ABSTRACT

HOW CHARGED RESIDUES INFLUENCE THE THERMAL STABILITY OF COLLAGEN: A STUDY WITH NATURAL AND NON-NATURAL AMINO ACIDS

By:

Patrick Daniel Banzon

December 2020

Director of Thesis: William E. Allen, PhD

Major Department: Chemistry

Triple-helical collagens are key structural proteins in mammals. Their ubiquity and diverse functions drive our interest into understanding their behavior at a fundamental level. This thesis describes a reductionist approach using novel collagen-related peptides (CRPs), into which one or more electrical charges have been imparted at known positions. One series of CRPs includes fluorescent pyrene tags at their N-termini, directly adjacent to the charged residues lysine (Lys, K) and glutamic acid (Glu, E). When in close contact, the fluorophores form excimers that emit low-energy light. Monitoring of the excimer intensity shows that nucleation of collagen peptides is critically dependent on the charge location. Another series of CRPs features pH-independent (permanent) positive charge close to the peptide backbone, via a synthetic proline derivative called “Map.” When in close contact, repulsion between Map residues overwhelms the natural tendencies of the peptides to fold. CD and fluorescence investigations into the thermodynamic and kinetic behaviors of these CRPs have been

supplemented with computational analyses, to shed light on the deleterious role of charge in trimer formation.

HOW CHARGED RESIDUES INFLUENCE THE THERMAL STABILITY OF COLLAGEN: A STUDY WITH NATURAL AND NON-NATURAL AMINO ACIDS

A Thesis

Presented to the Faculty of the Department of Chemistry

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Chemistry

By:

Patrick Daniel Banzon

December 2020

HOW CHARGED RESIDUES INFLUENCE THE THERMAL STABILITY OF COLLAGEN: A STUDY WITH NATURAL AND NON-NATURAL AMINO ACIDS

By:

Patrick Daniel Banzon

APPROVED BY:

DIRECTOR OF THESIS: William E. Allen, PhD

COMMITTEE MEMBER: Colin S. Burns, PhD

COMMITTEE MEMBER: Andrew L. Sargent, PhD

COMMITTEE MEMBER: Nathan E. Hudson, PhD

CHAIR OF THE DEPARTMENT OF CHEMISTRY: Andrew T. Morehead, PhD

DEAN OF THE GRADUATE SCHOOL: Paul J. Gemperline, PhD

Pluralitas non est ponenda sine necessitate. - William of Ockham

ACKNOWLEDGEMENTS

Predictably, I'd like to start by recognizing the sacrifice my parents have made for me over the years. As humble 1st generation immigrants, you both had grand aspirations to bring stability to each of your families. This sense of unconditional love has also encumbered my life and I hope it pays off.

To Caroline, my younger sister, I'd be lying your observations did not influence my actions. I hope this project and my time spent on it demonstrates the good fruits of hard work and diligence.

To the department, six crucial years of personal growth were refined from my time here. To see many gifted individuals excel in their respective craft that they love is a perpetual source of inspiration. You all are the reason why entering higher academia is such a privileged investment.

To my colleagues during my time here: Jared, Abby, Megan, Mikayla, Brennan, Emily, Justin, and Alex, there's something special into directing each and everyone's talents and personalities into becoming better thinkers. Please stay in touch. I'm a terrible texter though.

To Lucy, as the biggest influence to be a better version of myself, your drive to excel and your love for others fuel me to keep working and stay positive when life blindsides you. There's nothing more certain than my desire to explore life with you.

To Dr. Allen, I came to you as a wide-eyed undergraduate, where my ambitions and dreams were as volatile as plenty of the solvents I accidently inhaled. I cannot emphasize this enough; I will never stop learning because of you. Not only have you've made me a better chemist, you've made me a better person. Watching your craft as a professor, but especially as a human being, simplifies life for me. As my time in this program comes to a close, I haven't fully realized what you have done for me and my peers. Still, I'll attempt to express my gratitude to its fullest extent.

To those who will enter a similar position I once was, here's a few tips. Don't fight mother nature. Write your experimental sections early. Assist your peers over competing with them. Bring morale-boosting donuts to the office from time to time. Be patient with your growth; the program is doing its job. Hope this helps.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

A	Alanine
Amp	4-aminoproline
a.u.	Arbitrary Units
Ac	Acetyl group
Ala	Alanine
AzG	Aza-glycine
AzP	γ-aza-proline
˚C	Degrees in Celsius
C^{γ}	Gamma carbon (3^{rd} carbon)
CD	Circular dichroism
CRP	Collagen-related peptide
E	Glutamic acid
ESI+ QTOF	Electrospray ionization (positive) – quadrupole time of flight
FL	Fluorescence
Flp	(2S,4R)-4-Fluoroproline
flp	(2S,4S)-4-Fluoroproline
G	Glycine
ΔG°	Change in transition free energy
Glu	Glutamic acid
Gly	Glycine
H-bond	Hydrogen bond
HPLC	High performance liquid chromatography
Hyp	(2S,4R)-4-Hydroxyproline
hyp	(2S,4S)-4-Hydroxyproline
K	Lysine
Lys	Lysine
mL	Milliliter
mM	Millimolar

M/Z	Mass-to-charge
NMR	Nuclear magnetic resonance
O	(2S,4R)-4-Hydroxyproline
OI	Osteogenesis imperfecta
P	Proline
POG	Pro-Hyp-Gly
P4H	Proyl-4-hydroxylase
PPII	Polyproline II
ppm	Parts per million
Pro	Proline
Pyr-(0)	1-pyrenyl unit, directly attached to peptide
Pyr-(2)	1-pyrenyl unit, attached to peptide with a 2-carbon spacer
PyrATS	Pyrene-appended trimeric systems
R_f	Retention factor
s	Second
Sar	Sarcosine
SPPS	Solid phase peptide synthesis
$\boldsymbol{\Theta}_{224\ nm}$	Ellipticity at 224 nm
T	Temperature
$t_{½}$	Half-life
T_m	Melting (unwinding) temperature
Xaa	Amino acid in the X position of an Xaa-Yaa-Gly repeat in a collagen system
Yaa	Amino acid in the Y position of an Xaa-Yaa-Gly repeat in a collagen system
ΔH°	Change in enthalpy
µL	Microliter
µM	Micromolar
UV-Vis	Ultraviolet-visible spectroscopy
v/v	Percentage by volume

CHAPTER 1: Introduction to Collagen Structure, Function, and Elucidation

1.1 The Dominant Prevalence of Collagen

As the most abundant proteins in mammals, collagens have been conserved throughout time. Collagen's widespread and numerous biological functions drive its ubiquity within the evolution of mammalian systems. Remarkably, the presence of collagen has been detected in soft tissue fossils of a 68-million-year-old *Tyrannosaurus rex*.[1,2] If the discovery holds, it would be oldest detected protein to date, though certain reviews challenge these claims.[3,4] Otherwise, there have been widely accepted studies where intact collagen has been preserved in 600,000-year-old mastodon remnants to highlight its prevalence throughout our evolutionary history.[4]

The numerous biological functions of collagen include its debatably most important role: a scaffolding element of the mammalian extracellular matrix. Collagen accounts for three-quarters of dry weight in the human skin.[5]

1.2 Applications in Biomedicine & Materials

The desire to study collagen has largely split into two motives: understanding its role in disease treatment and its role in human pathology. As the pharmaceutical industry inches forward onto assigning proteins as viable vehicles for drug delivery, the interest in utilizing collagen has risen in recent years. Inspired from this idea, drug therapeutics were casted in collagenic shields to treat glaucoma.[5] This method allowed for sustained drug delivery, as common treatments often resulted in poor patient adherence. Another example integrates synthetic collagen mimics with gold nanoparticles to induce platelet aggregation (Figure 1.1), a feature where natural collagen is heavily involved.[6]

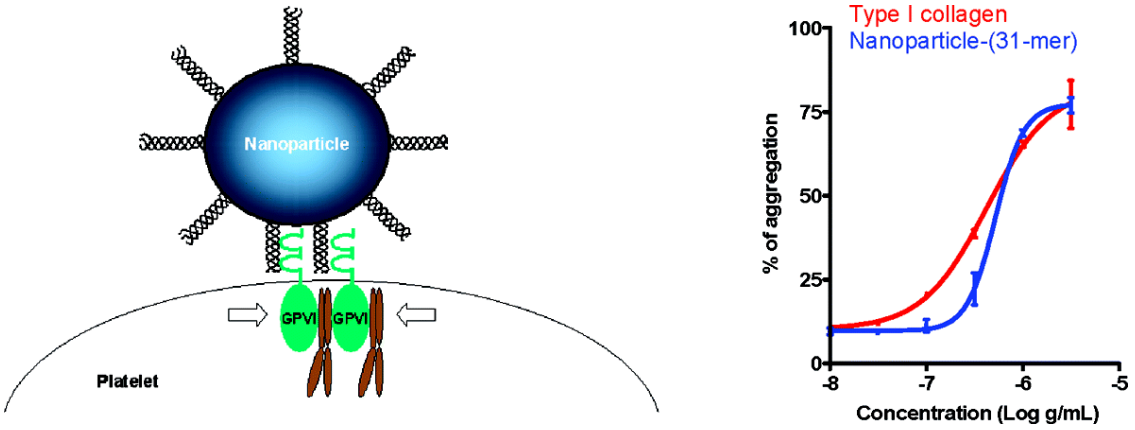

Figure 1.1: Clot inducing collagen-laced gold nanoparticles[6]

Decades of research have led to the deeper understanding of collagen-related pathologies. To highlight, genetic mutations create structural instabilities that cause a spectrum of diseases. For example, osteogenesis imperfecta (OI) and osteoporosis are caused by genetic mutations into type I collagen. Symptoms include brittle bones and often affect other tissues, causing thin skin, blue sclerae, and abnormal teeth. Other diseases linked to collagen types are tabulated from Raines et al (Table 1.1).[7]

Table 1.1: Vertebrate collagens, their distributions, and pathologies[7]

Type	Class	Composition	Distribution[b]	Pathology[c]
I	Fibrillar	$\alpha1[I]_2\alpha2[I]$	Abundant and widespread: dermis, bone, tendon, ligament	OI, Ehlers–Danlos syndrome, osteoporosis
II	Fibrillar	$\alpha1[II]_3$	Cartilage, vitreous	Osteoarthrosis, chondrodysplasias
III	Fibrillar	$\alpha1[III]_3$	Skin, blood vessels, intestine	Ehlers-Danlos syndrome, arterial aneurysms
IV	Network	$\alpha1[IV]_2\alpha2[IV]$ $\alpha3[IV]\alpha4[IV]\alpha5[IV]$ $\alpha5[IV]_2\alpha6[IV]$	Basement membranes	Alport syndrome
V	Fibrillar	$\alpha1[V]_3$ $\alpha1[V]_2\alpha2[V]$ $\alpha1[V]\alpha2[V]\alpha3[V]$	Widespread: bone, dermis, cornea, placenta	Ehlers-Danlos syndrome
VI	Network	$\alpha1[VI]\alpha2[VI]$ $\alpha3[VI]$[d] $\alpha1[VI]\alpha2[VI]$ $\alpha4[VI]$	Widespread: bone, cartilage, cornea, dermis	Bethlem myopathy
VII	Anchoring fibrils	$\alpha1[VII]_2\alpha2[VII]$	Dermis, bladder	Epidermolysis bullosa acquisita
VIII	Network	$\alpha1[VIII]_3$ $\alpha2[VIII]_3$ $\alpha1[VIII]_2\alpha2[VIII]$	Widespread: dermis, brain, heart, kidney	Fuchs endothelia corneal dystrophy
IX	FACIT[e]	$\alpha1[IX]\alpha2[IX]\alpha3[IX]$	Cartilage, cornea, vitreous	Osteoarthrosis, multiple epiphyseal dysplasia
X	Network	$\alpha1[X]_3$	Cartilage	Chondrodysplasia
XI	Fibrillar	$\alpha1[XI]\alpha2[XI]\alpha3[XI]$	Cartilage, intervertebral disc	Chondrodysplasia, osteoarthrosis
XII	FACIT	$\alpha1[XII]_3$	Dermis, tendon	—
XIII	MACIT	—	Endothelial cells, dermis, eye, heart	—
XIV	FACIT	$\alpha1[XIV]_3$	Widespread: bone, dermis, cartilage	—
XV	MULTIPLEXIN	—	Capillaries, testis, kidney, heart	—
XVI	FACIT	—	Dermis, kidney	—
XVII	MACIT	$\alpha1[XVII]_3$	Hemidesmosomes in epithelia	Generalized atrophic epidermolysis bullosa
XVIII	MULTIPLEXIN	—	Basement membrane, liver	Knobloch syndrome
XIX	FACIT	—	Basement membrane	—
XX	FACIT	—	Cornea (chick)	—
XXI	FACIT	—	Stomach, kidney	—
XXII	FACIT	—	Tissue junctions	—
XXIII	MACIT	—	Heart, retina	—
XXIV	Fibrillar	—	Bone, cornea	—
XXV	MACIT	—	Brain, heart, testis	Amyloid formation?
XXVI	FACIT	—	Testis, ovary	—
XXVII	Fibrillar	—	Cartilage	—
XXVIII[f]	—	—	Dermis, sciatic nerve	Neurodegenerative disease?

1.3 Elucidating Collagen Structure

Since the 1940's, the essence of collagen structure has been thoroughly debated. From proposing singly extended polypeptide chains[8,9] to mandating a *trans* conformation at every peptide bond[10], various spectroscopic techniques have been employed to acquire a greater understanding of its folding. These findings did not go without its spectroscopic limitations.

Native collagen is relatively large, insoluble, repetitive, and hierarchically complex, thus thwarting most biochemical and biophysical studies. Therefore, the physiochemical elucidation of collagen structure often relies on a reductionist approach using simpler collagen-related peptides (CRPs).[10]

Elucidating collagen structure at its fundamental level became highly approachable using CRPs. To emphasize, a high-resolution crystal structure of a CRP determined the presence of interstrand hydrogen bonds among $N–H_{(Gly)} \bullet\bullet\bullet O{=}C_{(Xaa)}$ atoms (Figure 1.2), a property that has been deemed the most essential electrostatic interaction for the triple helix to form.[11]

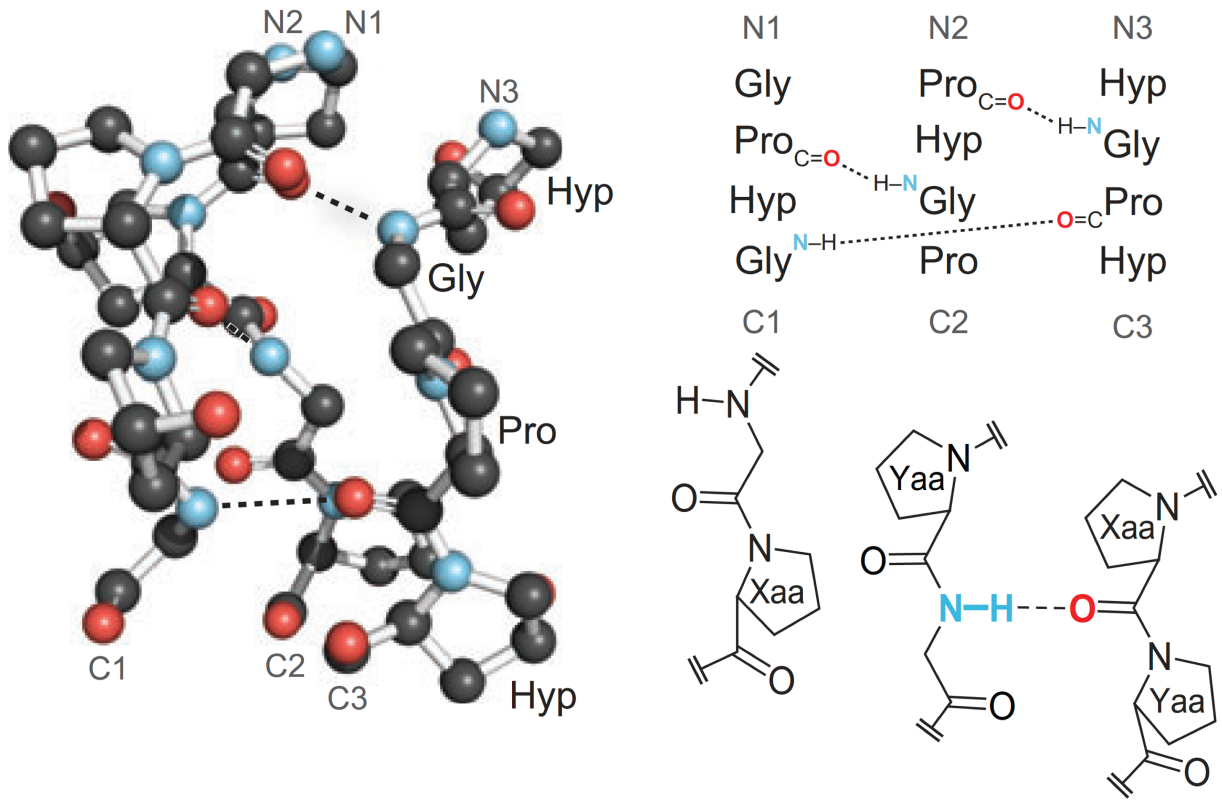

Figure 1.2: Ball-and-stick & staggered representations depicting interstrand hydrogen bonding.[7]

Because of this reductionist approach, it has been broadly realized that collagen structure stems from key underlying principles. Collagen peptides were revealed to be three

left-handed polypeptide strands, woven in a staggered pattern to create a higher-order right-handed triple helix. Often titled as polyproline type II (PPII) helices, it has also been established that underlying amino acid repeats of Xaa-Yaa-Gly must be the basis on the primary structure of these residues, where Xaa and Yaa represent the two adjacent amino acids to glycine (Gly, G). Xaa and Yaa often vary when sequencing natural collagen, but glycine is essential for the proper folding to occur.

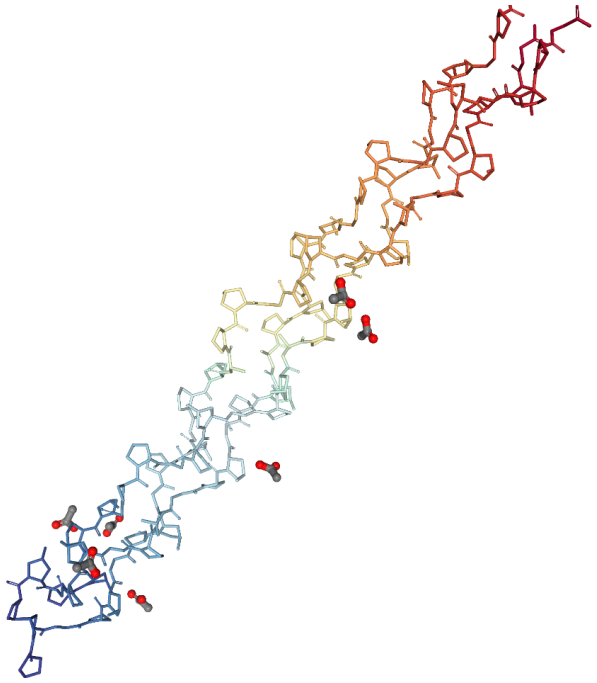

Figure 1.3: Computational model of a CRP with limited hydration interactions[12]

1.4 Structural Factors that Lead to Collagen Triple Helicity

The natural abundance of collagen is interwoven with its broad number of biological characteristics. Desirable characteristics such as thermal stability, biomolecular interactivity, and mechanical strength/flexibility underpin the drive to understand the basic structural unit of collagen.

The most important force in determining the structure of collagen is interstrand hydrogen bonding. The ubiquity of collagen makes such bonds the most abundant type of amide–amide hydrogen bonds in the kingdom Animalia. Using various CRPs, estimates of each hydrogen bond strengths fall around ΔH° = -2.0 kcal/mol[13], comparatively close to strengths of ΔH° = -1.4 kcal/mol in natural collagen[14]. Glycine substitutions, observed through x-ray crystallography, show the relationship between this residue and interstrand hydrogen bonding. On a structural basis, a central glycine substitution for alanine (Ala, A) causes a disruption in the formation of the triple helix.[15] This is due to the inability of the central triplet to form interstrand hydrogen bonding as effectively (Figure 1.4).

Figure 1.4: Disruption of a triple helix from the central triplet[7]

In Xaa-Yaa-Gly repeats, glycine mutations lead to the most damaging diseases, like previously mentioned osteogenesis imperfecta (OI).[15] To contrast, the Xaa and Yaa positions can be any natural residue. However, the most common residues to take the Xaa and Yaa positions of natural collagen are proline (Pro, P) and proline derivatives such as (2S,4R)-4-hydroxyproline (Hyp, O), at about 22% of all residues.[16] Altogether, Pro-Hyp-Gly (POG) triplets stand as the most common tri-residue repeats in collagen, accounting for 10.5%.[16]

Perhaps the prevalence of proline and its derivatives lies in its conformational restrictions. Higher order PPII folds are assisted by preorganization, thus decreasing the entropic cost.[17] However, proline and its derivatives also have detrimental energetic

consequences. While preorganization is an energy saving attribute, proline forms tertiary amides within proteins. Tertiary amides further stabilize the *cis* conformation, whereas collagen amide bonds are *trans*. Therefore, a relatively slow isomerization process is required to form a triple helix. To explore this relationship, results from CRPs containing isosteres such as *N*-methylalanine[18] and a *trans*-locked analog, Ψ[(*E*)CH═C)]-proline[19] (Figure 1.5), displayed destabilized helices. While these isosteres were designed to not interfere with interstrand hydrogen bonding, stability was still reduced, highlighting the difficulty of predicting trimer stability.

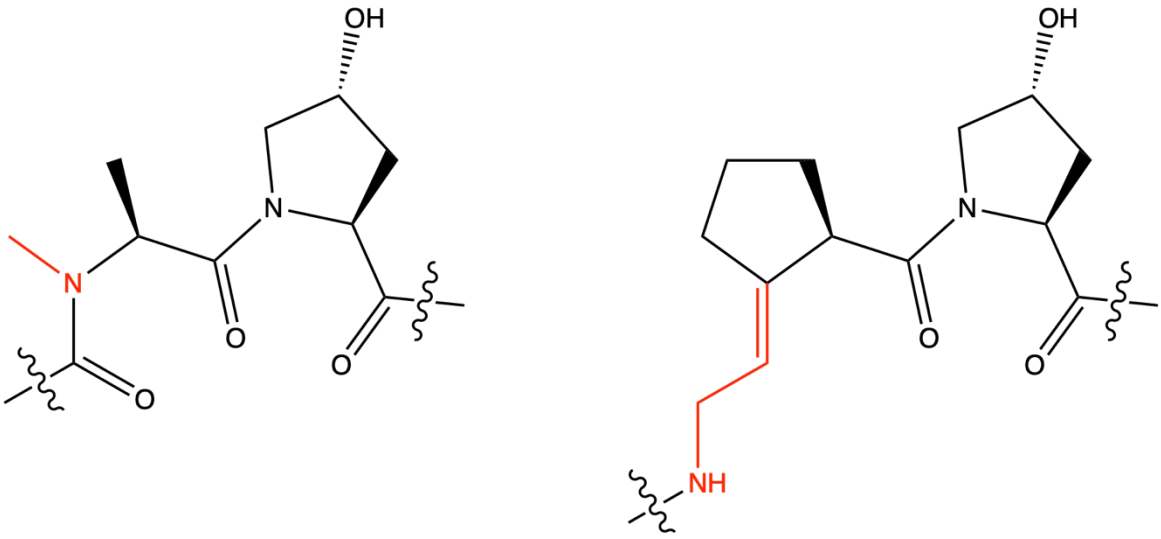

Figure 1.5: Isosteres, *N*-methylalanine and Ψ[(*E*)CH═C)]−Proline, to study the influence of *cis*-*trans* isomerization on collagen folding (modifications highlighted in red).[18,19]

Proline derivatives in natural collagen occur from post-modification. Commonly, proline residues in the Yaa position of collagen triplets are modified via prolyl 4-hydroxylase (P4H) to create (2S,4R)-4-hydroxyproline. In this location, the hydroxylation of proline dramatically stabilizes the collagen triple helix. Interestingly, the stereochemistry of the hydroxyl-group is vital for increasing stability. CRPs containing (2S,4S)-4-hydroxyproline (hyp) offer no formation

of the triple helix.[20,21] Initially, it was proposed that the *R*-configuration of the hydroxyl-group allowed for water-mediated hydrogen bonding along the trimers, often coined as water bridges. Subsequent studies removed that hydrogen bond capability by utilizing CRPs with (2S,4R)-4-fluoroproline (Flp) and suggested its lack of influence in helical stability, as it formed even stronger helicies.[22,23] Notably, CRPs containing isomeric (2S,4S)-4-fluoroproline (flp) failed to even form a triple helix.[24]

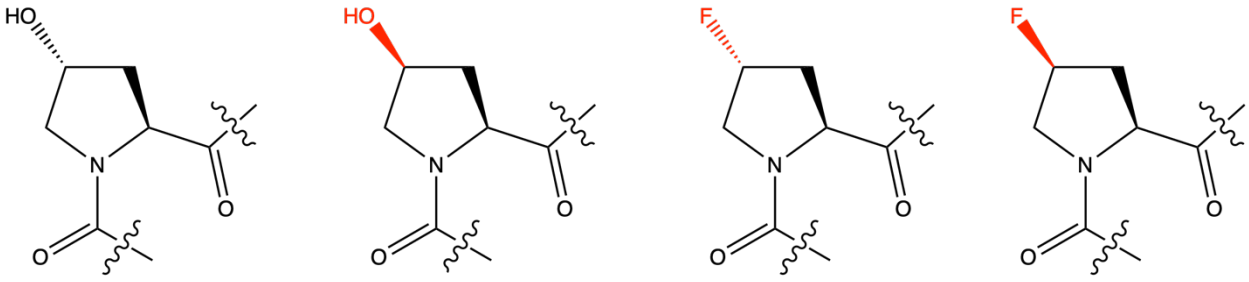

Figure 1.6: Proline derivatives to investigate the (2S,4R)-4-hydroxyproline role in triple-helical stability. From left to right: Hyp, hyp, Flp, flp (modifications highlighted in red).[20-24]

The failure of the "water bridge theory" became clear when sequencing data of various types of collagen became available. The infrequency of triplet repeats of Xaa-Hyp-Gly suggested that water bridges play little to no role in helical stability. Therefore, another explanation was needed for the relationship between Hyp and stability. When comparing CRPs that contain Hyp and Flp to hyp and flp in the Yaa position, helical formation occurred exclusively with *R*-proline derivatives. Yet, only one of those derivatives contain water bridging capability. Therefore, it was established that a stereoelectronic effect drove triple helical stability, as opposed to a purely inductive one.[24]

The third carbon of the pyrrolidine ring (C^{γ}) assigns a preference into ring puckering. Proline and its derivatives prefer one of two major pyrrolidine ring conformations, which are

termed C$^{\gamma}$-*exo* and C$^{\gamma}$-*endo*.[25] These preferences can be either beneficial or deleterious in collagen folding depending on the position in the Xaa-Yaa-Gly repeat. This is due to the optimization of torsion angles of a collagen twist, where the Xaa position prefers C$^{\gamma}$-*endo* and the Yaa position prefers C$^{\gamma}$-*exo* (Figure 1.7). Limited by gauche and steric effects, residues optimally turn to favor interstrand hydrogen bonding. To assist further, this optimization preorganizes the polyproline-II helices.

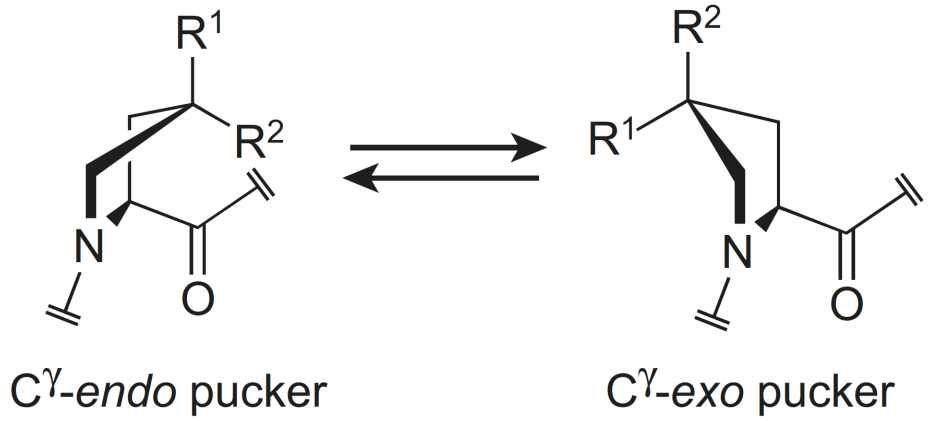

Figure 1.7: Major conformations of the pyrrolidine ring[7]

Stereoelectronics of these proline derivatives also influence the isomerization equilibrium constant ($K_{trans/cis}$), through n → π* interactions. When defining the effect of n → π* interactions, the carbonyl oxygen of one peptide bond (O_{i-1}) donates lone pair electron density into the antibonding orbital of the succeeding amide bond carbonyl ($C_{i'}{=}O_i$). In collagen, n → π* interactions can only occur if the peptide bond is *trans*. Therefore, this shifts main chain $K_{trans/cis}$ values with appropriate torsion angles which stabilizes *trans* conformations by an estimated $\Delta G^{\circ} = -0.7$ kcal/mol.[26,27]

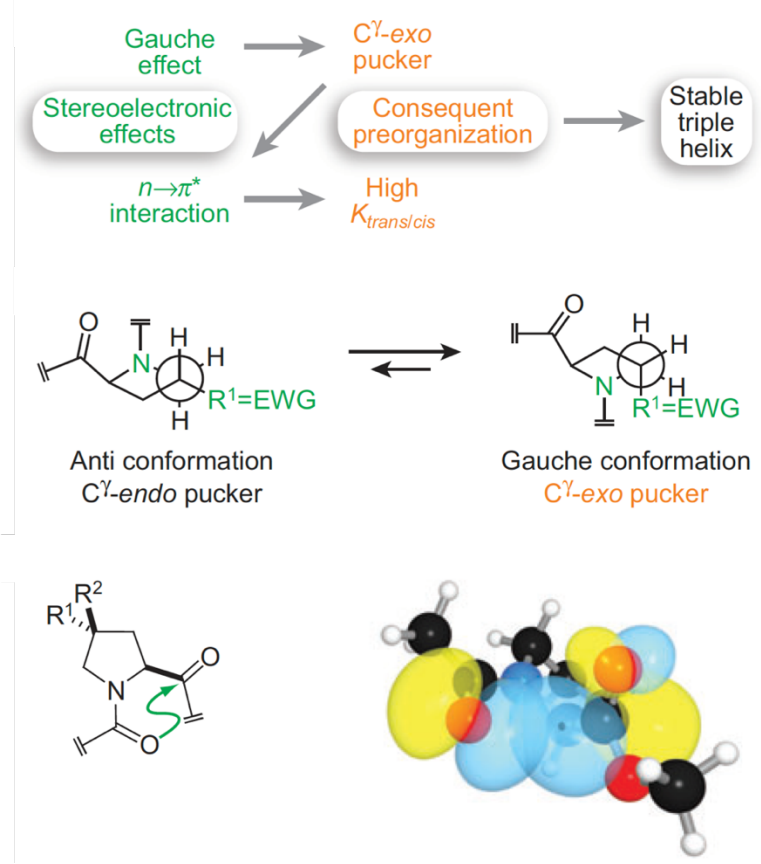

Figure 1.8: Stereoelectronic effects hold many responsibilities in favoring triple-helical stability.[7]

1.5 Modifications to Increase Triple-Helical Stability

Synthetic chemistry allows molecular editing beyond nature's capability. With this notion, there have been many attempts to modify CRPs in a way to prioritize triple-helical stability. Such attempts include backbone amendments, side chain alterations, and terminal capping with novel molecules. These studies can generate thermally resistant CRPs and further clarify our understanding on the physiochemical basis of collagen.

One attempt is the implementation of aza-glycine (AzG). This modification replaced an α-carbon of a glycine with an amino group (Figure 1.9). When AzG was substituted into central triplets of Ac-$(POG)_7$-NH_2 templates, the number of interstrand hydrogen bonds increased. As a result, it led to hyperstable triple helices, as judged by an approximated 10 ˚C increase in T_m.[28]

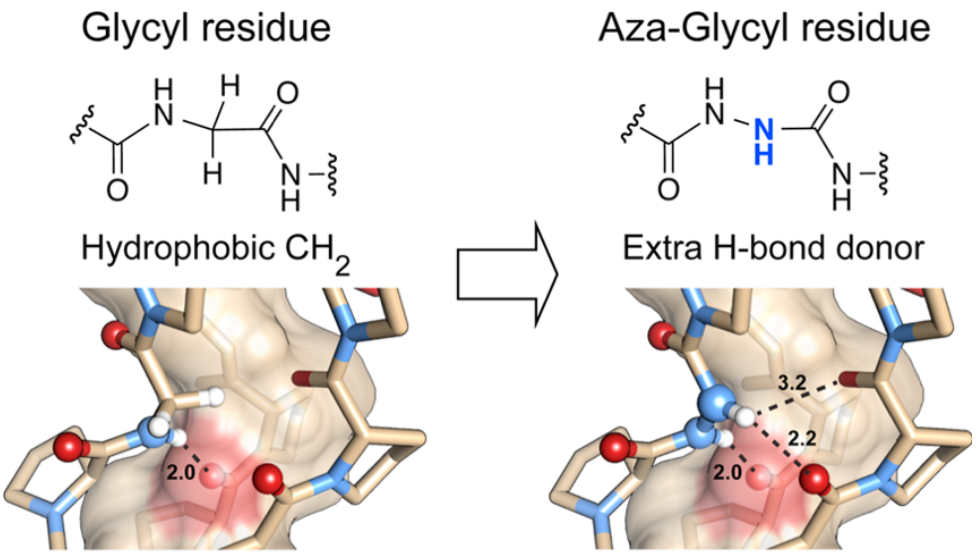

Figure 1.9: Aza-glycine confers hyperstability.[28]

Accordingly, AzG greatly enhanced enthalpic contributions to ΔG°, as the additional hydrogen bond donors were able to interact with adjacent interstrand carbonyls. Supplementary work studied AzG effects when substituted in CRPs in variety.[29,30] At its synthetic plateau, AzG-bearing CRP formed trimers with as few as 12 residues (4 repeats).[30]

Another attempt into improving triple helical stability utilized γ-aza-proline (AzP). This analogue residue replaced the γ-carbon of the proline ring for a nitrogen, allowing for hyperstability and further functionalization (Figure 1.10).[31-33] When substituted into central triplets of an Ac-$(POG)_7$-NH_2 templates, it showed little interference with helical stability. To add, the presence of imidazolidine ring added facile synthetic avenues to further modify the residue.[32]

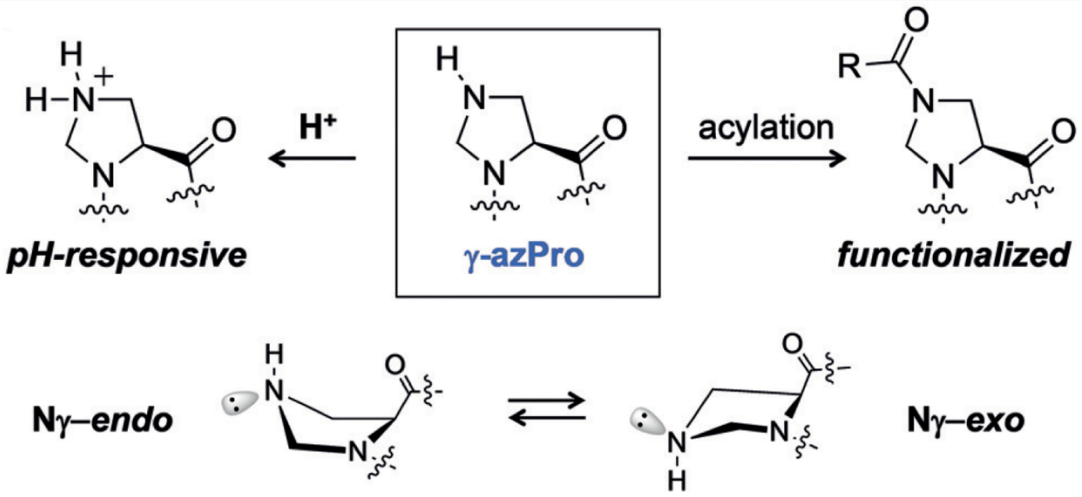

Figure 1.10: γ-Aza-proline confers functionalizability.[31]

In prior work from the Allen lab performed by Jared Keever, pyrene fluorescent tags were covalently attached to the N-termini of CRPs. Intending to modify CRPs to create hyperstable helices, this investigation highlighted key relationships among collagen folding, hydrophobicity, and critical micelle concentration. Peptide-peptide assembly driven by the association of lipophilic groups has been an understood feature and pyrene potentially offers this, along with fluorescent capabilities.[33] Pyrene has the ability to exhibit π - π stacking and potentially emit light at lower energies as a result. Broadly speaking, π - π stacking is a phenomenon where overlapping of π-electrons occur. This is particularly likely in water, where entropy favors aggregation.[34, 35] Individual pyrene molecules emit light from 370-430 nm upon excitation. However, pyrene exhibits excited state dimer or "excimer" at a certain proximity, emitting at approximately 485 nm (Figure 1.11).

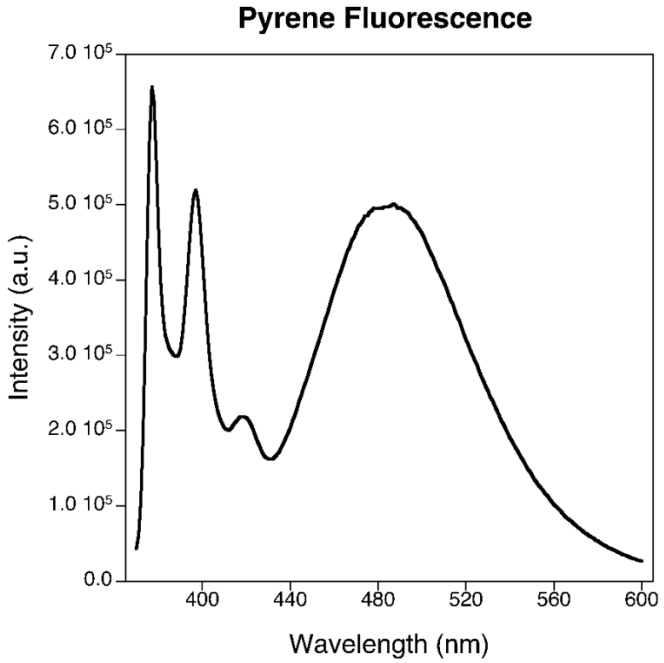

Figure 1.11: Typical fluorescence (FL) spectra of pyrene and its two-state emission.

When attached to the N-termini of CRPs, the intensity of the excimer should be proportional to the abundance triple helices in solution. Since excimers are distance-dependent, the intensity of excimer emission should be artificially inflated in conditions where the CRPs form triple helices. In prior work, CRPs containing this excimer ability were able to exhibit similar melting curves in both CD and FL spectroscopy (Figure 1.12), even exhibiting meting temperatures values (T_m) withing 10% of each other. Upon further inspection though, the curves from both studies differ in sharpness. This difference will be discussed at more depth in the following chapter.

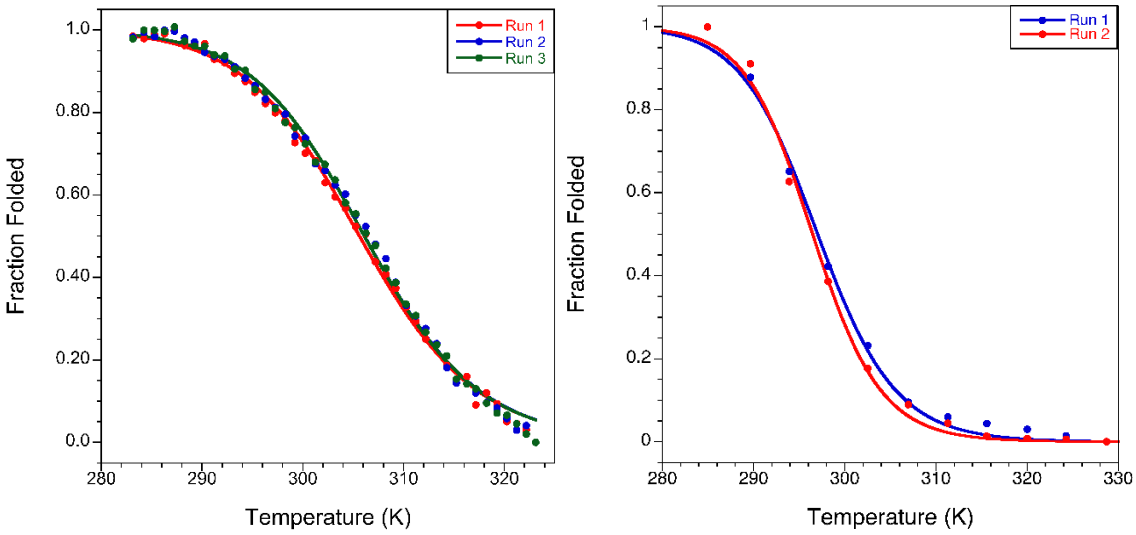

Figure 1.12: Sample CD & FL data of PyrATS system **Pyr-(0)-AcKK**, reported by Jared Keever

To briefly summarize, the residues alanine, glutamate, and lysine flanked the (Pro-Hyp-Gly)$_7$ core of the peptide monomers, at positions 1 and 23 in the primary sequence to impart varying degrees of "fraying." Triple helices derived from such peptides show a broad excimer emission at ~485 nm, indicative of close approach between pyrene units. The excimer intensity decreases upon heating. For sequences with a lysine residue adjacent to the pyrene probe, melting temperatures (T_m) obtained from fluorescence are within 3 ˚C of those determined using CD, even though the CD samples were typically four times more concentrated. The most significant disagreement between the two techniques (T_m = 65 ˚C and 41 ˚C for CD and fluorescence, respectively) was found for the peptide with the least helix-disruptive residue (i.e., Ala) adjacent to the fluorophore.

This system was computationally modeled in a water continuum using B3LYP and M06-2X density functionals. The former predicts the pyrene moieties to interact in an edge-to-face manner, while the latter favors parallel stacking. Calculated energies from B3LYP treatments match experimental trends better than the energies from M06-2X. It was determined that

pyrene probes offered a degree of additional helical stabilization when comparing the fluorescent capable systems with the control, predictably through an added hydrophobic effect among individual pyrenes. Perhaps more interestingly, fluorescent spectroscopy offered a different interpretation into the stability of the triple helices. Altogether, this investigation allowed for detailed exploration on the fundamental forces that act on collagen helices and their inability to override the dominant force of the hydrophobic effect.

1.6 Goals of this Research

As reviewed previously, collagen structure has been described in detail. Additionally, previous work in this lab was discussed to underscore our attempts for a deeper physiochemical understanding of collagen utilizing fluorescent probes. The following chapters of this thesis will highlight successive work of our synthetic strategies to develop and investigate relevant CRPs in order to gain further insight into role of charged residues in collagen structure. Parts of the thesis will discuss the exploration into adding fluorescent probes into various locations of CRPs and the implementation of a positively-charged proline derivative and its role in the formation of triple-helical formation.

CHAPTER 2: Further Developments into PyrATS: Pyrene-Appended Trimeric Systems

2.1 New Findings into Collagen Trimers Monitored by Pyrene Probes

Previous work on PyrATS described the synthesis and preliminary fitting of the thermodynamic and kinetics measurements of each system. This chapter aims to summarize new interpretations on the efficacy of pyrene probes as a monitor for helix unwinding and the countering effect of charged residues on helix formation.

From previous work, well characterized CRPs have been tagged at their N-termini with the fluorophore, pyrene. Since pyrene exhibits an excimer ability, this allows for alternative investigations into several fundamental questions in collagen research, through fluorescence spectroscopy. Self-quenching investigations between fluorescein dyes have been implemented similarly.[36] With this approach, concentration dependence, folding directionality, and local fraying, were addressed a relatively more sensitive and cost-effective manner than common CD experiments. Additionally, the pyrene probes provided a situational increase in thermal stability. Initially, the probes were intended to monitor the kinetics of triple-helical unwinding. Successive work, however, determined a deeper, perhaps more noteworthy understanding of the pyrene probes effect on the CRPs as a whole.

2.2 Peptide Design and Characterization

Several CRPs, derived from the well-studied Ac-(Pro-Hyp-Gly)$_7$-NH_2, were made through Fmoc-based solid-phase peptide synthesis (SPPS). Pyrene was appended to each N-terminus by using pyrene-containing carboxylic acids and subjecting each reagent to similar SPPS strategies. This is essentially a formal substitution of a hydrogen at the terminal acetyl group (Ac) with pyrene, with or without carbon-spacers. To add, flanking amino acids were attached at each

end of the (Pro-Hyp-Gly)$_7$ backbone to address two potential shortcomings upon spectroscopic analyses of these systems. First, pyrene as a polyaromatic hydrocarbon may introduce solubility issues when dissolved in aqueous buffers. Secondly, higher-order aggregation is likely to appear and may distort experimental results. Therefore, several CMPs featured at least one charged residue, like lysine (Lys, K) and glutamate (Glu, E), to counter these limitations.

Table 2.1: List of synthetic pyrene-labeled collagen systems, with associated abbreviations

Abbreviation	**Formulae**
AcKK	H-(CH$_2$)CO-K(POG)$_7$K-NH$_2$
Pyr-(0)-AcKK	Pyrene-(CH$_2$)CO-K(POG)$_7$K-NH$_2$
Pyr-(2)-AcKK	Pyrene-(CH$_2$)$_3$CO-K(POG)$_7$K-NH$_2$
Pyr-(2)-AcKA	Pyrene-(CH$_2$)$_3$CO-K(POG)$_7$A-NH$_2$
Pyr-(2)-AcAK	Pyrene-(CH$_2$)$_3$CO-A(POG)$_7$K-NH$_2$
Pyr-(2)-AcEA	Pyrene-(CH$_2$)$_3$CO-E(POG)$_7$A-NH$_2$
Pyr-(2)-AcKSarK	Pyrene-(CH$_2$)$_3$CO-K(POG)$_3$POSarcosine(POG)$_3$K-NH$_2$
Pyr-(2)-AcKSar$_3$K	Pyrene-(CH$_2$)$_3$CO-KPOSarcosine(POG)$_2$POSarcosine(POG)$_2$POSarcosineK-NH$_2$

Each peptide was purified by reverse-phase high-performance liquid chromatography (HPLC) while monitored by UV-vis detection at 355 nm. Major fractions collected from HPLC were then analyzed through mass spectrometry (MS) to characterize the peptide. Once isolated, each system was dissolved in phosphate-buffered saline (PBS; pH 7.4) solutions stored at 4 ˚C, at concentrations ranging from 5 to 200 μM for spectroscopic studies.

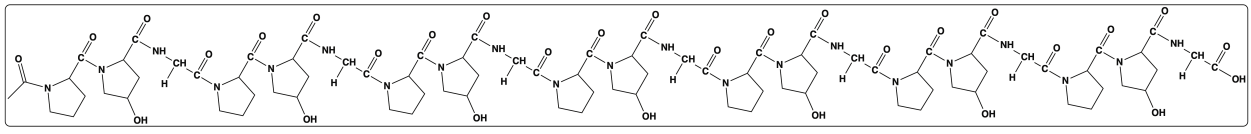

Figure 2.1: Well-studied Ac-(Pro-Hyp-Gly)$_7$-NH$_2$ CRP, often abbreviated Ac(POG)$_7$NH$_2$

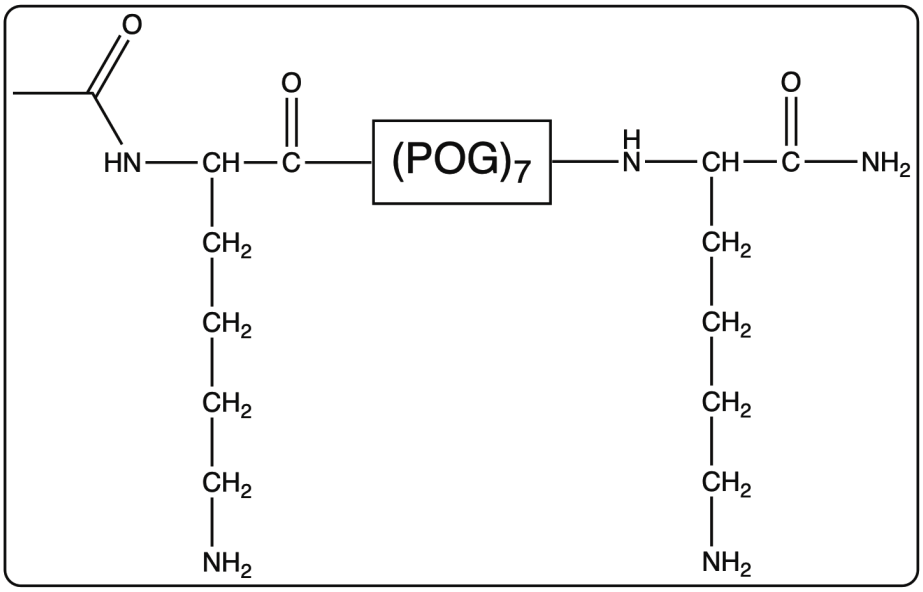

Figure 2.2: The **AcKK** system

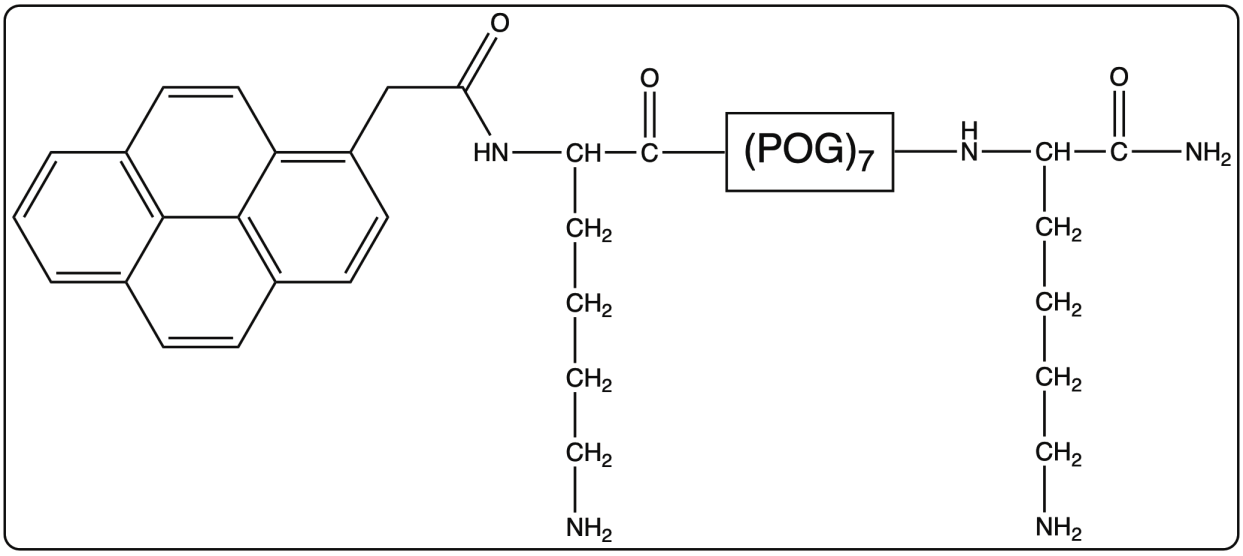

Figure 2.3: The **Pyr-(0)-AcKK** system

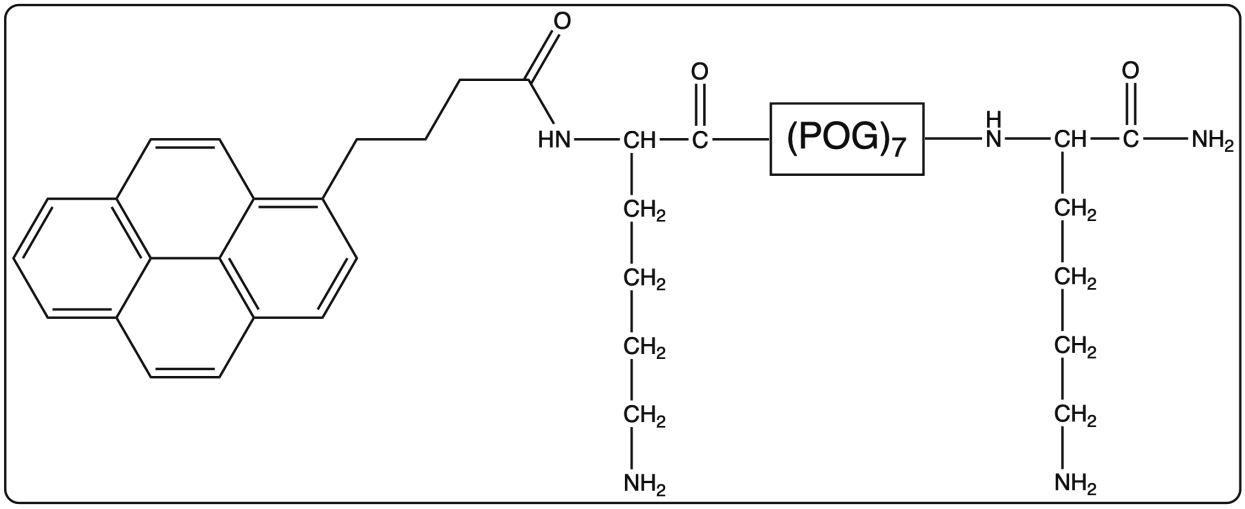

Figure 2.4: The **Pyr-(2)-AcKK** system

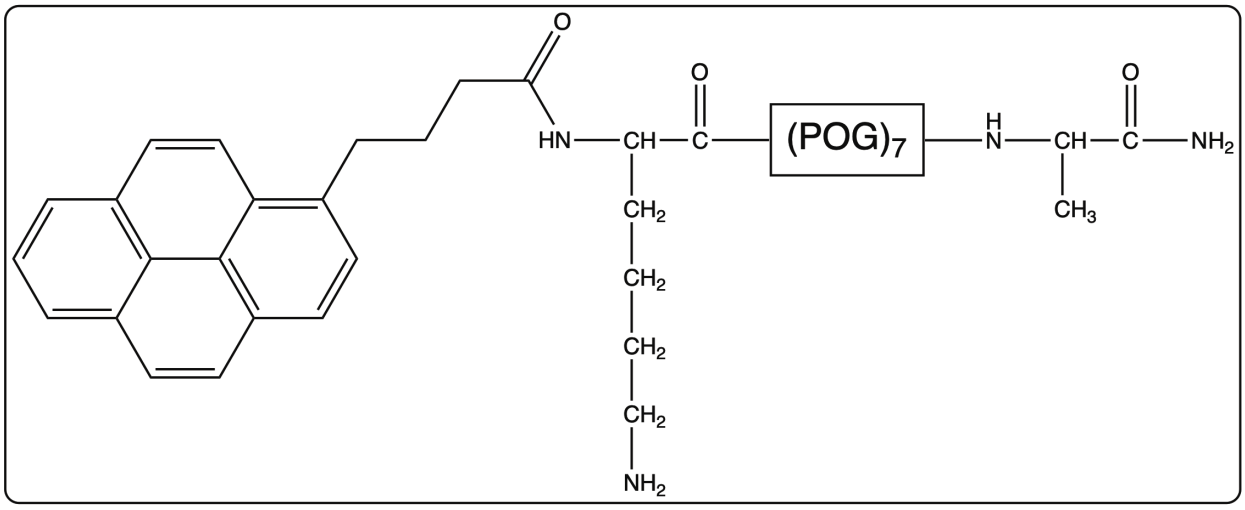

Figure 2.5: The **Pyr-(2)-AcKA** system

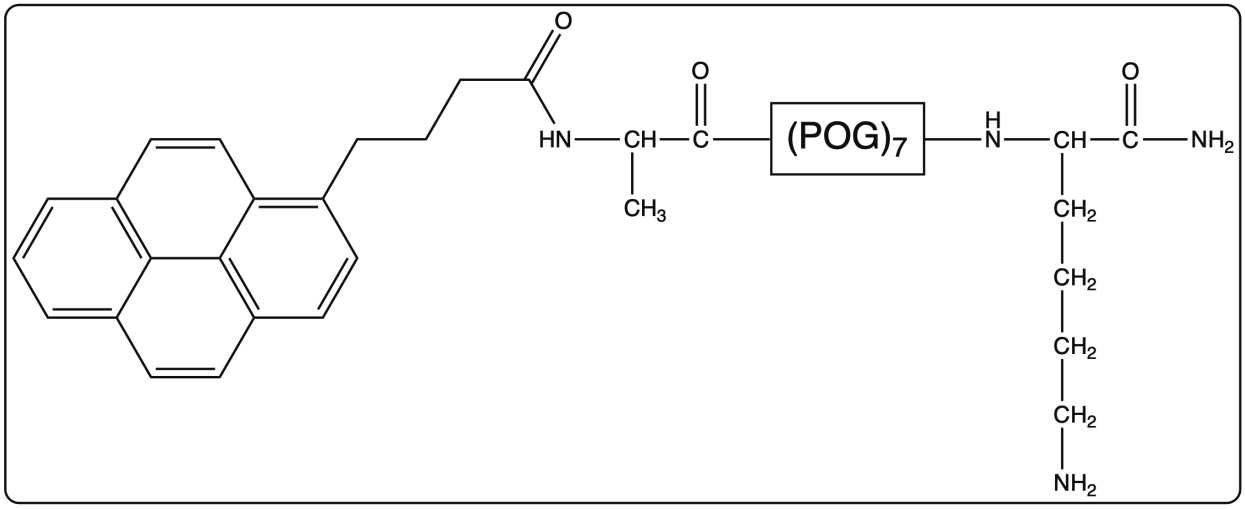

Figure 2.6: The **Pyr-(2)-AcAK** system

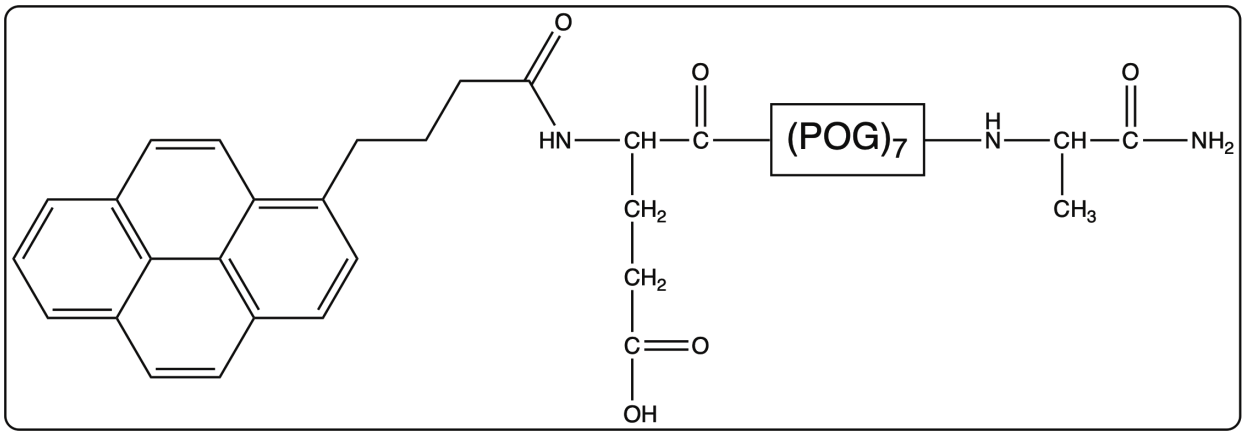

Figure 2.7: The **Pyr-(2)-AcEA** system

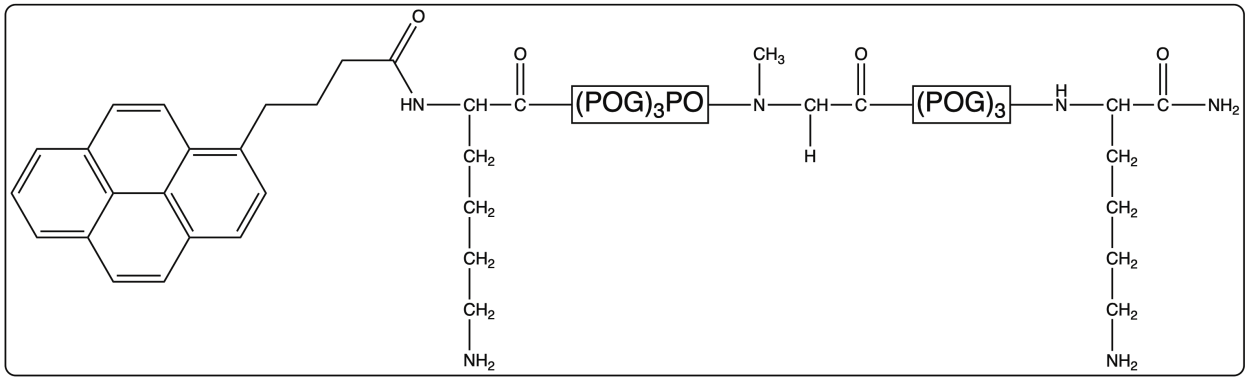

Figure 2.8: The **Pyr-(2)-AcKSarK** system

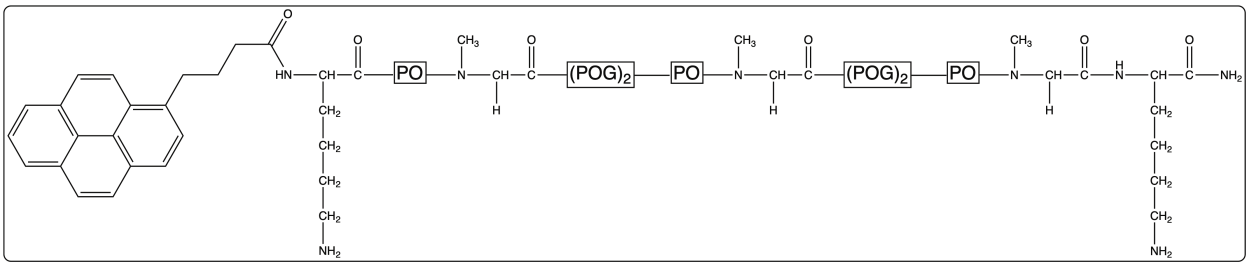

Figure 2.9: The **Pyr-(2)-AcKSar$_3$K** system

2.3 Thermodynamic Stabilities of Triple Helices as Measured by CD

CD spectroscopy was used to probe the triple helicity of each system. Firstly, all spectra measurements of each system at were prepared at the relatively high concentration of 200 μM, to ensure the presence of triple helices, though low enough to emit HT values at 224 nm under 400 V. This was qualitatively inferred through ellipticity maxima at 224 nm and minima at 200 nm, values similar to other CRP studies.[19,24,28] Predictably, deviations of this ellipticity coincided with modifications of the CRPs, such as Lys substitutions at conserved Gly registers or the flanking of the hydrophobic pyrene.

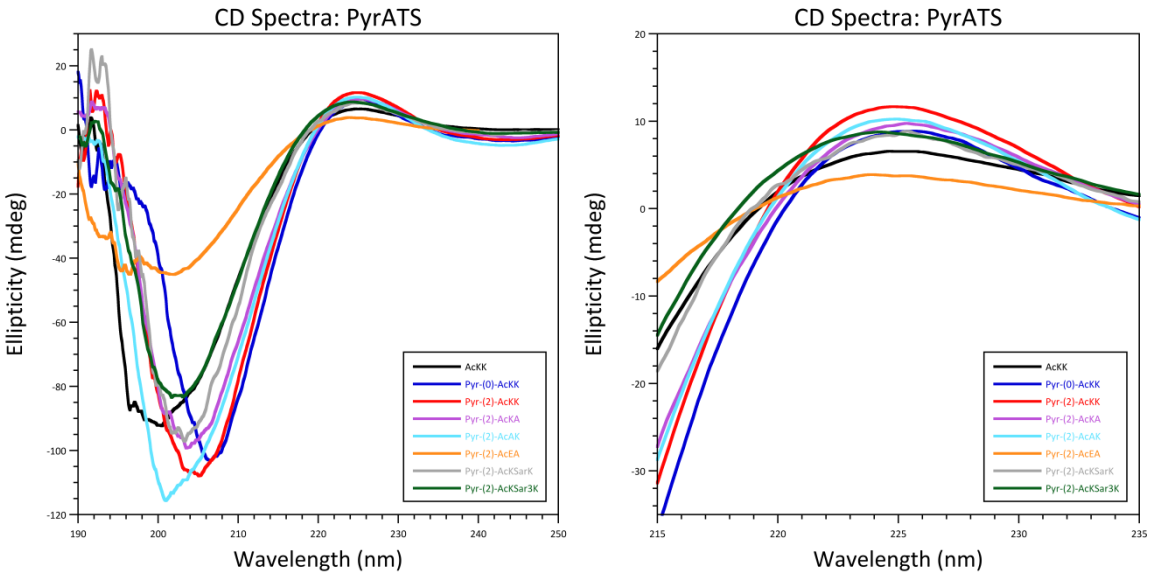

Figure 2.10: CD Spectra of various PyrATS at 0.20 mM in PBS. Full spectra show perturbations of PPII structure with additional modifications to the **AcKK** system (Left). Specific view at 224 nm highlights blue shifts in the **Pyr-(2)-AcKSar$_3$K** and the **Pyr-(2)-AcEA** systems (Right).

Secondly, temperature dependent measurements were conducted to measure the triple-helical stability of the CRPs. Each system was heated from 10 ˚C at a 0.2 ˚C/min rate measuring the change in ellipticity at 224 nm. Measurements were then fitted to a two-state model as described by previous studies[37], in which three monomeric chains coil into a helix. This model also offers an estimated transition enthalpy, ΔH°. It is important to note that the previous interpretation of the melting studies used a different two-state model. It was determined that this newly proposed model was more appropriate due to its proposed mechanism and its likeliness to occur in these systems.

Since these systems derive from Ac(POG)$_7$NH$_2$ assemblies, it is plausible to treat the transition into a triple helix as:

$$3C \rightleftharpoons H$$

Equation 2.1: Equilibrium between unfolded (monomer) and folded CRPs (trimers)

From previous studies[38], it was shown that the equilibrium constant is:

$$K = \frac{h}{c^3} = \frac{F}{3{c_o}^2(1-F)^3}$$

Equation 2.2: Equilibrium constant with respect to monomer and trimer

Variables *h* and *c* represent the concentrations of the triple-helix and monomeric coils, respectively. Variable c_o is the overall concentration of the CRP chains. Variable *F* is the fraction folded between the unfolded and folded state of CRPs.

The equilibrium constant can be related with the standard Gibbs free energy, Δ*G*°, the standard enthalpy, Δ*H*° and the standard entropy, Δ*S*°, by:

$$\Delta G° = \Delta H° - T\Delta S° = -RTlnK$$

Equation 2.3: Standard Gibbs free energy relationship to equilibrium constant

Specifically, the midpoint of the transition, where (*F* = 0.5) and the temperature, *T*, is represented as T_m, the equation is adjusted to:

$$T_m = \frac{\Delta H°}{\Delta S° + Rln(0.75{c_o}^2)}$$

Equation 2.4: Melting temperature determination

The temperature dependence of the equilibrium constant is given by the Van't Hoff equation:

$$\frac{dlnK}{dT} = \frac{\Delta H°}{RT^2}$$

Equation 2.5: Van't Hoff equation

This assumes a temperature independent transition enthalpy represented by Δ*H*°.

$$K = e^{\frac{\Delta H°}{RT}\times\left(\frac{T}{T_m}-1\right)-ln(0.75{c_o}^2)}$$

Equation 2.6: Final derived relation of thermal transition

This can then be expressed as a cubic equation and plotted on raw CD data to determine *ΔH*˚.

Table 2.2: Thermal denaturation data for PyrATS by CD

Abbreviation	T_m (CD) ˚C	*Δ*H˚ (CD) kcal/mol
AcKK	31.6 (± 1.2)	-63.1 (± 1.1)
Pyr-(0)-AcKK	33.7 (± 0.3)	-64.6 (± 1.0)
Pyr-(2)-AcKK	39.6 (± 0.3)	-60.3 (± 2.4)
Pyr-(2)-AcKA	39.5 (± 0.6)	-50.6 (± 2.0)
Pyr-(2)-AcAK	65.2 (± 0.2)	-88.8 (± 4.4)

Pyr-(2)-AcEA	31.5 (± 0.3)	-61.4 (± 1.7)
Pyr-(2)-AcKSarK	26.4 (± 0.3)	-81.3 (± 0.8)
Pyr-(2)-AcKSar$_3$K	-	-

Upon further inspection, the two flanking Lys imparted instability of the resultant 23-mer, **AcKK**, by approximately 8 ˚C when compared to a common control CRP, Ac(POG)$_7$NH$_2$ under identical studies.[38] In part, this is due to destabilizing factors such as like-charge repulsion[39] and a substitution for Gly at a location in the primary sequence that is strictly reserved for it.[40] When observing the **Pyr-(0)-AcKK** system, the increase in T_m was a modest 2 ˚C and the difference in ΔH° was within the margin of error. This suggests that the formal substitution of pyrene provided very little distortion with respect to helical folding. Perhaps, the absence of spacer atoms prevents the fluorophores from taking advantage of hydrophobic overlap, akin to studies where CRPs were tagged with a large carboxyfluorescein dyes (molar mass = 376) in a similar manner.[36a]

To contrast, observations into the CRPs with carbon spacers suggested a greater influence into folding stability (Figure 2.11). With the two-carbon spacer, **Pyr-(2)-AcKK** provided an additional 6 ˚C increase when compared to **Pyr-(0)-AcKK**. This may be due to the added conformational flexibility for the fluorophores to associate. What seems to drive this overlap is largely entropic, since ΔH° values for **Pyr-(0)-AcKK** and **Pyr-(2)-AcKK** are largely the same.

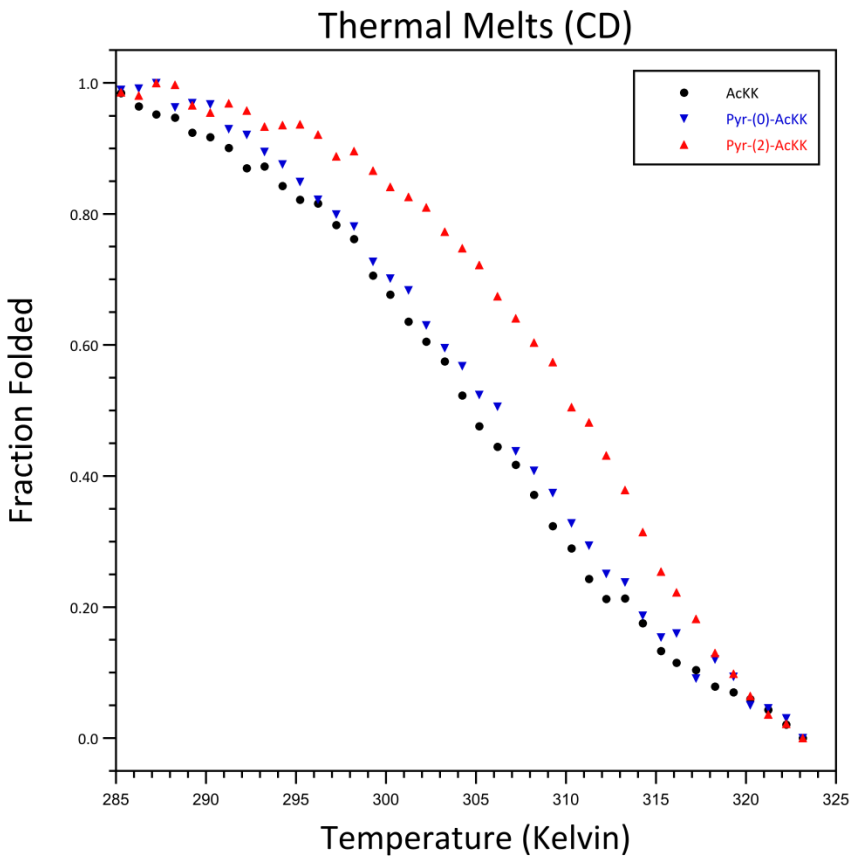

Figure 2.11: Comparisons among the thermal melts of **AcKK, Pyr-(0)-AcKK** and **Pyr-(2)-AcKK**

Switching one flanking amino acid from Lys → Ala caused noticeable changes in T_m and ΔH°. When comparing **Pyr-(2)-AcKA** system to **Pyr-(2)-AcKK** (Figure 2.12), a steep 10 kcal/mol decrease in ΔH° occurred while the overall stability remained largely unchanged. To contrast, **Pyr-(2)-AcAK** differs in T_m by 25 ˚C with its isomeric counterpart and essentially doubles its thermal stability when compared to **AcKK** (Figure 2.12). Unlike the other PyrATS, **Pyr-(2)-AcAK** contains an electrically neutral N-terminus, eliminating the large competitive force for pyrene overlap. This design may allow nucleation to form at the N-termini, a crucial initiation step for the triple helix to form. In native collagen, folding propagates from C-termini to N-termini[40a], but factors as simple as high concentrations[40b] or the presence of oligomerization domains[41]

have linked collagen strands starting from the N-termini. Flanking amino acids located on the N-termini sit on positions where, naturally, Gly would be. The higher stability of **Pyr-(2)-AcAK** may be due to a less disruptive Ala substitution at this position, when compared to Lys.[16]

The system **Pyr-(2)-AcEA** contains a negatively charged N-terminus. The presence of glutamic acid seems to be more detrimental to overall thermal stability than systems with positively charged N-termini, such as **Pyr-(2)-AcKA** (Figure 2.12). A possible explanation may be due again to a highly disruptive substitution on a position where glycine would normally be. A similar study using osteogenesis imperfecta CRPs revealed that negatively charged amino acid substitutions for Gly have been the most damaging for thermal stability.[16]

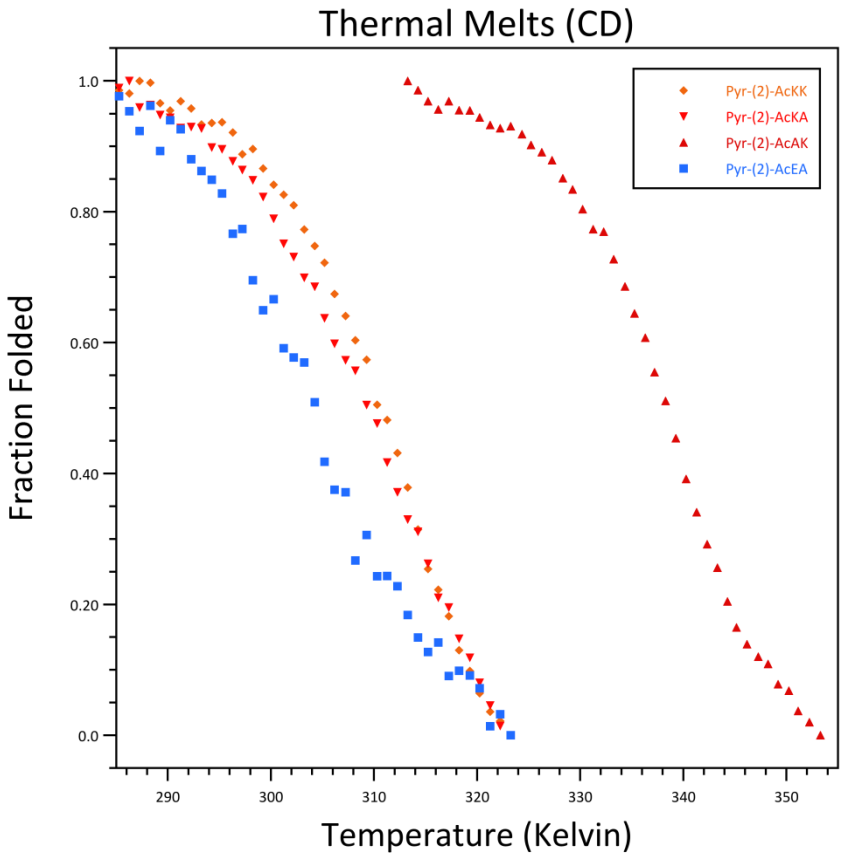

Figure 2.12: Comparing the influence of flanking amino acids on PyrATS

When implementing a Gly → Sar substitution at the central triplet, **Pyr-(2)-AcKSarK** provided a T_m of about 26 ˚C while cooperatively unfolding (Figure 2.13). It is interesting to note that even the removal of the central triplet N-H hydrogen bond donor, ΔH° increases. This trend was also present when a comparable modification was performed on Ac-$(POG)_8$-NH_2 CRPs.[40a] For **Pyr-(2)-AcKSar$_3$K**, a system that features three Gly → Sar substitutions, T_m and ΔH° were not determined due to its lack of sigmoidal behavior during unfolding.

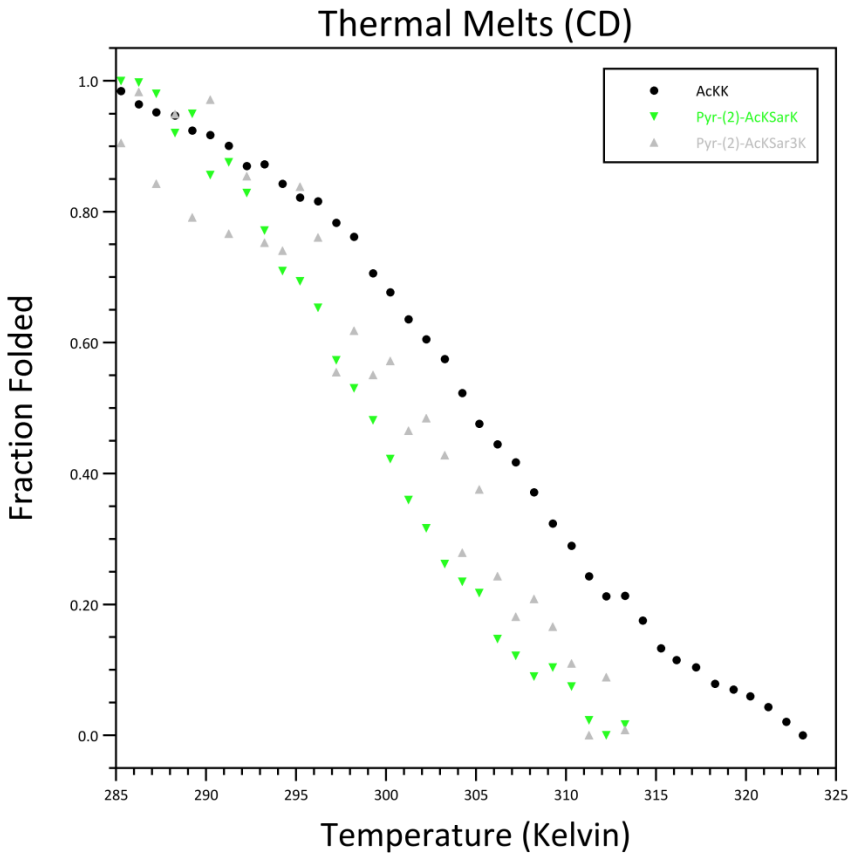

Figure 2.13: Sarcosine offers thermal instability in PyrATS.

2.4 Thermodynamic Stabilities of Triple Helices as Measured by FL

Concentrations routinely used for CD measurements are somewhat limiting when conducting more sensitive FL studies. Practical concerns such as the inner filter effect make it

difficult to acquire reliable emission data. Contrariwise, the sensitivity of fluorescence spectroscopy allows for investigations below concentrations where trimer formation would predominate. Thus, a selection of peptide samples was prepared to determine appropriate conditions for fluorescence sensitive melting studies.

To prepare for the auxiliary set of experiments, each sample was prepared in PBS stock solutions to cover concentration ranges from 5 to 100 µM. Each system was then stored at 4 ˚C overnight to maximize full equilibration at these low concentrations. Emission data were then acquired for each stock at room temperature. For **Pyr-(2)-AcKA**, the ratio of integrated intensities of excimer to monomer (I_E / I_M) was constant at 0.15 for the samples at 7 and 14 µM, then increased thereafter to a maximum value of 1.3 at ~50 µM (Figure 2.14, left image). This concentration was used for subsequent determination of T_m values. Little evidence of interaction between pyrene units was observed for sarcosine-containing **Pyr-(2)-AcKSarK** as I_E / I_M values remained under 0.2 throughout the concentration range (Figure 2.14, right image).

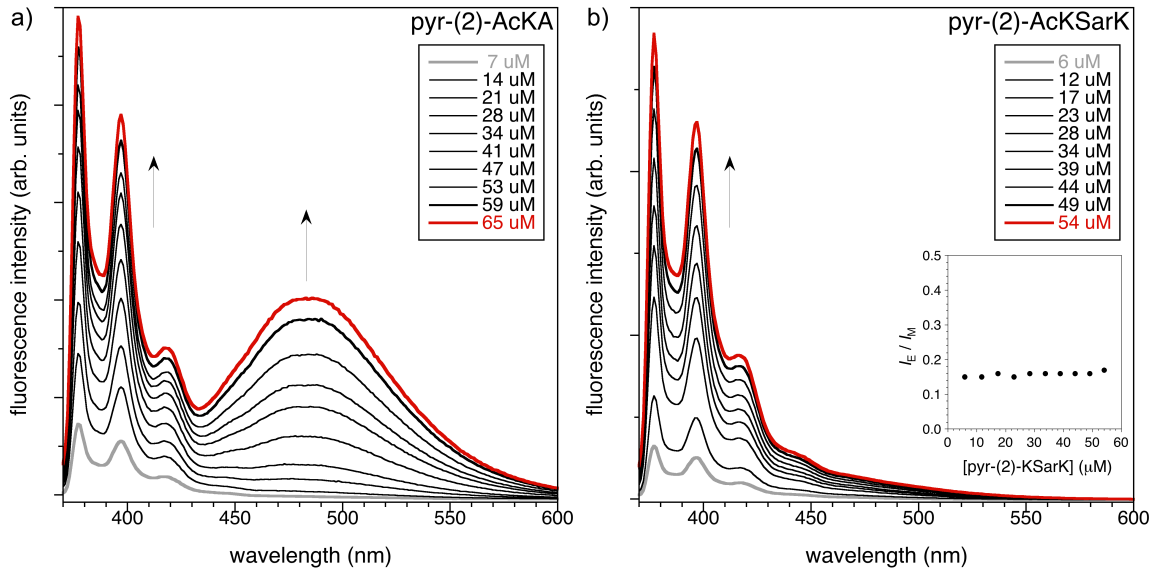

Figure 2.14: Concentration dependence of excimer formation. (a) Emission spectra of **Pyr-(2)-AcKA**. (b) Emission spectra of **Pyr-(2)-AcKSarK**. Inset: Ratio of integrated intensities (I_E / I_M) for **Pyr-(2)-AcKSarK**. To maintain consistency among peptides, the wavelength cutoff between monomer and excimer peaks were set at 435 nm.

Melting experiments were carried out by heating solutions of PyrATS from 10 ˚C to 60 ˚C. Emission signals from 370 to 435 nm, corresponding to monomeric pyrene moieties, grew in intensity in a sigmoidal manner as sample temperatures increased. These data were then converted to represent the fraction folded and plotted alongside the CD curves for comparison. To summarize, fitting to the same model gave T_m values within 10% of those derived from CD.

Table 2.3: Thermal denaturation data for PyrATS by CD & FL

Abbreviation	T_m (CD) ˚C	T_m (FL) ˚C
AcKK	31.6 (± 1.2)	-
Pyr-(0)-AcKK	33.7 (± 0.3)	33
Pyr-(2)-AcKK	39.6 (± 0.3)	36
Pyr-(2)-AcKA	39.5 (± 0.6)	38
Pyr-(2)-AcAK	65.2 (± 0.2)	41
Pyr-(2)-AcEA	31.5 (± 0.3)	32
Pyr-(2)-AcKSarK	26.4 (± 0.3)	-
Pyr-(2)-AcKSar$_3$K	-	-

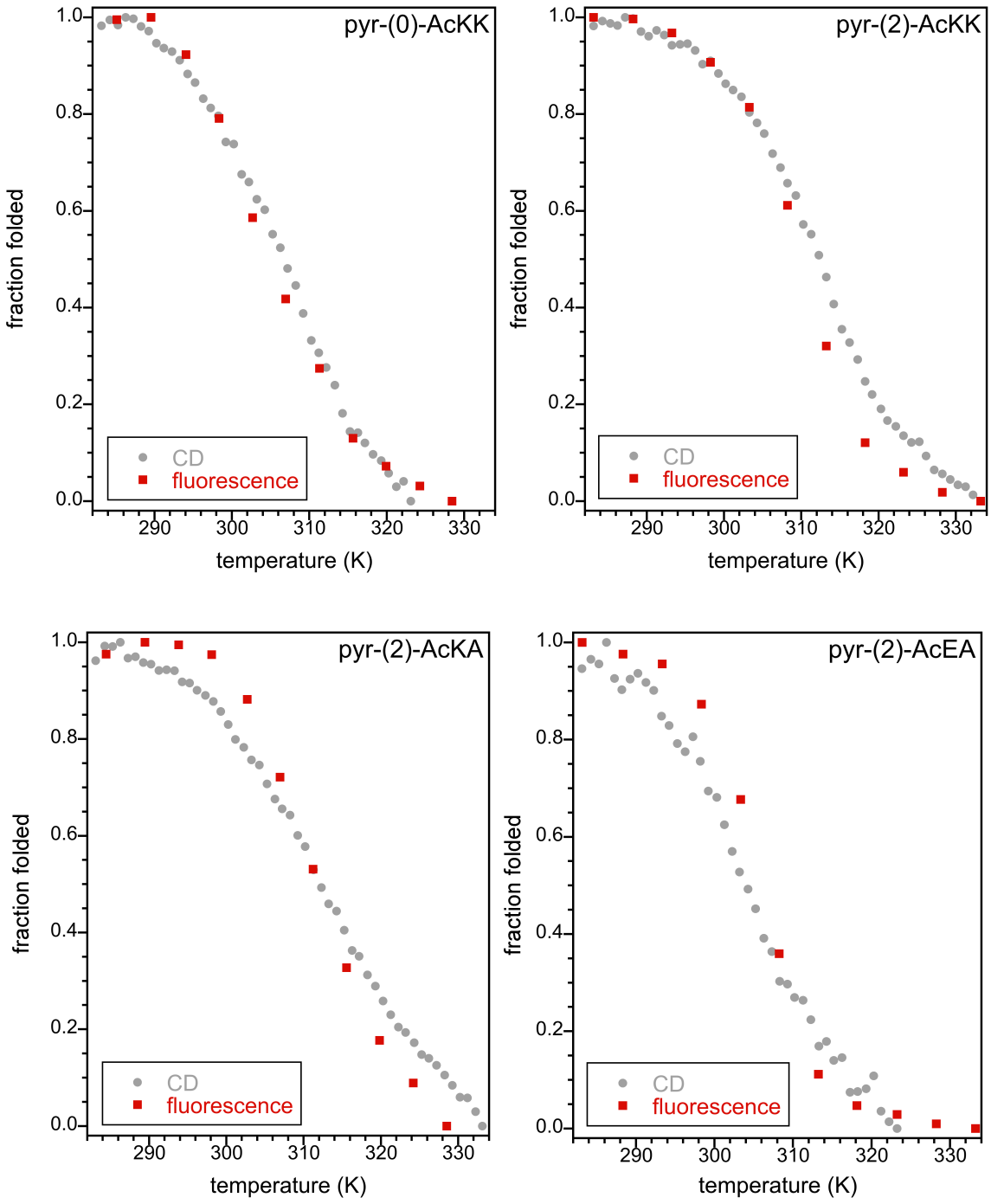

Figure 2.15: Comparisons of thermal studies offer T_m values with little deviation (10 %).

While the systems **Pyr-(0)-AcKK**, **Pyr-(2)-AcKK**, **Pyr-(2)-AcKA**, and **Pyr-(2)-AcEA** display similar T_m values, it is important to notice that the shape of the sigmoidal behavior can be visually different (Figure 2.15). A sharper turn demonstrates a two-state nature that is more

zero-sum. It can be interpreted that fitting the model described previously maybe unreliable in these studies.

The system **Pyr-(2)-AcAK** reveals two-state behavior, but with a significantly different T_m measured in CD (Figure 2.16). To reiterate, FL spectroscopy on PyrATS only measures changes directly influencing the fluorescent probe pyrene, while CD spectroscopy ultimately measures the chiral, stereoscopic nature of bulk peptides. It's obvious that this system has a different relationship between pyrene and the core residues.

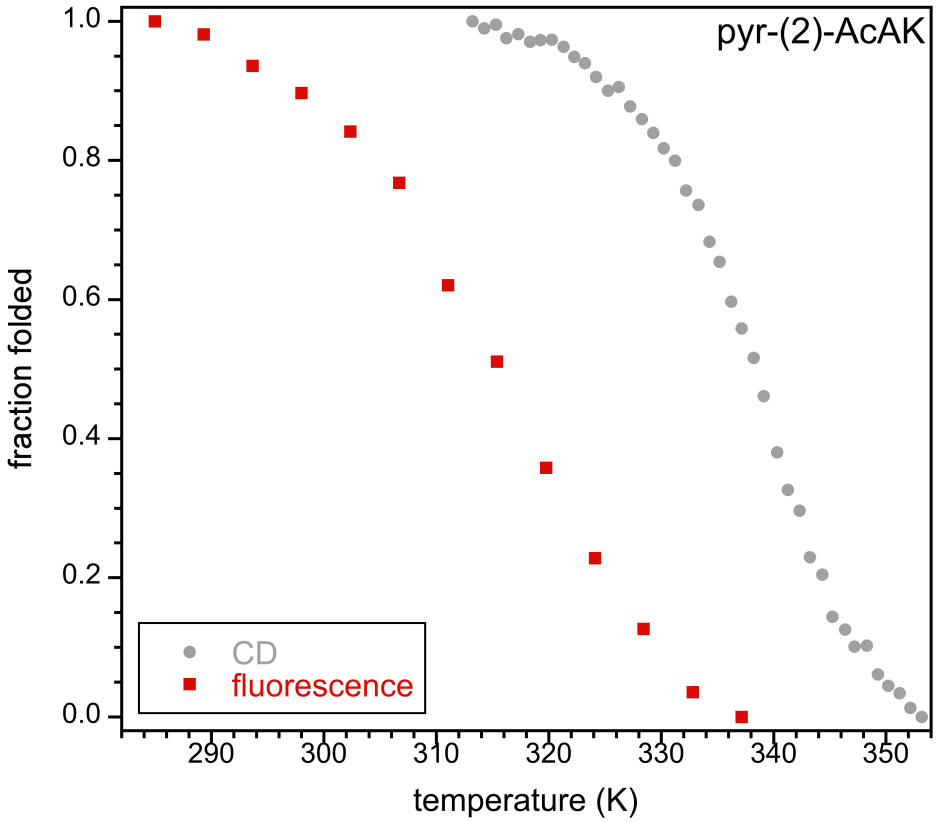

Figure 2.16: The melting studies from **Pyr-(2)-AcAK** reveal two dissimilar T_m values.

Thermodynamic studies into PyrATS have shown a glimpse on how the hydrophobic probe affects the free energy changes during unwinding. The addition of these probes has seemed to have little effect on interstrand bonding. This can be visualized through the lack of

major changes of ΔH° throughout each system when compared to the control **AcKK**. One major exception is the system **Pyr-(2)-AcAK**. With its drastically higher T_m determined by CD, the probe clearly has a large influence on its stability. Perhaps, the combination of the carbon spacer and the neutral flanking amino acid located on the N-terminus allows for maximal overlap on the fluorophores. This should be interpreted as hydrophobically driven as opposed to enthalpically, as the ΔH° for **Pyr-(2)-AcAK** did not experience a similar increase as the T_m. This can be further supported with the small enthalpic role π - π stacking has with similar systems discussed in previous studies.[34] The discrepancy in results between this system and the others make the relationship between the pyrene probes and collagen folding, unclear.

2.5: Kinetic Behavior of Triple Helices as Measured by FL

Investigating the kinetic behavior of PyrATS allows for a clearer interpretation of the effects of the probes on helical folding. A select number of systems were prepared in PBS solutions to concentrations of 25 μM and 50 μM. Each system was heated to 80 ˚C for 20 minutes to ensure full trimer dissociation. Subsequently, the system was then flash cooled down to room temperature and immediately transferred to a quartz cuvette for excitation. The cooling period (dead time) for each system from full dissociation to excitation was 90 seconds. Each scan measured the growth of excimer emission from 335 nm to 600 nm with respect to time. The number of scans were determined by visually inspecting of excimer intensity for indicators of plateauing out or lack of change in emission, which resulted to about 20 to 25 scans. Experiments lasted for at least 20 minutes, which are similar to the refolding times of similar systems in a different study.[28] The intensity within the emission range was integrated

and plotted to determine the order of kinetics. A rate constant was determined by fitting a linear curve upon the logarithmic conversion of the integrated intensity.

Table 2.4: Kinetic data for selected PyrATS by FL

Abbreviation	Concentration	Order	k_1	$t_{½}$
Pyr-(2)-AcAK	50 µM	1st	0.0047 s^{-1}	7 min
Pyr-(2)-AcKK	50 µM	>2nd	n.d.	17 min (est.)
Pyr-(2)-AcKA	50 µM	>2nd	n.d.	10 min (est.)
Pyr-(2)-AcAK	25 µM	>2nd	n.d.	10 min (est.)

For the systems, **Pyr-(2)-AcKK** and **Pyr-(2)-AcKA,** the rise in excimer emission displayed non-first order kinetics (Figure 2.17, left image). Interestingly, the change in excimer emission **Pyr-(2)-AcKK** system levels off slower than in **Pyr-(2)-AcKA.** A possible explanation with this discrepancy may be linked to the choice of flanking amino acids. **Pyr-(2)-AcKK** contains charged amino acids on both termini as opposed to one. This may shift the location on where each monomer associates.

For the **Pyr-(2)-AcAK** system, the rise in excimer emission displayed first order kinetics with a rate constant of 0.0047 s^{-1} (Figure 2.17, right image). This value is similar to a propagation rate constant of 0.002 s^{-1}, as determined in a previous study with a $(POG)_{10}$ CRP.[41] This behavior is likely contributed by the pyrene probe and the neutral N-termini; specifically, "chain-finding" steps must be fast for this system, such that the fluorescence study captures only helix propagation. This influence is concentration limited though, as **Pyr-(2)-AcAK** at 25 µM displays higher order kinetics.

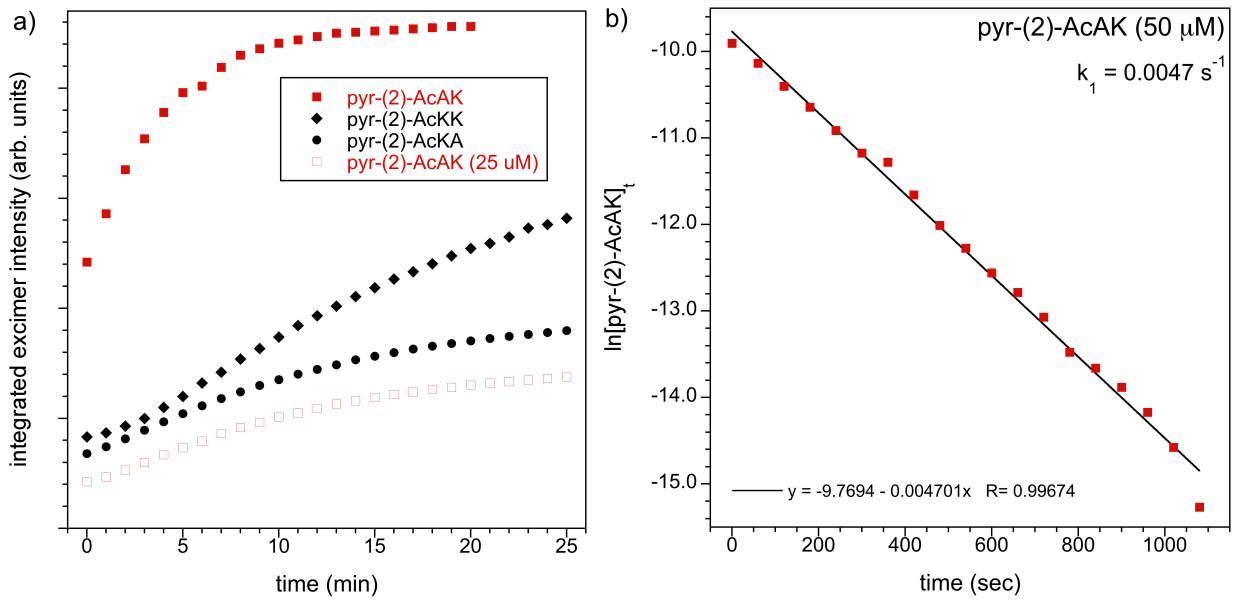

Figure 2.17: (a) Time course of excimer growth for selected peptides. Samples were heated to 80 ˚C for 20 minutes then rapidly cooled to room temperature for analysis. Dead time was 90 sec. (b) Fitting of data for **Pyr-(2)-AcAK** (at 50 μM) to first order kinetics.

To correlate this kinetic behavior with non-pyrene appended systems, a trusted, proposed mechanism of the rate of collagen folding[42] is presented as:

$$2C \leftrightharpoons C + D^* \longrightarrow H^* \longrightarrow H$$

Equation 2.7: Folding mechanism of collagen trimers

C represents the single monomeric chain, D^* represents the dimer state formed between two chains, *C*, H^* represents the nucleation of the triple helix, and *H* represents the full formation of the helix.

For all systems that displayed non-first order kinetics, it was presumed that the nucleation of the full helix ($C + D^* \rightarrow H^*$) was the rate-limiting step of the mechanism. This is supported by the interpretation of Engel et al.[41], as it was established that similarly low concentrations of analogous CRPs caused the folding behavior of the helix to be third order. To

contrast, **Pyr-(2)-AcAK** had a clear kinetic advantage in emission return. Even at low concentrations, this system was able to have a relatively quick folding half-life ($t_{½}$) of 7 min.

This incongruity can be explained through an entropic drive to associate the chains from the N-termini. In previous studies, certain foldons attached to the N-termini of various CRPs have caused nucleation to occur on the reverse sides of these collagen chains.[42] The association of pyrene molecules would be entropically favorable, especially with neutral N-termini, and in turn may preorganize the chains to associate in an untraditional manner.

2.6 Conclusions on PyrATS

All PyrATS, with the exception of **Pyr-(2)-AcKSar$_3$K**, display cooperative unfolding under thermal stress. The determined T_m values were very similar between CD and FL studies, with deviations well below 10%. Though the shape of the sigmoidal behavior were not congruent in some cases. The noticeable exception was **Pyr-(2)-AcAK**, as it formed helices easily discerned as hyperstable by CD but not by FL. In the kinetic studies, selected systems of PyrATS display folding behavior similar to non-pyrene appended systems at these concentrations. Though it is apparent that the same noticeable exception, **Pyr-(2)-AcAK**, had seemed to have folded in a much quicker fashion.

In conclusion, pyrene probes were annealed to CRPs to determine its viability as a tool for monitoring the thermodynamic and kinetic behaviors through fluorescence. While the probes provided little distortion in the strength of the intermolecular forces that helically bind the collagen chains, the FL determined I_E / I_M exhibited an unfolding behavior unparalleled to the measurements done by CD. The kinetic studies continue to show discrepancy between the two spectroscopic studies. For example, the t½ values for excimer emission were far quicker

than typical $t_{½}$ values in determining collagen refolding by CD, in this case performed by Chenoweth et al.[28] It is important to note that the kinetic studies were performed with relatively low concentrations. Refolding kinetics are usually concentration dependent, but low concentrations alone does not explain the discrepancy in this case. If non-pyrene appended CRPs were subjected in the same test conditions, the $t_{½}$ would presumedly be longer than what has been reported, rather than quicker.

Altogether, the pyrene probe seems to unreliably monitor the folding and unfolding behavior of certain charged collagenic molecules. While it lacks the specificity to quantitatively determine thermodynamic and kinetic principles of collagen unwinding, it has allowed for a deeper understanding on the competitive forces that act on collagen helices and the interdependent role between the probe and the peptides. The following chapters will attempt to address the role of charged residues in these discrepancies.

CHAPTER 3: Pyrene-Bearing Amino Acid Derivatives

3.1 Introduction into Synthetic Design of Amino Acid Derivatives

When our group first began work on pyrene-bearing CRPs, the goal was to develop a synthetic design that allows fluorophores to be attached at various locations. With PyrATS, a major limitation was their constraining the location of the probe to the N-termini, often offering information biased to one side of the peptide sequence. With a different, more flexible approach, a deeper understanding of collagen thermodynamics and kinetics may be revealed.

Substituting an aromatic molecule such as pyrene into an amino acid requires a high degree of selectivity. Amino acids contain multiple reactivity sites, therefore protecting groups are needed. In addition, these protecting groups may impair the reaction conditions that could favor the desired outcome. The basis of this new design is the use of amino acid derivatives that contain labile leaving groups. In turn, pyrene can then be covalently added as a compound that contains both the components of an amino acid and the fluorophore. The present chapter discusses the many synthetic directions taken and highlights the major roadblock encountered in this work, solubility.

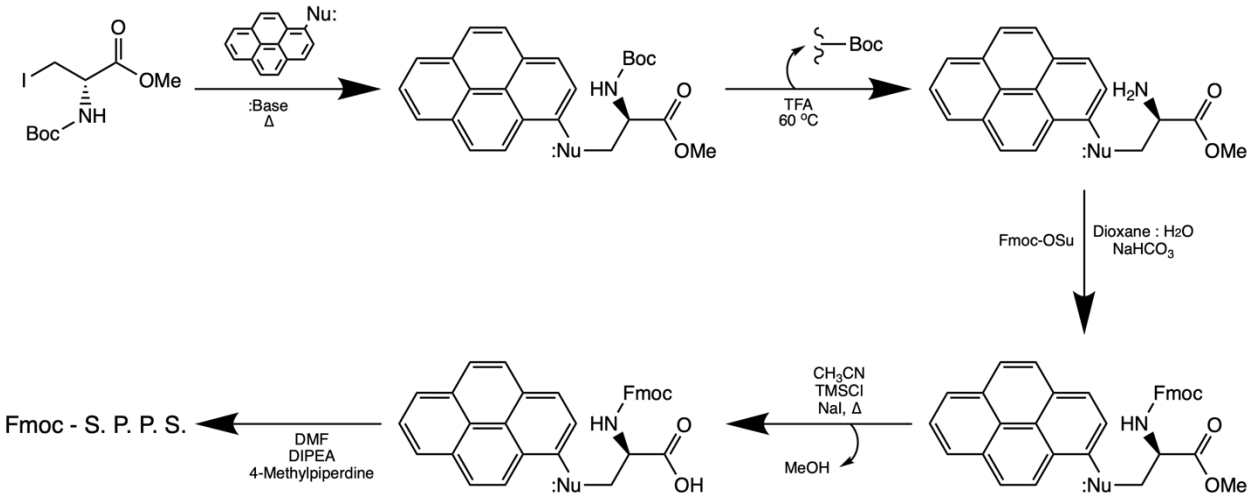

Figure 3.1: Synthetic strategy for pyrene-derived amino acids

3.2 Attempted Syntheses of Pyrene-Containing Amino Acids

Synthesis of Boc-Ala(Pyr)-OMe

The first attempt was based on Suzuki cross-coupling reactions.[43] $Pd_2(DBA)_3$ (0.150 mmol) and 4,5-bis(diphenylphosphino)-9,9-dimethylxanthene ("Xantphos", 0.300 mmol) were dissolved in THF (15 mL). Shortly thereafter, *N*-boc-3-iodo-L-alanine methyl ester (3.0 mmol) and pyrene-1-boronic acid (3.1 mmol) were added. Na_2CO_3 (1.0 M, 15 mL) was added as the activating base. The mixture was deoxygenated (5 min) with a gentle stream of argon, then heated under reflux for its proposed Suzuki reaction (~65 °C, 12 hours).[43] After cooling to room temperature, the reaction was partitioned between brine and CH_2Cl_2 in a separatory funnel and purified through flash chromatography (5:1, toluene : CH_2Cl_2).

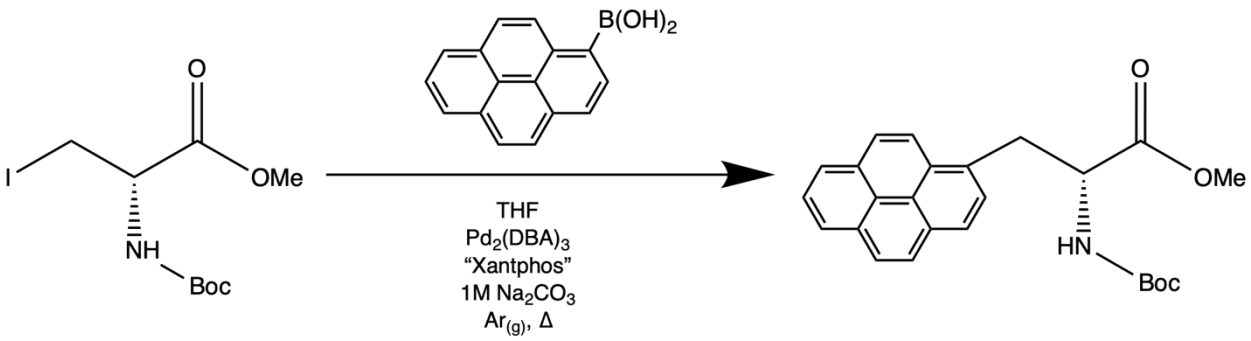

Figure 3.2: Suzuki cross-coupling synthetic scheme

Purification by flash chromatography proved to be difficult as products of the reaction had low solubility with common solvents of various polarities (hexane to methanol). Recovery of major product (R_f: 0.45) produced a low yield of 35%. ^{1}H NMR in $CDCl_3$ suggested that a beta-hydride elimination had occurred, as alkene peaks appeared and the aromatic peaks did not.

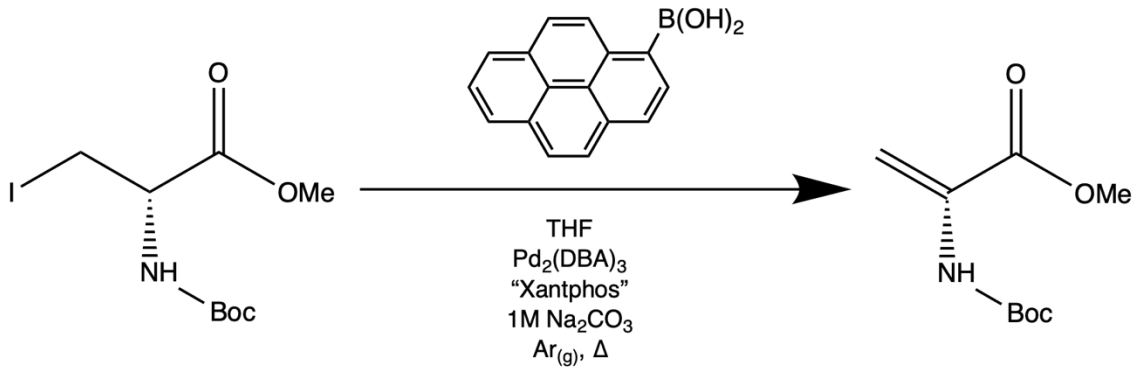

Figure 3.3: Synthetic conditions that promote beta-hydride elimination over cross-coupling.

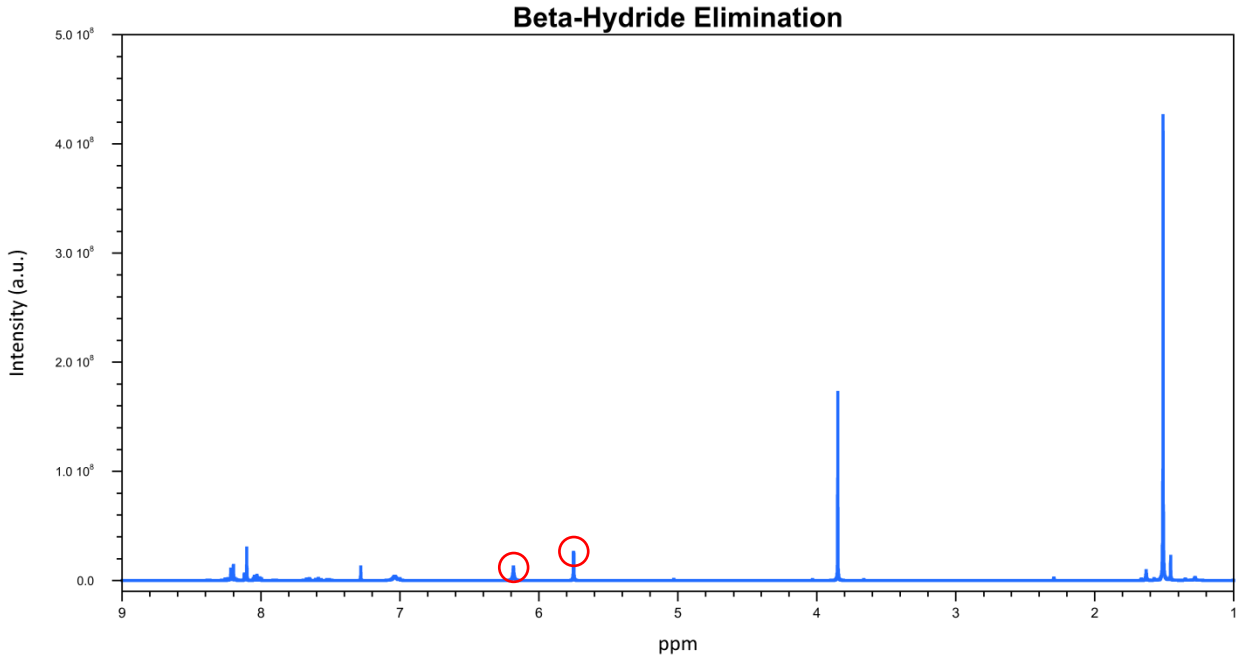

Figure 3.4: ^{1}H NMR ($CDCl_3$) of the elimination product. Alkene peaks appear at 5.78 and 6.21 ppm (2H, dd). Aromatic peaks at 7.00 to 8.50 ppm fail to integrate for 9H.

Synthesis of Boc-Ser(MePyr)-OMe

A different synthetic scheme based on S_N2 chemistry was evaluated next. *N*-boc-3-iodo-L-alanine methyl ester (2.00 mmol) and pyrene-1-carboxylic acid (2.00 mmol) were dissolved in THF (50 mL). Upon stirring, DIPEA (2.00 mol, 0.349 mL) was added as an activating base. The mixture was deoxygenated (5 min) with a gentle stream of argon, then heated to reflux to allow S_n2 reaction (~65 °C, 12 hours). After cooling to room temperature, the reaction contents were partitioned between brine and CH_2CL_2 in a separatory funnel and purified by flash chromatography (5:1, toluene : CH_2Cl_2). NMR analysis of the major product (R_f: 0.60) did not show evidence of pyrene incorporation, as all fractions investigated lacked aromatic peaks (6.5 to 8.0 ppm) and alpha-hydrogen peaks (~4.0 ppm) simultaneously. Additional attempts using more forcing conditions were also unsuccessful, presumably due to poor nucleophilicity of the aromatic carboxylate.

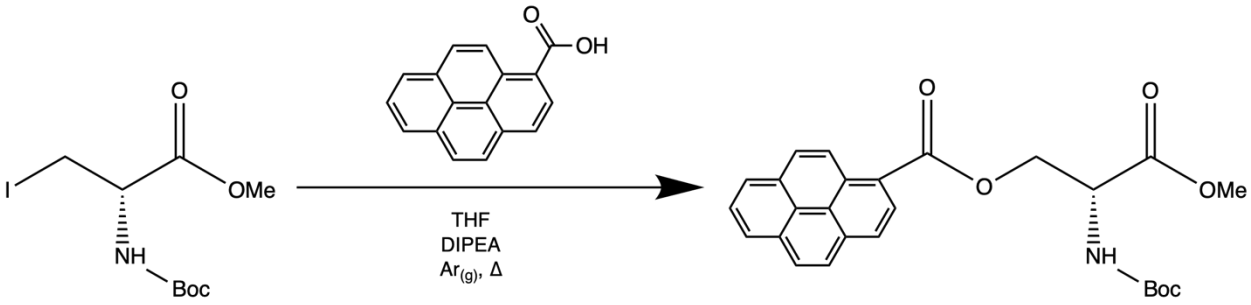

Figure 3.5: Attempted S_n2 approach to pyrene-substituted amino acid using DIPEA

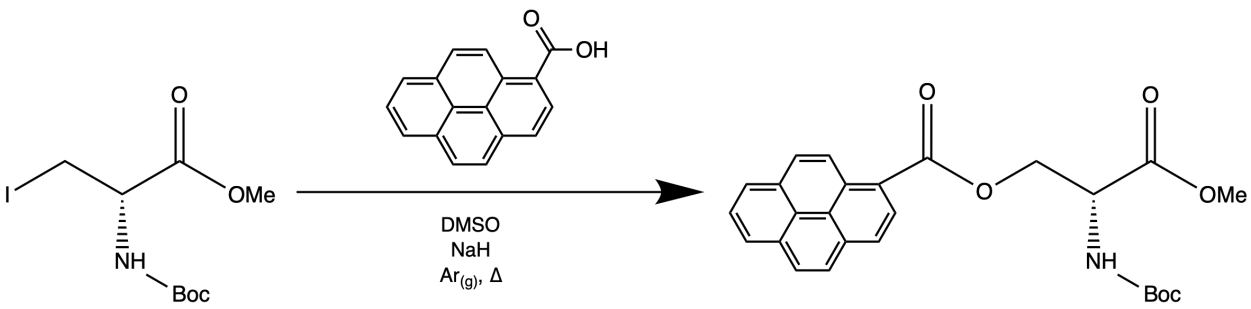

Figure 3.6: Attempted S_n2 approach to pyrene-substituted amino acid using NaH

Synthesis of Boc-Dap(MePyr)-OMe

A pyrene reagent featuring a better nucleophile was used. *N*-boc-3-iodo-L-alanine methyl ester (2.0 mmol) and pyrene-1-methylamine (2.0 mmol) were dissolved in THF (50 mL). DIPEA (5.0 mmol, 0.873 mL) was added as the activating base. The mixture was deoxygenated (5 min) with a gentle stream of argon, then heated to allow S_n2 reaction (~45 °C, 12 hours). After cooling to room temperature, the reaction mixture was partitioned between brine and CH_2Cl_2 in separatory funnel and purified by flash chromatography (10:1, EtOAc : CH_2Cl_2). This reaction produced the desired product in ~10 % yield, presumably due to low solubility of the perceived product. Conditions were altered to address the low yield and potentially avoid purification as a whole. Ultimately, these efforts were unsuccessful.

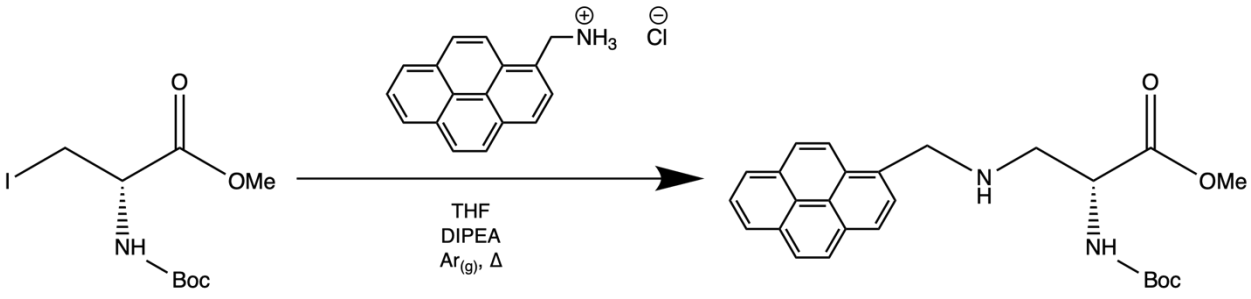

Figure 3.7: Attempted S_n2 approach to pyrene-substituted amino acid using a nucleophilic amine

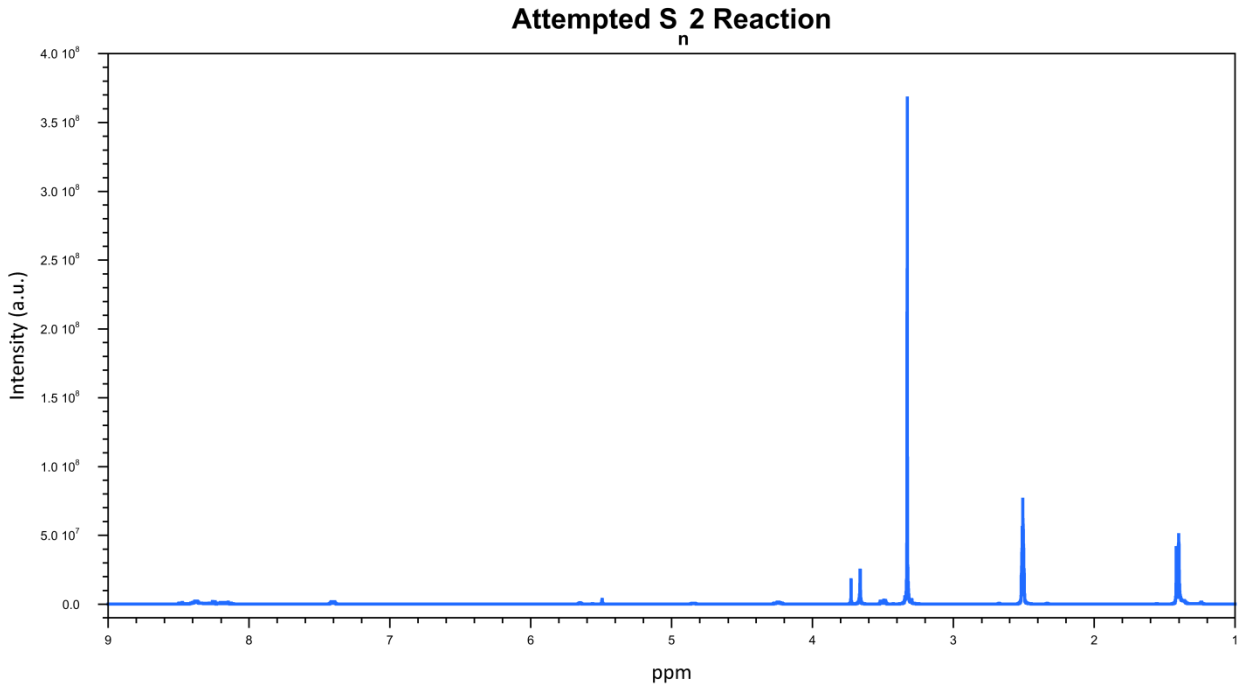

Figure 3.8: ^{1}H NMR (DMSO-d_6) of the attempted substitution product. There is no clear sign of substitution occurring, as aromatic peaks at 7.00 to 8.50 ppm fail to integrate for 9H.

Synthesis of Boc-Dap(MePyr)-OH

In response to the previously failed synthetic strategies, the sources of the nucleophile and the electrophile were switched. Boc-Dap-OH (2.0 mmol) was suspended in THF (20 mL). NaH (4.0 mmol) was then added as the activating base. Once activated, 1-bromomethylpyrene (2.0 mmol) was introduced. The mixture was deoxygenated (5 min) with a gentle stream of argon and stirred at room temperature (~22 °C, 12 hours).

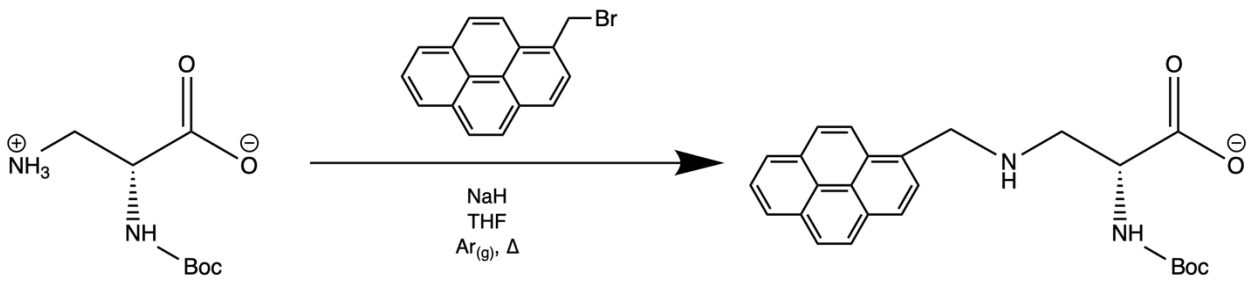

Figure 3.9: Attempted S_n2 approach to pyrene-substituted amino acid with Boc-Dap-OH

The reaction mixture was filtered by gravity and the solvent was removed under vacuum. Characterization of the crude product was completed by MS (ESI+ QToF) revealing a signal with *M/Z* of 419, among other peaks. The crude reaction was then subjected to flash chromatography, only to find issues in solubility with common purification solvents. Ultimately, the product could not be separated from the byproducts.

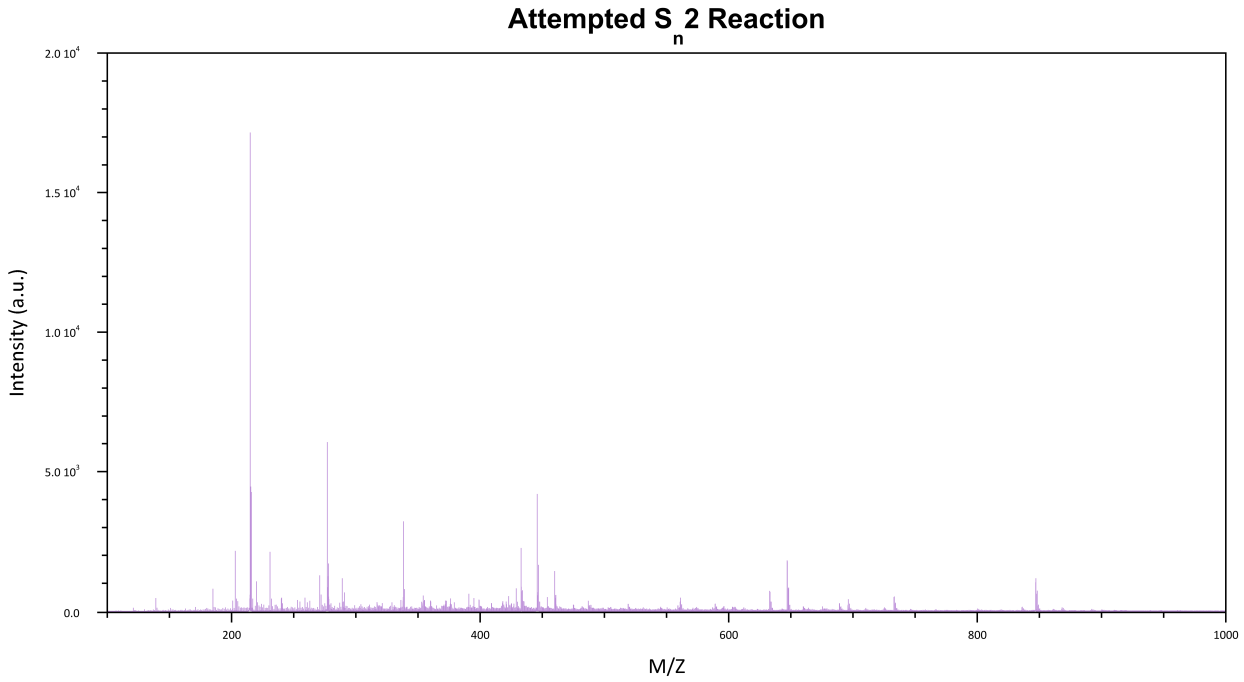

Figure 3.10: Mass spectrum (ESI+ QToF) of the best attempt of separating the product. A *M/Z* value of 419.34 (expected: 418.19) is observed for a minor peak.

3.3 Synthesis and Characterization of *N*-pyrenemethylglycine (Pmg)

The synthetic schemes discussed previously have reoccurring problems of insolubility. In the worst cases, the reaction produced a more favorable byproduct, such as the beta-hydride elimination. In other cases, the crude material was too insoluble to viably separate the products in bulk. This is most likely due to the coinciding presence of charged and hydrophobic ends.[44] Multiple alterations to each scheme attempted to generate a product of sufficient purity but

were unsuccessful. Fortunately, a novel, facile design described below produces pyrene-bearing glycines with high yields.

1-Pyrenemethylamine • HCl (4 mmol) was suspended in EtOH : H_2O (2:1, 45 mL). KOH (4 mmol) was added to the suspension and mixed until the amine was dissolved and activated. Sodium chloroacetate (4 mmol) was then introduced into the solution and heated to reflux (4 hours). After three hours, a white precipitate was observed. TLC analysis (EtOH : CH_3OH, 4:1) revealed a new spot (R_f: 0.05) with a faint blue appearance under a UV-lamp (365 nm). The immobility is most likely due to the product behaving in its zwitterionic form. Upon cooling the suspension to room temperature, the white precipitate was collected using a fine glass fritted funnel and was washed with EtOH and H_2O. Consequently, the product, dubbed *N*-pyrenemethylglycine (**Pmg**) was left to dry on weigh paper (86 %). ^{1}H NMR revealed a simultaneous presence of peaks in the aromatic region (6.5 – 8.0 ppm) and alpha-hydrogen range (~4.0 ppm). ^{13}C NMR accounted for all the projected carbons.

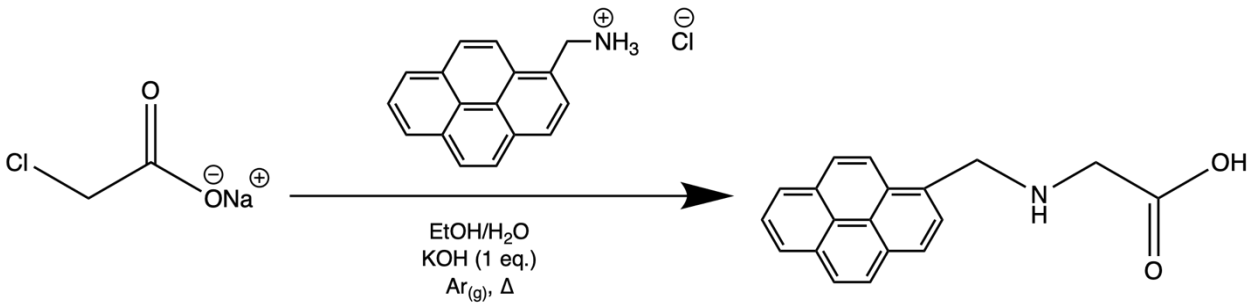

Figure 3.11: Synthetic scheme of *N*-pyrenemethylglycine

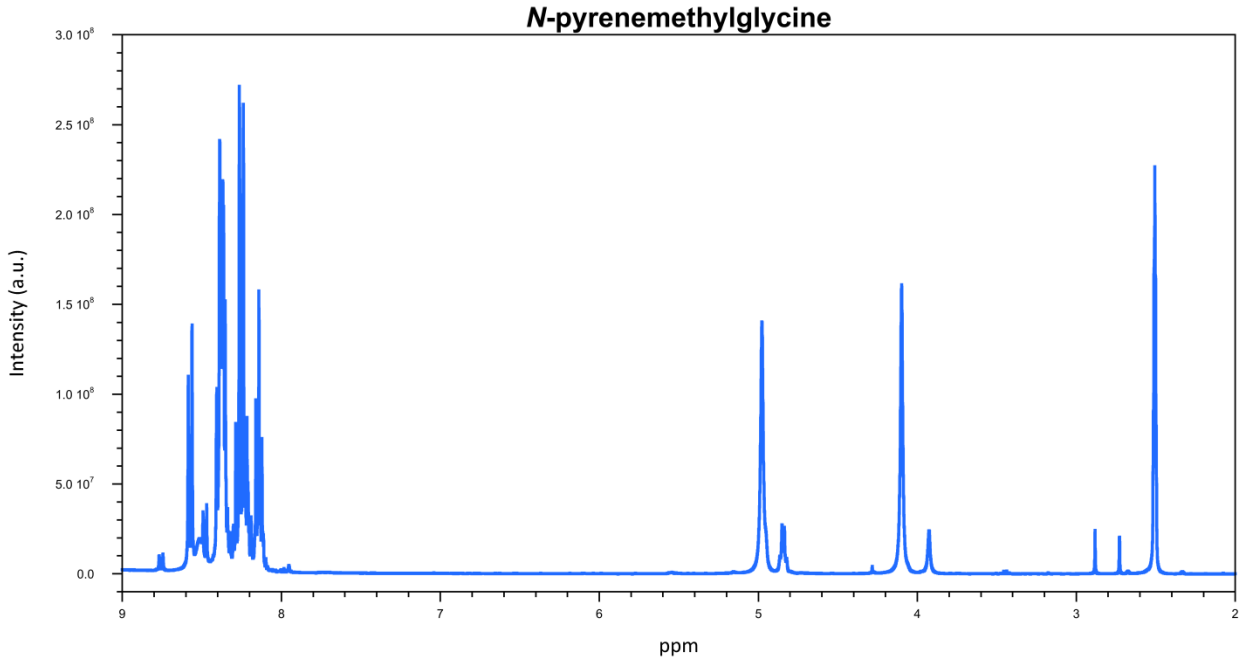

Figure 3.12: ^{1}H NMR (DMSO-d_6 + 0.1% TFA)of *N*-pyrenemethylglycine; DMSO-d_6 solvent peaks appear at 2.55 ppm (s).

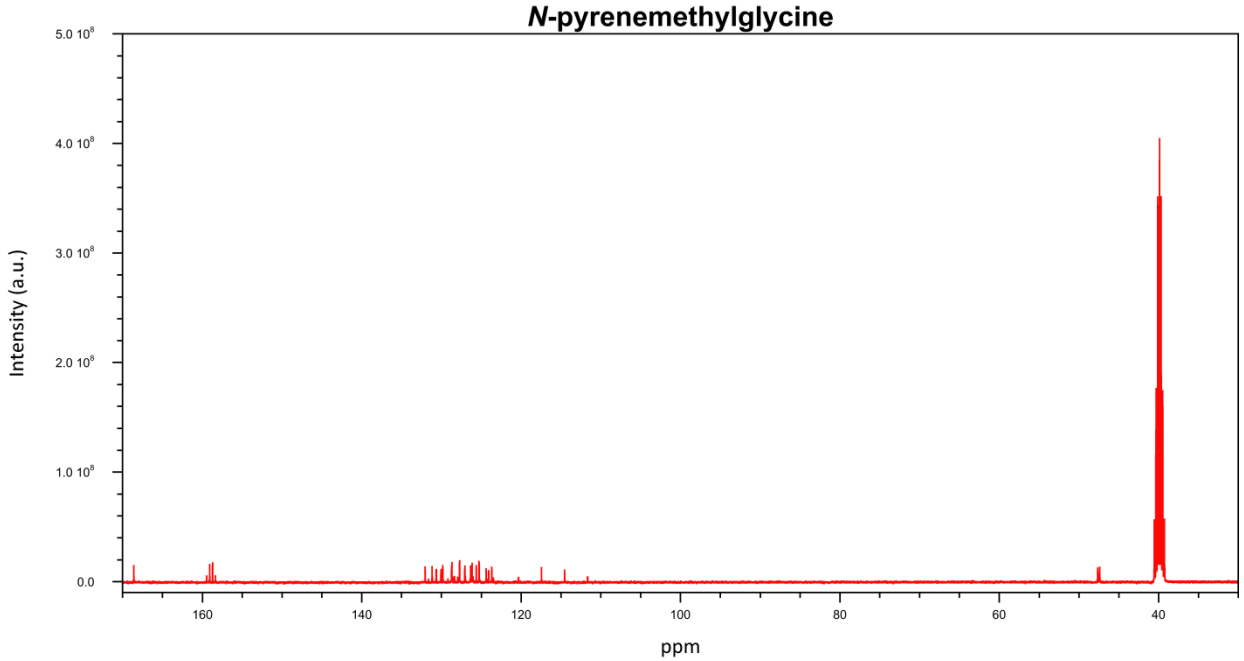

Figure 3.13: ^{13}C NMR (DMSO-d_6 + 0.1% TFA) of *N*-pyrenemethylglycine; DMSO-d_6 solvent peaks appear at 39.55 ppm (multiplet).

3.4: pK_a Determination of *N*-pyrenemethylglycine (Pmg)

A noticeable feature with *N*-pyrenemethylglycine is its ability to offer pH-responsiveness adjacent to the pyrene probe. This heightened functionalization could prove to be valuable, particularly in peptide chemistry.[45] For example, the Wennemers group utilized stereoisomers of aminoproline (Amp) to tune the thermal stability of trimers based on pH. The properties of amino versus ammonium groups led to the switching of the ring conformations of the proline derivatives.[45]

Separate vials containing Tris/HCl-buffered water ([Tris]: 0.12 M, 3.00 mL) were prepared, covering a pH range from 6.70 to 10.40. Using an accurate syringe, a stock solution of *N*-pyrenemethylglycine dissolved in DMSO ([**Pmg**]: 0.0050 M, 10.0 μL) was added to each vial. An electronic absorption spectrum was acquired (Figure 3.14) for the starting sample (pH: 10.40); it had a λ_{max} of 342 nm and Absorbance (*A*) less than 0.8. All samples were transferred sequentially to a quartz cuvette and were excited at 342 nm. The integrated emission intensity (*F*) from 410 to 525 nm was recorded. A plot of *F* versus pH was generated and was fit to the equation[31] (Figure 3.15):

$$F = F_{min} + \frac{F_{max} - F_{min}}{1 + e^{\frac{pH - pK_a}{\rho}}}$$

Equation 3.1: Two-state acid dissociation equation

The values of pK_a and rho (ρ) varied freely from initial guesses of pK_a: 8 and ρ: 0.4. *R* values for the three trials were > 0.995. This determination was completed in triplicate to provide a pK_a value of 8.46 ± 0.06.

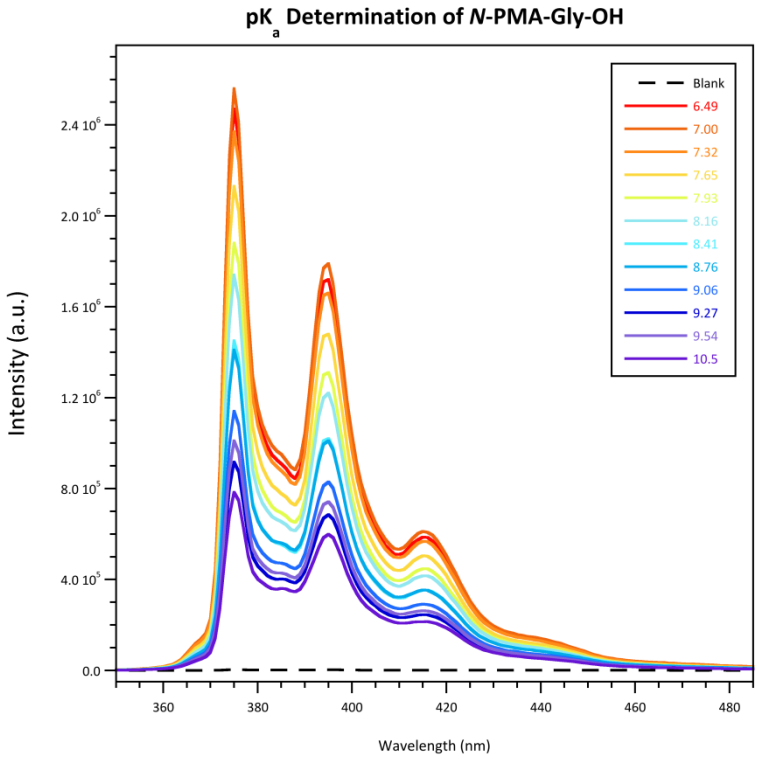

Figure 3.14: pH-dependent fluorescence emission study of **Pmg**

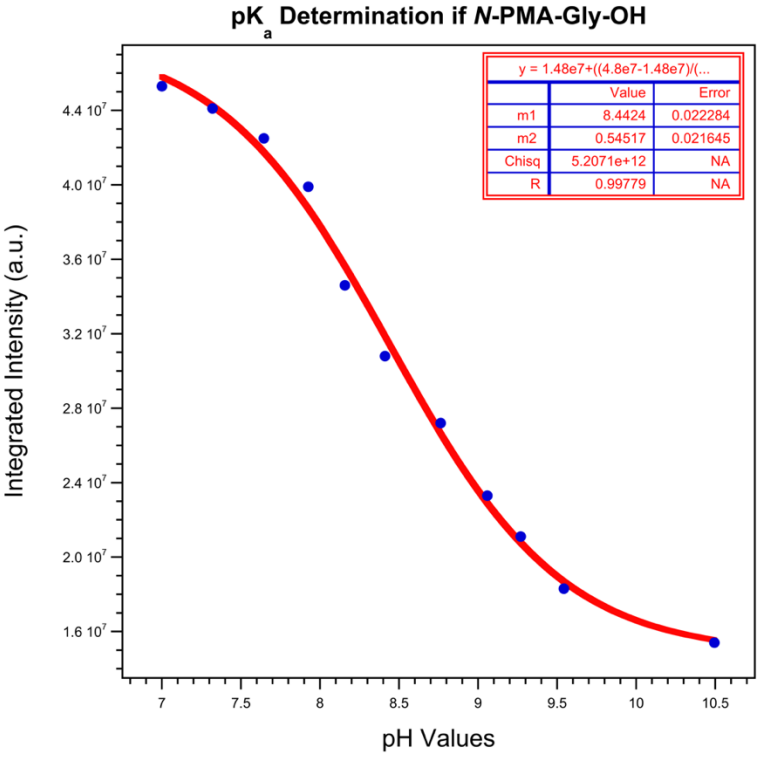

Figure 3.15: pK_a determination of **Pmg**

3.5 Synthesis and Characterization of *N*-pyrenemethylglycine-bearing CRPs

CRPs derived from the $Ac(POG)_7NH_2$ template were selected for study. Ultimately, **Pmg** was designed to be incorporated on various locations of the backbone. It was decided that three peptides containing **Pmg** on either the N-terminal, the central, or the C-terminal triplet would be synthesized.

Table 3.1: List of *N*-pyrenemethylglycine-bearing CRPs, with associated abbreviations

Abbreviation	**Formulae**
XPO_1	Pmg-Pro-Hyp-$(GPO)_6$-NH_2
XPO_4	$(GPO)_3$-Pmg-Pro-Hyp-$(GPO)_3$-NH_2
XPO_7	$(GPO)_6$-Pmg-Pro-Hyp-NH_2

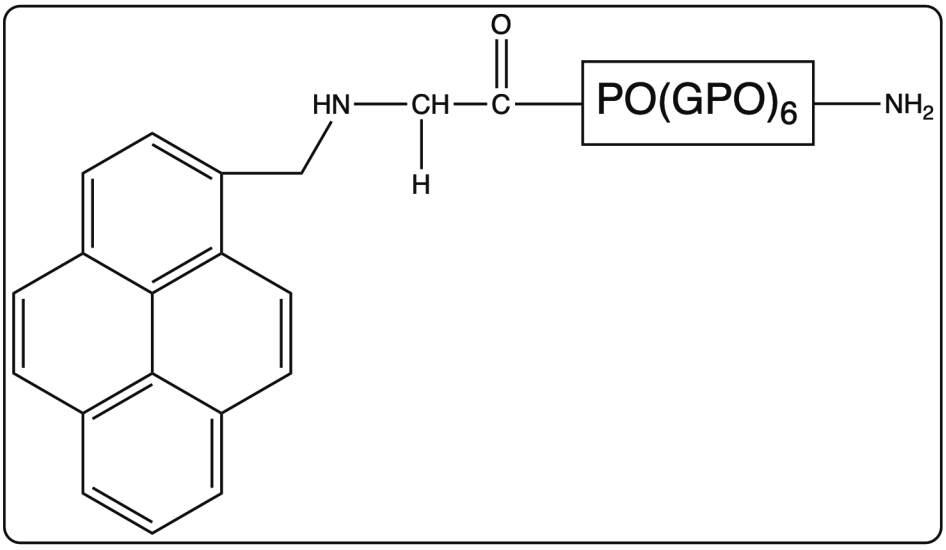

Figure 3.16: The **XPO_1** system

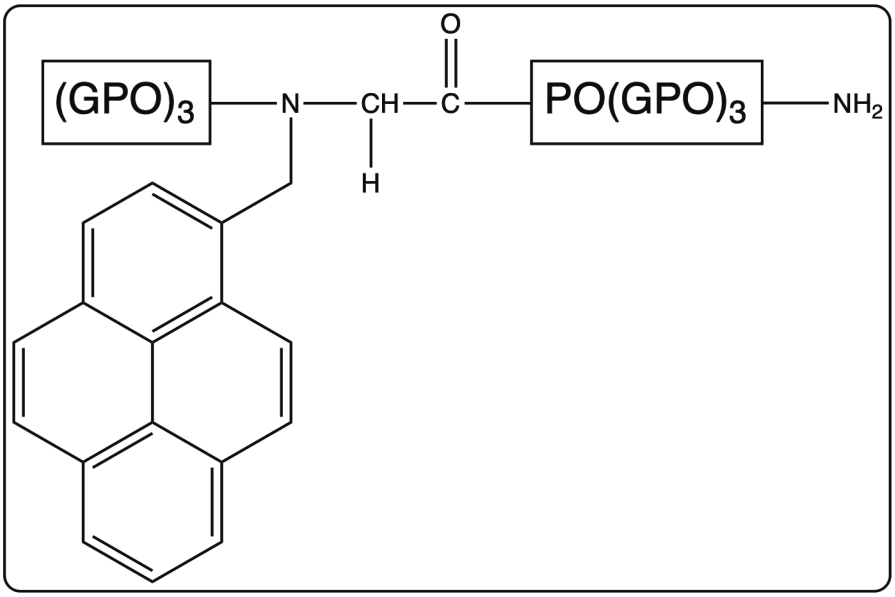

Figure 3.17: The **XPO_4** system

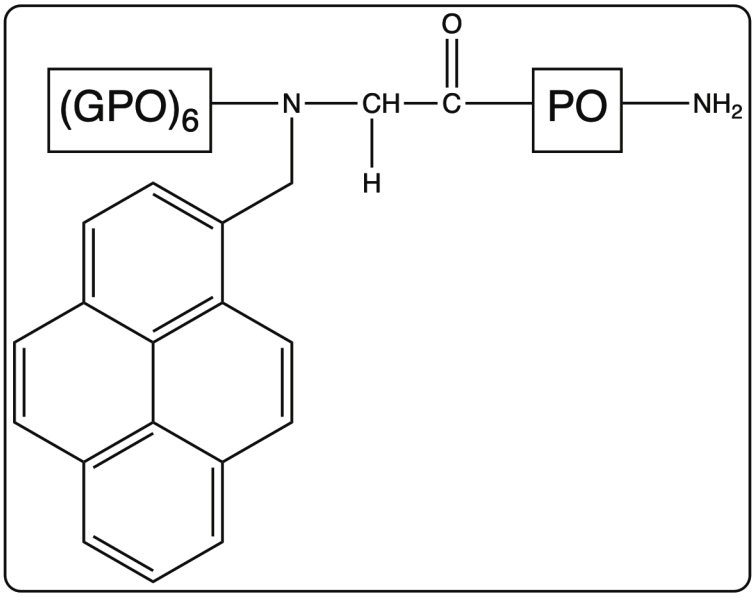

Figure 3.18: The **XPO_7** system

Peptides were synthesized using Fmoc-based SPPS on Rink amide resin. Fmoc-protected amino acids (4 eq.) were coupled in the presence of HBTU (4 eq.) using *N*,*N*-dimethylformamide (DMF) as a solvent, 20% 4-methylpiperidine in DMF as a deprotectant, and 7% *N*,*N*-diisopropylethylamine (DIPEA) in DMF as an activant, in an automated fashion. The derived

amino acid **Pmg** was manually coupled using double the amount of reagents relative to the automated protocol and allowed to mix for twice as long. An additional reagent was also introduced to mitigate the poor solubility of **Pmg**, tetraethylammonium bicarbonate. In the automated synthesis, each coupling step involving Pro and Hyp was programmed to occur twice without removal of the Fmoc protecting group from the first coupling step (double-coupled). C-terminal amino acids were also double coupled.

Following synthesis, the resulting crude resin was washed using suction filtration with glacial acetic acid, dichloromethane, and methanol, sequentially. Following washing, the resin was dried under vacuum for > 4 hours. Once dried, the resin was cleaved from the peptide with a mixture of trifluoroacetic acid, triethylsilane, and anisole (95:2.5:2.5) for 3 hours.

Following cleavage, the resin was removed, and the filtrate was introduced to chilled diethyl ether, and the suspension left to complete precipitation overnight. Following equilibration, the peptide was filtered out of solution by suction as a white solid. The precipitate was then purified by HPLC for analysis.

Purifications were performed on a *BioLogic DuoFlow* HPLC using and semi-preparatory C–18 columns to purify peptides. Mobile phases were composed of deionized water and acetonitrile containing trifluoroacetic acid (1% v/v). Each run was completed using a solvent gradient and was monitored by UV-Vis (214 nm, 254 nm, 280 nm, and 355 nm). Species could then be analyzed by mass spectrometry to determine the success of each synthesis.

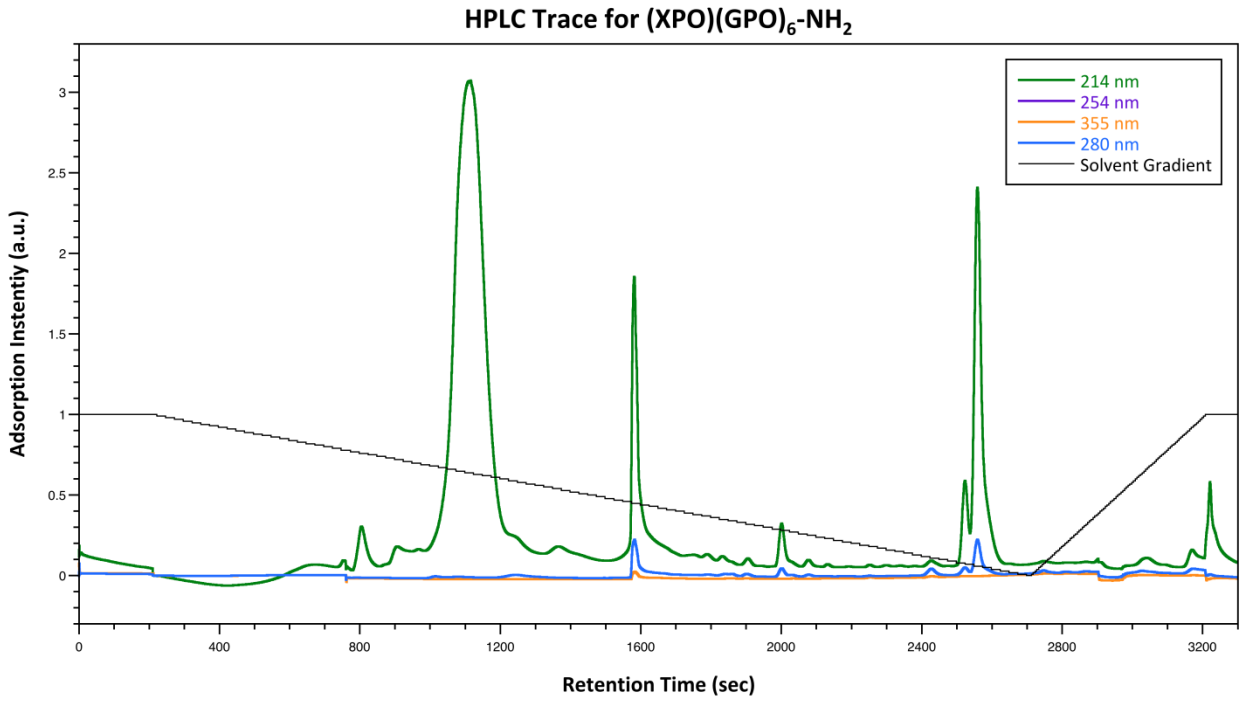

Figure 3.19: HPLC purification of **XPO$_1$**; presumed desired material was collected at 25 min. A gradient of 99.9% water/0.1% TFA to 99.9% acetonitrile/0.1% TFA was used.

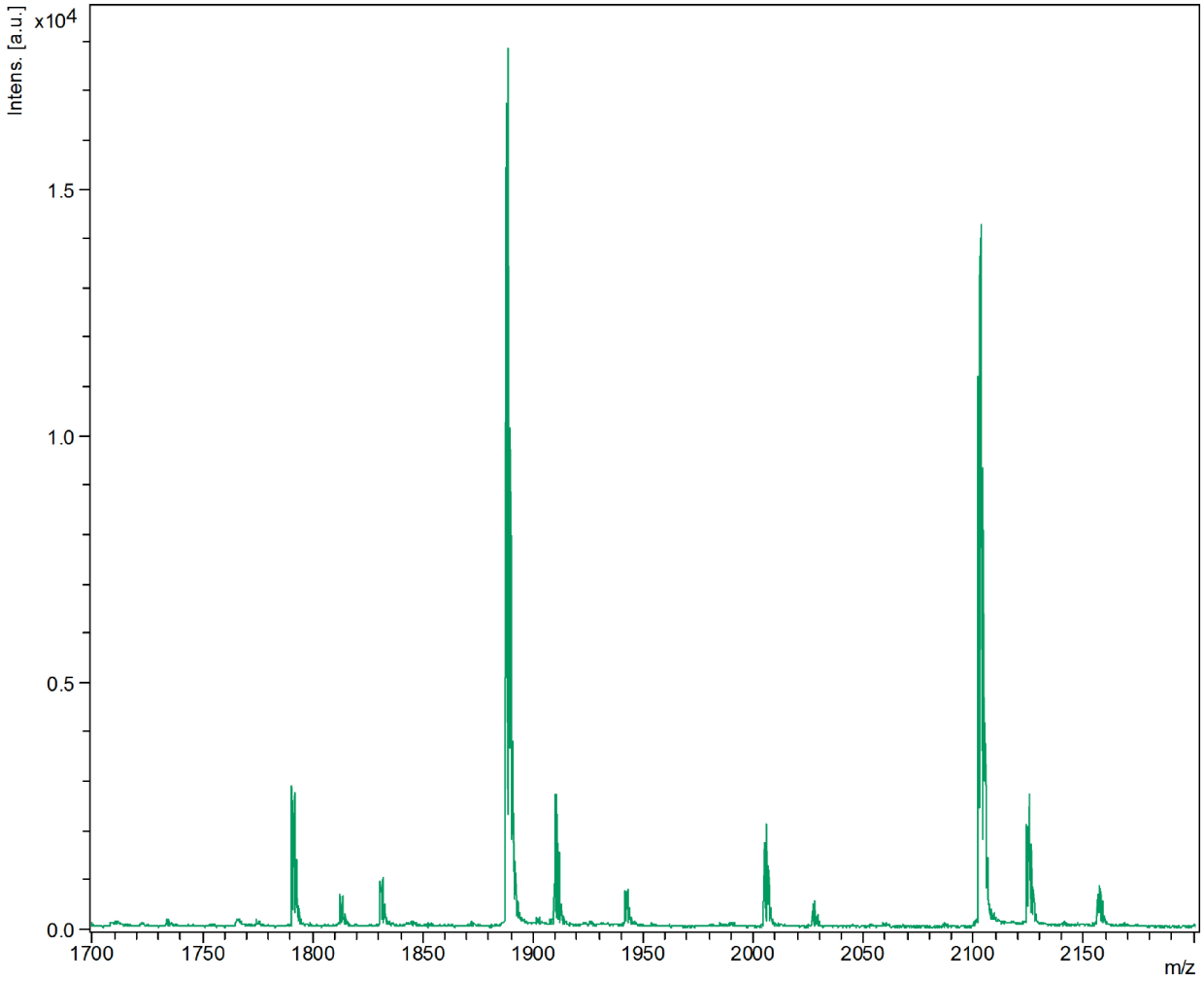

Figure 3.20: MS of the **XPO$_1$**; note the presence of *M/Z*: 2102.382 + 1H$^+$ (expected: 2102.97). A gradient of 99.9% water/0.1% TFA to 99.9% acetonitrile/0.1% TFA was used.

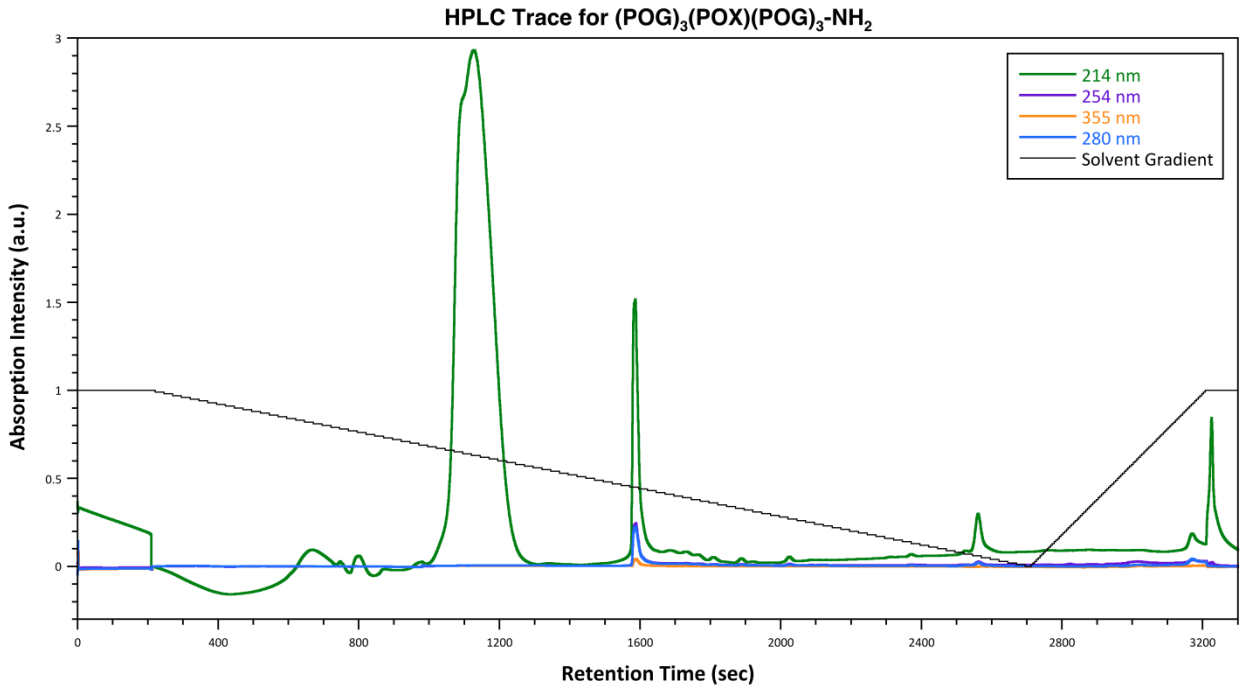

Figure 3.21: HPLC purification of **XPO$_4$**; presumed desired material was collected at 25 min. A gradient of 99.9% water/0.1% TFA to 99.9% acetonitrile/0.1% TFA was used.

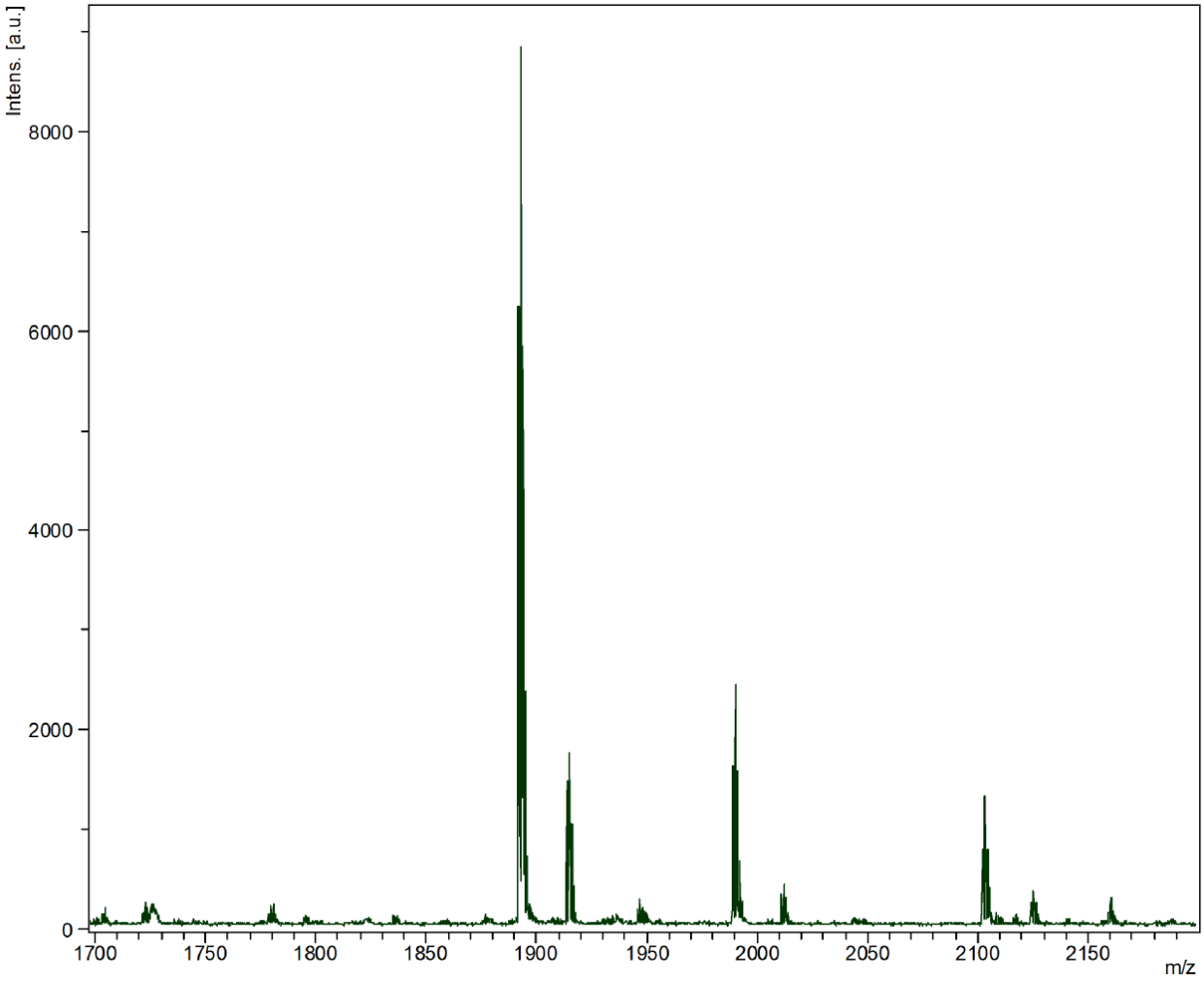

Figure 3.22: MS of the **XPO_4**; note the presence of *M/Z*: 2102.402 + $1H^+$ (expected: 2102.97). A gradient of 99.9% water/0.1% TFA to 99.9% acetonitrile/0.1% TFA was used.

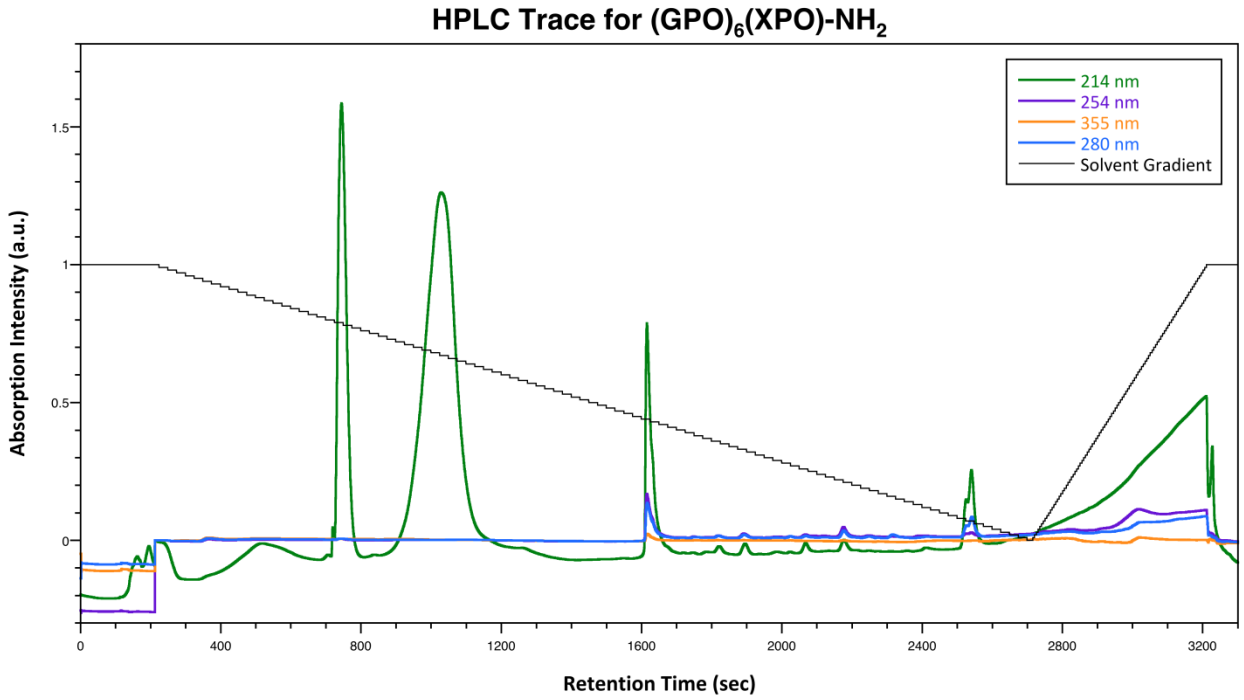

Figure 3.23: HPLC purification of **XPO_7**; presumed desired material was collected at 25 min.

A gradient of 99.9% water/0.1% TFA to 99.9% acetonitrile/0.1% TFA was used.

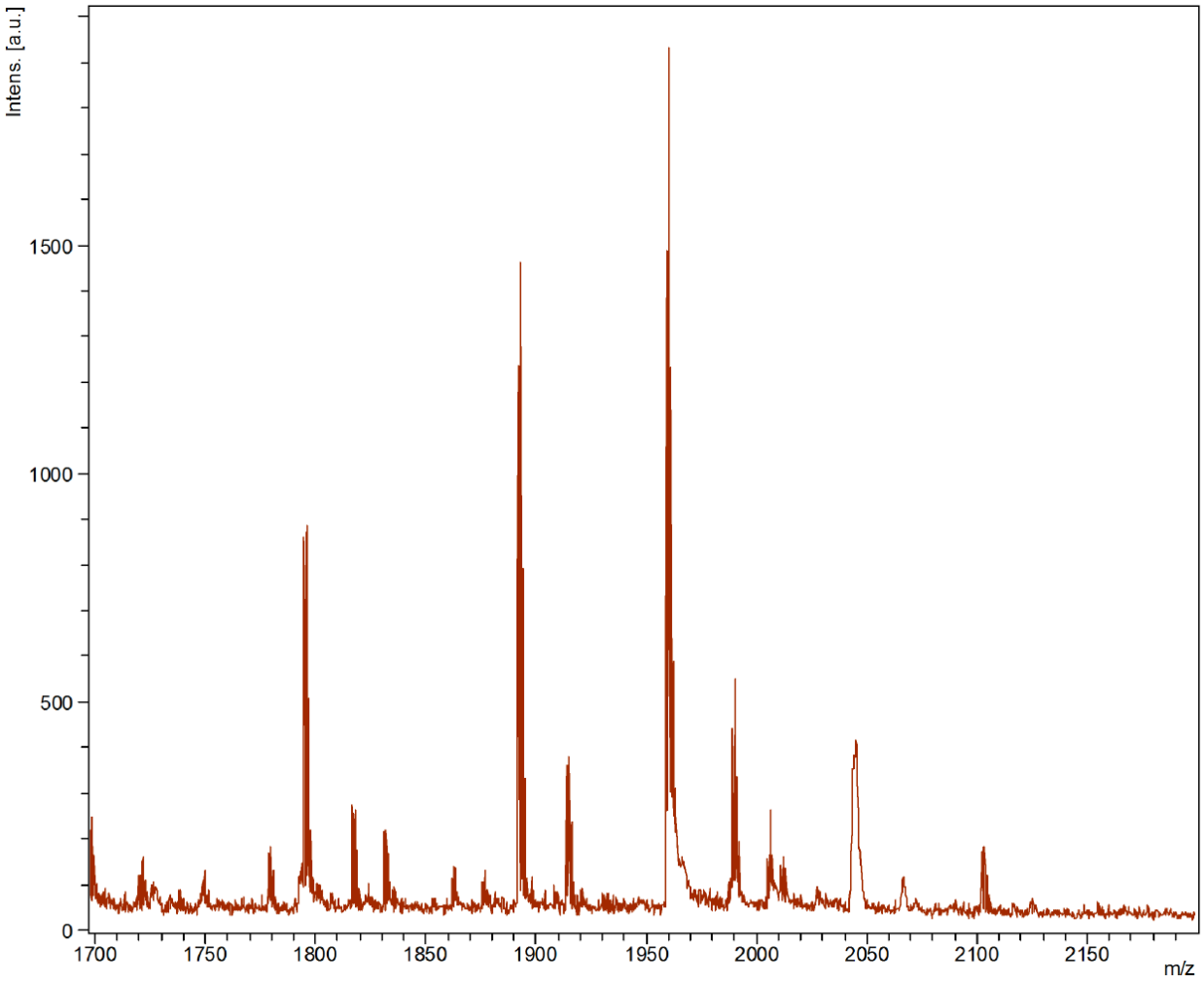

Figure 3.24: MS of the **XPO$_7$**; note the presence of *M/Z*: 2101.640 + 1H$^+$ (expected: 2102.97)

*Discussion on the difficulty of **Pmg** coupling*

Notably, the **Pmg** reagent has poor solubility in DMF. This may be due to the reasons discussed previously.[44] Several attempts into remaking the *N*-pyrenemethylglycine-bearing CRPs resulted in mixed/low yields of the desired product. It was also presumed that low yield during synthesis was due to in part of the lack of a Fmoc-protecting group on the compound, as it possess a secondary amine that can unwantedly deprotect in activating steps during SPPS .

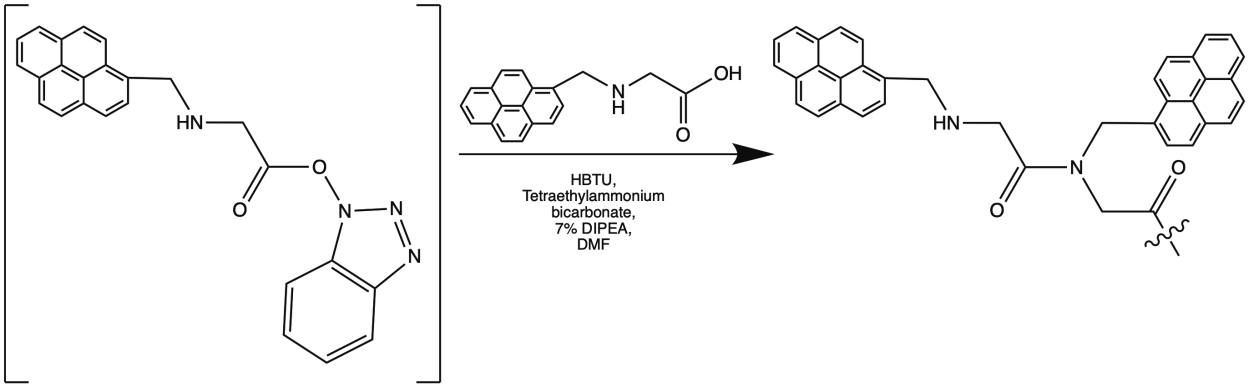

Figure 3.25: The **Pmg** self-coupling issue

In a final attempt to address the issue, efforts to add a Fmoc-protecting group were conducted. **Pmg** (2 mmol) and Fmoc-OSu (2 mmol) were suspended in dioxane and H_2O (2:1, 30 mL). Tetraethylammonium bicarbonate (2 mmol) and sodium bicarbonate (2 mmol) were added shortly for solubility and activation purposes, respectively. The suspension was allowed to stir overnight. Upon completion, it was presumed that the reaction had failed to occur due to the presence of precipitate after stirring overnight.

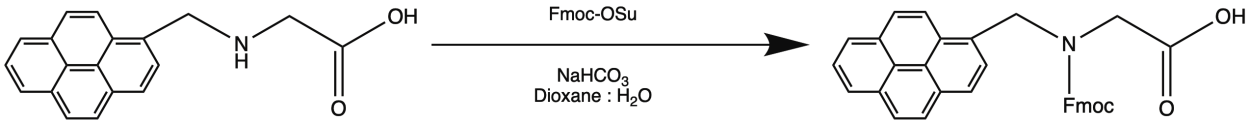

Figure 3.26: The protection step

3.6 Conclusions on Pmg

The addition of a pyrene tag to an amino acid, while synthetically difficult, can allow for the thermodynamic and kinetic study of peptides. If work into making such amino acid derivatives continues, addressing the simultaneous presence of hydrophobic and zwitterionic components directly could alleviate synthetic and purification roadblocks. For example, working with starting reagents with large protecting groups on the carboxylate ends eliminates the possibility of zwitterions upon addition of the pyrene probe. In turn, additional syntheses to

protect the N-termini for SPPS are viable. While technical limitations prevented the goal of implementing a fluorophore with charge capabilities throughout the backbone of CRPs, this work underpins the desires to investigate the role of localized charge and collagen behavior.

CHAPTER 4: Dimethylated ɣ-Azaprolines Induce Instability within Collagen Triple Helices

4.1: Introduction to ɣ-Carbon Modifications

Proline derivatives are valuable tools for studying and adjusting the structure and function of peptides and proteins. The ɣ-carbon within the pyrrolidine ring has a deep library of modifications.

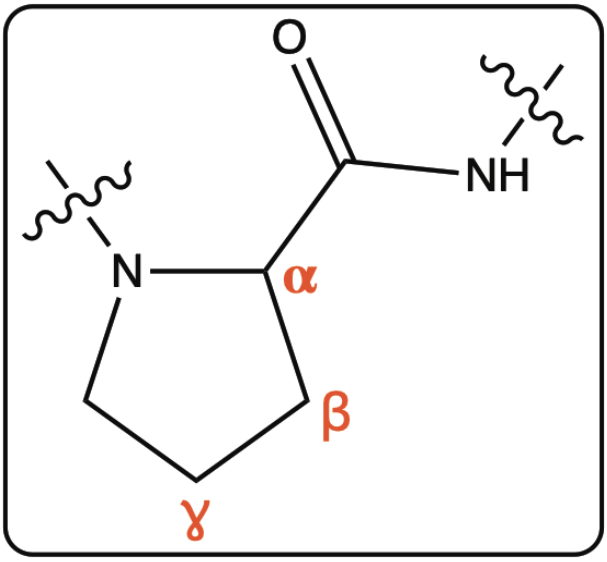

Figure 4.1: Nomenclature on atoms of the pyrrolidine ring

Such modifications can affect the conformation of the ring, leading to changes on peptide and protein folding as a whole. In previous studies, (2S,4R)-4-methylproline has a strong preference for the C_γ-endo pucker, whereas the stereoisomeric (2S,4S)-4-methylproline has a strong preference for the C_γ-exo pucker.[46] These methylated prolines were implemented within CRPs to induce varying degrees of instability.[46] In another study, 4-aminoproline residues underwent conformational change depending on the protonation state of the ɣ-amine.[45] This modification was also used to tune the thermal stability of collagen triple helices based on the acidity of the

environment.[45] This chapter details of a synthesis a novel proline derivative Fmoc-AzP(DM)-OH (**Map**), which features a persistent positive charge, and its thermodynamic effect on CRPs.

4.2: Synthetic Design of Fmoc-AzP(DM)-OH (Map)

Synthesis of Fmoc-AzP-(OH)

The overall approach to **Map** involves cyclization to form the five-membered ring. Fmoc-(L)-Dap-OH (4 mmol) was suspended in acetonitrile (60 mL) and heated for 5 min to assist in solvation. Paraformaldehyde (4 mmol) and *p*-TsOH (0.2 mmol) were added to the suspension which was heated to reflux (2.5 hours). After 1.5 hours, a yellow precipitate was observed. TLC analysis (MeOH) revealed a new spot (R_f: 0.90) with a faint appearance under a UV-lamp (254 nm) while starting material was still present (R_f: 0.20). After cooling the suspension to room temperature and then placing it in an ice bath (45 min), the yellow precipitate was collected using a fine glass frit and washed with chilled ACN. Consequently, the product, dubbed *Fmoc-AzP-(OH)* was left to dry under vacuum (75 %).

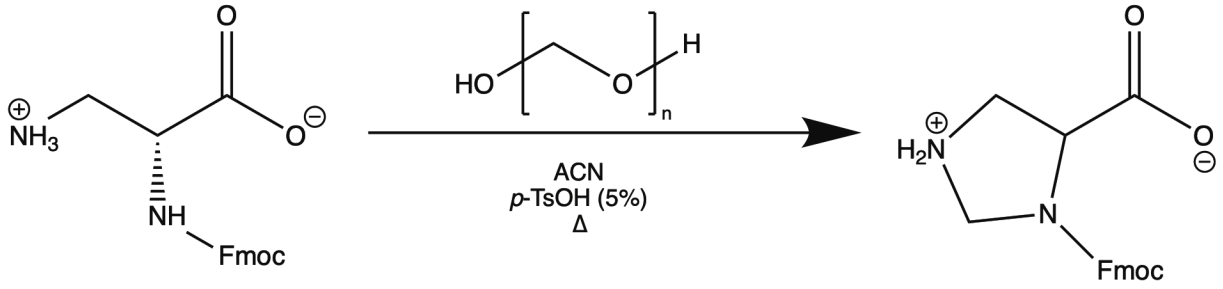

Figure 4.2: Synthetic Scheme of Fmoc-AzP-OH

^{1}H NMR revealed the simultaneous presence peaks in the aromatic region (6.5 – 8.0 ppm) and diastereoscopic hydrogens from the newly formed ring (~4.5 ppm). ^{13}C NMR accounted for all the projected carbons. A signal with *M/Z* = 341.134 was accounted for in MS.

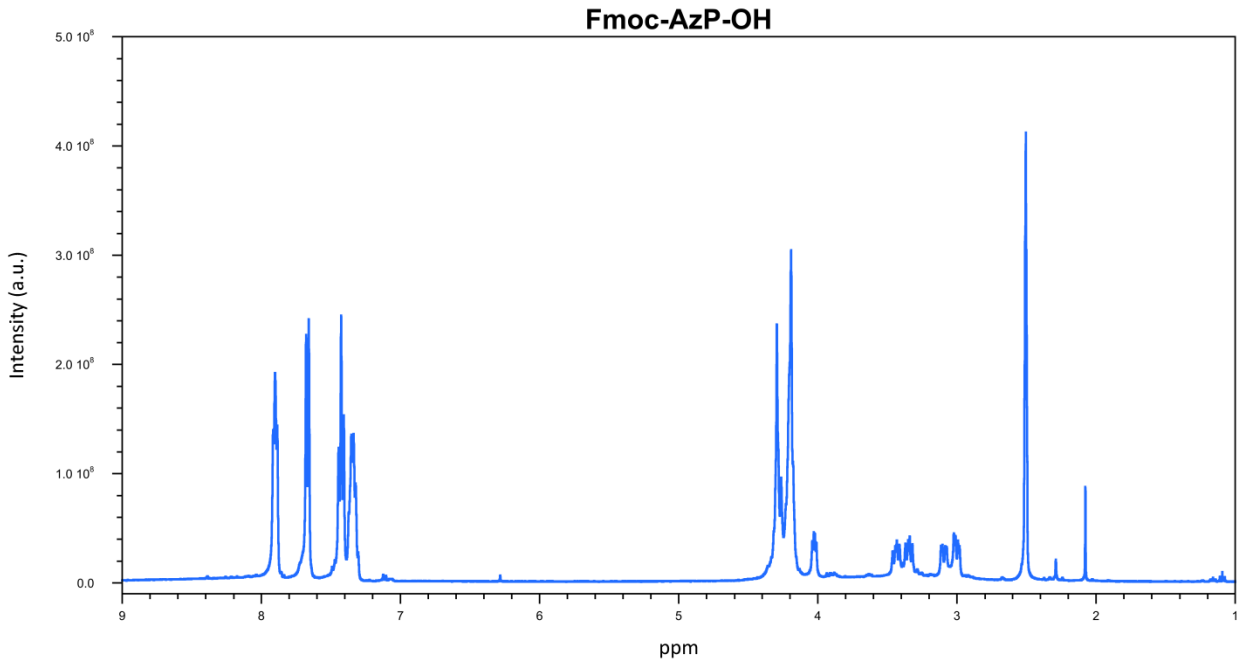

Figure 4.3: ^{1}H NMR (DMSO-d_6) of Fmoc-AzP-OH; DMSO-d_6 solvent peaks appear at 2.50 ppm (s).

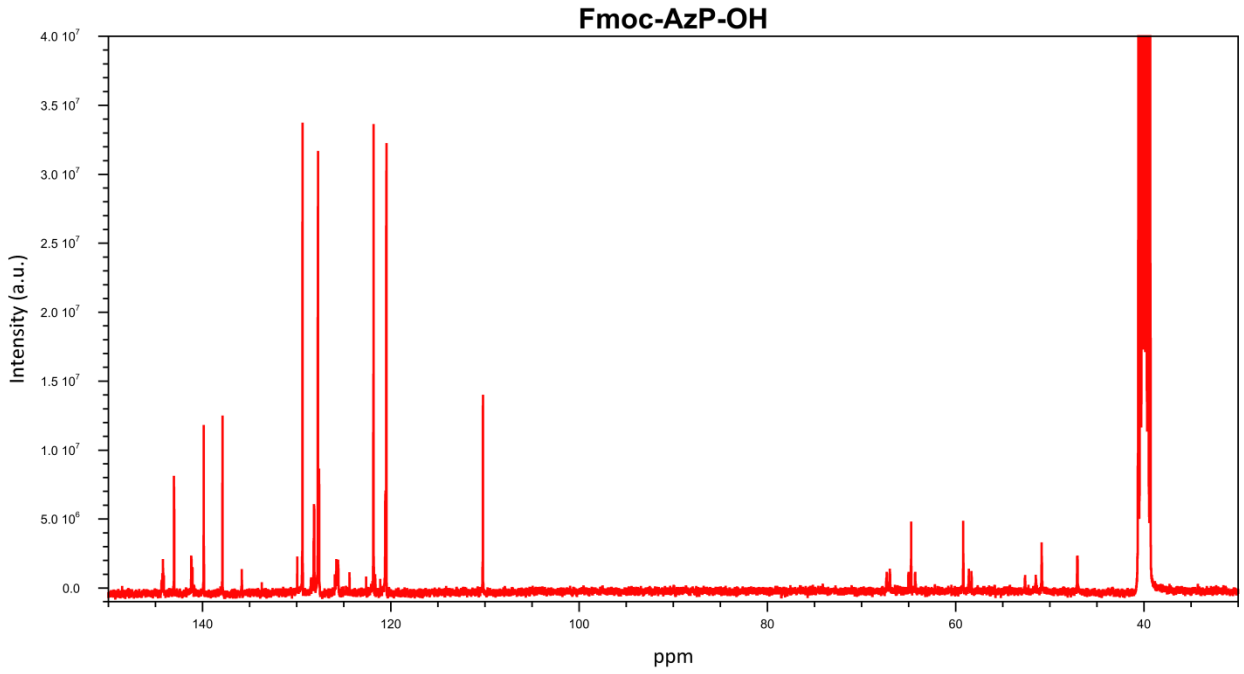

Figure 4.4: ^{13}C NMR (DMSO-d_6) of Fmoc-AzP-OH. DMSO-d_6 solvent peaks appear at 39.55 ppm (multiplet).

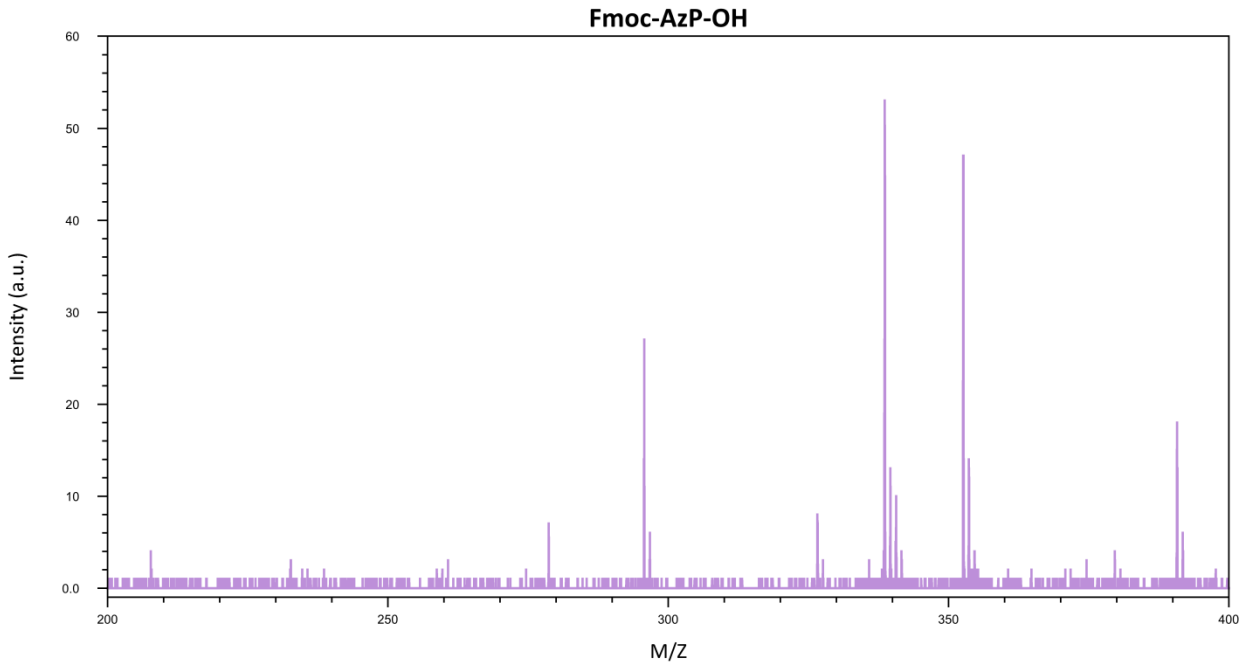

Figure 4.5: MS of Fmoc-AzP-OH; note the presence of *M/Z*: 338.43. + $1H^+$ (expected: 339.13)

*Synthesis of Fmoc-AzP(DM)-OH (**Map**)*

With the nitrogen-containing proline ring in hand, efforts shifted to methylation. Fmoc-AzP-OH (2 mmol), iodomethane (4 mmol) and DIPEA (4 mmol) was dissolved in DMF (15 mL) and heated for 4 hours (35 °C). After 1 hour, TLC analysis (MeOH) revealed a new faint spot (R_f: 0.10) under a UV-lamp (254 nm) while starting material was still present (R_f: 0.90). Upon cooling the reaction to room temperature, the reaction was neutralized with AcOH and the solvent was removed by rotary evaporation. The dark yellow residue, dubbed **Map**, was left to dry under vacuum.

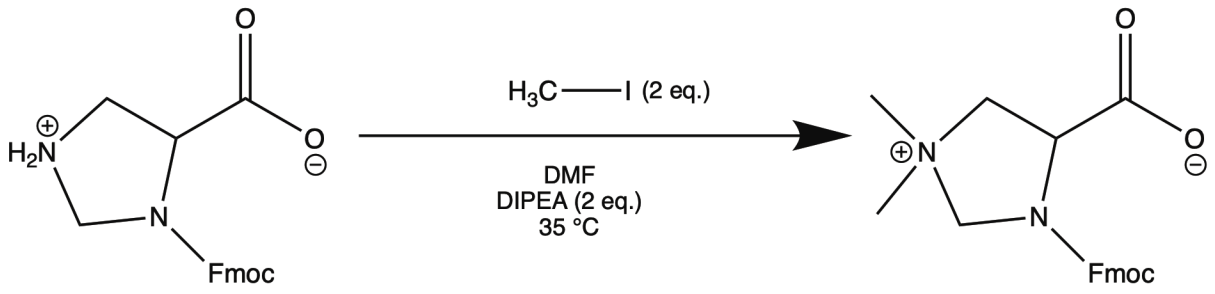

Figure 4.6: Synthetic Scheme of **Map**

^{1}H NMR analysis revealed a simultaneous presence peaks in the aromatic region (6.5 – 8.0 ppm) and methyl hydrogens (~2.5 ppm). ^{13}C NMR accounted for all the projected carbons. A *M/Z* signal of 367.434 + 1H^{+} was accounted for in MS.

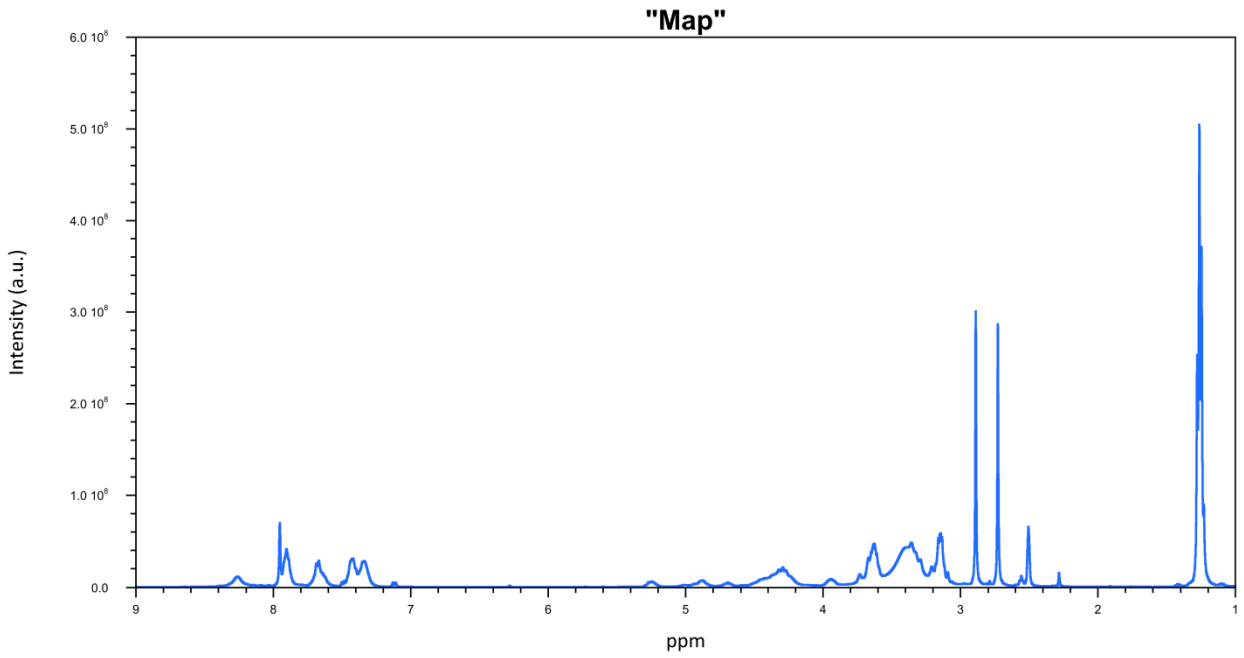

Figure 4.7: ^{1}H NMR of **Map**; residual hydrogens from DIPEA appear at 1.25, 3.38, and 2.41 ppm.

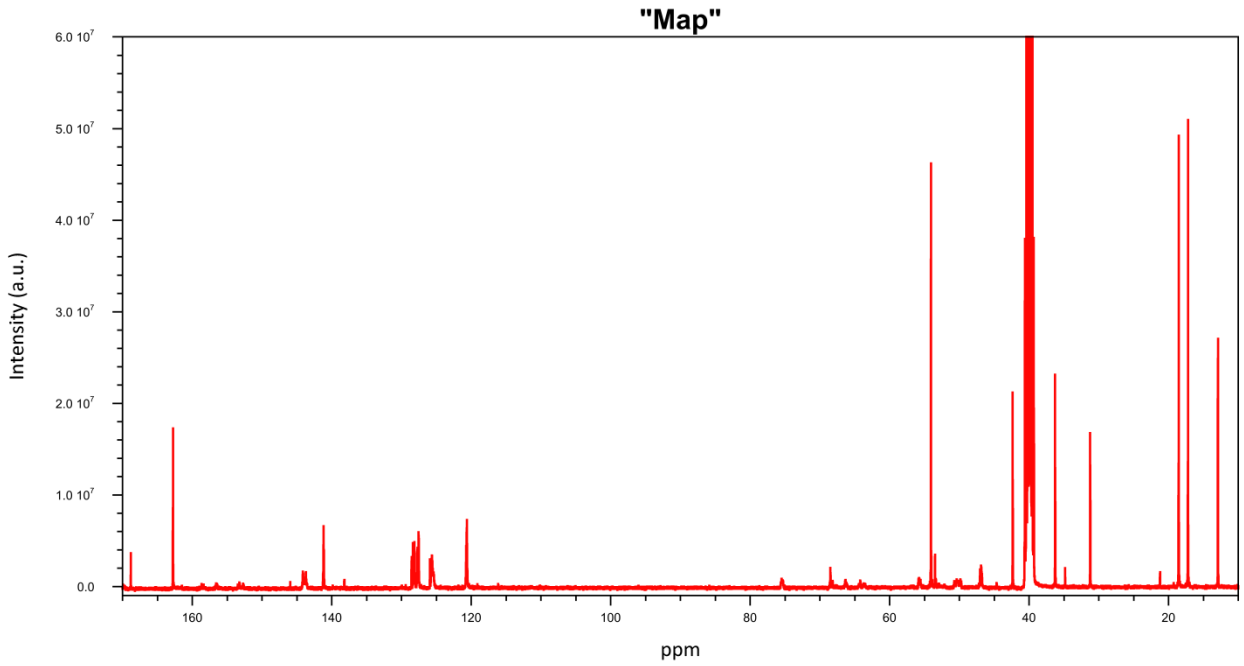

Figure 4.8: ^{13}C NMR of **Map;** DMSO-d_6 solvent peaks appear at 39.59 ppm (multiplet).

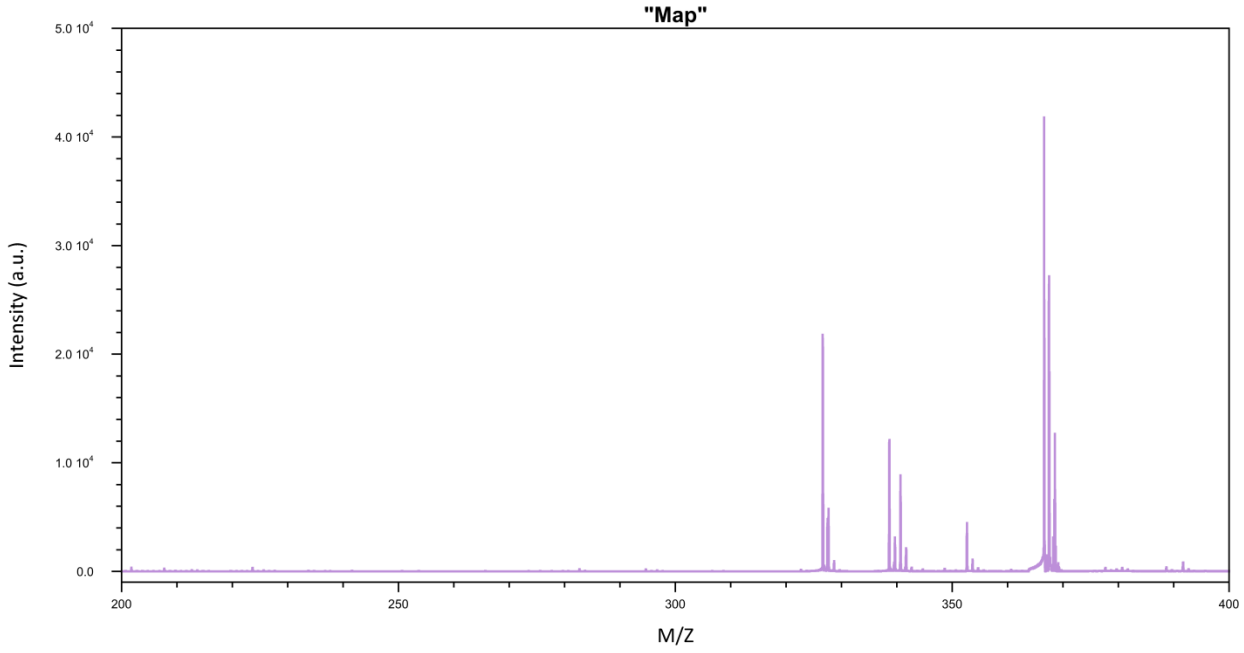

Figure 4.9: MS of **Map:** note the presence of *M/Z*: 367.427 + 1H$^+$ (expected: 367.17)

The difficulties in purifying Fmoc-AzP(DM)-OH was apparent when solubility issues returned upon the formation of the ammonium group. A compound with this group and the

hydrophobic Fmoc-protected group introduce the same issues described previously. However, NMR and MS suggested that the crude reaction provided a high enough yield where purification could be avoided.

4.3: Synthesis and Characterization of Map-Bearing CRPs

CRPs derived from the $Ac(POG)_7NH_2$ template were selected for study. Ultimately, **Map** was designed to be inserted at various locations of the backbone. It was decided that two peptides containing **Map** at either the N-terminal or the central triplet would be synthesized.

Table 4.1: List of **Map**-bearing CRPs, with associated abbreviations

Abbreviation	**Formulae**
POG_7	$Ac\text{-}(POG)_7\text{-}NH_2$
PXG_4	$Ac\text{-}(POG)_3\text{-Pro-Map-Gly-}(POG)_3\text{-}NH_2$
PXG_1	$Ac\text{-Pro-Map-Gly-}(POG)_6\text{-}NH_2$

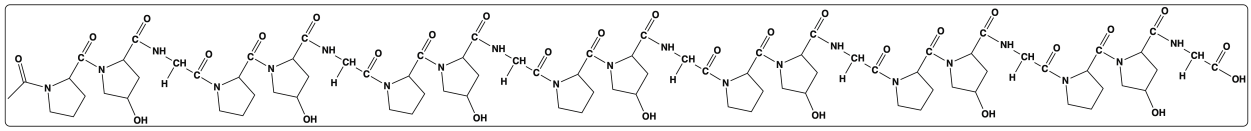

Figure 4.10: The **POG_7** system

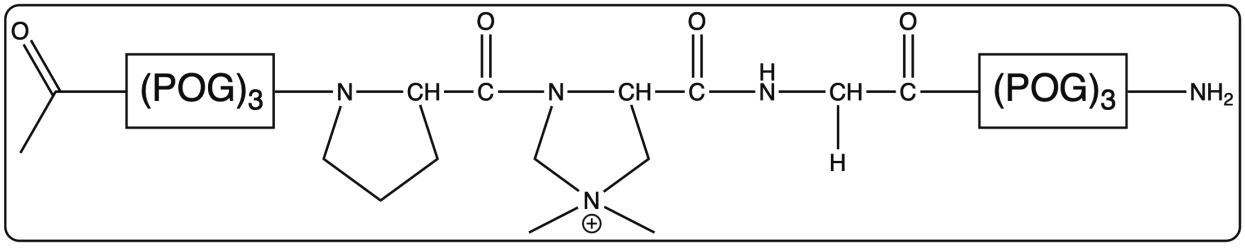

Figure 4.11: The **PXG_4** system

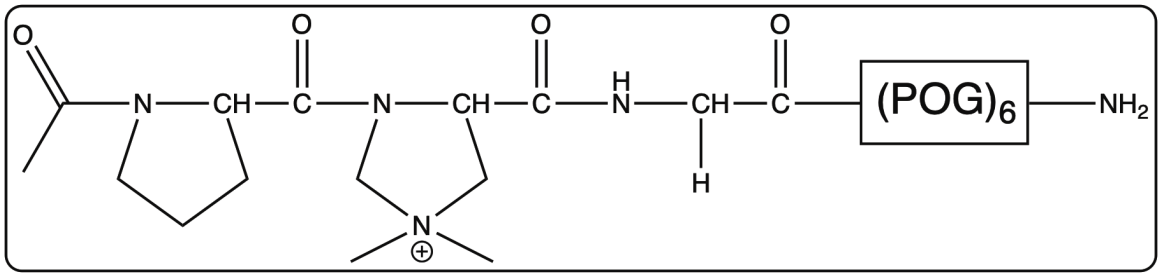

Figure 4.12: The **PXG$_1$** system

Peptides were synthesized using Fmoc-based SPPS on Rink amide resin. Fmoc-protected amino acids (4 eq.) were coupled in the presence of HBTU (4 eq.) using *N*,*N*-dimethylformamide (DMF) as a solvent, 20% 4-methylpiperidine in DMF as a deprotectant, and 7% *N*,*N*-diisopropylethylamine (DIPEA) in DMF as an activant in an automated fashion. The derived amino acid **Map** was manually coupled using double the amount of reagents relative to the automated protocol, and allowed to mix for twice as long. In the automated synthesis, each coupling step involving Pro and Hyp was programmed to occur twice without removal of the Fmoc-protecting group from the first coupling step (double-coupled). C-terminal amino acids were also double coupled.

Following synthesis, the resulting crude resin was collected by suction and washed over with glacial acetic acid, dichloromethane, and methanol, sequentially. Following washing, the resin was dried under vacuum for > 4 hours. Once dried, the resin was cleaved from the peptide with a mixture of trifluoroacetic acid, triethyl silane, and anisole (95:2.5:2.5), and by stirring for 3 hours.

Following cleavage, the resin was removed, and the filtrate was introduced to chilled diethyl ether and left to precipitate overnight. Following equilibration, the peptide was filtered

out of solution by suction as a white solid. The precipitate was then purified by HPLC for analysis.

Purifications were performed on a *BioLogic DuoFlow* HPLC using and semi-preparatory C–18 columns to purify peptides. Mobile phases were composed of deionized water and acetonitrile containing trifluoroacetic acid (1% v/v). Each run was completed over solvent gradient and was monitored by UV-Vis (214 nm, 254 nm, 280 nm, and 355 nm). Species could then be analyzed by mass spectrometry to determine the success of each synthesis.

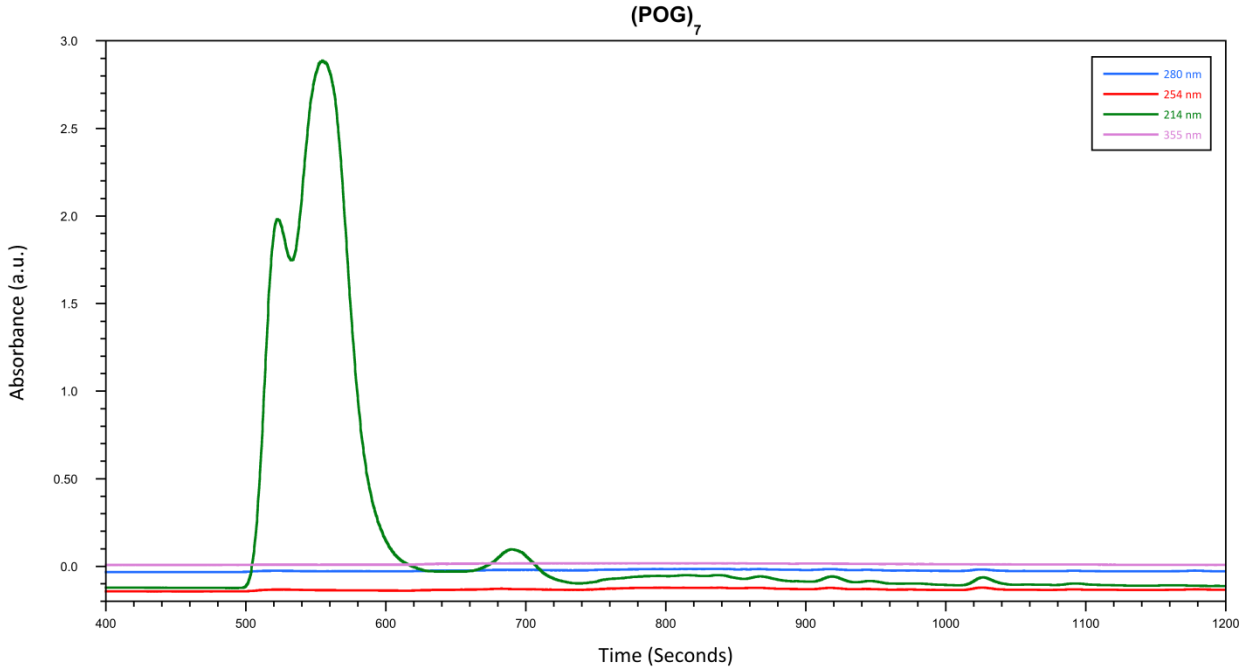

Figure 4.13: HPLC purification of **POG$_7$**; the peak collected at 8 min was subjected to MS. A gradient of 99.9% water/0.1% TFA to 99.9% acetonitrile/0.1% TFA was used.

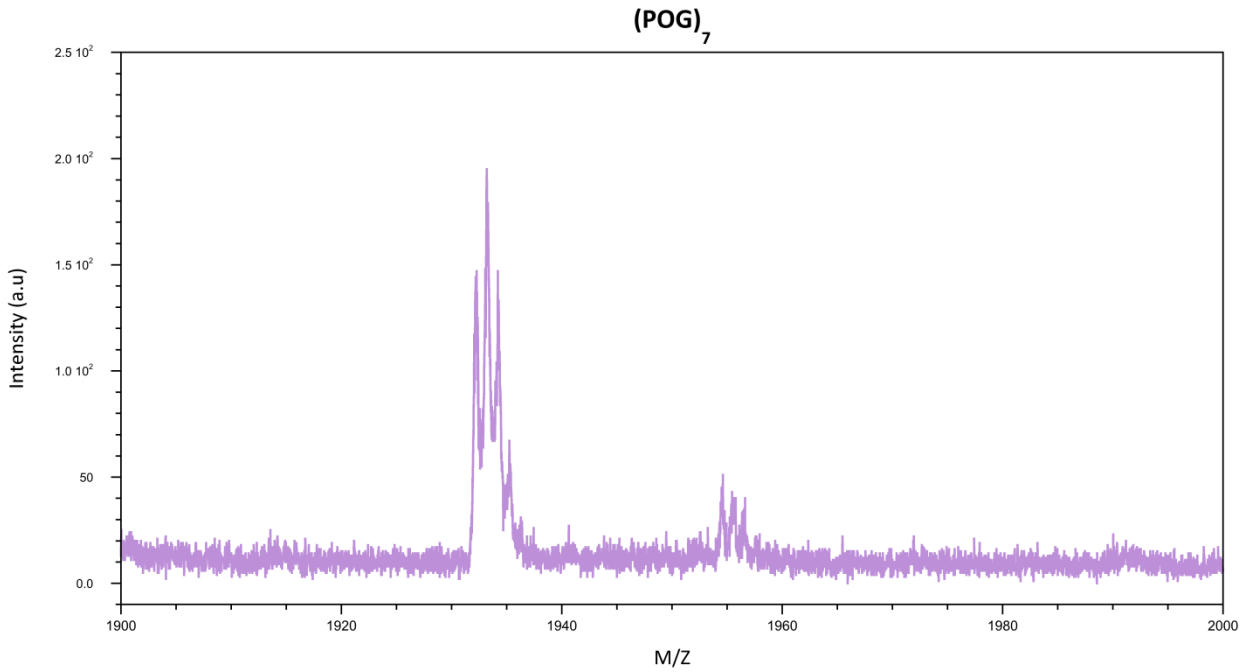

Figure 4.14: MS of **POG_7**; note presence of *M/Z*: 1930.12 + $1H^+$ (expected: 1929.87)

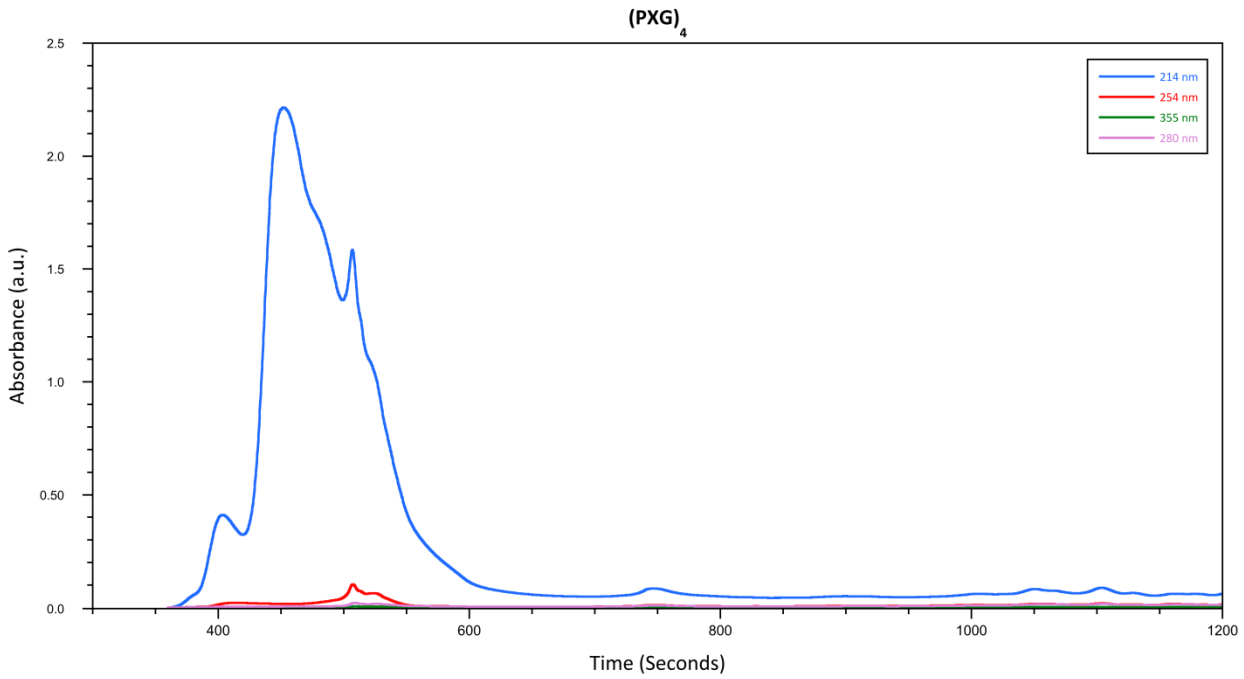

Figure 4.15: HPLC purification of **PXG_4**; the peak collected at 6 min was subjected to MS. A gradient of 99.9% water/0.1% TFA to 99.9% acetonitrile/0.1% TFA was used.

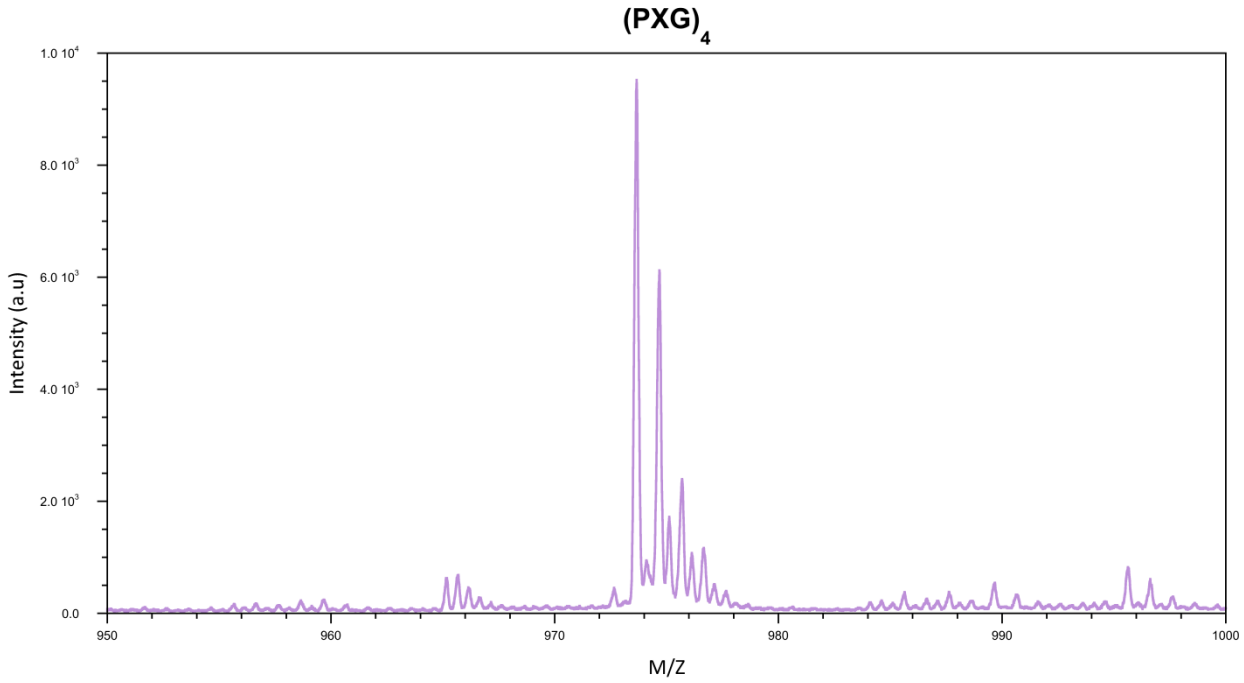

Figure 4.16: MS of **PXG$_4$**; note presence of *M/Z*: 974.14 + 2H$^+$ (expected: 1947.28)

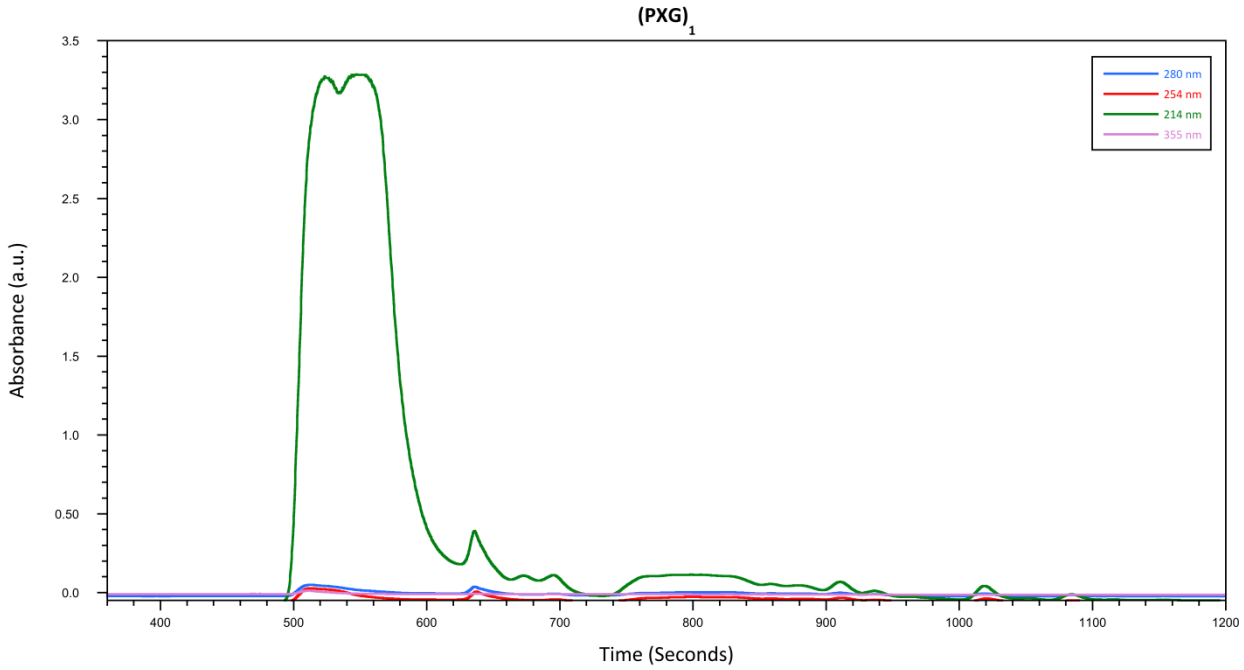

Figure 4.17: HPLC purification of **PXG$_1$**; the peak collected at 6min was subjected to MS. A gradient of 99.9% water/0.1% TFA to 99.9% acetonitrile/0.1% TFA was used.

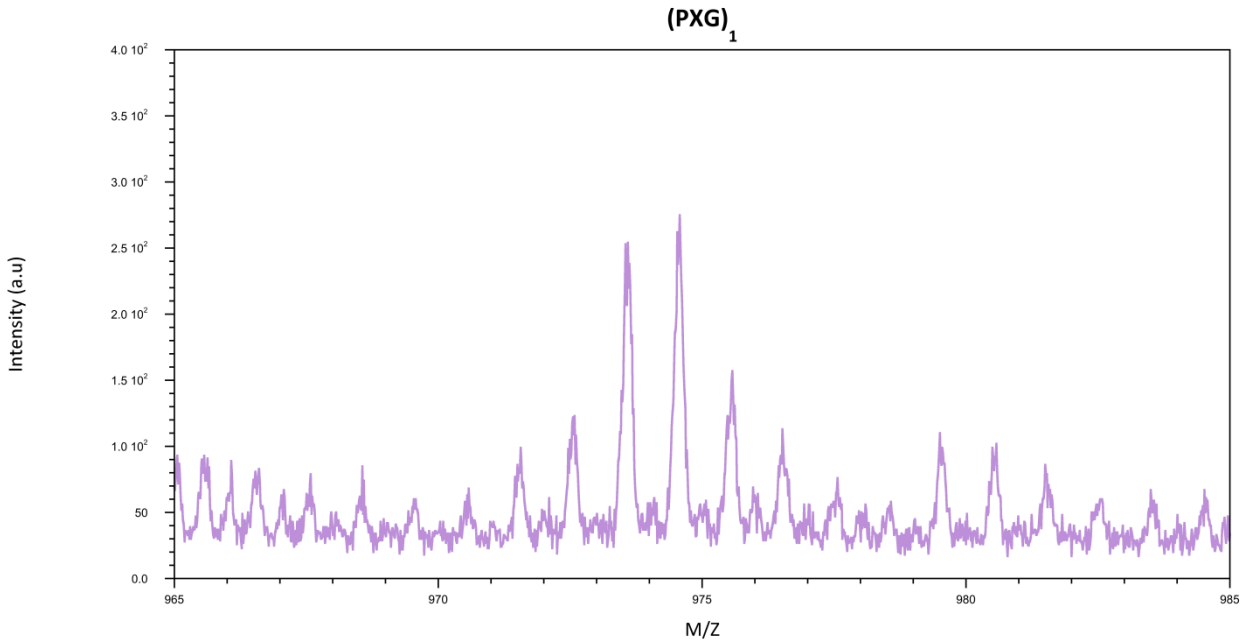

Figure 4.18: MS of **PXG$_1$**; note presence of *M/Z*: 974.12 + 2H$^+$ (expected: 1947.28)

4.4: Thermodynamic Stabilities of Triple Helices as Measured by CD

*CD Spectra of **Map**-Bearing CRPs*

CD spectroscopy was used to probe presence of the secondary structure of each system. First, samples of each system at were prepared in PBS solution at 200 μM for triple helical favorability with viable HT values of under 400 V at 224 nm. Systems were allowed to stand overnight at 4 ˚C in the dark to equilibrate before measurement. Spectra were acquired at 10 °C measuring the ellipticity from 190 - 260 nm. PPII helices were qualitatively inferred through presence of an ellipticity maximum at ~224 nm and minimum at ~200 nm (Figures 4.19 – 4.21).

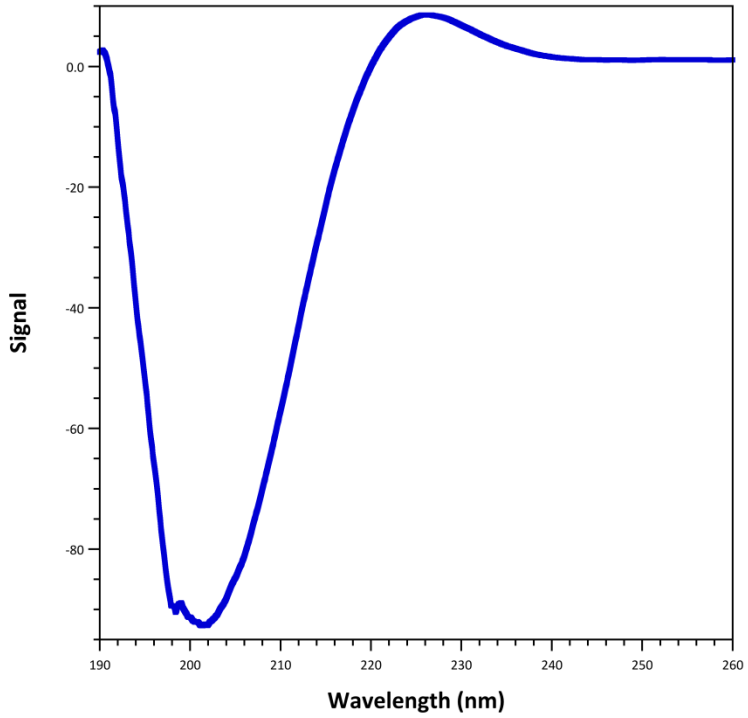

Figure 4.19: CD spectrum of **$(POG)_7$** at 0.20 mM in PBS at 10 ˚C

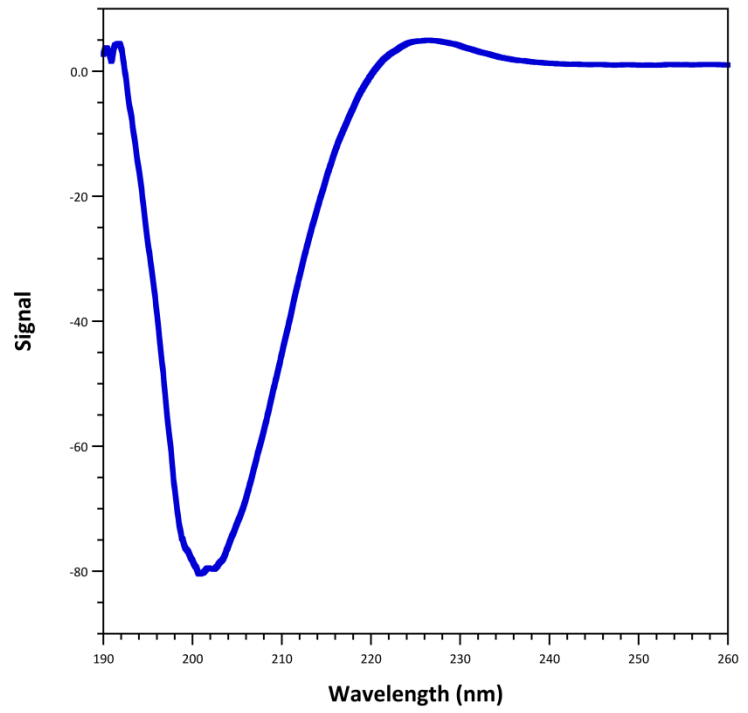

Figure 4.20: CD spectrum of **$(PXG)_4$** at 0.20 mM in PBS at 10 ˚C

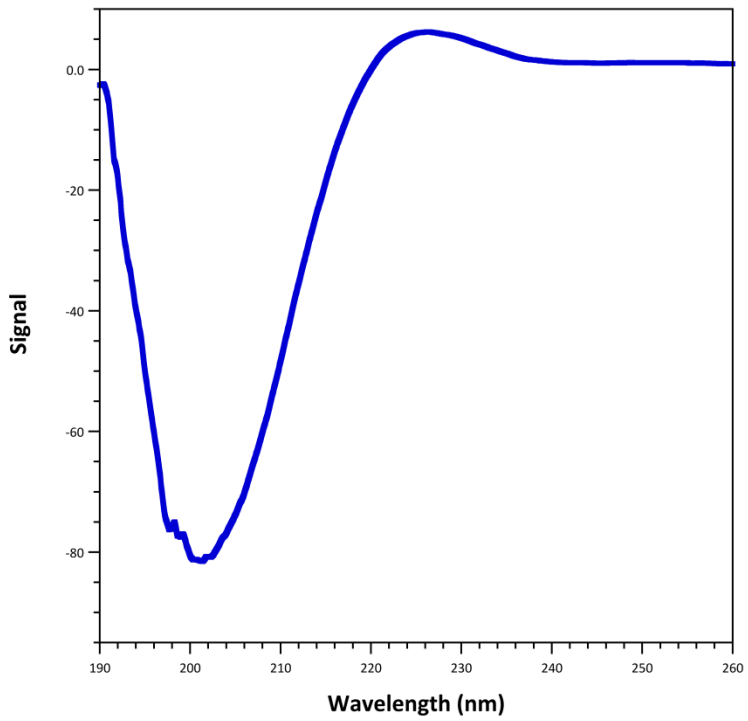

Figure 4.21: CD spectrum of **$(PXG)_1$** at 0.20 mM in PBS at 10 ˚C

*Melting Studies of **Map**-Bearing CRPs*

Temperature dependent measurements were conducted to measure the triple-helical stability of the CRPs. Each system was heated from 10 ˚C at a 0.2 ˚C/min rate while following the change in ellipticity at 224 nm. Measurements were then fitted to a two-state model (Figures 4.22 – 4.24) as described by previous studies[37], in which three monomeric chains coil into a helix. This model also offers an estimated transition enthalpy, ΔH°. All thermal denaturation experiments were performed in triplicate to ensure repeatability.

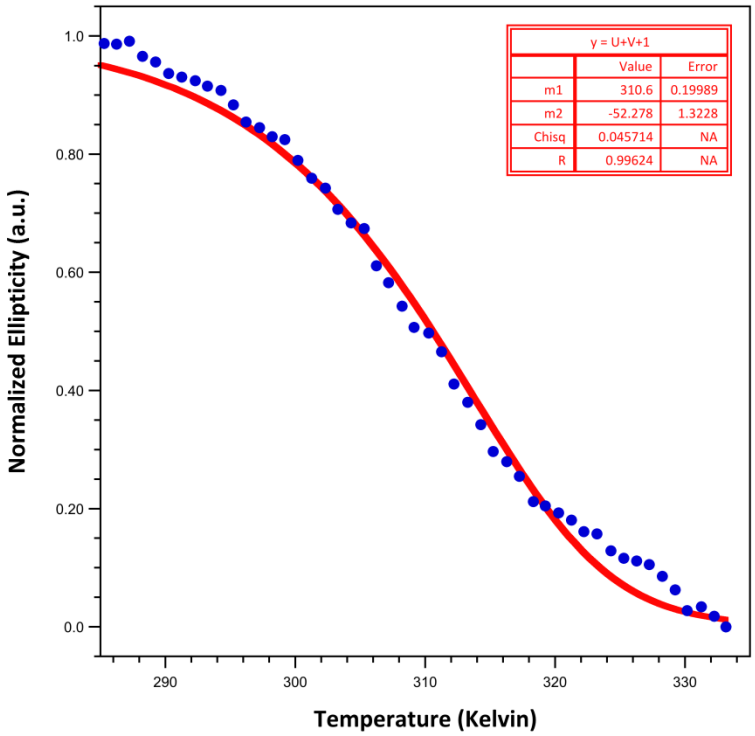

Figure 4.22: Melting study of **$(POG)_7$**; cooperative unfolding was indicated through the visible sigmoidal pattern.

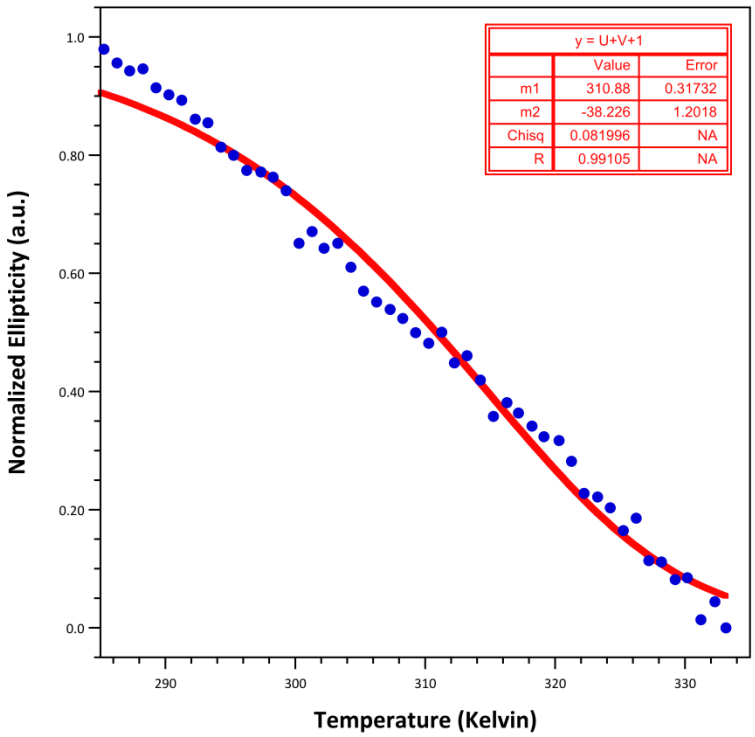

Figure 4.23: Melting study of **$(PXG)_4$**; sigmoidal fits could be obtained for the raw data but should be viewed with caution. It is inferred that these systems do not undergo cooperative melting transitions.

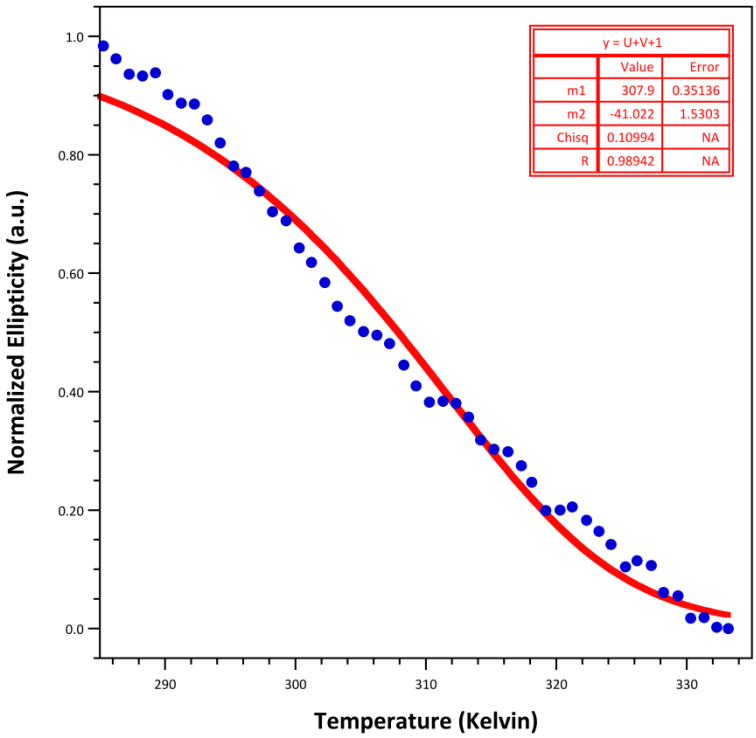

Figure 4.24: Melting studies of **$(PXG)_1$**; sigmoidal fits could be obtained for the raw data but should be viewed with caution. It is inferred that these systems do not undergo cooperative melting transitions.

Table 4.2: Thermal denaturation data for Map-bearing CRPs homotrimers by CD

Abbreviation	T_m (CD) ˚C	ΔH˚ (CD) kcal/mol
POG_7	36.2 (± 0.4)	-51.3 (± 0.9)
PXG_4	38.1 (± 0.3)	-40.6 (± 4.0)
PXG_1	n.d.	n.d.

The absence of sigmoidal behavior for **(PXG)$_4$** and **(PXG)$_1$** suggests that the ammonium charge and/or methyl groups prevent these peptides from associating into stable trimers.

*Melting Studies of **Map**-Bearing CRP Heterotrimers*

To determine whether intra-trimer charge repulsion was responsible for low stability heterotrimeric systems containing one or two **Map** units were prepared. Heterotrimer systems can potentially provide more insight into the influence of **Map** on trimer formation and thermal stability. Homotrimer samples of **(POG)$_7$** and **(PXG)$_4$** were combined to create 200 μM heterotrimer samples in PBS solutions. Each system was then heated to 75 ˚C for 25 minutes to ensure full trimer dissociation and then cooled to 4 ˚C overnight.

The conditions for the heterotrimer study are identical to the homotrimer study. Each system was heated from 10 ˚C at a 0.2 ˚C/min rate measuring the change in ellipticity at 224 nm. Measurements (Figure 4.26) were then fitted to a two-state model as described by previous studies.[37] All thermal denaturation experiments were performed in triplicate to ensure repeatability.

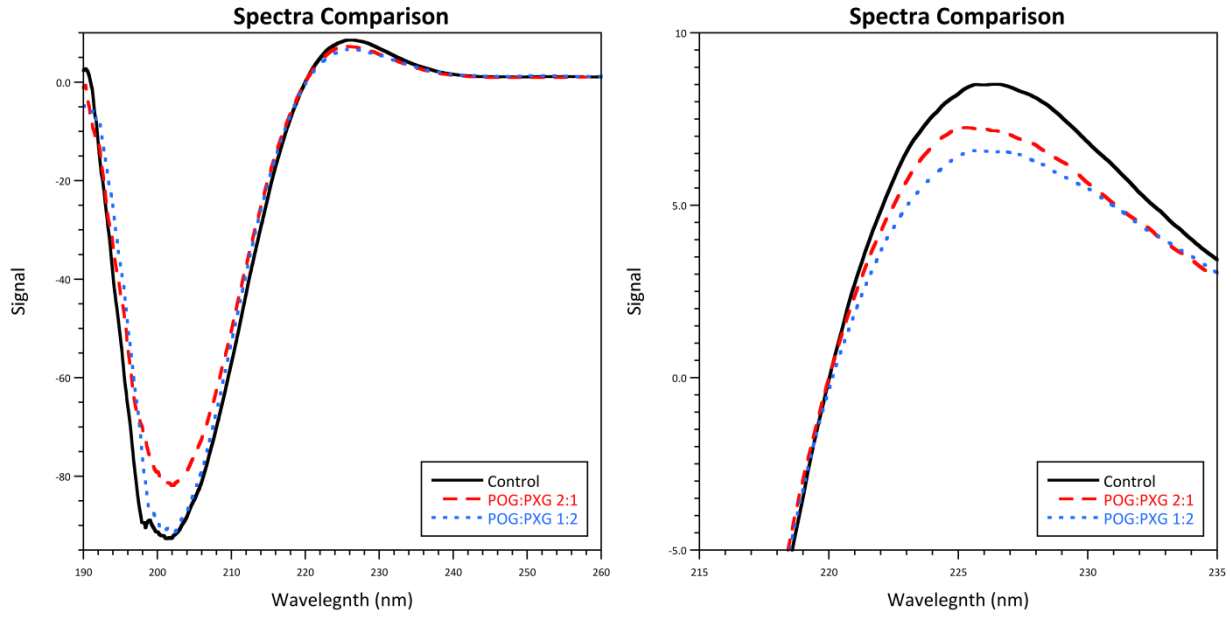

Figure 4.25: Comparisons among the CD spectra of **Map**-bearing CRPs heterotrimers

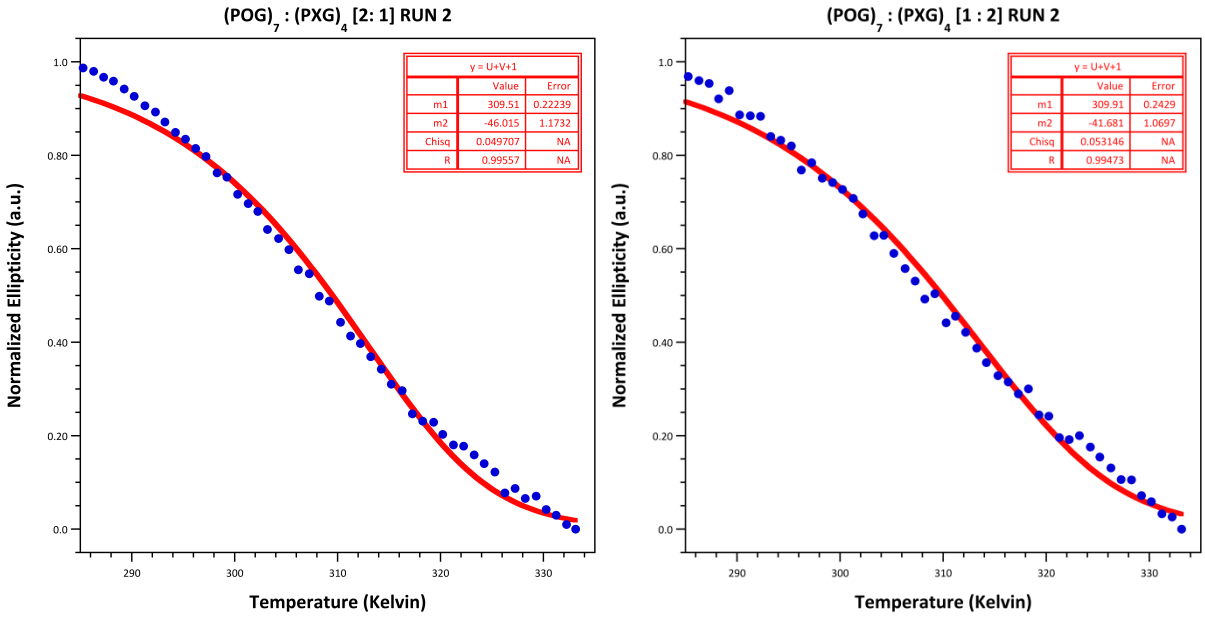

Figure 4.26: Comparisons among the melting studies of **Map**-bearing CRPs heterotrimers

Table 4.3: Thermal denaturation data for Map-bearing CRPs Heterotrimers by CD

Abbreviation	T_m (CD) ˚C	ΔH° (CD) kcal/mol	$\Theta_{224\ nm}$ (CD) mdeg
POG_7	36.2 (± 0.4)	-51.3 (± 0.9)	8.704
(2 : 1)	35.1 (± 0.9)	-46.4 (± 3.0)	7.237
(1 : 2)	36.1 (± 0.6)	-42.7 (± 0.8)	6.412
PXG_4	38.1 (± 0.3)	-40.6 (± 4.0)	4.072

4.5: Discussion on Thermodynamic Stabilities of Map-Bearing CRPs

*Discussion on the **Map**-bearing CRP homotrimers*

CD spectroscopy was used to probe the higher-order structures of **Map**-bearing CRPs. The appearance if an ellipticity maximum at 224 nm (Figure 4.25) suggests that all systems likely contain PPII helices, but not necessarily *triple* helices at 10 ˚C. Upon completing the melting study of each system, it was apparent that the incorporation of **Map** at the Yaa position

of either the 4th or 1st triplet of the CRPs was disruptive enough to prohibit trimer formation. The linearity in the melting profiles demonstrates a structure no more complex than PPII helices.

Previous work has demonstrated certain unnatural proline derivatives installed at the Yaa position prevent triple helix formation. A common feature of all these derivatives is the stereochemistry of substituents at the ɣ-position of the ring. Derivatives such as (2S,4S)-4-fluoroproline are isosteres of (2S,4R)-4-hydroxyproline, yet fully derail trimer formation.[24] As **Map** presents a methyl group on both "faces" of the 5-membered ring, perhaps it has similar behavior to (2S,4S)-4-fluoroproline, even though it lacks a stereocenter at that position. The source of this behavior would be a gauche effect within the triplet. This conformational effect largely impacts folding energetics in n → π* interactions and chain preorganization.[7] Notably, computational modeling of **Map** in an Ac-Pro-**Map**-Gly-NH_2 system predicts the dimethyl ammonium group does not alter the conformation of the ring, leaving it at a favorable C_γ-exo pucker at the Yaa position (Figure 4.27).

Another possible explanation into the disruption by **Map** is the introduction of a new hydration site within the ɣ-position of the ring. This quaternary ammonium likely invites the presence of nearby water molecules, as generally charged ends of amino acids interact with more water than the rest of the proteins.[47] However hydration through "water bridges" at this location can stabilize collagen mimics.[7]

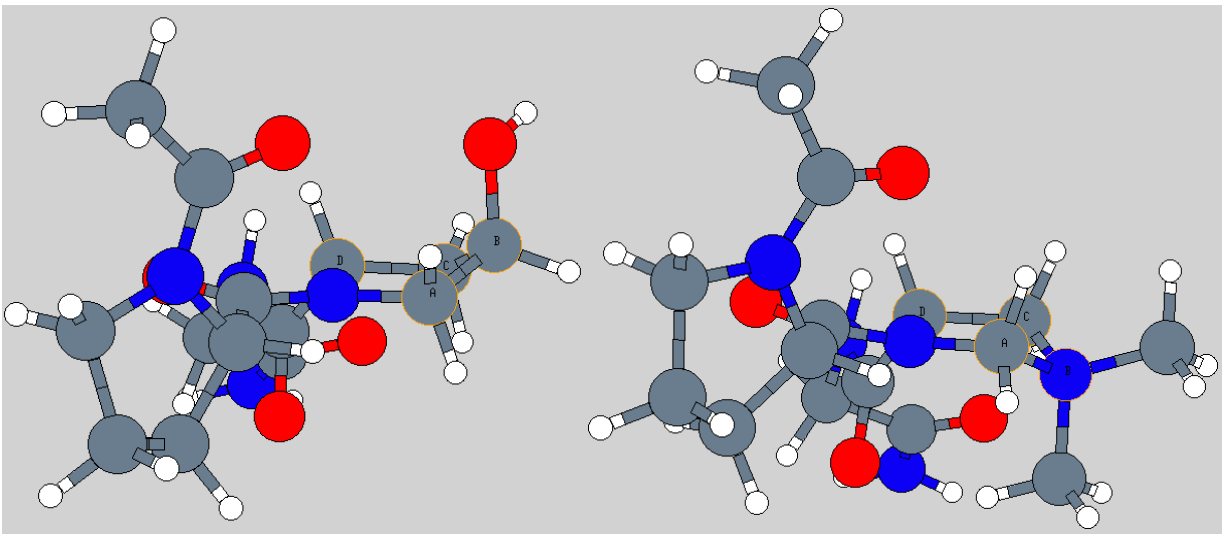

Figure 4.27: DFT calculations of Ac-Pro-Hyp-Gly-NH_2 (left) and Ac-Pro-**Map**-Gly-NH_2 (right) systems in a water continuum, performed by Andrew T. Sargent. Energetic values report indistinguishable differences in $K_{cis/trans}$ values and n → π* interactions.

*Discussion on the **Map**-bearing CRP heterotrimers*

This set of experiments was initially conducted to determine whether **Map**-bearing CRPs prioritize folding with the analogous neutral CRPs. Measuring the rate of unfolding can shed light on the relationship between the sigmodal behavior and the concentrations of **(POG)$_7$**. The **(2:1)** system contains an effective 133 μM of **(POG)$_7$** and consequently reduces the sigmoidal behavior along with lowering the maximum at 224 nm. The trend continues with decreasing the amount of **(POG)$_7$** to an effective 67 μM. This information suggests that an averaging of signals between homotrimeric **(POG)$_7$** and single-stranded **(PXG)$_4$** is occurring, indicative of both structures in the systems rather than a hybridization of the two.

The ratio between **(POG)$_7$** and **(PXG)$_4$** seems to correlate with the ΔH˚ values. It is interesting to note that the difference between each study accounts for enthalpic losses very similar to the loss of a single interstrand hydrogen bond (ΔH° = -3.0 kcal/mol)[7], though there

lacks a corresponding trend in T_m among the heterotrimeric systems. This relationship could be investigated further in CRPs with a larger number of triplets. Distortions in the hydrogen bonding network in the central position of a triple helix have demonstrated the greatest destabilizing affect than distortions in other triplets.[48] Additional triplets could manage the distortion in the center of the CRP and form helices under similar thermal stress.

CONCLUSIONS

This thesis describes various describe various approaches to develop a deeper physiochemical understanding of collagen. The synthetic designs into creating novel CRPs highlight the sheer complexity into the stability of this class of proteins. With our PyrATS systems, the pyrene probe seems to unreliably monitor the folding and unfolding behavior of collagenic molecules, yet these systems have revealed that the nucleation of the monomers can be influenced to occur away from the customary site at the C-termini. Our **Map**-bearing CRPs reinforce the deleterious consequence of perturbing the central triplet of collagen mimics.

A thematic similarity among the CRPs in this study and their thermodynamic and kinetic behaviors was found in their response to charge. When flanked by neutral amino acids, the bulky pyrene probes can dominate the folding stability and kinetics. However, this effect is somewhat mitigated when the probes are proximal to charged side chains. This deleterious relationship is further illuminated in **Map**-bearing CRPs when placing a local charge closer to the backbone of the peptides inhibit trimer formation. Consequently, both findings highlight the perpetual need for novel CRPs to comprehend the complexities and ubiquity of collagen.

REFERENCES

1. Schweitzer, M. H.; Suo, Z.; Avci, R.; Asara, J. M.; Allen, M. A.; Arce, F. T.; Horner, J. R. Analyses of Soft Tissue from Tyrannosaurus Rex Suggest the Presence of Protein. *Science* **2007**, *316* (5822), 277–280. https://doi.org/10.1126/science.1138709.

2. Asara, J. M.; Schweitzer, M. H.; Freimark, L. M.; Phillips, M.; Cantley, L. C. Protein Sequences from Mastodon and Tyrannosaurus Rex Revealed by Mass Spectrometry. *Science* **2007**, *316* (5822), 280–285. https://doi.org/10.1126/science.1137614.

3. Bern, M.; Phinney, B. S.; Goldberg, D. Reanalysis of Tyrannosaurus Rex Mass Spectra. *J. Proteome Res.* **2009**, *8* (9), 4328–4332. https://doi.org/10.1021/pr900349r.

4. Hill, R. C.; Wither, M. J.; Nemkov, T.; Barrett, A.; D'Alessandro, A.; Dzieciatkowska, M.; Hansen, K. C. Preserved Proteins from Extinct Bison Latifrons Identified by Tandem Mass Spectrometry; Hydroxylysine Glycosides Are a Common Feature of Ancient Collagen. *Molecular & Cellular Proteomics* **2015**, *14* (7), 1946–1958. https://doi.org/10.1074/mcp.M114.047787.

5. Kukkar, P. M.; Savkare, A. D.; Bhavsar M. R.; Gholap, V. D. A review on nanoparticle cross-linked collagen shield for sustained delivery of drug in glaucoma. *Int J Pharm Sci Res* **2017**, *8* (7), 2731-2739. https://doi.org/10.13040/IJPSR.0975-8232.8(7).2731-39.

6. Cejas, M. A.; Chen, C.; Kinney, W. A.; Maryanoff, B. E. Nanoparticles That Display Short Collagen-Related Peptides. Potent Stimulation of Human Platelet Aggregation by Triple Helical Motifs. *Bioconjug. Chem.* **2007**, *18* (4), 1025–1027. https://doi.org/10.1021/bc070105s.

7. Shoulders, M. D.; Raines, R. T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009**, *78*, 929–958. https://doi.org/10.1146/annurev.biochem.77.032207.120833.

8. Astbury, W.; Bell, F. Molecular Structure of the Collagen Fibres. *Nature* **1940,** *145*, 421–422 (1940). https://doi.org/10.1038/145421a

9. Ramachandran, G. N.; Kartha, G. Structure of Collagen. *Nature* **1954**, *174* (4423), 269–270. https://doi.org/10.1038/174269c0.

10. Fields G. B.; Prockop D. J. Perspectives on the synthesis and application of triple-helical, collagen model peptides. *Biopolymers* **1996** *40* (4), 345–57. https://doi.org/10.1002/(SICI)1097-0282(1996)40:4<345::AID-BIP1>3.0.CO;2-W.

11. Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. Crystal and Molecular Structure of a Collagen-Like Peptide at 1.9 Å Resolution. *Science* **1994**, *266* (5182), 75–81. https://doi.org/10.1126/science.7695699.

12. Bella, J.; Brodsky, B.; Berman, H. M. Hydration Structure of a Collagen Peptide. *Structure* **1995**, *3* (9), 893–906. https://doi.org/10.1016/S0969-2126(01)00224-6.

13. Jenkins, C. L.; Vasbinder, M. M.; Miller, S. J.; Raines, R. T. Peptide Bond Isosteres: Ester or (E)-Alkene in the Backbone of the Collagen Triple Helix. *Org. Lett.* **2005**, *7* (13), 2619–2622. https://doi.org/10.1021/ol050780m.

14. Boryskina, O. P.; Bolbukh, T. V.; Semenov, M. A.; Gasan, A. I.; Maleev, V. Ya. Energies of Peptide–Peptide and Peptide–Water Hydrogen Bonds in Collagen: Evidences from Infrared Spectroscopy, Quartz Piezogravimetry and Differential Scanning Calorimetry. *Journal of Molecular Structure* **2007**, *827* (1), 1–10. https://doi.org/10.1016/j.molstruc.2006.05.002.

15. Beck, K.; Chan, V. C.; Shenoy, N.; Kirkpatrick, A.; Ramshaw, J. A. M.; Brodsky, B. Destabilization of Osteogenesis Imperfecta Collagen-like Model Peptides Correlates with the Identity of the Residue Replacing Glycine. *PNAS* **2000**, *97* (8), 4273–4278. https://doi.org/10.1073/pnas.070050097.

16. Ramshaw, J. A.; Shah, N. K.; Brodsky, B. Gly-X-Y Tripeptide Frequencies in Collagen: A Context for Host-Guest Triple-Helical Peptides. *J. Struct. Biol.* **1998**, *122* (1–2), 86–91. https://doi.org/10.1006/jsbi.1998.3977.

17. Cram, D. J. The Design of Molecular Hosts, Guests, and Their Complexes. *Science* **1988**, *240* (4853), 760–767. https://doi.org/10.1126/science.3283937.

18. Kersteen, E. A.; Raines, R. T. Contribution of Tertiary Amides to the Conformational Stability of Collagen Triple Helices. *Biopolymers* **2001**, *59* (1), 24–28. https://doi.org/10.1002/1097-0282(200107)59:1<24::AID-BIP1002>3.0.CO;2-N.

19. Nan D.; Wang X. J.; Etzkorn F. A. The effect of a trans-locked Gly–Pro alkene isostere on collagen triple helix stability. *J. Am. Chem. Soc.* **2008,** *130* (16), 5396–5397. https://pubs.acs.org/doi/full/10.1021/ja711021m.

20. Inouye, K.; Sakakibara, S.; Prockop, D. J. Effects of the Stereo-Configuration of the Hydroxyl Group in 4-Hydroxyproline on the Triple-Helical Structures Formed by Homogenous Peptides Resembling Collagen. *Biochim. Biophys. Acta* **1976**, *420* (1), 133–141. https://doi.org/10.1016/0005-2795(76)90352-4.

21. Jiravanichanun, N.; Nishino, N.; Okuyama, K. Conformation of AlloHyp in the Y Position in the Host-Guest Peptide with the pro-pro-Gly Sequence: Implication of the Destabilization of (Pro-AlloHyp-Gly)10. *Biopolymers* **2006**, *81* (3), 225–233. https://doi.org/10.1002/bip.20405.

22. Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. Code for Collagen's Stability Deciphered. *Nature* **1998**, *392* (6677), 666–667. https://doi.org/10.1038/33573.

23. Holmgren, S. K.; Bretscher, L. E.; Taylor, K. M.; Raines, R. T. A Hyperstable Collagen Mimic. *Chem. Biol.* **1999**, *6* (2), 63–70. https://doi.org/10.1016/S1074-5521(99)80003-9.

24. Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. Conformational Stability of Collagen Relies on a Stereoelectronic Effect. *J. Am. Chem. Soc.* **2001**, *123* (4), 777–778. https://doi.org/10.1021/ja005542v.

25. Giacovazzo, C., Monaco, H.L., Artioli, G., Viterbo, D., Ferraris, G., Gilli, G., Zanotti, G., Catti, M. Molecules and molecular crystals. *Fundamentals of Crystallography (International Union of Crystallography Texts on Crystallography)*, 3rd; Oxford Science: United Kingdom, **2002**; 618-625.

26. Hinderaker, M. P.; Raines, R. T. An Electronic Effect on Protein Structure. *Protein Sci.* **2003**, *12* (6), 1188–1194. https://doi.org/10.1110/ps.0241903

27. Hodges, J. A.; Raines, R. T. Energetics of an n → Π* Interaction That Impacts Protein Structure. *Org. Lett.* **2006**, *8* (21), 4695–4697. https://doi.org/10.1021/ol061569t

28. Zhang, Y.; Malamakal, R. M.; Chenoweth, D. M. Aza-Glycine Induces Collagen Hyperstability. *J. Am. Chem. Soc.* **2015**, *137* (39), 12422–12425. https://doi.org/10.1021/jacs.5b04590

29. Zhang, Y.; Herling, M.; Chenoweth, D. M. General Solution for Stabilizing Triple Helical Collagen. *J. Am. Chem. Soc.* **2016**, *138* (31), 9751–9754. https://doi.org/10.1021/jacs.6b03823

30. Kasznel, A. J.; Zhang, Y.; Hai, Y.; Chenoweth, D. M. Structural Basis for Aza-Glycine Stabilization of Collagen. *J. Am. Chem. Soc.* **2017**, *139* (28), 9427–9430. https://doi.org/10.1021/jacs.7b03398

31. Aronoff, M. R.; Egli, J.; Menichelli, M.; Wennemers, H. γ-Azaproline Confers PH Responsiveness and Functionalizability on Collagen Triple Helices. *Angew. Chem. Int. Ed. Engl.* **2019**, *58* (10), 3143–3146. https://doi.org/10.1002/anie.201813048

32. Lindahl, F.; Hoang, H. N.; Fairlie, D. P.; Cooper, M. A. Facile Synthesis of Mono- and Bis-Methylated Fmoc-Dap, -Dab and -Orn Amino Acids. *Chem. Commun.* **2015**, *51* (21), 4496–4498. https://doi.org/10.1039/C4CC09780G

33. (a) Egli, J.; Siebler, C.; Köhler, M.; Zenobi, R.; Wennemers, H. Hydrophobic Moieties Bestow Fast-Folding and Hyperstability on Collagen Triple Helices. *J. Am. Chem. Soc.* **2019**, *141*, 5607-5611. https://doi.org/10.1021/jacs.8b13871
(b) Dehsorkhi, A.; Castelletto, V.; Hamley, I. W. Self-assembling amphiphilic peptides. *J. Pept. Sci.* **2014**, *20*, 453-467. https://doi.org/10.1002/psc.2633
(c) Versluis, F.; Marsden, H. R.; Kros, A. Power struggles in peptide-amphiphile nanostructures. *Chem. Soc. Rev.* **2010**, *39*, 3434-3444.

34. Martinez, C. R.; Iverson, B. L. Rethinking the Term "Pi-Stacking." *Chem. Sci.* **2012**, *3* (7), 2191–2201. https://doi.org/10.1039/C2SC20045G

35. Vargas-Uribe, M.; Rodnin, M. V.; Ojemalm, K.; Holgado, A.; Kyrychenko, A.; Nilsson, I.; Posokhov, Y. O.; Makhatadze, G.; von Heijne, G.; Ladokhin, A. S. Thermodynamics of Membrane Insertion and Refolding of the Diphtheria Toxin T-Domain. *J. Mem. Biol.* **2015***, 248*, 383-394. https://doi.org/10.1007/s00232-014-9734-0

36. (a) Sun, X.; Fan, J.; Hou, Y.; Liang, S.; Zhang, Y.; Xiao, J. Fluorescence characterization of the thermal stability of collagen mimic peptides. *Chin. Chem. Lett.* **2017**, *28*, 963-967. https://doi.org/10.1016/j.cclet.2016.11.029

(b) Sun, X.; Fan, J.; Zhang, S.; Liu, X.; Xiao, J. Colorimetric and fluorometric monitoring of the helix composition of collagen-like peptides at the nM level. *Chem. Commun.* **2016**, *52*, 3107-3110. https://doi.org/10.1039/C5CC09565D

37. (a) Erdmann, R. S.; Wennemers, H. Importance of Ring Puckering versus Interstrand Hydrogen Bonds for the Conformational Stability of Collagen. *Angew. Chem. Int. Ed.* **2011,** *50,* 6835-6838. https://doi.org/10.1002/anie.201008118
(b) Frank, S.; Kammerer, R. A.; Mechling, D.; Schulthess, T.; Landwehr, R.; Bann, J.; Guo, Y.; Lustig, A.; Bächinger, H. P.; Engel, J. Stabilization of Short Collagen-like Triple Helices by Protein Engineering. *J. Mol. Biol.* **2001,** *308,* 1081-1089. https://doi.org/10.1006/jmbi.2001.4644

38. Engel, J.; Chen, H.-T.; Prockop, D.J.; Klump, H. The triple helix ⇌ coil conversion of collagen-like polytripeptides in aqueous and nonaqueous solvents. Comparison of the thermodynamic parameters and the binding of water to $(L\text{-Pro-L-Pro-Gly})_n$ and $(L\text{-Pro-L-Hyp-Gly})_n$. *Biopolymers*. **1977,** *16*, 601-622. https://doi.org/10.1002/bip.1977.360160310

39. (a) Egli, J.; Erdmann, R. S.; Schmidt, P. J.; Wennemers, H. Effect of N- and C-terminal functional groups on the stability of collagen triple helices. *Chem. Commun.* **2017**, *53*, 11036-11039. https://doi.org/10.1039/C7CC05837C
(b) Venugopal, M. G.; Ramshaw, J. A. M.; Braswell, E.; Zhu, D.; Brodsky, B. Electrostatic Interactions in Collagen-like Triple-Helical Peptides. *Biochemistry* **1994**, *33*, 7948-7956. https://doi.org/10.1021/bi00191a023

40. (a) Chen, Y.-S.; Chen, C.-C.; Horng, J.-C. Thermodynamic and Kinetic Consequences of Substituting Glycine at Different Positions in a Pro-Hyp-Gly Repeat Collagen Model Peptide. *Biopolymers (Pept. Sci.)* **2011**, *96*, 60-68. https://doi.org/10.1002/bip.21470
(b) Bryan, M. A.; Cheng, H.; Brodsky, B. Sequence Environment of Mutation Affects Stability and Folding in Collagen Model Peptides of Osteogenesis Imperfecta. *Biopolymers (Pept. Sci.)* **2011**, *96*, 4-13. https://doi.org/10.1002/bip.21432
(c) Xiao, J.; Cheng, H.; Silva, T.; Baum, J.; Brodsky, B. Osteogenesis Imperfecta Missense Mutations in Collagen: Structural Consequences of a Glycine to Alanine Replacement at a Highly Charged Site. *Biochemistry* **2011**, *50*, 10771-10780. https://doi.org/10.1021/bi201476a

41. Boudko, S.; Frank, S.; Kammerer, R. A.; Stetefeld, J.; Schulthess, T.; Landwehr, R. Lustig, A.; Bächinger, H. P.; Engel, J. Nucleation and Propagation of the Collagen Triple Helix in

Single-chain and Trimerized Peptides: Transition from Third to First Order Kinetics. *J. Mol. Biol.* **2002**, *317*, 459-470. https://doi.org/10.1006/jmbi.2002.5439

42. Frank, S.; Boudko, S.; Mizuno, K.; Schulthess, T.; Engel, J.; Bächinger, H. P. Collagen Triple Helix Formation Can Be Nucleated at Either End. *J. Biol. Chem*. **2003**, *278*, 7747- 7750. https://doi.org/10.1074/jbc.C200698200

43. Smith, A.E.; Clapham, K.M.; Batsanov, A.S.; Bryce, M.R; Tarbit, B. (Dimethoxy- and Dihalopyridyl)boronic Acids and Highly Functionalized Heteroarylpyridines by Suzuki Cross-Coupling Reactions. *Eur. J. Org. Chem*. **2008**, 1458-1463. https://doi.org/10.1002/ejoc.200701156

44. Das, D.; Gerboth, D.; Postma, A.; Srinivasan, S.; Kern, H.; Chen, J.; Ratner, D. M.; Stayton, P. S.; Convertine, A. J. Synthesis of Zwitterionic, Hydrophobic, and Amphiphilic Polymers via RAFT Polymerization Induced Self-Assembly (PISA) in Acetic Acid. *Polym. Chem.* **2016**, *7* (39), 6133–6143. https://doi.org/10.1039/C6PY01172A

45. J. Egli, C. Siebler, B. Maryasin, R. S. Erdmann, C. Bergande, C. Ochsenfeld, H. Wennemers. pH-Responsive Aminoproline-Containing Collagen Triple Helices. *Chem. Eur. J.* **2017**, *23*, 7938. https://doi.org/10.1002/chem.201701134

46. Shoulders, M. D.; Hodges, J. A.; Raines, R. T. Reciprocity of Steric and Stereoelectronic Effects in the Collagen Triple Helix. *J. Am. Chem. Soc.* **2006**, *128* (25), 8112–8113. https://doi.org/10.1021/ja061793d

47. Biedermannová, L.; Schneider, B. Structure of the Ordered Hydration of Amino Acids in Proteins: Analysis of Crystal Structures. *Acta Crystallogr D Biol Crystallogr* **2015**, *71* (Pt 11), 2192–2202. https://doi.org/10.1107/S1399004715015679

48. Parmar, A.S.; Xu, F.; Pike, D.H.; Belure, S.V.; Hasan, N.F.; Drzewiecki, K.E.; Shreiber, D.I.; Nanda, V. Metal Stabilization of Collagen and de Novo Designed Mimetic Peptides. *Biochemistry* **2015**, *54*, 4987-4997. https://doi.org/10.1021/acs.biochem.5b00502