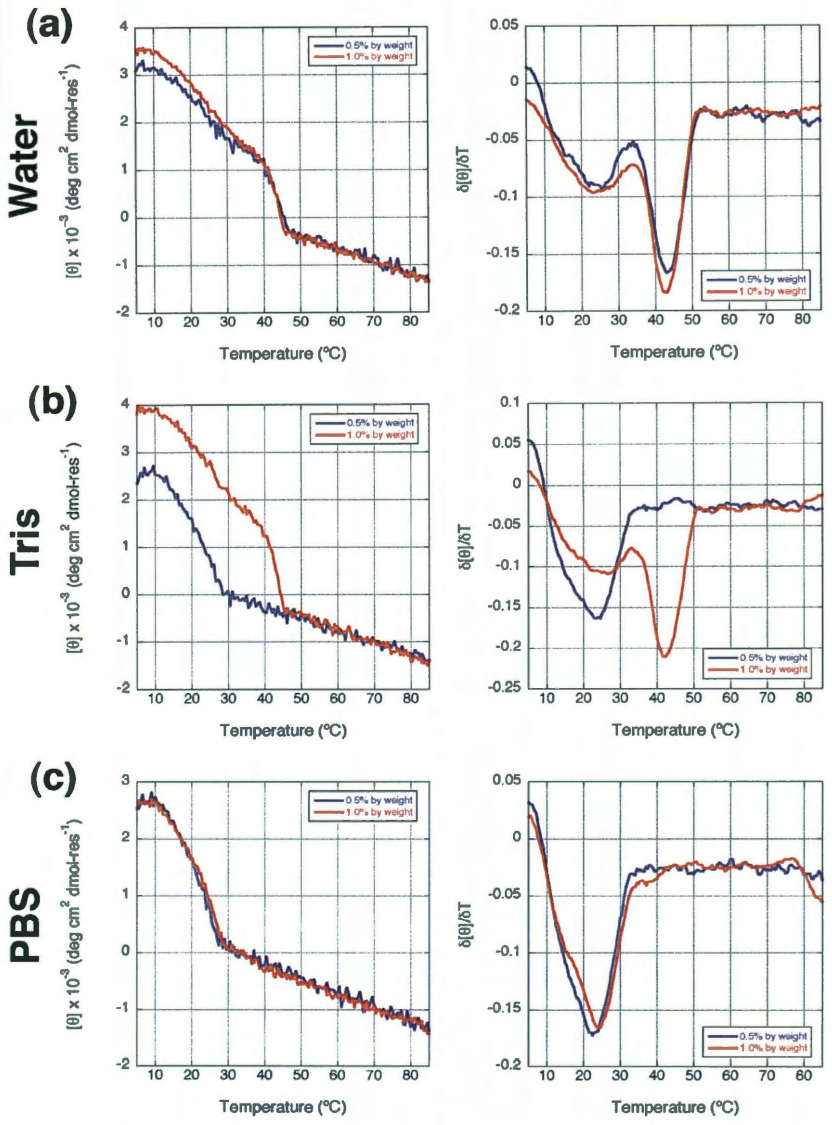

Figure 5.19. CD thermal unfolding curves for $(PKG)_4(POG)_4(DOG)_4$ in (a) water (adjusted to pH 7), (b) Tris and (c) PBS. Adapted from Figure S8 in reference 2. The melting profiles for each system are given as MRE vs. temperature in the left column and the first derivative of MRE vs. temperature in the right column with data for 0.5% by weight concentration in blue and 1.0% by weight concentration in red.[3]

Despite the differences in triple helical stabilities between buffers and between peptide concentrations within the same buffer, all samples formed nanofibers visible by AFM.[3] In water, the formed nanofibers had heights of 1.2 ± 0.3 nm and representative images are given in Figure 5.20. In Tris, shown in Figure 5.21, the observed fibers had heights of 1.1 ± 0.2 nm. Lastly, in PBS buffer, the fibers seen in AFM images had heights of 1.2 ± 0.2 nm (Figure 5.22). Therefore, the fiber height for $(PKG)_4(POG)_4(DOG)_4$ was similar for all buffers examined possibly indicating a similar mechanism for assembly in all buffers. In terms of fiber appearance, the fibers for gelled samples appeared very similar to those seen for phosphate sampled described above. The fibers seen in the non-gelled samples, however, had a more dense network of fibers than the gelled samples however the nanofiber lengths appeared shorter. Therefore, a hypothesis for the presence of nanofibers without visible gelation is that the nanofibers in 0.5% by weight Tris and both PBS samples were not long enough to form an organized hydrogel network thus no visible gelation was seen.[3]

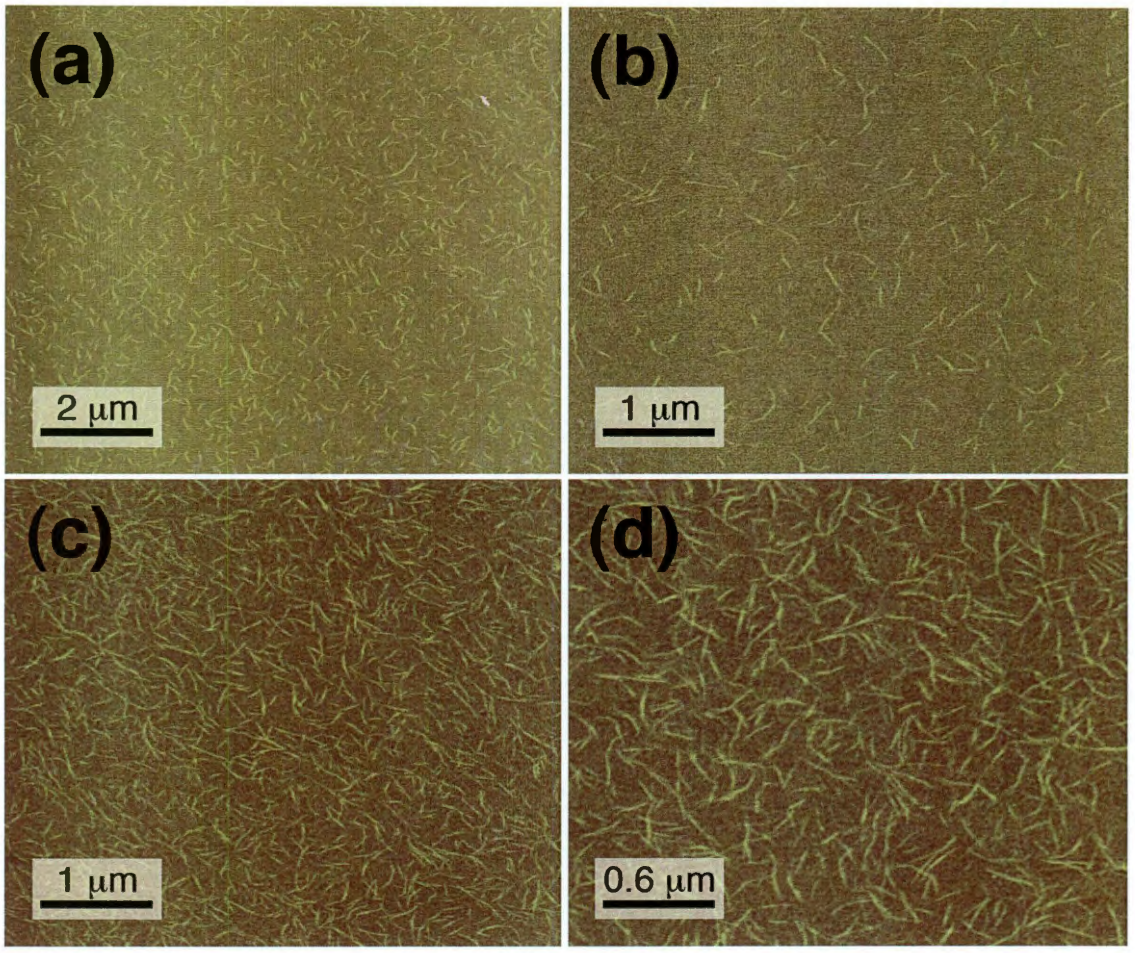

Figure 5.20. AFM of $(PKG)_4(POG)_4(DOG)_4$ nanofibers in water, pH 7, as observed after spin coating the gelled samples onto freshly cleaved mica at concentrations of (a and b) 0.5% by weight and (c and d) 1.0% by weight concentrations. Adapted from Figure S9 in reference 2.[3]

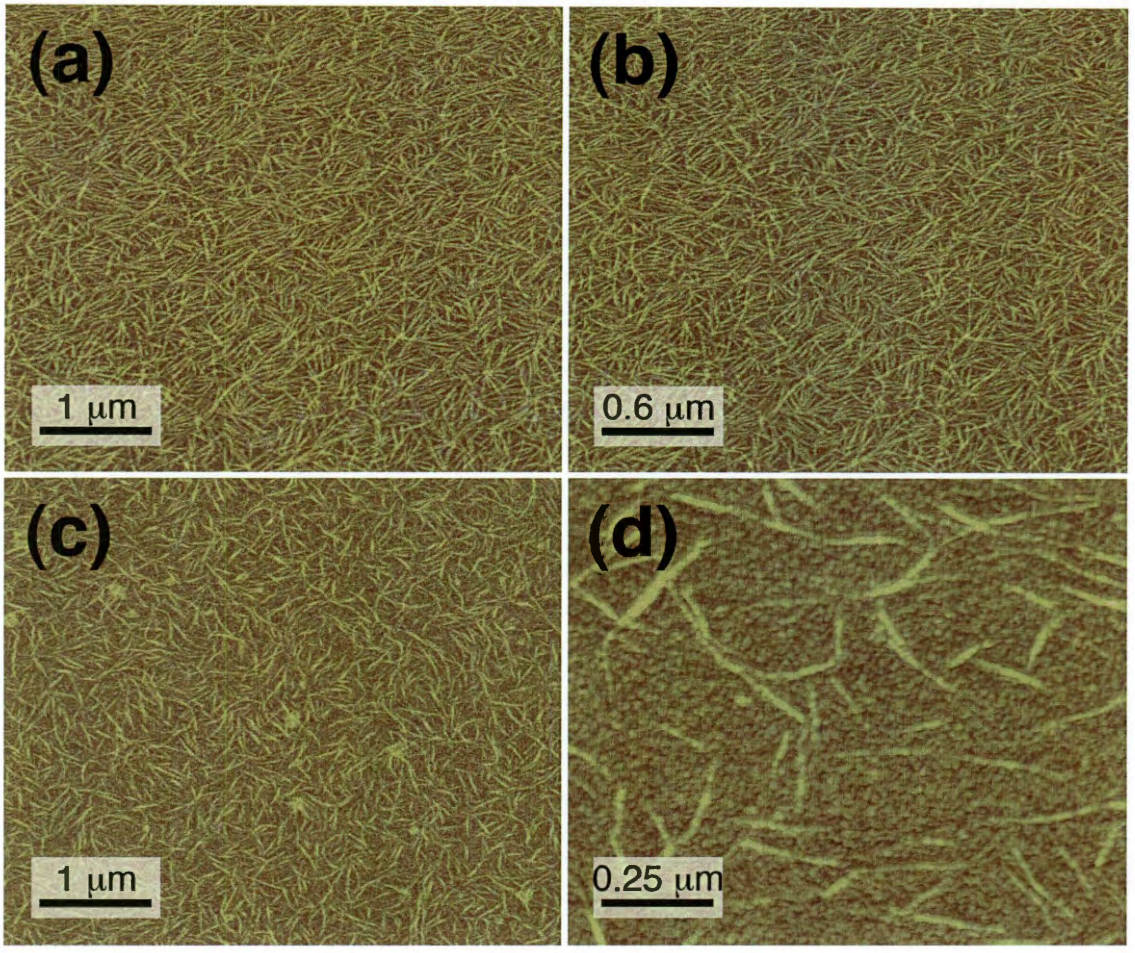

Figure 5.21. AFM of $(PKG)_4(POG)_4(DOG)_4$ nanofibers in 10 mM Tris, pH 7, as observed after spin coating onto freshly cleaved mica at concentrations of (a and b) 0.5% by weight and (c and d) 1.0% by weight concentrations. Adapted from Figure S10 in reference 2.[3]

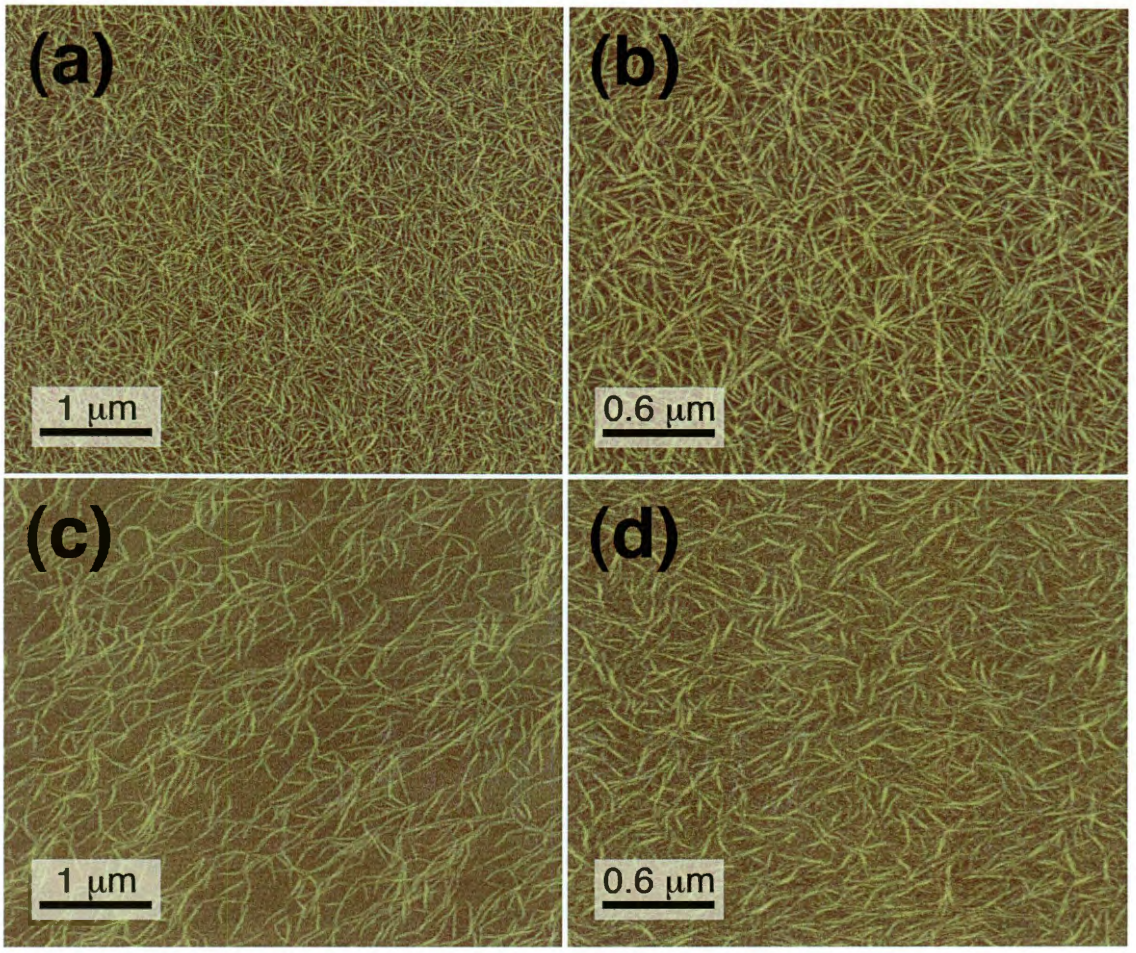

Figure 5.22. AFM of $(PKG)_4(POG)_4(DOG)_4$ nanofibers in PBS, pH 7, as observed after spin coating the samples onto freshly cleaved mica at concentrations of (a and b) 0.5% by weight and (c and d) 1.0% by weight concentrations. Adapted from Figure S11 in reference 2.[3]

Lastly, when the hydrogel properties were assessed for all three buffers, the ability of the peptide to form a hydrogel decreased as ionic strength increased.[3] For example, both peptide concentrations formed hydrogels in water and only the 1.0% by weight samples formed a hydrogel in Tris after the standard 8 hour incubation. PBS samples gelled very slowly, requiring more than a week for a 1.0% by weight sample, therefore rheological studies for this buffer were not performed. In addition, the

rheological properties decreased as buffer ionic strength increased as well. Figure 5.23 shows the strain and frequency sweep data for the water and Tris hydrogels. The G' values seen for the water samples were larger than those seen for the phosphate buffer (described previously) and we attributed this to the very low ionic strength of the deionized water samples. The rheological values seen for 1.0% by weight samples in Tris were very similar to the G' values reported for the 1.0% by weight sample in phosphate despite the fact that Tris samples made at concentrations below 1.0% by weight samples did not form gels. Last, the water and Tris samples showed a temperature dependence of the rheological properties, similar to phosphate samples.

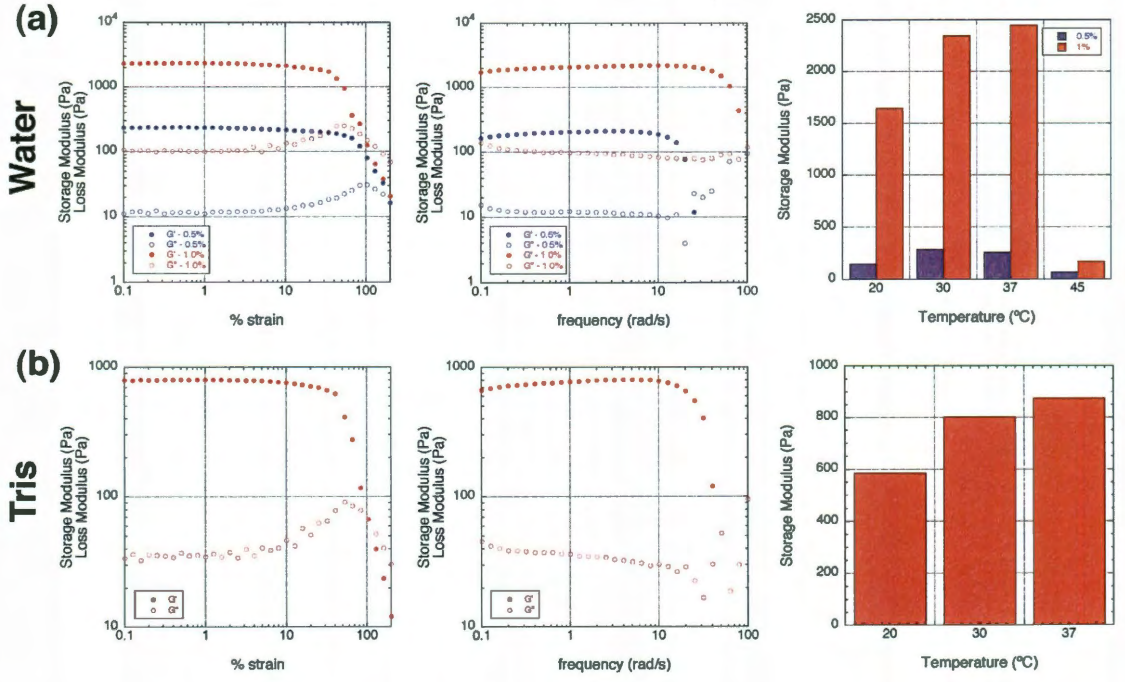

Figure 5.23. Rheology of the collagen-like peptide $(PKG)_4(POG)_4(DOG)_4$ in (a) water, pH 7 (0.5% and 1.0% by weight concentrations) and (b) Tris, pH 7 (1.0% by weight concentration). Adapted from Figure S12 in reference 2. The left column shows data from strain sweep studies at 1 rad/s (30 °C), the center column is frequency sweep data at 1% strain (30 °C) and the right column displays the temperature dependence of rheological properties in each buffer (data points acquired at 1 rad/s and 1% strain).[3]

5.4. Conclusions

Through the exploration of the peptide $(PRG)_4(POG)_4(EOG)_4$, designed by Chaikof *et al.*,[1] many key findings about collagen mimetic nanofiber formation were discovered. First, $(PKG)_4(POG)_4(DOG)_4$ was the only peptide within the library that assembled into organized nanofibers visible in the solution state and that formed a hydrogel. This peptide was the first reported system whose multi-hierarchical assembly mimicked native collagen and could be proven at each step.[3] Second, peptides designed around lysine-aspartate interactions showed superior assembly compared to those designed with arginine and glutamate. This conclusion complements previous results on AAB and ABC type collagen mimetic heterotrimers,[15,27,28] but was not connected to nanofiber formation until the $(PKG)_4(POG)_4(DOG)_4$ system was published.[3] Third, the arrangement of the positively charged region at the N-terminus and the negatively charged region at the C-terminus of the peptide yielded the assembly of nanofibers visible in dry-TEM when the inverse sequences with similar peptide lengths did not. The best hypothesis for this observation was that the K-D interactions between the lysine in triplet *n* and the aspartate in triplet *n+1* of the following strand play a major role in the stabilization of triple helical and nanofiber assemblies. To maximize these interactions, a slight stagger in $(PKG)_4(POG)_4(DOG)_4$ was necessary.[3] However, neither $(DOG)_4(POG)_4(PKG)_4$ nor $(DOG)_6(POG)_6(PKG)_6$ can form any of these stabilizing interactions due to the inversion of the placement of the charged regions resulting in the inferior nanofiber assembly for these two peptides.

The successful assembly of $(PKG)_4(POG)_4(DOG)_4$ into a triple helix, nanofiber and hydrogel marked this system as the first collagen mimetic peptide to replicate the

assembly of collagen.[3] The high quality hydrogels were degraded at a similar rate to rat-tail collagen and the diffraction patterns taken on the nanofibers showed similar peaks to previously reported work on kangaroo tendon.[3] Because of collagen's major role in critical functions such as tissue structure, repair and regeneration, this peptide, and those based on its design, have the potential to play an important role in regenerative medicine and drug delivery.

5.5. Experimental

Peptide Synthesis. $(PRG)_4(POG)_4(EOG)_4$ and $(PKG)_4(POG)_4(DOG)_4$ were synthesized using standard Fmoc chemistry for solid phase peptide synthesis on an Advanced Chemtech Apex 396 multi-peptide automated synthesizer at a scale of 0.15 mM on a glycine pre-loaded Wang resin. $(EOG)_3(POG)_3(PRG)_3$, $(DOG)_4(POG)_4(PKG)_4$ and $(DOG)_6(POG)_6(PKG)_6$ were previously synthesized by Varun Gauba using the same Fmoc chemistry. During the synthesis, amino acids were added in a 4:1 molar ratio to the growing peptide chain using the coupling agents O-benzotriazole-N,N,N',N'-tetramethyluroniumhexafluorophosphate (HBTU), 1-hydroxybenzotriazole hydrate (HoBt), and N,N-diisopropylethylamine (DiEA) in dimethylformamide (DMF) at molar ratios of 4:4:6 respectively. Amino acids were deprotected once coupled to the peptide chain using a 25% (by volume) solution of piperidine in DMF. The peptide was cleaved from the resin with a 38:1:1 mixture of trifluoroacetic acid (TFA), water and triisopropylsilane.

Mass Spectrometry. All peptides were examined by either MALDI/TOF mass spectrometry on a Bruker Autoflex mass spectrometer in positive ion mode or by ESI/TOF mass spectrometry on a Bruker microTOF to verify that the peptides were synthesized correctly. Spectra were analyzed using FlexAnalysis software.

Peptide Purification. Purification was performed on a Varian PrepStar220 HPLC using a preparative reverse phase C-18 column. The two HPLC solvents referred to as solvents A and B are water and acetonitrile respectively, each containing 0.05% TFA. The solvents were eluted through the column with a linear gradient ranging from a 1 to 3% increase in concentration of solvent B per minute. Once collected, the HPLC fractions were rotovapped down to remove the acetonitrile fraction and then lyophilized resulting in a peptide powder. Next, the peptides were dialyzed against deionized water in order to remove salts and then lyophilized again.

Sample Preparation. All peptide concentrations were measured by weight. All samples were adjusted to pH 7 with sodium hydroxide prior to the addition of buffer and then annealed for 15 minutes at 85 °C. Lastly, the samples were incubated at room temperature for at least 12 hours prior to characterization to ensure complete assembly.

Circular Dichroism. All spectra and thermal unfolding studies were performed on a Jasco J-810 spectropolarimeter equipped with a Peltier temperature control system. Quartz cells were used with pathlengths of 0.01 cm and 0.1 cm depending on the peptide concentration and buffer. Spectra were collected from 190 – 250 nm. Melting

experiments were performed from 5 to 85 °C, monitoring at 225 nm, and the first derivative of the thermal unfolding curve was taken in order to determine the melting temperature of the sample. The molar residual ellipticity (MRE) is calculated from the measured ellipticity using the equation:

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

where θ is the ellipticity in mdeg, m is the molecular weight in g/mol, c is the concentration in mg/ml, l is the pathlength of the cuvette in cm, and n_r is the number of amino acids in the peptide.

Atomic Force Microscopy. Samples were prepared and dropped onto freshly cleaved mica while spinning on a Headway Research, Inc. Photo-resist spinner. The sample was quickly rinsed with deionized water for 4 – 5 seconds and then spun for an additional 10 minutes. AFM images were collected on a Digital Instruments Nanoscope IIIa AFM in tapping mode under ambient conditions. Height profiles were obtained using Nanoscope software (20 measurements were taken per peptide concentration and buffer, averaged and the standard deviation calculated).

Transmission Electron Microscopy (TEM). Samples for TEM were prepared on Quantifoil® R1.2/1.3 holey carbon mesh on copper grids. For dry TEM, 0.5% by weight uranyl acetate (UA), pH 3.5, was used to positively stain the TEM grids. 2.0% by weight phosphotungstic acid (PTA), pH 6 was used to stain the TEM grids using negative staining techniques. All stains were made bi-weekly and syringe filtered prior to use. For positive staining, the peptide solution was added to the carbon side of a TEM grid,

allowed to dry for one minute and then was indirectly blotted with filter paper to remove excess solution. The grid was allowed to dry for 5 minutes before UA solution was dropwise added to the grid for 10 seconds and then immersed in water two times. The grid was then allowed to dry overnight. For negative staining, the peptide solution was added to the carbon side of a TEM grid, allowed to dry for one minute, then indirectly blotted with filter paper to remove excess solution. The grid was allowed to dry for 5 minutes before it was inverted onto an aliquot of PTA solution where it remained for 10 minutes. The grid was then placed on filter paper to dry overnight.

Vitreous ice TEM samples were prepared as follows. First, the TEM grids were glow discharged for one minute with a 5 mA discharge on a EMS 100 Glow Discharge Unit. The next stages of sample preparation were performed using a FEI Vitrobot type FP5350/60. The peptide solution was added to the grid and immediately blotted for 2 seconds before being immersed in liquid ethane. The grid was then manually transferred from liquid ethane to liquid nitrogen where it was stored until imaging. All TEM imaging was performed on a JEOL 2010 microscope (200 kV) and cryo-imaging was taken at a temperature of -176 °C using low dose conditions.

Scanning Electron Microscopy (SEM). 100 µl aliquots of each gel were placed in a 24-wellplate. Gels were dehydrated in a series of ethanol/water solutions progressing from 30% to 100% ethanol over the course of 24 hours. The dehydrated gels were critical point dried using an Electron Microscopy Sciences 850 critical point drier. They were then affixed to SEM pucks using conductive carbon tape. The pucks were sputter coated with 10 nm gold, rotated, and then sputter coated with an additional 5 nm gold using a

CRC-150 sputter coater. Samples were imaged using a FEI Quanta 400 ESEM at 20.00 kV.

Rheology. All rheological studies were performed on a TA AR-G2 rheometer. Strain and frequency experiments were performed using 12 mm stainless steel parallel plate geometry with a 500 mm gap size. Strain sweeps maintained a fixed frequency (1 rad/s) and a variable strain (0.01 – 200%). Frequency sweeps utilized a fixed strain (1%) and varying frequencies (0.1 – 200 rad/s).

X-ray diffraction. A freshly annealed 1.0% by weight sample of $(PKG)_4(POG)_4(DOG)_4$ was dried by placing 10 μl droplets between two capillaries held in the center of a custom magnet assembly as described by Sunde *et. al.* over a period of several days.[13] A dried peptide pellet attached to the end of the capillary was used for data collection. Data was collected at 1.54 Å using a Rigaku RUH3R rotating anode x-ray generator with a Rigaku R-axis IV++ detector. The detector was placed at a distance of 180.0 mm from the sample, which was cooled using a N_2 stream to 100 K. Diffraction patterns were acquired with exposure times ranging from 1 to 40 minutes, with the highest exposure time yielding the best pattern. The data was analyzed using the Fit2D software package.[29] The position of the beam stop was calculated using the ring 11.5 Å and a median filter was applied to the data. Radial integration was carried out to produce a 1D profile of the observed intensities as a function of D-spacing (Å) and angular integration to generate a plot of the observed intensities as a function of D-spacing (Å) and azimuthal angle.

Collagenase Degradation Study. Using the previously described gelation procedure, gels of $(PKG)_4(POG)_4(DOG)_4$ were prepared at a concentration of 2.0% by weight in 10 mM phosphate buffer. Directly after annealing, 100 μl of solution was pipetted into two wells of a 96-well cell culture plate and allowed to incubate at room temperature overnight. A 0.3% by weight collagenase type IV solution was prepared by dissolving 15.0 mg of non-sterile lyophilized enzyme into 5 ml of Hank's Balanced Salt Solution (HBSS). The solution was then filter-sterilized using a 0.2 μm filter attached to a syringe. Enzyme and buffer were from Invitrogen (Carlsbad, CA). After gelation was complete, 150 μl of collagenase was added on top of one well and 150 μL of HBSS was added on the other serving as a control. Three plates were prepared and incubated at room temperature (about 20 °C), 30 °C and 37 °C. Each condition was observed and imaged at 0, 1, 4, 6, 12, 24 and 48 hours after addition of the collagenase or HBSS. Plates containing rat-tail collagen gels at a concentration of 3.0% by weight were prepared and analyzed in an identical manner.

References

(1) Rele, S.; Song, Y.; Apkarian, R. P.; Qu, Z.; Conticello, V. P.; Chaikof, E. L. *J. Am. Chem. Soc.* **2007**, *129*, 14780-14787.

(2) Persikov, A. V.; Ramshaw, J. A.; Kirkpatrick, A.; Brodsky, B. *Biochemistry* **2000**, *39*, 14960-14967.

(3) O'Leary, L. E. R.; Fallas, J. A.; Bakota, E. L.; Kang, M. K.; Hartgerink, J. D. *Nature Chem.* **2011**, *3*, 821-828.

(4) Yang, Y. L.; Leone, L. M.; Kaufman, L. J. *Biophys. J.* **2009**, *97*, 2051-2060.

(5) Mi, K.; Wang, G.; Liu, Z.; Feng, Z.; Huang, B.; Zhao, X. *Macromol. Biosci.* **2009**, *9*, 437-443.

(6) Aulisa, L.; Dong, H.; Hartgerink, J. D. *Biomacromolecules* **2009**, *10*, 2694-2698.

(7) Greenfield, M. A.; Hoffman, J. R.; de, l. C., MO; Stupp, S. I. *Langmuir* **2010**, *26*, 3641-3647.

(8) Lamm, M. S.; Rajagopal, K.; Schneider, J. P.; Pochan, D. J. *J. Am. Chem. Soc.* **2005**, *127*, 16692-16700.

(9) Ozbas, B.; Kretsinger, J.; Rajagopal, K.; Schneider, J. P.; Pochan, D. J. *Macromolecules* **2004**, *37*, 7331-7337.

(10) Yokoi, H.; Kinoshita, T.; Zhang, S. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8414-8419.

(11) Zhang, S. G.; Holmes, T. C.; DiPersio, C. M.; Hynes, R. O.; Su, X.; Rich, A. *Biomaterials* **1995**, *16*, 1385-1393.

(12) Ottl, J.; Gabriel, D.; Murphy, G.; Knauper, V.; Tominaga, Y.; Nagase, H.; Kroger, M.; Tschesche, H.; Bode, W.; Moroder, L. *Chem. Biol.* **2000**, *7*, 119-132.

(13) Serpell, L. C.; Fraser, P. E.; Sunde, M. *Method. Enzymol.* **1999**, *309*, 526-536.

(14) Okuyama, K. *Connect. Tissue. Res.* **2008**, *49*, 299-310.

(15) Fallas, J. A.; Gauba, V.; Hartgerink, J. D. *J. Biol. Chem.* **2009**, *284*, 26851-26859.

(16) Pandya, M. J.; Spooner, G. M.; Sunde, M.; Thorpe, J. R.; Rodger, A.; Woolfson, D. N. *Biochemistry* **2000**, *39*, 8728-8734.

(17) Woolfson, D. N. *Biopolymers* **2010**, *94*, 118-127.

(18) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. *Science* **1994**, *266*, 75-81.

(19) Okuyama, K.; Okuyama, K.; Arnott, S.; Takayanagi, M.; Kakudo, M. *J. Mol. Biol.* **1981**, *152*, 427-443.

(20) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 272-281.

(21) Ramachandran, G. N.; Kartha, G. *Nature* **1955**, *176*, 593-595.

(22) Ramachandran, G. N.; Kartha, G. *Nature* **1954**, *174*, 269-270.

(23) Rich, A.; Crick, F. H. C. *J. Mol. Biol.* **1961**, *3*, 483-506.

(24) Ottani, V.; Martini, D.; Franchi, M.; Ruggeri, A.; Raspanti, M. *Micron* **2002**, *33*, 587-596.

(25) Ottani, V.; Raspanti, M.; Ruggeri, A. *Micron* **2001**, *32*, 251-260.

(26) Ruozi, B.; Tosi, G.; Leo, E.; Vandelli, M. A. *Talanta* **2007**, *73*, 12-22.

(27) Gauba, V.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 15034-15041.

(28) O'Leary, L. E. R.; Fallas, J. A.; Hartgerink, J. D. *J. Am. Chem. Soc.* **2011**, *133*, 5432-5443.

(29) Wess, T. J.; Hammersley, A.; Wess, L.; Miller, A. *J. Mol. Biol.* **1995**, *248*, 487-493.

Chapter 6: Conclusions

Collagen is the most abundant protein within the human body and due to its prevalence in the extracellular matrix (ECM), much research has been centered on understanding the stabilizing forces within this protein and replicating its assembly. Over the past four years, I have designed multiple peptide systems that utilized charged pair interactions to drive the self-assembly of collagen mimetic peptides into AAB type heterotrimers as well as homotrimeric nanofibers and hydrogels.

In order to direct the stable formation of AAB type heterotrimers, an initial design scheme focused on maximizing the number of lysine-aspartate salt bridges between peptide chains when assembled in an AAB type helix. However in this approach, we inadvertently made salt bridges within the homotrimers that stabilized these species preventing the formation of AAB type heterotrimers without the presence of residual homotrimer.

To improve upon the drawbacks of the first scheme, negative design was included. Heterotrimer formation was reinforced by favorable interactions between oppositely charged amino acids and homotrimers were minimized through the reduction of stabilizing POG triplets in each peptide as well as through the incorporation of charge repulsion. Two peptide mixtures from this design, $2(PKGPOG)_5 \bullet (EOG)_{10}$ and $2(PKGPOG)_5 \bullet (DOG)_{10}$, were the first reported heterotrimer systems to demonstrate compositional control, and attribute that makes them viable for use in the replication of AAB type collagens, such as collagen type I, including disease modeling and synthetic ECM mimetics.

Assembly of collagen mimetic nanofibers was initially examined through the incorporation of hydrophobic residues into highly charged peptide sequences so that salt bridges would stabilize the triple helix formation and the hydrophobic amino acids within the peptide chains would form interactions guiding nanofiber assembly. Although some of the designed models formed nanofibers visible in dry-TEM, cry-TEM of the peptide mixtures did not confirm the presence of these structures in the solution state. This inability to replicate the assembly of collagen as well as the difficulty seen in reproducing the dry-TEM results led to the termination of this design scheme.

Last, peptide assembly from triple helix through nanofiber and hydrogel, directed only by the presence of charged amino acids, was explored. From this study, a peptide, $(PKG)_4(POG)_4(DOG)_4$, was identified that folded into a stable triple helix, packed into organized nanofibers visible in dry-TEM, cry-TEM, AFM and SEM, and formed a hydrogel that degraded at a similar rate to rat-tail collagen. This was the first collagen mimetic peptide to report such assembly.

Despite the fact that the initial peptide design schemes for AAB heterotrimer and collagen mimetic nanofiber formation were unable to form the desired self-assembled species, key principles of peptide design were discovered through the failures of these systems that led to the successful work in subsequent design schemes. Based on the major role that collagen plays in tissue structure, repair and regeneration, the collagen mimetic peptides described in this thesis have the potential to play an important role in the replication of heterotrimeric collagen systems as well as in regenerative medicine and tissue engineering.

Appendix 1: Peptide Library

Peptide	Relevant Chapter(s)
$(PKGDOG)_5$	Chapter 2
$(POGDKG)_5$	Chapter 2
$(DKGPOG)_5$	Chapter 2
$(PRG)_{10}$	Chapter 3
$(PKG)_{10}$	Chapters 3 and 4
$(EOG)_{10}$	Chapter 3
$(DOG)_{10}$	Chapters 3 and 4
$(PRGPOG)_5$	Chapter 3
$(PKGPOG)_5$	Chapter 3
$(EOGPOG)_5$	Chapter 3
$(POGEOG)_5$	Chapter 3
$(POGDOG)_5$	Chapter 3
$WG(PKGPOG)_2PKGPOG*(PKGPOG)_2$[†]	Chapter 3
$WG(EOG)_{10}$	Chapter 3
$WG(DOG)_{10}$	Chapter 3
$PKGLKG(PKG)_7LIG$	Chapter 4
$(DOG)_3LOG(DIG)_2(DOG)_4$	Chapter 4
$PIG(POG)_4(LOG)_2PIG(POG)_2$	Chapter 4
$(PKG)_5LKG(PKG)_2LKGLIG$	Chapter 4
$LOGDIG(DOG)_3DIG(DOG)_4$	Chapter 4
$(POG)_2LOGPIG(POG)_2LOG(POG)_3$	Chapter 4
$(LOGPOG)_5$	Chapter 4
$(POGPIG)_5$	Chapter 4
$(LOGPIG)_5$	Chapter 4
$PKGLKG(PKG)_7LIG$	Chapter 4
$(DOG)_7LIG(DOG)_2$	Chapter 4
$PIGPOGLIG(POG)_7$	Chapter 4

[†] The asterisk indicates the presence of an N^{15}-label on the designated glycine residue.

Peptide	Relevant Chapter(s)
$(PRG)_4(POG)_4(EOG)_4$	Chapter 5
$(EOG)_3(POG)_3(PRG)_3$	Chapter 5
$(DOG)_6(POG)_6(PKG)_6$	Chapter 5
$(DOG)_4(POG)_4(PKG)_4$	Chapter 5
$(PKG)_4(POG)_4(DOG)_4$	Chapter 5

Appendix 2: Peptide Purification

All peptides were either analyzed by MALDI/TOF mass spectrometry on a Bruker Autoflex II or by ESI/TOF mass spectrometry on a Bruker microTOF, as described in previous chapters, and the acquired spectra were processed using FlexAnalysis software.

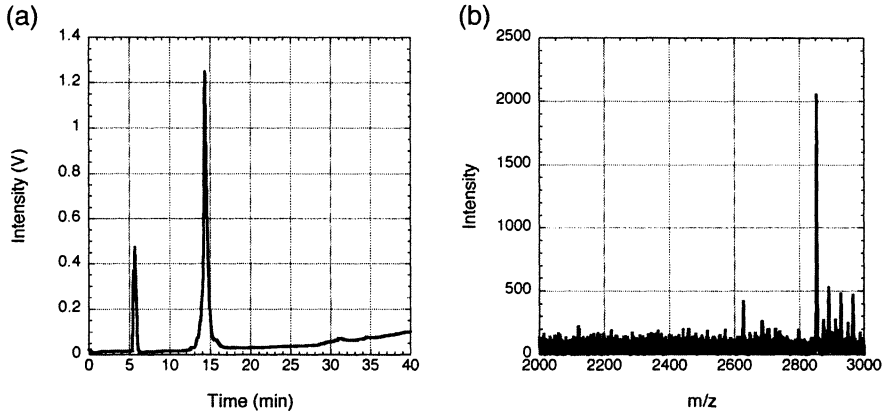

Figure A2.1. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $(PKGDOG)_5$. Expected mass: 2854.3 $[M+H]^+$, Observed mass: 2853.5 $[M+H]^+$.

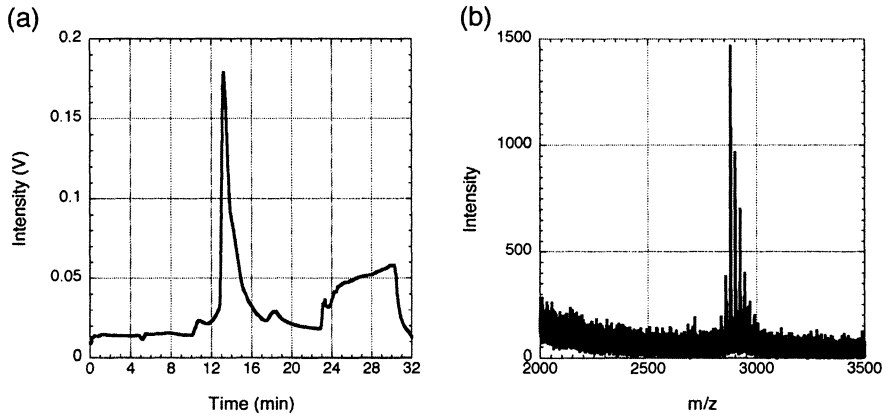

Figure A2.2. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $(POGDKG)_5$. Expected: 2879 $[M+Na]^+$, Observed: 2879.2 $[M+Na]^+$.

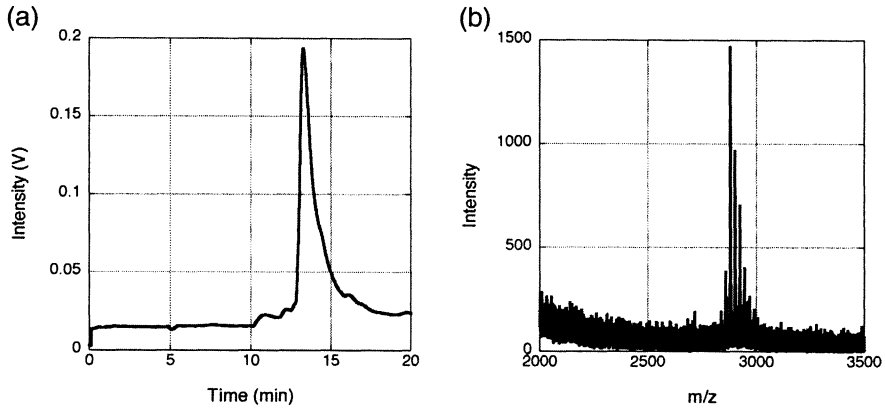

Figure A2.3. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $(DKGPOG)_5$. Expected: 2879.0 $[M+Na]^+$, Observed: 2878.58 $[M+Na]^+$.

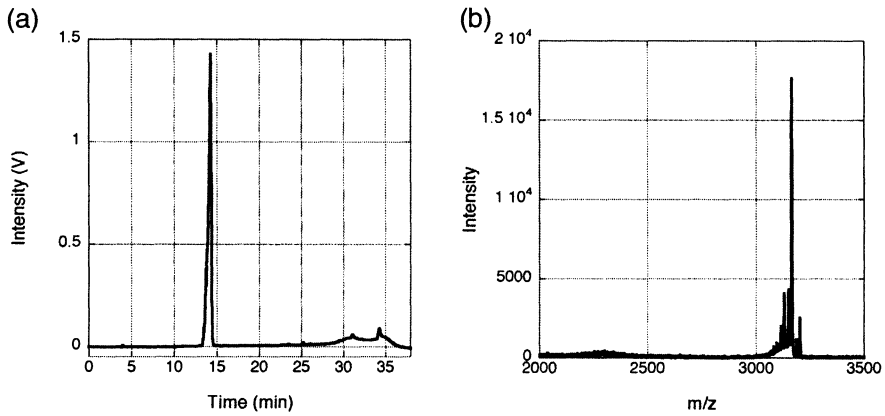

Figure A2.4. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $(PRG)_{10}$. Expected mass: 3163.6 $[M+H]^+$, Observed mass: 3163.3 $[M+H]^+$.

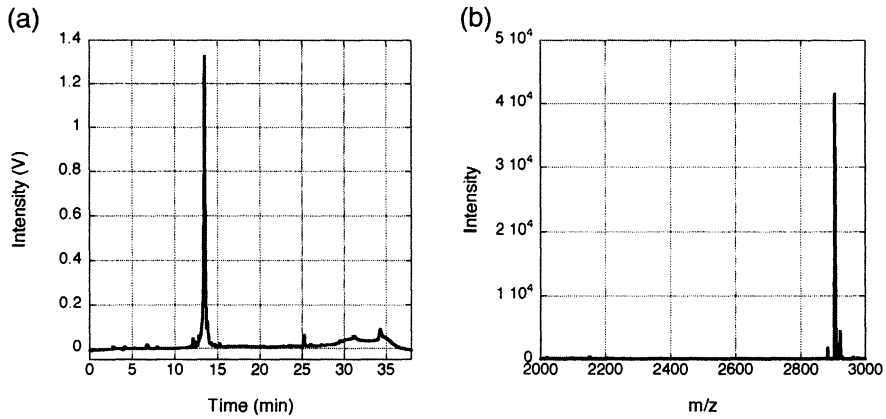

Figure A2.5. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $(PKG)_{10}$. Expected mass: 2905.5 $[M+Na]^+$, Observed mass: 2905.4 $[M+Na]^+$.

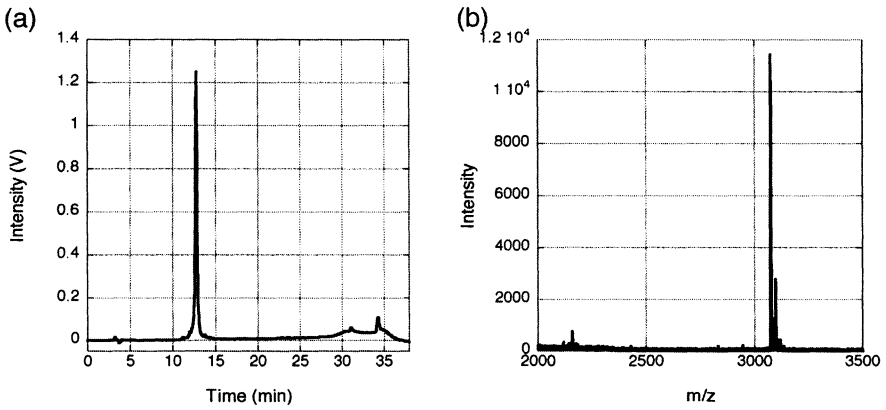

Figure A2.6. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $(EOG)_{10}$. Expected mass: 3073.1 $[M+Na]^+$, Observed mass: 3073.5 $[M+Na]^+$.

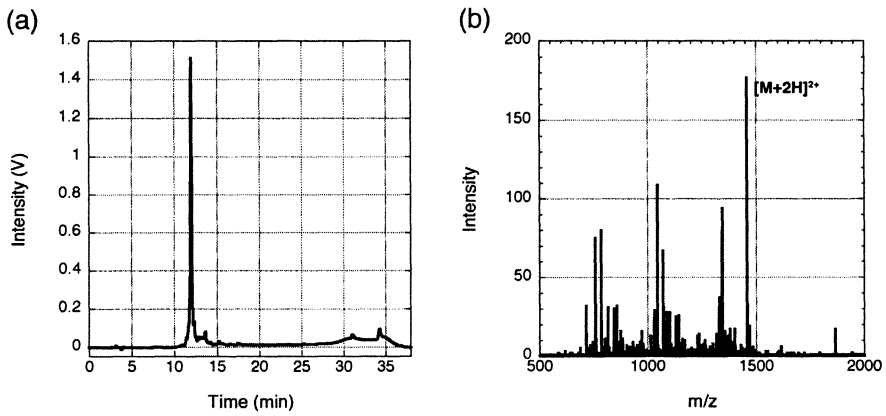

Figure A2.7. (a) HPLC and (b) ESI/TOF mass spectrometry data for $(DOG)_{10}$. Expected: 1456.7 $[M+2H]^{2+}$, Observed: 1456.6 $[M+2H]^{2+}$. Expected: 2912.6 $[M+H]^+$, Calculated: 2912.2 $[M+H]^+$.

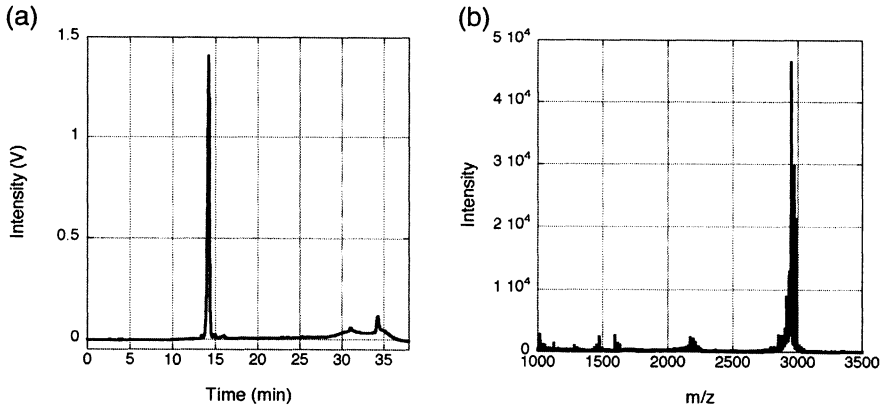

Figure A2.8. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $(PRGPOG)_5$. Expected: 2948.3 $[M+H]^+$, Observed: 2948.5 $[M+H]^+$.

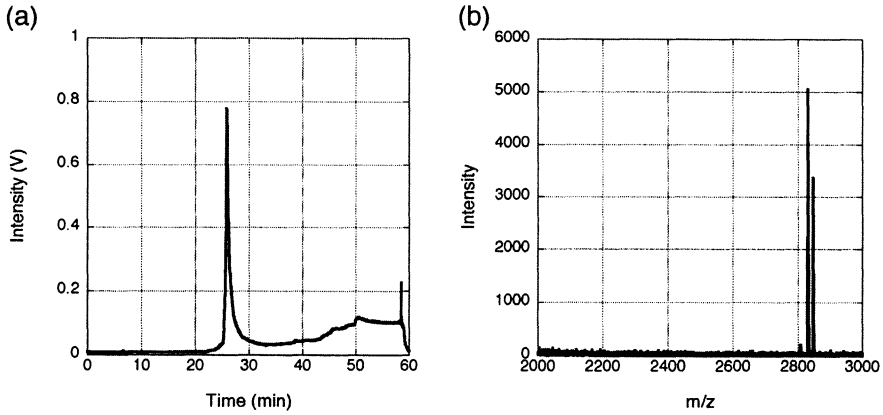

Figure A2.9. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $(PKGPOG)_5$. Expected: 2830.2 $[M+Na]^+$, Observed: 2830.8 $[M+Na]^+$.

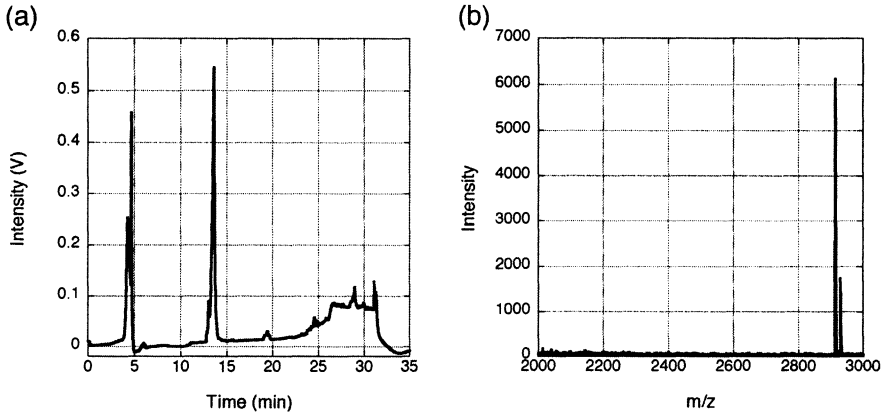

Figure A2.10. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $(EOGPOG)_5$. Expected mass: 2913.2 $[M+Na]^+$, Observed mass: 2913.8 $[M+Na]^+$.

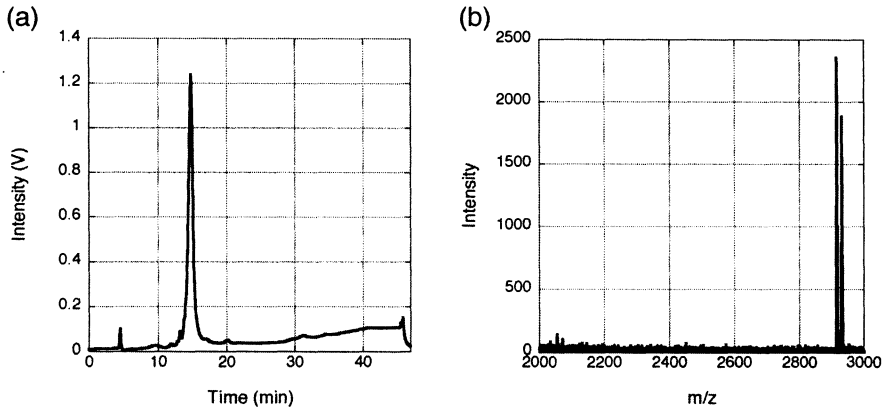

Figure A2.11. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $(POGEOG)_5$. Expected: 2914.9 $[M+Na]^+$, Observed: 2914.8 $[M+Na]^+$.

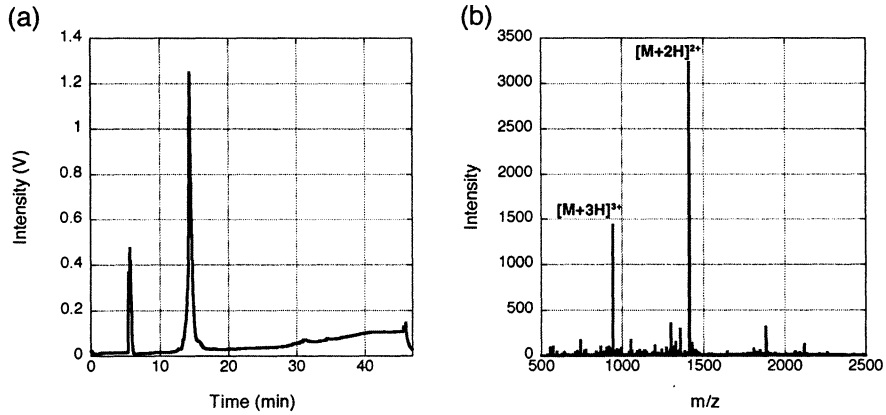

Figure A2.12. (a) HPLC and (b) ESI/TOF mass spectrometry data for $(POGDOG)_5$. Expected: 1411.8 $[M+2H]^{2+}$, Observed: 1411.6 $[M+2H]^{2+}$. Expected: 941.6 $[M+3H]^{3+}$, Observed: 941.4 $[M+3H]^{3+}$. Expected: 2822.8 $[M+H]^+$, Calculated: 2822.2 $[M+H]^+$.

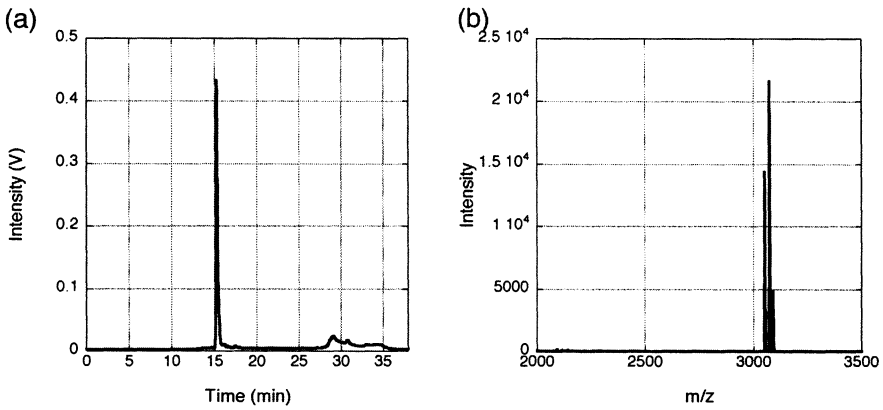

Figure A2.13. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $WG(PKGPOG)_2PKGPOG^*(PKGPOG)_2$ where G* is an N^{15}-labelled glycine residue. Expected: 3072.6 $[M+Na]^+$, Observed: 3072.4 $[M+Na]^+$.

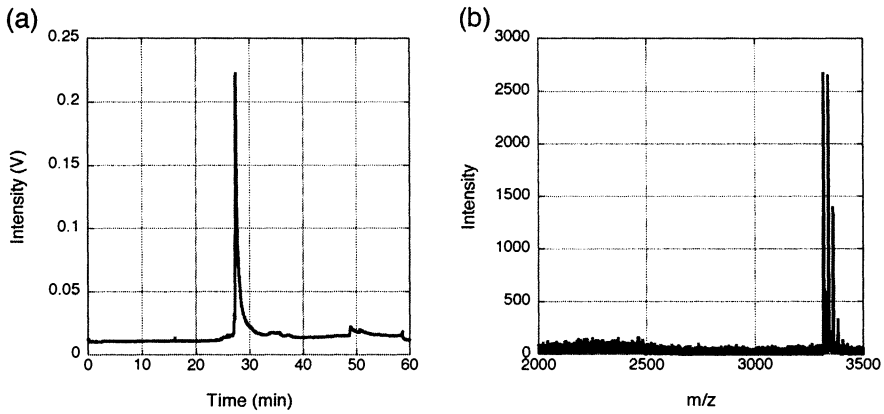

Figure A2.14. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $WG(EOG)_{10}$. Expected: 3316.2 $[M+Na]^+$, Observed: 3315.8 $[M+Na]^+$.

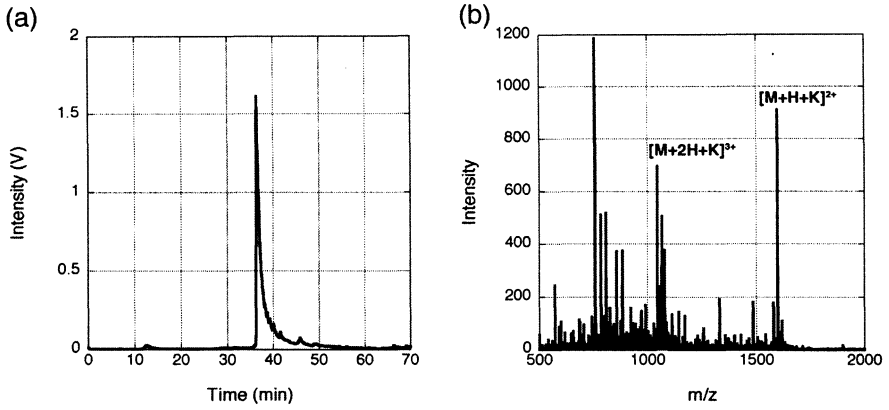

Figure A2.15. (a) HPLC and (b) ESI/TOF mass spectrometry data for $WG(DOG)_{10}$. Expected: 1597.5 $[M+H+K]^{2+}$, Observed: 1597.1 $[M+H+K]^{2+}$. Expected: 1065.3 $[M+2H+K]^{3+}$, Observed: 1065.1 $[M+2H+K]^{3+}$. Expected: 3194 $[M+K]^{+}$, Calculated: 3194.2 $[M+K]^{+}$.

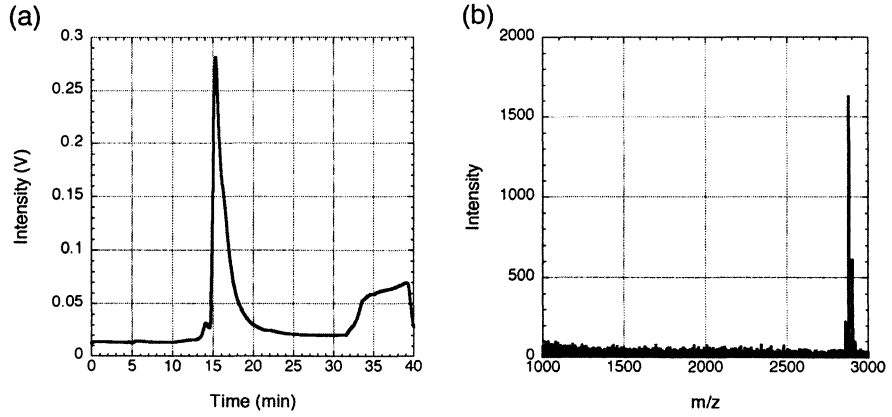

Figure A2.16. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $PKGLKG(PKG)_7LIG$. Expected mass: 2879.7 $[M+Na]^{+}$, Observed mass: 2879.7 $[M+Na]^{+}$.

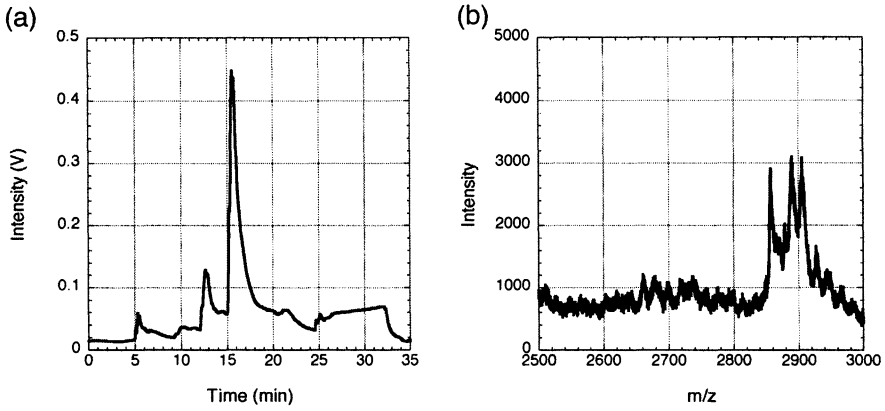

Figure A2.17. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $(DOG)_3LOG(DIG)_2(DOG)_4$. Expected mass: 2890.1 $[M+Na]^+$, Observed mass: 2889.4 $[M+Na]^+$.

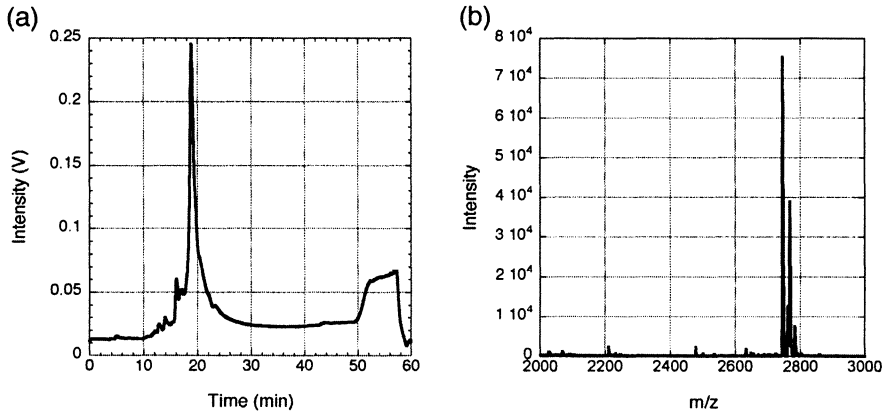

Figure A2.18. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $PIG(POG)_4(LOG)_2PIG(POG)_2$. Expected mass: 2746.0 $[M+Na]^+$, Observed mass: 2746 $[M+Na]^+$.

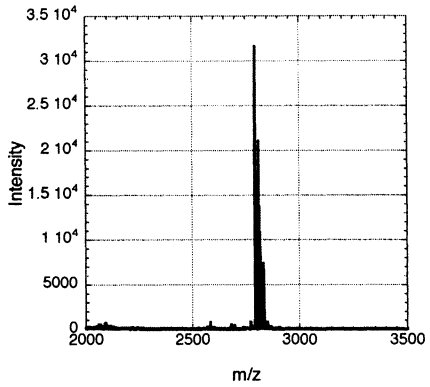

Figure A2.19. MALDI-TOF mass spectrometry data for $(LOGPOG)_5$. Expected mass: 2792.4 $[M+Na]^+$, Observed mass: 2792.7 $[M+Na]^+$.

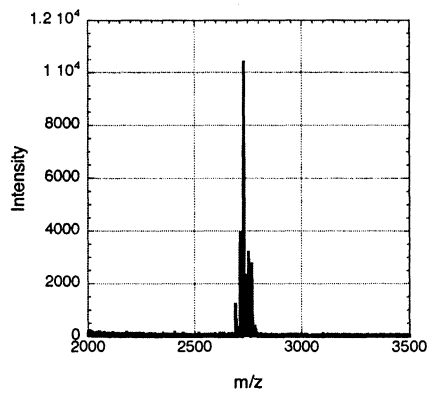

Figure A2.20. MALDI-TOF mass spectrometry data for $(POGPIG)_5$. Expected mass: 2728.4 $[M+K]^+$, Observed mass: 2728.5 $[M+K]^+$. This peptide was synthesized and purified previously by Varun Gauba therefore only mass spectrometry was performed in order to confirm purity.

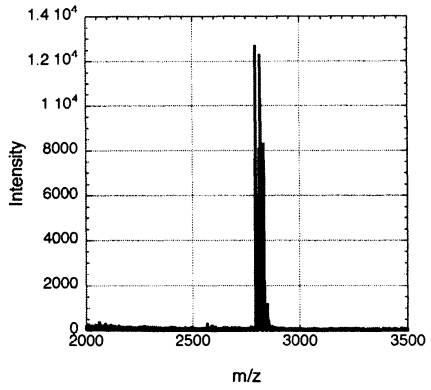

Figure A2.21. MALDI-TOF mass spectrometry data for $(LOGPIG)_5$. Expected mass: 2792.6 $[M+Na]^+$, Observed mass: 2792.6 $[M+Na]^+$. This peptide was synthesized and purified previously by Varun Gauba therefore only mass spectrometry was performed in order to confirm purity.

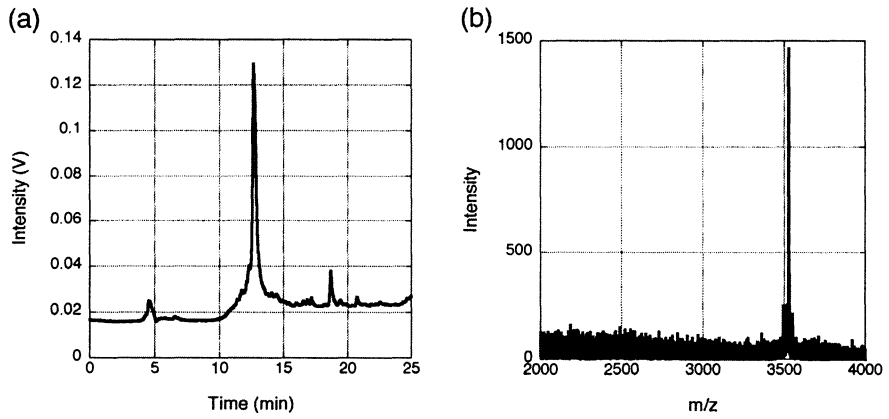

Figure A2.22. (a) HPLC and (b) MALDI-TOF mass spectrometry data for $(PRG)_4(POG)_4(EOG)_4$. Expected mass: 3524.65 $[M+H]^+$, Observed mass: 3525.30 $[M+H]^+$.

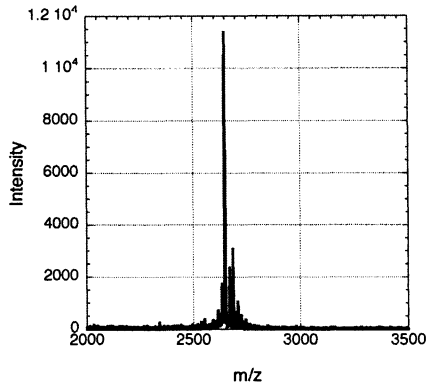

Figure A2.23. MALDI-TOF mass spectrometry data for $(EOG)_3(POG)_3(PRG)_3$. Expected mass: 2648.2 $[M+H]^+$, Observed mass: 2647.1 $[M+H]^+$. This peptide was synthesized and purified previously by Varun Gauba therefore only mass spectrometry was performed in order to confirm purity.

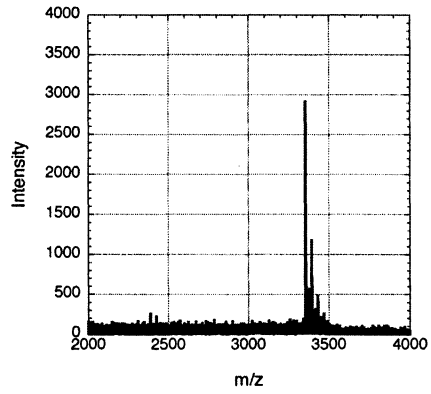

Figure A2.24. MALDI-TOF mass spectrometry data for $(DOG)_4(POG)_4(PKG)_4$. Expected mass: 3355.5 $[M+H]^+$, Observed mass: 3353.5 $[M+H]^+$. This peptide was synthesized and purified previously by Varun Gauba therefore only mass spectrometry was performed in order to confirm purity.

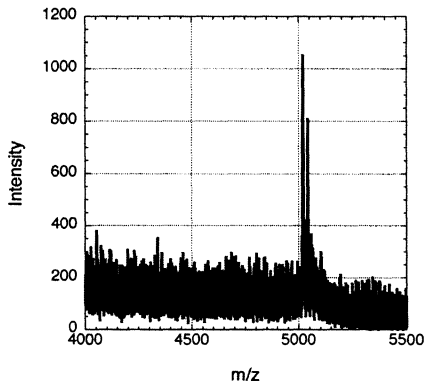

Figure A2.25. MALDI-TOF mass spectrometry data for $(DOG)_6(POG)_6(PKG)_6$. Expected mass: 5024.3 $[M+H]^+$, Observed mass: 5022.5 $[M+H]^+$. This peptide was synthesized and purified previously by Varun Gauba therefore only mass spectrometry was performed in order to confirm purity.

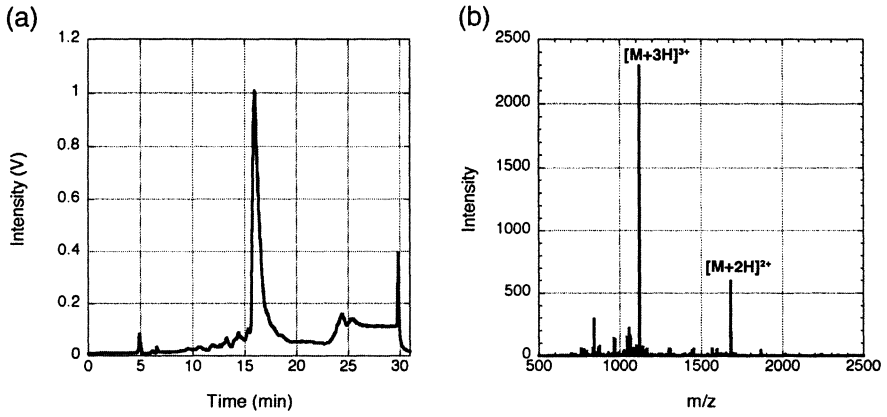

Figure A2.26. (a) HPLC and (b) ESI/TOF mass spectrometry data for $(PKG)_4(POG)_4(DOG)_4$. Expected: 1679.8 $[M+2H]^{2+}$, Observed: 1679.7 $[M+2H]^{2+}$. Expected: 1120.2 $[M+3H]^{3+}$, Observed: 1119.8 $[M+3H]^{3+}$. Expected: 3358.6 $[M+H]^+$, Calculated: 3358.6 $[M+H]^+$.